Genet Resour Crop Evol
https://doi.org/10.1007/s10722-020-01054-8

RESEARCH ARTICLE

Genetic diversity and population structure analysis of *Emmenopterys henryi* Oliv., an endangered relic species endemic to China

Yanli Niu · Arvind Bhatt · Yansong Peng · Wenxing Chen · Yuan Gao · Xuanhuai Zhan · Zhiyong Zhang · Wan Hu · Manzhen Song · Zhijun Yu

Received: 19 March 2020 / Accepted: 22 October 2020

Abstract A simple sequence repeat (SSR) marker was used to assess the genetic diversity and population structure of an endangered relic species *Emmenopterys henryi* Oliv., endemic to China. A total of 124 samples from six populations were analyzed using eight pairs of SSR primers. Total 114 alleles were detected with an average of 14.3 alleles per locus. The presence of null alleles can result in low genetic diversity parameters (H_e and allelic richness) for all loci and mostly likely caused the deviation from Hardy–Weinberg equilibrium (HWE). The Polymorphism information content (PIC) values were higher which indicates that selection of SSR markers were suitable choice for assessing the genetic diversity in *E. henryi*. Our results revealed that the natural populations of *E. henryi* have a high degree of genetic diversity. Genetic diversity amongst the six *E. henryi* populations in decreasing order were: GJY > LS > DWS > WYS > HS > JGS. Further, genetic structure and the Neighbor-Joining (NJ) cluster analysis indicated that there were cross-mixing among the 124 samples. Four populations (i.e., HS, DWS, JGS and WYS) were clustered in one group, whereas LS and GJY population were clustered separately as two groups. A Principal Component Analysis (PCA) also showed a similar clustering trend. The results of F_{st} (0.085) and AMOVA indicated that the genetic variation resided within the populations and existence of frequent gene exchanges among populations. Based on these results, populations with abundant genetic variation and rare alleles should be conserved in situ and ex situ. Furthermore, seeds from the different populations with high levels of genetic diversity could be collected and propagated in order to capture the maximum available genetic diversity. The present study on genetic diversity of *E.henryi* could be helpful in expanding the understanding on species survival and also for developing an effective long term conservation strategy.

Keywords *Emmenopterys henryi* · Relic plant · Genetic diversity · Germplasm resource · SSR

Y. Niu (✉) · A. Bhatt · Y. Peng · X. Zhan · Z. Zhang · W. Hu · M. Song · Z. Yu
Chinese Academy of Science, Lushan Botanical Garden, Jiangxi 332900, China
e-mail: nyl0601@126.com

W. Chen
Key Lab of Horticultural Plant Biology, Huazhong Agricultural University, Wuhan 430070, China

Y. Gao
Chinese Academy of Agricultural Sciences, Pomology Research Instituts, Xingcheng 25100, China

Introduction

Emmenopterys henryi Oliv. (Rubiaceae) is a monotypic endangered tree species (Guo and Li 2009). This species is native to subtropical China, where it occurs

Published online: 06 November 2020

in disjunct montane valleys, especially in warm-temperate deciduous forests (Chen et al. 2014). It perfers a warm or cool climate, grows well in fertile acidic or slightly acidic soils (Zhang 2016; Yang and Zhang 2007). The tree blooms every 2 to 4 years, but the first flowering time greatly differs between the introduced and wild individuals. For example, *E. henryi* planted at the Quarryhill Botanical Garden in Sonoma, Californiabloomed for the first time in the sixth year (McNamara 2005). It has large dark-green opposite leaves with petioles turning red in spring. A persistent white calyx develops lobes on each jasmine-like flower, and it turns red when the sweet-scented fruit ripens in December, alluding to its Chinese name "sweet-fruit tree" (Ma et al. 2018). *E. henryi* is valuable to the wood, pharmaceutical, landscaping and chemical industry (Hu and Cui 2011). In China, its roots and bark are effective in treating gastropathy (Zhou 2011). Unfortunately, the current distribution of *E. henryi* in China are fragmented and diminishing due to overlogging, tourism and habitat loss (Ma et al. 2018). Moreover, other factors such as insufficient pollination due to its specific pollinator requiments and high fruit/seed abortion rate, are also responsible for its endangered status (Cheng 2008; Guo et al. 2011a, b).

To date some ecological studies have been conducted to understand and identify factors which are responsible for endangered status of *E. henryi* among the natural populations (Yang et al. 2007; Man et al. 2008; Lu et al. 2018). Previous studies on population dynamics and structural features of *E. henryi* reported that natural populations were severely fragmented and with relatively few individuals in wild populations (98, 48 and 12) individuals in Jiugongshan Nature Reserve (Hubei Province), Qizimeishan National Nature Reserve and Tiantangzhai populations (Anhui Province), respectively (Xiong et al. 2006; Yang et al. 2007; Man et al. 2008). Similarly, Lu et al. (2018) also investigated the population composition and structural characteristics of *E. henryi* in the Gutianshan Valley between 600 and 850 m asl, where this species mainly occurs and has a high canopy density. All these studies reported that over-exploitation, habitat destruction and poor natural regeneration are the main contraints which are responsible for threatening natural populations. Information on genetic variation is vital for developing efficient conservation strategies for rare and endangered species. Therefore, understanding the genetic diversity of *E. henryi* is urgently required in order to develop long term conservation strategies.

Different molecular marker such as random amplified polymorphic DNA (RAPD), inter-simple sequence repeat (ISSR) and sequence-related amplified polymorphism (SRAP) have been used to assess the genetic diversity and genetic differentiation in different population of *E. henryi* (Li and Jin 2004, 2006; Zhang et al. 2007; Niu et al. 2017). Recently, Zhang et al. (2016) analyzed the genetic diversity in *E. henryi* based on cpDNA chlorotype and nrITS ribotype frequencies. Simple sequence repeats (SSRs) are reported to be highly effective DNA markers in population genetic studies due to their multi-allelic nature, high reproducibility and co-dominant inheritance (Powell et al. 1996). Hence, they are widely used for accessing the genetic diversity and genetic structure in various plants and animals species (Guo and Gao 2017; Ali et al. 2019; Agarwal and Khan 2019; Belalia et al. 2019; Nakintu et al. 2020; Rohini et al. 2020; Shahabzadeh et al. 2020). To date, no studies have used SSR markers for accessing the genetic diversity in *E. henryi* populations. Therefore, the present study investigated (1) the level and structure of genetic diversity in natural population of *E. henryi* using SSR markers, and (2) to provide suggestions for future conservation strategies.

Materials and methods

Plant materials

Leaves of 124 samples were randomly collected from Jinggangshan National Nature Reservation (Ji′anshi), Wuyishan National Nature Reservation (Yanshanxian), Ganjiangyuan National Nature Reservation (Shichengxian), Lushan National Nature Reservation (Jiujiangshi), and areas known to have natural populations namely Huashan (HS) and Daweishan (DWS), where habitat was destroyed severely in Jiangxi Province (Tongguxian and Yifengxian). The sampled trees were $\geq$ 8 m height with a $\geq$ 10 cm diameter at breast height, and a $\geq$ 3 m $\times$ 3 m crown width. Up to twenty plants were randomly sampled at each location and the distance between the sampled trees were, at least 100 m. Details of each population and specimens were recorded (Fig. 1 and Table 1). Voucher

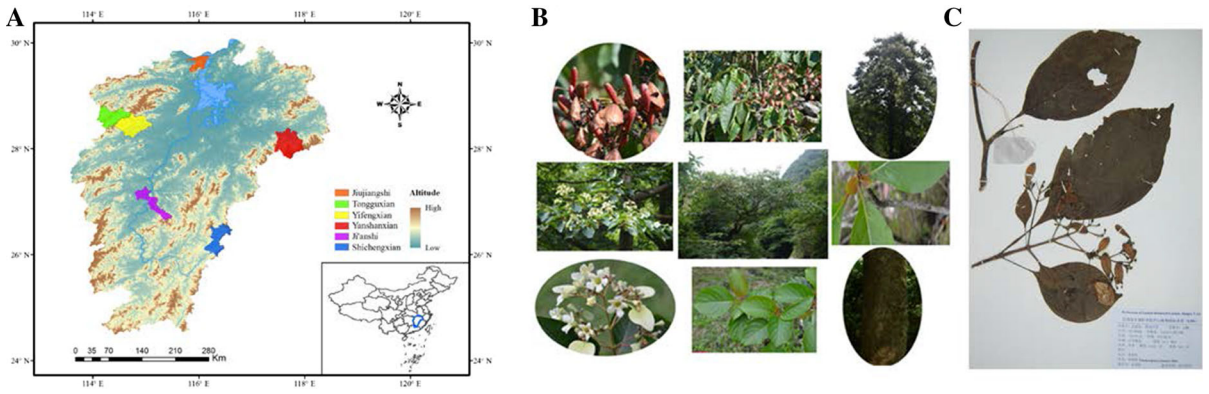

Fig. 1 Six populations and some morphological characteristics of *E. henryi* samples for genetic diversity and structure. A: Location of populations sampled B: morphological characteristics of *E. henryi* C: Voucher specimens

specimens were deposited at the herbarium of Lushan Botanical Garden, China.

Extraction of genomic DNA

Genomic DNA was extracted from leaf tissue using the cetyltrimethylammonium bromide (CTAB) method with some modifications (Doyle and Doyle 1987). The concentration and quality of DNA were defined using a spectrophotometer (Nano Drop 2000, Thermo Scientific). Isolated DNA was diluted to 20 ng/µl and subsequently stored at − 20 °C.

Fluorescent SSR analysis

Ten pairs of SSR primers were selected based on high polymorphism and clear amplification of a single amplicon in preparatory experiment as described by Ma et al. (2012). The primer sequences were synthesized by ABI (Huada Gene) with blue (FAM), green (HEX) and yellow (TMRA) fluorophores (Table 2). The optimal SSR-PCR amplifications were carried out in 20 µl reaction mixtures containing 40 ng template DNA, 0.4 µl dNTP (10 mM), 20 µM forward and reverse primers (20 µM), 1 unit of Taq DNA polymerase, 2 µl buffer (10X) and complete to 20 µl with sterile deionized water. The amplification program was as follows: initial denaturation at 94 °C for 5 min, 35 cycles of denaturation at 94 °C for 30 s, renaturation at 54 °C for 35 s, 40 s extension at 72 °C and a final extension at 72 °C for 3 min. PCR products were diluted by 10 times, mixed with GeneScan 500 LIZ size standard (MCLAB), run on capillary electrophoresis with 3730XL DNA Sequencer (ABI) and analyzed using Genemarker. The fragments were performed using GenemarkerV2.2 software.

Data analysis

SSR data were scored as two alleles per locus distinguished by their size. Observed number of alleles (*Na*), effective number of alleles (*Ne*), Nei's gene diversity (*H*), Shannon's Information index (*I*), observed heterozygosity (H_o), expected heterozygosity (H_e), Gene flow (Nm), Wright's F statistics parameters (F_{is}, F_{it} and F_{st}) gene difference coefficient (G*st*) were computed using POPGENE version 1.32. Allele frequency, number of private alleles, Np Nei' genetic identity (GI), Nei' genetic distances (GD), HWE (Hardy–Weinberg Equilibrium) test, Mantel test and AMOVA were calculated using GenAlEx6.5 (Peakall and Smouse 2012). Null allele frequencies were estimated using the models of Brookfield 1 (Brookfield 1996) methods of package Micro-Checker 2.2.3 (Van Oosterhout et al. 2004). Allelic richness (AR) of the sampled populations were calculated using the rarefaction procedure of HP-RARE 1.0 program (Kalinowski 2005). Polymorphism information content (PIC) was calculated using the program PIC-CALC0.6 (Liu et al. 2015). Based on the Nei genetic distance, a Neighbour-Joining (NJ) evolutionary tree of 124 samples was constructed using POPULATION 1.2 and drawn by iTOL (interactive tree of life) software (Letunic and Bork 2016). Genetic relationships among the genotypes were analyzed by

Table 1 Description of *Emmenopterys henryi* samples collected from different populations in Jiangxi Province

Population code	Number of samples	Habitat	Vegetation type	Latitude(N)/ Longitude(E)	Altitude (m) asl	Accompanying species
HS(1)	11	Midwood Wayside	Moso-Deciduous broad- leaved forest mixed	N:28°32′ E:114°12′	452–754	*Phyllostachys edulis*, *Cyclocarya paliurus*, *Taxus wallichiana* var. *mairei* *Acer davidii, Rubus trianthus*, *Cunninghamia lanceolata*, *Rhododendron latouchea*
DWS(2)	20	Midwood Brook	Bamboo forest Evergreen-Deciduous Broad-leaved forest mixed	N:28°27′ E:114°12′	767–1136	*Lindera reflexa*, *Cunninghamia lanceolata, Taxus wallichiana* var. *mairei*, *Photinia bodinier*, *Platanus* × *acerifolia*, *Acer davidii*, *Phyllostachys edulis*, *Cyclobalanopsis myrsinifolia*, *Cornus macrophylla*, *Cyclobalanopsis glauca*, *Dalbergia hupeana, Kalopanax septemlobus*, *Cerasus serrulata*, *Rubus trianthus*
WYS(3)	14	Midwood Brook	Evergreen-Deciduous broad leaved forest mixed Tea plantation Bamboo forest	N:27°49′ E:117°44′	1192–1438	*Liquidambar formosana*, *Cyclobalanopsis glauca*, *Pterostyrax corymbosus, Cephalotaxus sinensis*, *Padus avium*, *Padus buergeriana*, *Hovenia acerba*, *Rhododendron latoucheae*, *Acer sinense*, *Acer davidii*, *Camellia cuspidate, Tapiscia sinensis*, *Eleutherococcus nodiflorus, Triadica sebifera, Rhus chinensis*, *Viburnum plicatum*, *Sambucus williamsii*, *Glochidion wilsonii*, *Elaeagnus pungens*
JGS(4)	19	Midwood Roadside Open ground Sunny place	Evergreen Broad-leaved forest Evergreen-Deciduous broad leaved forest mixed Confier Broadleaf forest	N:26°37′ E:114°11′	598–883	*Liquidambar formosana*, *Schima argentea, Cunninghamia lanceolata*, *Cornus hongkongensis, Photinia davidsoniae, Hovenia acerba*, *Cornus controversa,Vernicia Montana, Machilus leptophylla*, *Daphniphyllum macropodum*, *Acer davidii*, *Phyllanthus glaucus, Pinus massoniana*, *Metasequoia glyptostroboides*, *Eleutherococcus nodiflorus, Firmiana simplex, Lindera reflexa*, *Morus alba*
GJY(5)	30	Midwood Roadside	Evergreen-Deciduous broad leaved forest	N:26°17′ E:116°27′	832–976	*Eurya hebeclados, Carpinus turczaninowii, Liquidambar formosana, Dendrobenthemia japonica, Lithocarpus harlandii, Ldesia polycarpa, Machilus thunbergii, Cyclocarya paliurus, Alniphyllum fortunei, Rhododendron latoucheae, Eurya japonica, Eurya hebeclados, Rhododendron ovatum, Trachelospermum jasminoides, Smilax china, Lophatherum gracile*
LS(6)	30	Midwood Brook	Coniferous and Broad- leaved mixed forest Bamboo forest	N:29°32′ E:115°58′	824–1060	*Cunninghamia lanceolata, Liquidambar formosana, Padus buergerianna, Hovenia acerba, Litsea elongata, Acer palmatum, Pterostyrax corymbosus, Rhododendron mariesii, Rhododendron simsii, Cephalotaxus fortunei, Lophatherum gracile, Phyllostachyspubescens, Platycarya strobilaceaS.et.Z.*

Note HS –Huashan, DWS-Daweishan, WYS-Wuyishan, JGS-Jinggangshan, GJY-Ganjiangyuan, LS-Lushan

Table 2 SSR primer sequences for analyzing the genetic diversity in *Emmenopterys henryi* populations

SSR primer pair	SSR primer sequence (5'-3')	Fluorescence Tag
Emm1	F: TTGGGATGAGAATTGGTATT	5'FAM
	R:ACACACACACACAGAGAGAGAG	
Emm2	F: GTCGCATAACCATGACCAAG	5'FAM
	R: ACACACACACACAGAGAGAGAG	
Emm3	F: TTGCACGATTGAAGAACC	5'TMRA
	R: ACACACACACACAGAGAGAGAG	
Emm4	F: ATCAAGCAGGAAAGACGG	5'HEX
	R: ACACACACACACAGAGAGAGAG	
Emm5	F: TAGTGGACTTGGCAGGAA	5'FAM
	R: ACACACACACACAGAGAGAGAG	
Emm6	F: AGGCTTCGCACAAGTTTA	5'FAM
	R: ACACACACACACAGAGAGAGAG	
Emm7	F: GGTCCCAGATTAACAACT	5'HEX
	R: ACACACACACACAGAGAGAGAG	
Emm8	F: GATGGGGATTGCCAGAAC	5'FAM
	R: ACACACACACACAGAGAGAGAG	
Emm9	F: AGGATGAACTGAGGGATG	5'FAM
	R: ACACACACACACAGAGAGAGAG	
Emm10	F: GCGTTAATAGCAATCAGG	5'FAM
	R: TCTCTCTCTCTCACACACACAC	

principal component analysis (PCA) in FactoMineR packages (Sebastien et al. 2008).

STRUCTURE (Pritchard et al. 2000) software was used to infer six *E. henryi* population structure based on SSR fluorescent markers. Ten independent runs were made with values of K set from 1 to 6, with three iterations for each value of K. The length of the burn-in period was set at 100,000 and the number of Markov Chain Monte Carlo (MCMC) repeats after burn-in was set at 100,000. Evanno's method (Evanno et al. 2005) was implemented with STRUCTURE HARVESTER (Earl and Vonholdt 2012), to determine the value of estimated Ln probability of data, Ln P(K) and to get the best fit value of K for the data. The significant differences between groups and samples were tested by analysis of molecular variance in GenAlEx 6.5 (Peakall and Smouse 2012).

Results

Genetic diversity analysis

Hardy–Weinberg equilibrium tests were used in the present study to detect deviations from the balance. Eight loci conformed to HWE test and two loci (EMM6 and EMM9) were found to be deviated from HWE, and these two loci were excluded from further population analyses. The main genetic parameters of the samples are presented in Table 3. The eight SSR markers generated a total of 114 alleles. The number of alleles per locus ranged from 8 to 21 with an average of 14.3 per locus. H_o and H_e ranged from 0.3468 to 0.7034 and 0.7588 to 0.9162, with an average value of 0.5388 and 0.8680, respectively. The frequencies of null alleles ranged from 0.07 to 0.2710. The PIC value spanned from a minimum of 0.7184 (EMM10) to a maximum of 0.9006 (EMM2) locus, with an average of 0.8496 and PIC value were significantly correlated with the number of alleles per locus. In all 8 SSR loci, PIC values were high, indicating that the selected markers were very informative and suitable for the genetic study.

Among the 124 samples from six populations, 114 alleles at 8 neutral SSR loci were detected, with the number of alleles in each population ranging from 46 to 83 (Table 4). The GJY population showed highest genetic diversity based on the maximum value of H_e. Whereas, LS population showed the highest values of *N*, *N*a, *N*e and *I*. However, the lowest values of *N*e, *I*,

Table 3 The characteristics of genetic diversity of the eight SSR loci

Locus	Na	Ne	*I*	H_o	H_e	*H*	*Fis*	*Fit*	*Fst*	Nm	PIC	NAF
EMM1	15	9.6305	2.4197	0.4508	0.8999	0.8962	0.4743	0.5405	0.1260	1.7344	0.8873	0.2119
EMM2	21	10.8989	2.6182	0.3846	0.9121	0.9082	0.6021	0.6470	0.1129	1.9650	0.9006	0.2710
EMM3	14	8.6301	2.3305	0.6777	0.8878	0.8841	0.1631	0.2578	0.1132	1.9591	0.8736	0.0738
EMM4	15	9.8093	2.4380	0.6210	0.9017	0.8981	0.2247	0.3185	0.1210	1.8167	0.8895	0.0982
EMM5	17	6.3501	2.2328	0.6911	0.8460	0.8425	0.2039	0.2866	0.1038	2.1588	0.8290	0.0893
EMM7	16	11.4084	2.5742	0.7034	0.9162	0.9123	0.1580	0.2567	0.1172	1.8834	0.9058	0.0700
EMM8	8	5.4924	1.7993	0.4355	0.8212	0.8179	0.4312	0.4801	0.0860	2.6585	0.7927	0.1855
EMM10	8	4.0937	1.6086	0.3468	0.7588	0.7557	0.4194	0.5561	0.2355	0.8116	0.7184	0.1469
Mean	14.3	8.2892	2.2527	0.5388	0.8680	0.8644	0.3320	0.4157	0.1253	1.7446	0.8496	0.1433

Note Na-Observed number of alleles, Ne-Effective number of alleles, *I*-Shannon'S Information index, H_o-Observed heterozygosity, H_e-Expected heterozygosity, *H*-Nei'S gene diversity, *Fis*-Inbreeding coefficient at the population level, *Fit*-Inbreeding coefficient at total populations, *Fst*-Proportion of differentiation among populations, Nm-Gene flow, PIC-Polymorphism Information Content, NAF-Null allele frequencies

H_o and H_e was dected for the JGS population. The value of PIC was greater than 0.5, which indicates that the studied populations of *E. henryi* are highly polymorphic. Generally, the frequency of private alleles used to assess the amount of gene flow between populations. The number of private alleles varied from 3 to 7 amongst the populations investigated and the value of allelic richness in each population ranged from 5.60 to 7.20 with an average of 6.49.

The results of an AMOVA indicated that genetic variation exist within the studied populations (Table 5), with 59% and 30% intra individual and among individuals variation, respectively. Moreover, 11% of the variation existed among populations. Genetic variation among populations was much lower than that of among individual within populations. The G_{st} and *Nm* were 0.524 and 2.014, respectively. The F_{st} value between populations ranged from 0.052 to 0.118 and with average values 0.085, which was lower than the mean F_{st} (0.22) of outcrossing plants (Table 6). The mean of gene flow between the populations was 2.842, which suggested that there were sufficient genetic exchange to prevent the genetic differentiation made by genetic drift among populations.

Bayesian genetic structure

Analysing the *E. henryi* population structure using STRUCTURE software with an Evanno correction, the optimal cluster (*K*) was identified based on maximum likelihood (Ln *K*) and posterior probability (Delta *K*) values. *K* was tested from one to six with ten replicated runs. The peak of Delta *K* was observed for $K = 2$ (Fig. 2a). Based on SSR markers, 124 samples were divided into two clusters. As shown in Fig. 2b, most samples from GJY population clustered into one group (red gene pool), whereas, samples from other populations clustered into another group (green gene pool). Based on the membership fractions, samples from different populations were categorized as pure or admixture (membership coefficient less than 0.9). The genetic structure pattern showed that there was not significant difference among populations and these results were consistent with POPGENE's results.

Cluster analysis

The cluster analysis of the six populations of *E. henryi* based on Nei's genetic distance found three groups which clustered based on geographical proximity (Fig. 3a). HS population and DWS population were relatively close, and clustered into a small group firstly, then into a group with JGS and finally into a group with WYS population together. Whereas, LS and GJY clustered separately forming two separate groups. NJ tree of 124 samples of *E. henryi* based on genetic distance is shown in Fig. 3b. Samples were also divided into three groups (blue, green and orange), but there were some cross-mixing within

Table 4 Genetic diversity parameters revealed by SSR markers on *E. henryi* populations

Population code	N	Na	Ne	Np	I	PIC	H_o	H_e	AR
HS	46	5.8 ± 0.453c	3.958 ± 0.406bc	3	1.492 ± 0.122b	0.681	0.468 ± 0.092bc	0.714 ± 0.050ab	6.73
DWS	68	8.5 ± 0.627ab	5.029 ± 0.463ab	7	1.798 ± 0.091ab	0.761	0.464 ± 0.060bc	0.788 ± 0.021a	5.75
WYS	64	8.0 ± 0.906bc	5.234 ± 0.788ab	5	1.760 ± 0.141ab	0.747	0.436 ± 0.043bc	0.775 ± 0.033a	5.60
JGS	61	7.6 ± 0.653bc	3.222 ± 0.300c	5	1.458 ± 0.112b	0.630	0.345 ± 0.059c	0.658 ± 0.050b	6.91
GJY	72	9.0 ± 1.035ab	5.701 ± 0.673a	6	1.874 ± 0.127a	0.779	0.678 ± 0.083a	0.805 ± 0.024a	6.74
LS	83	10.4 ± 0.999a	6.004 ± 0.774a	6	1.918 ± 0.164a	0.771	0.632 ± 0.079ab	0.790 ± 0.053a	7.20
Mean	65.7	8.208 ± 0.373	4.858 ± 0.271	5.3	1.717 ± 0.056	0.728	0.504 ± 0.032	0.755 ± 0.017	6.49

Note HS-huashan, DWS-daweishan, WYS-wuyishan, JGS-jinggangshan, GJY-ganjiangyuan, LS-lushan; N-Total number of observed alleles; Na-Observed number of alleles, Ne-Effective number of alleles, Np-Number of private alleles, H_o-Observed heterozygosity, H_e- Expected heterozygosity, *I*-Shannon'S Information index, Data expressed as mean ± SE, AR-Allelic richness

Table 5 Analysis of molecular variance (AMOVA) and genetic differentiation among six *E. henryi* populations

Source of variance	*Df*	*SS*	*MS*	Variance component	Total Variance	*Dst*	*Hs*	*Ht*	G_{st}	N_m
Among-population	5	104.309	20.862	0.412	11	–	–	–	–	–
Within-population	118	503.767	4.269	1.080	30	–	–	–	–	–
Intra-Individual	124	261.5	2.109	2.109	59	–	–	–	–	–
Total	247	869.577		3.601	100%	–	–	–	–	–
Mean	–	–	–	–	–	0.114	0.339	0.414	0.524	4 2.014

Note *df*-Degree of freedom, *SS*-Sum of squares, *MS*-Mean square, *Dst*-Proportion of differentiation among populations, *Hs*-Inbreeding coefficient at the population level, *Ht*-Inbreeding coefficient at total populations, G_{st}-gene difference coefficient, N_m-gene flow

Table 6 Fixation indices(F_{st}) and gene flow among six *E. henryi* populations

Population code	HS	DWS	WYS	JGS	GJY	LS
HS	***	0.079	0.084	0.095	0.087	0.098
DWS	2.921	***	0.074	0.090	0.079	0.052
WYS	2.734	3.130	***	0.096	0.075	0.062
JGS	2.393	2.524	2.342	***	0.116	0.118
GJY	2.633	2.908	3.096	1.907	***	0.066
LS	2.312	4.520	3.787	1.876	3.550	***

Note HS-huashan, DWS-daweishan, WYS-wuyishan, JGS-jinggangshan, GJY-ganjiangyuan, LS-lushan

Fixation indices (above diagonal) and gene flow (below diagonal)

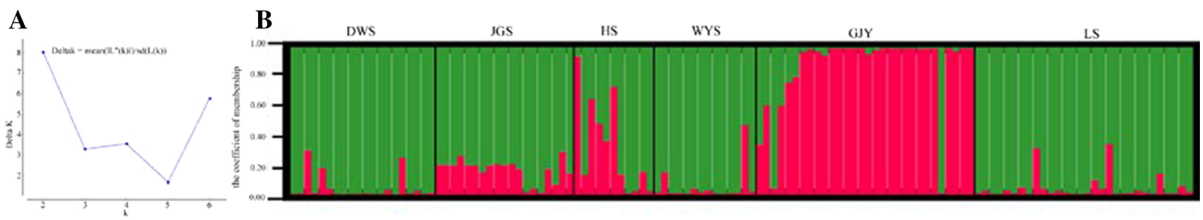

Fig. 2 Population stratification from STRUCTURE analysis based on consensus across 10 replications when K = 2. △ *K* estimates of the posterior probability distribution of the data for a given *K*. B. Bar plots representing population structure Vertical bars represent each sample and length of the colored bar shows the estimated proportion of membership when K = 2 proportion of membership

and among populations, which indicated that there were more genetic variation and the variation pattern was more complex.

Principal component analysis

A Principal component analysis (PCA) based on the genetic similarity matrix was performed further to understand the genetic relationships among the samples. The PCA analysis grouped the 124 samples into three major groups (Fig. 5). The two informative PCA components explained 7.86% of the total variation, and major variation came from PC1 and PC2, shared 4.39% and 3.47% variation, respectively. The degree of gene exchange was high among the six populations (Fig. 5) and overall, the results of PCA were consistent with the results of the population genetic structure and cluster analysis.

The genetic and geographical distance of six *E. henryi* populations were examined by Mantel test

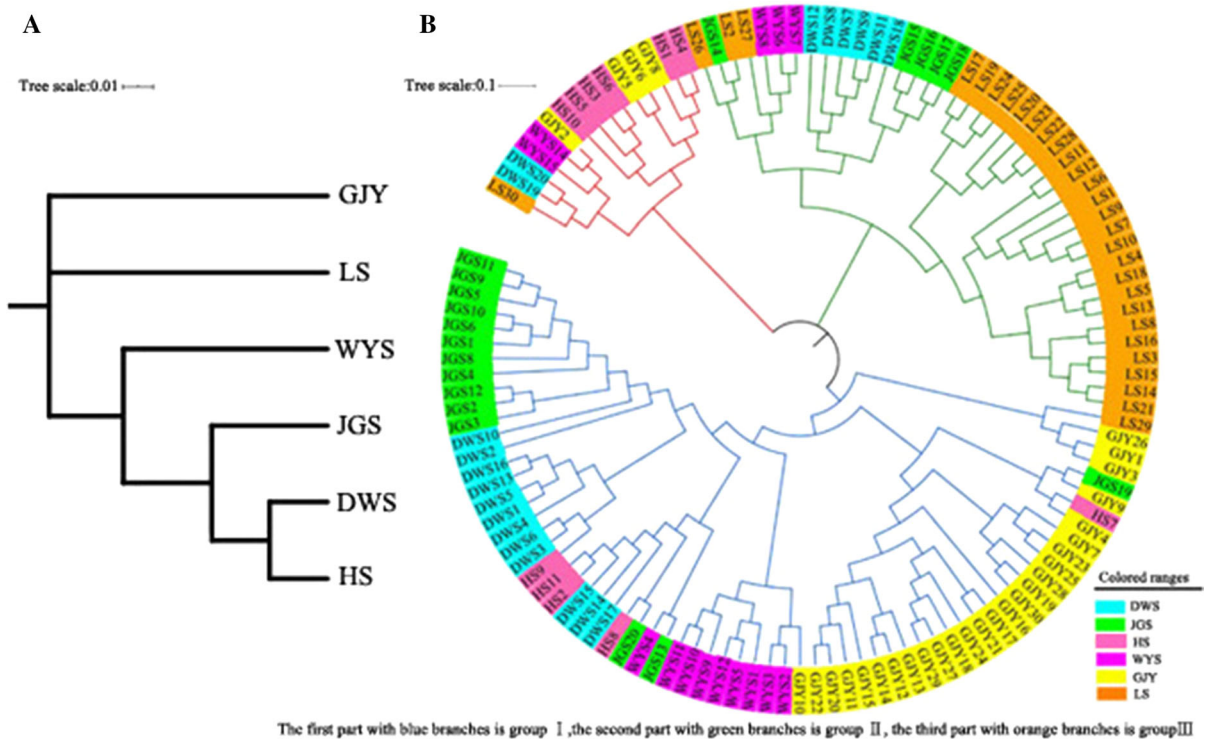

Fig. 3 The Neighbor-Joining phenogram of the 124 *E. henryi* samples based on Nei' genetic distance of SSR data. A: 6 populations of *E. henryi*; B: 124 samples of *E. henryi*

using GenAlEx 6.5 software. The result showed that genetic distance was positively correlated with geographical distance ($r = 0.1844$, $p = 0.2$) (Fig. 4), which were similar to the results of NJ tree and PCA analysis.

Discussion

Understanding the genetic diversity is essential for formulating conservation strategies of threathened species because it effects a population's ability to adopt to the changing environmental conditions (Kirk and Freeland 2011; Ilves et al. 2013). Usually, it is assumed that rare and endangered plant species have low levels of genetic diversity. Therefore, assessing

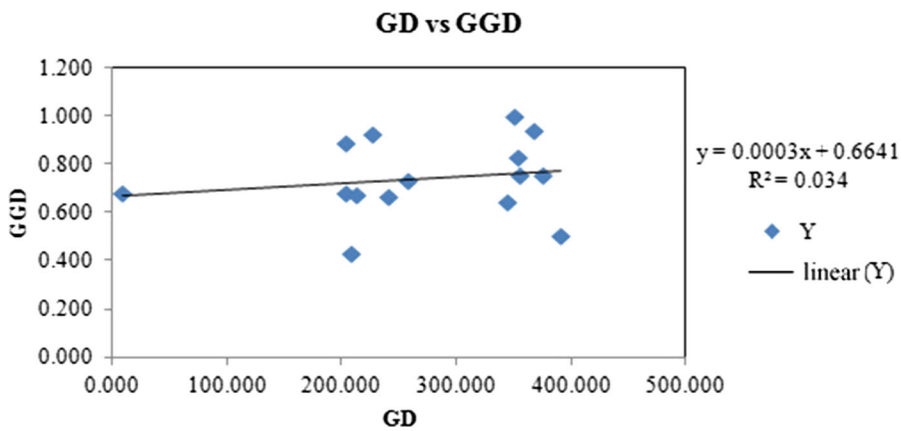

Fig. 4 The correlation between genetic distance and geographic distance for *E. henryi*. GD-geographical distance; GGD-genetic distance

the genetic diversity and genetic structure of rare and endangered plants is highly important for evaluating extinction risk and setting conservation plans. Evaluating genetic diversity could also be helpful in elucidating the evolutionary history of species and in determining if genetic factors/mechanisms which are responsible for controlling the likelihood of a species becoming extinct. The depletion in the population size would lead to decreased adaptability to the environmental changes, which had led to the decrease of genetic diversity of population (Montgomery 2000). In nature, reproductive biology and breeding systems are responsible for affecting the level of genetic diversity (Montgomery 2000).

A H_e = 0.755 is higher than the mean H_e (0.65) for outcrossing plants (Nybom 2004). In the present study, high value of H_e (> 0.65) for found for each population. This higher genetic diversity level of *E. henryi* populations indicates its better ability to adapt to the changing environmental conditions (the plant is a widespread speciess and distributed 430-1600 m asl). According to historical records, *E. henryi* was common in Jiangxi Province and distributed in 10 regions, with the highest population density in HS. However, rapid economical developmental activities (such as expansion of urban area, industrialization and tourism) and overexploitation causes habitat loss and thus it became endangered. It has also been confirmed by the local country annals that the original distribution of *E. henryi* had been replaced by bamboo forests for the WYS population in Jiangxi province. Compared to the native flora, endangered plant has weaker competitiveness and thus can not endure a series of environmental stresses, which ultimalely leads to their population degradation (Fig. 5).

In the present study, a high deviation between H_o and H_e at all loci is due to the presence of null alleles and also due to dropout of some alleles. This may be further lead to an underestimation of population genetic diversity and even a deviation from the HWE balance. Paetkau and Strobeck (1995) reported that the presence of null alleles significantly decreased the value of H_o and H_e. However, Barros et al. (2020) found that high null allele frequencies cause of deviation from HWE in all loci. Generally, PIC value for each marker revealed the genetic informativeness of marker among the accessions. In the present study, high PIC values indicated that SSR markers are very informative and suitable for the genetic study. Primer pair (Emm2) had the highest diversity, number of alleles and PIC, and was the most informative locii. In the present study, the value of H_e ranged between 0.658 and 0.805 and were higher than the mean H_e (0.65) of outcrossing plants (Nybom 2004), indicating the presence of higher genetic diversity in *E. henryi*. The ranking of population genetic diversity based on SSR markers was as follows: GJY > LS > DWS > WYS > HS > JGS. The genetic diversity level was high in all the studied populations of *E.henryi*. Usually, large samples are expected to have more alleles than smaller samples (Kalinowski 2005). Allelic richness of the LS and GJY populations were higher than other populations, however, for the HS population with minimum number of plants, the value of allelic richness was above average. In several studies assessing genetic diversity it was shown that H_e and allelic richness was related to the number of accessions (Montgomery et al. 2000; Zhang et al. 2007). Both GJY and LS population have higher number of plants and thus had higher level of genetic diversity. Whereas, HS population had a relatively low level of genetic diversity. This conclusion needs to be verified by a large number of populations and individuals in order to confirm this pattern.

Wright (1965) proposed that F_{is}, F_{it} and F_{st} indices to measure the degree of inbreeding genetic differentiation in populations. The degree of deviation from HWE is more obvious with the increasing value of F_{is}. F_{st} is usually concern with degree of genetic differentiation or inbreeding coefficient and thus it could be used for evaluating the genetic diversity among populations. In the present study, F_{st} value of six populations was far less than the mean F_{st} (0.22) of outcrossing plants (Nybom 2004). Moreover, the genetic variation among *E. henryi* populations was very small and the degree of genetic differentiation among the populations were far lower than that among individuals within the populations. This genetic differentiation amongst *E. henryi* populations might be due to human selective pressures and/or barriers to gene flow. Most of the genetic variation in *E. henryi* was maintained within populations as shown by AMOVA analysis and these results are in consistent with that in woody plants (Liu et al. 2012). Similar to our results, Xiong et al. (2006) found low genetic variations among the four populations of *E. henryi* from the Shennongjia region. Zhang et al. (2007) found higher genetic variations (74.76%) among nine

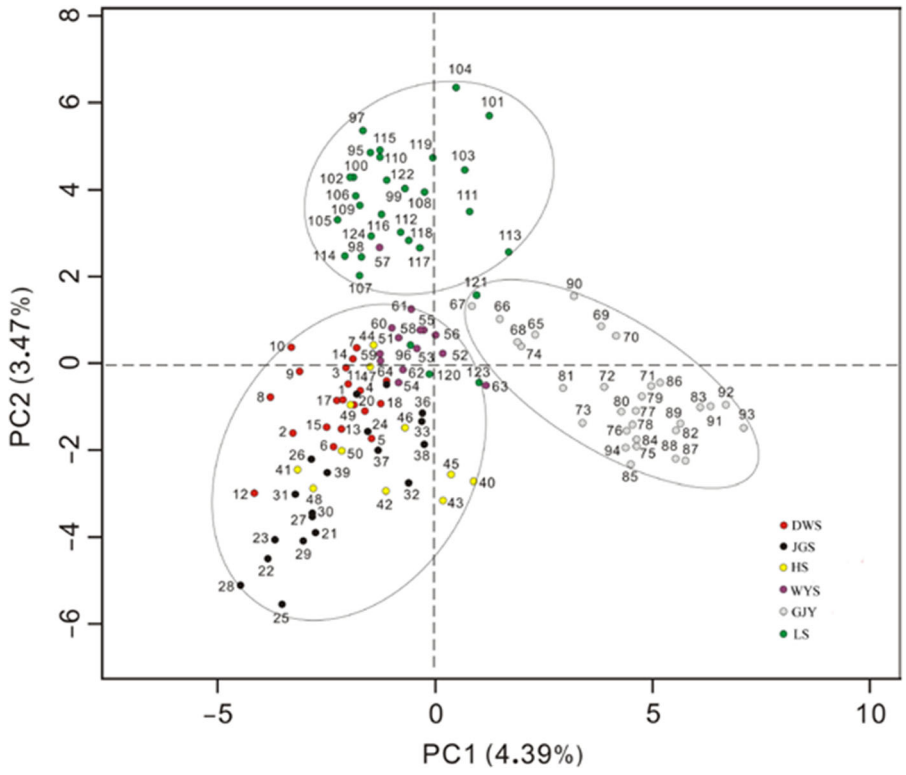

Fig. 5 PCA analysis of the 124 *E. henryi* samples based on SSR markers

populations of *E. henryi*. The low genetic genetic variations among the six studied populations in present case might be due to the limited number of sampled populations. Therefore, population size and its representatives need to be considered because it may affect the genetic structure of population. The highest F_{st} (0.118) value was between LS and JGS, indicates the lower probability of inbreeding. However, the lower F_{st} value (0.052) between LS and DWS populations indicates the higher probability of inbreeding (Li 2017). The size of gene flow affects the genetic structure of population, and it is a key factor for predicting the impact of environment change, human disturbance and population isolation. The Nm value was moderate (2.014), which indicates that the genetic structure of *E. henryi* among population is relatively stable. The stability in population genetic structure is considered to be important because the gene flow is critical for maintain population stability in the future (Wu et al. 2017). The gene flow among populations is mainly accomplished via the transmission of seeds and pollen dispersal. The mean of gene flow value was 2.808, which is comparatively high among the studied populations. However, the highest gene flow was between LS and DWS population (4.520), indicating the higher probability of inbreeding.

The genetic structure and the level of genetic differentiation mainly affected by the various factors such as evolutionary history of plants, habitat changes, genetic drift, bottleneck effects, natural selection, mating systems, gene flow, and so on and these factors have importance in order to establish the effective protection strategy of endangered plants (Li 2018). Differences in environmental factors such as altitude, temperature, etc. also affect the genetic structure of a species (Qian et al. 2014). Bayesian genetic structure analysis detected the complexed admixture among populations and in the present study gene exchange was found frequent among populations. Results of cluster analysis and PCA analysis showed similar trends. For instance, four population namely, HS, DWS, JGS and WYS were clustered in one group, whereas, LS and GJY clustered separately and form two separate group. Nevertheless, the result of clustering of 124 samples had different from that of PCA

analysis. Based on the results of structure, cluster and PCA analysis, we found that the studied samples were obviously mixed. The presented results also showed major degree of shared alleles among populations. Moreover, the higher level of gene flow (2.842) prevented gene differentiation that caused by genetic drift. Pollen and seeds are the major components of gene flow among populations and *E. henryi* is insect pollinated species and mainly reproduces by seeds (Cheng 2008).

We found that positive correlation between genetic distance and geographical distance, which indicate that genetic structure between populations can be analyzed and understood in terms of geographic proximity and similar habitat conditions (vegetation type, etc.). Two mountain, namely Dawei and Biandan (Yichun Tonggu county, Jiangxi province) separate the HS and DWS population, and although these populations are nearest in terms of distance, the frequency of gene exchange between these two populations was much lower than that between LS and DWS population. These results indicate that these mountains may act as a barrier to gene exchange among the HS and DWS populations. In addition to geographical distance, reproductive characteristics of the study species also played an important role in determining the genetic structure (Song et al. 2013; Song et al. 2013). During field studies, we observed very few seedlings of *E. henryi* in the natural population and those recorded in situ preferred valley situations. Moreover, this is insect pollinated species, and the winged seeds are very light and therefore they can dispesed to the long distance by wind. The LS and GJY populations were located far from each other geographically (392 km), however, results suggested that inbreeding exit between these populations. Generally, inbreeding may raise homozygote proportion and therefore, the possibility of these population have higher extinction risk due to existence of inbreeding and erosion of genetic diversity.

Generally, narrow distributions increased the probability of genetic isolation, genetic drift and lack of genetic mutations (Chen et al. 2014). We found that *E. henryi* had higher genetic diversity and results are consistent with the findings of Qi et al. (2012), who reported that wide spread species generally had higher genetic diversity. Therefore, further studies are needed to sampled more population of *E. henryi* from different geographical locations in order to capture all the possible genetic diversity for long term conservation and utilization. According to the local inhabitants, in the past the DWS population have the highest number of individuals of *E. henryi* in Jiangxi province. However, this population was severly depleted in recent past due to habitat alteration by construction of new roads and other anthropogenic activities. Other studied population of *E. henryi* may also faced similar recent challenges. Therefore, it is imperative to conserve all the populations in order to conseve the available genetic diversity of *E. henryi* germplasm resources. *E. henryi* is reported to be distributed in heterogenous habitat and environmental conditions (Yang and Zhang 2007), which indicates their adaptability to different environmental conditions. The higher adaptability of specis have been linked to the species genetic diversity i.e., adaptability is stronger when genetic diversity is higher (Zhang 2016). Maintaining the higher genetic diversity is core of plant genetic resource conservation program (Ming et al. 2013).

Habitat destruction and poor natural regeneration are responsible for threatening natural populations of *E. henryi* and therefore they are now restricted to only few location. Both in situ and ex situ efforts are needed to conserve the raming population of *E. henryi.* In situ conservation strategy could be used for the GJY population which has the highest genetic diversity among the investigated populations.

Acknowledgements This work was supported by the National Natural Science Foundation of China (No. 31460078), Natural Science Foundation Project of Jiangxi Province (20122BAB214031), Major National Science & Technology Specific Projects of China (2013FY11150-2-3) and Key Ministry Projects of Chinese Academy of Sciences (KFJ-3 W-N01). We are also very grateful to Dr. Syd Ramdhani Srinivasan, University of KwaZulu-Natal (UKZN),South Africa, for his critical revision of a final version of themanuscript.

Author contributions YL Niu conceived and designed the experiment; YS Peng, XH Zhan,Y Gao, ZY Zhang and W Hu performed the experiments; WX Chen, AB, MZ Song and ZJ Yu analyzed the data; YL Niu, WX Chen, MZ Song, ZJ Yu and AB wrote the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Agarwal S, Khan S (2019) Genetic diversity of 18 male and 18 female accessions of Jojoba [*Simmondsia chinensis* (link) Schneider] using EST-SSRs. Meta Gene 19:134–141

Ali A, Pan YB, Wang QN, Wang JD, Chen JL, Gao SJ (2019) Genetic diversity and population structure analysis of *Saccharum* and *Erianthus* genera using microsatellite (SSR) markers. Sci Rep 9:395

Barros J, Winkler FM, Velasco LA (2020) Assessing the genetic diversity in Argopecten nucleus (Bivalvia: Pectinidae), a functional hermaphrodite species with extremely low population density and self-fertilization: Effect of null alleles. Ecol Evol 10(9):3919–3931

Belalia N, Lupini A, Djemel A, Morsli A, Mauceri A, Lotti C, Khelifi-Slaoui M, Khelifi L, Sunseri F (2019) Analysis of genetic diversity and population structure in Saharan maize (*Zea mays* L.) populations using phenotypic traits and SSR markers. Genet Resour Crop Evol 66:243–257

Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. Mol Ecol 5:453–455

Chen C, Qi ZC, Xu XH, Comes HP, Koch MA, Jin XJ, Fu CX, Qiu YX (2014) Understanding the formation of Mediterranean-African disjunctions: evidence for Miocene climate-driven vicariance and recent long-distance dispersal in the Tertiary relict *Smilax aspera* (Smilacaeae). New Phytol 204:243–255

Cheng DM (2008) Study on the pollination biology of the state key protected species. Master's thesis. Henan Agricultural University

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Earl DA, vonHoldt BM (2012) Structure harvester: a website and program for visualizing structure output and implementing the Evanno method. Conserv Genet Resour 4:359–361

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Guo LN, Gao XF (2017) Genetic diversity and population structure of *Indigofera szechuensis* complex (Fabaceae) based on EST-SSR markers. Gene 624:26–33

Guo LJ, Li M (2009) Fractal analysis of the distribution pattern of an endangered plant, *Emmenopterys henryi* population in Wuyi mountain. Acta Bot Bor-Occid Sin 29:0867–0873

Guo LJ, Lin LW, Xiu WH, Wang AP (2011a) Characteristics of reproductive modules of *E henryi* natural population in Wuyi mountain. J Northwest Forest Univ 26(4):18–22

Guo N, Yang SH, Ge WY, Ge H (2011b) Phenotypic diversity of natural populations of Rosa laxa Retz in Tianshan Mountains of Xinjiang. Acta Hort Sin 38(3):495–502

Hu HQ, Cui TL (2011) Utilization value and cultivation techniques of *Emmenopterys henryi*. Forest Product Spec China 1:10. https://doi.org/10.13268/j.cnki.fbsic.2011.06.044

Ilves A, Lanno K, Sammul M, Tali K (2013) Genetic variability, population size and reproduction potential in *Ligularia sibirica* (L.) populations in Estonia. Conserv Genet 14:661–669

Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. Mol Ecol Notes 5:187–189

Kirk H, Freeland JR (2011) Applications and implications of neutral versus non-neutral markers in molecular ecology. Int J Mol Sci 12:3966–3988

Letunic I, Bork P (2016) Interactive tree of life (iTOL) V3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 8(44):242–245

Li YY (2017) Studies on conservation genetics of Chinese endemic species*Luculia yunnanensis* Hu. Ph. D. Chinese Academy of Forestry

Li TQ (2018) Conservation biology of the plant species with extremely small population *Rhododendron longipedicellatum*. Master Degree, Chinese Academy of Forestry

Li JM, Jin ZX (2004) Optimization of the RAPD conditions of *Emmenopterys henryi* and prrmary study on the genetic diversity. J. Fujian For Sci-Tech 31(2):36–40

Li JM, Jin ZX (2006) Optimization and screening of ISSR reaction system in endangered plant *Emmenopterys henryi*. J Anhui Agricult Univ 33(4):458–461

Liu YF, Xing M, Zhao W, Fan RJ, Luo S, Chen X (2012) Genetic diversity analysis of *Rhododendron aureum George* (Ericaceae) located on Changbai Mountain using ISSR and RAPD markers. Plant Syst Evol 298(5):921–930

Liu QW, Song Y, Liu L, Zhang MY, Sun JM, Zhang SL, Wu J (2015) Genetic diversity and population structure of pear (*Pyrus spp*) collections revealed by a set of core genome-wide SSR markers. Tree Genet Genomes 11:128

Lu W, Tang ZS, Zheng ZJ, Ye GH, Chen SW, Wang YQ, Ye D, Liu P, Chen JH (2018) The community composition and structural characteristics of *Emmenopterys henryi* forest, an endangered species in Gutianshan, zhejiang province. Ecol Environ Sci 27(6):1052–1059

Ma WW, Zhai SN, Zhang YH, Qiu YX (2012) Development and characterization of microsatellite markers for *Emmenopterys henryi* (Rubiaceae), a rare tree from china. Am J Bot 99(4):179–181

Ma MW, Wu YH, Zhang Y, Kang HJ, Chen ZL, Liu P (2018) Sprouting as a survival strategy for non-coniferous trees: Relation to population structure and spatial pattern of *Emmenopterys henryi* (*Rubiales*). Acta Ecola Sin 39:1–8

Man JS, Fang YP, Liu SX, Peng ZL (2008) The resource and protection of *Emmenopterys henryi* in Qizimei-shan National Nature Preserve. China. J Huanggang Normal Univ 28(3):44–46

McNamara W (2005) Emmenopterys henryi. Pacific Horticul 4(66):39–40

Ming L, Zhong Z, Xingjun M, Jingjing Z (2013) Genetic diversity and population structure of Siberian apri-cot (*Prunus sibirica* L.) in China. Int J Mol Sci 15:377–400

Montgomery ME, Woodworth LM, Nurthen RK, Gilligan DM, Briscoe DA, Frankham R (2000) Relationship between population size and loss of genetic diversity: comparisons of experimental results with theoretical predictions. Conserv Genet 1(1):33–43

Nakintu J, Albrecht C, Muller MC, Kagoro-Rugunda G, Andama M, Olet EA, Lejju JB (2020) Exploring the genetic diversity of jackfruit (*Artocarpus heterophyllus* Lam.) grown in Uganda based on SSR markers. Genet Resour Evol 67:605–619

Niu YL, Peng YS, Song L, Zhou SX, Wei ZX, Song MZ, Huang J, Zhang PX, Gao PX, Du J, Yu ZJ (2017) Establishment of reaction system and primer screening for SRAP marker on endangered plant *Emmenopterys henryi*. Mol plant breed 15(8):3136–3144

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol Eco 13(5):1143–1155

Paetkau D, Strobeck C (1995) The molecular basis and evolutionary history of a microsatellite null allele in bears. Mol Ecol 4:519–520

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparision of RFLP, RAPD, AFLP and SSR (microsatelite) markers for germplasm analysis. Mol Breeding 2:225–238

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 7:574–578

Qi XS, Chen C, Comes HP, Sakaguchi S, Liu YH, Tanaka N, Sakio H, Qiu YX (2012) Molecular data and ecological niche modeling reveal a highly dynamic evolutionary history of the East Asian Tertiary relict *Cercidiphyllum* (Cercidiphyllaceae). New Phytol 196:617–630

Qian S, SaitoW Mimura M, Kaneko S, Isagi Y, Mizumachi E, Mori AS (2014) Asymmetric gene flow and the distribution of genetic diversity in morphologically distinct *Abies mariesii* populations in contrasting eco-habitats. Plant Ecol 215:1385–1397

Rohini MR, Sankaran M, Rajkumar S, Prakash K, Gaikwad A, Chaudhury R, Malik SK (2020) Morphological characterization and analysis of genetic diversity and population structure in *Citrus* X *jambhiri* Lush. Using SSR markers. Genet Resour Crop Evol. https. https://doi.org/10.1007/s10722-020-00909-4

Sebastien L, Julie J, Francois H (2008) FactoMineR: an R package for multivariate analysis. J Statist Soft Ware 25:1–18

Shahabzadeh Z, Mohammadi R, Darvishzadeh R, Jaffari M (2020) Genetic structure and diversity analysis of tall fescue populations by EST-SSR and ISSR markers. Mol Bio Rep 47:655–669

Song J, Li SF, Liu LN, Li SF, Xie WJ, Guan WL (2013) Phenotypic diversity of natural populations of Michelia yunnanensis. Acta Bot Boreal-Occident Sin 33(2):272–279

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19:395–420

Wu FQ, Shen SK, Zhang X, Yang GS, Wang YH (2017) Inferences of genetic structure and demographic history of *Rhododendron protistum* var giganteum—the world's largest Rhododendron using microsatellite markers. Flora 233:1–6

Xiong D, Chen FJ, Li XP, Liang HW, He ZQ (2006) Genetic diversity of endangered *Emmenopterys* in Shennongjia Region of Hubei Province. Acta Bot Bor-Occid Sin 26(6):1272–1276

Yang KJ, Zhang XP (2007) Research advances of rare plant *Emmenopterys henryi* Oliy. Chin Wild Plant Resour 26(2):1–4

Zhang YH (2016) Phylogeography and landscape genetics of *Emmenopterys henryi* (Rubiaceae), a Tertiary relict species across China. Zhejiang University, Ph.D.Degree

Zhang WB, Jin ZX, Li JM (2007) Analysis of genetic diversity in the natural population of *Emmenopterys henryi* using RAPD. J Zhejiang Univ 33(1):61–67

Zhang YH, Wang IJ, Comes HP, Peng H, Qiu YX (2016) Contributions of historical and contemporary geographic and environmental factors to phylogeographic structure in a Tertiary relict species, *Emmenopterys henryi* (Rubiaceae). Sci Rep 6:24041

Zhou HB (2011) Studies on the Chemical Constituents and Bioactivity of *Emmenopterys henryi* Oliv. Master's thesis. The Second Military Medical University