Trees (2021) 35:1993–2007
https://doi.org/10.1007/s00468-021-02167-y

ORIGINAL ARTICLE

# Comparative phylogeography of *Juglans regia* and *J. mandshurica* combining organellar and nuclear DNA markers to assess genetic diversity and introgression in regions of sympatry

**Meng Dang[1] · Hui-Juan Zhou[2] · Keith E. Woeste[3] · Ming Yue[1,4] · Yi Zhang[1] · Gui-Fang Zhao[1] · Shuo-Xin Zhang[2] · Peng Zhao[1]**

Received: 23 February 2021 / Accepted: 24 June 2021 / Published online: 2 July 2021


## Abstract

***Key message*** **By comparing the phylogeography of *Juglans regia* and *J. mandshurica*, we found that two walnut species, even when sympatric, rarely introgress, suggesting that strong hybridization barriers exist between these species.**

**Abstract** The biogeographic investigation of temperate walnut (*Juglans*) trees is of great interest because of their ecological and economical importance. Our goal was to perform an in-depth investigation of the genetic and phylogeographic history of *J. regia* and *J. mandshurica*, two walnut species that are sympatric in parts of their ranges, including tests for gene flow and incomplete lineage sorting. We employed a 234 bp locus of mitochondrial DNA, a 1.8 Kbp locus of chloroplast DNA, 3 nuclear loci totaling 1740 bp in length, and 17 EST-SSRs. We sampled 559 individuals, 332 of *J. regia* and 227 of *J. mandshurica*, from 69 locations. We found that *J. regia* and *J. mandshurica*, even when sympatric, rarely introgress, suggesting that strong barriers to hybridization exist between these species. Niche overlap analyses for the two species found that they occupy distinct ecological niches and that the sympatric populations may be the result of recent postglacial population expansion.

**Keywords** Genetic differentiation · Phylogeography · Organellar DNA · Nuclear DNA · Introgression · *Juglans*

## Introduction

Phylogeography is the study of the geographical pattern of organisms from a phylogenetic perspective. Phylogeography correlates historical events such as gene flow, glaciation, and aridification to observed spatial−genetic patterns in a species or group of species (Bai et al. 2010, 2014; Fan et al.

Meng Dang and Hui-Juan Zhou contributed equally to this paper.

Communicated by José I Hormaza .

✉ Peng Zhao
pengzhao@nwu.edu.cn

Meng Dang
15339260798@163.com

Hui-Juan Zhou
ericguanyuzhao@163.com

Keith E. Woeste
woeste@purdue.edu

Ming Yue
yueming@nwu.edu.cn

Yi Zhang
yizhang247274030@163.com

Gui-Fang Zhao
gfzhao@nwu.edu.cn

Shuo-Xin Zhang
sxzhang@nwsuaf.edu.cn

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

[2] College of Forestry, Northwest A&F University, Yangling 712100, Shaanxi, China

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

[4] Xi'an Botanical Garden of Shaanxi Province, Xi'an, Shaanxi, People's Republic of China

2013; Avise 2000). Comparative phylogeography, when combined with a thorough consideration of ecological and life-history traits, can provide a better understanding of the relative contribution of historical, contemporary, biotic, and abiotic factors to the structure of genetic variation (Papadopouloua et al. 2016). Recent studies have improved our understanding of the phylogeography of walnut (*Juglans*) species by analyzing the effects of geography, ecology, and a changing climate on walnut diversity and spatial genetic structure (Bai et al. 2010, 2014, 2015; Fan et al. 2013; Wang et al. 2016; Han et al. 2016). The phylogeography of most temperate plant species in the northern hemisphere is strongly influenced by repeated cycles of climatic suitability during the Quaternary period (Qiu et al. 2011). A species' demographic, spatial, and genetic recovery from cold and dry periods can be strongly influenced by the location(s) and size of glacial refuges and what congeners may have been present there. Refugia for walnut species are not fully resolved, but improved genetic data can reveal the history of Asian walnut species (Chen et al. 2012).

Chloroplast DNA (cpDNA) and mitochondrial DNA (mtDNA) are widely used in phylogeographic analyses to help reveal evolutionary history and lineage divergence (Avise 1998, 2000; Casazza et al. 2016; Du et al. 2009; Gugger et al. 2010). Both cytoplasmic organelles are in most cases maternally inherited in angiosperms, so studies of the spatial genetic structure can provide insight at deeper evolutionary time scales (Avise 1998, 2009; Liu et al. 2012; Zhao et al. 2018; Sun et al. 2019). Data from nuclear DNA fragments [for example, barcodes such as the internal transcribed spacer (ITS)] are also commonly used for studies of population genetics and speciation. Nuclear DNA sequences can be used to reconstruct phylogenetic relationships among related lineages (Dang et al. 2019), to identify hybrid individuals (Zhao and Woeste 2011; Barton et al. 1993), and in some cases to reveal patterns of sympatric speciation (Sunnucks 2000; Van et al. 2008; Dang et al. 2019; Yuan et al. 2018). Nuclear DNA makers, such as SSRs, which are inherited biparentally, are also widely used in population genetic studies (Bai et al. 2015; Han et al. 2016; Yuan et al. 2018). The combination of cytoplasmic (mitochondrial and chloroplast) DNA sequence data with nuclear DNA sequence data provides a robust dataset for testing hypotheses related to phylogeography, population genetics, patterns of genetic variation, and interspecific gene flow.

All *Juglans* species are monoecious, wind-pollinated, and deciduous trees (Bai et al. 2014, 2015; Zhao et al. 2018). Common walnut, or *J. regia*, and *J. mandshurica* are both temperate species that bear nutritious nuts and produce high-quality timber, which makes them economically important (Bai et al. 2015; Han et al. 2016). *Juglans regia* has been cultivated throughout temperate parts of China for millennia, but its natural range there is uncertain (Pollegioni et al. 2014, 2015). *Juglans mandshurica* is widely distributed in China, ranging from the tropical South (< 22°N) and the Qinling Mountains to the Huai River line (~ 34°N) in hilly mid-elevation areas (Bai et al. 2014; Dang et al. 2015). *J. regia* and *J. mandshurica* differ in shell thickness, nut size, color of stigma, number of flowers per cluster, and number of leaflets per leaf (Lu et al. 1999a, b; Manning 1978). In China, local biogeographic history and climate have isolated and led to the diversification of the *J. regia* and *J. mandshurica* populations (Bai et al. 2014, 2015; Han et al. 2016; Pollegioni et al. 2015). *J. mandshurica* belongs to the section *Cardiocaryon* along with *J. hopeinesis* (Ma walnut), and *J. ailantifolia* (Japanese walnut) and, in some taxonomic delimitations, *J. cinerea* (American butternut) (Zhao et al. 2018; Dang et al. 2019). The first three species are native to East Asia, but *J. cinerea* (butternut) is native to eastern North America (Stanford et al. 2000; Zhao et al. 2018; Aradhya et al. 2007). The taxon *J. cathayensis* of Sect. *Cardiocaryon* is now considered invalid and all populations previously identified as *J. cathayensis* are now *J. mandshurica* (Zhao et al. 2018; Dang et al. 2015, 2019). *J. regia* (Common walnut) belongs to the section *Juglans*, along with *J. sigillata* (Iron walnut), a taxon native to southern China (Aradhya et al. 2007; Zhao et al. 2018; Feng et al. 2018a; Sun et al. 2019). Many species of *Juglans* can hybridize (Hoban et al. 2009; Zhao and Woeste 2011; Shu et al. 2016; Yuan et al. 2018; Mu et al. 2017). In North America, *J. nigra* and *J. cinerea* are sympatric and reproductively isolated; however, both species can hybridize with Asian *Juglans* and appear in cultivated populations but not in the wild (Woeste and Michler 2011). Differences in geographic range, climate, habitat, and the presence of natural geographical features have (with certain exceptions) been inferred to be important in maintaining genetic and spatial separation. To avoid the interference of cultivated trees on gene flow or introgression, we specifically sampled locations that might be expected to contain spontaneous wild hybrids.

We used mtDNA, cpDNA, nuclear DNA sequences, and EST-SSRs to investigate the biogeography of *J. regia* and *J. mandshurica*. Our goal was to understand *J. regia* and *J. mandshurica* genetic diversity and introgression in regions of sympatry. We analyzed the composition of mtDNA, cpDNA, nuclear DNA sequences, and microsatellite length variation within 69 populations of *J. regia* and *J. mandshurica*, 40 individuals from four other *Juglans* species, and 13 individuals from other genera in the Juglandaceae (*Carya cathayensis* and *Engelhardtia roxburghiana*). Because *J. regia* and *J. mandshurica* cooccur in several sympatric populations throughout China, we hypothesized that we would observe gene flow between these two walnut species in regions of sympatry. The objectives were to: (1) characterize mitochondrial and chloroplast haplotypes in Chinese *Juglans* and their distribution in two Chinese *Juglans*

species; (2) determine if spatial genetic structure can shed light on how evolutionary history shaped the current geographical distribution of these related, partially sympatric species; (3) determine if nuclear and cytoplasmic DNA indicate similar or discordant evolutionary histories for *J. regia* and *J. mandshurica*; and (4) determine if we could identify evidence of gene flow between *J. regia* and *J. mandshurica* and, if so, what the consequences might have been (parentage of hybrid species).

## Materials and methods

### Population sampling

We intensively sampled two closely related walnut species *Juglans regia* (43 populations 332 individuals) and *J. mandshurica* (26 populations; 227 individuals). We also sampled 40 individuals of four other *Juglans* species, including *J. hopeiensis*, *J. sigillata*, *J. cinerea*, and *J. ailantifolia* (Table S1). In addition to these six walnut (*Juglans*) species, we obtained sequence data from *Carya cathayensis* and *Engelhardtia roxburghiana* for use as outgroups (Tables S1). *J. cathayensis* is currently recognized as a synonym of *J. mandshurica* Lu et al. 1999a, b; http://www.theplantlist.org/tpl1.1/record/tro-16700183). In this study, our samples of *J. mandshurica* included samples from populations previously described as *J. cathayensis*. Altogether, a total of 599 leaf samples from six *Juglans* species, and samples from species of two additional genera were collected from 77 locations in mainland China, Taiwan, Japan, Canada, and the United States from July 2011 to July 2014 (Table S1). All sampled trees were mature adults, apparently healthy, growing in mountain forests, along forest roads, or near villages, but not in orchards or on farmed lands. Sampled trees were separated by at least 50 m. Fresh leaf samples were stored at 25 °C in silica gel until DNA extraction, as previously described (Zhao and Woeste 2011; Feng et al. 2018b). Using the geographic information system (Garmin), the latitude and longitude of collected locations were measured. We compiled spatial summary data of all sequence haplotypes using ArcMap.

### Genomic DNA extraction, PCR reaction, sanger sequencing, and EST–SSR genotyping

The genomic DNA was extracted from dry leaf samples using the protocols of Zhao and Woeste (2011) and Doyle and Doyle (1987) and stored at −20 °C. A total of 9 mitochondrial DNA markers, 18 chloroplast DNA markers, and 9 nuclear DNA markers were amplified from all samples (Table S2). After PCR amplification, we determined that only one mitochondrial locus (3–9), two chloroplast DNA markers (*trnL-F* and *trnS-G*), and three nuclear DNA loci (ITS, *15R-8*, and *Jr5680*) were polymorphic among individuals of *J. regi*a and *J. mandshurica*, so these were used for further analysis (Table S2). The internal transcribed spacer fragment ITS1 (primer Forward sequence: TCCGTA GGTGAACCTGCGG, Reverse sequence: TCCTCCGCT TATTGATATGC) was used in this study (Table S2). In the nuclear genome, the sequence variability at ITS (internal transcribed spacer) is commonly used for phylogenetic and population genetics studies (Wright 1978). The polymorphic nuclear locus *Jr5680* was developed from *J. regia* trancriptomic data in our lab (Table S2). The nuclear gene *Jr5680* has almost the same sequence (about 96 %) as the *JrPAL* (phenylalanine ammonia lyase) in GenBank (accession no. JX069977), with a total length of 2369 bp. The *PAL* gene is found widely in plants, as well as some yeast and fungi, with isoenzymes existing within many different species. The *PAL* gene was amplified by *Jr5680* primer (Wang et al. 2016; Table S2). A total of 17 EST-SSR primers were used to evaluate the phylogeography of the two walnut species (Table S2). The total volume of each reaction was 25 μL (2 μL DNA(50 ng/uL); 12.5 μL 2× PCR Taq Mix (0.05 u/L Taq DNA Polymerase: 3 mmol/L 4 mM $MgCl_2$; 500 umol/L 0.4 mM each dNTP; 100 mmol/L 500 mM KCl; Guangzhou Geneshun Biotech, Ltd., Guangzhou, China); 2.5 μL BSA (Bovine Serum Albumin, acetylated; Madison, WI, USA); 0.8 μL each primer (3–9, *trnL-F*, *trnS-G*, ITS, *15R-8*, and *Jr5680*) (Zhao and Woeste 2011; Wright 1978; Table S2); 6.4 μL sterilized distilled water). The PCR amplification procedure was as follows: (1) initial denaturation at 94 °C for 3 min, (2) denaturation for 15 s at 93 °C, (3) annealing temperature for 1 min (3–9, 50 °C; *trnL-F*, 55 °C; *trnS-G*, 55 °C; ITS, 53 °C; *15R-8*: 63 °C; *Jr5680*, 63 °C; EST-SSRs, 55 °C), (4) 30 s at 72 °C (32 cycles, from step 2 to step 4), and (5) final extension at 72 °C for 10 min (Zhao and Woeste 2011). The PCR products were sent to Sangon Biotech Co., Ltd. (Shanghai, China) for sequencing.

### Data analysis

The mitochondria, chloroplasts, and nuclear DNA sequences were trimmed and aligned using Bioedit version 7.0.8 (Hall 1999). Based on electropherograms of sequences, the presence of mutations was verified and sequences with erroneous base calls were manually corrected using MEGA 6.0 (Tamura et al. 2013). The electropherogram of the nuclear gene sequences (ITS, *15R-8*, and *Jr5680*) inherited from both parents occasionally showed a nested peak at some sites. If the weak signal in the nested peak signal reached one-quarter of the peak height of the strong signal, the site was considered to have base superposition (Fuertes 1999, 2003). The heterozygous sites in the sequence were replaced according to the annexation base coding information of

IUPAC. Dnasp version 5.0 (Librado and Rozas 2009) was used to split the nuclear gene sequence into two parent sequences for subsequent analysis (Bai et al. 2015) and to obtain haplotypes and calculate related genetic index ($\pi$, $h_T$). Each haplotype was illustrated using different colors in pie charts; we added a black border to the pie chart of *J. mandshurica*. Using NETWORK version 4.2.0.1, we constructed median-joining networks to determine the phylogenetic relationship of haplotypes within and among species (Bandelt et al. 1999). Arlequin version 3.11 was used to examine genetic distance, the neutrality of five fragments, and the AMOVA (analysis of molecular variance) of two species based on mtDNA, cpDNA, and nuclear DNA sequences (Excoffier and Lischer 2010). Departures from mutation-drift equilibrium were reflected in the values of Fu and Li's D (Fu and Li 1993). PERMUT 1.0 software (Pons and Petit 1996) was used to estimate the genetic differentiation coefficients of $N_{ST}$ and $G_{ST}$. The phylogeographic structure was estimated by comparisons between $G_{ST}$ and $N_{ST}$ and tested with 1000 random permutations of mitotype identity, as a significantly higher $N_{ST}$ compared to $G_{ST}$ supports the presence of structure. $N_{ST}$ and $G_{ST}$ differ only in that $N_{ST}$ considers the similarities between mitotypes (Pons and Petit 1996). AMOVA software was used to calculate the population genetic differentiation ($F_{ST}$) (Schneider et al. 2000). To detect whether the *J. regia* and *J. mandshurica* populations had experienced expansion events in their evolutionary history, Dnasp version 5.0 was used to show mismatch distributions (Librado and Rozas 2009).

The software GENEALEX ver. 6.5.0.1 was used to calculate the genetic index (*Na*, *Ne*, *Ho*, $H_E$, *uHe*, $F_{IS}$; Table 1, S3 and S4) and genetic distance between populations based on microsatellite data (Peakall and Smouse 2012). HP-RARE 1.0 software was used to calculate Allelic richness (*Rs*) and private allele richness ($P_{AR}$) values (Kalinowski 2005; Table S4). POPTREE2 (Takezaki et al. 2010) software was used to construct the unweighted pair group method with arithmetic means (UPGMA) tree, which was visualized with the software MEGA 6.0 (Schneider et al. 2000). We used the "ggplot" package (Wickham 2009) in R studio to draw the *PCA* (Principal Component Analysis) results based on microsatellite data, in which the genetic distance was calculated in GENEALEX ver. 6.5.0.1 software (Pons and Petit 1996). We used the "rmaverick" R package (developed by Robert Verity, 2019; software available from https://bobverity.github.io/rmaverick/) to calculate the genetic structure of 16 sympatric walnut (*Juglans*) populations based on a total of 17 EST-SSRs. The settings were as follows: length of burn-in period, 100,000; number of MCMC reps after burn-in, 500,000; $K = 1-9$; number of iterations, 15 (Evanno et al. 2005). We used the delta K method to estimate the optimal value of *K* through the program STRUCTURE HARVESTER (Earl and vonHoldt 2012) and function plot_posterior_K() in the "rmaverick" R package. Then we

**Table 1** Genetic diversity of *Juglans* using mtDNA sequences, cpDNA sequences, nuclear sequences and microsatellite markers

| Type | Group | $h_T$ | $\pi$ | Tajima's D | Fu & Li's D | No | $G_{ST}$ | $N_{ST}$ | $F_{ST}$ | Mitotypes/ Haplotypes |
|---|---|---|---|---|---|---|---|---|---|---|
| mtDNA | *J. regia* | 0.024 | 0.0001 | − 0.901 | 0.464 | 166 | − 0.006 | 0.006 | 0.053 | 5 |
| | *J. mandshurica* | 0.032 | 0.0002 | − 1.484 | − 2.219 | 189 | − 0.009 | − 0.016 | − 0.01 | 5 |
| cpDNA | *J. regia* | 0.077 | 0.0001 | − 1.782* | − 3.196 | 252 | – | – | 0.24 | 9 |
| | *J. mandshurica* | 0.512 | 0.0008 | − 1.090 | 1.436 | 193 | 0.670 | 0.349 | 0.30 | 13 |
| ITS | *J. regia* | 0.113 | 0.0002 | − 1.096 | 0.729 | 390 | 0.287 | 0.299 | 0.24 | 4 |
| | *J. mandshurica* | 0.052 | 0.0001 | − 1.346 | − 0.811 | 302 | 0.275 | 0.278 | 0.20 | 6 |
| *15R-8* | *J. regia* | 0.832 | 0.0089 | − 0.653 | 1.943** | 340 | 0.235 | 0.328 | 0.26 | 42 |
| | *J. mandshurica* | 0.661 | 0.0060 | − 1.758* | − 0.195 | 334 | 0.154 | 0.117 | 0.12 | 29 |
| *Jr5680* | *J. regia* | 0.370 | 0.0053 | 0.271 | − 0.067 | 476 | 0.301 | 0.479 | 0.43 | 25 |
| | *J. mandshurica* | 0.041 | 0.0006 | − 1.429 | − 2.298 | 242 | 0.197 | 0.206 | 0.16 | 7 |
| Type | Group | *Na* | *Ne* | *Ho* | $H_E$ | *uHe* | $F_{IS}$ | $R_S$ | $P_{AR}$ | |
| EST-SSR | *J. regia* | 2.581 | 1.755 | 0.291 | 0.346 | 0.360 | 0.161 | 2.36 | 0.061 | |
| | *J. mandshurica* | 4.191 | 2.597 | 0.462 | 0.455 | 0.470 | 0.003 | 7.36 | 0.139 | |

Total gene diversity ($h_T$), nucleotide diversity ($\pi$), among populations genetic differentiation ($G_{ST}$, $N_{ST}$), among populations differentiation ($F_{ST}$). number of different alleles (*Na*), number of effective alleles (*Ne*), observed heterozygosity (*Ho*), expected heterozygosity ($H_E$), unbiased expected heterozygosity (*uHe*), inbreeding coefficient ($F_{IS}$), allelic richness (*Rs*) and private allele richness ($P_{AR}$)

*Ns* Not significant

*$0.01 < P < 0.05$

**$0.001 < P < 0.01$

***$P < 0.001$

used function plot_qmatrix() to output the genetic structure results of "rmaverick" R package. To further estimate the historical gene flow between *J. regia* and *J. mandshurica*, we independently ran MIGRATE v. 3.6.1.1 (Beerli 2005) five times to estimate the mode and 95 % highest posterior density after checking for data convergence to ensure the validity of the results.

The geographical distribution characteristics of the total number of haplotypes and allelic richness (*Rs*) of *J. regia* and *J. mandshurica* were more directly reflected using the interpolation function in ArcGIS ver. 10.3 software (ESRI, Redlands, CA, USA) (Feng et al. 2018b). The geospatial interpolation of these indices enabled us to better compare the new spatial data representing the genetic diversity of *J. regia* vs. *J. mandshurica* populations. Based on our previous niche reconstruction results of *Juglans* (Zhao et al. 2018), we selected five climatic factors for niche overlap analysis of *J. regia* and *J. mandshurica*: bio4: Temperature Seasonality, bio9: Mean Temperature of the Driest Quarter, bio11: Mean Temperature of the Coldest Quarter, bio15: Precipitation Seasonality, and bio18: Precipitation of the Warmest Quarter. Based on environmental space (*E* spaces), the niche overlap statistics of two walnut species were analyzed using R packages (Broennimann et al. 2011; Petitpierre et al. 2012). Principal component analysis (*PCA*) analysis was used to transform occurrence and climate data into environmental axes (*PCA*-env) (Broennimann et al. 2011; Petitpierre et al. 2012; Dreyer et al. 2019). Niche overlap was determined according to D and the I index, which ranges from 0 (no overlap) to 1 (complete overlap). The contribution of the climatic variables on the two axes of the *PCA* and the percentage of inertia were explained by the two axes. All analyses were carried out using the R v3.6.1 (Dreyer et al. 2019).

## Results

### Diversity of mtDNA, cpDNA, and nuclear DNA loci

A total of 11 mitotypes (GenBank accession numbers: MZ150569-MZ150579) were identified among 612 individuals (in 8 species, 6 of which were in *Juglans*) collected from 77 locations (Table S1, S5 and Fig. 1a). The mitotype (gene) diversity $h_T$ was 0.024 and 0.032 for the *J. regia* samples and the *J. mandshurica* samples, respectively. The $h_T$ across all samples was 0.502. Nucleotide diversity for the same groups was $\pi = 0.00010$, 0.0002, and 0.0022 (Table 1).

A total of 30 chloroplast haplotypes (GenBank accession numbers: MZ291846-MZ291905) were identified among eight species (six *Juglans*) based on sequence variation of the *trnL-F* and *trnS-G* segment (aligned, trimmed length of 1806 bp) found in 612 individual tree samples collected from 77 locations (Table S1 and Fig. 1b). The total number of polymorphic sites was 191 (Table S6). Overall, the chloroplast haplotype (gene) diversity was similarly high for the *J. regia* samples, the *J. mandshurica* samples, and all other samples, at 0.077, 0.512, and 0.664, respectively. Nucleotide diversity for the same groups was $\pi = 0.0001$, 0.0008, and 0.0024 (Table ).

A total of 142 nuclear genotypes (22 ITS genotypes, 81 *15R-8* genotypes, and 39 *Jr5680* genotypes) from the 8 species (612 individuals) analyzed were obtained in this study (Fig. 2 and S1, S2; Tables S7, S8, and S9). The GenBank accession numbers of 22 ITS genotypes, 81 *15R-8* genotypes, and 39 *Jr5680* genotypes were MZ291824-MZ291845, MZ291704-MZ291784, and MZ291785-MZ291823, respectively. These were based on sequence variation of three segments (aligned, trimmed total length of 1740 bp). Sequences from some individuals could not be resolved because they did not produce amplification products; 403 samples were sequenced at ITS to obtain 639 bp sequence, 387 samples at *15R-8* to obtain 363 bp sequence, and 426 samples at *Jr5680* to obtain 639 bp sequence. Nucleotide diversity was generally higher in *J. regia* ($\pi = 0.0002$, 0.0089, and 0.0053) than in *J. mandshurica* ($\pi = 0.0001$, 0.0060, and 0.0006) for ITS, *15R-8*, and *Jr5680*, respectively (Table 1). Overall, the haplotype (gene) diversity in *J. regia* for ITS, *15R-8*, and *Jr5680* was high (0.113, 0.823, and 0.370, respectively), and larger than what we observed at the same loci for *J. mandshurica* (0.052, 0.661, and 0.041, respectively) (Table ).

Based on EST-SSR data, the values of *Na*, *Ne*, *Ho*, $H_E$, *uHe*, $F_{IS}$, *Rs*, and $P_{AR}$ of each collected population ranged from 2.059 to 5.706, 1.555 to 3.134, 0.203 to 0.527, 0.288 to 0.507, 0.301 to 0.532, −0.094 to 0.398, 2.03 to 10.86, and 0.000 to 0.270, respectively (Table S3). The *J. mandshurica* populations had a higher allele abundance and higher genetic diversity than *J. regia* (Table 1). The $F_{IS}$ value of *J. mandshurica* was lower than that of *J. regia*, which indicates lower levels of inbreeding for *J. mandshurica* (Table 1).

### Spatial analysis of the mitochondrion, chloroplast, and nuclear DNA sequences

Mitotype mapping did not show a strong pattern of geographic structure (Fig. 1a). Mitotypes mtH1 to mtH5 were found in *J. regia* in all parts of China, three haplotypes mtH9, mtH10, and mtH11 were found only in *J. mandshurica*, mtH6 was found only in *J. ailantifolia*, while mtH7 was found only in *J. cinerea* (Fig. 1a and Table S5). Mitotype mtH1 was widespread across China and even found in the North American species *J. cinerea*. Mitotype mtH1 was shared among *J. regia*, *J. sigillata*, *J. cinerea, C. cathayensis*, and *E. roxburghiana.* Mitotype mtH2, which differs from mtH1 by a single nucleotide (mutation G−T), was found in four *Juglans* species (*J. hopeiensis, J. mandshurica, J.*

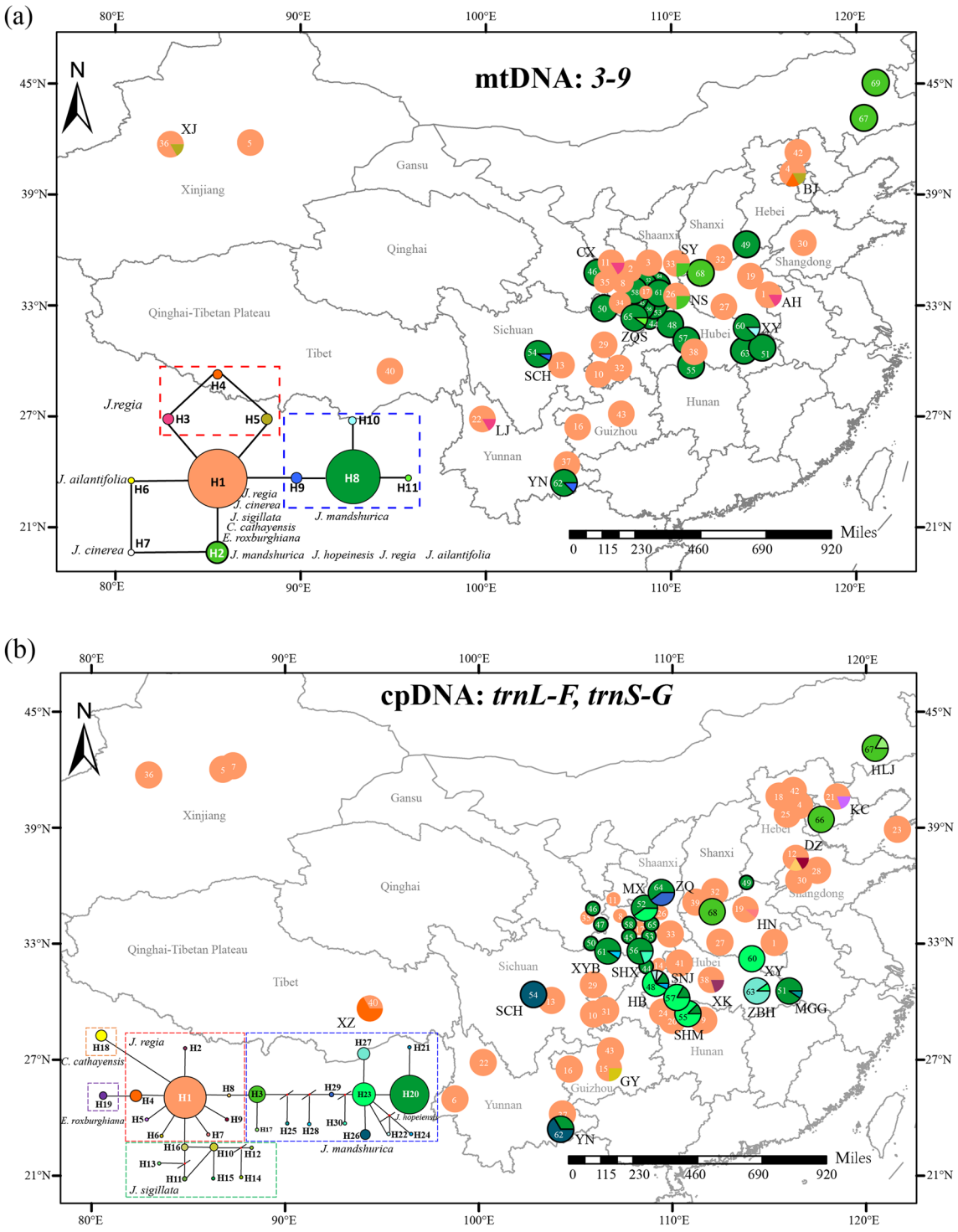


**Fig. 1** Map of the occurrence of mitochondrial and chloroplast haplotypes based on **a** mitochondrial locus *3–9* and **b** chloroplast loci. Abbreviations for samples, numbers of populations, and sampling details are in Table S1. Pie charts bordered in black indicate samples of *J. mandshurica*. Locations with more than one haplotype are indicated as pie charts that show the proportion of each haplotype present at the sampled location. The network of genealogical relationships for 11 mitochondrial haplotypes and 30 chloroplast haplotypes found in all samples in this study are indicated in the lower left of each panel. Black lines show the steps between the haplotypes based on parsimony criteria. The size of the circles in the network of genealogical relationships is roughly proportional to the number of alleles of that type sampled. Some circles on the map are reduced in size for ease of viewing

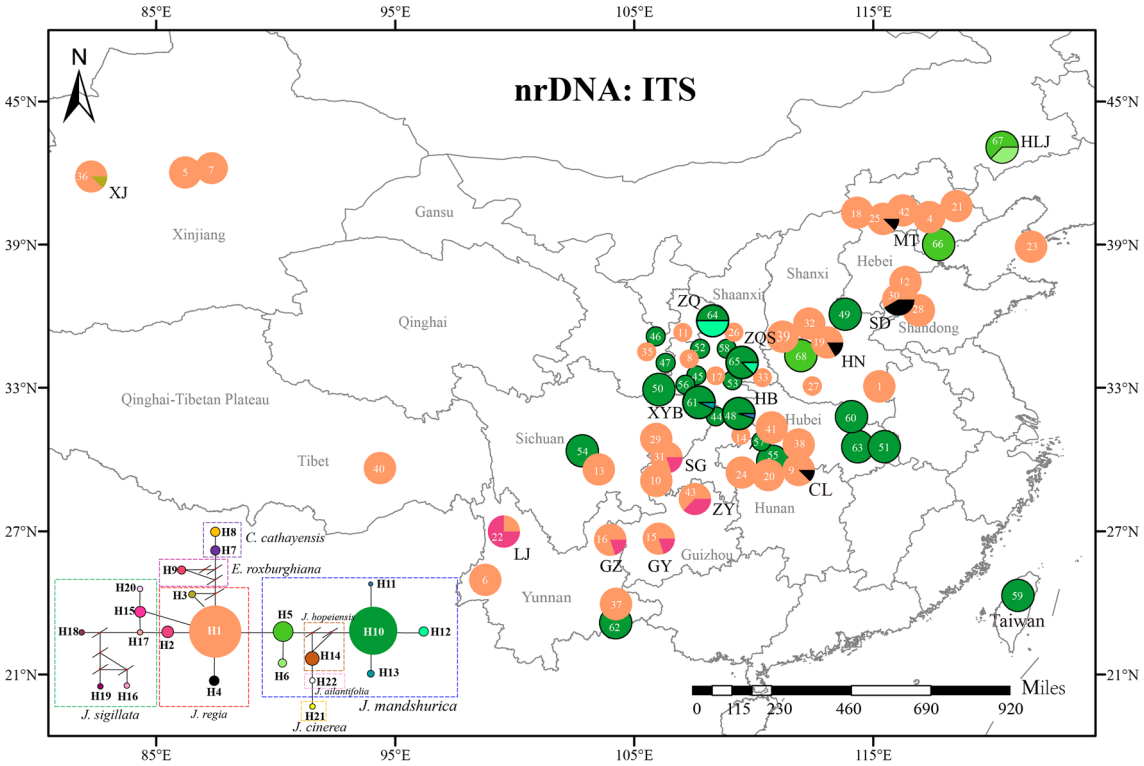


**Fig. 2** Geographic distribution of sequence variation based on the (nuclear) ITS region. Locations where more than one haplotype was encountered are represented as pie charts that show the proportion of each haplotype. Pie charts with a Amber circles indicate *J. regia* sample sites, green circles indicate *J. mandshurica*. We added a black border indicate to the pie chart of Chinese walnut. The network of genealogical relationships for the 22 alleles found among all samples in this study. The black lines show the inferred connections between the alleles based on parsimony optimization. Size of circles is roughly proportional to the number of alleles of that type sampled of that type encountered; the smallest circles represent one tree each. Some circles on the map were reduced in size for ease of viewing. Abbreviations for samples, numbers of populations, and sampling details are in Table S1

*regia*, and *J. ailantifolia*) in 11 populations (Fig. 1). Mitotype mtH8 was found only in *J. mandshurica*, while four populations of *J. mandshurica* contained mitotypes mtH9, mtH10, and mtH11 (SCH, XY, YN, and ZQS) and three populations of *J. mandshurica* only contained mitotypes mtH2 (HLJ, SM, and YC) (Fig. 1 and Table S5). *J. sigillata* had one mitotype (mtH1) among the samples in our study, which was shared with *J. regia*.

Parsimony analysis of the 11 mitotypes produced a network centered on mitotypes mtH1 and mtH8, with single-step branches to most of the other mitotypes. The mtH1 mitotype was widely distributed in *J. regia*, while mtH8 was widely distributed in *J. mandshurica* (Fig. 1a). Although mtH1 was shared by four species besides *J. regia*, it was not shared by *J. mandshurica*, and mtH8, the most common haplotype for *J. mandshurica*, was not shared by any other species. In *J. regia* populations, Gansu (CX), Anhui (AH), and Yunnan (LJ) in China contained the mitotype mtH3; Shaanxi (NS and SY) contained mtH2; Xinjiang (XJ) and Beijing (BJ) contained mitotype mtH5 and mtH4 (Fig. 1a).

Chlorotypes cpH1, cpH2, cpH4, cpH5, cpH6, cpH7, cpH8, and cpH9 were found in *J. regia* in China (Fig. 1b and Table S6). Chlorotype cpH20 was widespread and shared by *J. mandshurica* and *J. hopeiensis*, although chlorotypes cpH3, cpH23, cpH26, and cpH27 were common in certain *J. mandshurica* populations. Chlorotypes cpH3, cpH13, cpH21, cpH22, cpH23, cpH24, cpH25, cpH26, cpH27, cpH28, cpH29, and cpH30 were found in *J. mandshurica* only (Fig. 1b), but most of these chlorotypes were uncommon. The common haplotype of *J. regia* cpH1 was found in the BXS population of *J. mandshurica*, but it was only in one individual. *J. sigillata* had seven chlorotypes, cpH10, cpH11, cpH12, cpH13, cpH14, cpH15, and cpH16; all of these were found in *J. sigillata* only (Table S6).

Although the nuclear gene sequence is diploid, we used the software DNASP to split the nuclear gene sequence into two sequences, paternal and maternal. Then, haplotypes were analyzed using two sequences from each individual. Haplotypes of the three nuclear loci we analyzed showed a clear pattern of geographic structure corresponding to species delimitations (Fig. 2 and S1, S2). ITS haplotype $H1_{ITS}$,

*15R-8* haplotype $H15_{15R-8}$, and haplotype $H1_{Jr5680}$ at locus *Jr5680* were the most common in *J. regia*, while $H10_{ITS}$, $H38_{15R-8}$, and $H26_{Jr5680}$ were the most common in *J. mandshurica* (Fig. 2 and S1, S2). The haplotype network for ITS (Fig. 2) follows the taxonomic split between the two sections of *Dioscaryon* (section *Dioscaryon*/*Juglans*: *J. regia* and *J. sigillata* and section *Cardiocaryon*: *J. mandshurica*, *J. hopeinesis*, *J. ailantifolia*, and *J. cinerea*). Within species, there was also evidence of regional variability. Haplotype $H2_{ITS}$ in *J. regia* was distributed in Yunnan, Guizhou, and Sichuan Provinces (southwestern China), $H4_{ITS}$ was only found in *J. regia* populations in northern China, and $H3_{ITS}$ was only found in Xinjiang Province, northwestern China (Fig. 2). *J. sigillata*, a species from southwestern China, presented six haplotypes at ITS but none were shared with *J. regia*. We only observed variability at ITS for *J. mandshurica* in samples from the Qinling−Bashan Mountains (populations ZQ, ZQS, XYB, and HB) and Heilongjiang (HLJ). Except for the populations BXS and SM, which only had the $H5_{ITS}$ haplotype, all other samples had haplotype $H10_{ITS}$. Butternut (*J. cinerea*), a species closely related to *J. ailantifolia*, contained $H21_{ITS}$, a private haplotype one mutation removed from *J. ailantifolia* (Fig. 2). The outgroup species (*C. cathayensis* and *E. roxburghiana*) contained ITS sequences different to any *Juglans* species.

We observed particularly high levels of variation at the nuclear locus *15R-8*, as 81 haplotypes were found in the *Juglans* species we analyzed. *J. regia* and *J. mandshurica* (Fig. S1) were especially polymorphic at this locus. For *J. regia*, the haplotype $H6_{15R-8}$ and similar haplotypes (closely related haplotypes) were mainly distributed in Xinjiang Province, northwestern China; haplotype $H23_{15R-8}$ and similar haplotypes were mainly distributed in Yunnan; haplotype $H1_{15R-8}$ and related haplotypes were mainly distributed in Guizhou Sichuan, southwestern China; and the haplotype $H19_{15R-8}$ and similar haplotypes were mainly distributed in the area around the Qinling−Bashan Mountains (Fig. S1). Haplotypes $H1_{15R-8}$ to $H34_{15R-8}$ were only observed in *J. regia*. *J. mandshurica* also had high haplotype diversity at *15R-8*. Haplotype $H38_{15R-8}$ was most common and most widely distributed, although haplotypes $H35_{15R-8}$, $H40_{15R-8}$, and $H44_{15R-8}$ were also widely distributed. There were, however, many rare and private haplotypes; $H36_{15R-8}$, $H37_{15R-8}$, $H54_{15R-8}$, $H56_{15R-8}$, $H57_{15R-8}$, $H59_{15R-8}$, $H61_{15R-8}$, $H62_{15R-8}$, and $H64_{15R-8}$−$H68_{15R-8}$ were found in only a few *J. mandshurica* populations (Fig. S1 and Table S8). In general, *J. mandshurica* showed a less obvious genetic structure at this locus than *J. regia*.

Across all collection sites for *Jr5680* (*PAL* gene), $H1_{Jr5680}$ and $H26_{Jr5680}$ were the dominant haplotypes for *J. regia* and *J. mandshurica*, respectively (Fig. S2 and Table S9). The latter two species branched from the network at *J. mandshurica*. At locus *Jr5680* there appeared to be a geographic structure; haplotypes $H8_{Jr5680}$ and $H17_{Jr5680}$ were most common in the samples from Yunnan, Guizhou, and Sichuan (southwestern China). Haplotypes $H15_{Jr5680}$, $H5_{Jr5680}$, $H13_{Jr5680}$, and $H14_{Jr5680}$ were restricted to the populations DZ (northern China), BM (western China), CL (central China), and DZ (northern China), respectively, and were predicted to be part of a complex network in *J. regia* (Fig. S2). Compared to *J. regia* (25 haplotypes), there was much lower haplotype diversity in *J. mandshurica* populations at the *JR5680* locus; we observed seven haplotypes ($H26_{Jr5680}$ to $H32_{Jr5680}$), of which $H26_{Jr5680}$ was by far the most common (Fig. S2). The *PAL* gene diversity of *J. regia* and *J. mandshurica* populations in the Qinling−Bashan Mountains and Yunnan−Kweichow Plateau was higher than that of other regions. The maximum likelihood (ML) phylogenetic trees of *Juglans* were constructed using haplotypes of three nuclear DNA sequences (ITS, *15R-8*, and *Jr5680*; Fig. S3). The results supported that *J. regia* and *J. sigillata* (sect. *Dioscaryon*) were clustered and two species of sect. *Cardiocaryon* (*J. hopeiensis* and *J. mandshurica*) were clustered into one branch (Fig. S3).

## Neutrality tests and mismatch distribution analysis

The Neutrality test and mismatch distribution analysis showed that *J. regia* and *J. mandshurica* had different demographic patterns based on mitochondrial, chloroplast, and nuclear data (Fig. S4). The distribution of mismatches for the mitochondrial sequences contained a single peak, indicating a recent demographic expansion for both *J. regia* and *J. mandshurica* (Fig. S4a, b). The mismatch distribution of *J. regia* chloroplast sequences also revealed a single peak, indicating a recent demographic expansion, but the analysis of *J. mandshurica* chloroplast data revealed two peaks, which indicated no recent expansion or contraction (Fig. S4c, d). All cytoplasmic and nuclear markers showed negative values for Tajima's D in *J. mandshurica*, and nearly all for *J. regia*; however, the probability of significance of Tajima's D values based on cpDNA sequences in *J. regia* was $P < 0.05$, whereas it was $P < 0.05$ in *J. mandshurica* based on *15R-8* sequences, and the probability of significance of other Tajima's D values were not significant ($P > 0.1$, Table 1. A recent population expansion of *J. mandshurica* is strongly indicated by these results (Table 1). Fu and Li's D (1.943) was positive and significant (at $P < 0.01$) for *J. regia* based on *15R-8* sequences. For nuclear DNA, the test of mismatch distribution for the *15R-8* sequences and *Jr5680* sequences showed no recent contraction or expansion (Fig. S4g, i). For ITS, however, the mismatch distribution for ITS and *Jr5680* sequences showed a single peak (for *J. mandshurica*), further confirmation that the *J. mandshurica* population size may have experienced a recent expansion (Fig. S4f, j) that

was not indicated by analysis of the *15R-8* sequences (Fig. S4h).

### AMOVA (Analysis of molecular variance)

AMOVA showed that genetic variation was mainly within populations of *J. regia* (94.68 % using mtDNA, 75.85 %, using cpDNA, 76.5 % using ITS, 74.1 % using *15R-8*, 56.5 % using *Jr5680*, and 83.78 % using EST-SSRs, $P < 0.001$) (Table 2). For *J. mandshurica*, the structure of genetic variation was similar to *J. regia* and mainly within populations (99.74 % using mtDNA, 70.03 % using cpDNA, 79.9 % using ITS, 88.0 % using *15R-8*, 84.3 % using *Jr5680*, and 87.68 % using EST-SSRs, $P < 0.001$) (Table 2. The AMOVA among populations of same species, especially for the *Jr5680* in *J. regia*, is quite high (Table 2). As expected, there was a clear genetic differentiation between members of different sections of *Juglans* (Sect. *Dioscaryon* and Sect. *Cardiocaryon*). An AMOVA analysis showed 78.85 % of the variance for ITS, 77.42 % for *15R-8*, and 64.14 % for *Jr5680* between sections.

### Inverse distance weighted and niche comparison analyses on the environmental space

We conducted a genetic diversity inverse distance weighted (IDW) analysis based on mitochondrial DNA (mtDNA), chloroplast DNA (cpDNA), three nuclear DNA fragment (ITS, *15R-8*, and *Jr5680*), and 17 EST-SSRs, respectively (Fig. 3), to reveal the genetic diversity of mitochondrial mitotypes and chloroplast haplotypes and their distribution in two Chinese *Juglans* species. For *J. regia*, the results of mtDNA + cpDNA sequences showed that the greatest genetic diversity of *J. regia* was in samples from Xinjiang (Northwest China), Tibet, Qinling Mountains, and eastern China (Fig. 3a); the haplotype diversity based on three nuclear loci was the greatest in Xinjiang (northwest China) and southwestern China (Fig. 3c). Based on the microsatellite data, the populations of Qinling Mountains showed high allelic richness (*Rs*) (Fig. 3e). For *J. mandshurica*, the results of mtDNA + cpDNA sequences and three nuclear sequences indicated that the highest haplotype diversity was in the Shennongjia Mountains (Hubei provinces) (Fig. 3b, d). However, based on 17 pairs of EST-SSR loci, *J. mandshurica* had high allelic richness in Henan and Guizhou provinces (Fig. 3f). To reveal the ecological differentiation between *J. regia* and *J. mandshurica*, we analyzed the niche overlap statistics using five selected climatic factors (Fig. 3g–i). The results indicated that *J. regia* and *J. mandshurica* occupied distinct ecological niches (Fig. 3g). *PCA* results supported that *J. regia* and *J. mandshurica* were primarily separated based on five climate factors (bio4: Temperature Seasonality, bio9: Mean Temperature of the Driest Quarter, bio11: Mean Temperature of the Coldest Quarter, bio15: Precipitation Seasonality, and bio18: Precipitation of the Warmest Quarter) (Fig. 3h, i).

### Population structure of *J. regia* and *J. mandshurica* based on EST-SSR data

Based on EST-SSR data of 16 *J. regia* and *J. mandshurica* sympatric populations, the UPGMA tree resolved two distinct clades (Fig. 4b). Principal component analysis (*PCA*) analysis also resolved two separate clusters corresponding to *J. regia* and *J. mandshurica* (Fig. 4c). Lastly, the structural analysis also resolved two separate genetic clusters corresponding with taxonomic designation (Fig. 4). The population structure analysis showed that the optimal *K* was 2 (Fig. 4d, e). At increasing levels of *K*, subpopulations were delimited within each species but the initial split between the species remained at these higher levels of *K* (Fig. 4d). Overall, the genetic data, including the phylogenetic tree, principal component analysis (*PCA*), and structure analysis indicated that the two walnut sections in China were divided into two distinct genetic clusters based on EST-SSR data. The gene flow analysis produced $\theta$ and *M* values greater than 0 (Table S10). The $\theta$ value and the size of the immigration rate (*M*) revealed a highly symmetric historical gene flow between the two species. The migrate analysis results indicated that *J. regia* and *J. mandshurica* have strong barriers to gene flow (0.49 versus 0.40) (Table S10). These analyses indicated that that there was almost no gene flow between *J.*

**Table 2** Analysis of molecular variance (AMOVA) based on mtDNA, cpDNA, and nuclear DNA sequences

| Species | Source of Variation | mtDNA | cpDNA | ITS | *15R-8* | *Jr5680* | *EST-SSR* |
|---|---|---|---|---|---|---|---|
| *Juglans regia* | Among populations (%) | 5.32 | 24.15 | 23.5 | 25.9 | 43.5 | 16.22 |
| | Within populations (%) | 94.68 | 75.85 | 76.5 | 74.1 | 56.5 | 83.78 |
| | Fixation Index | 0.053* | 0.242* | 0.235* | 0.259* | 0.435* | 0.162* |
| *J. mandshurica* | Among populations (%) | 0.26 | 29.97 | 20.1 | 12.0 | 15.7 | 12.32 |
| | Within populations (%) | 99.74 | 70.03 | 79.9 | 88.0 | 84.3 | 87.68 |
| | Fixation Index | 0.007* | 0.300* | 0.201* | 0.120* | 0.157* | 0.123* |

mtDNA = 234 bp, cpDNA = 1,806 bp, ITS = 639 bp, *15R-8* = 363 bp, *Jr5680* = 639 bp

*$P < 0.001$. Genetically variable sites see Supplementary Table 5–9

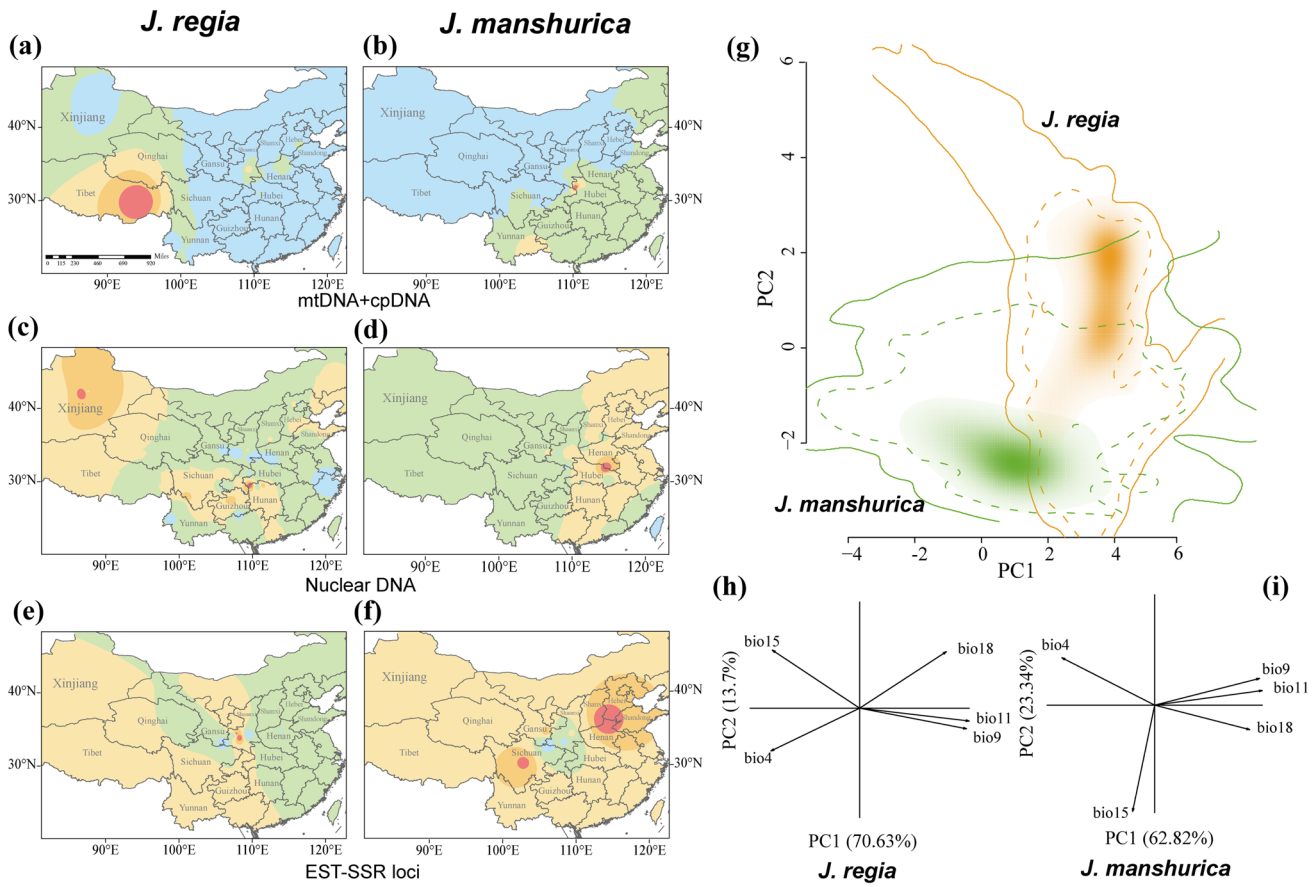


**Fig. 3** Inverse distance weighted interpolation of the haplotype diversity (based on the number of haplotypes using mtDNA + cpDNA and three nuclear loci (ITS, *15R-8*, and *Jr5680*)) and allelic richness (*Rs*) (using EST-SSR loci) of *J. regia* populations and *J. mandshurica* population in China. **a** mtDNA + cpDNA (*J. regia*); **b** mtDNA + cpDNA (*J. mandshurica*); **c** three nuclear loci (*J. regia*); **d** three nuclear loci (*J. mandshurica*); **e** EST-SSR loci (*J. regia*); **f** EST-SSR (*J. mandshurica*). The red, orange, pink, green, and blue regions indicate the haplotype diversity and allelic richness, from high to low. **g** Niche overlaps maps of *J. regia* (orange) and *J. mandshurica* (green) in the environmental space (*E spaces*) using R. The solid and dashed contour lines delimit the 100th and 75th quantiles, respectively, of the density at the available climate. **h**–**i** The contribution of the climatic variables on the two axes of the *PCA* and the percentage of inertia explained by the two axes. bio4: Temperature Seasonality (standard deviation ×100); bio9: Mean Temperature of the Driest Quarter; bio11: Mean Temperature of the Coldest Quarter; bio15: Precipitation Seasonality; bio18: Precipitation of the Warmest Quarter

*regia* and *J. mandshurica* based on evidence from EST-SSR loci (Fig. 4 and Table S10).

## Discussion

### Differences between nuclear, plastid, and mitochondrial datasets

To reveal genetic diversity, haplotype variations, and distribution of *J. regia* and *J. mandshurica* based on maternal genetic markers (*3–9*, *trnL-F* and *trnS-G*) and biparental genetic markers (ITS, *15R-8*, and *Jr5680*), we analyzed 612 individuals of these two walnuts from different locations in China. The genetic diversity of mitochondria sequences from 58 *Juglans* populations was low (Table 1) compared to other species. For example, in *J. regia*, $h_T = 0.024$ and in *J. mandshurica*, $h_T = 0.032$, whereas the mitotype diversity was 0.37 and 0.44 in the conifer species *Picea asperata* ($h_T = 0.37$) and *P. crassifolia* ($h_T = 0.44$) (Du et al. 2009). Based on variability in the mitochondrial *3–9* DNA fragment, only 11 mitotypes associated with eight mutations were found in *Juglans*, whereas we identified a total of 30 chlorotypes in *Juglans* (Fig. 1). We also found no apparent geographic structure to the genetic variability in the mitochondria of *J. mandshurica*, whereas there was clear spatial structure in chloroplasts (Fig. 1). Different cellular compartments can present different outcomes; this is especially the case for biparentally inherited DNA versus uniparentally inherited DNA. The average rate of synonymous substitution in the chloroplast genome is nearly three times higher than that in the mitochondrial DNA in higher plants (Wolfe

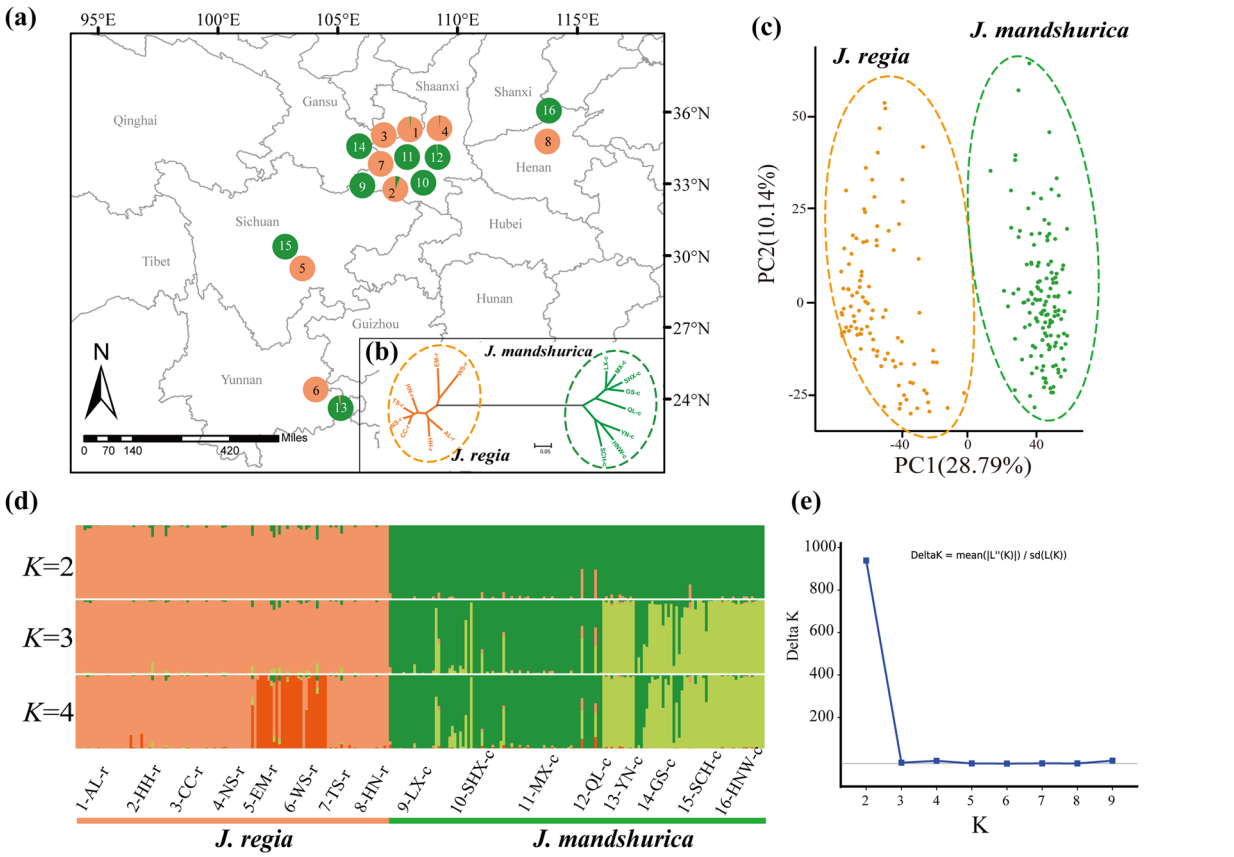


**Fig. 4** Genetic structure and geographical distribution of 16 sympatric walnut (*Juglans*) populations based on 17 EST-SSRs in China. **a** Geographic origin of the eight pairs of sympatric *J. regia* and *J. mandshurica* populations (16 populations) and their color-coded grouping at the highest supported $K=2$. **b** The unweighted pair group method with arithmetic means (UPGMA) tree of 16 walnut populations resolved into two clusters based on the EST-SSR datasets. **c** Principal component analysis (*PCA*) of all individuals of 16 walnut populations, resolved into 2 genotype groups based on the EST-SSR datasets. **d** Results of the sympatric *J. regia* and *J. mandshurica* populations via the "rmaverick" R package, shown from $K=2$, $K=3$, and $K=4$. **e** Results of the deltaK method of Evanno et al. 2005. The location of the pies indicates the sampled locations

et al. 1987), although it is still 4−5 times slower than the total synonymous substitution rate in plant nuclear genome (Wolfe et al. 1989; Nagata 2010) found that in about 20% of angiosperms, the control mechanisms affecting the entry of mitochondrial versus plastid organellar DNA into generative cells were independent (Nagata 2010).

Measures of mutation frequency ($\pi$) in *J. regia* and *J. mandshurica* shed light on their evolutionary history. In both mitochondria and chloroplast sequences, the π value of *J. mandshurica* was higher than that of *J. regia*, and it was considerably higher in cpDNA sequence (0.0008 versus 0.0001). The π value was about the same in nuclear genomes as in organellar genomes at ITS, but at *15R-8* the values of $\pi$ was 30- to 90-fold higher, with values for *J. regia* greater than those for *J. mandshurica* (0.832 versus 0.661). At *Jr5680*, the $\pi$ for *J. regia* was comparable to those for *J. regia* and *J. mandshurica* at *15R-8*, but for *J. mandshurica* at *Jr5680*, values were comparable to those for ITS and even those for organellar genomes. Yet, oddly, we observed negative values for Tajima's D at the locus *Jr5680* in *J. mandshurica*, which indicated an excess of low-frequency polymorphisms, possibly arguing against purifying selection. Tajima's D was not significantly negative, indicating that these groups have not experienced significant expansion (Tajima 1989; Waikhom et al. 2015). A significant positive value of Tajima's D indicates that the gene may have undergone balancing selection or that the species has experienced a demographic bottleneck. Population subdivision and balancing selection may lead to the accumulation of old mutations, which may lead to significantly positive tests of neutrality (Alonso and Armour 2001; Hu et al. 2017a). We observed that $F_{ST}$ was much higher for *J. regia* than *J. mandshurica* at *Jr5680* (a coding gene) and *15R-8* (non-genic), but not at ITS. The coding gene and noncoding nuclear sequence had different patterns of variation and evolutionary rates (Papadopouloua et al. 2016). This outcome may reflect local selection and gene evolution in native populations.

### Similarities and differences between *J. regia* and *J. mandshurica*

Genetic variation was high in both *J. regia* and *J. mandshurica* (Table 2). Because haplotype mtH4 was only found in BJ (Fig. 1), two hypotheses concerning the origin of mtH4 seem reasonable: this haplotype may have evolved locally and not spread (Liu et al. 2009), or it may represent a haplotype that was introduced from outside China. It cannot be ruled out that the current distribution of *J. mandshurica* and especially *J. regia* in China is the product of human selection and of longstanding and potentially long-distance human dispersal. Natural patterns and processes may have been substantially altered, as the importation and cultivation of *J. regia* in China has a long history (Pollegioni et al. 2015). Thus, we cannot be certain the BJ samples were wild trees and not escaped from cultivation. Based on mitochondrial sequences, the values of $G_{ST}$, $N_{ST}$ and $F_{ST}$ of *J. mandshurica* were negative. And $N_{ST}$ was significantly less than $G_{ST}$ indicating that the geographical distribution of haplotypes was probably not related to genetic distance. The negative value of $F_{ST}$ indicated that there was little genetic differentiation among populations of *J. mandshurica*.

Chlorotypes from cpH1, cpH2, and cpH4 to cpH9 were found in *J. regia* in China, while cpH3, and cpH20 to cpH30 were found in *J. mandshurica* populations (Fig. 1 and Table S1). These two walnut species did not share any chloroplast haplotypes. *J. regia* showed similar genetic diversity and similarly complex networks of haplotypes to *J. mandshurica* in China.

Walnuts in China do not have a spatial genetic structure corresponding to a 'core and periphery' model (Feng et al. 2018b), and our AMOVA results indicated that levels of differentiation among populations of conspecifics were relatively high. The mean genetic variation between populations was 31.0 % for *J. regia* populations based on the three nuclear loci, while the mean of molecular variation was 15.9 % between *J. mandshurica* populations (Table 2). The levels of population differentiation ($F_{ST} = 0.24$ (ITS), $F_{ST} = 0.26$ (*15R-8*), and $F_{ST} = 0.46$ (*Jr5680*)) of *J. regia* and *J. mandshurica* ($F_{ST} = 0.20$ (ITS), $F_{ST} = 0.12$ (*15R-8*), and $F_{ST} = 0.16$ (*Jr5680*)) are shown in Table 1. These findings indicate that *J. regia* generally evolved in isolation in China. The best evidence we have for this theory is the marked differentiation of *J. regia* samples from Xinjiang (BM, BZ, XJ), from near Yun-Gui-Chuan (BS, YN, GZ, GY, and ZY), and from the Qinling−Bashan Mountains (NS and CX) and northern China (HL, KC, MT, DZ, SD, LN, and QH) versus all other regions based on *PAL* (*Jr5680*) sequence (Fig. S2). A similar result was found in *J. regia* populations in northern versus southern China-based *PAL* haplotypes (Han et al. 2016). There was a biogeographic divide between the northern and southern parts of East Asia during the Neogene and into the Pleistocene (Bai et al. 2014; Feng et al. 2018b). The north/south phytogeographic divide in China influences *J. regia* and *J. mandshurica* in different ways (Bai et al. 2014; Han et al. 2016). The north/south phytogeographic division in China has been noted frequently, especially with respect to Chinese *Juglans*. Our results showed that both *J. regia* and *J. mandshurica* populations diverged into two subpopulations (Figs. 1 and 2); this may have been the result of ecological differentiation of species, long-term climate change, and/or human influence (Bai et al. 2015; Han et al. 2016; Feng et al. 2018b).

Ecological niche models of Chinese *Juglans* (Bai et al. 2015; Han et al. 2016; Zhao et al. 2018; Feng et al. 2018b) infer refugia for both *J. regia* and *J. mandshurica* in eastern China during the Quaternary. However, *J. regia* may have had an additional refugium in western China (Pollegioni et al. 2015). The IDW results using the haplotype number based on sequence data and the allele richness based on EST-SSR data showed that the centers of genetic diversity for the two walnut species were different, which was consistent with the results of ENMs (Fig. 3) (Zhao et al. 2018). Niche overlap statistics demonstrated that the two species (*J. regia* and *J. mandshurica*) occupied distinct ecological niches based on the environmental space, suggesting that sympatric populations may have arisen recently (*E* spaces; Fig. 3 g). The simultaneous formation of mountain ranges and local climate changes will accelerate niche divergence (Antonelli et al. 2018). However, *J. regia* and *J. mandshurica* can be sympatric distribution and only rarely have gene flow or introgression, which seems to indicate that the two walnut species evolved a hybridization barrier that was maintained over time. Many studies have shown that Quaternary glaciation plays a key role in driving species formation and differentiation (Han et al. 2016; Levsen et al. 2012; Wang et al. 2014). The Xinjiang province, the Qinling−Bashan Mountains, and southwestern China are considered to be important refugia for *J. regia* in China (Bai et al. 2010; Han et al. 2016; Feng et al. 2018b), while *J. mandshurica* refuges were mainly distributed in the Changbai Mountains, the Qinling Mountains, and Taihang Mountains (Bai et al. 2010). The absence of common haplotypes for *J. regia* and *J. mandshurica* probably indicates that these two walnut species evolved separately, and did not share refugia (Bai et al. 2014, 2015; Pollegioni et al. 2015), or, if they did, reproductive barriers kept them distinct. Moreover, in *J. regia*, the higher genetic diversity regions were in western, northern, and southern China, while eastern China was where the genetic diversity of *J. mandshurica* was highest (Figs. 3 and 4).

### The forces of hybridization within a section versus reproductive isolation between sections in natural populations: implications for maintenance of species differences

In China, gene flow and introgression are common for sympatric populations of *Juglans* (Yuan et al. 2018; Zhao et al. 2018). However, only a few instances of gene flow were found between walnut species in different sections, e.g., between *Cardiocaryon* (Asian butternuts) and *Dioscaryon* (Dang et al. 2015; Pollegioni et al. 2013). *J. regia* populations and *J. mandshurica* populations showed little evidence of gene introgression in sympatric populations in the Qinling Mountains (Dang et al. 2015). Neither was it observed in southern China based on mitochondrial, chloroplast, nuclear DNA sequences, or EST-SSRs in this study (Figs. 1, 2 and 4 and S1−2). Gene flow is much more likely between species from different sections (*Cardiocaryon* and *Dioscaryon*) (Figs. 1, 2 and 4 and S1−2). Within the section *Cardiocaryon*, all populations are divided into two clades (northern subpopulations and southern subpopulations) based on genetic data, with a hybrid zone located in northern China (Bai et al. 2015). *J. hopeiensis*, a horticultural variety with limited spatial distribution in northern China, is a hybrid of *J. regia* and *J. mandshurica* (Zhao et al. 2018). There is frequent gene flow and introgression between *J. regia* and *J. sigillata* (Yuan et al. 2018; Zhao et al. 2018). *J. regia* did not shared any haplotypes with *J. mandshurica*, but *J. sigillata* and *J. regia* did share some haplotypes at *Jr5680*. The structure and migrate analysis based on EST-SSRs data showed that there was almost no gene flow between *J. regia* and *J. mandshurica*, even if they were sympatric, which was consistent with the haplotype distribution of five fragments (Fig. 4). All *Juglans* species are wind-pollinated and deciduous trees (Bai et al. 2014, 2015; Zhao et al. 2018); however, the forces of hybridization within sections and reproductive isolation between sections in natural populations appear to maintain the species differences, especially in regions where section *Juglans* are sympatric (Lowry et al. 2008). This complex phenomenon needs to be confirmed in the future.

## Conclusion

We studied the phylogeographic history of *J. regia* and *J. mandshurica* using nuclear and cytoplasmic DNA collected from a total of 559 individuals. Based on all sequence data, *J. regia* rarely shared any haplotypes with *J. mandshurica*. The structure and migrate analysis based on EST-SSR data showed that there was almost no gene flow between *J. regia* and *J. mandshurica*, even in instances of sympatry, which was also consistent with the haplotype distribution. Both *J. regia* and *J. mandshurica* had high genetic diversity and shared refugia in eastern China during the Quaternary period. However, *J. regia* was inferred to have had an additional refugium in western China. Niche overlap statistics demonstrated that the two species (*J. regia* and *J. mandshurica*) occupied distinct ecological niches based on the environmental space (*E* spaces). The forces of hybridization within a section and reproductive isolation between sections in natural populations may preserve the species as distinct, even in areas of sympatric distribution.

**Author contribution statement** PZ designed the project. MD and HZ performed the experiments. PZ, MD, HZ, YZ, and MY did the fieldwork. PZ, HZ, YZ, and MD collected the leaf samples. MD, HZ, and PZ did the DNA extraction and PCR reaction. MD and PZ completed the DNA sequence alignment and analyzed the data. MD, HZ, and PZ finished the EST-SSR genotyping experiment and data analysis. PZ and MY supported the software resources. MD, HZ, KW, and PZ conceived the study. PZ, MD, HZ, MY, SZ, and KW wrote the draft manuscript. PZ and KW improved the writing of the manuscript. PZ, MD, GZ, and SZ revised the manuscript. All the authors have approved the final manuscript.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s00468-021-02167-y.

**Acknowledgements** We thank Nan Hou, Yiheng Hu, Tao Zhou, and Lei Wang for assisting with sampling. 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.

**Funding** This work was supported by the National Natural Science Foundation of China (32,070,372, 41,471,038, and 31,200,500), Shaanxi Academy of Science Research Funding Project (2019 K-06), the Natural Science Foundation of Shaanxi Province of China (2019JM-008), the Opening Foundation of the Key Laboratory of Resource Biology and Biotechnology in Western China (Northwest University), the Ministry of Education (ZSK2018009), and the Program for Excellent Young Academic Backbones, funded by Northwest University.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

## References

Alonso S, Armour J (2001) A highly variable segment of human subterminal 16p reveals a history of population growth for modern humans outside Africa. Proc Natl Acad Sci USA 98:864–869

Antonelli A, Kissling WD, Flantua SGA, Bermúdez MA et al (2018) Geological and climatic influences on mountain biodiversity. Nat Geosci 11:718–725

Avise JC (1998) Pleistocene phylogeographic effects on avian populations and the speciation process. Proc R Soc Lond B Biol Sci 265:457–463

Avise JC (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambridge

Avise JC (2009) Phylogeography: retrospect and prospect. J Biogeography 36:3–15

Aradhya MK, Potter D, Gao F et al (2007) Molecular phylogeny of *Juglans* (Juglandaceae): a biogeographic perspective. Tree Genet Genomes 3:363–378

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 188:892–901

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 209:1757–1772

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

Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. In: Harrison RG Hybrid Zones and the Evolutionary Process. Oxford University Press, Ed, New York NY USA 13–45

Beerli P (2005) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. Bioinformatics 22:341–345

Broennimann O, Fitzpatrick MC, Pearman PB et al (2011) Measuring ecological niche overlap from occurrence and spatial environmental data. Global Ecol Biogeogr 21:481–497

Casazza G, Grassi F, Zecca G, Minuto L (2016) Phylogeographic insights into a peripheral refugium: the importance of cumulative effect of glaciation on the genetic structure of two endemic plants. PloS One 1:e0166983

Chen DM, Zhang XX, Kang HZ, Sun X, Yin S, Du HM, Yamanaka N et al (2012) Phylogeography of *Quercus variabilis* based on chloroplast DNA sequence in East Asia: multiple glacial refugia and mainland-migrated island populations. PloS One 7:e47268

Dang M, Liu ZX, Chen X et al (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

Dang M, Yue M, Zhang M, Zhao GF, Zhao P (2019) Gene introgression among closely related species in sympatric populations: a case study of three walnut (*Juglans*) species. Forests 10:965

Dreyer JBB, Higuchi P, Silva AC (2019) Ligustrum lucidum W. T. Aiton (broad-leaf privet) demonstrates climatic niche shifts during global-scale invasion. Sci Rep 9:3813

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

Du FK, Petit RJ, Liu JQ (2009) More introgression with less gene flow: chloroplast vs. mitochondrial DNA in the Picea asperata complex in China, and comparison with other conifers. Mol Ecol 18:1396–1407

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

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

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

Fan DM, Yue JP, Nie ZL, Li ZM, Comes HP, Sun H (2013) Phylogeography of *Sophora davidii* (Leguminosae) across the 'Tanaka-Kaiyong Line', an important phytogeographic boundary in Southwest China. Mol Ecol 22:4270–4288

Feng XJ, Yuan XY, Sun YW et al (2018a) Resources for studies of Iron walnut (*Juglans sigillata*) gene expression, genetic diversity and evolution. Tree Genet Genomes 14:51

Feng XJ, Zhou HJ, Saman Z et al (2018b) The phytogeographic history of common walnut in China. Front Plant Sci 9:1399

Fuertes AJ, Nieto FG (2003) Additive polymorphisms and reticulation in an ITS phylogeny of thrifts (Armeria, Plumbaginaceae). Mol Phylogenet Evol 28:430–447

Fuertes AJ, Rosselló JA, Nieto FG (1999) Nuclear ribosomal DNA (nrDNA) concerted evolution in natural and artificial hybrids of Armeria (Plumbaginaceae). Mol Ecol 8:1341–1346

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Gugger PF, Sugita S, Cavender-Bares J (2010) Phylogeography of Douglas-fir based on mitochondrial and chloroplast DNA sequences: testing hypotheses from the fossil record. Mol Ecol 19:1877–1897

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Sympos Series 41:95–98

Han H, Woeste KE, Hu YH et al (2016) Genetic diversity and population structure of common walnut (*Juglans regia*) in China based on EST-SSRs and the nuclear gene phenylalanine ammonia-lyase (*PAL*). Tree Genet Genomes 12:111–122

Hoban SM, McCleary T, Schlarbaum S et al (2009) Geographically extensive hybridization between the forest trees American butternut and Japanese walnut. Biol Lett 5:324–327

Hu YH, Dang M, Feng XJ, Woeste KE, Zhao P (2017a) Genetic diversity and population structure in the narrow endemic Chinese walnut *Juglans hopeiensis* Hu: implications for conservation. Tree Genet Genomes 13:91

Kalinowski ST (2005) HP-Rare: A computer program for performing rarefaction on measures of allelic diversity. Mol Ecol 5:187–189

Levsen ND, Tiffin P, Olson MS (2012) Pleistocene speciation in the genus *Populus* (Salicaceae). Syst Biol 61:401–412

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

Liu JQ, Sun YS, Ge XJ, Gao LM, Qiu YX (2012) Phylogeographic studies of plants in China: advances in the past and directions in the future. J Syst Evol 50:267–275

Liu Y, Wang Y, Huang H (2009) Species-level phylogeographical history of *Myricaria* plants in the mountain ranges of western China and the origin of *M. laxiflora* in the Three Gorges mountain region. Mol Ecol 18:2700–2712

Lowry DB, Modliszewski JL, Wright KM (2008) The strength and genetic basis of reproductive isolating barriers in flowering plants. Philos Trans R Soc B Biol Sci 363:3009–3021

Lu AM, Stone DE, Grauke LJ (1999a) Juglandaceae. In: Wu ZY, Raven PH (Eds) Flora of China. Science Press. Beijing 4: 282–283

Lu AM, Stone DE, Grauke LJ (1999) Juglandaceae. Flora China 4:277–285

Manning WE (1978) The classification within the Juglandaceae. Ann Missouri Bot Gard 65:1058–1087

Mu XY, Sun M, Yang PF et al (2017) Unveiling the identity of wenwan walnuts and phylogenetic relationships of Asian *Juglans* species using restriction site-associated DNA-sequencing. Front Plant Sci 8:1708

Nagata N (2010) Mechanisms for independent cytoplasmic inheritance of mitochondria and plastids in angiosperms. J Plant Res 123:193–199

Papadopouloua A, Knowlesa LL (2016) Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. P Natl Acad Sci USA 113:8018–8024

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

Petitpierre B, Kueffer C, Broennimann O, Randin C et al (2012) Climatic niche shifts are rare among terrestrial plant invaders. Science 335:1344–1348

Pollegioni P, Woeste KE, Chiocchini F et al (2014) Landscape genetics of Persian walnut (*Juglans regia* L.) across its Asian range. Tree Genet Genomes 10:1027–1043

Pollegioni P, Olimpieri I, Woeste KE et al (2013) Barriers to interspecific hybridization between *Juglans nigra* L. and *J. regia* L. species. Tree Genet Genomes 9:291–305

Pollegioni P, Woeste KE, Chiocchini F et al (2015) Ancient humans influenced the current spatial genetic structure of common walnut populations in Asia. PLoS One 10:e0135980

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

Qiu YX, Fu CX, Comes HP (2011) Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. Mol Phylogenet Evol 59:225–244

Schneider S, Roessli D, Excoffier L (2000) Arlequin: a software for population genetics data analysis. User Manual Ver 2:2496–2497

Shu Z, Zhang X, Yu DQ et al (2016) Natural hybridization between Persian walnut and Chinese walnut revealed by simple sequence repeat markers. J Am Soc Hortic Sci 141:146–169

Stanford AM, Harden R, Parks CR (2000) Phylogeny and biogeography of *Juglans* (Juglandaceae) based on matK and ITS sequence data. Am J Bot 87:872–882

Sun YW, Hou N, Woeste KE et al (2019) Population genetic structure and adaptive differentiation of iron walnut *Juglans regia* subsp. *sigillata* in southwestern China. Ecol Evol 9:14154–14166

Sunnucks P (2000) Efficient genetic markers for population biology. Trends Ecol Evol 15:199–203

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Takezaki N, Nei M, Tamura K (2010) POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with windows interface. Mol Biol Evol 27:747–752

Tamura K, Stecher G, Peterson D et al (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 197:2725

Van LM, Joseph JA, Heinze B, Fay MF, Lexer C (2008) Clonality and spatial genetic structure in Populus×canescens and its sympatric backcross parent P. alba in a central European hybrid zone. New Phytol 177:506–516

Waikhom B, Anirudh K et al (2015) Nucleotide diversity analysis of three major bacterial blight resistance genes in rice. Plos One 10:e0120186

Wang WT, Xu B, Zhang DY, Bai WN (2016) Phylogeography of postglacial range expansion in *Juglans mandshurica* (Juglandaceae) reveals no evidence of bottleneck, loss of genetic diversity, or isolation by distance in the leading-edge populations. Mol Phylogenet Evol 102:255–264

Wang J, Kallman T, Liu J, Guo Q et al (2014) Speciation of two desert *poplar* species triggered by Pleistocene climatic oscillations. Heredity 112:156–164

Wickham H (2009) ggplot2, elegant graphics for data analysis. Springer, New York

Woeste K, Michler C (2011) Genomic and breeding resources. In: Chittaranjan K (ed) Wild Crop Relatives. Springer, Berlin/Heidelberg, Germany

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84:9054–9058

Wolfe KH, Sharp PM, Li WH (1989) Rates of synonymous substitutions in plant nuclear genes. J Mol Ecol 29:208–211

Wright S (1978) Variability within and among natural populations. University of Chicago Press, Chicago

Yuan XY, Sun YW, Bai XR, Dang M et al (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 P, Woeste KE (2011) DNA markers identify hybrids between butternut (*Juglans cinerea* L.) and Japanese walnut (*Juglans ailantifolia* Carr.). Tree Genet Genomes 7:511–533

Zhao P, Zhou HJ, Potter D, Hu YH, Feng XJ, Dang M et al (2018) Population genetics, phylogenomics, and hybrid speciation of *Juglans* in China determined from whole chloroplast genomes, transcriptomes, and genotyping-by-sequencing (GBS). Mol Phylogenet Evol 126:250–265