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

**Qian Liu**
Nanchang University
**Xiang Liu**
Nanchang University
**Xi Gong**
Nanchang University
**Bicai Guan** ( guanbicai@ncu.edu.cn )
Nanchang University

**Research Article**

**Keywords:** Sassafras tzumu, landscape genetics, environmental distance, genetic diversity, genetic differentiation

**Posted Date:** October 10th, 2023

**DOI:** https://doi.org/10.21203/rs.3.rs-3407218/v1

# 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

In this investigation, we employed genotyping-by-sequencing technology to scrutinize the genetic diversity and structure of 106 individuals of *S. tzumu*, carefully gathered from 13 naturally occurring habitats in China. The primary results could be summarized as follows: the genetic diversity of *S. tzumu* was inferior to those of most other woody plants. Moreover, *S. tzumu* exhibited a moderate level of genetic differentiation ($F_{ST}$=0.103), with the preponderance of genetic variation residing within populations (71%). Our findings unveiled that the 13 populations could be categorically classified into four distinct genetic clusters (K = 4). Notably, the populations from Mount Hengshan (HS) and Mount Shaoshan (SS) coalesced into a single genetic cluster, while the populations from Mount Lushan (LS) and Mount Meiling (ML) formed another genetic cluster, with the remaining two populations occupying unique genetic clusters. Utilizing the multiple matrix regression with randomization (MMRR) analysis method, we unveiled that genetic distance was concurrently influenced by both geographical distance and environmental distance ($r$ = 0.57, $p$ < .01). However, it was noteworthy that the regression coefficient of environmental distance was nearly threefold greater than that of geographical distance, thus underscoring the prominence of environmental distance in shaping genetic distance ($\beta_E$ = 0.46, $p$ < .01; $\beta_D$ = 0.16, $p$ < .01).

## Conclusions

S. tzumu had a moderate level of genetic differentiation and low genetic diversity in our study. 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*.


# 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 adept a species becomes in navigating its environment, thus bestowing it with a heightened advantage in terms of survival and evolutionary potential [3]. Factors such as the geographic 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* belongs to the Lauraceae genus *Sassafras*, which is a deciduous tree. *S. tzumu* is widely distributed among the south of the Yangtze River Basin and other southern areas in China, at an altitude of 100–1900 meters and often in sparse or dense forests [12]. 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 [13]. 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 [12]. Regrettably, due to its immense commercial value, certain regions had suffered from rampant deforestation and theft of *S. tzumu* specimens (Retrieved from Qianjiang Evening news, http://qjwb.thehour.cn/html/2017-07/18/content_3547987.htm?div=-1; Retrieved from Official Website of Jiande Procuratorate, 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* [14].

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 [13, 15, 16]. 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 [16]. In order to evaluate the genetic diversity and genetic structure of five *S. tzumu* populations at varying altitudes in Mount Tianmushan, 13 pairs of SSR primers were employed. This endeavor disclosed the distribution pattern of genetic variation across the diverse altitudinal populations [13]. 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 during the driest quarter, precipitation during the wettest month, temperature seasonality, and mean temperature during the 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 [17]. 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) [18]. 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 [19]. Notably, GBS offered the advantage of requiring fewer steps for database construction and enabled the establishment of databases for a large number of samples [20]. In certain genomic studies, the utilization of genotyping-by-sequencing (GBS) technology allowed for the elucidation of species' genetic diversity and hybridization characteristics [21].

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 [22, 23]. 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

### 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% ($p$<.01) among population, 19% ($p$<.01) among individual, and the primary variation in *S. tzumu* was attributed to differences within populations (71%, $p$<.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, $p$<.01). The inbreeding coefficient ($F_{IS}$) was calculated to be 0.215 ($p$<.01). The number of migrants ($N_m$) between the 13 populations exceeded 1

($N_{m=}$2.172; Table 3). This 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). By partitioning all groups into four distinct clusters, we were able to achieve optimal clustering of all populations together (Fig. 1c). The results of genetic structure analyses indicated that the populations fromed Hunan Province (SS and HS samples) separated from the southern ones (Jiangxi Province; LS and ML samples). A few samples from Mount Longchishan (LCS), Mount Tiantaishan (TTS), and Mount Guanzhaishan (GZS) clustered together. Additionally, a considerable number of samples (Mount Guanzhaishan (GZS), Mount Tianmushan (TMS), Mount Jiuhuashan (JHS), Mount Jigongshan (JGS), Mount Fuding (FD), Mount Maoshan (MS), Mount Wuyishan (WYS), Mount Tiantaishan (TTS), Mount Longchishan (LCS) also formed a cluster.

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 ($p$<.01). The results showed that SS and HS clustered, LS and ML clustered, LCS, GZS and TTS clustered, and others segregated from them (Fig. 2).

### 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 ($r$=0.56, $p$<.01; Fig. 3a). Additionally, a significant association between genetic and environmental distances was observed ($r$=0.56, $p$<.01; Fig. 3c). Within the 13 *S. tzumu* populations, the regression coefficient of environmental distance ($\beta_E$=0.46, $p$<.01) exceeded that of geographical distance ($\beta_D$=0.16, $p$<.01) by approximately threefold, implying that IBE predominantly accounted for the genetic distance. However, IBD also made a considerable contribution (Fig. 3b). Furthermore, the outcomes of MMRR analyses demonstrated a significant association between environmental distances and geographical distances ($r$=0.97, $p$<.01; Fig. 3d). 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

The greater the genetic variation of a species, the stronger its capacity to adapt to intricate and ever-changing environments [24]. Consequently, examining the genetic diversity of species is instrumental in analyzing their biological characteristics. In this study, the average values of $H_E$ and $H_O$ of *S. tzumu* were recorded as 0.469 and 0.515, respectively. Notably, the value of $H_O$ matched that of *Litsea auriculata*, a nationally protected plant of the Lauraceae family ($H_E$=0.515) [25], but fell short in comparison to other Lauraceae plants, such as *Ocotea rotundata* ($H_E$=0.76 and $H_O$=0.65) [26]. Furthermore, the values of $H_E$

and $H_O$ of 13 *S. tzumu* populations were inferior to those of other woody plants, including *Xanthoceras sorbifolium* ($H_E$=0.53 and $H_O$=0.72) and *Olea europaea* L. ($H_E$=0.60 and $H_O$=0.75) [27, 28].

The predominant distribution of genetic variation within the *S. tzumu* population was observed to be within populations (71%, $p < .01$), whereas the genetic variation among populations was minimal (10%, $p < .01$). These findings aligned with the outcomes of previous studies on perennial plants [29] and further supported the notion that the genetic differentiation of *S. tzumu* primarily occurred within populations (93.6%, $p < .001$) [13]. 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%) [30]. Furthermore, the overall genetic differentiation ($F_{ST}$) was calculated to be 0.103, suggesting a moderate level of differentiation. 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 may result in reduced genetic diversity and an increase in genetic structure within the 13 *S. tzumu* populations [31]. The number of migrants ($N_m$=2.172) indicated a moderate level of gene flow occurring among populations. The primary reproductive strategy of the Camphoraceae plant is out-crossing, a characteristic that is also observed in *S. tzumu*. *S. tzumu* possesses amphimerotic flowers, and its pollen is dispersed by insects, while birds aid in the dispersion of fruits. These two modes of transmission serve as the main sources of gene flow among populations. The transmission of this information occurs through the fruits, contributing to gene flow [15]. Based on the genetic structure analysis results of 13 populations of *S. tzumu*, *S. tzumu* could be divided into four most suitable clusters: cluster 1 was distributed in Hunan Province (HS and SS populations), cluster 2 was distributed in Jiangxi Province (LS and ML populations), cluster 3 was mostly located in the eastern part of the populations, and cluster 4 was partly from the eastern part of the populations (Fig. 1c). The PCA results also supported the division of cluster 1 and cluster 2 (Fig. 2). But a small number of these populations had already had some genetic exchange. The results of PCA also supported the situation. The composition of LS, ML, SS and HS populations was mostly similar, but there were differences, which may be related to the geographical boundary of Mount Luoxiaoshan, a mountain range bordering between Hunan Province and Jiangxi Province. In addition, most groups in the east were similar. This was related to the eastern hilly area and less high mountains. As a result, these groups exchanged genes to a large extent. Compared with the Hunan population, the Jiangxi population had partial gene exchange with most eastern populations, because the geographical distance between them was relatively close, and the eastern region was relatively flat, so gene flow did not become a major geographical barrier. However, the Hunan population was far away from the eastern population, and was blocked by mountains and the Yangtze River, so there was no similar gene exchange.

The Multiple Matrix Regression with Randomization (MMRR) analysis is a linear regression model that aims to quantify the impact of multiple explanatory variables, such as environmental factors and geographical distance, on genetic diversity. The results of MMRR analyses demonstrated that genetic

distance was influenced by both geographical distance and environmental distance ($r = 0.57$, $p < .01$; Fig. 3b). Notably, IBE ($r = 0.56$, $p < .01$; Fig. 3c) 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 [15], 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 [32]. Research indicated that relative humidity (accounting for 26.2% of permutation importance), average temperature during the driest quarter (16.6%), annual precipitation (12.6%), and mean diurnal temperature range (10.3%) were the primary factors influencing the distribution of *S. tzumu* in China [14]. Moreover, a portion of the genetic variation could be accounted for geographical distance ($r = 0.56$, $p < .01$; Fig. 3a). The *S. tzumu* populations in HS and SS, as well as LS and ML, exhibited gene exchange and were distinct from other populations due to geographic distance. This could explain the results of the previous genetic structure analysis and principal component analysis. But environmental distance of *S. tzumu* in this study had a greater and profounder impact on its genetic diversity than geographical distance ($\beta_E = 0.46$, $p < .01$; $\beta_D = 0.16$, $p < .01$; Fig. 3b), which could be attributed to its preference for warm and humid habitats based on its biological characteristics [33].

To effectively elucidate the genetic structure of *S. tzumu* populations, we screened and developed 11,862 SNP loci. The findings revealed that *S. tzumu* populations exhibited a moderate 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 scientifically informed 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 plant material samples from 13 naturally existing populations of *S. tzumu* across its range in China (Fig. 4). We made sure to collect individuals that were at least 50 meters apart, with a collection of 10-15 individuals from each population. Ultimately, we carefully selected a total of 106 samples of *S. tzumu* for analysis (Table 5). To maintain the plant material's integrity, we carefully dried the leaves in sealed plastic bags with silica gel and stored them in a freezer at -20°C.

### DNA extraction, library preparation, and sequencing

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 *EcoR*I-HF and *Msp*I 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 Illumina HiSeq platform.

### Single nucleotide polymorphisms (SNPs) calling

Sequence data obtained from the Hiseq Xten platform were processed for SNP identification using the GBS pipeline integrated within Stacks 3.0.166 [34]. In brief, variants were filtered using vcftools based on criteria such as a minimum depth of coverage (>5), quality score (>30), and an initial maximum missingness of 50%. 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 was converted into compatible formats using PGDspider 2.0.9.0 [35] Genetic differentiation ($F_{ST}$) was calculated for each SNP using GenAlEx 6.501 [36], 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 [37] was employed. The genetic structure analysis was conducted with three independent replicates, ranging from K=2 to K=13. Each run consisted of $5.0\times10^4$ burn-in iterations and $3.0\times10^4$ MCMC replications under the admixture model. The resulting output was submitted to the structure harvester website for further analysis. The optimal K value, indicative of the optimal number of clusters, was determined. 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 [38] and visualized using DISTRUCT [39] and Adobe Illustrator 2021. To explore the clustering patterns between populations, principal components analysis (PCA) was performed using GenAlEx 6.501 [36].

### The multiple matrix regression with randomization analysis

The geosphere package in R software [40] was utilized to compute the euclidean distance of the latitude and longitude information from 13 populations. This enabled the generation of a euclidean distance matrix, which depicted the distances between diverse populations. Additionally, MEGA 4 [41] was employed to determine the genetic distance among these populations. The value of the environmental layer in sampling locations was extracted from the matrix format utilizing Arcmap 10.8 [42]. The isolation

by distance (IBD) and isolation by environment (IBE) were quantified using the "MMRR" function in R software [43], with the response variable and geographical and environmental distances serving as explanatory variables. To elucidate 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.

# Declarations


### Acknowledgements

We thank the other members of our research group from Jiangxi Nanchang University for their assistance.


### Authors' contributions

Conceptualization, Q.L., X.L., X.G. and Bc.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, Q.L., X.L., X.G. and Bc. 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 National Natural Science Foundation of China (Grant No. 32160053) and the Natural Science Foundation of Jiangxi Province (Grant No.20212BAB205030).


### Data Availability

All data generated or analysed during this study are included in this article.

### Ethics approval and consent to participate

Not applicable to this study.

### Consent for publication

Not applicable to this study.

### Competing interests

The authors declare that there are no competing interests.

# References


1. 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.

2. 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;526:574-589.
3. Agashe D, Falk JJ, Bolnick DI. Effects of founding genetic variation on adaptation to a novel resource. Evolution. 2011;65:2481-2491.
4. 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.
5. Hamrick JL, Godt MJW. Allozyme diversity in plant species. Crop Sci. 1997;37:26-30.
6. Baker HG. Self-compatibility and establishment after “long-distance” dispersal. Evolution. 1955;9:347-349.
7. Grossenbacher D, Runquist RB, Goldberg EE, Brandvain Y. Geographic range size is predicted by plant mating system. Ecol. Lett. 2015;18:706-713.
8. Pannell JR. Evolution of the mating system in colonizing plants. Mol. Ecol. 2015;24:2018-2037.
9. Stebbins GL. Self fertilization and population variabilityin the higher plants. Am. Nat. 1957;91:337-354.
10. Loveless MD, Hamrick JL. Ecological determinants of genetic structure in plant populations. Annu Rev Ecol Evol Syst. 1984;15:65-95.
11. 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;23:376-384.
12. Li XW, Li J, van der Werff H. *Sassafras* J. Presl in Berchtold & J. Presl Prir. Rostlin 2(2): 30. 1825. Flora of China. Beijing, St. Louis: Science, Missouri Botanical Garden; 2008. p.159-161.
13. 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;27:1829-1836.
14. Zhang KL, Zhang Y, Jia DW, Tao J. Species distribution modeling of *Sassafras tzumu* and implications for forest management. Sustainability. 2020;12:4132.
15. 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.
16. Ding XF, Chen HL, Cao J, Koozo K, Masatosi U, Masanori O. Preliminary study on genetic structure of three natural populations in *Sassafrass Tzumu*. Hubei Forestry Science and Technology. 2006;1-2.
17. 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;36:320-326.
18. 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. Journal of Yunnan Agricultural University (Natural Science). 2022;37:285-293.
19. Poland JA, Rife TW. Genotyping-by-sequencing for plant breeding and genetics. The Plant Genome. 2012;5:92-102.

20. Rowe HC, Renaut S, Guggisberg A. RAD in the realm of next-generation sequencing technologies. Mol. Ecol. 2011;20:3499-3502.
21. 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.
22. 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;68:1143-1149.
23. Zhao Y, Li GY, Yang J. Conservation and utilization of wild relatives of cultivated plants. Biodiversity Science. 2018;26:414-426.
24. 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. Journal of Heredity. 2012;103:853-862.
25. Geng QF, Sun L, Zhang PH, Wang ZS, Qiu YX, Liu H, Lian CL. Understanding population structure and historical demography of *Litsea auriculata* (Lauraceae), an endangered species in east China. Sci. Rep. 2017;7:17343.
26. Marques I, Draper D, Riofrío L, Naranjo C. Early signs of the effects of forest fragmentation on the genetic diversity and structure of the threatened ecuadorian tree *Ocotea rotundata* (Lauraceae). Forests. 2022;13:1940.
27. 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.
28. 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;26:33-39.
29. 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;301:1055-1064.
30. 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;44:317-324.
31. Cole CT. Genetic variation in rare and common plants. Annu Rev Ecol Evol Syst. 2003;34:213-237.
32. 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.
33. 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;39:884-894.
34. Catchen JL, Hohenlohe PA, Bassham SS, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. Mol. Ecol. 2013;22:3124-3140.
35. Lischer HEL, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics. 2012;28:298-299.

36. Smouse PE, Whitehead MR, Peakall R. An informational diversity framework, illustrated with sexually deceptive orchids in early stages of speciation. Mol Ecol Resour. 2015;15:1375-1384.
37. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945-959.
38. 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;23:1801-1806.
39. Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes. 2003;4:137-138.
40. Karney CFF. Algorithms for geodesics. J Geod. 2013;87:43-55.
41. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 2007;24:1596-1599.
42. Brown JL. SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. Methods Ecol. Evol. 2014;5:694-700.
43. Legendre P, Lapointe FJ, Casgrain P. Modeling brain evolution from behavior: A permutational regression approach. Evolution. 1994;48:1487-1499.

# Tables

**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 | |

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

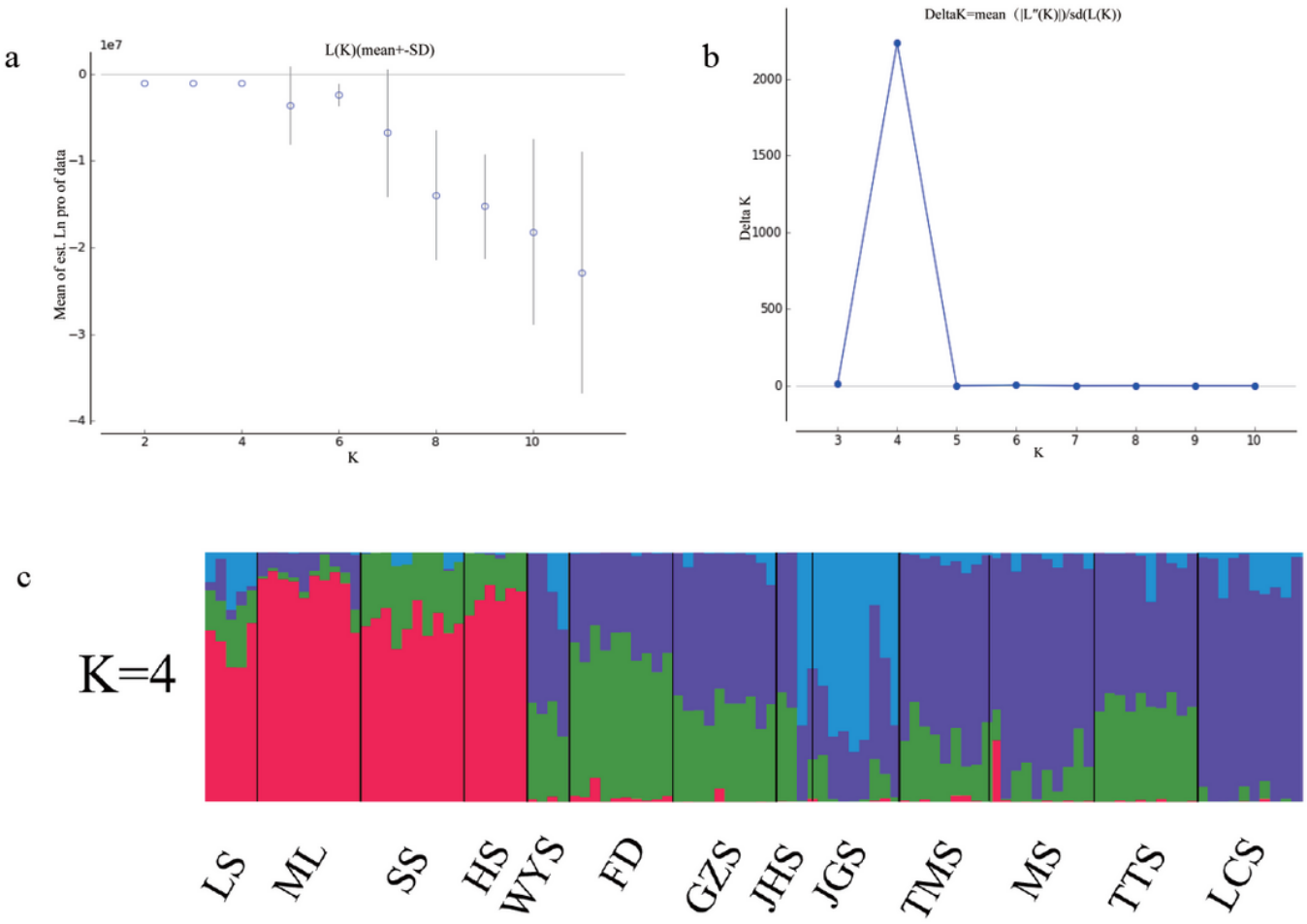
| file1 | file2 | *r*. squared | *p*-value |
|---|---|---|---|
| gen | geo | 0.58 | $< 2.2e^{-16}$ |
| gen | env | 0.56 | $9.831e^{-16}$ |
| gen | geo and env | 0.57 | $3.61e^{-16}$ |
| env | geo | 0.97 | $<2e^{-16}$ |

(Abbreviations: gen, genetic distance; geo, geographical distance; env, envrionmental distance)

**Table 5** The sampling locations of *Sassafras tzumu.*

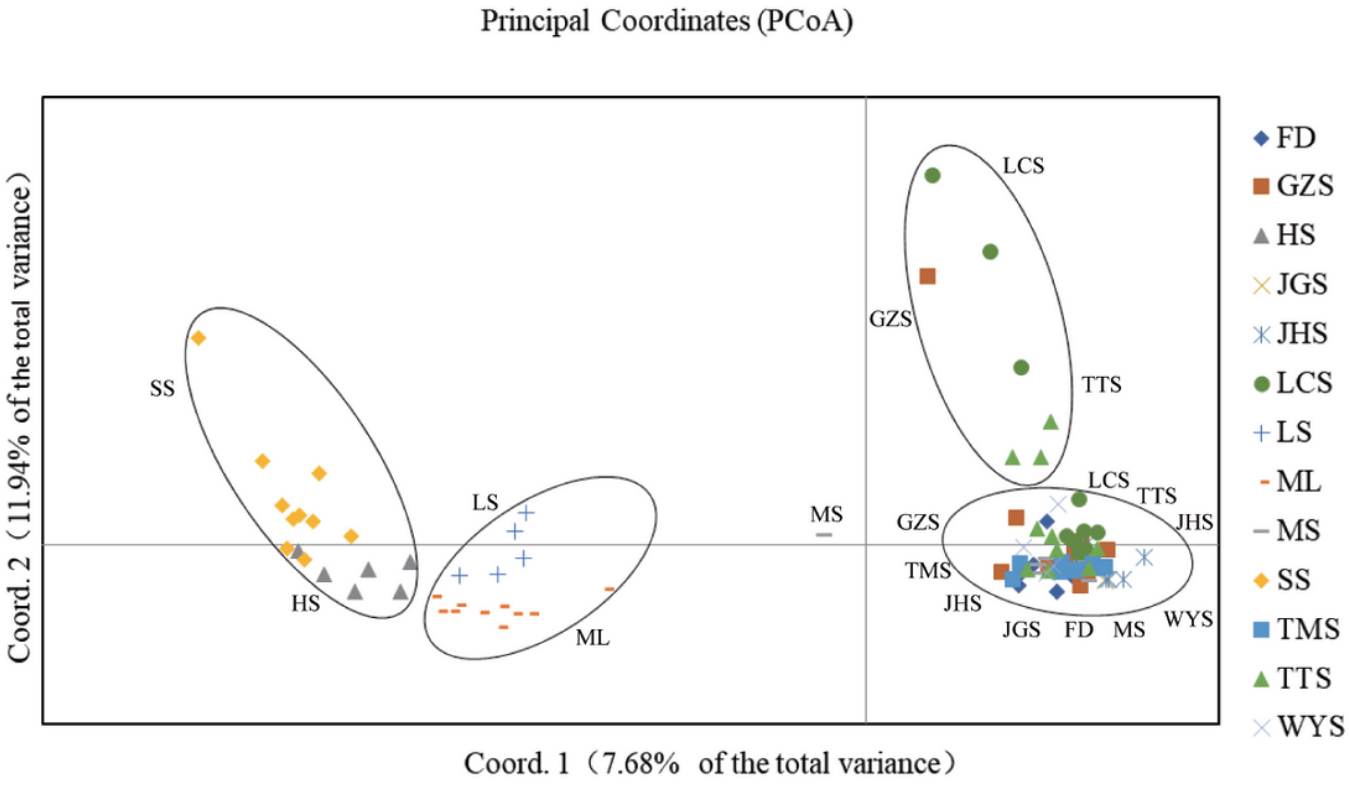
| Code | Longitude | Latitude | Number | Temperature | Precipitation | Altitude (m) |
|---|---|---|---|---|---|---|
| FD | 120.0398 | 27.13686 | 10 | 16.48 | 1686 | 497 |
| GZS | 116.7378 | 25.4839 | 10 | 19.03 | 1708 | 157 |
| HS | 112.7029 | 27.27414 | 6 | 16.03 | 1763 | 598 |
| JGS | 114.0766 | 31.81327 | 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 |

# Figures



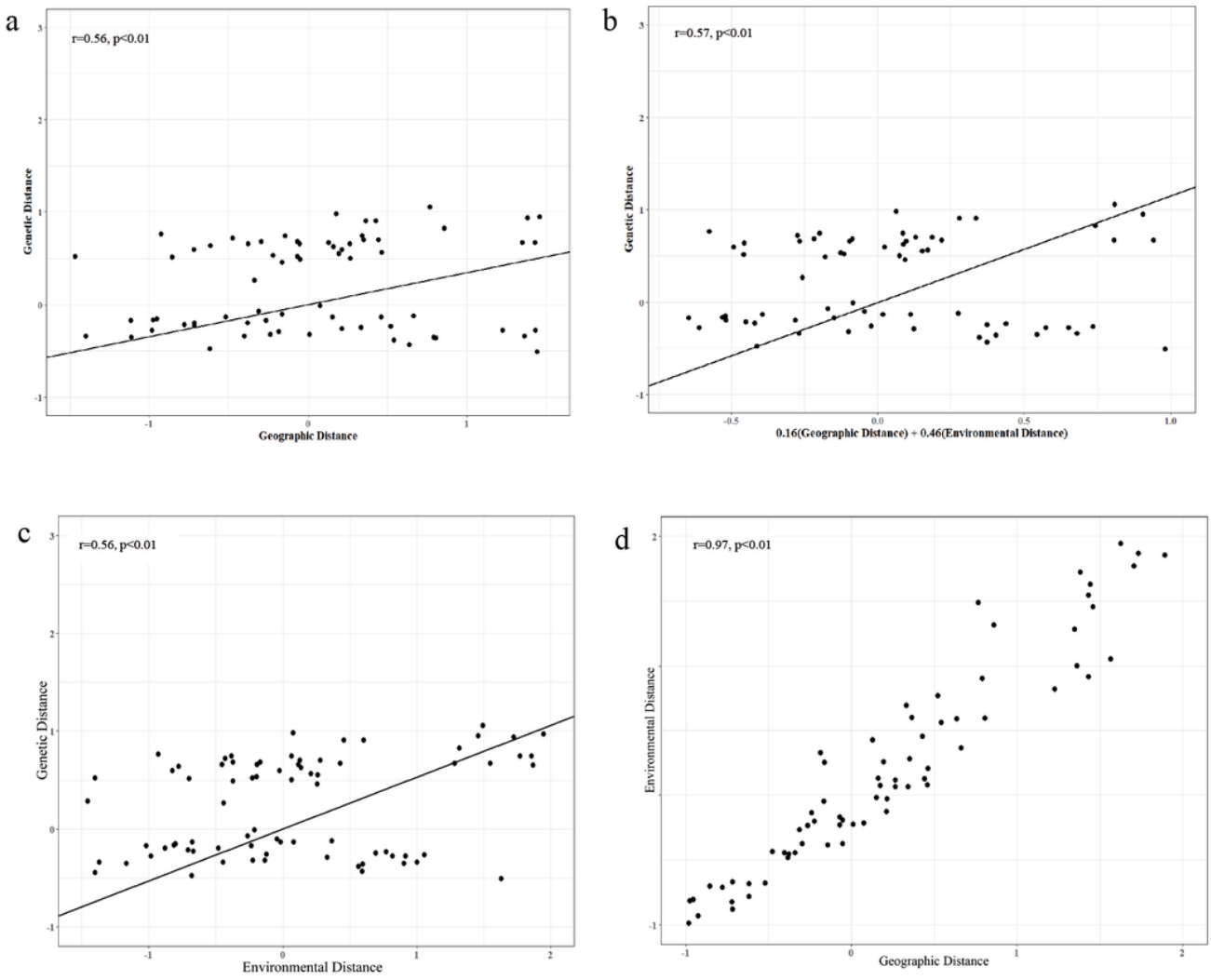

## Figure 1

Results of genetic structure analyses in *Sassafras tzumu* populations. (a) Relationship between mean LnK and K value. (b) Relationship between Delta K and K value. (c) Summaries of genetic structure analyses (K=4) on SNP data.



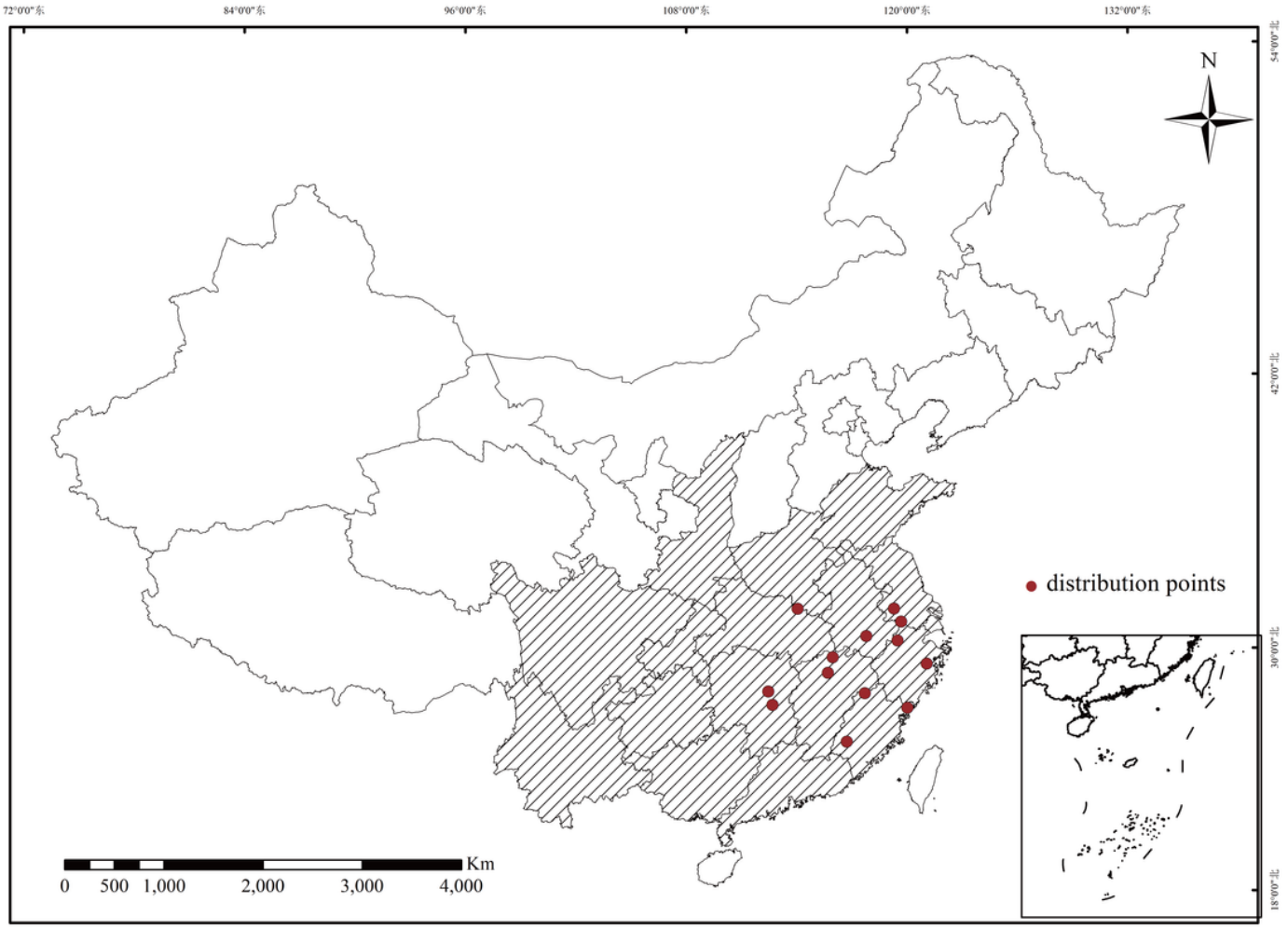

## Figure 2

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.




## Figure 3

(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.




## Figure 4

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).