© The Author(s), 2020. Published by
Cambridge University Press on behalf of NIAB
ISSN 1479-2621

Plant Genetic Resources: Characterization and Utilization (2020) **18**(6); 417–426
doi:10.1017/S1479262120000441

# Genetic diversity of *Juglans sigillata* Dode germplasm in Yunnan Province, China, as revealed by SSRs

Shaoyu Chen[1]*, Tao Wu[1], Liangjun Xiao[2], Delu Ning[2] and Li Pan[2]

[1]*Yunnan Provincial Key Laboratory of Cultivation and Exploitation of Forest Plants, Yunnan Academy of Forestry and Grassland, Kunming 650204, Yunnan, China and* [2]*Economic Forest Institute, Yunnan Academy of Forestry and Grassland, Kunming 650204, Yunnan, China*

**Received 29 April 2020; Accepted 22 October 2020 – First published online 17 November 2020**

## Abstract

Iron walnut, *Juglans sigillata* Dode, restricted to southwestern China, has its centre of distribution in Yunnan Province which has a varied climate, geography and rich plant diversity. Yunnan contains abundant *J. sigillata* germplasm. In this study, a provincial-scale set of walnut germplasm resources (14 populations comprising 1122 individuals) was evaluated for genetic diversity based on 20 simple sequence repeat (SSR) loci. The number of SSR alleles per locus ranged from 7 to 27, with an average of 17.55. Mean allelic richness and mean private allelic richness ranged from 3.40 to 4.62 and 0.11 to 0.36, with average of 3.93 and 0.30, respectively. Expected heterozygosity ($H_e$) varied from 0.26 to 0.78, with an average of 0.57. Polymorphism information content ranged from 0.22 to 0.79, with an average of 0.57. Genetic differentiation ($F_{ST}$) was 0.05, indicating that only 5% of total genetic variability was inter-populational, a finding supported by an analysis of molecular variance and STRUCTURE analysis. Relatively high gene flow ($N_m = 6.70$) was observed between populations. A unweighted pair-group method with arithmetic analysis classified the 14 populations into two major groups. Mantel testing uncovered a significant correlation between geographic distance and genetic distance ($r = 0.33$, $P = 0.04$). Overall, the research revealed a moderately high level of genetic diversity in the germplasm and low genetic differentiation among populations, which showed great potential for further development and exploitation of this resource.

**Keywords:** genetic differentiation, genetic diversity, *Juglans sigillata* Dode, SSR markers, UPGMA clustering

## Introduction

Walnut (*Juglans* spp., Juglandaceae) comprises an economically important group of dual-purpose nut and timber trees worldwide. Nuts are included as one of the important elements of healthy dietary guidelines because they are considered a good dietary source of nutrients including unsaturated fatty acids. The walnut production of world was 3,662,507 tons in 2018, while the production of China reached 1,586,367 tons (FAOSTAT, 2012). Eight walnut species are found in China, two of which, common walnut: *Juglans regia* and iron walnut: *Juglans sigillata*, are widely cultivated species available as walnut breeding resources. *J. regia* is distributed globally, whereas *J. sigillata* is restricted to southwestern China (Wang *et al.*, 2007; Feng *et al.*, 2018; Yuan *et al.*, 2018). *J. sigillata* is cultivated in Yunnan Province, where the native species is economically important.

Yunnan Province, a distinct alpine region with various climatic and geographical features harbouring rich plant diversity, is a centre of origin and endemism of *J. sigillata* (Xi and Zhang, 1996). *J. sigillata* can be found in all 127 counties of the province at elevations of 850–2900 m (Lu, 2009). The unique geography and climate of Yunnan are responsible for superior agronomic properties and valuable traits,

*Corresponding author. E-mail: sherry9872@163.com

https://doi.org/10.1017/S1479262120000441 Published online by Cambridge University Press

such as high oil and protein contents, kernel quality and strong disease like cold and drought resistances (Xi, 1987), that have significant potential value for variety improvement.

Genetic diversity results from the long-term existence, evolution and eco-adaption of a species; consequently, assessment of the genetic diversity of a species greatly aids its utilization and conservation. Most walnut trees are open-pollinated, which would lead to accumulate abundant genetic diversity. As a result of long-term environmental pressures and increasing human activities during the past decades, however, walnut resources have been depleted, thereby reducing genetic diversity. Although analysis on characteristics and economic quality of walnut nut in Yunnan has been done (Lu, 2009) and molecular-based genetic analysis on some germplasm resource was finished (Ning *et al.*, 2011; Xiao *et al.*, 2018), the overall evaluation on the genetic diversity has not been systematically evaluated.

Several molecular marker techniques have been applied to examine the genetic diversity of walnut. Examples include methods based on randomly amplified polymorphic DNA (Wu *et al.*, 2000; Nazeer *et al.*, 2012), restriction fragment-length polymorphism (Fjellstrom and Parfitt, 1994), amplified fragment-length polymorphism (Bayazit *et al.*, 2007; Chen *et al.*, 2008; Wang *et al.*, 2011), inter-simple sequence repeats (SSRs) (Potter *et al.*, 2002; Li *et al.*, 2011; Liu *et al.*, 2017), and SSRs (Wang *et al.*, 2007, 2008; Shah *et al.*, 2018; Shamlu *et al.*, 2018). Lack of systematic study of the genetic diversity of a germplasm resource seriously restricts its genetic improvement and further utilization. In this paper, we therefore, drew on the work of the previous resource survey and analysed the SSR-based genetic diversity and genetic structure of *J. sigillata* germplasm sampled from 14 prefectures (cities) of Yunnan. The results should serve as a valuable reference for the conservation and breeding programmes.

## Materials and methods

### Plant material

A total of 1122 iron walnut individuals sampled from 14 prefectures of Yunnan were used for the analysis. All of these individuals were derived from wild plants growing naturally in the mountains or in villages, except for samples from KM (the capital of Yunnan). To better reflect geographic characteristics of districts and the effects of human activities on germplasm, we considered individuals from the same administrative district to be one population; consequently, 14 populations (namely, 14 administrative districts) were analysed (Fig. 1). In each population, 18–221 naturally growing, adult trees of seedling origin (>60 cm diameter at breast height) were sampled; to reduce the probability of sampling from family clusters, individuals separated from each other by a distance of more than 50 m were selected. Fresh mature leaves were randomly sampled from each individual, stored in plastic bags, and taken to the laboratory for SSR analysis. Locations and sampling sizes of the 14 populations in the study are listed in Table 1.

### DNA extraction

Genomic DNA was extracted from stored leaves using an improved cetyltrimethylammonium bromide method (Yang *et al.*, 2005). DNA quality and concentration were determined by gel electrophoresis and ultraviolet–visible photometry (Thermo Scientific, USA). Purified DNA was diluted with Tris-EDTA (Ethylene diamine tetraacetic acid) buffer to $\sim 20\, ng\, \mu l^{-1}$ and preserved in a freezer at −20°C for use in SSR PCR (Polymerase chain reaction) amplification.

### SSR-PCR amplification

To amplify SSRs of all samples, 20 primer pairs developed for *J. sigillata* (Chen *et al.*, 2017) were synthesized by Aoke Biological Technology Co. (Beijing, China). Information on the 20 primers is listed in the online Supplementary Table S1. PCR amplifications were performed in 20-μl reaction volumes containing 10.0 μl of 2 × ES TaqMaster mix, 0.4 μl each of forward and reverse primers, 1.0 μl template DNA and 8.2 μl RNase-free water. Touchdown PCR amplifications were carried out as follows: initial denaturation for 5 min at 94°C, followed by 35 cycles of 30 s at 94°C, 1 min at an annealing temperature of 53–60°C (based on the specific primer used, see online Supplementary Table S1) and 1 min at 72°C, with a final extension of 10 min at 72°C. The amplification products were detected by capillary electrophoresis (3730xl DNA analyser) (Chen *et al.*, 2014).

### Data statistics and genetic diversity analysis

#### Diversity of SSR loci

The sizes of detected amplification fragments were input into Excel, and GenAlEx 6.5 was used to calculate the number of alleles per SSR locus ($N$) and the following parameters: observed heterozygosity ($H_o$), expected heterozygosity ($H_e$), Shannon's information index ($I$), Nei's genetic distance ($D$) and gene flow ($N_m$) (Peakall and Smouse, 2012).

#### Population genetic diversity and genetic differentiation

The number of private alleles, mean allelic richness (AR), mean private allelic richness (PAR), $H_e$ and $I$ were used to measure the diversity of the populations. The number

https://doi.org/10.1017/S1479262120000441 Published online by Cambridge University Press

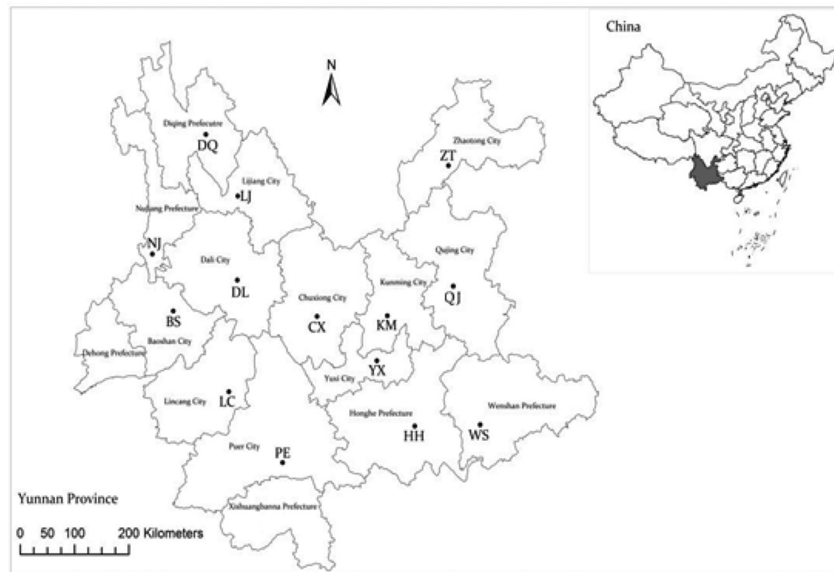


**Fig. 1.** Geographic locations of sampling sites of the 14 iron populations in this study. Note: The full names of abbreviated locations are given in Table 1.

of private alleles, AR and PAR were calculated with HP-RARE version 1.0 (Kalinowski, 2005), which uses rarefaction and hierarchical sampling to adjust for uneven sample sizes. Eight individuals (16 allele) were sampled at random from each population to match the smallest LC regional sample size. GenAIex 6.5 was used to calculate $H_e$ and $I$ of the populations and it was also used to measure population genetic differentiation by calculating the within-population inbreeding coefficient ($F_{IS}$) and the among-population genetic differentiation coefficient ($F_{ST}$). Among-and within-population genetic differences were further estimated by analysis of molecular variance (AMOVA) of population genetic variation in Arlequin v3.1 (Testolin *et al.*, 2000).

### Population structure and cluster analysis

The program structure, version 2.3.2 (Pritchard *et al.*, 2000) was used to detect the number of populations within germplasm collection of 1122 genotypes. The membership of each genotype was tested from $K = 1$ to $K = 6$ with admixture model, along with a burn-in period of 100,00 and a run length of 100,000. Based on the results, the $\Delta K$ parameter was estimated to identify the optimal number of clusters, as described in Evanno *et al.* (2005). A value of $K$ was selected when the estimate of $\Delta K$ peaked in the range of 1–6 and inferred ancestry estimates of individuals (membership coefficients) were derived for the selected populations (Pritchard *et al.*, 2000).

The unweighted pair-group method with arithmetic (UPGMA) means was used to construct a dendrogram based on Nei's distances in NTsys 2.10 (Rohlf, 2000). Correlation of population genetic distance with geographic distance was estimated by a Mantel test (Mantel, 1967).

## Results

### SSR-PCR amplification and diversity of SSR loci

As an example, amplification products detected by capillary electrophoresis were shown for the case of primer CUJRA123 amplified from samples W092–W095 in online Supplementary Figure S1.

A total of 351 alleles were detected from 1122 samples using 20 pairs of primers, all of which were polymorphic with lengths ranging from 96 to 250 bp (online Supplementary Table S1). The number of alleles detected by each primer varied from 7 (JSI73) to 27 (CUJRB307), with an average of 17.55 ($N$, online Supplementary Table S2). Genetic diversity parameters for each locus were listed in online Supplementary Table S2. For all 20 loci, observed heterozygosity ($H_o$) and expected heterozygosity ($H_e$) ranged from 0.18 to 0.75 and 0.26 to 0.78, respectively, with corresponding averages of 0.46 and 0.57. Shannon's information index ($I$) ranged from 0.55 to 1.86, with an average of 1.19. Polymorphism information content (PIC) varied from 0.22 to 0.79, with an average of 0.57. With regard to specific loci, CUJRB307 had the highest values of $H_e$ (0.78), $I$ (1.861) and PIC (0.79), while JUG13 exhibited the lowest values ($H_e = 0.26$, $I = 0.55$ and PIC = 0.22).

### Population genetic diversity

According to estimates of genetic diversity of the 14 iron walnut populations in Yunnan (Table 2), AR and PAR ranged from 3.40 to 4.62 and 0.11 to 0.36, with an average of 3.93 and 0.30, respectively. $H_e$, $I$, and PIC ranged from 0.49

https://doi.org/10.1017/S1479262120000441 Published online by Cambridge University Press

**Table 1.** Locations and sampling sizes of the 14 iron walnut populations in this study

| Location | Population code | Elevation range (m) | Climate type | Annual average temperature(°C) | Sample size |
|---|---|---|---|---|---|
| Baoshan | BS | 1580–2291 | Mid-subtropical low latitude monsoon | 21 | 32 |
| Chuxiong | CX | 1620–2586 | Sub-tropical monsoon | 14.8–21.9 | 50 |
| Dali | DL | 1494–2847 | Sub-tropical monsoon | 15 | 90 |
| Diqing | DQ | 1996–2760 | Temperate/cold temperate monsoon | 2.4–15 | 129 |
| Honghe | HH | 1585–2129 | Subtropical plateau monsoon | 15.1–22.6 | 41 |
| Kunming | KM | 1900–2295 | Subtropical plateau monsoon | 14.8 | 82 |
| Lingcang | LC | 1350–2100 | Subtropical low latitude monsoon | 16.5–21.5 | 18 |
| Lijiang | LJ | 1610–2600 | Low latitude warm temperate monsoon | 12.6–19.8 | 89 |
| Nujiang | NJ | 1033–2733 | Sub-tropical mountain monsoon | 14.8–20 | 90 |
| Puer | PE | 1390–1951 | Mid-subtropical mountain monsoon | 15.3–20.2 | 42 |
| Qujing | QJ | 1280–2240 | Subtropical plateau monsoon | 12.7–16.3 | 221 |
| Wenshan | WS | 1052–2063 | Subtropical plateau monsoon | 15.8–19.3 | 60 |
| Yuxi | YX | 1411–2230 | Subtropical plateau monsoon | 15.6–23.8 | 25 |
| Zhaotong | ZT | 950–2523 | Three-D climate | 18–21 | 153 |
| Total | | | | | 1122 |

to 0.65, 0.84 to 1.40 and 0.41 to 0.60, with corresponding averages of 0.58, 1.19 and 0.53, respectively. These parameters differed among the 14 populations, with the DQ population having the highest values of AR, $H_e$, $I$ and PIC (4.62, 0.65, 1.40 and 0.60, respectively) and the LC population having the lowest values (AR = 3.40, $H_e = 0.49$, $I = 0.84$, PIC = 0.41). As demonstrated by these results, the genetic diversity of *J. sigillata* in the 14 populations was highest in the DQ population and lowest in the LC population.

## Population genetic differentiation

The inbreeding coefficient for the total population ($F_{IT}$) per locus varied from −0.42 to 0.54, with an average of 0.19. The within-population inbreeding coefficient ($F_{IS}$) ranged from −0.41 to 0.59 and the average value was 0.20. The among-population genetic differentiation coefficient ($F_{ST}$) ranged from 0.01 to 0.09, with an average of 0.05 (online Supplementary Table S2), thus indicating low differentiation among populations (i.e. only 5.00% of the total genetic variability being inter-populational). The level of gene flow ($N_m$) was estimated to be 6.70 (online Supplementary Table S2), which implied a relatively high level of gene flow among populations and thus prevention of gene differentiation among populations. As shown in online Supplementary Table S2, values of $F_{IT}$ and $F_{IS}$ were positive for most loci, thus indicating a deficiency of heterozygosity at these loci.

The among-population *P*-value returned by the AMOVA was <0.001, which indicated the difference was extremely significant (online Supplementary Table S3). Most genetic diversity was within-population, accounting for 94.85% of the total variance, with only 5.15% existing among populations. The results of the AMOVA regarding population differentiation were consistent with the above-estimated *F*-statistics.

## Population structure and cluster analysis

STRUCTURE program was used to analyse the genetic structure of the germplasm accessions. As showed in Fig. 2(a), $\Delta K$ parameter reached the sharp peak when $K = 2$, suggesting that all accessions were divided into two genetic clusters. In Fig. 2(b), showing from the results of the STRUCTURE analysis, each accession was represented by a vertical histogram with two colours segments that represent the individual's membership fraction in two clusters. The results indicated that all accessions contained a certain membership fraction in two clusters and no obvious differentiation among the 14 populations.

To reveal genetic relationships between populations and further clarify among-population genetic differentiation,

https://doi.org/10.1017/S1479262120000441 Published online by Cambridge University Press

**Table 2.** Genetic diversity of 14 iron walnut populations in Yunnan

| Populations | AR | PAR | $H_o$ | $H_e$ | $I$ | PIC | $F_{IT}$ |
|---|---|---|---|---|---|---|---|
| BS | 3.93 | 0.30 | 0.50 | 0.55 | 1.10 | 0.49 | 0.07 |
| CX | 3.63 | 0.11 | 0.47 | 0.54 | 1.06 | 0.48 | 0.12 |
| DL | 3.86 | 0.12 | 0.43 | 0.54 | 1.12 | 0.50 | 0.22 |
| DQ | 4.62 | 0.34 | 0.46 | 0.65 | 1.40 | 0.60 | 0.29 |
| HH | 4.28 | 0.23 | 0.45 | 0.60 | 1.24 | 0.55 | 0.23 |
| KM | 4.57 | 0.24 | 0.49 | 0.64 | 1.36 | 0.59 | 0.21 |
| LC | 3.40 | 0.15 | 0.42 | 0.49 | 0.84 | 0.41 | 0.06 |
| LJ | 3.78 | 0.19 | 0.44 | 0.54 | 1.11 | 0.50 | 0.23 |
| NJ | 4.53 | 0.24 | 0.44 | 0.63 | 1.35 | 0.59 | 0.30 |
| PE | 3.61 | 0.22 | 0.44 | 0.51 | 1.00 | 0.46 | 0.14 |
| QJ | 4.34 | 0.25 | 0.48 | 0.60 | 1.31 | 0.56 | 0.21 |
| WS | 4.08 | 0.20 | 0.44 | 0.58 | 1.20 | 0.53 | 0.25 |
| YX | 4.38 | 0.25 | 0.49 | 0.61 | 1.23 | 0.55 | 0.17 |
| ZT | 4.40 | 0.36 | 0.47 | 0.61 | 1.33 | 0.56 | 0.24 |
| Average | 3.93 | 0.30 | 0.46 | 0.58 | 1.19 | 0.53 | 0.19 |

AR, mean allelic richness; PAR, mean private allelic richness; $H_o$, obseved heterozygosity; $H_e$, expected heterozygosity; $I$, Shannon's information index; PIC, polymorphic information content; $F_{IT}$, total population inbreeding coefficient.

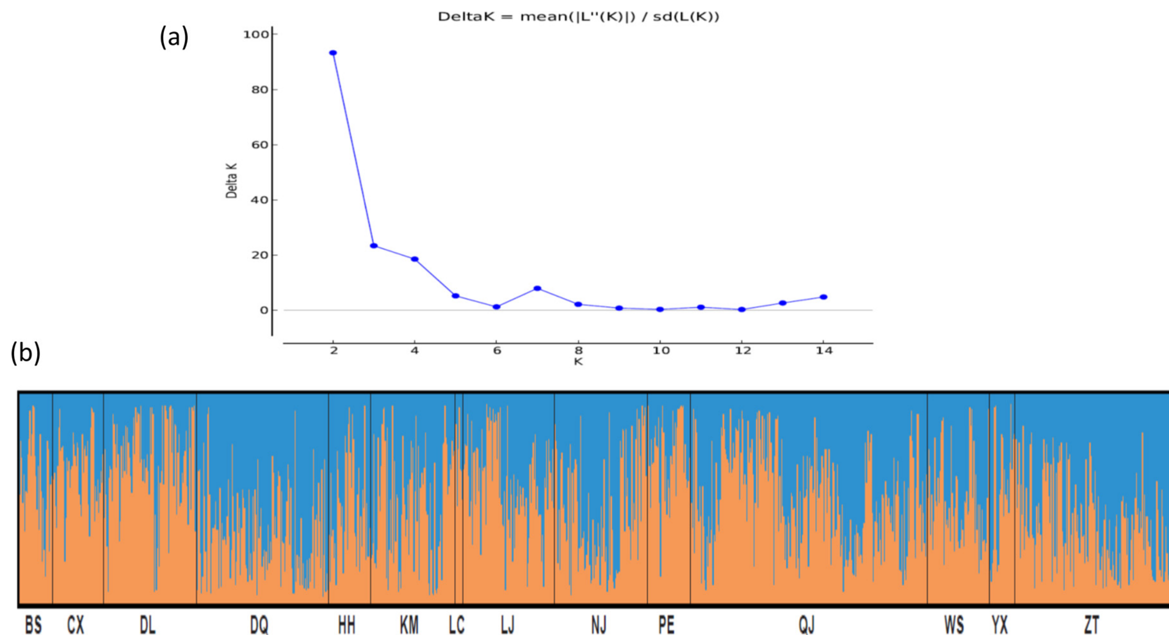


**Fig. 2.** Prediction of sub-populations ($K = 2$) for collection of 1122 iron walnut genotypes using Evanno method in (a) STRUCTURE analysis and (b) STRUCTURE plot depicting sub-populations. Note: In (b), each individual is represented by a vertical line and populations are separated by a vertical black line. Different colours in the same line indicate the individual's estimated membership percentage in $K$ clusters.

Nei's unbiased genetic identity coefficients were analysed (Table 3). Genetic identities varied from 0.85 to 0.98, with an average of 0.94, which corresponded to a relatively high genetic similarity among populations. However, populations of *J. sigillata* germplasm in Yunnan had a relatively low level of among-population

https://doi.org/10.1017/S1479262120000441 Published online by Cambridge University Press

**Table 3.** Nei's unbiased genetic identity (above diagonal) and genetic distance (below diagonal) among 14 iron walnut populations

| Populations | BS | CX | DL | DQ | HH | KM | LC | LJ | NJ | PE | QJ | WS | YX | ZT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BS | | 0.95 | 0.96 | 0.92 | 0.97 | 0.93 | 0.95 | 0.98 | 0.96 | 0.95 | 0.98 | 0.97 | 0.96 | 0.94 |
| CX | 0.05 | | 0.97 | 0.89 | 0.94 | 0.93 | 0.91 | 0.94 | 0.94 | 0.96 | 0.95 | 0.93 | 0.96 | 0.90 |
| DL | 0.04 | 0.03 | | 0.90 | 0.95 | 0.94 | 0.95 | 0.97 | 0.96 | 0.98 | 0.96 | 0.95 | 0.97 | 0.91 |
| DQ | 0.08 | 0.12 | 0.11 | | 0.94 | 0.92 | 0.85 | 0.93 | 0.96 | 0.86 | 0.94 | 0.91 | 0.90 | 0.95 |
| HH | 0.03 | 0.07 | 0.05 | 0.07 | | 0.95 | 0.92 | 0.97 | 0.95 | 0.93 | 0.98 | 0.97 | 0.97 | 0.95 |
| KM | 0.08 | 0.08 | 0.07 | 0.08 | 0.05 | | 0.88 | 0.94 | 0.95 | 0.92 | 0.97 | 0.95 | 0.95 | 0.92 |
| LC | 0.06 | 0.09 | 0.05 | 0.16 | 0.08 | 0.13 | | 0.95 | 0.91 | 0.95 | 0.93 | 0.95 | 0.94 | 0.90 |
| LJ | 0.03 | 0.06 | 0.03 | 0.07 | 0.04 | 0.07 | 0.06 | | 0.96 | 0.96 | 0.98 | 0.97 | 0.96 | 0.93 |
| NJ | 0.04 | 0.07 | 0.04 | 0.04 | 0.05 | 0.06 | 0.10 | 0.05 | | 0.92 | 0.96 | 0.95 | 0.95 | 0.96 |
| PE | 0.05 | 0.04 | 0.02 | 0.15 | 0.07 | 0.09 | 0.06 | 0.04 | 0.08 | | 0.94 | 0.93 | 0.96 | 0.89 |
| QJ | 0.02 | 0.05 | 0.04 | 0.06 | 0.02 | 0.04 | 0.07 | 0.02 | 0.04 | 0.06 | | 0.98 | 0.97 | 0.96 |
| WS | 0.03 | 0.07 | 0.05 | 0.09 | 0.03 | 0.05 | 0.05 | 0.03 | 0.05 | 0.07 | 0.02 | | 0.96 | 0.95 |
| YX | 0.04 | 0.04 | 0.03 | 0.10 | 0.03 | 0.05 | 0.07 | 0.04 | 0.05 | 0.04 | 0.03 | 0.04 | | 0.94 |
| ZT | 0.06 | 0.10 | 0.10 | 0.05 | 0.05 | 0.08 | 0.11 | 0.07 | 0.04 | 0.12 | 0.04 | 0.05 | 0.07 | |

genetic differentiation. The highest identity coefficient was between populations DL and PE (genetic identity = 0.98), indicating their close genetic relationship, while the lowest identity coefficient was that between populations DQ and LC (genetic identity = 0.85). A UPGMA dendrogram based on genetic distance, constructed to further illustrate population relationships, clustered the 14 populations into two major groups under the genetic distance of 0.075 (Fig. 3). The first group (I) contained three populations (ZT, NJ and DQ) distributed mainly in northwestern and southeastern Yunnan. The second group (II) included other 11 populations. The group III was divided into two sub-groups under the genetic distance of 0.065, LC as one sub-group, and the second consisting of other ten populations.

Mantel testing of correlations between pairwise linearized genetic differentiation values and the natural logarithm of between-site geographic distances revealed that geographic distance was significantly correlated with genetic distance at $P < 0.05$. We found a significant correlation between geographic distance and genetic distance ($r = 0.33$, $P = 0.04$), which implied that the geographic distance between populations affects population differentiation to some extent.

## Discussion

Shannon's information index ($I$) and expected heterozygosity ($H_e$) were used to evaluate genetic diversity (Shannon and Weave, 1959; Nei, 1973). $I$ can effectively reflect population genetic diversity (Huang, 2013), whereas $H_e$ can be applied not only to evaluate genetic diversity, but also to indicate the richness and evenness of an allele in populations (Wang *et al.*, 2007). As it is less influenced by sample size, $H_e$ is more suitable for estimating population genetic diversity, with a higher $H_e$ corresponding to richer genetic diversity (Frankham *et al.*, 2002). In the present study, species-level $H_e$ and $I$ were 0.578 and 1.190, respectively. Compared with other *Juglans* species assessed using the same molecular markers, the parameters estimated in this study were higher than those of *J. regia* and cultivars in Xinjiang ($H_e = 0.282$, $I = 0.4225$; Liu *et al.*, 2008) and *J. regia* and *J. sigillata* in Shanxi, Sichuan, Chongqing and Yunnan ($H_e = 0.525$; Wang *et al.*, 2008), similar to those of *J. regia* and *J. sigillata* in Xizang ($H_e = 0.586$; Wang *et al.*, 2015), and lower than those of *J. regia* and *J. sigillata* in Shanxi, Henan, Sichuan and Yunnan ($H_e = 0.617$; Wang *et al.*, 2007), *J. regia* in the Daba mountains of Sichuan ($H_e = 0.657$, $I = 1.324$; Xu *et al.*, 2016), *J. nigra* in the central hardwood region of the United States ($H_e = 0.793$; Victory *et al.*, 2006) and *J. mandshurica* in central and southwestern China ($H_e = 0.806$; Wang *et al.*, 2008). PIC is another important index of genetic diversity.

https://doi.org/10.1017/S1479262120000441 Published online by Cambridge University Press

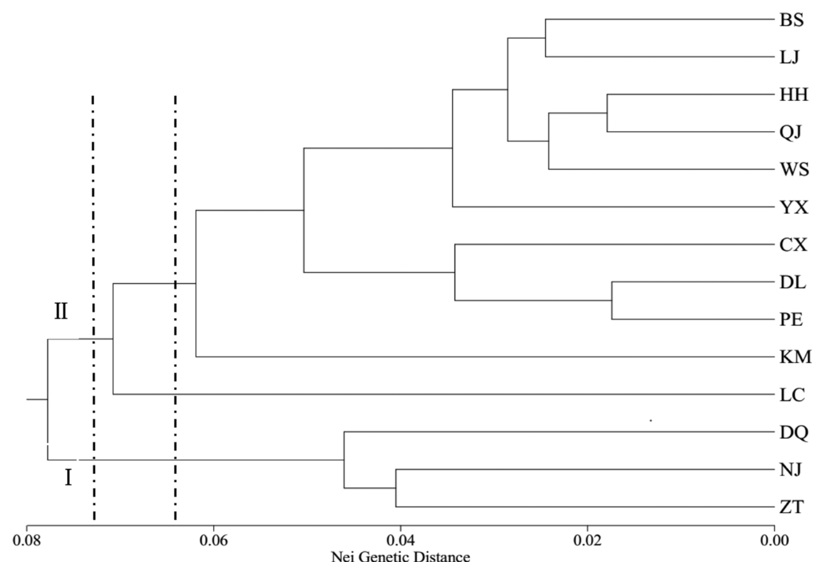


**Fig. 3.** UPGMA dendrogram for 14 iron walnut populations in Yunnan.

According to Botstein *et al.* (1980), a locus is highly polymorphic if PIC > 0.5 and moderately polymorphic if 0.25 < PIC < 0.5, with a PIC < 0.25 corresponding to low polymorphism. In the present study, the average PIC of 20 loci was 0.570, with 13 loci having PICs > 0.5. The average PIC of the 14 populations was 0.5265. The PICs of 12 populations were near or >0.5; the exceptions were LC (0.4104) and PE (0.4576). AR is a better measure of diversity and informative when the population size vary significantly. The mean AR for 14 populations was 3.93, with the highest value of 4.62 (DQ) and the lowest of 3.40 (LC). According to our comprehensive analysis, the provincial-wide genetic diversity of *J. sigillata* germplasm resources in Yunnan is moderately high, which implies that great potential exists for their further exploitation and development.

Internal factors (e.g. breeding systems, genetic drift, gene mutation and gene flow), as well as external ones (e.g. population isolation and habitat fragmentation resulting from environmental changes and human activities), will interactively influence species genetic diversity. Hamrick and Loveless (1989a, 1992) have stated that the genetic diversity level of a species is related to its life history characteristics and ecological traits. Rare and endemic species and species of narrow distribution have low levels of genetic diversity, whereas those characterized by a broad distribution, outcrossing and animal-dispersed seeds should possess high genetic diversity (Hamrick and Godt, 1996). Although *J. sigillata* is only distributed in southwestern China (mainly in Yunnan, Guizhou, western Sichuan and southern Tibet) (Wang *et al.*, 2009), this distribution area includes varied ecological and geographic environments with different climatic features, a large range of growing elevations (from 800 to 3200 m) and complicated topographic and geomorphic conditions. The ability to adapt to various environments results in relatively high genetic diversity (Wen *et al.*, 2010). Another key factor leading to relatively high genetic diversity is breeding system. Iron walnut is wind-pollinated, a situation in which pollen is spread more easily compared with insect pollination. Furthermore, walnut is a dichogamous, monoecious plant with unisexual flowers, which promotes outcrossing and discourages selfing, thereby helping to maintain relatively high levels of genetic diversity. Finally, two other characteristics, overlapping generations and a long reproductive life span, should contribute to the maintenance of relatively high genetic diversity.

With respect to spatial distribution, the genetic diversities of the 14 populations were different. The DQ population had the highest level of genetic diversity (AR = 4.62; $H_e$ = 0.648; $I$ = 1.402), whereas that of LC was the lowest (AR = 3.40; $H_e$ = 0.489; $I$ = 0.838). When arranged from highest to lowest genetic diversity based on AR, $H_e$ and $I$, the relative order of populations was as follows: DQ, KM, NJ, ZT, QJ, HH, YX, WS, DL, LJ, BS, CX, PE and LC. This order may be mostly due to geographic location, germplasm type, population size and the influence of human activities. DQ, NJ and ZT are located in northern and northwestern Yunnan, a region characterized by complicated geographic and climatic conditions, and, most importantly, they adjoin Tibet, Guizhou and Sichuan, where *J. sigillata* is reportedly native (Gunn *et al.*, 2010). In addition, these three populations include a greater number of naturally growing trees in high mountains and valleys that are less influenced from human activities. Kunming, the capital of Yunnan, had a higher level of genetic diversity; this is because it serves as a collection centre, with

https://doi.org/10.1017/S1479262120000441 Published online by Cambridge University Press

germplasm from throughout the province concentrated to some extent. The low genetic diversity of LC may be a consequence of small sampling size.

Private alleles, which are generally low-frequency alleles often reflects the selective advantage of populations. Thus, estimating PAR would be useful to reveal the genetic distinctiveness of populations. The mean PAR of 14 populations was 0.30, and ZT had the highest value of 0.36, then DQ (0.34) and BS (0.30). CX showed the lowest PAR of 0.11. That implied that ZT, DQ and BS contained more private alleles and should take more attention in the purposed genetic selection.

$F_{ST}$ is the main index reflecting population differentiation (Templeton, 2010). In the present study, the $F_{ST}$ of the 14 populations was 0.05 (online Supplementary Table S2), thus indicating that only 5% of the total variation existed among populations. This finding was confirmed by the results of AMOVA showing that the genetic variation in walnut germplasm in the area was mainly inter-populational (within-population) and also confirmed by STRUCTURE analysis showing that high genetic diversity among individuals not among populations. The genetic differentiation uncovered in this study was relatively low compared with that found in other studies of related walnut populations, such as *J. nigra* ($F_{ST}$ = 0.017, Victory *et al.*, 2006), *J. regia* ($F_{ST}$ = 0.103) and *J. sigillata* in Tibet ($F_{ST}$ = 0.111; Wang *et al.*, 2015), eight *J. regia* populations in China ($F_{ST}$ = 0.196; Wang *et al.*, 2007) and eight populations in northern China ($G_{ST}$ = 0.4934; Ji *et al.*, 2014). Evaluation of gene flow ($N_m$), a very important factor influencing within and among-population genetic variation (Slatkin and Barton, 1989), led to the same conclusion. Generally speaking, species with a high $N_m$ have low among-population genetic variation. When $N_m > 1$, according to Hamrick *et al.* (1995), gene flow can overcome genetic drift and prevent differentiation among populations. Conversely, when $N_m < 1$ genetic drift would lead to the among-population differentiation. The average $N_m$ in this study was 6.70, which was >1, thus indicating the existence of relatively high gene flow among populations. Also, that would prevent population genetic differentiation. This result is consistent with the findings of Chen *et al.* (2008), who investigated the genetic diversity of walnut germplasm in Sichuan Province.

Bussell (1999) summarized $F_{ST}$ values of 35 species and concluded that the genetic differentiation of inbreeding species is obviously higher than that of outbreeding ones. Hamrick *et al.* (1989b) statistically analysed numerous seed plants and found that the genetic differentiation of long-lived, wind-pollinated, and outcrossing plants was generally the lowest. In our study, the relatively low genetic differentiation of *J. sigillata* in Yunnan can be primarily attributed to the long lifespan and wind-pollination and outcrossing traits of this species. Germplasm exchange and seed spread is another important reason for the low differentiation. Yunnan is a mountainous province where traditional agriculture is practiced. Iron walnut, an economically important local crop, has been grown and cultivated for hundreds of years in this area because of its economic value and cultural appeal. During the course of walnut cultivation, human-mediated selection, germplasm exchange and seed spread have frequently occurred. Given these facts, relatively strong human interference may be another primary reason for the low genetic differentiation.

On the basis of genetic similarity, 14 populations were divided into two major groups under the genetic distance of 0.075 and genetic distance was found to be significantly correlated with geographic distance. In other words, most populations that were geographically close to one another generally clustered in the same group, such as DQ and NJ (northwestern Yunnan) in one group, and CX, DL and PE (central Yunnan) in another. Geographic location, environment, resource state and genetic diversity were related to group compositions. For example, members of the first group, consisting of DQ, NJ and ZT, were collected from the border of the province, which is characterized by complex landforms of high mountains, valleys and rivers and various climates. Given the underdeveloped status of this area, samples obtained for our investigation were from natural growing populations rather than products of human-mediated selection. The three populations constituting the first group also had the highest genetic diversities in 14 populations. In contrast, the remaining nine populations (except for LC and KM) comprising the second group were located in the provincial interior, which is characterized by flatter land and a better-developed economy. Although some samples in this group were also natural growing in villages, they might be more or less interfered by human activities, e.g. they might be planted by previous generations of hundreds of years ago, which would be expected to have low levels of genetic diversity and higher genetic identities. The LC population, which was adjacent to PE, CX and DL, had the smallest number of samples and the lowest genetic diversity, but the genetic difference of the samples in this group were relatively high. More samples should be taken from LC for further study. The KM population represented resources gathered from throughout the province; it thus displayed high levels of genetic diversity and constituted its own sub-group. When taking 1122 individuals as a total or with in population, the average Nei' genetic distance between 1122 genotypes was 0.5578, the minimum and maximum values were 0.0149 and 0.9984, respectively.

Improvement programmes of iron walnut have been carried out through natural selection and crossbreeding during the past 20 years, and superior cultivars and strains have been officially registered (Fang *et al.*, 1998; Fan *et al.*, 2005; Zhao *et al.*, 2007). Recently, phenotypic traits and

https://doi.org/10.1017/S1479262120000441 Published online by Cambridge University Press

diversity of nuts, genetic diversity of resources in some area of Yunnan have been evaluated (Xiao *et al*., 2018; Chen *et al*., 2019). However, the present research is the first time comprehensive characterization of germplasm in Yunnan has been carried out, which may help future genetic improvement.

In conclusion, genetic diversity is the foundation of germplasm exploitation, conservation and utilization. In the present study, a moderately high level of genetic diversity in *J. sigillata* germplasm in Yunnan was uncovered which suggests that great potential exists for the further exploitation and development of these germplasm resources, especially those from populations containing richer genetic diversity, such as DQ, NJ and ZT. Further systematic study needs to be carried out on these resources, and conservation measures, such as *in situ* conservation or establishment of a resource collection nursery, are even required. Given that the genetic variation in the iron walnut germplasm in the study area is mainly within populations, future breeding and conservation efforts should focus on the selection of individuals and families within populations, especially those within DQ, ZT and BS populations of high PAR, in addition to the selection of populations with high genetic diversity.

**Supplementary material**

The supplementary material for this article can be found at https://doi.org/10.1017/S1479262120000441.

## Acknowledgements

The authors thank the staff members of the local Forestry Bureau and Forest Research Institutes of various prefectures and cities for providing information on walnut distribution and assisting with the collection of experimental materials. This work was financially supported by the National Natural Science Foundation of China (31660214), the Forestry Innovation Project in Yunnan ([2014]CX01) and the Yunnan Provincial Science and Technology Major Project (2018ZG002-1).

## References

Bayazit S, Kazan K, Gulbitti V, Evik C, Ayanoglu H and Ergul A (2007) AFLP analysis of genetic diversity in low chilling requiring Walnut (*Juglans regia* L.) genotypes from Hatay, Turkey. *Scientia Horticulturae* 111: 394–398.

Botstein D, White RL, Skolnick M and Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphism. *American Journal of Human Genetics* 32: 314–331.

Bussell JD (1999) The distribution of random amplified polymorphic DNA (RAPD) diversity amongst populations of *Isotoma petraea* (Lobeliaceae). *Molecular Ecology* 8: 775–789.

Chen LH, Hu TX, Zhang F and Li GH (2008) Genetic diversity of four *Juglans* Populations revealed by AFLP in Sichuan Province, China. *Journal of Plant Ecology* 32: 1362–1372, (In Chinese).

Chen LN, Ma QG, Chen YK, Wang BQ and Pei D (2014) Identification of major walnut cultivars grown in China based on nut phenotypes and SSR markers. *Scientia Horticulturae* 168: 240–248.

Chen SY, Ning DL, Wu T, Xiao LJ, Lin HD, Zhu YF, He N and Pan L (2017) Development of SSR markers in *Juglans sigillata* and its application in genetic diversity analysis. *Journal of Northwest Forestry University* 32: 91–96, (In Chinese).

Chen SY, Wu T and Xiao LJ (2019) Genetic diversity and distribution characteristic of *Juglans sigillata* resources along Jinsha river in three parallel rivers belt of Yunnan. *Journal of Southern Agriculture* 50: 2656–2664, (In Chinese).

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

Fan ZY, Xi XL, Fang WL, Zhang Y and Zhao TS (2005) Breeding on three new promising and early fruiting walnut cultivars developed by species hybridization. *Journal of Zhejiang Forestry College* 22: 518–523, (In Chinese).

Fang WL, Yang ZB, Huang Q, Zhang XX, Fan ZY and Qi SL (1998) Selective breeding of precocious, prolific and quality walnut strains. *Economic Forest Researches* 16: 6–10, (In Chinese).

FAO (Food and Agricultural Organization of the United Nations) (2012) *FAO Statistical Databases and DataSets*. FAO (Food and Agricultural Organization of the United Nations). Available online: http://faostat.fao.org (accessed on 10 October 2020).

Feng XJ, Yuan XY, Sun YW, Hu YH, Zulfiqar S, Ouyang XH, Dang M, Zhou HJ, Woeste K and Zhao P (2018) Resources for studies of iron walnut (*Juglans sigillata*) gene expression, genetic diversity, and evolution. *Tree Genetics & Genomes* 14: 51.

Fjellstrom RG and Parfitt DE (1994) Genetic diversity determined by restriction fragment length polymorphisms. *Genome* 37: 690–700.

Frankham R, Ballou JD and Briscoe DA (2002) *Introduction to Conservation Genetics*. Cambridge: Cambridge University Press.

Gunn BF, Aradhya M, Salick JM, Miller AJ, Yang YP, Liu L and Hai X (2010) Genetic variation in walnut (*Juglans regia* and *J. sigillata*; Juglandaceae): species distinctions, human impacts, and the conservation of agrobiodiversity in Yunnan, China. *American Journal of Botany* 97: 660–671.

Hamrick JL and Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of Royal Society of London B* 351: 1291–1298.

Hamrick JL and Loveless MD (1989a) The genetic structure of tropical tree populations: associations with reproductive biology. In: Bock JU, Linhart YB (eds) *Plant Evolutionary Ecology*. Boulder: Westview Press, pp. 129–146.

Hamrick JL and Loveless MD (1992) Factors influencing levels of genetic diversity in woody plants species. *New Forests* 6: 95–124.

Hamrick JL, Godt MJW and Brown AHD (1989b) Allozyme diversity in plant species. *Plant Population Breeding & Genetic Resources* 28: 43–63.

Hamrick JL, Godt MJW and Sherman-Broyles LS (1995) Gene flow among plant populations: evidence from genetic markers. In: Hoch DC, Stephnon AG (eds) *Experimental and Molecular Approaches to Plant Biosystematics*. St Louis: Missouri Botanical Garden, pp. 215–232.

Huang Y (2013) Analysis of *Camellia meiocarpa* genetic diversity based on SRAP markers. *Scientia Silvae Sinicae* 49: 43–50, (In Chinese).

Ji AQ, Wang YN, Wu GL, Wu WJ, Yang HY and Wang QH (2014) Genetic diversity and population structure of North China mountain walnut revealed by ISSR. *American Journal of Plant Sciences* 5: 3194–3202.

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

Li GT, Ai CX, Zhang LS, Wei HR and Liu QZ (2011) ISSR Analysis of genetic diversity among seedling Walnut (*Juglans* Spp.) populations. *Journal of Plant Genetic Resources* 12: 640–645, (In Chinese).

Liu XL, Chen XS, Zhang MY, Chen XL, He TM, Zhang LJ and Zhang CY (2008) Population genetic structure analysis of *Juglans regia* using SSR markers. *Journal of Fruit Science* 25: 526–530, (In Chinese).

Liu BY, Liu XL, Wei HB and Gu WY (2017) Genetic diversity of germplasm resources of Walnut in Qinghai. *Journal of Northwest Forestry University* 32: 130–135, (In Chinese).

Lu B (2009) Characteristics and quality of *Juglans sigillata*. *Nonwood Forest Research* 27: 137–140, (In Chinese).

Lu SJ, Lu B, Wu SB, Duan LJ, Fan LY and Yi XY (2015) Analysis of the main economic characters of walnut nuts in Yunnan. *Northern Horticulture* 19: 18–21, (In Chinese).

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* 27: 209.

Nazeer A, Mir JI, Mir RR, Nazir AR, Rizwan R, Shabir HW, Wajida S, Hidayatullah M and Sheikh MA (2012) SSR And RAPD analysis of genetic diversity in walnut (*Juglans regia* L.) genotypes from jammu and Kashmir, India. *Physiology and Molecular Biology of Plant* 18: 149–160.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy Sciences* 70: 3321–3323.

Ning DL, Ma QG, Zhang Y, Wang H, Lu B and Pei D (2011) FISH-AFLP analysis of genetic diversity on walnut cultivars in Yunnan Province. *Forest Research* 24: 189–193, (In Chinese).

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

Potter D, Gao F, Aliello G, Leslie C and McGranahan G (2002) Intersimple sequence repeat markers for fingerprinting and determining genetic relationship of Walnut (*Juglans regia*) cultivars. *Journal of the American Society for Horticultural Science* 127: 75–81.

Pritchard JK, Stephens M, Rosenberg NA and Donelly P (2000) Association mapping in structured populations. *The American Journal of Human Genetics* 37: 170–181.

Rohlf FJ (2000) *NTSYS-PC: Numerical Taxonomy and Multivariate Analysis System Version 2.1*. Setauket New York: distribution by Exeter software.

Shah UN, Mir JI, Ahmed N and Fazili KM (2018) Assessment of germplasm diversity and genetic relationships among walnut (*Juglans regia* L.) genotypes through microsatellite markers. *Journal of the Saudi Society of Agricultural Sciences* 17: 339–350.

Shamlu F, Rezaei M, Lawson S, Ebrahimi A, Biabani A and Khan-Ahmadi A (2018) Genetic diversity of superior Persian walnut genotypes in Azadshahr, Iran. *Physiology and Molecular Biology of Plants* 24: 939–949.

Shannon CE and Weave W (1959) *The Mathematical Theory of Communication*. Illiois: University of Illiois Press, pp. 20–29.

Slatkin M and Barton NH (1989) A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43: 1349–1368.

Templeton AR (2010) *Introduction to Conservation Genetics*. Cambridge: Cambridge University Press. 56.

Testolin R, Marrazzo T, Cipriani G, Quarta R, Verde I, Dettori MT, Pancaldi M and Sansavini S (2000) Microsatellite DNA in peach (*Prunus persica* L. Batsch) and its use in fingerprinting and testing the genetic origin of cultivars. *Genome* 43: 512–520.

Victory ER, Glaubitz JC, Rhodes OE and Woeste KE (2006) Genetic homogeneity in *Juglans nigra* (Juglandaceae) at nuclear microsatellites. *American Journal of Botany* 93: 118–126.

Wang H, Hao JM, Wang BQ and Pei D (2007) SSR Analysis of genetic diversity of eight natural walnut populations in China. *Scientia Silvae Sinicae* 43: 120–124, (In Chinese).

Wang H, Pei D, Gu RS and Wang BQ (2008) Genetic diversity and structure of Walnut populations in Central and Southwest China revealed by microsatellite markers. *Journal of the American Society for Horticultural Science* 133: 197–203.

Wang H, Yan YB, Zhang JP and Pei D (2009) Application of ITS sequence and SSR markers to study the relationship between *Juglans regia* And *Juglans sigillata*. *Journal of Nanjing Forestry University: Natural Science Edition* 33: 35–38, (In Chinese).

Wang HX, Zhao SG, Gao Y, Zhang ZH and Xuan LC (2011) Genetic diversity of *Juglans regia* L. Cultivars revealed by AFLP analysis. *Scientia Agricultura Sinica* 44: 1434–1442, (In Chinese).

Wang H, Pan G, Ma QG, Zhang JP and Pei D (2015) The genetic diversity and introgression of *Juglans regia* and *Juglans sigillata* in Tibet as revealed by SSR markers. *Tree Genetics & Genomes* 11: 1–11.

Wen YF, Han WJ and Wu S (2010) Plant genetic diversity and its influencing factors. *Journal of Central South University of Forestry & Technology* 30: 80–87, (In Chinese).

Wu YM, Pei D, Xi SK and Li JR (2000) A study on the genetic relationship among species in *Juglans* L. Using RAPD markers. *Acta Horticulturae Sinica* 27: 17–22, (In Chinese).

Xi SK (1987) Gene resources of *Juglans* and genetic improvement of *Juglans regia* in China. *Scientia Silvae Sinicae* 23: 342–349, (In Chinese).

Xi RT and Zhang YP (1996) *Fruit Trees of China: Walnut Volume*. Beijing, China: Chinese Forestry Press. (In Chinese).

Xiao LJ, Chen SY, Ning DL, Wu T and He N (2018) Study on the genetic diversity of germplasm resources of *Juglans sigillata* in Northwest Yunnan. *Journal of Yunnan Agricultural University (Natural Science)* 33: 597–604, (In Chinese).

Xu YJ, Han HB, Wang H, Chen NL, Ma QG and Pei D (2016) Phenotypic and genetic diversities of nuts of Walnut (*Juglans regia*) populations originated from seedlings in Daba Mountains. *Scientia Silvae Sinicae* 52: 111–118, (In Chinese).

Yang E, Chen SY, Zhang Y, Fang ZY and Xi XL (2005) Comparison on effect between two methods of DNA extraction from leaves of *Juglans sigillata*. *Journal of West China Forestry Science* 34: 72–75, (In Chinese).

Yuan XY, Sun YW, Bai XR, Dang M, Feng XJ, Zulfiqar S and Zhao P (2018) Population structure, genetic diversity, and gene introgression of two closely related walnuts (*Juglans regia* and *J. sigillata*) in southwestern China revealed by EST-SSR markers. *Forests* 9: 646.

Zhao TS, Fang WL, Fan ZY, Xi XL and Zhang Y (2007) Yunxin 90303, a promising new early walnut selection. *Journal of Fruit Science* 24: 252–253, (In Chinese).

https://doi.org/10.1017/S1479262120000441 Published online by Cambridge University Press