

Assessing population genetic structure and diversity and their driving factors in Phoebe zhennan populations

Jian Peng

Sichuan Academy of Forestry

Jiaxin Xie

Sichuan Agricultural University

Yunjie Gu

Sichuan Academy of Forestry

Hongying Guo

Sichuan Academy of Forestry

Shuaiying Zhang

Sichuan Agricultural University

Xin Huang

Sichuan Agricultural University

Xiandan Luo

Sichuan Agricultural University

Jianghong Qian

Sichuan Agricultural University

Minhao Liu

Sichuan Academy of Forestry

Xueqin Wan

Sichuan Agricultural University

Lianghua Chen

Sichuan Agricultural University

Xiong Huang

Sichuan Agricultural University

Fan Zhang

Sichuan Agricultural University

Fang He

Sichuan Agricultural University

Peng Zhu

Sichuan Agricultural University

Yu Zhong

Sichuan Agricultural University
Hanbo Yang
yanghanbo6@sicau.edu.cn

Sichuan Agricultural University

Research Article

Keywords: population structure, genetic diversity, Phoebe zhennan, nSSR, cpSSR, genetic conservation

Posted Date: September 30th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-4798641/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Plant Biology on November 18th, 2024. See the published version at <https://doi.org/10.1186/s12870-024-05810-1>.

Abstract

Background

Phoebe zhennan, commonly known as “golden-thread nanmu,” is one of the most valuable and protection tree species in China. An accurate understanding of the population genetic structure and its environmental factors is of significance for the protection and selection of new *P. zhennan* varieties.

Results

Sixteen nSSR and six cpSSR markers were used to determine the genetic diversity and population structure of *P. zhennan* and the effect of environmental factors on the genetic structure. Low inbreeding ($F = 0.237$) and high genetic diversity ($I = 2.294$, $h = 0.412$) were observed in *P. zhennan* populations. The genetic differentiation was mainly detected within populations (only 28% and 13% of the variance being between populations according to the nSSR and cpSSR datasets). Among them, the HNSZX ($h = 0.469$) and SCSZZ ($I = 1.943$) populations exhibited the highest level of genetic diversity, while the HNXXT ($h = 0.041$) and SCLJS ($I = 0.943$) populations exhibited the lowest level of genetic diversity. The average genetic differentiation coefficient (F_{st}) and gene flow (Nm) were 0.022–0.128 and 1.698–11.373, respectively, which indicated a moderate level of genetic differentiation and a high level of gene flow. STRUCTURE, neighbor-joining clustering, and principal coordinate analysis divided 543 individuals into two or three categories based on the nSSR or cpSSR datasets. Four temperature, three precipitation, five chemical, five physical, and one soil texture variable showed significant effects on the genetic structure and distribution of *P. zhennan* populations. Compared to nSSR, the genetic differentiation among populations based on cpSSR datasets conformed to the geographic isolation model, suggesting that geographic and genetic distances should be considered for further genetic conservation and breeding utilization. The importance of *in situ* conservation units, such as populations with a high level of genetic diversity, more private alleles, and haplotypes (e.g., population SCGTS, SCYFS, and YNYJX), should be emphasized. Additionally, breeding, along with artificially assisted population regeneration and restoration, should also be carefully planned, taking into account climate and soil properties at the same time.

Conclusions

In conclusion, this study provided genetic background information for the genetic conservation, management, and utilization of *P. zhennan*.

Introduction

Phoebe zhennan is unique and valuable tree species in China, which mainly distributed in subtropical evergreen broad-leaved forests below 1500 m above sea level with warm and humid environment [1]. *P.*

zhennan is naturally distributed in Sichuan, Guizhou, Chongqing, Yunnan, Hunan, and Hubei province, China, of which Sichuan is the main distribution and production area [1, 2]. The *P. zhennan* is the main source of “golden-thread nanmu” and its wood has very high economic value with the characteristics of attractive visible golden thread pattern, unique fragrance, and high durability [2–4]. Due to these qualities, *P. zhennan* has been utilized in the production of furniture, various high-class furnishing materials, cosmetics, and health products before 221 BC [2, 3]. During the pre-Qin period, the distribution area of *P. zhennan* was mainly located at N28–35° and E103–121°, whereas the current distribution area has moved southward by more than 1° latitude based on the coldness of China’s historical climate and the southward movement of the subtropical north boundary. Due to the change of climate, indiscriminate deforestation, pilferage, and war damage, especially during the Ming and Qing Dynasties, the distribution range of *P. zhennan* natural population decreased sharply. As a result of habitat loss and logging, serious population isolation, and obvious habitat fragmentation, *P. zhennan* currently sporadically distributed in Sichuan basin and its surroundings [5, 6]. Although some ancient trees (more than 100 years old) are preserved at scenic spots, temples, and villages, however, these trees are often poorly maintained. Therefore, *P. zhennan* has been listed as endangered in the Chinese Plant Red Book because of the drastically decreased population [6]. Although there is an increase of the area of artificial cultivation due to its immense ecological and economic value, scant information is available on the population genetic structure. Thus, it is necessary to assess the genetic diversity, population structure, and the variation under environmental change scenarios. This information of genetic background will contribute to the development of protection strategy, breeding program, and genetic resource conservation.

The protection of the evolutionary potential of plant species, i.e., the ability to evolve and adapt to new environments, is changing, and understanding the amount and distribution of genetic variation will be key in developing conservation and utilization strategies. Genetic diversity is a key component of biodiversity and plays an important role in determining the plant origins, adaptability, and suitable distribution areas [7]. Analyzing genetic diversity and variation is the key to effectively and continuously protect and utilize plant genetic resources [8]. Genetic diversity can be evaluated using various approaches, including phenotypic characteristics, physicochemical indicators, as well as molecular markers [9]. Of which, molecular markers serve as a robust and dependable approach for conducting genetic analyses, enabling researchers to identify and examine genetic diversity directly at the DNA level [10]. By using these markers, scientists can accurately pinpoint differences in genetic sequences, which is essential for understanding the complexities of genetics. This method of molecular marker not only enhances the precision of genetic studies but also contributes significantly to the field of genomics by offering insights into genetic diversity. Of these, simple sequence repeat (SSR) markers with the characteristics of excellent repeatability, high polymorphism, and robust stability, are widely used in population genetic diversity analysis, breeding research, and kinship identification in plants, e.g., *P. shearerii* [11], *Ginkgo biloba* [12], *Juglans mandshurica* [13], *Vicia amoena* [14], and *Pandanus odorifer* [15]. Moreover, the SSR loci found in chloroplast DNA (cpSSR) offer benefits such as high conserved genomes, maternal transmission, and a degree of independent evolutionary progression. Proven to be

effective in examining genetic diversity, phylogeny, and evolution, cpSSR has emerged as a valuable and essential tool for plant species assessment [10, 16–18]. Currently, significant advancements have been achieved in research on the genetic diversity and population structure analysis, as well as in the classification of populations and their biogeographic distribution within the plant species through cpSSR markers, such as *Paeonia suffruticosa* [10], *Physalis angulate* [19], and Cupressaceae species (*Platycladus orientalis*, *Sabina chinensis*, *Juniperus formosana*, and *Cupressus torulosa*) [20]. Consequently, the combination of nSSR and cpSSR markers is preferred, given that nSSRs provide excellent genome coverage and exhibit significantly polymorphic patterns both within and between plant species [15, 21]. And, cpSSR markers exhibit more sequence variation compared to coding regions, and are distinguished by a slow evolutionary rate and a nearly absent recombination rate [20, 22].

As *P. zhennan* is considered vulnerable, it requires conservation. An previous investigation into genetic diversity, which utilized amplified fragment length polymorphisms (AFLPs) in *P. zhennan* (92 resources), revealed a low level of genetic diversity and notable divergence [23]. In contrast, high level of genetic diversity along with a moderate degree of genetic differentiation were detected within the *P. zhennan* breeding population (102 resources) in our previous study through SSR and phenotypic markers [1]. The results of restriction site-associated DNA sequencing (RAD-seq) indicated ring diversification of *P. zhennan* populations (72 individuals in 12 populations) with the barrier of Sichuan basin to gene flow [3]. However, the sample sizes of the populations and sample types limit the acquisition of comprehensive genetic information, and it remains uncertain if historical climate variations significantly influenced the genetic diversity. Therefore, much genetic information about natural *P. zhennan* populations has yet to be revealed. A better understanding of its population genetics is necessary to guide appropriate protective actions, sustainable management, and resource utilization. This study used nSSR and cpSSR markers to analyze the population genetic structure and genetic differentiation in 27 *P. zhennan* populations (543 individuals) across their entire distribution area and to speculate on the potential driving factors of the population structure. The purposes of this research are: (1) to determine the level of genetic diversity and population structure of *P. zhennan*; and (2) to identify the environmental factors that affect population structure.

Materials and methods

Plant materials and DNA isolation

Leaves were gathered from 27 distinct natural populations that span the distribution range of *P. zhennan* throughout China (Fig. 1, Table S1). Among these, the populations from Sichuan Province are the core distribution area, while those from Chongqing, Yunnan, and Guizhou, constitute the northwest regional distribution area. The populations of Hubei and Hunan Provinces constitute the southeastern regional distribution area. The distance between each population was at least 1.0 km, and the number of sampled individuals in each population ranged from 4 to 30, totaling 543 individuals. The 543 individuals were collected in the field. DNA was extracted from frozen leaves using a modified CTAB method, and the DNA quality and concentration was determined through electrophoresis on 0.8% agarose gel and with a

NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). DNA sample concentration was normalized for all samples.

nSSR and cpSSR genotyping

Sixteen nSSR and six cpSSR fluorescent-labeled primers were used to label and detect polymorphisms (Table S2) [1]. For PCR amplification, a 15- μ L reaction mixture was prepared containing 2.4 μ L of DNA (100 ng/ μ L), 0.3 μ L of primers, 4.5 μ L of ddH₂O, and 7.5 μ L of PCR mix (Takara, Dalian, China), was prepared. PCR was performed using a BIO-RAD T100 thermal cycler (Hercules, CA, USA) according to the procedure described by Zhu et al. (2022). PCR product separation was performed using an ABI 3500 DNA capillary electrophoresis analyzer from Applied Biosystems (Foster, CA, USA).

Genetic analysis

For cpSSR markers, GeneAEx software was used to calculate the genetic diversity parameters: the number of alleles (N_a), effective number of alleles (N_e), and Shannon's information index (I) [24]. The polymorphism information content (PIC) was calculated by PowerMarker ver. 3.25 [25]. The number of haplotypes (A), number of private haplotypes (P), effective number of haplotypes (N_{eh}), haplotypic richness (R_h), genetic diversity (H), mean genetic distance between individuals (D_{sh}^2), diversity due to differentiation (D_{st}), and interpopulation differentiation coefficient (F_{st}) were calculated by Haplotype Analysis ver. 1.05 [26]. The haplotype network was determined by Network 10 with the median-joining method. For nSSR markers, GenAEx ver. 6.501 software was used to calculate the genetic diversity parameters: N_a , N_e , I , observed heterozygosity (H_o), expected heterozygosity (H_e), and fixation index (F), and the PIC was calculated using Cervus version 3.0.7 [27].

Analysis of molecular variance (AMOVA) was used to estimate the distribution of genetic variation between and within populations. The F-statistics, including the population-level inbreeding coefficient (F_{is}), total population inbreeding coefficient (F_{it}), and F_{st} , and gene flow (Nm), were estimated to determine the actual proportion of genotypes in the population using GenAEx ver. 6.501.

Population structure

First, the nSSR and cpSSR data were converted into the STRUCTURE data format by the CONVERT 1.31 program [28]. Then, the population genetic structure was determined using the model-based program in STRUCTURE 2.3.4 software with a Bayesian approach [29]. The most likely number of populations (K) was identified as 1–10, and 10 interactions were performed for each K value. The length of the burn-in Markov chain Monte Carlo (MCMC) replications was set to 100,000, with 100,000 MCMC replications in each run. The optimal K capturing the major structure was determined using Structure Harvester [30], and the results were plotted using CLUMPP [31] and DISTRUCT [32].

Analysis of potential environmental variables' effect on genetic structure

Standard 19 bioclimatic variables were obtained from the WorldClim website (<https://www.worldclim.org/>) for the period from 1970 to 2000 at a spatial resolution of 30 s under the representative concentration pathway (RCP) 4.5 scenario. The soil datasets were provided by the National Cryosphere Desert Data Center. (<http://www.ncdc.ac.cn>). The variance inflation factor (VIF) was used to remove environmental and soil factors with correlation coefficients > 0.8 using the R package “usdm.” The following environmental factors were retained as predictors: mean diurnal range (bio2), temperature seasonality (bio4), maximum temperature of the warmest month (bio5), minimum temperature of the coldest month (bio6), precipitation during the wettest month (bio13), precipitation during the driest month (bio14), precipitation during the wettest quarter (bio16), precipitation during the warmest quarter (bio18), and precipitation during the coldest quarter (bio19). The following soil properties were retained as predictors: topsoil gravel content (T_GRAVEL), topsoil sand fraction (T_SAND), topsoil silt fraction (T_SILT), topsoil clay fraction (T_CLAY), topsoil organic carbon (T_OC), topsoil pH (H₂O) (T_PH_H2O), topsoil CEC (clay) (T_CEC_CLAY), topsoil CEC (soil) (T_CEC_SOIL), topsoil sodicity (ESP) (T_ESP), subsoil gravel content (S_GRAVEL), subsoil USDA texture (S_USDA_TEX), subsoil organic carbon (S_OC), and subsoil CEC soil (S_CEC_SOIL). The contribution of environmental and soil variables to the genetic structure of the population was determined by redundancy analysis (RDA).

Results

Characterization of cpSSR and nSSR markers

The number of alleles (N_a) of cpSSR locus ranged from 2 to 5, with an average of 1.8 (Table 1). The effective number of alleles (N_e) varied from 1.538 (cpSSR4) to 2.002 (cpSSR10), with an average of 1.716. The average value of shannon’s information index (I) and genetic diversity (H), which reflecting the level of genetic diversity, were 0.671 and 0.412, respectively. Of which, cpSSR2 had the highest I value (0.903), cpSSR10 had the highest H value (0.500), while the lowest values of I and H were recorded at 0.567 for cpSSR12 and 0.350 for cpSSR4 and cpSSR18. The average value of PIC was 0.344, which varied from 0.296 to 0.427. A total of 451 N_a was detected by 16 pairs of nSSR markers (Table 2). The N_a and N_e ranged from 13 to 55 and 2.481 to 23.823 with the average value of 28.188 and 8.540, respectively. The observed heterozygosity (H_o) and expected heterozygosity (H_e) were 0.624 (0.369–0.861) and 0.821 (0.597–0.958), respectively. I averaged 2.294, with a minimum at SSR17 (1.379) and a maximum at SSR16 (3.471). The average F value was 0.237, with a minimum at site SSR3 (– 0.090) and a maximum at SSR2 (0.573). In addition, the 16 nSSR locus showed high informative scores with an average value of 0.803 (0.577–0.957 > 0.500).

Table 1
Characteristics of six cpSSR markers based on genotyping results of natural *Phoebe zhennan* populations

cpSSR marker	<i>Na</i>	<i>Ne</i>	<i>I</i>	<i>H</i>	<i>PIC</i>
cpSSR2	5	1.845	0.903	0.458	0.427
cpSSR4	5	1.538	0.587	0.350	0.296
cpSSR7	2	1.764	0.625	0.433	0.339
cpSSR10	3	2.002	0.715	0.500	0.380
cpSSR12	2	1.610	0.567	0.379	0.307
cpSSR18	3	1.539	0.631	0.350	0.314
Mean	3	1.716	0.671	0.412	0.344
Note: <i>Na</i> : Number of different alleles; <i>Ne</i> : Number of effective alleles; <i>I</i> : Shannon's information index; <i>H</i> : Genetic diversity; <i>PIC</i> : Polymorphism information content. The same below.					

Table 2
Characteristics of nSSR markers based on genotyping results from natural *Phoebe zhennan* populations

nSSR marker	<i>Na</i>	<i>Ne</i>	<i>I</i>	<i>Ho</i>	<i>He</i>	<i>F</i>	<i>PIC</i>
MN-g3	49	17.899	3.213	0.798	0.944	0.155	0.942
MN-g5	39	10.533	2.782	0.388	0.905	0.571	0.899
MN-g18	21	11.689	2.633	0.861	0.914	0.059	0.909
MN-g30	54	13.954	3.141	0.808	0.928	0.129	0.925
CL20730Contig1	21	3.915	1.722	0.483	0.745	0.351	0.708
Unigene29601	13	4.104	1.755	0.449	0.756	0.407	0.729
MN-e96	16	2.481	1.389	0.444	0.597	0.257	0.577
CL4747Contig1	18	3.986	1.858	0.559	0.749	0.254	0.722
SSR17	14	2.837	1.379	0.595	0.648	0.080	0.607
SSR2	20	7.355	2.269	0.369	0.864	0.573	0.851
SSR3	25	3.960	1.759	0.815	0.748	-0.090	0.714
SSR1	20	6.161	2.156	0.579	0.838	0.309	0.822
SSR19	26	3.749	1.851	0.612	0.733	0.165	0.701
SSR20	40	10.616	2.813	0.566	0.906	0.376	0.900
SSR21	20	9.578	2.512	0.820	0.896	0.084	0.888
SSR16	55	23.823	3.471	0.846	0.958	0.117	0.957
Mean	28.188	8.540	2.294	0.624	0.821	0.237	0.803
Note: <i>Ho</i> : Observed heterozygosity; <i>He</i> : Expected heterozygosity; <i>F</i> : Fixation index. The same below.							

Genetic diversity parameters among populations

The *Na* and *Ne* of cpSSR in 27 populations varied from 1.167 to 2.667 and 1.054 to 1.974 (Table 3, Table S3). The highest level of diversity exhibited in YNYJX (*I* = 0.729, *H* = 0.456) and HNSZX (*I* = 0.722, *H* = 0.469). The *Na* and *Ne* of nSSR in 27 populations varied from 3.438 (GZCSS) to 10.688 (SCSZZ) and from 2.399 (SCLJS) to 6.377 (SCSZZ), respectively (Table 4, Table S4). The minimum and maximum *Ho* of 0.532 and 0.744 were obtained for SCLYS and GZCSS populations, respectively. The lowest and highest *He* were obtained for populations SCXFS (0.490) and SCSZZ (0.802), respectively. The *F* value was greater than 0, except in four populations (CQWGS, SCLJS, GZCSS, and HNXXT), indicating that there was a loss of heterozygosity in these populations. In contrast to the cpSSR results, the highest level of genetic diversity exhibited in SCSZZ with the highest *I* value of 1.943.

Table 3
Characteristics of cpSSR markers in populations

Populations	N	<i>Na</i>	<i>Ne</i>	<i>I</i>	<i>H</i>
SCXFS	22	2.000	1.503	0.458	0.291
SCGTS	21	2.000	1.570	0.453	0.283
SCLYS	22	1.833	1.272	0.315	0.187
SCPZS	26	2.000	1.614	0.482	0.314
SCJFS	21	2.000	1.721	0.521	0.339
SCYFS	24	2.167	1.583	0.505	0.311
CQWGS	22	2.167	1.640	0.535	0.340
CQYCQ	22	2.167	1.340	0.383	0.223
SCBZL	22	1.833	1.517	0.436	0.284
SCJLJL	25	1.667	1.282	0.289	0.183
SCLJS	4	1.167	1.167	0.116	0.083
SCSSS	17	2.000	1.668	0.520	0.344
SCXYX	15	1.333	1.299	0.221	0.157
GZTZX	14	1.167	1.142	0.109	0.077
GZCSS	10	1.333	1.073	0.108	0.060
SCSZZ	16	2.333	1.775	0.644	0.413
YNYJX	20	2.667	1.897	0.729	0.456
HNSZX	7	2.333	1.974	0.722	0.469
HNHYZ	18	2.333	1.632	0.546	0.341
HNYLX	12	2.333	1.547	0.526	0.317
HNJDX	11	2.000	1.464	0.450	0.281
HNYSX	11	1.500	1.210	0.211	0.132
HNXXT	7	1.167	1.054	0.068	0.041
HNAHX	7	1.333	1.108	0.137	0.082
HBLFX	20	1.667	1.383	0.329	0.216
GZQDN	6	1.333	1.128	0.150	0.093
GZSNX	7	1.667	1.497	0.373	0.245

Table 4
Characteristics of nSSR markers in populations

Populations	<i>Na</i>	<i>Ne</i>	<i>I</i>	<i>Ho</i>	<i>He</i>	<i>F</i>
SCXFS	5.813	2.604	1.015	0.593	0.490	0.026
SCGTS	8.313	4.193	1.443	0.558	0.619	0.172
SCLYS	5.625	3.298	1.192	0.532	0.598	0.194
SCPZS	6.125	3.090	1.234	0.613	0.616	0.006
SCJFS	6.375	2.999	1.284	0.549	0.638	0.164
SCYFS	10.125	4.991	1.656	0.676	0.698	0.031
CQWGS	9.063	4.431	1.578	0.716	0.693	-0.049
CQYCQ	9.688	5.457	1.788	0.640	0.760	0.144
SCBZL	8.750	3.973	1.571	0.561	0.707	0.196
SCJJL	10.625	5.063	1.832	0.651	0.767	0.144
SCLJS	3.188	2.399	0.943	0.692	0.547	-0.189
SCSSS	9.750	5.532	1.802	0.661	0.762	0.136
SCXYX	6.750	4.452	1.507	0.601	0.697	0.137
GZTZX	5.188	2.858	1.197	0.605	0.608	0.011
GZCSS	3.438	2.686	1.022	0.744	0.588	-0.250
SCSZZ	10.688	6.377	1.943	0.650	0.802	0.190
YNYJX	8.688	3.994	1.574	0.569	0.711	0.191
HNSZX	4.625	3.332	1.276	0.632	0.659	0.040
HNHYZ	10.188	5.048	1.828	0.648	0.774	0.154
HNYLX	5.000	3.350	1.287	0.656	0.661	0.023
HNJDX	5.938	3.707	1.367	0.630	0.650	0.029
HNYSX	7.438	5.306	1.686	0.707	0.751	0.061
HNXXT	4.375	3.429	1.150	0.607	0.574	-0.092
HNAHX	5.188	3.822	1.371	0.606	0.679	0.100
HBLFX	6.500	3.413	1.329	0.642	0.648	0.014
GZQDN	4.000	3.165	1.149	0.617	0.608	0.003
GZSNX	5.375	4.086	1.452	0.708	0.710	0.015

Chloroplast DNA haplotype variation

Seventy-eight haplotypes (2–15 haplotypes per population) were detected in 27 population, while 35 private haplotypes inhibited across populations (Table 5). Among the studied populations, the YNYJX population had the most private haplotypes. N_{eh} and R_h ranged from 1.324 to 14.222 and 0.571 to 2.950, respectively, with averages of 6.023 and 2.076, respectively. The H value varied from 0.286 (HNXXT) to 0.992 (SCSZZ), and D^2_{sh} exhibited more variability, ranging from 4.762 (HNXXT) to 284.127 (HNSZX), with a mean value of 86.387. The majority of haplotypes were identified in a single individual, leading to a more uniform distribution across the populations (Fig. 2). Especially, there were six haplotypes (haplo-52, haplo-64, haplo-67, haplo-68, haplo-69, and haplo-70) shared by more than 20 individuals, and 32 haplotypes were present in a single individual. There were 32 haplotypes contained in only one sample and 13 haplotypes contained in two samples (Fig. 2, 3). Haplo-64 was found in the most samples, with 51 samples from 14 populations, ranging from 1 (SCJFS, SCBZL, SCJLJ, and HNHYZ) to 9 samples (GZTZX). Haplo-67 ranked second, with 42 samples from 11 populations, ranging from 1 (SCPZS and CQWGS) to 9 (SCJLJ).

Table 5
Haplotype diversity in 27 natural *Phoebe zhennan* populations based on six cpSSR markers

Population	<i>A</i>	<i>P</i>	<i>N_{eh}</i>	<i>R_h</i>	<i>D²_{sh}</i>
SCXFS	13	2	6.914	2.463	85.786
SCGTS	11	1	8.647	2.589	101.206
SCLYS	8	0	3.408	1.828	63.276
SCPZS	14	0	9.941	2.632	118.205
SCJFS	14	0	10.756	2.726	137.778
SCYFS	14	2	9.000	2.603	110.326
CQWGS	14	3	9.308	2.638	153.535
CQYCQ	8	0	4.321	2.042	62.987
SCBZL	14	0	6.050	2.383	116.378
SCJJL	7	0	4.085	1.921	54.111
SCLJS	2	0	2.000	1.000	11.111
SCSSS	13	1	10.704	2.785	106.618
SCXYX	4	0	3.571	1.806	24.762
GZTZX	2	0	1.849	0.869	8.242
GZCSS	3	0	1.515	0.800	18.148
SCSZZ	15	7	14.222	2.950	136.410
YNYJX	15	8	10.526	2.733	238.947
HNSZX	6	5	5.444	2.714	284.127
HNHYZ	12	3	9.000	2.662	127.233
HNXLX	10	1	9.000	2.818	113.131
HNJDX	8	1	5.261	2.439	77.576
HNYSX	4	0	2.574	1.500	21.818
HNXXT	2	0	1.324	0.571	4.762
HNAHX	3	0	1.815	1.143	11.111

Note: *A*: Number of haplotypes; *P*: Number of private haplotypes; *N_{eh}*: Effective number of haplotypes; *R_h*: Haplotypic richness; *D²_{sh}*: Mean genetic distance between individuals.

Population	<i>A</i>	<i>P</i>	<i>N_{eh}</i>	<i>R_h</i>	<i>D²_{sh}</i>
HBLFX	7	1	4.545	2.064	54.386
GZQDN	2	0	1.385	0.667	22.222
GZSNX	6	0	5.444	2.714	68.254
Note: <i>A</i> : Number of haplotypes; <i>P</i> : Number of private haplotypes; <i>N_{eh}</i> : Effective number of haplotypes; <i>R_h</i> : Haplotypic richness; <i>D²_{sh}</i> : Mean genetic distance between individuals.					

Private alleles

To examine the genetic variation among populations, we detected locally common alleles (frequency $\geq 5\%$) in 25% or fewer of the populations, locally common alleles (frequency $\geq 5\%$) in 50% or fewer of the populations, and private alleles which identified in just one population (Fig. 4). The results showed that most locally common alleles (frequency $\geq 5\%$) were presented in 25% or fewer populations. A total of 77 private alleles were detected in 15 nSSR loci (Table S7). Nine populations, of which six populations were from Sichuan (including SCXFS, SCGTS, and SCLYS), two populations were from Chongqing (CQWGS and CQYCQ), and one population was from Hunan Province (HNYLX), contained private alleles. The SCYFS population contained the most private alleles (1.063), followed by CQYCQ (0.813), SCJFS (0.688), and SCPZS (0.688).

Genetic differentiation of *P. zhennan* natural populations

Based on the cpSSR data, the 28% of genetic variation existed among the populations, while 72% of the genetic variation was detected within populations, indicating that the genetic variation mainly occurred among individuals within the population in *P. zhennan* populations (Table 6). A similar result was obtained from the nSSR data. Genetic differentiation mainly took place among individuals within the population, and only 13% of the genetic variation occurred among populations. Interpopulation differentiation coefficient (*F_{st}*) analysis showed that 12.8% (nSSR) and 2.2% (cpSSR) of genetic variation occurred among populations. The inbreeding coefficient was predicted to be moderate (*F_{is}* = 0.230). Diversity due to differentiation was also predicted to be moderate (*D_{st}* = 0.217). The gene flow (*Nm*) derived from the nSSR and cpSSR data was 1.698 and 11.373, respectively, suggesting relatively high gene flow among *P. zhennan* populations.

Table 6
Analysis of molecular variance (AMOVA) within and among populations

	Source of variation	<i>df</i>	<i>SS</i>	<i>MS</i>	Variance component	Percentage of variance (%)
cpSSR	Among populations	26	22840.061	878.464	47.833	28
	Within populations	402	49661.626	123.536	123.536	72
	Total	428	72501.688		171.370	100
	<i>Fst</i>	0.022				
	<i>Dst</i>	0.217				
	<i>Nm</i>	11.373				
nSSR	Among populations	26	1089.683	41.911	0.870	13
	Among individuals within populations	516	3748.195	7.264	1.358	20
	Within individuals	543	2470.000	4.549	4.549	67
	Total	1085	7307.878		6.776	100
	<i>Fst</i>	0.128				
	<i>Fis</i>	0.230				
	<i>Fit</i>	0.329				
	<i>Nm</i>	1.698				
	Note: <i>df</i> : Degrees of freedom; <i>SS</i> : Sum of squares; <i>MS</i> : Mean square; <i>Dst</i> : Diversity due to differentiation; <i>Fst</i> : Interpopulation differentiation coefficient; <i>Fis</i> : Inbreeding coefficient at the population level; <i>Fit</i> : Inbreeding coefficient of total populations; <i>Nm</i> : Gene flow.					

Genetic relationship among populations

Principal coordinate analysis (PCoA) revealed obvious genetic variation among *P. zhennan* populations (Fig. 5A, C). Similar to PCoA, the Nei's genetic distance between 27 populations ranged from 0.005 (SCBZL and YNYJX) to 0.828 (SCLJS and GZTZX) and from 0.003 (HNXXT and HNAHX) to 0.971 (SCJFS and HNXXT) based on nSSR and cpSSR markers, respectively (Tables S5 and S6). Based on UPGMA analysis, all populations were separated to two or three categories base on the Nei's genetic distance using nSSRs and cpSSRs (Fig. 5B, D). For UPGMA clustering of nSSRs, the first major category included 19 populations (including SCJJL, HNHYZ, CQWGS, YNYJX, HBLFX, and GZSNX) from six provinces. The second category comprised six populations (SCXFS, SCGTS, SCJFS, SCYFS, SCLYS, and SCPZS), which

were all from Sichuan Province. SCLJS and HNYLX were in the third category, although there was a long geographical distance between them. For the UPGMA cluster of cpSSR, the members of the first category from 15 populations in similar regions, such as SCXFS, CQWGS, and GZTZ, were distributed in three regions adjacent to Sichuan, Chongqing, and Guizhou Provinces. The second category mainly contained populations from Hunan Province, including HNHYZ, HNSZX, and HNJD. There was one population each from Yunnan (YNYJ), Sichuan (SCSZ), and Hubei Provinces (HBLFX), and two populations were from Guizhou Province (GZQD and GZSN).

Mantel test between genetic distance and geographic distance

There was no correlation between genetic distance and geographic distance in *P. zhennan* populations based on nSSR data ($r = 0.03$, $P = 0.290$) (Fig. 6). The genetic distance of neither the northwest (core distribution populations) nor southeastern regional populations was significantly correlated with geographic distance (Fig. S1). In contrast, a significant correlation was detected between genetic distance and geographic distance based on cpSSR data ($r = 0.477$, $P = 0.010$) (Fig. 6). Additionally, the Mantel test indicated a significant positive correlation between genetic distance and geographic distance ($r = 0.434$, $P = 0.010$) within the core distribution populations (northwest regional), whereas no such correlation ($r = 0.022$, $P = 0.730$) was determined among the populations of southeastern regional (Fig. S2).

Genetic structure among populations

The neighbor-joining (NJ) trees for *P. zhennan* individuals were constructed based on nSSR and cpSSR datasets, and three and two clusters were identified (Fig. 7A, B). For the nSSR analysis, cluster 1 (red) mainly included individuals mainly from SCSS, SCGTS, SCJFS, SCYFS, SCXFS, SCPZS, SCLYS, CQYCQ, and CQWGS populations. Cluster 2 (yellow) mainly included individuals from HBLFX, HNYLX, HNHYZ, HNSZX, HNJD, HNYSX, GZQD, and GZSN populations. Individuals from SCBZL, CQWGS, HNHYZ, HNJJL, HBLFX, GZCCS, HNAHX, SCXYX, GZTZ, YNYJ, SCSZ, HNYSX, and HNXXT populations formed the third cluster (green) (Fig. 7A). For cpSSR analysis, the individuals in cluster 1 (green) was mainly came from Sichuan (nine populations, including SCXFS, SCGTS, and SCJFS), Chongqing (CQWGS and CQYCQ), Guizhou (GZTZ and GZCS), and Hunan Provinces (HNSZX). Individuals forming the second cluster (red) were from populations from Hunan (six populations, including HNHYZ, HNYLX, and HNJD), Sichuan (six populations, including SCLYS, SCYFS, and SCLJS), Guizhou (GZQD and GZSN), Hubei (HBLFX), and Yunnan Provinces (YNYJ) (Fig. 7B).

Similar to the results of NJ cluster analysis, the delta K value of STRUCTURE analysis reached a maximum when $K = 3$ and $K = 2$ for nSSR and cpSSR datasets, respectively (Fig. S3). The 27 *P. zhennan* populations could then be clustered to three groups based on the Bayesian clustering model for the nSSR dataset. The first group (red) mainly comprised individuals from SCXFS, SCLYS, SCPZS, SCYFS, and CQWGS populations. The second group (green) mainly included individuals from SCGTS, CQYCQ,

SCBZL, SCYXY, YNYJX, HNSZX, and HNXXT populations. The third group (blue) was mainly composed of individuals from SCJLJL, SCLJS, SCSSS, GZTZ, GZCSS, SCSZZ, HNHYZ, HNYLX, HNJD, HNYSX, HNAHX, HBLFX, GZQDN, and GZSNX populations (Fig. 7C). The *P. zhennan* populations could be clustered to two groups based on the cpSSR datasets. The individuals in first group (red) mainly came from GZSNX, GZQDN, HBLFX, HNAHX, HNXXT, HNYSX, HNJD, HNYLX, HNHYZ, HNSZX, HNSZX, YNYJX, and SCSZZ, and the individuals in second group (green) was mainly came from Sichuan (including SCXFS, SCGTS, and SCLYS), Chongqing (CQWGS and CQYCQ), and Guizhou Provinces (GZTZ and GZCSS) (Fig. 7D).

The influences of potential environmental variables on genetic structure

RDA was conducted to determine the influence of potential environmental factors on the genetic diversity of *P. zhennan* populations. The percentage variances in the connection between environmental factors and genetic parameters based on the nSSR datasets were 20.80 and 17.64% in RDA1 and RDA2, respectively (Fig. 8A). Based on cpSSR datasets, the percentage variances in the connection between environmental factors and genetic parameters were 56.19 and 17.57% in RDA1 and RDA2, respectively (Fig. 8B). In RDA1 and RDA2, the percentage variances in the connection between genetic parameters and soil variables were 19.20% and 18.14%, respectively, based on nSSR datasets (Fig. 8C) and 46.72 and 17.24%, respectively, based on cpSSR datasets (Fig. 8D). Consistent with the results of STRUCTURE and cluster analysis, the results of RDA indicated that these soil variables and environmental factors also significantly divided 27 populations into two and three groups of nSSR and cpSSR datasets. The contribution of each environmental factor and soil variable varied among the *P. zhennan* populations both of which are based on the nSSR and/or cpSSR datasets. Three environmental factors (two temperature factors: bio2 and bio5, and one precipitation factor: bio13) and four environmental factors (two temperature factorss: bio4 and bio6, and two precipitation factors: bio14 and bio18) were significantly correlated with the genetic structure according to the nSSR and cpSSR datasets (Table S8). For nSSR and cpSSR datasets, five (two chemical properties: topsoil CEC clay and subsoil CEC soil; two physical properties: topsoil sodicity and subsoil gravel content; and one soil texture: subsoil USDA texture) and nine soil variables (four chemical properties: topsoil organic carbon, topsoil CEC clay, topsoil CEC soil, and subsoil organic carbon; four physical properties: topsoil sand fraction, topsoil silt fraction, subsoil gravel content, and topsoil clay; and one soil texture: subsoil USDA texture), respectively, were significantly correlated with genetic structure (Table S9).

Discussion

The long-term evolution of species leads to genetic diversity, which is crucial for the survival and evolution of population [33, 34]. And, the assessment of genetic diversity is the key point for biodiversity conservation and breeding utilization of tree plants [34, 35]. In present study, the sixteen nSSR and six cpSSR markers were used to evaluate the genetic diversity and population structure of *Phoebe zhennan* populations. Although the previously study based on AFLP markers and RAD-seq reported a low genetic

diversity in *P. zhennan* germplasm resources and low-level genomic diversity within populations [3, 23]. Our previous study indicated a high genetic diversity level in *P. zhennan* breeding populations (102 germplasm resources) using SSR and phenotypic makers [1]. Similar to our previous study on the *P. zhennan* breeding population, the high genetic diversity level was also determined in the present study for 27 *P. zhennan* populations. This could be related to the different genetic markers, population sizes, and plant samples. Whatever, the distribution of species, population size, and global or regional climate change all may affect the genetic diversity [36]. Populations that are spread out over extensive regions are better positioned to preserve allelic diversity compared to those that occupy specific niches [37]. The *P. zhennan* population showed higher level of genetic diversity than *P. shearereri* and *P. chekiangensis*, which is consistent with the difference in distribution region and environmental heterogeneity between *P. zhennan* and the other two *Phoebe* species [11, 38, 39]. Heterozygosity frequently serve as an indicator of genetic diversity and can offer valuable insights for the conservation of threatened species [40]. Our research indicates that the average observed and expected heterozygosity ($H_o = 0.624$ and $H_e = 0.821$) exhibit a moderate degree of genetic variation within the *P. zhennan* population. These findings align with earlier studies on *P. shearereri* ($H_o = 0.609$, $H_e = 0.614$) [11], yet they surpass the heterozygosity level found in the tree species of *Pandanus odorifer* ($H_t = 0.268$) [15] and *Taxus chinensis* ($H_o = 0.263$, $H_e = 0.261$), while remaining below those observed for *Ginkgo biloba* ($H_e = 0.808$) [12]. This result might be related to living habits, life history characteristics, geographical distribution, and sampling size [11, 36, 41]. Furthermore, the value of H_o was less than H_e , suggesting a heterozygous deficiency that may result from inbreeding, non-random reproduction, or disruption of population structure [42].

Subsequently, the genetic diversity of 27 *P. zhennan* populations revealed various differences. The population of SCSZZ ($I = 1.943$) and SCLJS ($I = 0.943$) showed the highest and lowest level of genetic diversity according to the nSSR datasets, respectively. In contrast, the HNSZX population showed the highest diversity ($h = 0.469$), while the HNXXT population had the lowest ($h = 0.041$), according to the cpSSR dataset. The different characteristics of transmission characteristics may lead to the differential of intraspecific genetic structures [19]. The inconsistencies between cpSSR and nSSR markers may be linked to biological characteristics, such as significant pollen and seed exchange, that could partially postpone the genetic imprinting associated with prolonged isolation [43, 44]. However, the nuclear genome has high recombination rate and genetic variability, while the chloroplast genome is maternally inherited, with high stability and a low mutation rate. As a result, nSSR markers tend to show higher polymorphism and genetic diversity than cpSSR markers [45, 46]. The SCJJL population is located in the key conservation and core distribution areas of *P. zhennan* and is well protected. The high level of genetic diversity using nSSR dataset in the population of SCJJL would be correlated to the less human disturbance. The low population size and genetic diversity of some populations (e.g., SCLJS, SCXFS, and GZCSS population) could be due to the thousands of years of indiscriminate cutting and war damage. The high level of genetic diversity in the population of HNSZX based on cpSSR datasets could be closely related to early introduction from Sichuan Province to Hunan Province. The value of fixed index (F) in most populations was higher than zero, suggested excess of pure heterozygotes and slight phenomenon of breeding in those populations, expect the population of CQWGS, SCLJS, GZCSS, and HNXXT [47].

Diverse environmental factors and isolation of populations can hinder the co-adaptive combinations of alleles from parents, consequently diminishing the adaptability of offspring in distant locations [48]. Thus, further analysis of genetic conservation and population restoration is warranted. In the course of the population expansion, alleles with low frequency have a higher likelihood of being eliminated compared to those with high frequency, which positions the former as a significant foundation for studies on the origins of populations [12]. Typically, more rare and/or unique alleles were possess in the origin populations than those in normal populations [12]. The high frequency of private alleles in SCYFS, CQYCQ, SCJFS, and SCPZS populations suggests that the populations in Sichuan and Chongqing Province contained more unique genetic information. Combined with the results of haplotype analysis, haplo-39 displayed the most connections with other haplotypes, which only existed in the SCSST population. Our results supported the view of ring diversification reported by Xiao et al. [3] and speculated that there was a new branch: spreading from the distribution center (Sichuan Province) through Guizhou Province to southeast China (e.g., Hubei and Hunan Provinces). The application of cpSSR makers enabled the unique haplotypes differentiation within populations. Interestingly, smaller populations exhibited a haplotype count that aligned with sample size, whereas 12 out of 27 populations had distinct haplotypes, particularly the populations of YNYJX, SCSZZ, and HNSZX. These findings improve our comprehensive of the complex genetic composition and variation among *P. zhennan* resources.

The ascertain of genetic differentiation is key point to promoting the forests genetic improvement [49]. Consisted with the results for tree species such as *Juglans mandshurica*, *P. sheareri*, and *P. chekiangensis*, the results of AMOVA indicated that genetic differentiation among populations was lower than that within populations in *P. zhennan* for both nSSR and cpSSR markers [11, 13, 38]. Consequently, individuals within the populations with high level of genetic diversity should be focused on when selecting superior germplasms, as this is advantageous for the genetic enhancement of *P. zhennan*. In this study, the mean value of *Fst* for the *P. zhennan* populations was 0.128 based on nSSR markers, indicating a moderate level of genetic differentiation. In contrast, a lower degree of genetic differentiation was observed when employing cpSSR markers (*Fst* = 0.022). Differences may be attributed to the lower polymorphism in cpSSRs than that in nSSR markers, the reason would be due to the characteristics of cpSSRs: size homoplasy, lack of recombination, and uniparental inheritance [10, 50]. These findings align with tree species that possess extensive geographic distributions and utilize an outcrossing reproductive strategy, leading to greater genetic variability within populations compared to that observed between them [41]. Gene flow (*Nm*) is an important index to measure population genetic variation. The *Nm* based on nSSRs and cpSSRs was higher than 1.00, indicating that the effects of genetic drift and differentiation among the *P. zhennan* populations could be effectively eliminated [51]. *P. zhennan* is an insect-pollinated tree species, and its seeds are mainly transmitted by birds [3]; therefore, gene exchange in various geographic regions may have been accomplished by *P. zhennan* through the mechanisms of long-distance pollen or seed dispersal, which in turn could diminish geographic isolation and lessen genetic differentiation across populations.

Understanding the population structure is helpful for revealing the genetic diversity distribution patterns among and within populations; this knowledge is essential for identifying strategies to protect endangered species and effectively utilize germplasm resources [52, 53]. In this study, PCoA, UPGMA, NJ, and STRUCTURE analyses based on nSSRs divided *P. zhennan* populations/individuals into three categories, and populations from the northwest region (e.g., Sichuan and Chongqing Provinces) were the corresponding potential distribution centers. The populations and/or individuals did not cluster populations strictly according to their geographical sources; this could be attributed to the long-distance seed dispersal by frugivorous birds [3] and frequent seed/seedling exchange, especially in the Chengdu plain area of Sichuan Province, China. The results of mantel test ($r = 0.03$, $P = 0.290$) based on nSSR makers also supported these views. Both cluster and STRUCTURE analyses based on cpSSRs divided populations/individuals into two categories, which were approximately in line with geographic sources and consistent with the mantel test ($r = 0.477$, $P = 0.010$). A too-long distance between populations would restrict pollen movement of pollinators and limit dispersal triggers. Moreover, in the majority of angiosperms, chloroplast DNA is inherited uniparentally via maternal sources and is exclusively passed on through seeds, which means it does not carry any information related to pollen [16]. However, some individuals also showed the same ancestry with populations outside the sampling population, which was mostly likely due to the exchange or introduction of seeds between provinces. Further study is needed to elucidate the factors triggering the current spatial population structure, such as “point-by-point” sources based on uniparental maternal DNA sequencing [54].

Environmental factors (including monsoons, mountain tectonics and various historical or ecological processes), genetic drift, and the breeding system influence the distribution, genetic diversity, and population structure of many species [19, 55, 56]. Revealing how environmental factors affect genetic diversity and population structure is an important method to test ecological adaptability [57]. In this study, temperature and precipitation were identified as key environmental factors driving population adaption. This finding is consistent with previous studies, which have underscored the significance of these two environmental factors in affecting plant distribution and genetic structure formation [11, 58, 59]. While climate factors were the primary indicators, soil characteristics also played a significant role in the distribution of plant life on a continental scale [60]. Soil properties are one of the most important environmental factors influencing seed germination, seedling establishment, and plant distribution [61]. The soil chemical and physical properties were identified as vital factors affecting the genetic structure of *P. zhennan* populations. The findings bolster the theory that the characteristics of soil properties significantly influence the distribution of plants and their population structure. Therefore, climate factors and soil properties should be considered when predicting plant population structure and adaption areas.

Human actions have resulted in the division of the habitat for the population and have restricted the exchange of genes among its populations [62]. The in-situ conservation of *P. zhennan* populations revealed fragmented biological communities within the distribution range, although it has been listed as a protected tree species that is ancient and famous trees in China. Only a few *P. zhennan* seeds germinate and survive to become adults due to the low germination rate of seeds and the fact that most of the populations are located in areas with high human disturbance, such as temples. Therefore, the

improvement of the natural regeneration ability with artificial assistance is key to fragment population recovery. The extent to which a species or population can adapt and evolve in response to its environment is influenced by the degree of its genetic diversity [11]. The primary approach for *P. zhennan* is in-situ conservation, as the genetic diversity found within populations significantly exceeds that observed between them. The populations distributed in the northwest area (distribution center of *P. zhennan*), such as SCGTS, SCYFS, and SCJFS populations, which exhibited high genetic diversity level along with a greater number of private alleles. Additionally, the populations of SCSZZ, YNYJX, and HNSZX showcased comparatively high genetic diversity level and an increased count of private haplotypes, which could be safeguarded through the establishment of nature reserves. Furthermore, strengthening breeding work, such as setting up progeny trials and genetic resource nurseries, in the main distribution region is an important method to expand the range of *P. zhennan* populations. This can be done through artificial breeding, thereby expanding weak populations, such as SCLJS, GZCSS, and HNANX. Establishing a gene bank of germplasm resources is the key method for genetic conservation; it can not only preserve diversified genotypes but also provide core germplasm resources for long-term breeding of *P. zhennan*.

Conclusion

In this study, genetic diversity was determined using 543 individuals from 27 *P. zhennan* populations using 16 nSSR and six cpSSR markers. Elevated genetic variation and minimal inbreeding were noted across the populations, with significant genetic differentiation largely occurring within these populations, potentially linked to temperature, precipitation, as well as both chemical and physical soil factors. The *P. zhennan* individuals were separated into two or three groups according to STRUCTURE and cluster analysis based on nSSRs and cpSSRs, respectively. Combined with geographic and genetic distance, these data should be considered in further genetic conservation and breeding utilization efforts. Regarding the genetic characteristics of the *P. zhennan* populations, there is a need to formulate in situ conservation and population renewal, for example, by prioritizing populations that exhibit greater genetic diversity and unique alleles, as well as by enhancing seedling reintroduction efforts to bolster these populations, we aim to sustain the genetic diversity of *P. zhennan* populations to the fullest extent. This research provides a basis for the population protection strategy and theoretical guidance for the development and utilization of *P. zhennan* genetic resources.

Declarations

Funding

This work was supported by Science and Technology Innovation Team Project of Sichuan Forestry and Grassland (2023LCTD0203; 2024LCTD0213); the Key Research and Development Project of Sichuan Province (2021YFYZ0032); the National Science Foundation of Sichuan Province (2022NSFSC1062); the Ecosystem Improvement in the Upper Reaches of the Yangtze River Basin Program (510201202038467).

Authors' contributions

Jian Peng, Yunjie Gu, and Hanbo Yang design of the work; Jiaxin Xie, Hongying Guo, Shuaiying Zhang, and Hanbo Yang the acquisition, analysis, Xin Huang, Xiandan Luo, Jianghong Qian, and Jian Peng, Minhao Liu, Xueqin Wan, Lianghua Chen, Xiong Huang interpretation of data; Hanbo Yang, Fan Zhang, Fang He, Peng Zhu, and Yu Zhong have drafted the work or substantively revised it.

Competing interests

Not applicable.

Ethics approval and consent to participate

Not applicable.

Availability of data and materials

Not applicable.

Acknowledgements

Not applicable.

References

1. Zhu Y, An W, Peng J, Li J, Gu Y, Jiang B, Chen L, Zhu P, Yang H: **Genetic Diversity of Nanmu (*Phoebe zhennan* S. Lee. et F. N. Wei) Breeding Population and Extraction of Core Collection Using nSSR, cpSSR and Phenotypic Markers**. In: *Forests*. vol. 13; 2022.
2. Yang H, Zhang S, Gu Y, Peng J, Huang X, Guo H, Chen L, Jiang Y, Liu M, Luo X *et al*: **Identification and variation analysis of the composition and content of essential oil and fragrance compounds in *Phoebe zhennan* wood at different tree ages**. *Frontiers in Plant Science* 2024, **15**.
3. Xiao J-H, Ding X, Li L, Ma H, Ci X-Q, van der Merwe M, Conran JG, Li J: **Miocene diversification of a golden-thread nanmu tree species (*Phoebe zhennan*, Lauraceae) around the Sichuan Basin shaped by the East Asian monsoon**. *Ecology and Evolution* 2020, **10**(19):10543-10557.
4. Yang H, An W, Gu Y, Peng J, Jiang Y, Li J, Chen L, Zhu P, He F, Zhang F *et al*: **Integrative metabolomic and transcriptomic analysis reveals the mechanism of specific color formation in *Phoebe zhennan* heartwood**. *International Journal of Molecular Sciences* 2022, **23**(21):13569.
5. Ding X, Xiao JH, Li L, Conran JG, Li J: **Congruent species delimitation of two controversial gold-thread nanmu tree species based on morphological and restriction site-associated DNA sequencing data**. *Journal of Systematics and Evolution* 2019, **57**(3):234-246.
6. Fu LG: **China plant red data book: Rare and endangered plants (Vol. I)**: Beijing: Science Press (in Chinese); 1992.

7. Muriira NG, Muchugi A, Yu A, Xu J, Liu A: **Genetic Diversity Analysis Reveals Genetic Differentiation and Strong Population Structure in Calotropis Plants.** *Scientific Reports* 2018, **8**(1):7832.
8. Sun Q-B, Li L-F, Li Y, Wu G-J, Ge X-J: **SSR and AFLP Markers Reveal Low Genetic Diversity in the Biofuel Plant *Jatropha curcas* in China.** *Crop Science* 2008, **48**(5):1865-1871.
9. Kaler AS, Purcell LC, Beissinger T, Gillman JD: **Genomic prediction models for traits differing in heritability for soybean, rice, and maize.** *BMC Plant Biology* 2022, **22**(1):87.
10. Guo Q, Xue X, Wang D, Zhang L, Liu W, Wang E, Cui X, Hou X: **Genetic diversity and population genetic structure of *Paeonia suffruticosa* by chloroplast DNA simple sequence repeats (cpSSRs).** *Horticultural Plant Journal* 2024.
11. Wang Y, Ma X, Lu Y, Hu X, Lou L, Tong Z, Zhang J: **Assessing the current genetic structure of 21 remnant populations and predicting the impacts of climate change on the geographic distribution of *Phoebe sheareri* in southern China.** *Science of The Total Environment* 2022, **846**:157391.
12. Zhou Q, Mu K, Ni Z, Liu X, Li Y, Xu L-a: **Analysis of genetic diversity of ancient Ginkgo populations using SSR markers.** *Industrial Crops and Products* 2020, **145**:111942.
13. Zhang Q, Zhang X, Yang Y, Xu L, Feng J, Wang J, Tang Y, Pei X, Zhao X: **Genetic Diversity of *Juglans mandshurica* Populations in Northeast China Based on SSR Markers.** *Frontiers in Plant Science* 2022, **13**.
14. Wu F, Zhang S, Gao Q, Liu F, Wang J, Wang X: **Genetic diversity and population structure analysis in a large collection of *Vicia amoena* in China with newly developed SSR markers.** *BMC Plant Biology* 2021, **21**(1):544.
15. Nasim N, Sandeep IS, Sahoo A, Das S, Panda MK, Acharya L, RamaRao VV, Nayak S, Mohanty S: **Population genetic structure and diversity analysis in economically important *Pandanus odorifer* (Forssk.) Kuntze accessions employing ISSR and SSR markers.** *Industrial Crops and Products* 2020, **143**:111894.
16. Dumolin S, Demesure B, Petit RJ: **Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method.** *Theoretical and Applied Genetics* 1995, **91**(8):1253-1256.
17. Xiong Y, Xiong Y, Shu X, Yu Q, Lei X, Li D, Yan J, Bai S, Ma X: **Molecular Phylogeography and Intraspecific Divergences in Siberian Wildrye (*Elymus sibiricus* L.) Wild Populations in China, Inferred From Chloroplast DNA Sequence and cpSSR Markers.** *Frontiers in Plant Science* 2022, **13**.
18. Niu Z, Hou Z, Wang M, Ye M, Zhang B, Xue Q, Liu W, Ding X: **A comparative plastomics approach reveals available molecular markers for the phylogeographic study of *Dendrobium huoshanense*, an endangered orchid with extremely small populations.** *Ecology and Evolution* 2020, **10**(12):5332-5342.
19. Lu Z-J, Wang T-R, Zheng S-S, Meng H-H, Cao J-G, Song Y-G, Kozłowski G: **Phylogeography of *Pterocarya hupehensis* reveals the evolutionary patterns of a Cenozoic relict tree around the Sichuan Basin.** *Forestry Research* 2024, **4**(1).

20. Huang L-S, Sun Y-Q, Jin Y, Gao Q, Hu X-G, Gao F-L, Yang X-L, Zhu J-J, El-Kassaby YA, Mao J-F: **Development of high transferability cpSSR markers for individual identification and genetic investigation in Cupressaceae species.** *Ecology and Evolution* 2018, **8**(10):4967-4977.
21. Arif IA, Bakir MA, Khan HA, Al Farhan AH, Al Homaidan AA, Bahkali AH, Al Sadoon M, Shobrak M: **A Brief Review of Molecular Techniques to Assess Plant Diversity.** *International Journal of Molecular Sciences* 2010, **11**(5):2079-2096.
22. Powell W, Morgante M, Andre C, McNicol JW, Machray GC, Doyle JJ, Tingey SV, Rafalski JA: **Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome.** *Current Biology* 1995, **5**(9):1023-1029.
23. Gao J, Zhang W, Li J, Long H, He W, Li X: **Amplified fragment length polymorphism analysis of the population structure and genetic diversity of Phoebe zhennan (Lauraceae), a native species to China.** *Biochemical Systematics and Ecology* 2016, **64**:149-155.
24. Peakall R, Smouse PE: **GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update.** *Bioinformatics* 2012, **28**(19):2537-2539.
25. Liu K, Muse SV: **PowerMarker: an integrated analysis environment for genetic marker analysis.** *Bioinformatics* 2005, **21**(9):2128-2129.
26. Eliades N-G, Eliades DG: **HAPLOTYPE ANALYSIS: Software for analysis of haplotype data.** .
Distributed by the authors Forest Genetics and Forest Tree Breeding, Georg-August University Goettingen, Germany URL <http://www.uni-goettingen.de/en/134935html> 2009.
27. Kalinowski ST, Taper ML, Marshall TC: **Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment.** *Molecular Ecology* 2007, **16**(5):1099-1106.
28. Glaubitz JC: **convert: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages.** *Molecular Ecology Notes* 2004, **4**(2):309-310.
29. Falush D, Stephens M, Pritchard JK: **Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies.** *Genetics* 2003, **164**(4):1567-1587.
30. Earl DA, vonHoldt BM: **STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method.** *Conservation Genetics Resources* 2012, **4**(2):359-361.
31. 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**(14):1801-1806.
32. Rosenberg NA: **distruct: a program for the graphical display of population structure.** *Molecular Ecology Notes* 2004, **4**(1):137-138.
33. González AV, Gómez-Silva V, Ramírez MJ, Fontúrbel FE: **Meta-analysis of the differential effects of habitat fragmentation and degradation on plant genetic diversity.** *Conservation Biology* 2020, **34**(3):711-720.
34. Govindaraj M, Vetriventhan M, Srinivasan M: **Importance of Genetic Diversity Assessment in Crop Plants and Its Recent Advances: An Overview of Its Analytical Perspectives.** *Genetics Research*

International 2015, **2015**(1):431487.

35. Millar CI, Westfall RD: **Allozyme markers in forest genetic conservation.** *New Forests* 1992, **6**(1):347-371.
36. Rubio-Moraga A, Candel-Perez D, Lucas-Borja ME, Tiscar PA, Viñegla B, Linares JC, Gómez-Gómez L, Ahrazem O: **Genetic Diversity of *Pinus nigra* Arn. Populations in Southern Spain and Northern Morocco Revealed By Inter-Simple Sequence Repeat Profiles.** *International Journal of Molecular Sciences* 2012, **13**(5):5645-5658.
37. Sanchez M, Ingrouille MJ, Cowan RS, Hamilton MA, Fay MF: **Spatial structure and genetic diversity of natural populations of the Caribbean pine, *Pinus caribaea* var. *bahamensis* (Pinaceae), in the Bahaman archipelago.** *Botanical Journal of the Linnean Society* 2014, **174**(3):359-383.
38. Ding Y, Zhang J, Lu Y, Lin Ep, Lou L, Tong Z: **Development of EST-SSR markers and analysis of genetic diversity in natural populations of endemic and endangered plant *Phoebe chekiangensis*.** *Biochemical Systematics and Ecology* 2015, **63**:183-189.
39. Booy G, Hendriks RJJ, Smulders MJM, Van Groenendael JM, Vosman B: **Genetic Diversity and the Survival of Populations.** *Plant Biology* 2000, **2**(4):379-395.
40. Schmidt TL, Jasper M-E, Weeks AR, Hoffmann AA: **Unbiased population heterozygosity estimates from genome-wide sequence data.** *Methods in Ecology and Evolution* 2021, **12**(10):1888-1898.
41. Hamrick JL, Godt MJW, Sherman-Broyles SL: **Factors influencing levels of genetic diversity in woody plant species.** *New Forests* 1992, **6**(1):95-124.
42. Liao R, Luo Y, Yisilam G, Lu R, Wang Y, Li P: **Development and characterization of SSR markers for *Sanguinaria canadensis* based on genome skimming.** *Applications in Plant Sciences* 2019, **7**(9):e11289.
43. Li X, Ruhsam M, Wang Y, Zhang H-Y, Fan X-Y, Zhang L, Wang J, Mao K-S: **Wind-dispersed seeds blur phylogeographic breaks: The complex evolutionary history of *Populus lasiocarpa* around the Sichuan Basin.** *Plant Diversity* 2023, **45**(2):156-168.
44. Bai W-N, Wang W-T, Zhang D-Y: **Contrasts between the phylogeographic patterns of chloroplast and nuclear DNA highlight a role for pollen-mediated gene flow in preventing population divergence in an East Asian temperate tree.** *Molecular Phylogenetics and Evolution* 2014, **81**:37-48.
45. Duan B, Kang T, Wan H, Liu W, Zhang F, Mu S, Guan Y, Li Z, Tian Y, Kang X: **Microsatellite markers reveal genetic diversity and population structure of *Portunus trituberculatus* in the Bohai Sea, China.** *Scientific Reports* 2023, **13**(1):8668.
46. Wang S, Wang Y, Zhou J, Li P, Lin H, Peng Y, Yu L, Zhang Y, Wang Z: **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**(20):2706.
47. Wright S: **Genetical Structure of Populations.** *Nature* 1950, **166**(4215):247-249.
48. Hufford KM, Krauss SL, Veneklaas EJ: **Inbreeding and outbreeding depression in *Stylidium hispidum*: implications for mixing seed sources for ecological restoration.** *Ecology and Evolution* 2012, **2**(9):2262-2273.

49. Li X, Zhao M, Xu Y, Li Y, Tigabu M, Zhao X: **Genetic Diversity and Population Differentiation of *Pinus koraiensis* in China.** *Horticulturae* 2021, **7**(5):104.
50. Wheeler GL, Dorman HE, Buchanan A, Challagundla L, Wallace LE: **A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology.** *Applications in Plant Sciences* 2014, **2**(12):1400059.
51. Slatkin M: **Gene Flow in Natural Populations.** *Annual Review of Ecology, Evolution, and Systematics* 1985, **16**(Volume 16):393-430.
52. Melo ATdO, Coelho ASG, Pereira MF, Blanco AJV, Franceschinelli EV: **High genetic diversity and strong spatial genetic structure in *Cabralea canjerana* (Vell.) Mart. (Meliaceae): implications to Brazilian Atlantic Forest tree conservation.** *Natureza & Conservação* 2014, **12**(2):129-133.
53. Vinceti B, Loo J, Gaisberger H, van Zonneveld MJ, Schueler S, Konrad H, Kadu CAC, Geburek T: **Conservation Priorities for *Prunus africana* Defined with the Aid of Spatial Analysis of Genetic Data and Climatic Variables.** *PLOS ONE* 2013, **8**(3):e59987.
54. Zhou B, Zhang Z, Zhang H, Li Y, Ma Y, Zhang S, Niu S, Li Y: **"Point by point" source: The Chinese pine plantations in North China by evidence from mtDNA.** *Ecology and Evolution* 2024, **14**(6):e11570.
55. Willis KJ, Bennett KD, Walker D, Hewitt GM: **Genetic consequences of climatic oscillations in the Quaternary.** *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* 2004, **359**(1442):183-195.
56. Qiu Y-X, Fu C-X, Comes HP: **Plant molecular phylogeography in China and adjacent regions: Tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora.** *Molecular Phylogenetics and Evolution* 2011, **59**(1):225-244.
57. Lu M, Krutovsky KV, Loopstra CA: **Predicting Adaptive Genetic Variation of Loblolly Pine (*Pinus taeda* L.) Populations Under Projected Future Climates Based on Multivariate Models.** *Journal of Heredity* 2019, **110**(7):857-865.
58. Nedlo JE, Martin TA, Vose JM, Teskey RO: **Growing season temperatures limit growth of loblolly pine (*Pinus taeda* L.) seedlings across a wide geographic transect.** *Trees* 2009, **23**(4):751-759.
59. Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, González-Martínez SaC, Neale DB: **Patterns of Population Structure and Environmental Associations to Aridity Across the Range of Loblolly Pine (*Pinus taeda* L., Pinaceae).** *Genetics* 2010, **185**(3):969-982.
60. Ni M, Vellend M: **Soil properties constrain predicted poleward migration of plants under climate change.** *New Phytologist* 2024, **241**(1):131-141.
61. Alamusa, Yang TT, Cao J, Wang YC, Liu Y: **Soil moisture influences vegetation distribution patterns in sand dunes of the Horqin Sandy Land, Northeast China.** *Ecological Engineering* 2017, **105**:95-101.
62. HAAG T, SANTOS AS, SANA DA, MORATO RG, CULLEN JR L, CRAWSHAW JR PG, DE ANGELO C, DI BITETTI MS, SALZANO FM, EIZIRIK E: **The effect of habitat fragmentation on the genetic structure of a top predator: loss of diversity and high differentiation among remnant populations of Atlantic Forest jaguars (*Panthera onca*).** *Molecular Ecology* 2010, **19**(22):4906-4921

Figures

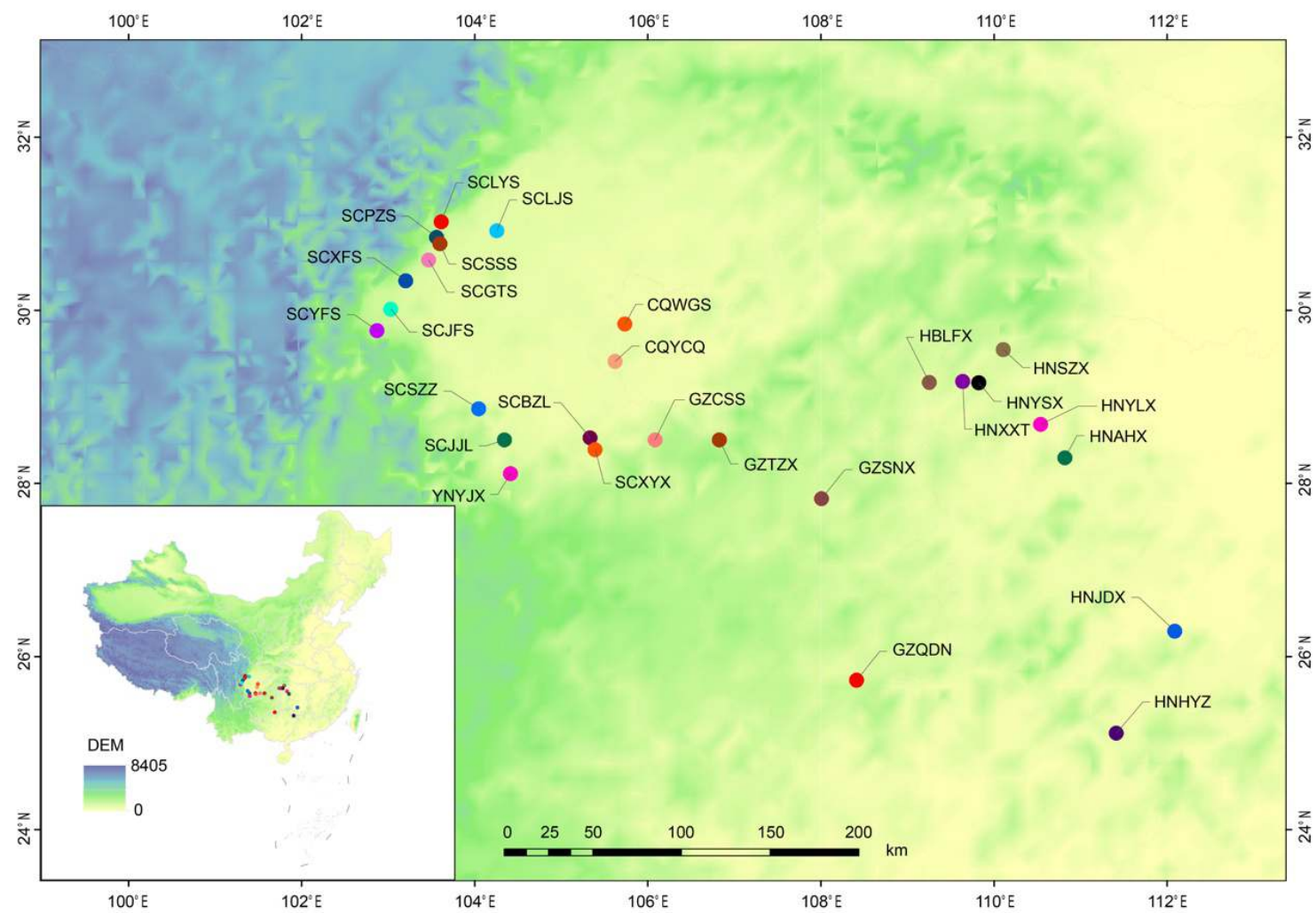


Figure 1

Geographical distribution of 27 natural *Phoebe zhennan* populations

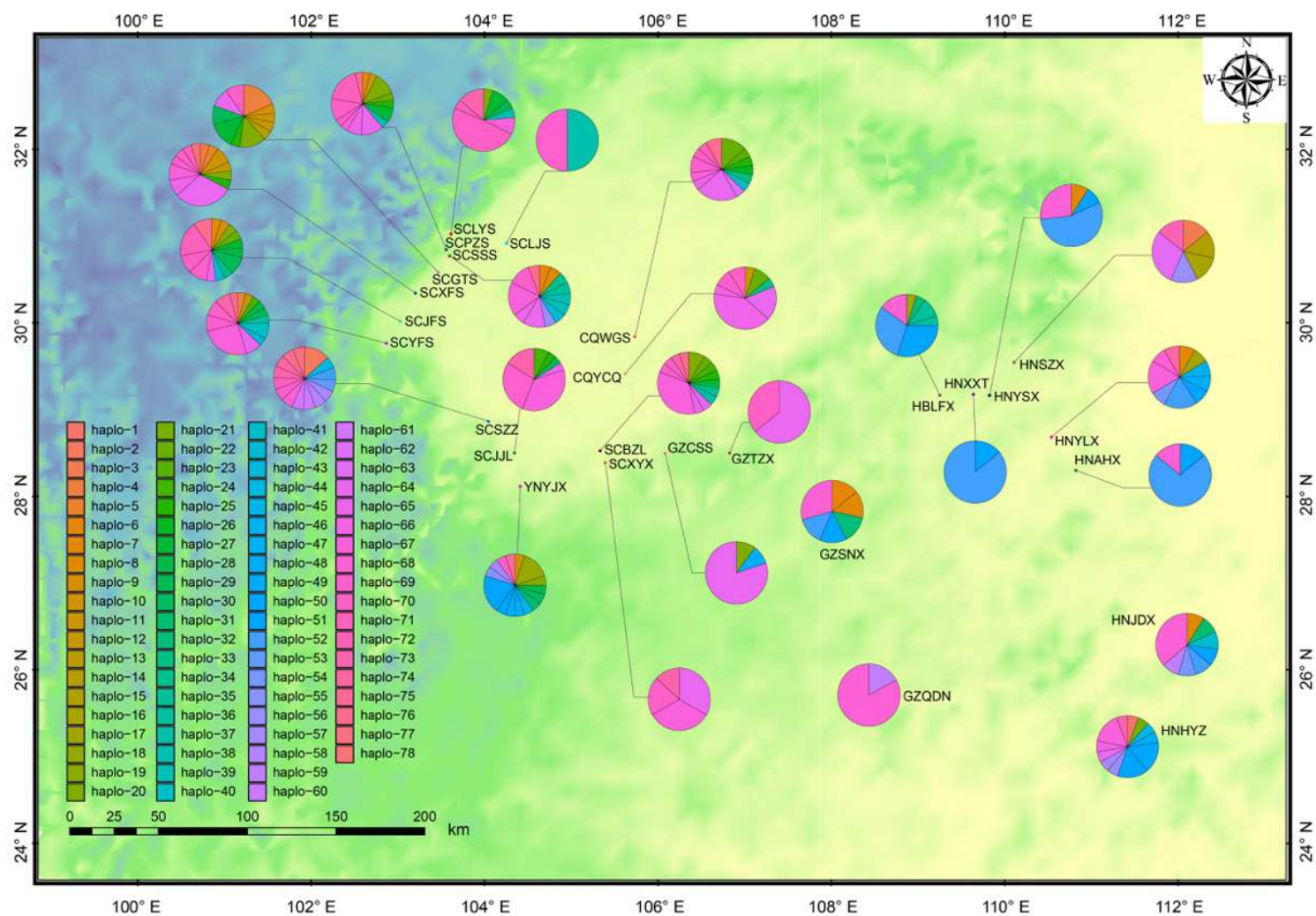


Figure 2

Haplotype frequency analysis of natural *Phoebe zhennan* populations using six cpSSR markers

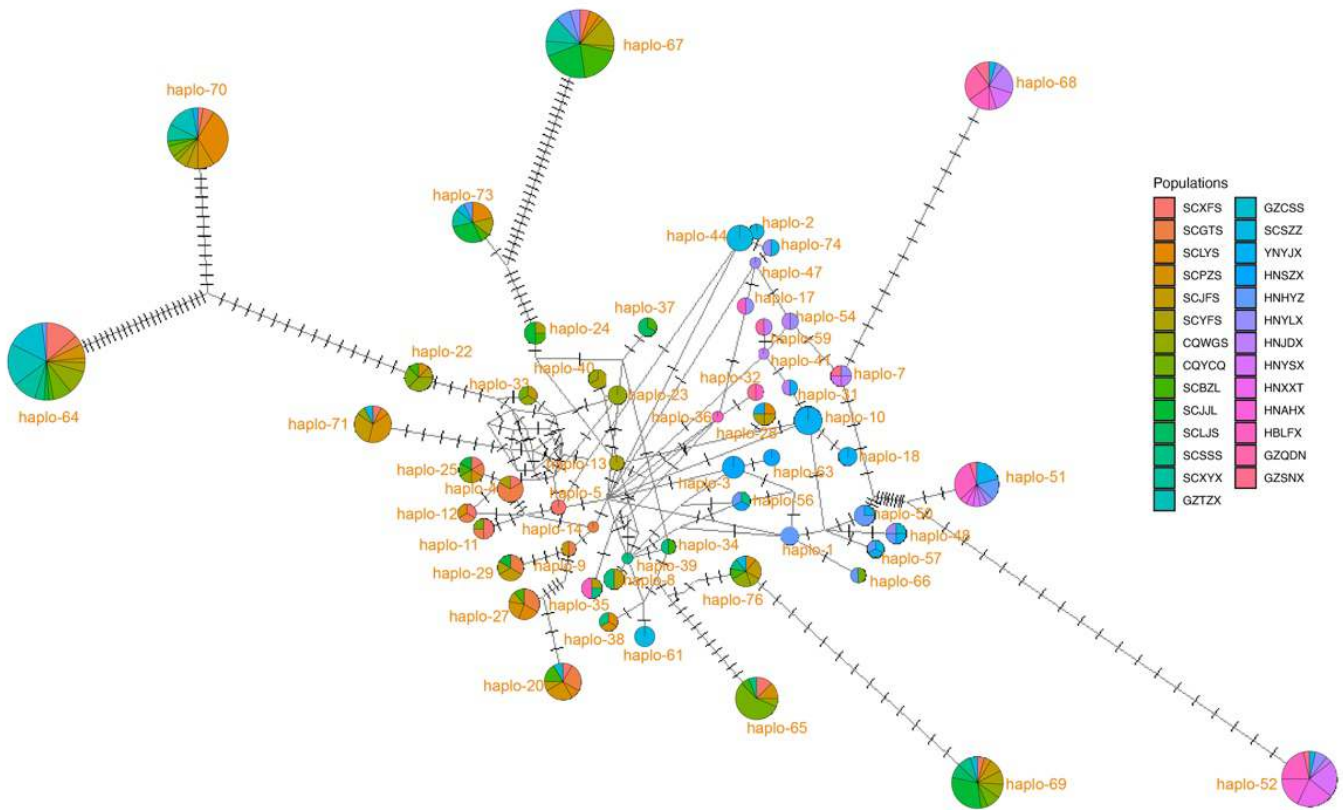


Figure 3

Network of cpDNA haplotypes constructed using the median-joining method

Note: The different colors in the circle represent the different populations. The size of the circles corresponds to the frequency of haplotypes, with larger circles indicating more frequent haplotype occurrences and smaller circles representing haplotypes that are found less frequently. The size of pies in the circles corresponds to the number of haplotypes in the population. The numerals (such as haplo-20) near circles are sequential numbers of haplotypes. The long lines represent the connection lengths. The short tangent segment representation on a long line represents mutated position lines.

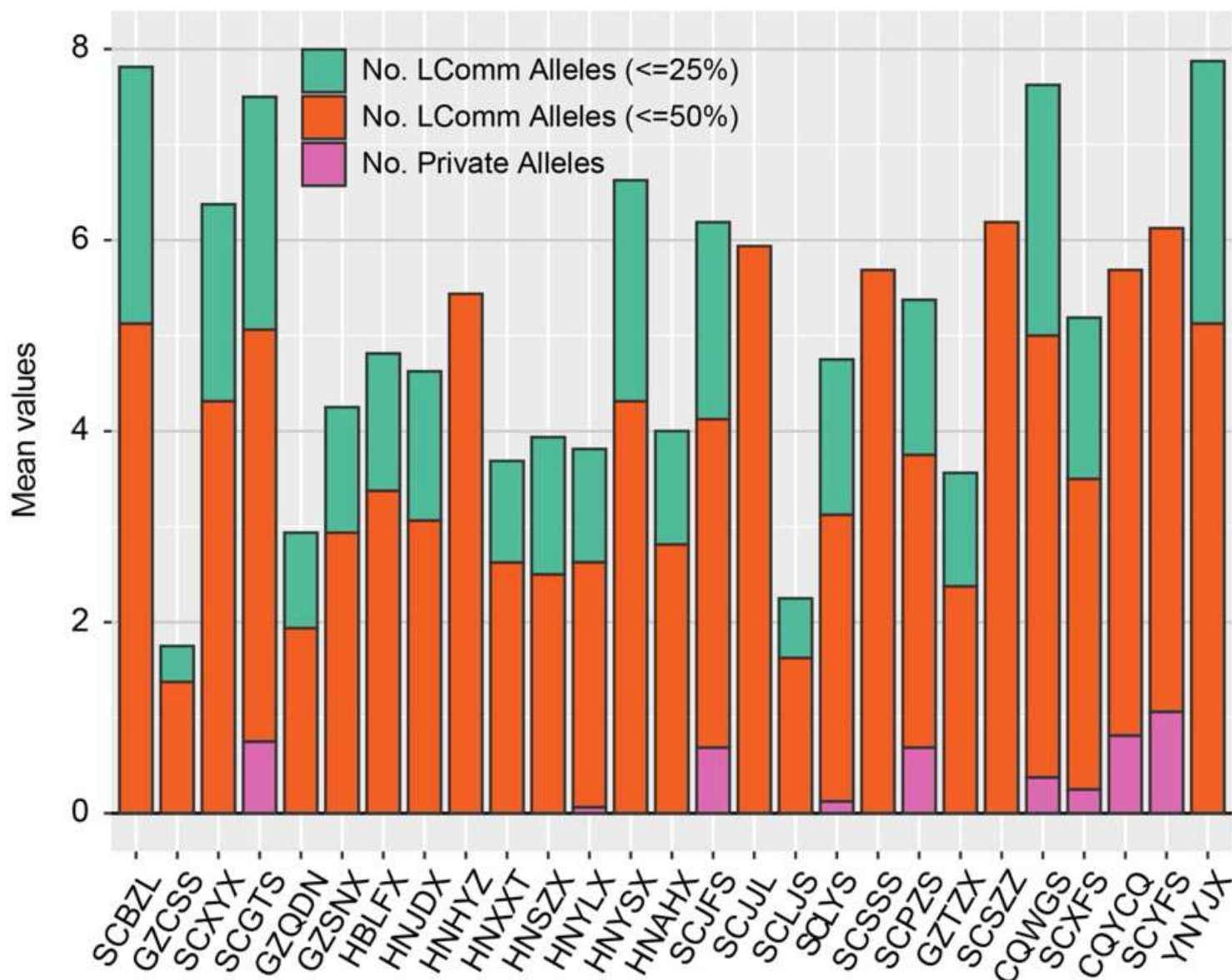


Figure 4

Number of rare and private alleles within populations

Note: No. Private Allele: No. of alleles unique to a single population; No. LComm Alleles ($\leq 25\%$): No. of locally common alleles (freq. $\geq 5\%$) found in 25% or fewer populations; No. LComm Alleles ($\leq 50\%$): No. of locally common alleles (freq. $\geq 5\%$) found in 50% or fewer populations.

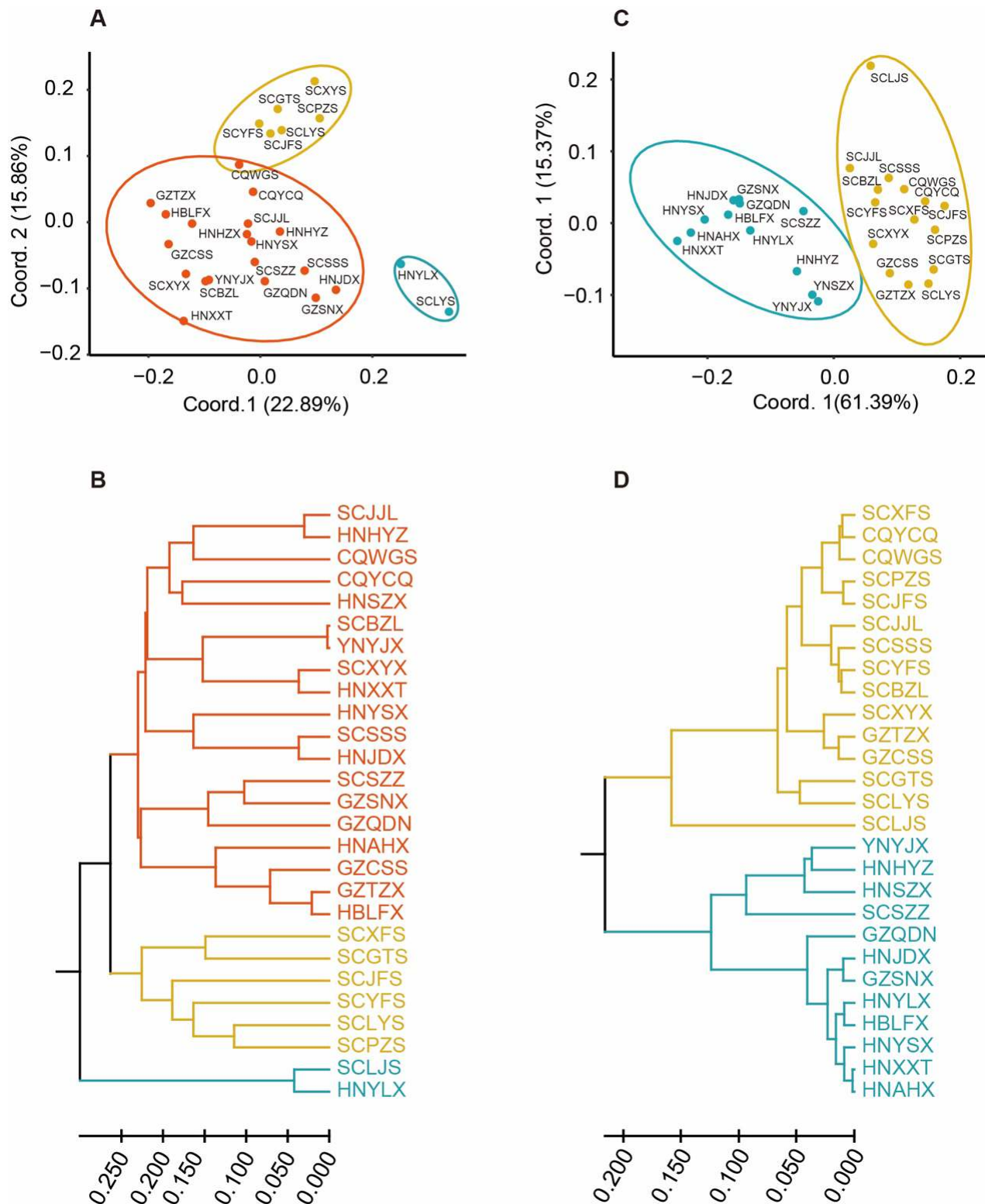


Figure 5

Distribution and unweighted pair group method with arithmetic mean (UPGMA) dendrogram of *Phoebe zhennan* populations based on genotyping data from nSSR and cpSSR markers

Note: A: Population distribution based on genotyping data from nSSR markers using the first two PCoA coordinates; B: UPGMA dendrogram for populations based on Nei's genetic distance from nSSR data; C:

Population distribution based on genotyping data from cpSSR markers using the first two PCoA coordinates; D: UPGMA dendrogram for populations based on Nei's genetic distance from cpSSR data.

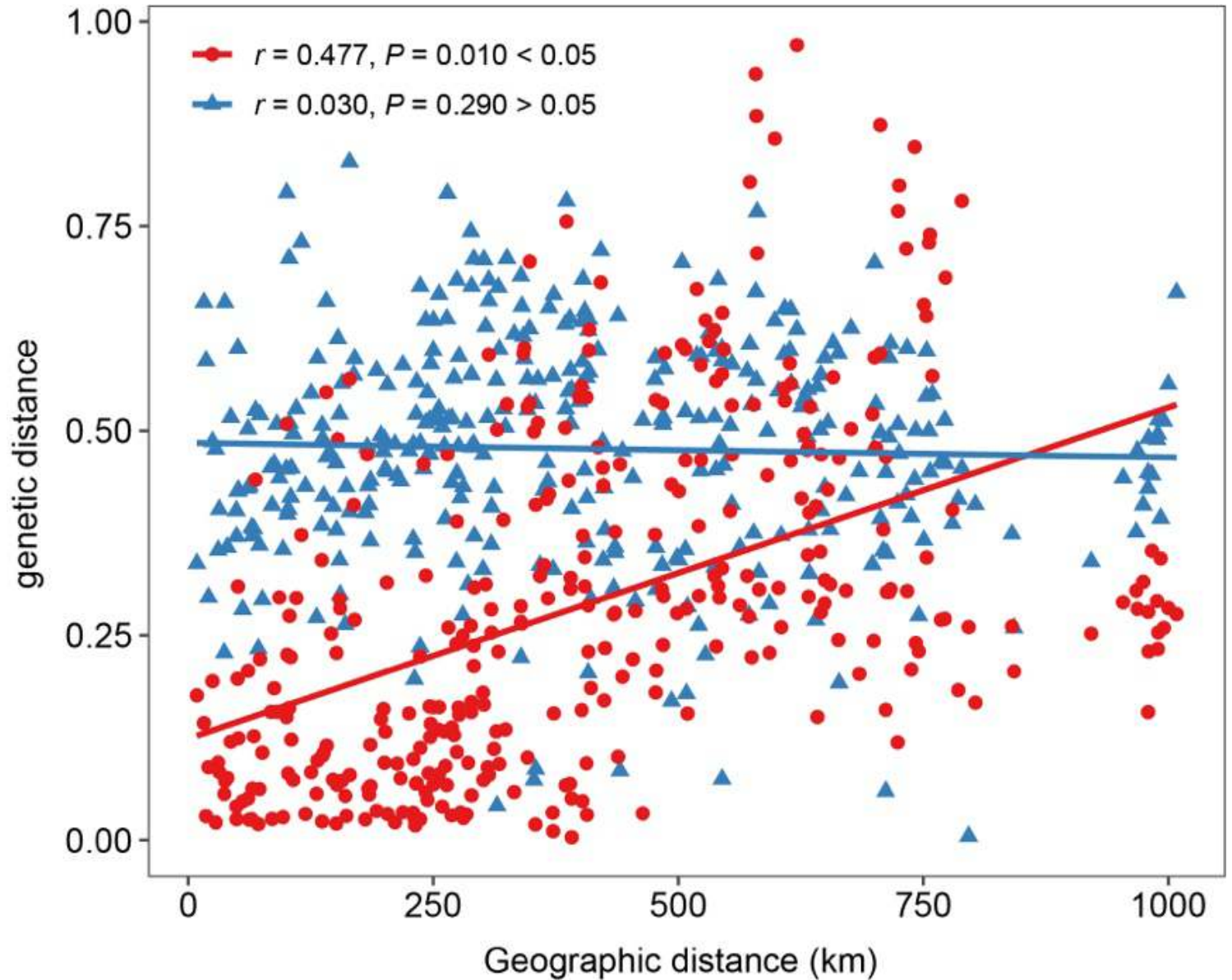


Figure 6

Mantel test of the correlation between Nei's genetic distance and geographic distance among the studied populations.

Note: The red line and dots represent the correlation between Nei's genetic distance based on cpSSR and geographic distance, while the blue line and triangles represent the correlation between Nei's genetic distance based on nSSR and geographic distance.

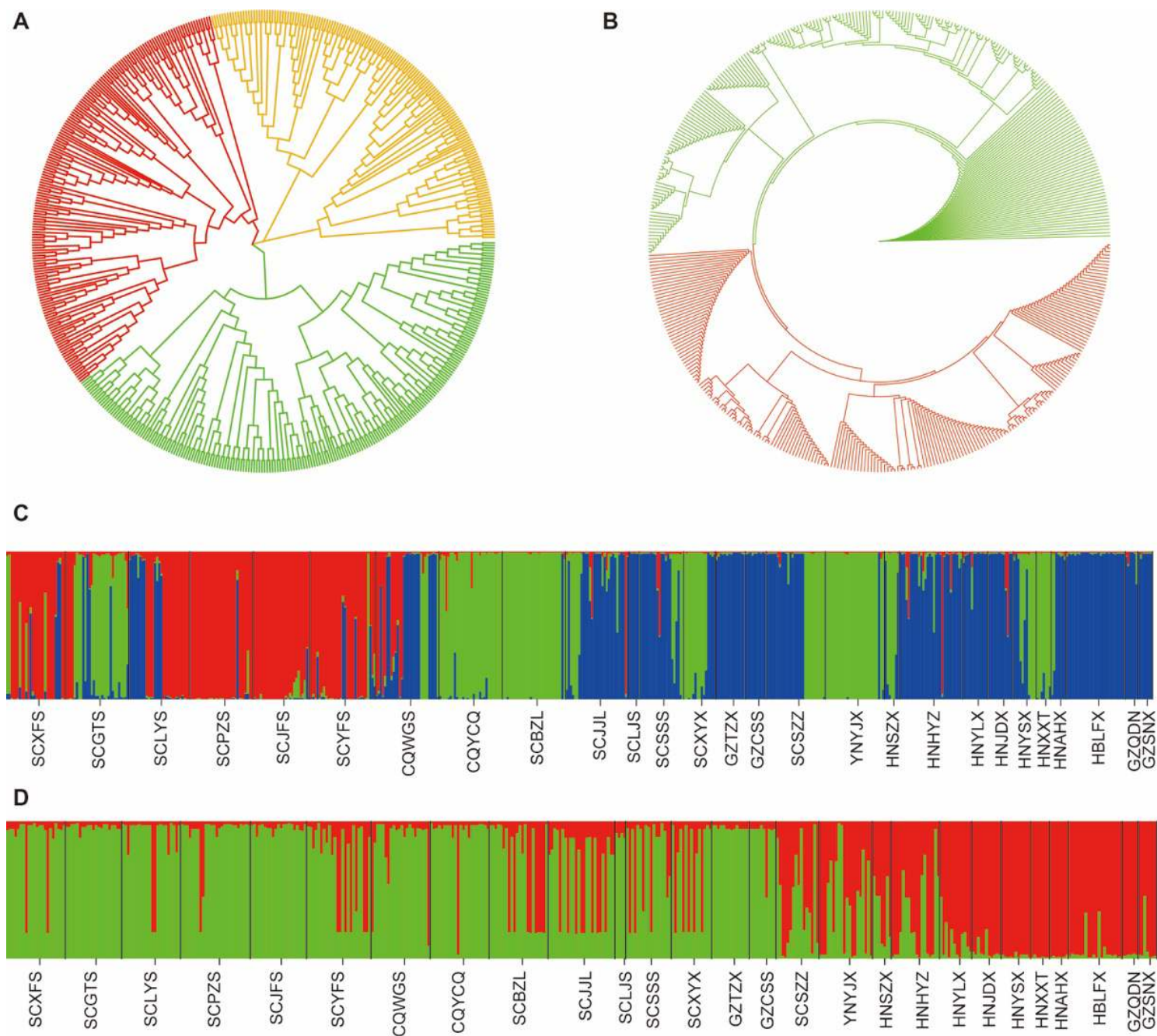


Figure 7

Neighbor-joining tree and population structure of 27 natural *Phoebe zhennan* populations

Note: A: Neighbor-joining tree based on nSSR datasets; B: Neighbor-joining tree based on cpSSR datasets; C: Population genetic structure based on the Bayesian clustering model based on the nSSR dataset; D: Population genetic structure based on the Bayesian clustering model based on the cpSSR dataset.

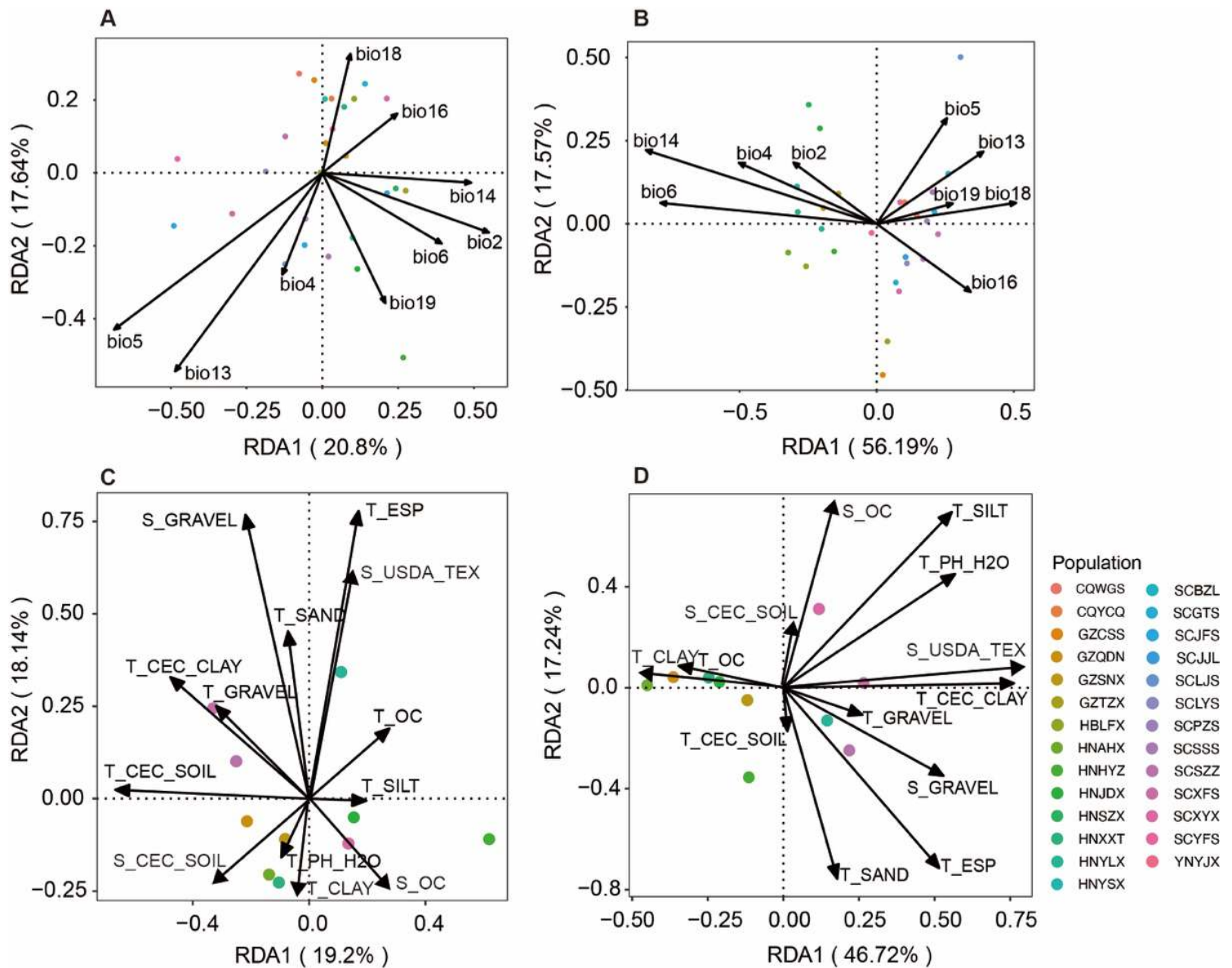


Figure 8

RDA of *Phoebe zhennan* showing relative contribution of each environmental and soil variable shaping population genetic structure

Note: The biplot depicts the eigenvalues and lengths of eigenvectors in RDA; A: RDA of the relative contribution of environmental variables based on the nSSR dataset; B: RDA of the relative contribution of environmental variables based on the cpSSR dataset; C: RDA of the relative contribution of soil properties based on the nSSR dataset; D: RDA of the relative contribution of soil properties based on the cpSSR dataset.

Supplementary Files

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

- [supplementaryfiles.zip](#)