Population genetics, phylogenomics and hybrid speciation of *Juglans* in China determined from whole chloroplast genomes, transcriptomes, and genotyping-by-sequencing (GBS)

Peng Zhao<sup>a,⁎,1</sup>, Hui-Juan Zhou<sup>a,1</sup>, Daniel Potter<sup>c</sup>, Yi-Heng Hu<sup>a</sup>, Xiao-Jia Feng<sup>a</sup>, Meng Dang<sup>a</sup>, Li Feng<sup>a</sup>, Saman Zulfiqar<sup>a</sup>, Wen-Zhe Liu<sup>a</sup>, Gui-Fang Zhao<sup>a</sup>, Keith Woeste<sup>b</sup>,<sup>n</sup>

<sup>a</sup>Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education, College of Life Sciences, Northwest University, Xi’an, Shaanxi 710069, China

<sup>b</sup>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

<sup>c</sup>Department of Plant Sciences, University of California, Davis, CA 95616, USA

**Keywords:** Population genetics, phylogenomics and hybrid speciation of *Juglans* genus, gene introgression, phylogeography, hybridization speciation, *Juglans*

**Abstract**

Genomic data are a powerful tool for elucidating the processes involved in the evolution and divergence of species. The speciation and phylogenetic relationships among Chinese *Juglans* remain unclear. Here, we used results from phylogenomic and population genetic analyses, transcriptomics, Genotyping-By-Sequencing (GBS), and whole chloroplast genomes (Cp genome) data to infer processes of lineage formation among the five native Chinese species of the walnut genus (*Juglans*, Juglandaceae), a widespread, economically important group. We found that the processes of isolation generated diversity during glaciations, but that the recent range expansion of *J. regia*, probably from multiple refugia, led to hybrid formation both within and between sections of the genus. In southern China, human dispersal of *J. regia* brought it into contact with *J. sigillata*, which we determined to be an ecotype of *J. regia* that is now maintained as a landrace. In northern China, walnut hybridized with a distinct lineage of *J. mandshurica* to form *J. hopeiensis*, a controversial taxon (considered threatened) that our data indicate is a horticultural variety. Comparisons among whole chloroplast genomes and nuclear transcriptome analyses provided conflicting evidence for the timing of the divergence of Chinese *Juglans* taxa. *J. cathayensis* and *J. mandshurica* are poorly differentiated based on our genomic data. Reconstruction of *Juglans* evolutionary history indicate that episodes of climatic variation over the past 4.5 to 33.80 million years, associated with glacial advances and retreats and population isolation, have shaped Chinese walnut demography and evolution, even in the presence of gene flow and introgression.

**1. Introduction**

Walnuts and butternuts (*Juglans*) are known for their edible nuts and high-quality wood (Manning, 1978; Aradhya et al., 2007). The genus *Juglans* includes about 21 species distributed in Asia, southern Europe, North America, Central America, western South America, and the West Indies (Manning, 1978; Stanford et al., 2000; Aradhya et al., 2007). Species of *Juglans* are diploid, with a karyotype of 2n = 2x = 32 (Woodworth, 1930). All *Juglans* are monoecious, wind-pollinated, temperate deciduous trees (Manning, 1978). *J. regia* (Common walnut), *J. sigillata* (Iron walnut), *J. cathayensis* (Chinese walnut), *J. hopeiensis* (Ma walnut), and *J. mandshurica* (Manchurian walnut) grow in China (Manning, 1978; Aradhya et al., 2007). Chinese *Juglans* species are divided into two sections (sect. *Dioscaryon*, and sect. *Cardiocaryon*) based on species’ geographical distribution, leaf, flower, and fruit morphology (Manning, 1978) and molecular evidence (Fjellstrom and Parfitt, 1995; Stanford et al., 2000; Aradhya et al., 2007). *J. regia* and *J. sigillata* belong to sect. *Dioscaryon*, and the other three species (*J. cathayensis*, *J. hopeiensis*, and *J. mandshurica*) belong to sect. *Cardiocaryon* (Stanford et al., 2000; Aradhya et al., 2007). Phylogeny based on complete chloroplast genomes, protein coding sequences (CDS), and the introns and spacers (IGS) data (Hu et al., 2017) strongly supported division of the five Chinese walnut species into two previously recognized sections (*Juglans/Dioscaryon* and *Cardiocaryon*) with a 100%...
Juglans regia, while Iron walnut (J. cathayensis) is narrowly distributed in northern China in the Iberian Peninsula (Manning, 1978; Draine and Hiden 1998; Martínez-García et al., 2016; Pollegioni et al., 2017). J. regia is native to the mountainous regions of central Asia (Pollegioni et al., 2015; Martínez-García et al., 2016), while Iron walnut (J. sigillata) is indigenous to China, distributed mainly in southwestern China (Wang et al., 2015) sympatric with J. regia. J. cathayensis is widely distributed in southern China (Bai et al., 2014), while J. mandshurica is mainly distributed in North China, Northeast China and the Korean Peninsula (Wang et al., 2016). J. hopeiensis is narrowly distributed in northern China in the hilly, mid-elevation area between Hebei province, Beijing, and Tianjin (Hu et al., 2015). A strongly supported phylogeny of these five species is not available due to a lack of informative molecular markers (Fjellstrom and Parfit, 1995; Stanford et al., 2000; Aradhya et al., 2007). Studies of gene introgression and introgression have concluded J. regia and J. sigillata are particularly closely related, and some have questioned whether they are distinct (Wang et al., 2008; Wang et al., 2015). Aradhya et al. (2007) used ITS, RFLP, and cpDNA sequence data to suggest J. regia and J. sigillata are distinct species. Grimshaw (2003) considered J. sigillata distinct and valid based on morphology.

The relationships among species of sect. Cardio Caryon are unsettled. For example, the relationship of Ma walnut (J. hopeiensis) to other members of the section Cardio Caryon, especially J. mandshurica (Lu et al., 1999; Aradhya et al., 2007) is disputed. Although the previous phylogenetic study concluded Ma walnut is a well-defined lineage and a sister clade to J. altantolia, J. mandshurica and J. cathayensis within section Cardio Caryon (Stanford et al., 2000; Aradhya et al., 2007), evidence from randomly amplified polymorphic DNA (RAPD) markers, isozymes, and karyotype analysis indicated that J. hopeiensis might have arisen from the recent hybridization of J. regia and J. mandshurica (Wu et al., 1999; Mu et al., 1990). Grimshaw (2003) considered J. hopeiensis and J. cathayensis to be synonyms of J. mandshurica. J. cathayensis and J. mandshurica were also combined into one species in Flora of China (English version) (Lu et al., 1999), which does not consider J. hopeiensis (Kuang and Lu 1979; Aradhya et al., 2007) a valid taxon. Others have suggested that J. hopeiensis is a variant of J. mandshurica based on another characteristics and morphology (Lu et al., 1999). Several important studies have drawbacks/shortages in sampling because their authors
did not have access to adequate samples of all Chinese Juglans (Fjellstrom and Parfitt, 1995; Stanford et al., 2000; Aradhye et al., 2007).

There are many potential advantages to using datasets that span entire nuclear and chloroplast genomes (Cp genome) to sort out phylogenetic and evolutionary history (McCormack et al., 2013; Stöltling et al., 2015; Daniell et al., 2016). Large, genome-scale datasets can be used to address phylogenetic relationships among closely related species and, at the same time, examine patterns of lineage sorting and historical hybridization (Escudero et al., 2014; Dodsworth et al., 2015; Daniell et al., 2016). The uniparental inheritance (maternal transmission), haploid state, and general absence of recombination of the Cp genome (limiting gene flow to seed dispersal only, Moore et al., 2010) make Cp genome sequences particularly useful for studies of plant population genetic and phylogeography (Perdereau et al., 2017). Analysis of genetic variability within the nuclear genome using orthologous genes and Genotyping-by-Sequencing (GBS) has the potential to resolve divergence that straddles the population-species boundary (Davey et al., 2011; Nicotra et al., 2016; Mattila et al., 2012; Yang et al., 2014).

Here, we use population genomic approaches to clarify the evolutionary relationships among the five Chinese Juglans species and to gain insight into intraspecific variation within each of the two sections of Juglans native to China (Stanford et al., 2000; Aradhye et al., 2007). Other aims include to determine the nature and extent of gene flow between sect. Dioscaryon and sect. Cardiocaryon and its consequences, and to determine the relationship between J. regia and J. sigillata. Finally, we wanted to determine if J. hopeiensis arose from a recent hybridization and, if so, how and when this event occurred.

2. Materials and methods

2.1. Sample collection, DNA extraction, and RNA extraction

For whole chloroplast genome (Cp genome) research, fresh leaves of 34 healthy tree of five Juglans species were collected from different locations in China (eleven Juglans regia, six J. sigillata, five J. hopeiensis, six J. cathayensis, and six J. mandshurica, Table 1) and silica gel-dried and stored at −4°C. High-quality genomic DNA was extracted using a modified CTAB method (Zhao and Woeste, 2011). For transcriptome research, fresh leaves, buds, flowers were collected from single, mature, healthy-appearing J. regia (Qinling Mountains), J. sigillata (Yunnan province), J. hopeiensis (Laishui, Beijing), J. cathayensis (Qinling Mountains), and J. mandshurica (Xiaolongmen, Beijing) and immediately frozen in liquid nitrogen prior to storage at −80°C. Total RNA was extracted using a Plant RNA Kit (OMEGA Bio-Tek, Norcross, GA, USA). RNA degradation and contamination was monitored on 1% agarose gels (details see Hu et al., 2017).

For Genotyping by sequencing (GBS), five Chinese walnut (Juglans) species were collected from field sites across their biological ranges (Table 1). A total of 140 individuals (65 Juglans regia, 13 J. sigillata, 14 J. hopeiensis, 32 J. cathayensis, and 16 J. mandshurica) from 47 locations (18 populations of J. regia, 5 populations of J. sigillata, 5 populations of J. hopeiensis, 15 populations of J. cathayensis, and 4 populations of J. mandshurica) were collected for analysis (Table 1). Each sampled tree was a mature adult, 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 100 m. Sampled locations were mapped using ArcGIS (version 10.0; ESRI, 2010) (Fig. 1). Fresh leaves were dried with silica gel prior to DNA extraction. DNA was extracted following the methods described by Doyle and Doyle (1987) and Zhao and Woeste (2011). The DNA was quantified and its quality evaluated using three methods: (i) Agarose gel electrophoresis to test DNA purity and integrity, (ii) Nanodrop (Wilmington, DE, USA) test DNA purity (OD260/OD280), (iii) Qubit® DNA Assay Kit in Qubit® 2.0 Fluorometer (Life Technologies, CA, USA) to measure DNA concentration.

2.2. Chloroplast genome sequencing and phylogenetic analysis

We previously sequenced the complete chloroplast genome of J. regia with the Illumina HiSeq sequencing platform was performed by Novogene Bioinformatics Technology Co., Ltd., Beijing, China (www.novogene.cn). The complete chloroplast genomes of all samples (Table 1) were sequenced using Illumina HiSeq 2500 and assembled via reference-guided assembly based on the Cp genome of J. regia (Hu et al., 2016, NCBI Accession number: KT963008) (Table S1). Raw sequence reads obtained from whole genome sequencing were aligned to the J. regia chloroplast genome using Bowtie2 v2.2.6 (Langmead and Salzberg, 2012) and assembled into genomes using sPaDses v3.6 (Bankevich et al., 2012) assembler. Gaps and boundaries were extended using mitobim v. 1.7 (Hahn et al., 2013) based on mira v. 4 (Chevreux et al., 2004). Haplotype diversity (Hd) and nucleotide diversity (π) were calculated using DNASP v.5 (Librado and Rozas, 2009). To generate median-joining (MJ) haplotype networks, all chloroplast genome sequence variation was analyzed using Popart v1.7 (Forster and Ro, 1994; Leigh and Bryant, 2015). Heterozygous bases were removed and sequences were aligned using MAFFT v7.017 (Katoh and Standley, 2013). Statistical selection of a TVM+I substitution model was made using ModelTest v3.7 (Posada and Crandall 1998) and a phylogenetic tree was then estimated using MrBayes v3.2.6 (Ronquist et al., 2012). Run parameters were: a chain length of 1,100,000 and a burn-in length of 100,000 over 4 heated-chains (chain temp 0.2). Phylogenetic analyses of the Cp genome were performed using all five Chinese Juglans species, J. nigra, Carya sinensis, and Castanea mollissima (Jansen et al., 2011, NCBI accession number: NC_014674.1) (Details see Hu et al., 2017).

2.3. GBS sequencing and SNP analysis

Genotyping-by-Sequencing (GBS) was performed as described by Elshire et al. (2011) at Novogene Bioinformatics Technology Co., Ltd., Beijing, China (www.novogene.cn). Briefly, genomic DNA from individual samples was digested with Msel and HaeIII restriction enzyme. Digested DNA was ligated with one of 96 uniquely barcoded sequencing adaptor pairs (Elshire et al., 2011; Morris et al., 2011). Library amplifications between 250 and 600 bp were extracted from an agarose gel and sequenced in a HiSeq2500-PE125 using 150 bp Paired End protocol.

The original image data obtained by high throughput sequencers was transformed to raw data that was filtered to good reads containing adapters. The remaining high quality reads were mapped to the walnut reference genome (Martinez-Garcia et al., 2016, http://dendrome.ucdavis.edu/ftp/Genome_Data/genome/Reju/) using BWA v0.7.8 (Li and Durbin, 2009) with the command ‘mem-4-k32-M’. After alignment, we performed SNP calling on a population scale using a Bayesian approach as implemented in the package SAMtools v1.3.1 (Li et al., 2009). We then calculated genotype likelihoods from reads for each individual at each genomic location, and the allele frequencies in the sample with a Bayesian approach. The ‘mpileup’ command was used to identify SNPs with the parameters as ‘-q 1 -C 50 -t SP -t DP -m2 -F 0.002’. Then, to exclude SNP calling errors caused by incorrect mapping or Indels, only high quality SNPs (coverage depth ≥2 and ≤80, RMS mapping quality ≥20,maf ≥0.05, miss ≤0.1) were kept for phylogeny analysis. Consequently, 45,548 SNPs were left for further analysis after filter from 4,343,085 raw SNPs.

2.4. Population genetic polymorphism using GBS sequencing data

Phylogenetic analyses were performed on 45,548 SNP loci present in at least half of the individuals we sequenced (but present in all species). After the SNP detection, the individual SNPs were used to calculate the distance among populations. The software TreeBest v1.9.2 was used to calculate the distance matrix, and on this basis, a phylogenetic tree was constructed by neighbor-joining. Guide values
(bootstrap values) were calculated more than 1000 times (Li et al., 2006). We also constructed a Maximum Likelihood (ML) phylogenetic tree using all 140 individuals of the five Chinese Juglans species (Fig. S4). Based on concatenated SNP sequences, divergence times were estimated by MCMCtree package in PAML 4.5 (Yang, 2007), using soft fossil constraints under various molecular clock models. The time unit was 100 Million years (Mya). We added the calibration of the most recent common ancestor of Cardiocaryon and Rhysocaryon [0.1, 0.2] (Bai et al., 2016), to edit the tree. The clock variable was set to 2. The process was run to sample 10,000 times using HKY85 as the substitution model, with the sample frequency set to 50 after a burn-in of 50,000 iterations.

Fig. 1. Species ranges, geographical distribution of haplotypes, and phylogenetic relationship among five Chinese walnut (Juglans) taxa. (a) Geographical distribution of 18 chloroplast haplotypes in five Chinese Juglans and sampled locations. Shading indicates the native range of each of the five Chinese Juglans based on the native distribution of Juglans (Aradhya et al., 2007). Colors of haplotypes correspond of those of the small Fig. in the left corner of the map. Haplotypes H9 and H12 are represented using two colors to indicate that J. hopeiensis and J. mandshurica shared H9, and J. sigillata and J. regia shared H12. The triangles indicate J. regia and J. sigillata, circles indicate J. hopeiensis, J. mandshurica, and J. cathayensis. The perimeter color of triangles and circles indicates species. (b) The minimum spanning network of 18 chloroplast genome haplotypes rooted by Juglans nigra and Carya sinensis. In the network diagram, small red circles indicate intermediate haplotypes not detected in the dataset; red split lines and numbers indicate mutation steps supported by indels, while black lines indicate one mutation step supported by indels. Colors of haplotypes correspond to taxa as indicated in the inset box of Fig. 1a. The yellow (H18) and black circles (H17) indicate the haplotypes of Carya sinensis and J. nigra, respectively.
2.5. Population structure and principal components analysis

Population genetic structure was analyzed via an expectation maximization algorithm, as implemented in the program FRAPPE v1.170 (http://med.stanford.edu/tanglab/software/frappe.html). This approach involves use of multiple loci GB data to cluster samples into groups (K). The number of genetic clusters/groups, K, was tested from 1 to 10 and 20 independent runs with 100,000 burn-in iterations, followed by 500,000 Markov Chain Monte Carlo (MCMC) iterations for each K value. The most likely number of genetic clusters (optimal K) was determined using ADMIXTURE (Alexander et al., 2009). Analysis was performed at several different levels of differentiation; we initially used all samples, and then hierarchically tried smaller structure groups consisting of subgroups of species. The principal component analysis (PCoA) of 140 samples was performed based on SNPs among individual genomes using GCTA (Yang et al., 2011) and GAPIT software (Lipka et al., 2012).

2.6. Transcriptome sequencing and phylogenetic analysis

Transcriptome sequencing was performed on Illumina HiSeq2000 by Novogene Bioinformatics Technology Co., Ltd., Beijing, China (http://www.novogene.cn). Orthologous groups were constructed from the BLASTP results with OrthoMCL v2.0.9 (Fischer et al., 2011) using default settings and used for the phylogenetic tree reconstruction.

2.7. Ecological niche modeling

Geographical information for the presence of the five Chinese Juglans species (J. regia, n = 1, 200 individuals; J. sigillata, 500 individuals; J. hopeiensis, 50 individuals; J. mandshurica, 500 individuals; J. cathayensis, 600 individuals) was obtained from sampling locations, the distribution records for J. hopeiensis sourced from the National Specimen Information Infrastructure (http://www.nsci.org.cn/), and China Virtual Herbarium (http://www.cvh.org.cn/). Ecological niche modeling was employed to provide evidence for the location and extent of habitat for each of the five species during last glacial maximum (LGM, 21 ka BP), the inter-glacial period (LIG, 130 ka BP), and recent past. Species distribution models for all five Chinese walnuts were generated using MAXENT v3.3.3 (Phillips et al., 2006). Prior to modeling, all records were mapped and examined to identify and exclude records with obvious geo-referencing errors and misidentifications.

To explore the potential distribution of five Chinese walnut species under current climatic conditions and predict where suitable conditions were present in the past based on paleoclimate environmental layers, we used 19 biologically meaningful climate variables (BIO1-19, Table S9). The 19 bioclimatic layers were downloaded from WorldClim website at 2.5 arc-min resolution (http://www.worldclim.com). A current distribution model was then projected onto the set of climatic variables simulated 20 times by both the Community Climate System Model (CCSM3.0; Collins et al., 2006) and the Model for Interdisciplinary Research on Climate (MIROC, Otto-Bliesner et al., 2006) to infer the extent of suitable habitat during the last glacial maximum (LGM, 21 ka BP). The paleo-coastlines during the LGM were estimated assuming a 130 m lower sea level. In addition, we projected the model to the last stage of the inter-glacial period (LIG, 130 ka BP) using the climate model of Otto-Bliesner et al. (2006). In order to examine the relative importance of the climate variables on the species distribution, we evaluated percent contribution, permutation importance, and jackknife tests (Ornelas et al. 2016). To assess the degree of ecological niche overlap among Juglans species and different lineages, we performed pairwise analyses, examining the niche space between J. regia/J. sigillata, J. regia/J. hopeiensis, J. regia/J. mandshurica, J. mandshurica/J. hopeiensis, J. mandshurica/J. cathayensis, and sub-cluster I of J. regia (Northwestern China, Xinjiang province)/ sub-cluster II of J. regia (mostly southern and eastern China). We evaluated niche model overlap using Schoener (1968) and Hellingier’s D calculated in ENMTools (Warren et al., 2010; Dowell and Hekkala, 2016). Schoener’s D and Hellingier’s I range from 0 (no overlap) to 1 (complete overlap) (Warren et al., 2008). Significance of niche similarity metrics was tested in ENMTools, assessed with a one-sided Wilcoxon test, and plotted using the ggplot2 package (Wickham, 2009) in R v3.1.3 (R Core Team, 2014).

2.8. Impact of environmental factors on genetic structure (isolation by environment)

In order to evaluate the effect of present climatic conditions on the observed pattern of genetic differentiation, we tested for the relationship between pairwise Fst and climatic distance while controlling for geographic distance for the 47 populations. Nineteen bioclimatic layers for the occurrence points as used earlier in ecological niche modeling (ENM) were summarized into the first two axes of a principal coordinate analyses (PCoA) using R 3.1.0. We computed climatic (Euclidian) distance matrices based on population scores for both PCoA axes (PC1 and PC2), and for each bioclimatic layers. Tests were performed for the whole data set, including both northern and southern lineages, using partial Mantel tests (‘mantel.partial’ function; R Core Team 2013) based on 10,000 permutations.

3. Results

3.1. Sequence assembly and SNP detection

In total, 49.57 GB quality reads of transcriptome were generated, resulting in a total of 46.37 GB of clean data for assembly (Dang et al., 2015, 2016; Hu et al. 2015, 2016, details see Table S2). The mean Q30 percentage (sequencing error rate < 0.05%) ranged from 90.99 to 91.88 across all five species, overall mean Q30 percentage was 91.35. GC content percentage ranged from 45.16 to 46.64, mean GC percentage was 45.52 (Table S2). De novo assemblies generated 1,039,412 transcripts including 451,242 unigenes. The mean length of the transcripts varied from 1212 bp to 2,577 bp, with an average of 1361 bp, and mean N50 value was 2282 bp. The number of unigenes varied from 81,909 to 103,167 with an average size of 670 bp and mean N50 value of 1186 bp. Full-length coding DNA sequences (CDS) (N = 94,933) longer than 500 bp accounted for about 45.38%. When we aligned the CDS with a threshold of 1e-5 by performing a BLASTX search against diverse protein databases, a total of 43,807 gene families were obtained, including 6421 Core Orthologs (Table S3). When we extracted and aligned the putative CDSs, a total of 190 orthologous unigenes were found and annotated for all five Chinese Juglans based on comparison with ten other plant species including Oryza sativa, Glycine max, Malus domestica, Vitis vinifera, Ricinus communis, Populus trichocarpa, Theobroma cacao, Castanea mollissima, and Arabidopsis thaliana (Table S4).

Using GBs, 1179.4 Gb raw reads and 1172.4 Gb clean reads were generated. For each sample, 30,213 high-quality tags were identified from 4.07 G pairwise-end reads (Table S5). The sequence data was high-quality (Q20 ≥ 90% and Q30 ≥ 85%). The mean GC content was 36.7% (Table S6). A total of 8,095,840 SNPs with an MAF ≥ 0.05 were identified from 47 populations of 140 individuals with a mean of 57,827 SNPs (Table S7) in each of the five Chinese Juglans. A total of 45,548 SNPs were identified in the nuclear genomes of Chinese Juglans using genotyping by sequencing data.

3.2. Phylogenetic analysis using chloroplast genome

The complete plastome sequences of 34 Juglans individuals plus J. nigra and Carya sinensis were aligned; their length averaged 160,350 bp, containing 1,849 polymorphic sites, 1,293 singleton variable sites, and 556 parsimony informative sites (PICs, 0.41%) were detected across five Chinese Juglans plus J. nigra and Carya sinensis. Among the 34 Juglans individuals and the outgroup Carya sinensis we detected 18
chloroplast haplotypes based on complete genome sequences (Fig. 1); two haplotypes were unique to J. hopeiensis (Fig. S1). The haplotype diversity (hD) and nucleotide diversity (π) for all samples were 0.867 and 0.0025, respectively. The 18 haplotypes were highly differentiated among species, but haplotype richness was low across the geographic range within species. Chinese walnut (J. cathayensis) contained the highest haplotype diversity. Each individual we sequenced contained a unique chloroplast haplotype (H1, H2, H3, H4, H5, and H6), while J. mandshurica had three haplotypes (H9, H10, and H11) (Fig. 1). The statistical parsimony network of 18 chloroplast genome haplotypes showed that two samples of Ma walnut (J. hopeiensis) shared haplotype H9 with J. mandshurica, while other samples of J. hopeiensis contained haplotypes H8 and H7, which were not found in any sample of J. mandshurica (Fig. 1). Common walnut (J. regia), the most widespread species, only had four haplotypes (H12, H13, H14, and H16) among 11 individuals, while four samples of J. sigillata shared haplotype H12 with five J. regia individuals (XJ, XJH, GS, BJ, and GZ). One J. sigillata sample (LJ) contained a private haplotype (H15) (Fig. 1).

The minimum spanning network of 16 chloroplast genome haplotypes (omitting J. hopeiensis haplotypes) clearly separated the sect. Juglans/Dioscaryon clade and the sect. Cardiocaryon clade, while J. regia and J. sigillata shared haplotype H12 (Fig. S1). The chloroplast-based phylogeny resolved lineage relationships with high statistical support (> 95%), indicating that there were four clades, one associated with J. regia and J. sigillata (sect. Juglans/Dioscaryon clade), one associated with J. nigra (black walnut clade as outgroup), one associated with outgroup Carya sinensis, and one associated with J. mandshurica and J. cathayensis (sect. Cardiocaryon clade).

The phylogenetic relationships among the species based on chloroplast sequences were depicted with a divergence time tree (Fig. 2a; Fig. S2). A point estimate for the coalescent time among three Juglans chloroplast genome clades was dated to 44.98 Mya/44.99 Mya, while J. regia and J. sigillata diverged much more recently (0.25 Mya/0.85Mya), and J. cathayensis diverged from J. mandshurica before 13.46 Mya/15.68 Mya (95% HPD: 5.12–26.52 Mya, Fig. 2a; Fig. S2). The tree topology are strong similar using Maximum likelihood (ML) and Bayesian inference (BI) methods (Fig. 2a, Fig. S2). However, the divergence time estimates between sect. Juglans/Dioscaryon and sect. Cardiocaryon chloroplasts differed considerably depending on whether estimates of divergence was anchored by the separation of Juglans from Castanea mollissima versus the separation of Juglans from Carya sinensis. In the former case, divergence times between species of two sections were estimated at 0.84 Mya (Fig. 2b) versus 36.15 Mya when the anchor was C. sinensis (Fig. 2c), while the divergence between the J. regia and J. sigillata was estimated to have occurred 0.02 Mya (Fig. 2b) versus 0.22 Mya (Fig. 2c). J. cathayensis diverged from J. mandshurica 0.17 Mya (C. mollissima anchor, Fig. 2b) or 23.93 Mya (C. sinensis, Fig. 2c) based on whole chloroplast genome sequences.

3.3. Population and phylogenomic analysis by Genotyping-By-Sequencing (GBS) based on 45,548 SNPs

The (nuclear) genome-wide phylogeny clearly resolves the two walnut sections (Dioscaryon and Cardiocaryon) (Fig. 3; Figs. S3, S4). All phylogenies are congruent and significant and similar to PCA and FRAPPE output (Fig. 3; Figs. S3, S4). Assignment of all individuals from all geographic localities to genetic clusters using FRAPPE revealed an optimum K = 4 genetic clusters (Fig. S3). When K = 4 genetic clusters (Fig. 3a), J. cathayensis and J. mandshurica samples were joined into a single cluster, with J. hopeiensis clearly admixed. Common walnut (J. regia) divided into two genetic sub-clusters: one major cluster contained samples from all over China, and an additional (yellow) cluster dominated by samples from Northwestern China (Fig. 3b). At K = 2, iron walnut (J. sigillata) samples were joined with J. regia, but at K > 2 they became a coherent genetic group with signs of admixture with J. regia (Fig. 3a; Fig. S3). The phylogeny shows iron walnut (J. sigillata) samples appear completely embedded within common walnut (J. regia) (Fig. S4). When samples within sect. Dioscaryon were analyzed separately, they differentiated into K = 2 genetic clusters (Fig. 4a). Common walnut (J. regia) and iron walnut J. sigillata samples divided into distinct groups, with some individuals of mixed ancestry, probably largely from gene introgression (Fig. 4a). J. sigillata emerges as a distinct lineage within J. regia in an ML tree based on ~100,000 nuclear SNPs (Fig. 4c). The PcoA separated the two Dioscaryon species, although J. sigillata and J. regia exhibited considerable overlap (Fig. 4c). The three major PCs from the PcoA explained 18.58% (8.53%, 6.26% and 3.79%, respectively) of the total variance.

Samples of the endangered Ma walnut (J. hopeiensis) did not form a coherent genetic cluster at any K < 5 and all remained a nearly 50%/50% admixture between Dioscaryon and Cardiocaryon (Fig. S3). At K = 5, J. hopeiensis samples appeared to be J. cathayensis/J. mandshurica × J. regia hybrids. FRAPPE analysis of the J. regia, J. hopeiensis, and J. mandshurica (n = 4) samples (excluding J. sigillata and J. cathayensis) revealed an optimal ΔK for K = 2 genetic clusters (Fig. 4d). When the same three species were analyzed at K = 3, a second population of J. regia emerged and J. hopeiensis remained admixed; strongly indicating that J. hopeiensis is a hybrid species (Fig. 4b, d). The results of PcoA confirm the result from FRAPPE; the cloud of points representing J. hopeiensis is midway between J. mandshurica and J. regia in the space delimited by the three largest PCs (Fig. 3c). In fact, every analytical approach supported the hybrid origin of J. hopeiensis (Fig. 3; Fig. 4c,d; Fig. S3, S4).

Phylogenies based on Maximum Likelihood (ML) were strongly similar (Fig. S4). In those trees, J. cathayensis and J. mandshurica are intermixed, and J. hopeiensis appears as intermediate between J. regia and the other members of sect. Cardiocaryon (Fig. 3d; Fig. S4). J. hopeiensis, which is sympatric with J. mandshurica (Fig. 3b) constituted two genotypic clusters within section Cardiocaryon based on chloroplast data (Fig. S2), but at nuclear SNPs J. hopeiensis samples were admixed with roughly equal contributions from species in sect. Dioscaryon and Cardiocaryon (Figs. 5a, 4c, d, Fig. S4). In phylogenetic trees, J. hopeiensis was between the two branches representing Dioscaryon and Cardiocaryon (Figs. 3d, 4b).

We constructed a phylogenetic tree, based on 45,548 nuclear SNPs by the Bayesian MCMC method (Fig. 5a). To gain an understanding of the time scale of the phylogeny of GBS data, we used 140 individuals of all five Chinese Juglans. The MCMC phylogenetic tree showed that the all samples Juglans could be classified into two lineages: lineage I (J. regia and J. sigillata), lineage II (J. hopeiensis, J. mandshurica, and J. cathayensis). A point estimate for the coalescent time between two sections of Juglans was dated to 15.72 Mya, while J. regia and J. sigillata diverged from 8.85 to 12.39 Mya, and J. cathayensis diverged from J. mandshurica before 8.86 Mya (Fig. 5a).

3.4. Analysis of molecular variance (AMOVA)

AMOVA based on (45,548) nuclear SNPs revealed a clear species–species separation: FST = 0.081, P < 0.001, J. regia-J. sigillata; FST = 0.118, P < 0.001, J. sigillata-J. hopeiensis; FST = 0.043, P < 0.001, J. cathayensis-J. mandshurica; FST = 0.153, P < 0.001, J. hopeiensis-J. mandshurica; FST = 0.213, P < 0.001, J. cathayensis-J. hopeiensis; FST = 0.847, P < 0.001, J. cathayensis-J. regia; FST = 0.296, P < 0.001, J. hopeiensis-J. regia, and FST = 0.068, P < 0.001, the northern versus southern populations of Chinese J. regia (Table S8).

3.5. Phylogenetic analysis using transcriptome data

The average number of genes in each gene family (Table S3), the number of unique gene families (Fig. S5a), and number of genes in unique gene families (Fig. S5b) of the five Chinese walnuts were less than those of Oryza sativa, Glycine max, and other plants with more complete genomic resources. Nevertheless, 6421 orthologous groups
were shared by all 14 species used in our analysis (Fig. S5a), which is comparable to previous studies (Fischer et al., 2011). Among the 6421 core orthologous groups, 190 contained only one ortholog in each species (single copy, Fig. S5b).

The alignments of each of the 6421 orthologous genes were separated into four datasets corresponding to each of the three codon
positions in the CDS and whole CDS were used to estimate phylogeny. The four datasets resulted in four strongly similar topology structure maximum likelihood trees (Fig. S6). Notably, the clades leading to *Juglans* species had 100% bootstrap support values (Fig. 3a; Fig. S6). Ma walnut was not included in this phylogenetic analysis because of its apparent origin from recent hybridization. The remaining four taxa separated into two sections (*Dioscuraryon* and *Cardiocaryon*) sister to *C. mollissima* (Fagaceae). As shown in Fig. S7a, the divergence between section *Dioscuraryon* and section *Cardiocaryon* appears to have occurred ~12.01 million years ago (Mya), during the middle Miocene (based on protein coding sequences) (Fig. 5a), although the timing of their divergence differed considerably if the analysis was based on different codon positions [6.91 Mya, 13.67 Mya, and 33.47 Mya using the first (a), second (b), and third (c) codon positions, respectively (Fig. S7)].

### 3.6. Ecological niche modeling

The MAXENT models had a high predictive power and were highly accurate (AUC = 0.99). For sect. *Cardiocaryon*, the projection of the model over the present bioclimatic conditions showed that *J. mandshurica* has suitable habitats in northeastern China between 33°N and 48°N, while *J. cathayensis* has habitats in southern China between 20°N and 40°N (Fig. 6a, c). Habitats of *J. mandshurica* and *J. cathayensis* overlapped between 33°N and 40°N (Fig. 1a). *J. hopeiensis* is narrowly distributed in northern China in the hilly, mid-elevation area near Beijing and Tianjin, including parts of Hebei province (Lu et al., 1999; Hu et al., 2015), and it is sympatric with *J. mandshurica*, which may explain why most *J. hopeiensis* contain *J. mandshurica* chloroplast types. For sect. *Dioscuraryon*, the ecological niche model (ENM) showed habitat suitability for *J. regia* is between 80°E and 140°E latitude and between 20°N and 44°N longitude. *J. sigillata* is sympatric with *J. regia* and distributed in southern China in Yunnan, Sichuan, Tibet, and Guizhou provinces (Fig. 6d, e).

The reconstructed historical distributions based on climate showed that the predicted range of *J. mandshurica* contracted considerably from the LIG to LGM under the influence of cooler, drier climate, while the ENM predicted it to be climatically suitable. Because we trained ENMs on individual species, this type of range over-prediction indicates that the climatic variables we examined are not the main factors limiting *Juglans* species’ dispersal and gene flow. Inspection of the spatial overlap between ENMs (Fig. 6) revealed that factors other than those described in the ENM maintain parapatry for *J. regia* and *J. mandshurica*, *J. cathayensis* and *J. hopeiensis* (Fig. S8), while *J. hopeiensis* is climatically less restricted than other *Juglans* species (Fig. S8).

### 4. Discussion

Our results provide important insight into the evolution and biogeography of Chinese *Juglans*. Prior studies of Chinese *Juglans* species based on chloroplast sequences and SSRs from many individual trees (Bai et al., 2014, 2016; Wang et al., 2016) or genomic sampling from a few individuals (Wang et al., 2008, 2015; Bai et al., 2016) revealed deep phylogeographic structure associated with major landscape features. Our Cp genome and GBS data is consistent with the same phylogeographic breaks, but we also found that post-Quaternary dispersal of *J. regia* and its interaction with other species and formerly distinct *J. regia* lineages led to the human propagation of novel taxa.

### 4.1. The status of *Juglans* hopeiensis

Genomic data from GBS and orthologous sequences analyzed using ML, BI, NJ, and FRAPPE strongly indicated that *J. hopeiensis* is a hybrid (*J. mandshurica × J. regia*) or (*J. cathayensis × J. regia*) species (Figs. 1, 3, Figs. S3, S4), most likely maintained as such by human selection for nut phenotypes valued in commercial trade. A similar conclusion was suggested based on randomly amplified polymorphic DNA (RAPD) markers, isozyme, and karyotype analysis (Redher, 1940; Wu et al., 1999; Mu et al., 1990), although *J. mandshurica* was generally considered the female parent and *J. regia* the male (Wu et al., 1999; Mu et al., 1990) a conclusion contradicted by our genomic analysis of chloroplasts (Fig. 1). The presence of *J. cathayensis × J. regia* hybrids among *J. hopeiensis* samples was unexpected and to our knowledge has not been previously reported. The origin of *J. hopeiensis* as an intersectional hybrid probably explains its low fertility (Dai et al., 2014). It is not surprising that *J. hopeiensis* occupies sites that are bioclimatically distinct from *J. regia* and *J. mandshurica* (Fig. S8) because other *Juglans* hybrids have been shown to do so (Crystal et al., 2016). The presence of a private chloroplast haplotype within *J. hopeiensis* derived from both *J. cathayensis* and *J. mandshurica* may indicate that distinct lineages of these species were found in the area now occupied by *J. hopeiensis*. These lineages may have been isolated from other conspecifics during glaciations and subject to gene flow from *J. regia* after that species was introduced to northern China by human dispersal (Pollegioni et al.,
Fig. 3. Geographical distribution and population structure of Chinese Juglans. (a) Population structure analysis of genotypes of five Chinese Juglans at (the predicted optimal) $K = 4$ based on FRAPPE software; samples included 140 individuals from 47 locations (Table 1). Analyses were based on 45,548 SNPs with 20 iterations for each $K$ for 500,000 iterations. Each individual is represented by a vertical bar. The most likely number of genetic clusters (optimal $K = 4$) was determined using ADMIXTURE (Alexander et al., 2009). (b) Geographical distribution of sample sites of five Chinese Juglans; colors of four clusters correspond of those in the legend in the right corner of the map, for each population, percent membership in each of the four genetic clusters is indicated with a pie chart. (c) Principal coordinate analyses (PCoA) based on GBS SNPs. (d) Maximum Likelihood (ML) tree (unrooted) based on nuclear SNPs. Colored branches represent the non-admixed individuals within each corresponding group, colors are as indicated in (c). All major nodes have 100% bootstrap support.
Our data show the chloroplast haplotypes in *J. hopeiensis* were derived from existing haplotypes H9 (*J. mandshurica*) about 1 Mya, and H7 (*J. cathayensis*) about 1.3 Mya (Fig. S2). The recent derivation of *J. hopeiensis* (timed with the arrival of *J. regia* and subsequent hybridization of *J. regia* with local sect. *Cardiocaryon* species), and its derivation from both *Cardiocaryon* species argue that *J. hopeiensis* should be considered a horticultural variety, rather than a species. *J. hopeiensis* does not occupy an ecological niche that strongly separates it from other *Juglans* species, which could potentially explain the gene flow between geographically adjacent populations of the two sections (Fig. S8). Future studies assessing the fine-scale ecological characteristics of *Juglans* species in relation to genetic patterns could clarify the role of environmental factors in limiting gene flow among populations in different sections of *Juglans*.

4.2. The status of *Juglans sigillata*

Our data provide strong evidence that *J. sigillata* is a sub-species or, perhaps, a landrace of *J. regia*. *Juglans sigillata* was first proposed as a species by Dode (1906), who distinguished it from *J. regia* (as well as *J. fallax*, *J. orientalis*, and *J. sinensis* which Rehder later identified as *J. mandshurica × J. regia*, a generally accepted conclusion) primarily based on characteristics of the nut. Of all the members of sect. *Dioscaryon* described in detail by Dode, only *J. regia* and *J. sigillata* remain widely accepted (Flora of China, 1999). Whether *J. sigillata* is distinct from *J. regia*, or an ecotype has been controversial. Manning (1978) lumped them together based on morphology, Grimshaw (2003) did not. Gunn et al. (2010) concluded that the species were morphologically distinct, but not separable based on 14 microsatellites. Wang et al., 2015 agreed the species are morphologically distinct, but also found about 8% of total variance at 12 SSR loci was between species, and based on Bayesian analysis of genetic structure concluded the species were genetically distinct as well. Our data showed that *J. sigillata* contained at least one chloroplast haplotype not shared by *J. regia* (H15, Table 1), but that this haplotype was derived from a *J. regia* haplotype and its estimated divergence time from *J. regia* was only 0.25 Mya/0.02 Mya/0.22 Mya (Fig. 2). Unlike Wang et al. (2015), our genetic structure analysis was based on (45,548) SNPs, but like Wang et al. (2015), we found that the two species were clearly separated with some admixture (Figs. 3, 4a,b, Fig. S3 at all K ≥ 3). Structure analysis is not designed to identify species, however, but differences among populations. A maximum likelihood phylogeny based on 45,548 SNPs...
Fig. 5. Phylogenetic timetrees of Chinese Juglans based on genotype by sequencing (GBS) data and transcriptome CDS (protein coding sequences). (a) Phylogenetic timetree of 140 Juglans individuals based on 45,548 SNPs constructed using the Bayesian MCMC method. Divergence of sect. Dioscaryon and sect. Cardiocaryon was estimated at 15 ± 5 Mya (Manchester and Garden, 1987; Bai et al., 2016). Divergence times (Mya) are shown at each node. (b) Phylogenetic timetree based on CDS. The purple bars at the nodes indicate 95% posterior probability intervals. Divergence of Castanea mollissima and Juglans was estimated at 64.4 ± 0.5 Mya (Manchester and Dilcher, 1997; Zhang et al., 2013). The geological time scale is in millions of years. Paleoc, Paleocene; Plioc, Pliocen; Q, Quaternary.
showed all *J. sigillata* samples were embedded within *J. regia*. The same result was obtained using ML based on whole genome data (Fig. S4). Divergence time estimates based on protein coding sequences and GBS SNPs showed *J. sigillata* and *J. regia* separated about the same time as *J. mandshurica* and *J. cathayensis* (Fig. 5a, b; Table S10), but in sect. Dioscaryon the chloroplast data show *J. sigillata* to be a genetic subset of *J. regia*. We found no clear evidence that *J. sigillata* was a lineage independent of *J. regia*. *Juglans sigillata* appears to be maintained as a distinct landrace, possibly by ecological isolation (Wang et al., 2015), although we found that the bioclimatic envelopes of the taxa overlap (Figs. 6, 7a). It is more likely *J. sigillata* is kept distinct from *J. regia* by human selection (Gunn et al., 2010).

### 4.3. *Juglans regia* in China

In common walnut (*J. regia*), the SNP-based phylogeny and analysis of genetic structure showed two subpopulations in China, with a clear geographic genetic break corresponding to divergence between the Xinjiang Province and other regions (Figs. 3, 4; Pollegioni et al., 2015). This north/south divide in *J. regia* may reflect a broader divide in east Asian phylogeography (Bai et al., 2016), the general genetic structure of *J. regia* in Asia (Pollegioni et al., 2014), or distinct introductions of *J. regia* into China. Although the origin of *J. regia* is obscure (Martínez-García et al., 2016), historical biogeography and presumed locations of Quaternary glacial refugia led Aradhya et al. (2007) to suggest the species had multiple centers of origin. None of the proposed refugia are...
in central or southern China, however, but instead in Tibet and possibly Xinjiang, so the genetic structure of *J. regia* in China has probably been determined by (relatively recent) human selection and dispersal (Pollegioni et al., 2015). Common walnuts from far-western China (Xinjiang province) were likely the ancestors of populations distributed in central and northern China (Fig. 1; 3b), and *J. regia* from southern China (especially in Yunnan province) may represent a separate source population for *J. regia* in China (Fig. 1, Fig. S4; Pollegioni et al., 2015).

Confi rming the origins and history of *J. regia* in Asia represents an important goal, especially for regions where cultivation and selection or isolation and adaptation may have been potent drivers of genetic change. Research of this type is a particular challenge without a publicly available, high-quality genome assembly. Once such a resource is developed, the loci underlying population divergence between closely relative species and within species can be identified (Schlötterer et al., 2014).

### 4.4. *Juglans cathayensis* and mandshurica

*J. mandshurica* and *J. cathayensis* are considered the two Chinese members of sect. *Cardiocaryon*, but some have lumped them into a single species (Lu et al., 1999; Grimshaw, 2003) along with *J. collapsa* Dode, *J. draconis* Dode, and *J. stenocarpa* Maxim. The independence of *J. mandshurica* and *J. cathayensis* has been asserted historically (‘Ye Hu tao’, *J. cathayensis* versus ‘Hu tao qiu’, *J. mandshurica*), in classical taxonomy (Dode, 1909, although Dode was without doubt an unreformed taxonomic “splitter”), and in recent molecular analyses (Aradhya et al., 2007). Our samples of *J. cathayensis* contained a total of...
six chloroplast haplotypes, and each of the six sampled populations was fixed for a single haptype, displaying the pronounced phylogeographic structure previously remarked upon by Bai et al. (2016) (Fig. 1, Table 1). Bai et al. (2016) identified nine haplotypes in J. cathayensis, all distinct from those of J. mandshurica. These results, and ecological niche models, are evidence that J. cathayensis populations survived the latest glaciation in situ, although they were probably more fragmented than today (Fig. 6c; Fig. S8; Bai et al., 2014). The timing of the divergence of the chloroplast lineages common to all J. cathayensis from those of J. mandshurica was probably the late Miocene (Bai et al., 2016), which was confirmed by our analysis (13.45 Mya (Fig. 2a)), but this reconstruction depended heavily on the fossil calibration (0.17 Mya in Fig. 2b and 22.93 Mya in Fig. 2c), a conclusion shared by Bai et al. (2016).

The genetic structure of Cardiocaryon in China at nuclear loci (based on GBS data) indicated J. cathayensis and J. mandshurica were separated at the optimal K = 2. Samples from populations HP, BX and SNJ from Hunan, Sichuan, and Hubei, showed subtle admixture with J. mandshurica when the analysis was expanded to include sect. Dioscaryon (samples at K = 5 and K = 8; Fig. S3). The ML trees (Fig. 3d, Fig. 54), and PCoA (Fig. 3c) also showed samples of J. cathayensis from the same populations (BX, HP, and SNJ) clustered with J. mandshurica. Bai et al. (2014, 2016) found evidence of weak genetic structure within the nuclear genome of J. cathayensis, but it was associated with admixture with J. mandshurica in northern populations, not with samples of BX, HP, or SNJ that were sampled by Bai et al. (2016). They reported an FST of 0.11 among J. cathayensis, considerably higher than our estimate of 0.043 between J. cathayensis and J. mandshurica, although we sampled fewer and less dispersed populations. In any case, nuclear gene flow among J. cathayensis populations is probably sufficient to keep them “well-connected” (Bai et al., 2014), so we have no explanation for the subtle structure we observed in our J. cathayensis samples.

Manchurian walnut (J. mandshurica) is distributed parapatrically with J. cathayensis; their predicted habitats potentially overlap based on our bioclimate models (Fig. 5), but in fact they are adapted to distinct climates [Fig. 6, Fig. S8; see also Bai et al. (2016)]. J. mandshurica is allopatric with J. altantifolia, its closest relative, which is native to Japan. The three Asian butternuts (sect. Cardiocaryon) are closely related, with small morphological differences (Grishmaw, 2003): J. mandshurica has often been characterized as a species with abaxially glabrescent leaflets and a fruiting spike with four or five nuts, whereas J. cathayensis is described as having tomentose leaflets and a flowering spike that typically bears six to ten nuts (Lu et al., 1999; Bai et al., 2016); J. cathayensis nuts were described as larger than those of J. mandshurica, with a more acute apex and a less sharp external roughness (Dode, 1909). The split between the closely related species J. cathayensis and J. mandshurica occurred 0.17 to 23.93 Mya based on chloroplast genomes; the most recent estimate is derived when C. mollissima as the outgroup, the oldest estimate is derived when C. siennis is used (Table S10). The analysis of their divergence time based on nuclear data is nearly as inconclusive; an estimate of 2.0–26.05 Mya is obtained when CDS (orthologous genes), first, second, or third codon positions are analyzed (Table S10). Where J. mandshurica and J. cathayensis are placed in the continuum between ecological races or varieties and species is not obvious from the data (Mallet, 2008). The observation that the two taxa can hybridize and produce fertile progeny is probably irrelevant, as intersectional (and intercontinental) hybrids are often fertile in Juglans (Pollegioni et al., 2009).

We identified three chloroplast haplotypes within J. mandshurica, and perhaps an additional J. mandshurica haplotype private to J. hopeiensis (Fig. 1). This diversity was significant because Bai et al. (2016) considered the northern (i.e., Chinese) populations of J. mandshurica to have undifferentiated chloroplasts and therefore postulated a single refugium for them. It is certainly possible that J. mandshurica had multiple and isolated Chinese refugia (similar to J. cathayensis) but that all were separated from those of J. cathayensis after the species diverged in the late Miocene (Fig. S2, Bai et al., 2016), maintaining the phylogeographic break suggested in Bai et al. (2016). We did not observe any evidence of genetic substructure at nuclear loci within J. mandshurica based on our limited samples, although Bai et al. (2016) suggested a break between two populations at about 125°E longitude that is different than the Korean versus Chinese J. mandshurica genetic pools indicated in chloroplast phylogeny.

The evidence indicates sect. Cardiocaryon divided into J. mandshurica (northern China) and J. cathayensis (southern China) during the Neogene, possibly as late as the Pliocene. Gene introgression by pollen flow kept the species connected. In sect. Dioscaryon, J. regia and J. sigillata divided more recently than the split in sect. Cardiocaryon (Figs. 2, 5, Fig. S7), and it appears human management of both taxa in the section will determine whether J. sigillata remains distinct (Gunn et al. 2010). Notwithstanding the presence of J. hopeiensis, which is probably only viable as a horticultural variety, we found no evidence for substantial gene flow between sect. Cardiocaryon and sect. Dioscaryon in natural populations (Figs. 1, 3; Fig. 7a; Fig. S8), although their distributions overlap in China.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 41471038, 31205001, and J2120063), Changjiang Scholars and Innovative Research Team in University (No. IRT1174), and the Program for Excellent Young Academic Backbone Funding by Northwest University (No. 338050070). 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 accessibility

The sequences reported in this paper were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) and Transcriptome Shotgun Assembly (TSA). Raw paired-end reads are available through the NCBI SRA under accession numbers: SRX1295882, SRX1734262. The complete Cp genome of all five Juglans species were deposited in NCBI GenBank (Hu et al., 2016, 2017; accession numbers, KT963008, KX671976, KX671977, KX671975, and KT963008).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ympev.2018.04.014.

References


Crystal, P.A., Lichten, N.L., Weose, K.E., Jacobs, D.F., 2016. Vegetative and adaptive traits...


