Tree Genetics & Genomes (2016) 12: 111
DOI 10.1007/s11295-016-1064-1

ORIGINAL ARTICLE

# Genetic diversity and population structure of common walnut (*Juglans regia*) in China based on EST-SSRs and the nuclear gene phenylalanine ammonia-lyase (*PAL*)

**Huan Han[1] · Keith E. Woeste[2] · Yiheng Hu[1] · Meng Dang[1] · Tian Zhang[1] · Xiao-Xiao Gao[1] · Huijuan Zhou[1] · Xiaojia Feng[1] · Guifang Zhao[1] · Peng Zhao[1]**

Received: 13 April 2016 / Revised: 20 September 2016 / Accepted: 3 October 2016 / Published online: 9 November 2016


**Abstract** Common walnut (*Juglans regia* L.) is an economically important temperate tree species valued for both its nut and wood. We investigated the genetic diversity and population structure of *J. regia* germplasm from 13 locations in China using 10 markers derived from expressed sequences (EST-SSR) and sequence polymorphisms within the phenylalanine ammonia-lyase (*PAL*) gene. Analysis of the population genetic structure based on EST-SSRs showed distinct populations in northern versus southern China that were not reflected in the spatial distribution of *PAL* haplotypes. High levels of population differentiation were probably the result of reproductive isolation and in southern China, hybridization with *Juglans sigillata*. The results indicate the possible presence of distinct evolutionary lineages of *J. regia* in the Qinling and Daba Mountains of China and in Yunnan province that may require ecological management if they are to be retained as in situ resources.

**Keywords** Qinling-Daba MountainsPersian walnutGenetic structure*Juglans sigillata*Microsatellites

## Introduction

*Juglans regia* (Persian walnut or common walnut) is a diploid tree crop species valued since ancient times for both its timber and its edible nut (Vinson and Cai 2012; Hayes et al. 2015; Tsoukas et al. 2015; Pollegioni et al. 2015). It is native to the

Communicated by Y. Tsumura

**Electronic supplementary material** The online version of this article (doi:10.1007/s11295-016-1064-1) contains supplementary material, which is available to authorized users.

✉ Peng Zhao
pengzhao@nwu.edu.cn

Huan Han
1515459965@qq.com

Keith E. Woeste
kwoeste@fs.fed.us

Yiheng Hu
191133254@qq.com

Meng Dang
a1193371050@qq.com

Tian Zhang
1260583247@qq.com

Xiao-Xiao Gao
1535564566@qq.com

Huijuan Zhou
514335229@qq.com

Xiaojia Feng
332146704@qq.com

Guifang Zhao
gfzhao@nwu.edu.cn

[1] Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education, College of Life Sciences, Northwest University, Road 229, Xi'an, Shaanxi 710069, China

[2] USDA Forest Service Hardwood Tree Improvement and Regeneration Center (HTIRC), Department of Forestry and Natural Resources, Purdue University, 715 West State Street, West Lafayette, IN 47907, USA

mountainous regions of central Asia but is now the most widespread tree nut in the world (Bayazit et al. 2007; Chen et al. 2014; Pollegioni et al. 2015). China is the third largest walnut (in shell) producer, with a production of 1,700,000 tons (FAO 2012; Golge et al. 2016). Walnuts have been cultivated in China for thousands of years, and China is one of the major centers of walnut genetic diversity, serving as a germplasm source for walnut breeding (Chen et al. 2014). Starting in the Western Han Dynasty (206 BCE to 9 CE), cultivated walnuts were likely derived by selection of seedlings from geographically distinct natural populations and spread by trade and military conquest (Molnar et al. 2011; Pollegioni et al. 2015).

Despite their importance, little is known about the genetic diversity and structure of wild populations of common walnut in China. Most studies of *J. regia* in China have focused on cultivar development (Chen et al. 2014), the characterization of breeding germplasm (Wang et al. 2011; Ning et al. 2011; Li et al. 2011), or the relationship of *J. regia* in China to the rest of Asia (Pollegioni et al. 2015). Wild populations of *J. regia* in the Qinling Mountains of central China showed low genetic differentiation among sampled sites based on ITS (internal transcribed spacer) sequences; genetic variation was mainly within populations (Hu et al. 2014).

Whether *J. regia* and *Juglans sigillata* are legitimately distinct taxa in China has been controversial, but what is clear is that the two species freely hybridize and represent closely related gene pools. There is evidence of introgression between the two species in regions of sympatry, or else there is incomplete lineage sorting (Gunn et al. 2010; Wang et al. 2008; Wang et al. 2015).

Floristic and phylogeographic zones in China are sketchy, but the flora of East Asia can be divided into northern and southern regions (Xiang and Soltis 2001; Milne 2006). Species in the genus *Juglans* have been shown to have distinct genetic breaks across China, but without a consistent north–south divide. For example, the genetic data consistently identified two clades for Asian butternuts, one northern, comprising *Juglans mandshurica* and *Juglans ailantifolia*, and one southern, comprising *Juglans cathayensis*. Demes of *J. mandshurica* clustered into northeastern and northern Chinese populations (Bai et al. 2010), whereas demes of *J. cathayensis* separated into western and eastern populations (Bai et al. 2014). Dang et al. (2015) found that *J. cathayensis* in China had a pronounced north–south differentiation despite high levels of pollen mediated gene flow that reduced genetic structure at the nuclear level, although chloroplast lineages remained distinct and isolated (Bai et al. 2014).

The Quinling and Daba mountains in west-central China have an important role in shaping the biogeography and genetic structure of Chinese flora. The Qinling Mountains run east–west and act as a watershed divide between the two great rivers of China, the Yangtze (to the south), and the Yellow. The mountains also constitute a transition zone between a cool subtropical climate zone and a warm temperate zone (Guo et al. 2013; Cheng et al. 2013). The topographic and climatic variability of the region produces a wide range of ecological conditions that support exceptionally high levels of biodiversity (Wang et al. 2008; Bai et al. 2010; López-Pujol et al. 2011; Zhang et al. 2015; Cheng et al. 2013). Bai et al. (2010) suggested that sites in the Qinling Mountains could have been glacial refugia for temperate deciduous species.

Although microsatellites are especially useful tools for the study of intraspecific genetic variability and the genetic relationships among closely related species (Zeng et al. 2011; Wei et al. 2013; Pollegioni et al. 2015), sequence data from nuclear genes can provide important insights as well (Li et al. 2013; Wang et al. 2016; Du et al. 2015). The phenylalanine ammonia-lyase (*PAL*; EC 4.3.1.24) gene family serves as the introductory step of the phenylpropanoid pathway, which has roles in plant growth, development, responses to environmental stress, and disease resistance (Huang et al. 2010; Tonnessen et al. 2015). The *PAL* gene of *J. regia* has been cloned (Xu et al. 2012) and its expression studied (Christopoulos and Tsantili 2015).

In our study, we analyzed genetic diversity and population structure of wild *J. regia* from 13 sites in China using 10 EST-SSR markers and sequence polymorphism within the phenylalanine ammonia-lyase (*PAL*) ortholog *Jr5680*. Our sampled sites included the Daba and Qinling Mountains, regions known to support high levels of genetic diversity, and Yunnan Province, where *J. regia* and *J. sigillata* are sympatric, as well as regions of contact between *J. regia* and other *Juglans* native to China. Our objectives were to (1) determine the genetic diversity and population structure of *J. regia* in eastern China, (2) estimate the degree of population differentiation and genetic structure of *J. regia* populations from the north versus south of the Qinling Mountains, and (3) to evaluate whether any regions of China contained exceptional levels of genetic diversity for *J. regia* and to evaluate whether hybridization with congeners, especially *J. sigillata*, has affected the diversity of *J. regia* in China.

## Materials and methods

### Sample collections and DNA extraction

In 2014, leaf samples of 144 *J. regia* individuals were collected from 13 geographically distinct regions in China (Table 1). We also sampled 30 *J. sigillata* from three populations in southern China and 10 individuals of *J. mandshurica* from Beijing (Table S1). All sampled trees were mature adults, apparently healthy, growing in a mountain forest, along a forest road, or near a village but not in an orchard or on farmed land. Sampled trees were separated by at least 50 m. Sampled locations were mapped using ArcGIS (version 10.0; ESRI

**Table 1** Source information for 13 *Juglans regia* populations sampled in China

| Code | Sample site | Samples (N) | Latitude (N) | Longitude (E) | $N_A$ | *Ho* | *He* | *uHe* | *F*is |
|---|---|---|---|---|---|---|---|---|---|
| KC | Chengde, Hebei[a] | 8 | 40°36′41″ | 118°29′07″ | 1.70 ± 0.21 | 0.34 ± 0.10 | 0.31 ± 0.07 | 0.32 ± 0.07 | −0.01 ± 0.21 |
| YC | Xinzhou, Shanxi[a] | 7 | 38°24′12″ | 112°43′24″ | 1.80 ± 0.33 | 0.37 ± 0.12 | 0.32 ± 0.08 | 0.33 ± 0.09 | −0.06 ± 0.21 |
| SD | Jinan, Shandong[a] | 21 | 36°33′33″ | 116°43′35″ | 2.00 ± 0.26 | 0.36 ± 0.07 | 0.39 ± 0.07 | 0.40 ± 0.07 | 0.12 ± 0.09 |
| SX | Qinshui, Shanxi[a] | 6 | 35°41′24″ | 112°11′15″ | 1.80 ± 0.25 | 0.33 ± 0.10 | 0.35 ± 0.07 | 0.37 ± 0.08 | 0.07 ± 0.19 |
| HN | Luoyang, Henan[a] | 6 | 34°45′55″ | 113°45′13″ | 1.60 ± 0.16 | 0.45 ± 0.12 | 0.35 ± 0.07 | 0.38 ± 0.08 | −0.53 ± 0.08 |
| CC | Baoji, Shaanxi[b] | 16 | 34°12′27″ | 107°18′14″ | 1.60 ± 0.22 | 0.38 ± 0.11 | 0.29 ± 0.08 | 0.29 ± 0.08 | −0.34 ± 0.10 |
| SY | Shangluo, Shaanxi[b] | 15 | 33°29′59″ | 109°53′09″ | 2.20 ± 0.33 | 0.31 ± 0.08 | 0.38 ± 0.07 | 0.39 ± 0.07 | −0.30 ± 0.16 |
| GM | Guanmenshan, Hubei[b] | 2 | 31°44′19″ | 110°40′50″ | 1.10 ± 0.10 | 0.15 ± 0.10 | 0.08 ± 0.05 | 0.20 ± 0.10 | −1.00 ± 0.10 |
| SC | Nanchong, Sichuan[c] | 10 | 30°52′33″ | 105°55′30″ | 1.70 ± 0.22 | 0.28 ± 0.10 | 0.39 ± 0.08 | 0.42 ± 0.09 | 0.42 ± 0.23 |
| EM | Emei, Sichuan[c] | 5 | 29°35′29″ | 103°31′21″ | 1.90 ± 0.28 | 0.33 ± 0.11 | 0.39 ± 0.08 | 0.44 ± 0.09 | 0.25 ± 0.23 |
| LJ | Lijiang, Yunnan[c] | 14 | 26°51′18″ | 100°13′40″ | 2.00 ± 0.26 | 0.30 ± 0.08 | 0.32 ± 0.08 | 0.33 ± 0.08 | 0.13 ± 0.12 |
| GZ | Guiyang, Guizhou[c] | 14 | 26°30′33″ | 104°40′37″ | 1.70 ± 0.26 | 0.25 ± 0.07 | 0.29 ± 0.07 | 0.30 ± 0.07 | 0.27 ± 0.17 |
| BS | Baoshan, Yunnan[c] | 20 | 25°7′12″ | 99°10′48″ | 1.80 ± 0.29 | 0.42 ± 0.13 | 0.34 ± 0.08 | 0.35 ± 0.08 | −0.33 ± 0.23 |

[a] Northern China

[b] Indicates Qinling mountains and Bashan mountains

[c] Indicates populations from Southern China

2010). All fresh leaves were dried with silica gel prior to DNA extraction, which followed the methods described by Doyle and Doyle (1987) and Zhao and Woeste (2011). DNA was diluted with water to a final concentration of 10 ng/μL and stored at −20 °C.

### EST-SSRs primer design and PCR amplification

Ten polymorphic EST-SSRs were selected from the congeners *J. hopeiensis* (JH84548, JH89978, Hu et al. 2015a), *J. mandshurica* (JM61666, JM68820, JM78331, and JM5969, Hu et al. 2015b), and *J. cathayensis* (JC8125, Dang et al. 2015); two other loci were previously unpublished (JR4616, JR6439, and JR3773, Table 2). All the EST-SSR-containing unigene sequences were searched using BLAST in the NCBI database to identify their genic context (Table S2). We mapped the source species (*J. hopeiensis*, *J. mandshurica*, and *J. cathayensis*) EST-SSR sequence data onto the *J. regia* transcriptome to verify that these SSRs were also in expressed sequences of *J. regia* (Table S3). All primers were synthesized by Sangon Biotech in Shanghai, China (Table 2). PCR amplification was carried out on a SimpliAmp Thermal Cycler (Applied Biosystem, Hercules, CA, USA) in 10-μL reaction volumes (5 μL 2× PCR Master Mix,0.4 μL each primer, 1 μL BSA, 1 μL of 10 ng/μL DNA). The PCR was programmed for 3 min at 95 °C followed by 32 cycles of 15 s at 93 °C, 1 min at annealing temperature (Tm; Table 2), 30 s at 72 °C and extension of 10 min at 72 °C. PCR products were separated by size on 10 % polyacrylamide gels and visualized by silver staining. Fragment sizes of each locus were estimated using Quantity One Software (Bio-Rad Laboratories, ) and a 50 bp DNA ladders size standard (pBR322 DNA/MspI marker; Tiangen, Beijing, China). We verified polyacrylamide results by re-genotyping a sample of all genotypes using an ABI 3730ABI and fluorescently labeled primers. We performed the PCR amplification using a Veriti 96-Well Thermal Cycler (Applied Biosystems, Foster City, CA, USA). The upper primers were labeled with fluorescent dye, 6-FAM, HEX, TAMRA or ROX (Sangon, Shanghai, China), 1 μL BSA, 2 μL of 10 ng/μL DNA. ABI3730 DNA analyzer (Applied Biosystems) was used to score genotypes and the software GeneMarkers version 2.0 was used for binning (Holland and Parson 2011).

### *PAL* gene marker primer design and PCR amplification

We investigated genetic variation within *Jr5680*, a unigene identified in our lab using RNA-seq. The total unigene sequence of *Jr5680* was 2369 bp (Table S4). *Jr5680* and the *JrPAL* gene sequence in GenBank (accession no. JX069977, Xu et al. 2012) are 96 % identical. Primers to amplify the *PAL* gene were identified using Primer3 (F:TGGACATG GCTAGTGACTGG; R: TCCCTCTTGCCTACATTGC; Sangon Biotech in Shanghai, China) based on RNA sequence data for *Jr5680*. The primers were designed to amplify between nucleotides at position 838 and 1661 (Table S4). The *Jr5680* sequence (from *J. regia*) was compared to (unpublished) transcript sequences for *J. hopeiensis, J. mandshurica, J. sigillata,* and *J. cathayensis* produced in our Lab. PCR amplification was carried out on a SimpliAmp Thermal Cycler (Applied Biosystem, USA) in 20 μL reaction volumes (10 μL 2× PCR Master Mixincluding0.1 U Taq polymerase/μL; 500 μM each dNTP; 20 mMTris-HCl (pH 8.3); 100 mMKCl; 3.0 mM $MgCl_2$ (Tiangen, Beijing,

**Table 2** Characteristics of 10 EST-SSR primers used in this study

| Primer | Primer sequence(5′–3′) | *Tm* | $N_A$ | *Ae* | *Ho* | *He* | *I* | Source |
|---|---|---|---|---|---|---|---|---|
| JR4616 | F:AGCCCTTTTGCATCGGCTAT<br>R:AGCTGACCGATCGATCAACA | 55 | 2 | 2.5 | 0.400 | 0.497 | 0.372 | Unpublished |
| JR6439 | F:TCGATGCGATCATCTCCGTG<br>R:CGGCACCAAAACAGAACTCG | 55 | 4 | 3.4 | 0.381 | 0.624 | 0.542 | Unpublished |
| JR3773 | F:GGTGGTTTGACCCTTAATTCTGT<br>R:ACCCTGCCACAATGACCAAA | 55 | 4 | 3.1 | 0.207 | 0.248 | 0.226 | Unpublished |
| JM61666 | F:AACTGTTGCCGGAGCTTTCT<br>R:TGGGATAACACCACATGCAGT | 55 | 2 | 4.1 | 0.158 | 0.187 | 0.189 | Hu et al. 2015b |
| JM68820 | F:TCCTTCTGTGTGAGTGCGTG<br>R:GGTCAGGTGAGTGGAGCAAA | 55 | 4 | 2.7 | 0.319 | 0.616 | 0.535 | Hu et al. 2015b |
| JM78831 | F:GCAGTGCGCTCCTTTTTCAA<br>R: TTCTCGGGTTGAAGCCACAA | 58 | 2 | 1.2 | 0.123 | 0.156 | 0.168 | Hu et al. 2015b |
| JM5969 | F:ACAATAGTCTCTGCACCGCC<br>R: AGCTTGTACTTACCGCCGAC | 55 | 6 | 7.9 | 0.577 | 0.594 | 0.527 | Hu et al. 2015b |
| JC8125 | F: AGCAACCAGAGCAGAGCATT<br>R: AACCTCAACACCAACTATGCT7 | 55 | 4 | 3.9 | 0.261 | 0.393 | 0.358 | Dang et al. 2015 |
| JH84548 | F:TCTGAGGAAGCTGCATGGAA<br>R: AACTCTGGACACATGCCGC | 58 | 3 | 1.5 | 0.050 | 0.049 | 0.048 | Hu et al. 2015a |
| JH89978 | F:ACCTTCCCTGCTCCTCTCTT<br>R:GAGCCTTGTGGAAGCAAACG | 55 | 3 | 3.8 | 0.257 | 0.451 | 0.406 | Hu et al. 2015a |
| Mean | | | 3.4 | 3.4 | 0.273 | 0.381 | 0.337 | |

*Tm* annealing temperature, *A* number of alleles, *Ae* effective number of alleles, $H_O$ observed heterozygosity, $H_E$ expected heterozygosity, *I* Shannon's information Index

China),0.5 μL each primer, 2 μL BSA, 2 μL of 10 ng/μL DNA). The PCR was programmed for 3 min at 94 °C followed by 35 cycles of 15 s at 93 °C, 1 min at annealing temperature (60 °C), 30 s at 72 °C and extension of 10 min at 72 °C. After PCR amplification, fragments were sequenced by Sangon Biotech (Shanghai, China).

## Data analysis

Genetic diversity per locus and population were evaluated based on the following descriptive summary statistics: number of alleles ($N_A$), number of effective alleles (*Ne*), observed ($H_O$), and expected ($H_E$) heterozygosity, Shannon's diversity index (*I*),genetic differentiation coefficient ($F_{ST}$), fixation index ($F_{IS}$), and proportion of polymorphic loci (*PPL*) using GenAlEx6.5 (Peakall and Smouse2012). GENEPOP version 4.2 (Rousset 2008) was used to test Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for all loci (Table 2). Their significance was assessed with 1500 permutations. CERVUS version 3.0 (Kalinowski et al. 2007) was used to calculate polymorphic information content (PIC). Null alleles were predicted using MICRO-CHECKER 2.2.3 (Van Oosterhout et al. 2004). Genetic structure analysis was performed using the software STRUCTURE version 2.3.4. We used a burn-in of 100,000 Markov Chain Monte Carlo iterations, followed by 1,000,000 iterations and 10 replicates per run for *K* from 2 to 10 clusters with the admixture model (Pritchard et al. 2000; Evanno et al. 2005). The program STRUCTURE HARVESTER was used to calculate the optimal value of *K* using the delta *K* criterion (Earl 2012), the inferred clusters were drawn as colored box plots using program DISTRUCT (Rosenberg 2004). The overall genetic variation based on EST-SSR data was partitioned using principal coordinates analysis (*PCoA*) using GenALEx 6.5 (Peakall and Smouse 2012). The number of clusters was determined by using the k-means method (Peakall and Smouse 2012). IBD software (http://ibdws.sdsu.edu/~ibdws/) was used to perform the Mantel test of association between geographic distance and genetic distance based on the IBDWS method (Bohonak 2002; Jensen et al. 2005). A plot of genetic distance against geographic distance was generated using the allele data and geographic information from IBD. Neighbor joining (NJ) based on Nei's genetic distance (Nei 1978) was performed using GENEPOP version 1.2 (Raymond and Rousset 1995). AMOVA based on pairwise differences were performed in Arlequin v. 3.5 (Excoffier and Lischer 2010).

The *PAL* gene sequence data were edited and aligned using Bioedit version 7.0.8 (www. mbio.ncsu.edu/BioEdit/bioedit.html). DnaSP version 5.0 (Librado and Rozas 2009) was used to calculate the number of haplotypes (Table 4), nucleotide diversity ($\pi$), haplotype diversity ($H_d$), Tajima's *D*, and Fu's *Fs*. We used NETWORK v 4.6.1 (Fluxus Technology Ltd) to generate minimum spanning tree using median-joining (Bandelt et al. 1999). $G_{ST}$ and $N_{ST}$ were estimated to take into account the evolutionary distance of haplotypes using Permut (Pons and Petit 1996). Neutrality and mismatch distribution was detected using DnaSP version 5.0

(Librado and Rozas 2009). Arlequin v. 3.5 (Excoffier and Lischer 2010) was used to calcluate the genetic differentiation coefficient ($F_{ST}$) for the *PAL* gene. The Spearman and Pearson correlation between $F_{ST}$ values based on EST-SSRs and *PAL* was calculated using SPSS (Coakes and Steed 2009).

## Results

### Genetic diversity and population structure based on EST-SSRs

A total of 34 alleles were detected from 144 samples in 13 demes based on 10 EST-SSRs (Table 2). SSRs were only evaluated in *J. regia*, not *J. sigillata* or any other *Juglans* species. A private allele was identified at high frequency (0.4) at the BS sample site; in YC at JM68820 and JM 78331 (allele frequencies were 0.10 and 0.20, respectively); in GZ at JR6439, JC8125, and JH84548 (allele frequencies were 0.143, 0.156, and 0.148 respectively); in CC at JR3773, JM61666, and JC8125 (allele frequencies were 0.095, 0.143, and 0.095, respectively); and in SY at JM78331 (allele frequency was 0.50). The numbers of alleles per locus ($N_A$) ranged from 2 to 6, with a mean of 3.4. Locus JM5969 had the largest number of alleles ($N_A = 6$). The observed heterozygosity ($H_O$) and expected heterozygosity ($H_E$) varied from 0.050 to 0.577 ($\bar{x} = 0.273$) and from 0.049 to 0.594 ($\bar{x} = 0.382$), respectively. Shannon's information Index (I), ranged from 0.048 to 0.542, with a mean of 0.337. Three loci showed significant departure from Hardy-Weinberg equilibrium (HWE), JC8125 (six populations), JR3773 (three populations), and JR6439 (six populations). Genetic differentiation among populations was high ($F_{ST} = 0.280$, $P < 0.001$), as was the fixation index ($F_{IS} = 0.127$), although $F_{IS}$ was not different than zero for eight of the 13 tested populations (Table 3). The Bayesian clustering software STRUCTURE separated the samples into four populations based on EST-SSRs (Evanno et al. 2005, Figs. 1 and Fig. S1). Most of the samples clustered into populations that corresponded to the geographic region from which they were sampled (Fig. 1). Principal coordinate analyses accounted for 54.95 % of the observed variance on the first two coordinate axes (Fig. 3) and revealed three clusters of samples similar to the groups identified by STRUCTURE (Figs. 1a–b, 3, S1). A NJ tree based on Nei's (1978) genetic distance (Fig. 1c) supported clustering the samples into three groups as well, but the groups did not exactly correspond to those identified using principal coordinate analysis and k-means method (Figs. 3 and S5). Both STRUCTURE and NJ separated population BS, which was from the extreme southwestern edge of the sampled range (red in Fig. 1a, b), from other southern populations (LJ, EM, GZ, and SC). The remaining trees comprised two northern populations, based on STRUCTURE; trees from CC, SY, HN, and KC (a mostly central China cluster, green in Fig. 1a, b), and a generally northeastern population (yellow in Fig. 1a, b, comprised mostly of trees from GM, SX, SD, and YC). The NJ analysis did not separate the central and northeastern populations (yellow and green in Fig. 1a, b). The AMOVA (northern vs southern regions as described in Table 1), the results of STRUCTURE at $K = 2$ (Fig. S1; Table S6), and the neighbor-joining tree (Fig. 1, Table 1) supported a north–south genetic divide (Table S6, Fig. S1). A north/south division was also clear when demes with 10 or fewer individuals were removed from the analysis, although the location SC flips from "southern" to "northern" (Fig. S2). A Mantel test for matrix correlation between genetic distance and geographic distance was significant ($r = 0.45$, $P < 0.001$, Fig. S3), possibly indicating isolation by distance contributed to the current genetic structure.

### *PAL* gene haplotype diversity and differentiation

The *Jr5680* sequence, an apparent ortholog of the nuclear gene phenylalanine ammonia-lyase (*PAL*), had a total length of 823 bp. Polymorphism within the sequence of *Jr5680* constituted four haplotypes in *J. regia*, H1–H4 (Fig. 2, Table 4). The total haplotype diversity ($H_d$) for the *J. regia* samples was 0.345, and the nucleotide diversity ($\pi$) was 0.00085. When each sampled site was considered separately, the haplotype ($H_d$) and nucleotide ($\pi$) diversities ranged from 0.000 to 0.667, and 0.000 to 0.00169, respectively. The haplotype diversity was highest in population SD, followed by GZ, with no diversity found in seven *J. regia* populations (Fig. S2). All nucleotide changes that contributed to haplotype diversity resulted in predicted amino acid changes (Table 4).

Haplotype H1 was the dominant haplotype across all the sampled sites; it was found in the southernmost and northernmost sites, and H1 was a minority haplotype in only three sites. Haplotype H1 was also the ancestral haplotype based on network analysis that included *J. mandshurica* as the outgroup. Four of the five northern demes were fixed for the most common haplotype (H1), the only exception was samples from Dezhou (SD), which contained a private haplotype (H4) and a few individuals with haplotype H2. Haplotype H2 was the second most commonly observed. It was prominent only in the southern demes SC and BS, but was also found in the Qinling mountains and in the northern deme SD. Haplotype 3, while not restricted to samples from the Qinling and Daba mountains (CC and SY), was present in the southern deme Guiyang (GZ) only as a small minority. GZ and SD were the only demes to contain three haplotypes, and while several populations contained only H1, no population contained only H2 or H3. Haplotype H3 was between H1 and H2 based on network analysis (Fig. 2). Haplotypes H1 and H2 were shared by *J. regia* and its close relative *J. sigillata*, but *J. sigillata* also showed private *PAL* haplotypes H6 and H7 that were different from all *J. regia* samples

**Table 3** Pairwise $F_{ST}$ matrix for 13 *Juglans regia* populations in China based on EST-SSRs and *PAL* sequence

| Population | BS | GZ | LJ | EM | SC | GM | SY | CC | HN | SX | SD | YC | KC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| BS | – | 0.213 | −0.017 | 0.225 | 0.129 | 0.227 | 0.443 | 0.339 | 0.481 | 0.565 | 0.325 | 0.518 | 0.567 |
| GZ | 0.210 | – | 0.396 | 0.124 | 0.325 | 0.365 | 0.597 | 0.532 | 0.538 | 0.258 | 0.452 | 0.025 | 0.386 |
| LJ | 0.216 | 0.164 | – | 0.000 | 0.562 | 0.000 | 0.362 | 0.496 | 0.000 | 0.000 | 0.128 | 0.000 | 0.000 |
| EM | 0.274 | 0.261 | 0.328 | – | 0.687 | 0.000 | 0.405 | 0.781 | 0.000 | 0.000 | 0.395 | 0.000 | 0.000 |
| SC | 0.244 | 0.306 | 0.277 | 0.001 | – | 0.121 | 0.450 | 0.622 | 0.688 | 0.549 | 0.811 | 0.126 | 0.210 |
| GM | 0.210 | −0.204 | −0.204 | −0.250 | 0.135 | – | 0.231 | 0.279 | 0.000 | 0.000 | 0.068 | 0.000 | 0.000 |
| SY | 0.435 | 0.406 | 0.398 | 0.265 | 0.153 | 0.091 | – | 0.036 | 0.328 | 0.326 | 0.319 | 0.098 | 0.258 |
| CC | 0.349 | 0.540 | 0.495 | 0.374 | 0.209 | 0.320 | 0.357 | – | 0.532 | 0.631 | 0.388 | 0.514 | 0.595 |
| HN | 0.335 | 0.359 | 0.304 | 0.263 | 0.172 | 0.052 | 0.281 | 0.263 | – | 0.000 | 0.251 | 0.000 | 0.000 |
| SX | 0.305 | 0.320 | 0.374 | 0.109 | 0.084 | −0.115 | 0.230 | 0.227 | 0.130 | – | 0.124 | 0.000 | 0.000 |
| SD | 0.305 | 0.267 | 0.300 | 0.087 | 0.003 | −0.141 | 0.177 | 0.148 | 0.117 | 0.034 | – | 0.056 | 0.128 |
| YC | 0.479 | 0.365 | 0.314 | 0.163 | 0.236 | −0.064 | 0.213 | 0.156 | 0.173 | 0.126 | 0.061 | – | 0.000 |
| KC | 0.397 | 0.442 | 0.459 | 0.329 | 0.205 | 0.208 | 0.004 | 0.316 | 0.316 | 0.259 | 0.195 | 0.268 | – |

Upper matrix is pairwise $F_{ST}$ based on *PAL* gene, lower matrix is pairwise $F_{ST}$ based on EST-SSRs. The Pearson correlation of pairwise $F_{ST}$ values based on EST-SSR and those based on *PAL* was low but significant ($r = 0.281$, $P = 0.013$), and the Spearman correlation was similar ($r = 0.272$, $P = 0.016$)

(Fig. 2). *J. sigillata* was only species that shared haplotypes with *J. regia*. All 10 samples of *J. mandshurica* were haplotype H5 (Fig. 2). Some individuals from GZ (20 %) and BS (21 %) were heterozygous for haplotypes H1 and H2. The permutation test for haplotype structure was non-significant ($G_{ST}$ was 0.373) probably because the two most common haplotypes (H1 and H2) were found in both northern and southern China. $N_{ST}$ (0.372) was almost exactly equal to $G_{ST}$, so haplotypes were phylogenetically equivalent (Pons and Petit 1996). Tajima's $D$ was 0.98 ($P > 0.05$), and Fu's $Fs$ test statistic was 1.22, indicating no effect of a relatively ancient bottleneck. No significant deviation from neutrality and mismatch distribution was detected using the Hudson–Kreitman–Aguade tests ($\chi^2 = 0.12$, $P = 0.23$). The genetic differentiation at *Jr5680* among all sampled sites was moderately high ($F_{ST} = 0.164$, $P < 0.001$, based on *PAL* sequence only). There was no evidence for IBD at the *PAL* locus.

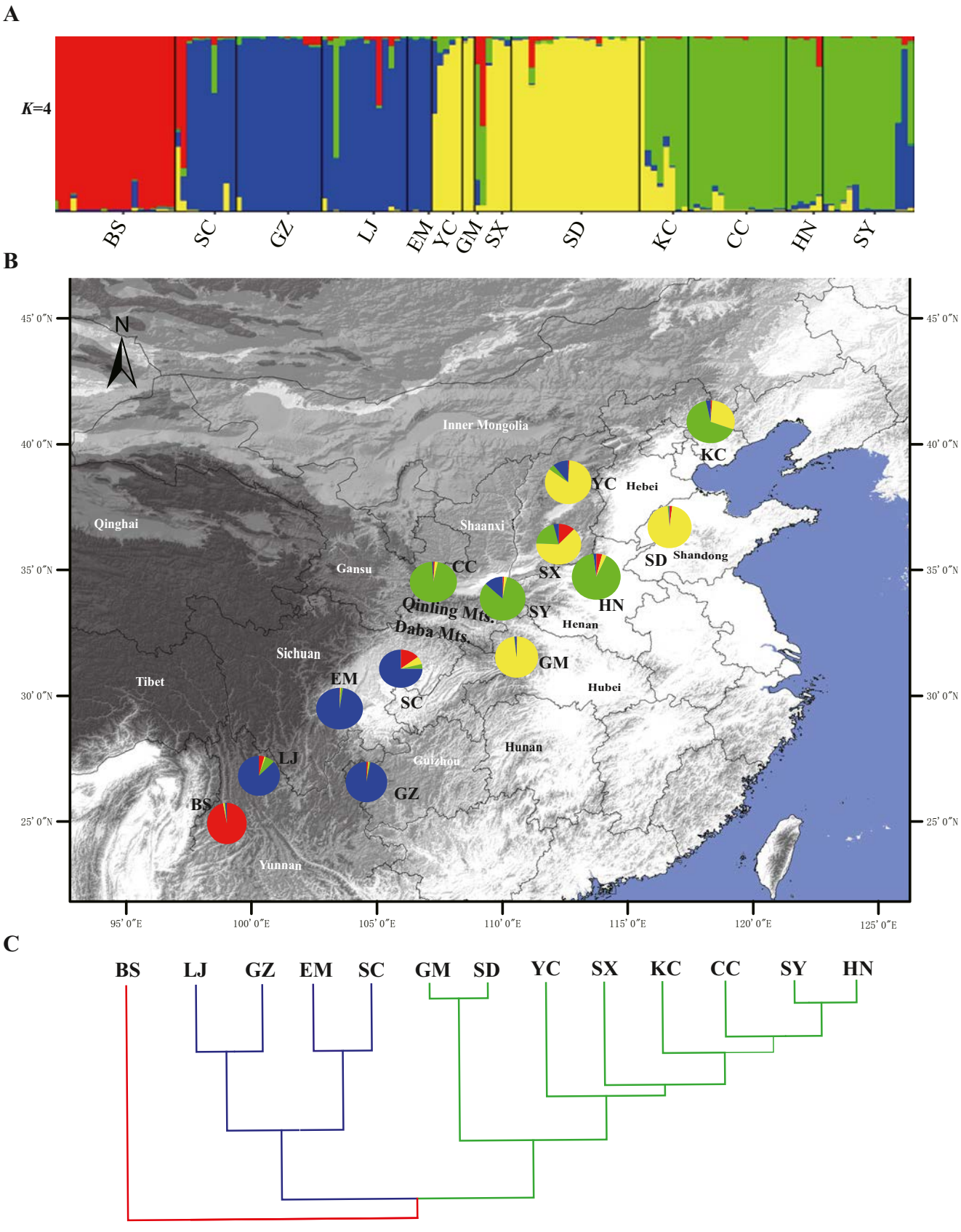
A search of transcriptomic data from *J. hopeiensis*, *J. sigillata*, and *J. cathayensis* found that transcripts highly similar to *Jr5680* were present in all three congeners. The high level of identity and level of haplotype sharing among *Juglans* species gave us confidence that we compared homologs (within species) and orthologs (across species). The available transcript from *J. cathayensis* was 316 bp; it was 98.1 % identical to one portion of the *J. regia PAL* gene (*Jr5680*, 2369 bp, Tables S4, S5; Fig. S4). The *J. hopeiensis* transcript was 345 bp, it was nearly 82.6 % identical to a portion of the *J. regia PAL* gene (*Jr5680*, 2369 bp, Table S4, S5; Fig. S4). The *J. sigillata* transcript was 2407 bp; it was highly similar to the sequence from the transcriptome of *J. regia* (*PAL* gene; 2369 bp, Table S4; Fig. S4). Two of four *J. sigillata* haplotypes based on predicted protein sequences were also found in *J. regia* (H1 and H2; Table S5).

## Discussion

In general, quantification and description of genetic diversity is essential for understanding plant species evolution and adaptation (Brady et al. 2005; Jiang et al. 2012). In this experiment, 10 EST-SSRs and sequence polymorphism within the nuclear gene phenylalanine ammonia-lyase were used to understand the genetic diversity and structure of common walnut (*J. regia*) in China. Small sample sizes from some demes limit the strength of our conclusions related to population differentiation, but several lines of evidence point to clear spatial genetic structure among the sampled locations.

The presence of genetic structure among *J. regia* populations in China is clearly evident from the results of the STRUCTURE analysis (Fig. 1a, b), *PCoA* (Fig. 3), $F_{ST}$ (Table 3), and the NJ tree (Fig. 1c), but the historical and biogeographical causes of the existing genetic structure can only be inferred in rough outlines. In general, natural processes including isolation, selection/adaptation, hybridization, and drift can lead to the formation of genetic structure, and these can be augmented or ameliorated by human intervention. Human dispersal typically contributes to a general reduction of genetic structure of cultivated plants, but human dispersal

**Fig. 1** Genetic structure and geographical distribution of 13 apparently natural populations based on 10 EST-SSRs. **a** Results of STRUCTURE 2.3.4 analysis (Pritchard et al. 2000) showing $K = 4$. **b** Geographic origin of the 13 populations and their color-coded grouping at the most likely $K = 4$ as determined using the delta$K$ method of Evanno et al. (2005). The location of the pie indicates the sampled locations. The proportion of colors in each pie reflects the proportion of genetic affiliation within each of the four populations determined by STRUCTURE averaged over all samples from that location. **c** Dendrogram of NJ tree based Nei's (1978) genetic distance


A
K=4
BS
SC
GZ
LJ
EM
YC
GM
SX
SD
KC
CC
HN
SY
B
N
45' 0"N
40' 0"N
35' 0"N
30' 0"N
25' 0"N
95' 0"E
100' 0"E
105' 0"E
110' 0"E
115' 0"E
120' 0"E
125' 0"E
Inner Mongolia
Qinghai
Shaanxi
Gansu
Hebei
Shandong
Henan
Hubei
Sichuan
Tibet
Hunan
Guizhou
Yunnan
Qinling Mts.
Daba Mts.
KC
YC
SD
SX
CC
HN
SY
GM
EM
SC
LJ
GZ
BS
C
BS
LJ
GZ
EM
SC
GM
SD
YC
SX
KC
CC
SY
HN

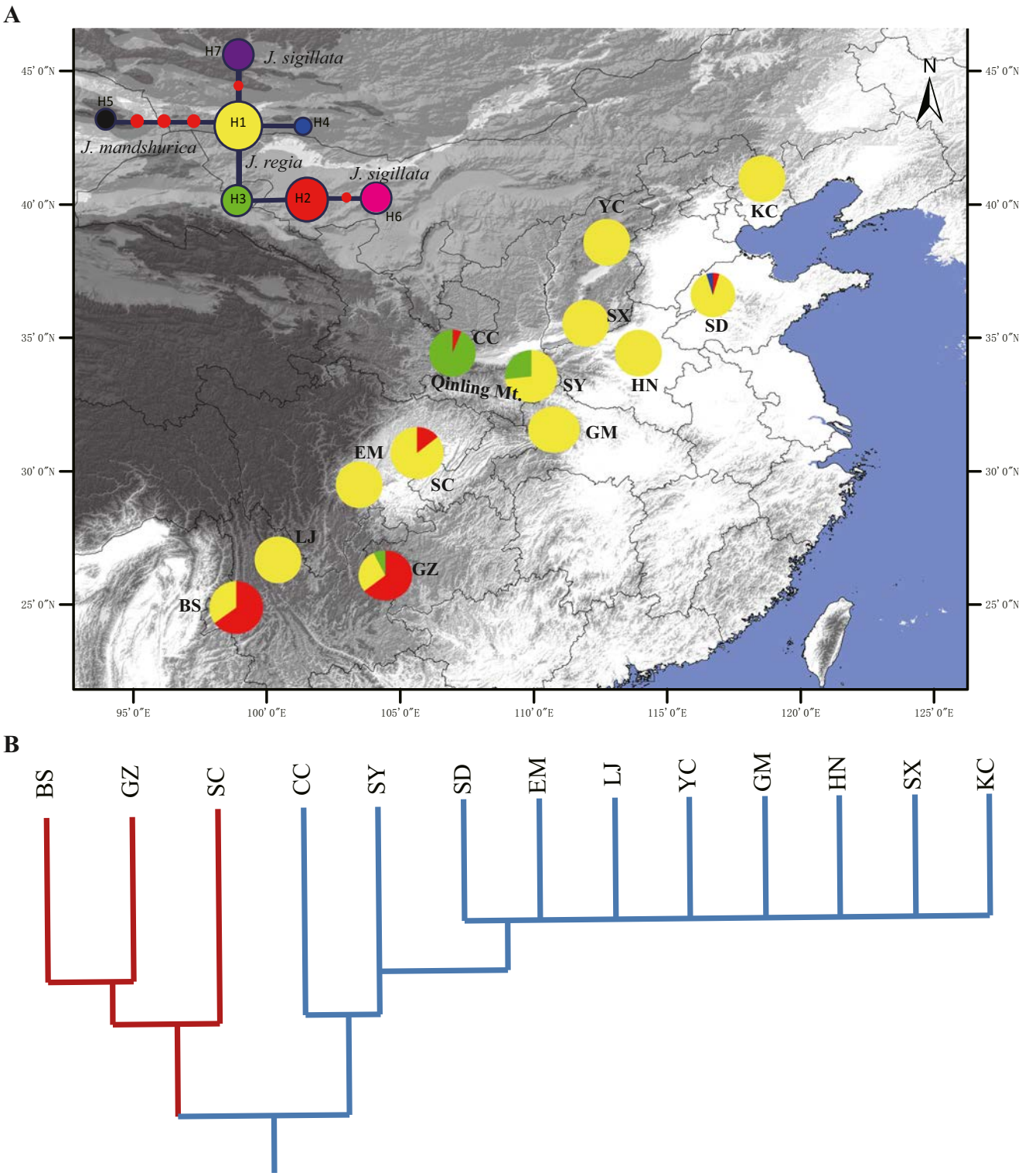



**Fig. 2** Distribution of haplotypes and NJ tree of the nuclear gene phenylalanine ammonia-lyase (*PAL*). **a** Population codes are identified in Table 1. The *upper left panel* shows the minimum spanning network of the four *J. regia* haplotypes. *Circle sizes* in the network are proportional to the number of samples per haplotype. **b** Dendrogram of NJ tree based Nei's (1978) genetic distance

can also move cultivated plants into new habitats. For example, it is not certain that *J. regia* is native to many of the parts of China we sampled (Pollegioni et al. 2015). If the dispersal of *J. regia* across the parts of China we sampled is a relatively recent event, then its current genetic structure is likely the result of population bottlenecks caused by human selection,

**Table 4** Four nuclear haplotypes in *J. regia* based on amino acid sequence variation within *PAL* gene

| Haplotype | Amino acid position | | | | | | |
|---|---|---|---|---|---|---|---|
| | 46 | 52 | 69 | 141 | 166 | 200 | 265 |
| H1 | F | Q | E | G | H | L | Q |
| H2 | V | Q | E | G | H | L | Q |
| H3 | V | R | G | G | R | P | R |
| H4 | V | Q | E | D | H | L | Q |

Amino acid positions based on sequence of *Jr5680*, NCBI accession numbers: KT820730, KT820731, KT820732, KT820733

possibly followed by hybridization, rather than natural evolutionary or biogeographic forces.

The mating system of walnuts is predominantly out-crossing, typically resulting in high diversity within populations (Victory et al. 2006; Pollegioni et al. 2014). So we were surprised when our data indicated relatively high levels of genetic differentiation among geographically isolated demes of *J. regia* (among populations of *J. regia* ($F_{ST}$ = 0.280, $P < 0.001$ based on EST-SSR; $F_{ST}$ = 0.164, $P < 0.001$, based on *PAL*). This high level of differentiation is unusual when compared with *Juglans* populations sampled in other studies, and studies of other wind pollinated species. The relatively high genetic differentiation and the correlation of genetic and geographic distance probably indicate that some IBD is present (Fig. S3); perhaps wild populations of *J. regia* in China are isolated except for gene flow from nearby cultivated walnuts.

Phenylalanine ammonia-lyase (*PAL*) catalyzes the first step of the phenylpropanoid pathway, which produces precursors to a variety of important secondary metabolites (Andersen et al. 2007; Huang et al. 2010). The importance of *PAL* in multiple metabolic pathways made it an attractive candidate for examining genetic diversity. There was evidence of spatial structure based on the *PAL* sequence data. *PAL* haplotype H3 was found in the Qinling Mountains (populations CC and SY). This indicates that gene flow has not homogenized all nuclear alleles, possibly. Because H3 recently evolved, because H3 represents an ancient haplotype preserved in only certain refugial lineages, or possibly because H3 represents an adaptation not well suited to other Chinese environments. These populations, as well as the Daba Mountains population (GM) appeared to cluster with northern Chinese demes based on EST-SSRs (Figs. 1 and 2). Thus, the samples from the Qinling Mountains were not distinct based on SSR data but they were distinct based on *PAL* sequence polymorphism (Figs. 1 and 2). Other evidence for regional genetic structure included the concentration of *PAL* haplotype H2 in southern Chinese populations, especially in BS and GZ, and the near absence of H2 in northern China (it was found only in two individuals in SD; Fig. 2).

The Qinling Mountain range is particularly rich in plant diversity (Wang et al. 2008; Bai et al. 2010; Chen et al. 2014; Guo et al. 2013). Both CC and SY contained private alleles, and both Qinling sites had distinctive genetic differentiation based on EST-SSR and *PAL* sequences compared to samples from elsewhere (Table 3, Fig. 2). It has been suggested that the Qinling Mountains were refugia for other temperate deciduous species (*Ostryopsis davidiana* and *J. mandshurica*) with reproductive biology similar to that of common walnut (Tian et al. 2009; Bai et al. 2010). Thus, the distinctive genetic diversity of *J. regia* in central China may be a consequence of its descent from a disjunct refugium, the region's considerable topographic, climatic, and ecological variability (Bai et al. 2010; López-Pujol et al. 2011; Zhang et al. 2015), or human dispersal, but our data did not show evidence that the Qinling Mountains represent a topographic barrier to *J. regia* gene flow.

The presence of a northern/southern divide in *J. regia*, as suggested for some other Chinese species, cannot be finally confirmed without additional data. Based on STRUCTURE, two populations (red and blue in Fig. 1) were found in warmer

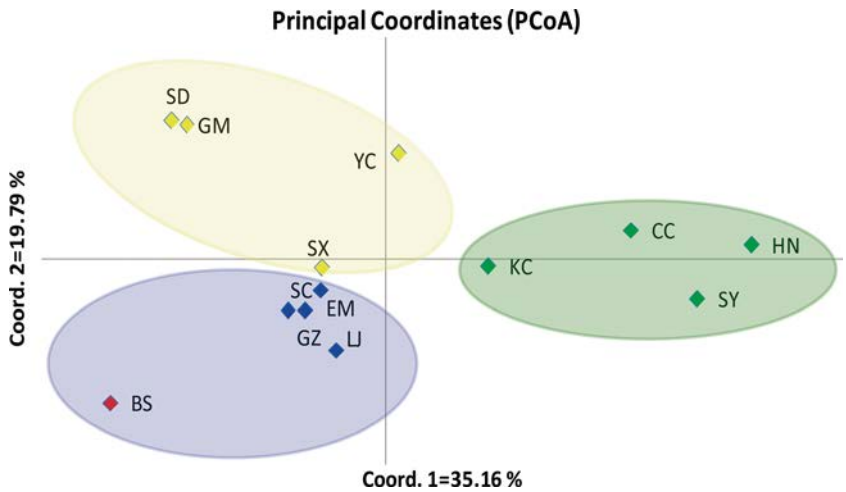



**Fig. 3** Principal coordinate analyses (PCoA) of 13 common walnut (*J. regia*) populations resolved into four genotype groups based on 10 microsatellite loci. *Yellow circle*: cluster I, *red circle*: cluster II, *blue circle*: cluster III; *green circle*: cluster IV

and more southern populations, and two different genetic clusters (green and yellow) in northern China (Fig. 1). A similar division of the populations was identified by *PCoA*, and a north/south divide also appeared in the AMOVA and NJ tree. But the assignment of some populations (north or south) was ambiguous, especially if only locations with a large number of samples were analyzed, and no north/south divide was apparent in the *PAL* data.

As mentioned previously, *J. regia* may not have a long history at some or all of the sites we sampled, and the role of long distance dispersal by humans in shaping the current spatial genetic structure of Common walnut in China is uncertain (Gunn et al. 2010). Humans have used and selected *J. regia* in Asia for over 6800 years, and evidence points to long-range dispersal by humans for both its timber and nuts (Pollegioni et al. 2015; Beer et al. 2008; Joly and Visset 2009). For example, we found that some *PAL* haplotypes were shared between distant (southern and northern) populations such as YC, KC, and SD. The reason may be selection and dispersal by humans of seeds or scion wood (Gunn et al. 2010; Pollegioni et al. 2011; Pollegioni et al. 2015), especially into northern China.

Hybridization may have played a role in the genetic differentiation we observed among common walnut populations. Hybridization within *Juglans* is well-attested, and common walnut is sympatric with *J. cathayensis*, *J. sigillata*, and *J. mandshurica* in various parts of China. There was overlap in haplotypes between *J. regia* and *J. sigillata* but not other *Juglans* species. Thus, gene flow from *J. sigillata* may have contributed to local or regional differentiation of *J. regia* populations (Wang et al. 2008, 2015; Gunn et al. 2010), but this may not be true in other regions where *J. regia* is sympatric with other Chinese *Juglans* species. This possibility is reinforced by the distinctiveness of *J. regia* in Yunnan, the only province where *J. regia* and *J. sigillata* naturally co-occur. The BS samples from Yunnan Province contained a private allele and BS samples were highly differentiated from other sampled populations based on EST-SSRs. They were similar to only one other (southern China) deme (GZ) based on *PAL* gene sequence polymorphism (Fig. 1, Fig. 2, Table 3). Hybrids between *J. regia* and *J. sigillata* are known to be fertile (Fan and Xi 2002; Yang and Xi 1989). Based on a sample of 220 walnut trees from six Tibetan villages, Gunn et al. (2010) showed that although *J. regia* and *J. sigillata* are morphologically distinct, the two species were indistinguishable based on microsatellite data. The nature of the genetic relationships among common walnuts in Tibet, Yunnan Province, and the rest of China remains uncertain, and additional collection and analysis is warranted. Outside of possible hybridization with *J. sigillata*, it is also possible the Tibetan *J. regia* gene pool may be considerably different from that of most of China (Pollegioni et al. 2011, 2015).

This study was the first large-scale, population genetic analysis of the germplasm of *J. regia* in China. Especially surprising was the striking spatial patterns for haplotypes of the *PAL* gene and the unambiguous spatial genetic structure of *J. regia* based on EST-SSRs. The results provide guidance for germplasm collection, and increase the likelihood that genetic variance for valuable traits can be identified in uncultivated demes. Finally our study identified distinct regions such as Yunnan and Qinling Mountains where in situ genetic conservation and sampling for use by breeding programs may be warranted. A more complete picture of the biogeography and spatial genetic structure of common walnut is needed and possibly could be obtained by studying patterns of variation in the chloroplast genome of *J. regia* as previously reported for other *Juglans* species by Bai et al. (2010, 2015).

**Acknowledgments** The authors wish to thank Jia Yang, Li Feng, and Tao Zhou for sample collection. This work was supported by the National Natural Science Foundation of China (nos. 41471038, 31200500, J1210063), the Program for Excellent Young Academic Backbones funding by Northwest University (no. 338050070), the Northwest University Training Programs of Innovation and Entrepreneurship for Undergraduates (nos. 2015159 and 2016171), Changjiang Scholars and Innovative Research Team in University (no. IRT1174). Mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the U.S. Department of Agriculture and does not imply its approval to the exclusion of other products or vendors that also may be suitable.

**Data archiving statement** The different haplotypes of amino acid positions based on sequence of *Jr5680* was submitted on the National Center for Biotechnology Information (NCBI), the accession numbers were: KT820730, KT820731, KT820732, KT820733. The Unigene sequences and BLAST search results for 11 SSR-containing ESTs were provided as Supplementary Table S2. The total unigene sequence of *Jr5680* was provided as Supplementary Table S4.

## References

Andersen JR, Zein I, Wenzel G, Krützfeldt B, Eder J, Ouzunova M, Lübberstedt T (2007) High levels of linkage disequilibrium and associations with forage quality at a phenylalanine ammonia-lyase locus in European maize (*Zea mays* L.) inbreds. Theor Appl Genet 114:307–319

Bai WN, Liao WJ, Zhang DY (2010) Nuclear and chloroplast DNA phylogeography reveal two refuge areas with asymmetrical gene flow in a temperate walnut tree from East Asia. New Phytol 188(3):892–901

Bai WN, Wang WT, Zhang DY (2014) Contrasts between the phylogeographic patterns of chloroplast and nuclear DNA highlight a role for pollen-mediated gene flow in preventing population divergence in an East Asian temperate tree. Mol Phylogenet Evol 81:37–48

Bai WN, Wang WT, Zhang DY (2015) Phylogeographic breaks within Asian butternuts indicate the existence of a phytogeographic divide in East Asia. New Phytol Doi. doi:10.1111/nph.13711

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Bayazit S, Kazan K, Golbitti S, Cevik V, Ayanogla H, Ergul A (2007) AFLP analysis of genetic diversity in low chill requiring walnut (*Juglans regia* L.) genotyping from Hatay Turkey. Sci Hortic 111: 394–398

Beer R, Kaiser F, Schmidt K, Ammann B (2008) Vegetation history of the walnut forests in Kyrgyzstan (Central Asia): natural or anthropogenic origin? Quaternary Sci Rev 27:621–632

Bohonak AJ (2002) IBD (isolation by distance): a program for analyses of isolation by distance. J Hered 93:153–154

Brady KU, Kruckeberg AR, Bradshaw HD Jr (2005) Evolutionary ecology of plant adaptation to serpentine soils. Annu Rev Ecol Evol S 36:243–266

Chen L, Ma Q, Chen Y, Wang B, Pei D (2014) Identification of major walnut cultivars grown in China based on nut phenotypes and SSR markers. Sci Hortic 168:240–248

Cheng F, Peng X, Zhao P, Yuan J, Zhong C, Cheng Y, Cui C, Zhang S (2013) Soil microbial biomass, basal respiration and enzyme activity of main forest types in the Qinling Mountains. PLoS One 8:e67353–e67353

Christopoulos MV, Tsantili E (2015) Participation of phenylalanine ammonia-lyase (*PAL*) in increased phenolic compounds in fresh cold stressed walnut (*Juglans regia* L.) kernels. Postharvest Bio Tech 104:17–25

Coakes SJ, Steed L (2009) SPSS: Analysis without anguish using SPSS version 14.0 for Windows. John Wiley & Sons, Inc.

Dang M, Liu ZX, Chen X, Zhang T, Zhou HJ, Hu YH, Zhao P (2015) Identification, development, and application of 12 polymorphic EST-SSR markers for an endemic Chinese walnut (*Juglans cathayensis* L.) using next-generation sequencing technology. Biochem Syst Ecol 60:74–80

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

Du S, Wang Z, Ingvarsson PK, Wang D, Wang J, Wu Z, Tembrock LR, Zhang J (2015) Multilocus analysis of nucleotide variation and speciation in three closely related *Populus* (Salicaceae) species. Mol Ecol 24:4994–5005

Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

ESRI (2010) ArcGIS version 10.0. Environmental Systems Research Institute, Redlands

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

Excoffier L, Lischer HEL (2010) Arlequin suite v. 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567

Fan ZY, Xi XL (2002) Five new early bearing fruit *Juglans* varieties developed by crossing *Juglans sigillata* Dode and *Juglans regia* Linn. Hunan For Sci Technol 29:66–68 in Chinese

FAO (Food and Agricultural Organization of the United Nations) (2012) FAO statistical databases and data sets. Available from: http://faostat.fao.org

Golge O, Hepsag F, Kabak B (2016) Determination of aflatoxins in walnut sujuk and Turkish delight by HPLC-FLD method. Food Control 59:731–736

Gunn BF, Aradhya M, Salick JM, Miller AJ, Yongping Y, Lin L, Xian H (2010) Genetic variation in walnuts (*Juglans regia* and *J. sigillata*; Juglandaceae): species distinctions, human impacts, and the conservation of agrobiodiversity in Yunnan, China. Am J Bot 97:660–671

Guo Y, Li G, Hu Y, Kang D, Wang D, Yang G (2013) Regeneration of *Betula albosinensis* in strip clearcut and uncut forests of the Qinling Mountains in China. PLoS One 8:e59375

Hayes D, Angove MJ, Tucci J, Dennis C (2015) Walnuts (*Juglans regia*) chemical composition and research in human health. Crit Rev Food Sci. doi:10.1080/10408398.2012.760516

Holland MM, Parson W (2011) GeneMarker® HID: a reliable software tool for the analysis of forensic STR data. J Forensic Sci 1:29–35

Hu YH, Dang M, Zhang T, Luo GC, Xia HL, Zhou HJ, Hu DF, He L, Ma ZH, Zhao P (2014) Genetic diversity and evolutionary relationship of *Juglans regia* wild and domesticated populations in Qinling Mountains based on nrDNA ITS sequences. Scientia Silvae Sinicae 50:47–55 In Chinese

Hu YH, Zhao P, Zhang Q, Wang Y, Gao XX, Zhang T, Zhou HJ, Dang M, Woeste KE (2015a) De novo assembly and characterization of transcriptome using Illumina sequencing and development of twenty five microsatellite markers for an endemic tree *Juglans hopeiensis* Hu in China. Biochem Syst Ecol 63:201–211

Hu Z, Zhang T, Gao XX, Wang Y, Zhang Q, Zhou HJ, Zhao GF, Wang ML, Woeste KE, Zhao P (2015b) De novo assembly and characterization of the leaf, bud, and fruit transcriptome from the vulnerable tree *Juglans mandshurica* for the development of 20 new microsatellite markers using Illumina sequencing. Mol Gen Genomics:1–14

Huang J, Gu M, Lai Z, Fan B, Zhou YH, Yu JQ, Chen Z (2010) Functional analysis of the Arabidopsis *PAL* gene family in plant growth, development, and response to environmental stress. Plant Physiol 153:1526–1538

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. BMC Genet 6:13

Jiang Z, Chen Y, Bao Y (2012) Population genetic structure of *Tamarix chinensis* in the Yellow River Delta, China. Plant Syst Evol 298: 147–153

Joly C, Visset L (2009) Evolution of vegetation landscapes since the Late Mesolithic on the French West Atlantic coast. Rev Palaeobot Palyno 154:124–179

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106

Li GT, Ai CX, Zhang LS, Wei HR, Liu QZ (2011) ISSR analysis of genetic diversity among seedling walnut (*Juglans* spp.) populations. J Plant Genet Resour 12:640–645 in Chinese

Li L, Abbott RJ, Liu B, Sun Y, Li L, Zou J, Wang X, Miehe G, Liu J (2013) Pliocene intraspecific divergence and Plio-Pleistocene range expansions within Picea likiangensis (*Lijiang spruce*), a dominant forest tree of the Qinghai-Tibet Plateau. Mol Ecol 22:5237–5255

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

López-Pujol J, Zhang FM, Sun HQ, Ying TS, Ge S (2011) Centres of plant endemism in China: places for survival or for speciation? J Biogeogr 38:1267–1280

Milne RI (2006) Northern hemisphere plant disjunctions: a window on tertiary land bridges and climate change? Ann Bot 98:465–472

Molnar TJ, Zaurov DE, Capik JM, Eisenman SW, Ford T, Nikolyi LV, Funk CR (2011) Persian walnuts (*Juglans regia* L.) in Central Asia. Annu Rep North Nut Grow Assoc 101:56–69

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590

Ning D, Ma Q, Zhang Y, Wang H, Liu B, Pei D (2011) FISH-AFLP analysis of genetic diversity on walnut cultivars in Yunnan Province. For Res 24:189–193 in Chinese

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

Pollegioni P, Woeste K, Olimpieri I, Marandola D, Cannata F, Malvolti ME (2011) Long-term human impacts on genetic structure of Italian walnut inferred by SSR markers. Tree Genet Genomes 7:707–723

Pollegioni P, Woeste KE, Chiocchini F, Olimpieri I, Tortolano V, Clark J, Hemery GE, Mapelli S, Malvolti ME (2014) Landscape genetics of Persian walnut (*Juglans regia* L.) across its Asian range. Tree Genet Genomes 10:1027–1043

Pollegioni P, Woeste KE, Chiocchini F, Del Lungo S, Olimpieri I, Tortolano V, Clark J, Hemery GE, Mapelli S, Malvolti ME (2015) Ancient humans influenced the current spatial genetic structure of common walnut populations in Asia. PLoS One 10:e0135980

Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. Genetics 144:1237–1245

Pritchard J K, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics. 155:945–959

Raymond M, Rousset F (1995) GENEPOP, version 1.2: population genetics software for exact tests and ecumenicism. J Hered 86:248–249

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138

Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software for windows and Linux. Mol Ecol Resour 8: 103–106

Tian B, Liu R, Wang L, Qiu Q, Chen K, Liu J. (2009) Phylogeographic analyses suggest that a deciduous species (Ostryopsis davidiana Decne., Betulaceae) survived in northern China during the Last Glacial Maximum. J Biogeogr 36:2148–2155

Tonnessen BW, Manosalva P, Lang JM, Baraoidan M, Bordeos A, Mauleon R, Oard J, Hulbert S, Leung H, Leach JE (2015) Rice phenylalanine ammonia-lyase gene *OsPAL4* is associated with broad spectrum disease resistance. Plant Mol Biol 87:273–286

Tsoukas MA, Ko BJ, Witte TR, Dincer F, Hardman WE, Mantzoros CS (2015) Dietary walnut suppression of colorectal cancer in mice: mediation by miRNA patterns and fatty acid incorporation. J Nut Biochem 26:776–783

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

Victory ER, Glaubitz JC, Rhodes OE Jr, Woeste KE (2006) Genetic homogeneity in *Juglans nigra* (Juglandaceae) at nuclear microsatellites. Am J Bot 93:118–126

Vinson JA, Cai Y (2012) Nuts, especially walnuts, have both antioxidant quantity and efficacy and exhibit significant potential health benefits. Food Function 3:134–140

Wang H, Pei D, Gu RS, Wang BQ (2008) Genetic diversity and structure of walnut populations in central and southwestern China revealed by microsatellite markers. J Am Soc Hortic Sci 133:197–203

Wang HX, Zhao SG, Gao Y, Zhang ZH, Xuan LC (2011) Genetic diversity of *Juglans regia* L. Cultivars revealed by AFLP analysis. Sci Agric Sin 44:1434–1442 in Chinese

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

Wang Q, Liu J, Allen GA, Ma Y, Yue W, Marr KL, Abbot RJ (2016) Arctic plant origins and early formation of circumarctic distributions: a case study of the mountain sorrel, *Oxyriadigyna*. New Phytol 209:343–353

Wei Z, Du Q, Zhang J, Li B, Zhang D (2013) Genetic diversity and population structure in Chinese indigenous poplar (*Populus simonii*) populations using microsatellite markers. Plant Mol Biol Rep 31:620–632

Xiang QYJ, Soltis DE (2001) Dispersal-vicariance analyses of intercontinental disjuncts: historical biogeographical implications for angiosperms in the Northern Hemisphere. Int J Plant Sci 162:S29–S49

Xu F, Deng G, Cheng S, Zhang W, Huang X, Li L, Cheng H, Rong X, Li J (2012) Molecular cloning, characterization and expression of the phenylalanine ammonia-lyase Gene from *Juglans regia*. Molecules 17:7810–7823

Yang ZX, Xi SK (1989) A study on isozymes of peroxidase of 10 species in *Juglans* L. Acta Phytotaxon Sinica 27:53–57 in Chinese

Zeng Y, Liao W, Petit RJ, Zhang D (2011) Geographic variation in the structure of oak hybrid zones provides insights into the dynamics of speciation. Mol Ecol 20:4995–5011

Zhang ZX, Niu HY, Guo X, Wang D, Eaton WD (2015) Genetic diversity and genetic structure of *Corydalis tomentella* Franch. (Papaveraceae), an endangered herb species from Central China. Biochem Syst Ecol 63:27–33

Zhao P, Woeste K (2011) DNA markers identify hybrids between butternut (*Juglans cinerea* L.), and Japanese walnut (*Juglans ailantifolia* Carr.). Tree Genet Genomes 7:511–533