

RESEARCH ARTICLE

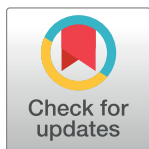
Domestication and selection footprints in Persian walnuts (*Juglans regia*)

Xiang Luo^{1,2,3}, Huijuan Zhou^{2,4,5}, Da Cao^{3,6}, Feng Yan², Pengpeng Chen², Jiangtao Wang², Keith Woeste⁷, Xin Chen⁸, Zhangjun Fei⁹, Hong An¹⁰, Maria Malvolti¹¹, Kai Ma¹², Chaobin Liu⁶, Aziz Ebrahimi⁷, Chengkui Qiao³, Hang Ye², Mengdi Li², Zhenhua Lu³, Jiabao Xu^{13*}, Shangying Cao^{3*}, Peng Zhao^{2*}

1 College of Agriculture, Henan University, Kaifeng, Henan, China, **2** Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education, College of Life Sciences, Northwest University, Xi'an, Shaanxi, China, **3** Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou, China, **4** Xi'an Botanical Garden of Shaanxi Province, Xi'an, China, **5** College of Forestry, Northwest A&F University, Yangling, Shaanxi, China, **6** Laboratory of Functional Plant Biology, Department of Biology, Ghent University, Ghent, Belgium, **7** USDA Forest Service Hardwood Tree Improvement and Regeneration Center (HTIRC), Department of Forestry and Natural Resources, Purdue University, West Lafayette, Indiana, United States of America, **8** Shandong Institute of Pomology, National Germplasm Repository of Walnut and Chestnut, Tai'an, China, **9** Boyce Thompson Institute for Plant Research, US Department of Agriculture (USDA) Robert W. Holley Center for Agriculture and Health, Cornell University, Ithaca, New York, United States of America, **10** Bioinformatics and Analytics Core, University of Missouri, Columbia, Missouri, United States of America, **11** Research Institute on Terrestrial Ecosystems, National Research Council, Porano, Terni, Italy, **12** Xinjiang Academy of Agricultural Sciences, Urumqi, China, **13** BGI Genomics, BGI-Shenzhen, Shenzhen, China

* These authors contributed equally to this work.

* jiabao.xu@bgi.com (JX); s.y.cao@163.com (SC); pengzhao@nwu.edu.cn (PZ)



OPEN ACCESS

Citation: Luo X, Zhou H, Cao D, Yan F, Chen P, Wang J, et al. (2022) Domestication and selection footprints in Persian walnuts (*Juglans regia*). PLoS Genet 18(12): e1010513. <https://doi.org/10.1371/journal.pgen.1010513>

Editor: Li-Jia Qu, Peking University, CHINA

Received: July 20, 2022

Accepted: November 7, 2022

Published: December 7, 2022

Copyright: © 2022 Luo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: DATA AVAILABILITY STATEMENT The Illumina whole genome re-sequence data that support the finding of this study have been deposited in China National GeneBank (CNCB) Nucleotide Sequence Archive database under accession number CNP0001209 (<https://db.cngb.org/search/project/CNP0001209/>). The 'Chandler' walnut reference genome (v2.0) from the NCBI Genbank (https://www.ncbi.nlm.nih.gov/assembly/GCF_001411555.2) under accession PRJNA291087. All other relevant data are contained within the paper and its Additional files.

Abstract

Walnut (*Juglans*) species are economically important hardwood trees cultivated worldwide for both edible nuts and high-quality wood. Broad-scale assessments of species diversity, evolutionary history, and domestication are needed to improve walnut breeding. In this study, we sequenced 309 walnut accessions from around the world, including 55 *Juglans* relatives, 98 wild Persian walnuts (*J. regia*), 70 *J. regia* landraces, and 86 *J. regia* cultivars. The phylogenetic tree indicated that *J. regia* samples (section *Dioscaryon*) were monophyletic within *Juglans*. The core areas of genetic diversity of *J. regia* germplasm were southwestern China and southern Asia near the Qinghai-Tibet Plateau and the Himalayas, and the uplift of the Himalayas was speculated to be the main factor leading to the current population dynamics of Persian walnut. The pattern of genomic variation in terms of nucleotide diversity, linkage disequilibrium, single nucleotide polymorphisms, and insertions/deletions revealed the domestication and selection footprints in Persian walnut. Selective sweep analysis, GWAS, and expression analysis further identified two transcription factors, *JrbHLH* and *JrMYB6*, that influence the thickness of the nut diaphragm as loci under selection during domestication. Our results elucidate the domestication and selection footprints in Persian walnuts and provide a valuable resource for the genomics-assisted breeding of this important crop.

Funding: This research was supported by the National Natural Science Foundation of China (32070372 to PZ and 41471038 to PZ), the National Key Research and Development Project (2018YFD1000604 to SC), Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2015-ZFRI to SC), Natural Science Foundation of Shaanxi Province of China (2019JM-008 to PZ), and Opening Foundation of Key Laboratory of Resource Biology and Biotechnology in Western China (Northwest University), Ministry of Education (ZSK2018009 to PZ). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Walnut (*Juglans*) species are economically important hardwood trees. The Persian walnut as a woody nut crop is limited by an incomplete understanding of the diversity and spatial genetic structure of the species' wild germplasm, and the relationship of the secondary germplasm pool of walnut (wild and landrace walnuts) to improved cultivars. To better understand walnut's breeding, we sequenced the genomes of 309 walnut accessions from around the world. The results revealed core areas of genetic diversity of Persian walnut germplasm were southwestern China and southern Asia near the Qinghai-Tibet Plateau and the Himalayas. The genome-wide scan for selective sweeps, genome-wide association studies (GWAS), and expression analysis identified *JrbHLH* and *JrMYB6* transcript factors related to the thickness of the diaphragm that were under selection during domestication. Our results elucidate the genetic diversity of walnut and its evolution as a species and a crop. We identified centers of wild genetic diversity that could contribute to breeding and sustainable improvement.

Introduction

Persian or common walnut (*Juglans regia* L.; Juglandaceae, Fagales) is a large, wind-pollinated, diploid ($2n = 32$), monoecious, heterodichogamous, long-lived tree. It is grown worldwide [1–3] for its edible nuts and high-quality wood [4–6]. Nearly 500 varieties and cultivars of Persian walnut have been recorded around the world [6–10]. The relationships between the current distribution of genetic diversity, its ancient centres of diversity, and its histories of domestication and spread by humans remain open questions [6–10]. The authors of a recent study suggested that *J. regia* originated as an ancient hybrid between American and Asian walnut lineages in the late Pliocene (3.45 million years ago, Mya) [11]. It is generally accepted that after the last glaciation, *J. regia* survived and grew in isolated stands in the foothills of the Western Himalayas from the Kashmir region to Tajikistan and Kyrgyzstan [9–13] and in South Asia [9]. Persian walnut was also a part of the ancient Chinese flora [14]. C-dated leaf fossils and carbonized nuts found in Shandong and Hebei provinces were shown to be ca.

$7,335 \pm 100$ years old [12,14]. Some researchers have suggested that the presence of additional centres of diversity of walnuts were in Eurasia and even southern Europe [6,7,10,12,15]. Whatever the distribution of walnut before its human-mediated dispersal, the current species range is the product of long-term and complex interactions between biogeography, climate, and human forces [10,12,16–17]. Indeed, the biogeography of *J. regia* is complicated by its long history of human cultivation and its use as a traditional nut crop in many Old World cultures. Because of its ancient association with humans and the movement of walnuts via regional- and even continental-scale trade routes, apparently wild or (possibly) feral populations of walnuts are now found in isolated favourable habitats from China to the mountains of the Iberian Peninsula [9,10,12]. Population genetics and comparative genomics help the footprints of domestication and migration, which have been applied successfully to both annual crops (e.g., rice, soybean, and cotton) [18–21] and perennials (e.g., apple, pear, peach, and chestnut) [22–25].

Programs to produce improved, clonal varieties of Persian walnut began in the mid-20th century, so modern cultivars have short pedigrees, although the selection and dispersal of walnuts by humans have been occurring in Asia and Europe for millennia [7,10,12,16]. Traits considered desirable for modern international markets include a light kernel colour and a tight shell seal. It is still unclear which other traits may have been selected by humans in diverse and dispersed villages across the species' range. The walnut landraces propagated by local

communities as seedlings represent a valuable resource as a secondary germplasm pool for breeders. Theoretically, germplasm pools propagated as seedlings are more genetically diverse than those created with controlled crosses within a pedigree [5,26–27]. Landraces might also contain linkage blocks of genes for adaptations to biotic and abiotic stresses, particularly in the source environment [28–30]. The value of these landraces is compounded by their preselection for traits desired by human consumers and cultivators. The secondary germplasm pool for walnut is not well characterized; although some wild or apparently wild germplasm has been collected [31], most of the resources are still maintained *in situ* in remote locations, and the relationship between landrace samples and already characterized populations is unclear. Wild or landrace walnut populations may be descendants of truly autochthonous local sources, feral trees, descendants of trees dispersed from distant locations by humans, or a combination of these sources. The origins of walnut populations and the effects of human selection during domestication may be revealed by a detailed genetic comparison of wild, landrace, and cultivar populations. The availability of single nucleotide polymorphism (SNP) markers for rapid and highly automated genotyping makes them ideal for genetic diversity studies, genomic selection analysis, and genome-wide association studies (GWAS).

Here, we sequenced 309 diverse walnut accessions from around the world, including 55 *Juglans* relatives, 98 (presumed) wild *J. regia* accessions, 70 *J. regia* landraces, and 86 *J. regia* cultivars. We provide genomic evidence for elucidating the intraspecific genetic relationships of walnut source populations, the historical dynamics of *J. regia* cultivation, and candidate loci under selection during Persian walnut domestication.

Results

Genetic variation of walnut accessions

To understand the diversity of Persian walnut, we sequenced 309 *Juglans* accessions, including wild walnuts, landrace walnuts, cultivars, and related species (S1 Table). We evaluated a total of 4,643 Gb of high-quality, cleaned sequence data, with an average of 15.03 Gb per accession, which is equivalent to an approximately 29× coverage of the ~ 568 Mb walnut genome [5,26]. These sequences were aligned to the genome of the Persian walnut ‘Chandler’ [5,26], with an average mapping rate of 98.42% (S1 Table). We identified 25,063,036 single nucleotide polymorphisms (SNPs) and 3,845 insertions and deletions (InDels) at the genome level (S1C and S1D Fig).

Genetic diversity and linkage disequilibrium in the genomes of walnuts

The nucleotide diversity (π) of walnut at the whole-genome level across all walnut accessions was 4.5×10^{-3} (Table 1). The patterns of differences in diversity (π) and SNP and InDel frequencies revealed genetic differentiation between wild walnuts and domesticated walnuts (cultivars and landraces) (S1C, S1D, and S1F Fig). Wild walnut genomes contained notably higher nucleotide diversity (4.56×10^{-3}) than landrace walnuts (4.44×10^{-3}) and cultivated walnut (4.39×10^{-3}) genomes (Table 1). The genome-level comparisons showed that the population differentiation (F index, F_{ST}) between wild walnuts and landrace walnuts was similar to that

Table 1. Nucleotide diversity and linkage disequilibrium (LD) decays in wild walnuts, walnut landraces, and walnut cultivars.

Groups	Number of samples	SNP (N)	SNP/kb	$\theta\pi$	Tajima's D	r^2	Half distance (Kb)
Wild walnuts	98	8,387,336	14.39	4.56×10^{-3}	2.29	0.2926	3.8
Landraces	70	8,367,894	14.40	4.44×10^{-3}	2.30	0.2894	7.6
Cultivars	86	8,307,806	14.47	4.39×10^{-3}	2.19	0.2925	7.4

<https://doi.org/10.1371/journal.pgen.1010513.t001>

between wild walnuts and cultivars, but there was a relatively low F_{ST} value between landraces and cultivars based on SNPs and InDels, indicating that modern breeding has not substantially influenced the genome of walnuts beyond the initial effects of domestication.

The walnut populations showed relatively short linkage disequilibrium (LD) distances and relatively rapid LD decay (Fig 1D). The LD levels of the population were relatively low, with an average r^2 value of 0.225 (S2 Table). The average distance over which LD decayed to ~50% of its maximum value was much shorter (3.8 kb) in wild walnuts than in landrace walnuts (7.6 kb) and cultivated walnuts (7.4 kb) (S2 Table). Cultivars and landraces were similar in terms of rates of LD decay (Fig 1D).

The results that support the distinctiveness of the wild samples of *J. regia* compared to the cultivated and landrace samples include differences in the π value and LD distance, but similar levels between cultivated and landrace walnuts (Fig 1C and 1D; Table 1). This provided evidence of diversity loss associated with domestication but a relatively weak additional loss of diversity due to cultivar selection.

Phylogenetic relationships and population structure of all walnut accessions

To explore the relationships between various walnut accessions, we constructed a phylogenetic tree of all 309 accessions based on ~7.8 M SNPs. The tree revealed the presence of three distinct groups for all 309 accessions, namely, other Juglandaceae species (wild relatives of *J. regia*, including 55 samples), mainly wild *J. regia* (98 samples), and cultivar and landrace walnuts (156 samples) (Fig 1B).

To further reveal the relationships of 254 *J. regia* accessions, we performed phylogenetic analysis, principal component analysis (PCA), and population structure analysis (Fig 2). As a result, cultivars and landraces (local varieties) were identified as belonging to a common gene pool (Fig 2A and 2C). We speculate that these individuals represent the mixing of germplasm pools as a consequence of human-mediated dispersal over large distances, possibly as part of commercial activities. The modern cultivars that we sampled are representatives of breeding programs that have drawn on a diverse base of landraces preselected for traits by local communities in geographically distinct regions. The gene flow analysis showed strong gene flow ($N_m = 61.33$) between landraces and cultivars by the BABA-ABBA test and N_m calculation (Fig 2D). The gene flow of wild comparing to landrace was 2.71, while gene flow was 2.90 between wilds and cultivars (Fig 2D). We also detected the same gene flow patterns by the TreeMix analysis (S2 Fig).

Origin and differentiation of Persian walnuts

To reveal the origin and differentiation of Persian walnuts, we focused on only wild genotypes ($N = 98$) and excluded all cultivated and landrace walnut accessions. The filtered SNP dataset of wild walnut accessions (8,387,336 SNPs) was subjected to phylogenetic analysis, population structure analysis (Fig 3A), and PCA (S3 Fig). The ΔK values revealed an optimal K of four sub-populations (S4 Fig), which was similar to the pattern observed in our PCA score plot (S3 Fig). The first sub-populations consisted of East Asian (EA) walnuts representing a group of samples from central and northern China. The second group contained samples from the Yunnan-Kweichow Plateau, southwestern China (YG: Yunnan and Guizhou provinces). The third group included Xinjiang walnuts represented by eight wild trees from Gongliu County in Xinjiang Province, western China (XJ: Gongliu County, Xinjiang Province), which is an apple diversity centre [22], and the fourth group consisted of South Asian (SA: Pakistan, India, and Nepal), western Asian (WA), and European (EU) walnuts. Among wild walnuts, the second

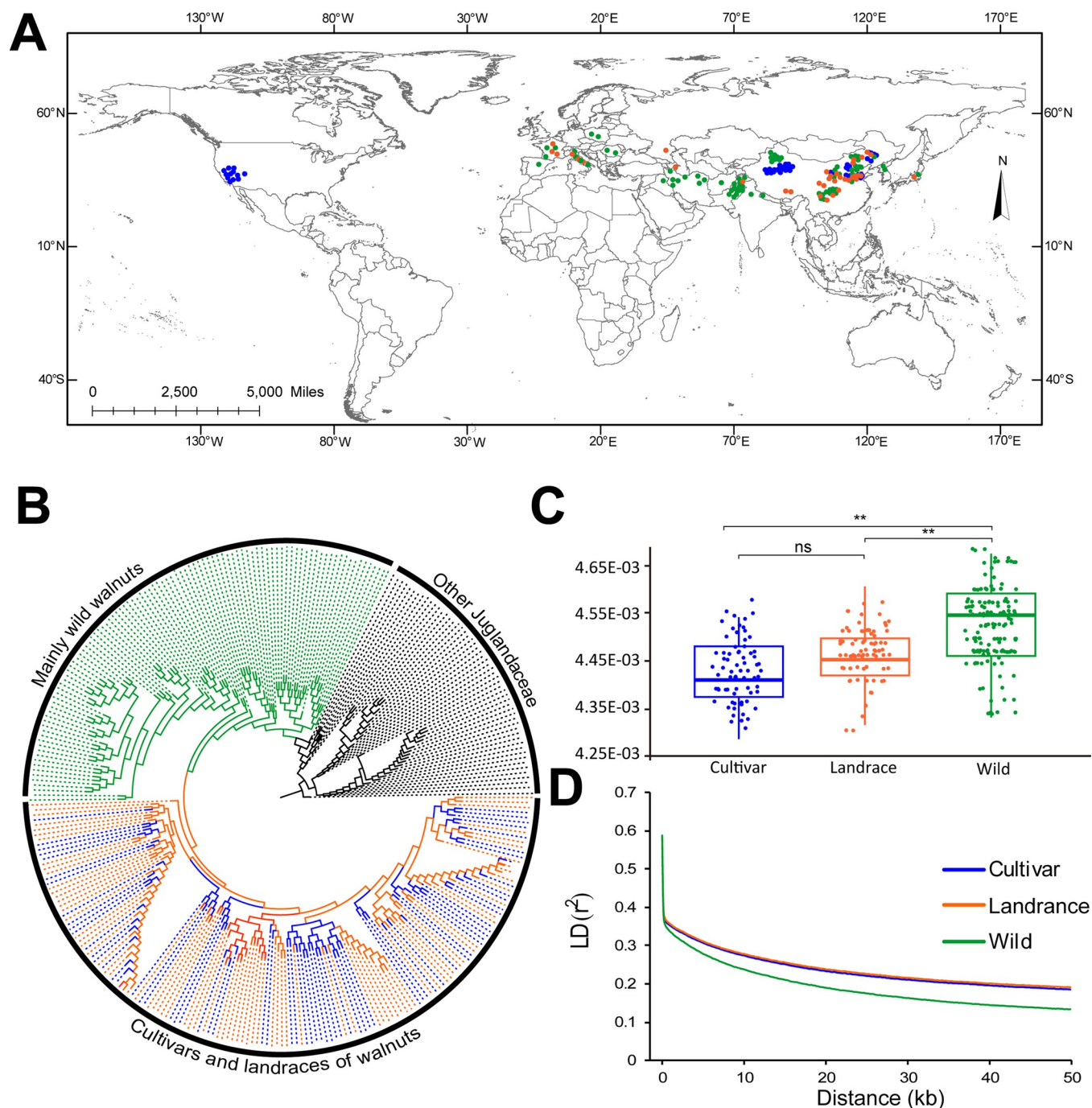


Fig 1. Phylogeny and population genetics of cultivated walnuts and uncultivated walnuts. (A) 309 walnut accessions distribution in different geographical regions. Green: wild; orange: landraces; blue: cultivars; and black: related *Juglans* species. The basemap and distribution of all sampling points were plotted in ArcGIS v10.2 (ESRI Inc., Redlands, California, USA). The world map was downloaded from the Natural Earth (<https://www.naturalearthdata.com/downloads/10m-cultural-vectors>). The grey line indicates the country region border shape in the map. (B) A neighbor-joining tree of 309 walnut accessions (C) Summary of nucleotide diversity (π) across wild walnuts, landraces, and cultivars. (D) Decay of linkage disequilibrium (LD) in different walnut groups.

<https://doi.org/10.1371/journal.pgen.1010513.g001>

group (YG) had higher nucleotide diversity (6.88×10^{-3}) than the first group (EA, 4.01×10^{-3}) and the fourth group (SA, 5.76×10^{-3} ; WA: Iran, Afghanistan, and Iraq, 4.91×10^{-3} ; EU, 4.39×10^{-3}). The lowest nucleotide diversity was observed among the wild

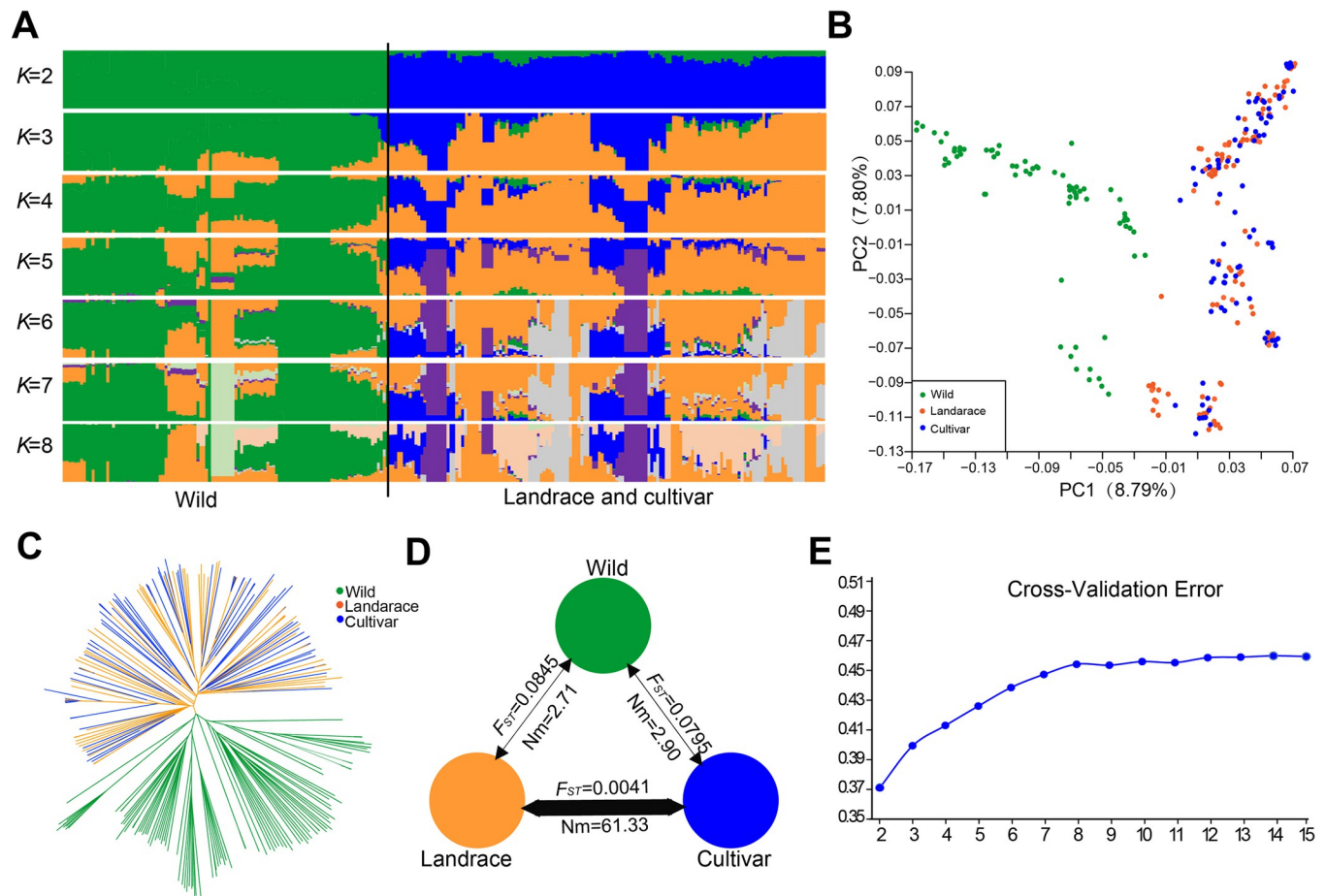


Fig 2. Population analysis of 254 Persian walnuts (*Juglans regia*). (A) Population structure analysis of ancestry components for 254 Persian walnuts (*Juglans regia*). Each color corresponds to a single population as noted. Each accession is represented by a vertical bar. The Y-axis refers to the proportion of the genetic background, and the height of each line with different colors represents the probability of an accession belonging to a different genetic background. (B) PCA of 254 Persian walnuts (*Juglans regia*). The first two principal components were accounting for 16.59% of variations. (C) A neighbor-joining tree of 254 walnut accessions. (D) D statistics for comparisons among Persian walnut groups. The black arrow line indicates the Z-score between the two groups. The thickness of the line indicates the value of Nm. (E) Estimated Delta K of possible clusters (k) from 2 to 15.

<https://doi.org/10.1371/journal.pgen.1010513.g002>

walnuts in the third group (XJ, 2.48×10^{-3}) (Fig 3B and Table 2). Our data indicated that southwestern China and southern Asia near the Qinghai-Tibet Plateau and Himalayas represent core areas of genetic diversity of Persian walnut.

The results of the pairwise sequentially Markovian coalescent (PSMC) model indicated that the Persian walnut demes shared strongly similar trajectories of effective population size (N_e) before 1.0 Mya (million years ago), but the shapes of the N_e curves diverged c. 0.4–0.5 Mya (Fig 3D). Most lineages' N_e started to increase from c. 40 to c. 50 Mya and decreased quickly beginning at c. 20 Mya, while walnuts from Qinling mountains (Qi) and Guizhou Province (Gu) had different N_e curves during c. 10 Mya to c. 5 Mya (Fig 3D). During the period from 20.0 to 1.0 Mya (early Pleistocene), the similarity of the N_e dynamics of all Persian walnuts was pronounced; most walnuts reached their largest population size ($N_e \sim 1.5\text{--}2.0 \times 10^4$) c. 10.0 Mya. During the Quaternary glacial (1–2 Mya) in Asia, all of the Persian walnut populations that we studied showed expected glacially induced decline (Fig 3D).

High levels of gene introgression were detected between EA and YG, between WA and SA, and between WA and EU (Fig 3C; S3 Table), a result supported by the STRUCTURE results

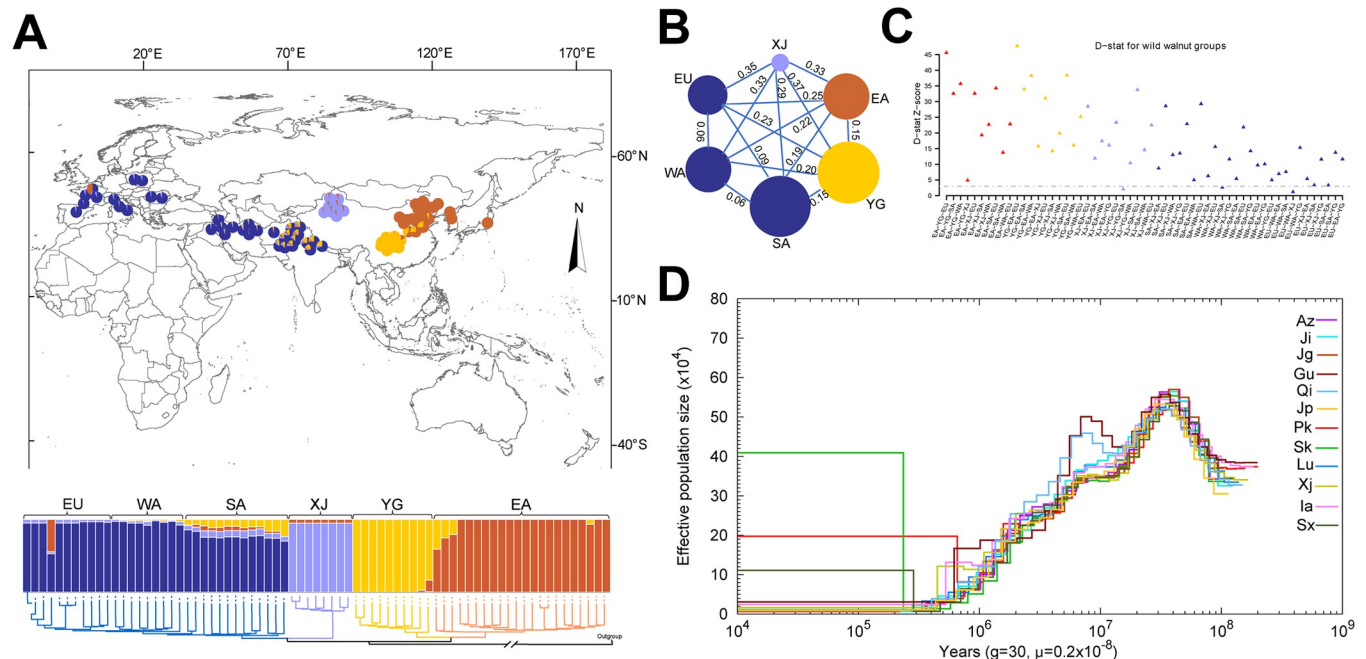


Fig 3. Phylogeny, population structure, diversity, and dynamic history of wild walnuts. (A) Geographic distribution, population structure, and phylogenetic tree analysis of wild Persian walnut accessions plus outgroups. Each of the six regional walnut populations are indicated by different colors. EA: Eastern Asia; YG: Yunnan-Kweichow Plateau; XJ: Xinjiang province; SA: southern Asia; WA: western Asia; EU: Europe. The basemap and distribution of all sampling points were plotted in ArcGIS v10.2 (ESRI Inc., Redlands, California, USA). The world map was downloaded from the Natural Earth (<https://www.naturalearthdata.com/downloads/10m-cultural-vectors>). The grey line indicates the country region border shape in the map. (B) Summary of nucleotide diversity ($\theta\pi$) and population divergence (F_{ST}) across groups. The size of each circle represents the nucleotide diversity ($\theta\pi$) for the group, and values on the line between pairs indicate the population divergence (F_{ST}). (C) Comparisons among wild Persian walnut groups (EA, YG, XJ, SA, WA, and EU) based on D statistics. The grey dotted line indicates where a Z-score is more than 3. (D) Pairwise sequentially Markovian coalescent (PSMC) estimates of the changes in the effective population size over time for 12 individuals of Persian walnut. Each line represents one individual. $10^6 = 1.0$ Ma; Persian walnut Ne curves converged 1.0 million years ago, probably reflecting the time when all sources of *J. regia* last shared a common ancestor. "Az"—Landrace from Azerbaijan; Ji—Cultivar Jiling province, China; Jg—cultivar from Beijing, China; Gu—wild walnut from Guizhou province, China; Qi—wild walnut from Qinling mountains, China; Jp—landrace from Japan; Pk—wild walnut from Pakistan; Sk—landrace from South Korea; Lu—cultivar from Shandong province, China; Xj—cultivar from Xinjiang province, China; Ia—wild walnut from Iran; Sx—wild walnut from Shaanxi province, China.

<https://doi.org/10.1371/journal.pgen.1010513.g003>

(Figs 3A and S3). The XJ population showed the lowest gene flow with other walnut populations (Fig 3C; S3 Table). We detected the highest genetic differentiation ($F_{ST} = 0.33$) value between XJ samples and samples from other populations; the lowest F_{ST} values were found between WA and SA and between WA and EU (Fig 3B). The PCA results showed that the XJ

Table 2. Genetic diversity and Tajima's D in wild Persian walnuts.

Walnut accessions- groups	Number of samples	SNP number	SNP/kb	θ_w	$\theta\pi$	Tajima's D
Eastern Asian (EA)	39	10,912,185	19.01	4.62×10^3	4.01×10^3	-0.56
Yungui (YG)	14	13,559,894	23.62	6.13×10^3	6.88×10^3	1.54
Xinjiang (XJ)	8	3,291,229	5.73	1.74×10^3	2.48×10^3	0.47
SA (Southern Asian)	17	13,282,721	23.14	6.10×10^3	5.76×10^3	-0.28
WA (Western Asian)	9	8,991,135	15.66	4.58×10^3	4.91×10^3	0.26
EU (Europe)	11	8,223,014	14.33	3.95×10^3	4.39×10^3	0.42

Note: EA: Eastern Asian; YG: Yunnan-Kweichow Plateau; XJ: Xinjiang province; SA: southern Asian; WA: western Asia; EU: Europe. SNP: single nucleotide polymorphism was comparison with 'Chandler'; SNP/kb: The numbers of SNP per kilobasepairs; θ_w —the Watterson's estimator; $\theta\pi$ —the average pairwise divergence within a population

<https://doi.org/10.1371/journal.pgen.1010513.t002>

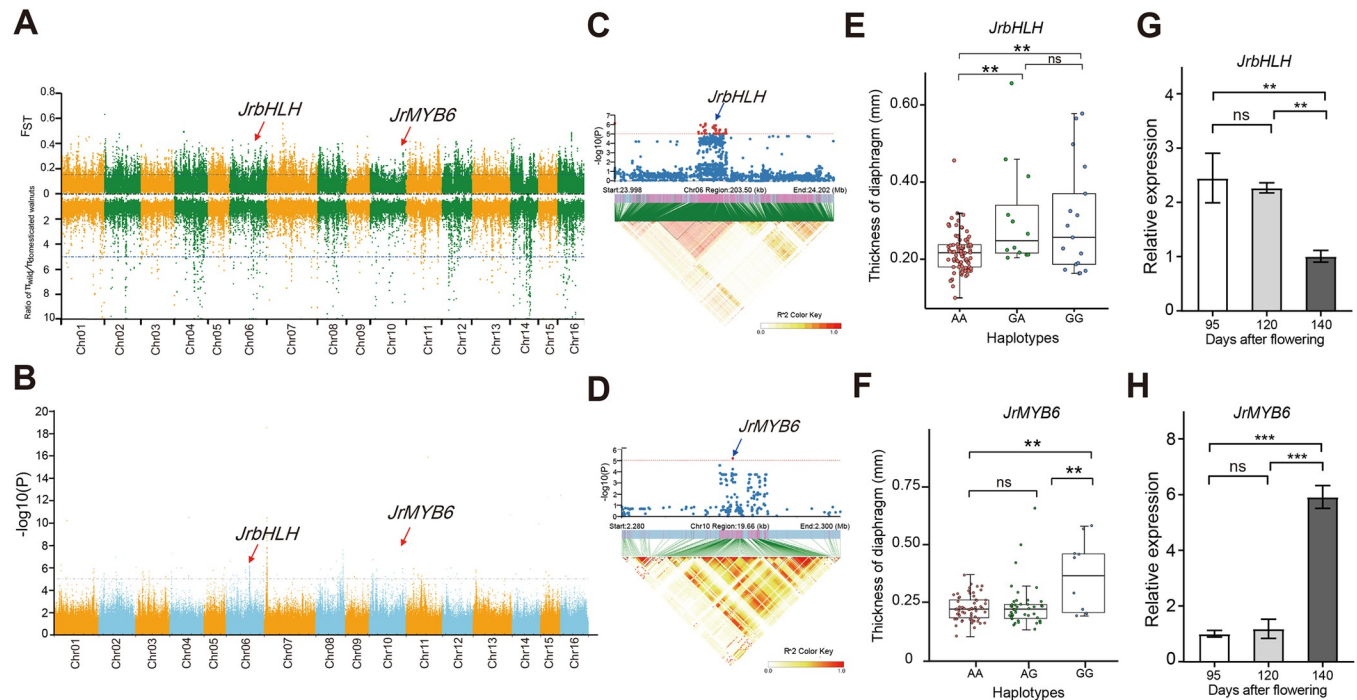


Fig 4. Selective sweeps and GWAS of diaphragm thickness. (A) Distribution of the F_{ST} and nucleotide diversity (ROD) for the wild group and cultivar +landrace group ($\pi_{wild}/\pi_{cultivar+landrace}$; domesticated) in windows along chromosomes. Horizontal dotted lines represent the cutoff fulfilling the requirement for regions under selection. The dashed line indicates the $F_{ST} = 0.15$. The red arrow indicates the selected loci in *JrbHLH* and *JrMYB6*. (B) Manhattan plot for GWAS on the thickness of the diaphragm in the full population. The dashed line indicates the threshold $-\log(P) = 5$. The red arrow indicates the SNP in *JrbHLH* and *JrMYB6*. Manhattan plot (top) and linkage disequilibrium heat map (bottom) for *JrbHLH* (C) and *JrMYB6* (D). Box plots show the thickness of the diaphragm in three haplotypes (Hap.) of *JrbHLH* (E) and *JrMYB6* (F). Relative expression levels of TF *JrbHLH* (G) and *JrMYB6* (H) in three development stages of the walnut diaphragm. ** $P < 0.01$, *** $P < 0.001$, ns = not significant, Student's test-test.

<https://doi.org/10.1371/journal.pgen.1010513.g004>

population was surprisingly distinct from all other populations; the distinctiveness of the XJ samples was also evident in the STRUCTURE results at $K = 3$ to $K = 6$ (S4 Fig). The sample number of wild trees from XJ was small because wild trees are difficult to access there; conclusions based on this small sample must be considered tentative. The population structure analysis showed that the samples from South Asia (SA) exhibited considerable admixture from all other sources (east and west) (Figs 3A and S4). It is possible that a more comprehensive sampling of remote locations will reveal the presence of genetic diversity not shared by other sources. The European and western Asian populations displayed low levels of gene introgression from SA and YG (Figs 3A and S4). Walnut diversity in Eurasia is globally governed by the separation between populations in eastern and southwestern China (EA and YG) and all other populations.

Genome-wide selection and GWAS of genes during domestication

As in most fruit tree species, an important step in the domestication of wild walnut into a nut crop must have involved the (partial) adaptation to cultivation and local environments through human selection. Genome-wide selection analysis indicated that approximately 3% (16.4 Mb) of the walnut genome containing 866 genes were shaped by selection based on a comparison of wild versus cultivar+landrace groups (Fig 4A; S4 Table). Among the genes under selection, *JrbHLH* was located 119 bp downstream of the associated SNP (Loc: 24,101,955) on chromosome 6, and *JrMYB6* was located 951 bp upstream of the associated

SNP (Loci: 2,290,478) on chromosome 10 based on GWAS for the thickness of the nut diaphragm (Fig 4B, 4C and 4D). These two SNPs could successfully distinguish the A and G alleles in the association panel with the thickness of the nut diaphragm (Fig 4E and 4F). The nut diaphragm thickness of the lines with the G allele were significantly higher than in those with the A allele ($P < 0.01$). Sequence comparison analysis indicated that the *JrbHLH* was homologs to that of *Arabidopsis thaliana*, *Glycine max*, *Populus trichocarpa*, and *Malus domestica* (S5 Fig). While *JrMYB6* of *J. regia* was homologs to that of *Zea mays*, *Arabidopsis thaliana*, *Solanum lycopersicum*, *Malus domestica*, *Oryza sativa*, *Glycine max*, and *Populus trichocarpa* (S6 Fig). The Expression analysis further denoted that *JrbHLH* was expressed decreased during walnut diaphragm development (July 27, August 18, and September 6), while *JrMYB6* was expressed increased during three walnut diaphragm developmental stages (Fig 4G and 4H). Thus, these two genes are selected and may play important roles during the development of the walnut diaphragm.

Discussion

Here, we sequenced the genomes of 309 walnut (*Juglans*) samples collected around the world, with a focus on *J. regia*, the most valuable nut crop in China. The study revealed that the core areas of genetic diversity of Persian walnut germplasm were southwestern China and southern Asia near the Qinghai-Tibet Plateau and the Himalayas. A comparison of the genomic diversity of wild, landrace, and improved Persian walnuts showed evidence of loss of diversity associated with domestication but a relatively weak additional loss of diversity from cultivar selection. The resources generated in this study provide a genomic framework for future germplasm use and walnut improvement.

Genetic diversity in Persian walnuts

When we began this research, we expected that close examination would show that breeders had reduced the genetic diversity of walnut as they advanced germplasm into the final stages of cultivar development. In general, this was not true; the genetic diversity, LD distance, and gene flow provided a relatively weak artificial selection in the evolution process, which is different from the main annual crops like rice and soybeans [18–20]. This could be explained by the outcrossing system, domestication history, and modern breeding history. First, walnut is wind pollination and heterodichogamous characteristics of the breeding system [5–6,12], perennial and long juvenile phase, which is similar to some other woody fruit trees (e.g., pear, apple, peach, and jujube) [22–25,32–36]. Second, from the point of domestication history, the divergence time between wild walnuts vs. landrace and cultivar was much later than annual crops such as rice and soybean. *J. regia* is considered as food used by humans back to Persia (7,000 BCE) [7,9,12,16], however, as an important annual crop, both rice and soybean ancient human domestication started from ~8,000 BCE [19–20,37–39]. Moreover, the walnut population had declined by the bottleneck scenario [40], which limited the expansion of genetic diversity. Third, intensive breeding may not be necessary where landraces already expressed desirable commercial traits to a considerable degree. Modern cultivars have a short pedigree and are generally developed by several breeding programs around the world from a geographically and genetically diverse group of landraces. In addition, most modern walnut cultivars have relatively short pedigrees because breeding programs have managed only a few generations of selection in the past 70 years or so, as walnut is perennial woody and has a long generation time. Importantly, the phenomenon of weak selection reminds us that it should be cautious during artificial selection breeding as wild material can also be feral accessions.

Phylogenetic relationships and gene introgressions of walnuts

Within *Juglans*, the phylogenetic tree showed that the *J. regia* samples (section *Dioscaryon*) were monophyletic (Fig 1B). Distinct and parallel clades were seen for the sect. *Cardiocaryon*, sect. *Trachycaryon*, and sect. *Rhysocaryon* based on our walnut samples (Fig 1B). The systematic position of Persian walnut in the genus *Juglans* has been well resolved based on morphology and molecular evidence, including sequence data from the internal transcribed spacer (ITS), five chloroplast DNA spacer sequences, a hypervariable *matK*, restriction fragment length polymorphisms (RFLPs), and whole chloroplast genomes [4–5,15,41–44]. The origin of sect. *Dioscaryon* is a topic of active research; however, it may be that *J. regia* and other members of *Dioscaryon* were derived from the hybridization of ancestors that became what is now sect. *Rhysocaryon* and sect. *Cardiocaryon*, a conclusion based on nuclear SNPs [11,45]. Differences in the evolutionary rate, inherited background, and complex speciation history of *Juglans* may explain why the phylogeny of Persian walnut differs depending on whether it is based on nuclear SNPs or chloroplast and mitochondrial sequences [46–48]. The tree that we generated based on SNPs was consistent with previous conclusions based on nuclear data and chloroplast genomes [43–48]. Therefore, our whole-genome resequencing data offer a valuable resource for examining the taxonomy of *J. regia* within the genus *Juglans*.

We observed that the landraces and cultivars were mixed and showed signs of introgression, certainly because of the use of landraces for cultivar breeding, as mentioned above, but probably also because seeds or scion wood of cultivars was propagated by villagers, which permitted gene exchange through pollen [2]. The gene flow analysis showed strong gene flow between landraces and cultivars by the BABA-ABBA test, Nm calculation, and TreeMix analysis (Figs 2D and S2). Based on studies of other tree species, local gene flow between wild and cultivated individuals is an important evolutionary factor [49–52]. Walnut landraces propagated by local communities as seedlings represent a valuable resource for breeders as a secondary germplasm pool [8].

The demographic history and past ecological distribution shifts of Persian walnut

Our results are supported by previous genetic and ecological data showing that Persian walnut survived the LGM in glacial refugia spread across a wide geographic area from 30°N to 45°N latitude in the Balkans, western Europe, Xinjiang Province, northeastern China, central China, and southwestern China [7,10,12]. During the Pleistocene, the differentiation of populations may have been influenced by repeated extinction and colonization during the Quaternary climate oscillations [12,53].

The presence of walnut habitat in the mountains of southern Asia was supported by a recent archaeological study of several new sites in northeastern India containing carbonized walnut samples ascribed to an earlier date (~4500 B.P.) than previously reported (~4000–3500 B.P.) [54]. The wide distribution of common walnuts in broadly similar but fragmented environments across Asia during the Late Miocene and Pleistocene climate changes [7,12] may have resulted in regionally adapted populations with an effective population size continually constrained by climate fluctuations and the long generation time of this perennial nut crop [22,54–55]. The formation of the Qinghai-Tibet Plateau and climate changes were the important drivers of the long-term population dynamics of Persian walnut [56–57].

Persian walnut is considered a relict species of the Tertiary Period [7,10,12] native to the mountain ranges of Asia [11,27,47]. Here, we found that the Qinghai-Tibet Plateau in southwestern China (including southern Asia and southwestern China) may be the core region of genetic diversity (Fig 3B; Tables 1 and S2). Taking these results together, it is inferred that

Persian walnut found worldwide differentiated from an ancestral Qinghai-Tibet Plateau population. This region is where the only other taxon in *Dioscaryon* (*J. sigillata*) is found [58]. The current distribution of walnut is the result of expansion/contraction from multiple refugia during the uplift of the Himalayan Mountains, climate changes, ecological adaptation [53,59–60], and later human exploitation [61–63].

Genes under selection

To the best of our knowledge, the diaphragm is a typical domestication trait that protects the nuts during natural environmental selection, thereby maintaining the continuity of the species. In addition, the diaphragm affects the ease of kernel extraction, and a thin diaphragm shows the great commercial value and thus is attractive to breeders. In the study, selection analysis and GWAS revealed that *JrbHLH* and *JrMYB6* were associated with the thickness of the diaphragm in walnut. The homologs of *JrbHLH* in poplar was related to cell wall development [64–65]. The homologs of *JrMYB6* in *Malus domestica* is involved in relative lignin metabolism [66–67]. It is known that lignin is a principal structural component of cell walls in higher terrestrial plants [68]. Thus, *JrbHLH* and *JrMYB6* may regulate cell wall formation and lignin biosynthesis during xylem development in walnut. Additionally, expression analysis indicated that *JrMYB6* was upregulated in walnut diaphragm development, but *JrbHLH* was downregulated during walnut diaphragm development. Collectively, *JrbHLH* and *JrMYB6* might differentially express to participate in the xylem development to affect the thickness of the diaphragm in walnut. Thus, these two TFs provide potential targets for improving Persian walnut nut quality by molecular marker-assisted selection or genetic manipulations.

Materials and methods

Sample collection and preparation

Wild samples were collected from apparently healthy mature adults growing in mountain forests, near villages, or along forest roads but not in orchards or on farmed land. The sampled trees were separated by at least 50 m. Landrace samples were collected from different local communities, while the cultivar samples were collected from orchards containing the named cultivars. A total of 55 Persian walnut relatives, including 23 black walnuts (*J. nigra*), 20 butternuts (8 *J. cathayensis*, 8 *J. mandshurica*, and 4 *J. cinerea*), 11 hybrids (*J. hopeiensis*), and *Carya illinoensis* (S1 Table), were obtained from improvement programs from many countries and were presumably selected to produce high-quality nuts in a wide range of production environments. The distribution of all sampling points was plotted with ArcGIS v10.2 (ESRI Inc., Redlands, California, USA). DNA extraction was performed using the improved CTAB method of Zhao and Woeste (2011) [69]. From 2018 to 2019, we measured the thickness of the diaphragm in Persian walnut accessions grown in Tai'an, Shandong Province, China (S1 Table). Three nuts from each accession in each replication were used for the measurements of diaphragm thickness.

DNA sequencing and detection of variations

In total, 5169 Gb of raw data was obtained using paired-end 150 bp reads. Libraries were processed with the Illumina Cluster Generation Station following the manufacturer's recommendations and sequenced on a HiSeq 4000 instrument. The CASAVA 1.7.0 version of the Illumina pipeline was used to process raw data [70].

To obtain clean reads, we filtered the reads with three features: 1) reads with adapters, 2) reads with low quality (<10) base percentage greater than 20% compared to the whole

genome, and 3) reads with N percentages greater than 5%. After filtering, we mapped the clean reads to the reference genome using BWA V0.7.17 software with default parameters [71]. Using the *J. regia* 'Chandler' (updated version 2.0) chromosome-level genome as a reference [26], clean data were mapped to detect SNPs and InDels for all 309 genotypes. Based on high-quality alignment (Map Quality > 20), we used Genome Analysis ToolKit V4.1.1.0 model *haplotypcaller* to identify SNPs [72]. Finally, to ensure data quality, we kept the SNPs that met the following criteria: 1) biallelic SNPs, 2) SNPs with a quality score > 30, 3) SNPs with a missing rate (MR) < 0.25, and 4) SNPs with a minor allele frequency (MAF) > 0.05.

Genetic diversity, population structure, and LD analysis

To compare the diversity of *J. regia* groups around the world, we wrote an in-house Perl script (<https://github.com/xujiabao127/Walnut/>) to calculate the diversity parameters, including the SNP number, the average pairwise divergence within a population (θ_π), Watterson's estimator (θ_w) and *Tajima's D*. Along the genome, a sliding window of 10 kb with a 5 kb step was used to estimate the θ_π , θ_w , and *Tajima's D* values. The population differentiation parameter F_{ST} and reduction of diversity (ROD) were computed in the same windows based on a pairwise SNP sequence [73]. PHYLIP software (version 3.696) was used to produce a cladogram based on the genetic distance matrix using the p-distance formula [74]. The algorithm we chose was the neighbour-joining (NJ) method. We set *C. illinoensis* as the outgroup. Figtree was used to graph the phylogenetic tree. We performed PCA using PLINK software (version 1.90) with default settings [75]. The top four eigenvectors were selected for the graph shown by EIGENSOFT software [76]. The population structure of 98 wild Persian walnut samples and 254 wild, cultivar, and landrace samples was investigated using ADMIXTURE software (version 1.3.0) [77] with a marginal likelihood model. We ran 10,000 iterations, and the number of clusters (K) was set from 2 to 10, representing the assumed groups of the simulated ancestral population. The best K was inferred based on the delta K method using the Structure Harvester program [78]. LD was calculated using PLINK software (version 1.07) [75] with the following settings:—file—r2—ld-window 99999—ld-window-kb 200—out. Then, values for the r^2 statistics were obtained. LD decay was calculated based on r^2 between two SNPs and the distance between the two SNPs.

Population history

We used the PSMC model to infer the population size history of different groups. First, heterozygous sites were generated by SAMtools version 1.6 [79]. The parameters were “*samtools mpileup -C 50 -Q 20 -u -v*” and “*vcfutils.pl vcf2fq -d 4 -D 100*”. Second, the diploid consensus sequence was obtained by fq2psmcfa as an input of the PSMC, and the parameter was “fq2psmcfa -q20”. Third, PSMC V0.6.5 was run with the parameter “psmc -N25 -t15 -r5 -p 4 +25*2+4+6”. To analyze the results, we used 30 years as the generation time, and the mutation rate was set as 2.09E-8 substitutions per site per generation [20].

Gene flow analysis between wild, landrace, and cultivar walnuts

To understand gene flow, we calculated the D-statistics of four populations with ADMIXtools version 4.1 software [77]. The SNPs in vcf format were converted to EIGENSOFT format via VCFtools version 0.1.13 and CONVERTF provided by ADMIXtools [77]. The D-statistics were computed by qpDstat. Sample s99 (*C. illinoensis*) was set as the outgroup (X) for all four population tests. For any four populations (W, X, Y, and Z), the gene flow occurred either between W and Y or between X and Z when the Z score was more than 0. Conversely, gene flow occurred between either W and Z or X and Y when the Z score was smaller than 0. A

gene flow signal was considered significant when the Z score was more than 3 or less than -3. To study gene flow among the wild samples, landraces, and cultivars, we calculated two parameters: the D-statistic and Nm (estimate of gene flow). The D-statistics were generated by ADMIXtools software based on the fixation index (F_{ST}). Nm was calculated according to Wright's (1931) formula $Nm = (1 - F_{ST}) / (4 * F_{ST})$ [80]. TreeMix22 version 1.12 was used to model the gene flow between the outgroups, wild walnuts, landrace walnuts, and cultivar walnuts [81].

Detection of selective sweeps

To identify evidence of selection in Persian walnut genomes associated with domestication and improvement, we screened for genomic regions with a sharp ROD between wild and domesticated walnut (cultivars + landraces) groups ($\pi_{wild}/\pi_{cultivar+landrace}$; domesticated). Genomic regions under selection often showed a decrease in genetic diversity. We identified genomic regions selected by domestication by comparing the F_{ST} and ROD for the pool of wild trees compared to the cultivar+landrace pool using a 100-kb sliding window with a step size of 10 kb using VCFtools software (v0.1.13) [73,79,81]. The candidate selected regions associated with human cultivation were identified with the following criteria: 1) $F_{ST} > 0.15$ (the genetic differentiation between groups is large), 2) the top 10% of F_{ST} values, and 3) $ROD > 0.2$. In the domestication-related genomic regions, we selected genes meeting these three criteria as candidate domestication genes [73,78].

GWAS and identification of the candidate genes

The SNPs with an $MAF < 0.05$ and a missing rate < 0.25 were used for the GWAS in SHAPEIT (4.0) [82–83]. To obtain a high confidence value for each trait, we calculated average values among different years. Then, we deleted the sample with a trait value greater than mean+2 standard deviations (SDs) or less than the mean-2 SDs. To estimate the kinship among samples, we used the software tassel (V5.0) to calculate the kinship matrix [84]. We used admixture (v 1.3.0) to calculate the population structure and Q matrix for the GWAS [77]. Based on the SNPs, trait values, kinship matrix, and Q matrix, we performed the GWAS analysis by the FarmCPU model in GAPIT (V3.0) software [85]. According to the associated loci determined by the GWAS, SNP types and locations were identified using the reference genome [26]. Haploview v4.2 and LDBlockShow v1.40 software were used to construct and visualize haplotype maps [86–87].

Gene expression analysis by qRT-PCR

To verify the expression patterns of two candidate transcription factors (*JrbHLH* and *JrMYB6*), we collected common walnut diaphragms from one adult tree growing at Xi'an (Northwest University campus) on 95 days after flowering (DAF), 120 DAF, and 140 DAF, respectively. Total RNA from walnut diaphragms was processed using RNA Library Prep Kit (Beverly, MA, USA). We then performed quantitative real-time PCR (qRT-PCR) reactions in a common walnut diaphragm during three development stages (July 27, August 18, and September 6) (primers see details in S6 Table). Before the PCR experiment, primer specificities and corresponding melting curves were verified. In each qRT-PCR experiment, three replicates were performed. Real-time amplification responses were performed on an Applied Biosystems (USA) 7500 quick real-time PCR system [88–89]. The relative concentration of expression of each gene was calculated using the 2-Ct method [88–89].

Supporting information

S1 Table. Summary of 309 walnut accessions sequenced and mapped against the ‘Chandler’ (*J. regia*) reference.

(XLSX)

S2 Table. Linkage disequilibrium (LD) between cultivar, landrace, and wild walnut samples.

(XLSX)

S3 Table. D statistics for comparisons among wild walnut populations.

(XLSX)

S4 Table. Genes identified as candidates for selection by F_{ST} ratio of wild and cultivated accessions, and their functional annotation.

(XLSX)

S5 Table. Identification and annotation of candidate genes associated with the thickness of the diaphragm.

(XLSX)

S6 Table. Primers used for quantitative real-time PCR.

(XLSX)

S1 Fig. Variation across the chromosomes of *Juglans regia* based on 300Kb windows. (A) sixteen chromosomes of *Juglans regia*. (B) Gene density represented as a grayscale heatmap; i.e., blacker areas had a higher gene density (max value was 56 single nucleotide polymorphisms (SNPs) per 300 kbp window). (C) Number of SNPs in 300-kb windows. SNPs among wild walnuts (green), cultivars (light blue), and landraces (red) (max value was 36 SNPs per 300 kbp window). (D) Insertions and deletions (indels) in wild walnuts (green), cultivars (light blue), and landraces (red). (E) The genomic nucleotide diversity (π) of the nuclear genomes of wild walnuts (green), cultivars (light blue), and landraces (red). (F) Heatmap of the genetic differentiation (F_{ST}) between wild walnuts and selected walnuts (cultivars and landraces). (G) Heatmap of the coverage of selected regions between wild walnuts and selected walnut (cultivars and landraces).

(TIF)

S2 Fig. Gene-flow patterns that were detected among walnut groups using TreeMix.

JY = outgroup, Wild = wild walnuts, Landrace = landrace walnuts, Cultivated = cultivar walnuts.

(TIF)

S3 Fig. The principal component analysis (PCA) of wild walnuts. Principal components plots for wild walnut accessions, including samples from Eastern Asian (EA, N = 39), Yunnan-Kweichow Plateau (YG, N = 14), Xinjiang Province (XJ, N = 8); southern Asia (SA, N = 17); western Asia (WA, N = 9); Europe (EU, N = 11).

(TIF)

S4 Fig. Population structure ($K = 2-6$) of wild walnuts. (A) Each color corresponds to a single population as noted. Each walnut accession is represented by a vertical bar. The Y-axis refers to the proportion of the genetic background, and the height of each line with different colors represents the probability of an accession belonging to a different genetic background. (B) Delta K showed a peak at 4, suggesting four clusters as the most appropriate option, which supports the phylogenetic tree and PCA result of “four major discrete clusters of Eastern Asian

(EA), Yunnan-Kweichow Plateau (YG), Xinjiang province (XJ), and southern Asian (SA)+western Asia (WA)+ Europe (EU) groups were detected”.

(TIF)

S5 Fig. Alignment of bHLH proteins from *Juglans regia*, *Arabidopsis thaliana*, *Glycine max*, *Populus trichocarpa*, and *Malus domestica*. The protein information of these plants as follows: *J. regia* (XP_035546602.1), *A. thaliana* (AT1G69010.1), *G. max* (KRH03185), *P. trichocarpa* (PNT00429), and *M. domestica* (mRNA:MD06).

(TIF)

S6 Fig. Alignment of MYB proteins from *Juglans regia*, *Zea mays*, *Arabidopsis thaliana*, *Solanum lycopersicum*, *Malus domestica*, *Oryza sativa*, *Glycine max*, and *Populus trichocarpa*. The protein information of these plants as follows: *J. regia* (XP_018812236.1), *Zea mays* (Zm00001eb392230.), *Arabidopsis thaliana* (AT3G47600.1), *Solanum lycopersicum* (Soly06g069), *Malus domestica* (mRNA:MD06), *Oryza sativa* (Os12t0125000-01), *Glycine max* (KRH64960.1), and *Populus trichocarpa* (PNS90469).

(TIF)

Acknowledgments

We thank Professor Shuoxin Zhang and Dr. Fan Hao for suggestions of the manuscript and for helpful discussions. We are grateful to Dr. Hanif Khan and Dr. Irfan Ullah for collecting samples. We acknowledge the experimental technical support from Jiayu Ma, Ruimin Xi, and Yujie Yan. Mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the U.S. Dept. of Agriculture and does not imply its approval to the exclusion of other products or vendors that also may be suitable.

Author Contributions

Conceptualization: Xiang Luo, Peng Zhao.

Data curation: Xiang Luo, Huijuan Zhou, Da Cao, Xin Chen, Maria Malvolti, Chaobin Liu, Aziz Ebrahimi, Chengkui Qiao, Hang Ye, Peng Zhao.

Formal analysis: Huijuan Zhou, Feng Yan, Pengpeng Chen, Jiangtao Wang, Hang Ye, Mengdi Li, Jiabao Xu, Peng Zhao.

Funding acquisition: Xiang Luo, Peng Zhao.

Investigation: Xiang Luo, Huijuan Zhou, Zhenhua Lu, Shangying Cao, Peng Zhao.

Methodology: Xiang Luo, Keith Woeste, Peng Zhao.

Project administration: Peng Zhao.

Resources: Peng Zhao.

Supervision: Peng Zhao.

Visualization: Peng Zhao.

Writing – original draft: Peng Zhao.

Writing – review & editing: Xiang Luo, Keith Woeste, Zhangjun Fei, Hong An, Kai Ma, Shangying Cao, Peng Zhao.

References

- Martínez ML, Mattea MA, Maestri DM. Pressing and supercritical carbon dioxide extraction of walnut oil. *J Food Eng*. 2008; 88:399–404.
- Hu Y, Dang M, Zhang T, Luo G, Xia H, Zhou H, et al. Genetic diversity and evolutionary relationship of *Juglans regia* wild and domesticated populations in Qinling Mountains based on nrDNA ITS Sequences. *Sci Sil Sin*. 2014; 50:47–55.
- Abdallah IB, Tlili N, Martínez-Force E, Rubio AGP, Camino MCP, Albouchi A, et al. Content of carotenoids, tocopherols, sterols, triterpenic and aliphatic alcohols, and volatile compounds in six walnuts (*Juglans regia* L.) varieties. *Food Chem*. 2015; 173:972–978.
- Wayne EM. The classification within the *Juglandaceae*. *Ann Mo Bot Gard*. 1978; 65:1058.
- Martínez-García PJ, Crepeau MW, Puiu D, Gonzalez-Ibeas D, Whalen J, Stevens KA, et al. The walnut (*Juglans regia*) genome sequence reveals diversity in genes coding for the biosynthesis of nonstructural polyphenols. *Plant J*. 2016; 87:507–532.
- Bernard A, Lheureux F, Dirlwanger E. Walnut: past and future of genetic improvement. *Tree Genet Genomes*. 2018; 14:1.
- Pollegioni P, Woeste KE, Chiocchini F, Lungo SD, Cioffi M, Olimpieri I, et al. Rethinking the history of common walnut (*Juglans regia* L.) in Europe: Its origins and human interactions. *PLoS One*. 2017; 12:e0172541.
- Dangl GS, Woeste K, Aradhya M, Pitcher AMK, Simon CJ, Potter D, et al. Characterization of 14 micro-satellite markers for genetic analysis and cultivar identification of walnut. *J Am Soc Hortic Sci*. 2005; 130:348–354.
- Beer R, Kaiser F, Schmidt K, Ammann B, Carraro G, Grisa E, et al. Vegetation history of the walnut forests in Kyrgyzstan (central Asia): natural or anthropogenic origin? *Quaternary Sci Rev*. 2008; 27:621–632.
- Aradhya M, Velasco D, Ibrahimov Z, Toktoraliev B, Maghradze D, Musayev M, et al. Genetic and ecological insights into glacial refugia of walnut (*Juglans regia* L.). *PLoS One*. 2017; 12:e0185974.
- Zhang BW, Xu LL, Li N, Yan PC, Jiang XH, Woeste KE, et al. Phylogenomics reveals an ancient hybrid origin of the Persian walnut. *Mol Biol Evol*. 2019; 11:2451–2461. <https://doi.org/10.1093/molbev/msz112> PMID: 31163451
- Feng XJ, Zhou HJ, Zulfikar S, Luo X, Hu YH, Feng L, et al. The phytogeographic history of common walnut in China. *Front Plant Sci*. 2018; 9:1399. <https://doi.org/10.3389/fpls.2018.01399> PMID: 30298084
- Zeven AC, Zhukovsky PM. Dictionary of cultivated plants and their centers of diversity excluding ornamentals, forest trees, and lower plants. Wageningen: Centre for Agricultural Publishing and Documentation; 1975.
- Xi RT. Discussion on the origin of walnut in China. *Acta Horticulturae*. 1990; 284:353–362.
- Aradhya MK, Potter D, Gao FY, Simon CJ. Molecular phylogeny of *Juglans* (*Juglandaceae*): A biogeographic perspective. *Tree Genet Genomes*. 2007; 3:363–378.
- Pollegioni P, Woeste KE, Chiocchini F, Lungo SD, Olimpieri I, Tortolano V, et al. Ancient humans influenced the current spatial genetic structure of common walnut populations in Asia. *PLoS One*. 2015; 10:e0135980. <https://doi.org/10.1371/journal.pone.0135980> PMID: 26332919
- Han H, Woeste KE, Hu Y, Dang M, Zhang T, Gao XX, et al. 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*. 2016; 12:111.
- Lam HM, Xu X, Liu X, Chen WB, Yang GH, Wong FL, et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat Genet*. 2010; 42:1053–9. <https://doi.org/10.1038/ng.715> PMID: 21076406
- Xu X, Liu X, Song G, Jensen JD, Hu FY, Li X, et al. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat Biotechnol*. 2012; 30:105–11.
- Meyer R, Choi JY, Sanches M, Plessis A, Flowers JM, Amas J et al. Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nature Genet*. 2016; 48:1083–1088. <https://doi.org/10.1038/ng.3633> PMID: 27500524
- Du X, Huang G, He S, Yang Z, Sun G, Ma X, et al. Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nat Genet*. 2018; 50:796–802. <https://doi.org/10.1038/s41588-018-0116-x> PMID: 29736014
- Duan NB, Bai Y, Sun HH, Wang N, Ma YM, Li MJ, et al. Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nat Commun*. 2017; 8:249. <https://doi.org/10.1038/s41467-017-00336-7> PMID: 28811498

23. Wu J, Wang YT, Xu JB, Korban SS, Fei ZJ, Tao ST, et al. Diversification and independent domestication of Asian and European pears. *Genome Biol.* 2018; 19:77. <https://doi.org/10.1186/s13059-018-1452-y> PMID: 29890997
24. Li Y, Cao K, Zhu GR, Fang WC, Chen CW, Wang XW, et al. Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. *Genome Biol.* 2019; 20:36. <https://doi.org/10.1186/s13059-019-1648-9> PMID: 30791928
25. LaBonte NR, Zhao P, Woeste K. Signatures of selection in the genomes of Chinese chestnut (*Castanea mollissima* Blume): the roots of nut tree domestication. *Front Plant Sci.* 2018; 9:810.
26. Marrano A, Britton M, Zaini PA, Zimin AV, Workman RE, Puiu D, et al. High-quality chromosome-scale assembly of the walnut (*Juglans regia* L.) reference genome. *GigaScience.* 2020; 9:giaa050.
27. Zhang JP, Zhang WT, Ji FY, Qiu J, Song XB, Bu DC, et al. High-quality walnut genome assembly reveals extensive gene expression divergences after whole-genome duplication. *Plant Biotechnol J.* 2020; 18:1848–1850.
28. Charrier G, Chuine I, Bonhomme M, Améglio T. Assessing frost damages using dynamic models in walnut trees: exposure rather than vulnerability controls frost risks. *Plant Cell Environ.* 2017; 41:1008–1021. <https://doi.org/10.1111/pce.12935> PMID: 28185293
29. Mercè G, Guillaume C, Antoni V, Robert S, Thierry A, Neus A. Genetics of frost hardiness in *Juglans regia* L. and relationship with growth and phenology. *Tree Genet Genomes.* 2016; 12:83.
30. Charrier G, Ngao J, Saudreau M, Améglio T. Effects of environmental factors and management practices on microclimate, winter physiology, and frost resistance in trees. *Front Plant Sci.* 2015; 6:259. <https://doi.org/10.3389/fpls.2015.00259> PMID: 25972877
31. Germplasm Resources Information Network [Internet]. Beltsville (MD): United States Department of Agriculture, Agricultural Research Service. [2020]. Available from: <http://www.ars-grin.gov/>.
32. Guo M, Zhang Z, Li S, Lian Q, Fu P, He Y, et al. Genomic analyses of diverse wild and cultivated accessions provide insights into the evolutionary history of jujube. *Plant Biotechnol J.* 2021; 19:517–531. <https://doi.org/10.1111/pbi.13480> PMID: 32946650
33. Chen P, Li Z, Zhang D, Shen W, Xie Y, Zhang J, et al. Insights into the effect of human civilization on *Malus* evolution and domestication. *Plant Biotechnol J.* 2021; 19:2206–2220.
34. Cao K, Wang B, Fang W, Zhu G, Chen C, Wang X, et al. Combined nature and human selections reshaped peach fruit metabolome. *Genome Biol.* 2022; 23:146. <https://doi.org/10.1186/s13059-022-02719-6> PMID: 35788225
35. Cornille A, Gladioux P, Smulders MJ, Roldán-Ruiz I, Laurens F, Le Cam B, et al. New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genet.* 2012; 8:e1002703. <https://doi.org/10.1371/journal.pgen.1002703> PMID: 22589740
36. Sun X, Jiao C, Schwaninger H, Chao CT, Ma Y, Duan N, et al. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat Genet.* 2020; 52:1423–1432. <https://doi.org/10.1038/s41588-020-00723-9> PMID: 33139952
37. Molina J, Sikora M, Garud N, Flowers JM, Rubinstein S, Reynolds A, et al. Molecular evidence for a single evolutionary origin of domesticated rice. *Proc Natl Acad Sci U S A.* 2011; 108:8351–8356. <https://doi.org/10.1073/pnas.1104686108> PMID: 21536870
38. Choi JY, Zaidem M, Gutaker R, Dorph K, Singh RK, Purugganan MD. The complex geography of domestication of the African rice *Oryza glaberrima*. *PLoS Genet.* 2019; 15:e1007414.
39. Lu S, Zhao X, Hu Y, Liu S, Nan H, Li X, et al. Natural variation at the soybean J locus improves adaptation to the tropics and enhances yield. *Nat Genet.* 2017; 49:773–779. <https://doi.org/10.1038/ng.3819> PMID: 28319089
40. Ding YM, Cao Y, Zhang WP, Chen J, Liu J, Li P, et al. Population-genomic analyses reveal bottlenecks and asymmetric introgression from Persian into iron walnut during domestication. *Genome Biol.* 2022; 23:145. <https://doi.org/10.1186/s13059-022-02720-z> PMID: 35787713
41. Hu YH; Woeste KE; Zhao P. Completion of the chloroplast genomes of five Chinese *Juglans* and their contribution to chloroplast phylogeny. *Front Plant Sci.* 2017; 7:1955.
42. Dong WP, Xu C, Li WQ, Xie XM, Lu YZ, Liu YL, et al. Phylogenetic resolution in *Juglans* based on complete chloroplast genomes and nuclear DNA sequences. *Front Plant Sci.* 2017; 8:1148.
43. Fjellstrom RG, Parfitt DE. Phylogenetic analysis and evolution of the genus *Juglans* (*Juglandaceae*) as determined from nuclear genome RFLPs. *Plant Syst Evol.* 1995; 197:19–32.
44. Stanford AM, Harden R, Parks CR. Phylogeny and biogeography of *Juglans* (*Juglandaceae*) based on matK and ITS sequence data. *Am J Bot.* 2000; 87:872–882.

45. Mu XY, Sun M, Yang PF, Lin QW. Unveiling the identity of Wenwan walnuts and phylogenetic relationships of Asian *Juglans* species using restriction site-associated DNA-Sequencing. *Front Plant Sci.* 2017; 8:1708.
46. Soltis DE, Albert VA, Savolainen V, Hilu K, Qiu YL, Chase MW, et al. Genome-scale data, angiosperm relationships, and 'ending incongruence': a cautionary tale in phylogenetics. *Trends Plant Sci.* 2004; 10:477–483. <https://doi.org/10.1016/j.tplants.2004.08.008> PMID: 15465682
47. Zhao P, Zhou HJ, Potter D, Hu YH, Feng XJ, Dang M, et al. Population genetics, phylogenomics and hybrid speciation of *Juglans* in China determined from whole chloroplast genomes, transcriptomes, and genotyping-by-sequencing (GBS). *Mol Phylogenet Evol.* 2018; 126:250–265. <https://doi.org/10.1016/j.ympev.2018.04.014> PMID: 29679714
48. Morris AB, Shaw J. Markers in time and space: A review of the last decade of plant phylogeographic approaches. *Mol. Ecol.* 2018; 27:2317–2333. <https://doi.org/10.1111/mec.14695> PMID: 29675939
49. Xanthopoulou A, Manioudaki M, Bazakos C, Kissoudis C, Farsakoglou AM, Karagiannis E, et al. Whole genome re-sequencing of sweet cherry (*Prunus avium* L.) yields insights into genomic diversity of a fruit species. *Hortic Res.* 2020; 7:60.
50. Liang ZC, Duan SC, Sheng J, Zhu SS, Ni XM, Shao JH, et al. Whole-genome resequencing of 472 *Vitis* accessions for grapevine diversity and demographic history analyses. *Nat Commun.* 2019; 10:1190.
51. Wu DZ, Liang Z, Yan T, Xu Y, Xuan LJ, Tang J, et al. Whole-genome resequencing of a worldwide collection of rapeseed accessions reveals the genetic basis of ecotype divergence. *Mol Plant.* 2019; 12:30–43. <https://doi.org/10.1016/j.molp.2018.11.007> PMID: 30472326
52. Huang J, Zhang C, Zhao X, Fei Z, Wan K, Zhang Z, et al. The jujube genome provides insights into genome evolution and the domestication of sweetness/acidity taste in fruit trees. *PLoS Genet.* 2016; 12(12):e1006433. <https://doi.org/10.1371/journal.pgen.1006433> PMID: 28005948
53. Hoban SM, Borkowski DS, Brosi SL, McCreary TS, Thompson LM, McLachlan JS, et al. Range-wide distribution of genetic diversity in the North American tree *Juglans cinerea*: a product of range shifts, not ecological marginality or recent population decline. *Mol. Ecol.* 2010; 19:4876–4891.
54. Roor W, Konrad H, Mamadjanov D, Geburek T. Population differentiation in common walnut (*Juglans regia* L.) across major parts of its native range—insights from molecular and morphometric data. *J Hered.* 2017; 108:391–404.
55. Ortego J, Knowles LL. Incorporating interspecific interactions into phylogeographic models: A case study with Californian oaks. *Mol. Ecol.* 2020; 29:4150–4524. <https://doi.org/10.1111/mec.15548> PMID: 32657460
56. Zhang DF, Fengquan L, Jianmin B. Eco-environmental effects of the Qinghai-Tibet Plateau uplift during the Quaternary in China. *Environ Geol.* 2000; 39:1352–1358.
57. Wu GH, Terol J, Ibanez V, López-García A, Pérez-Román E, Borredá C, et al. Genomics of the origin and evolution of *Citrus*. *Nature.* 2018; 554:311–316.
58. Wang H, Pan G, Ma QG, Zhang JP, Pei D. The genetic diversity and introgression of *Juglans regia* and *Juglans sigillata* in Tibet as revealed by SSR markers. *Tree Genet Genomes* 2015; 11:804–814.
59. McCammon MT, Hartmann MA, Bottema CD, Parks LW. Sterol methylation in *Saccharomyces cerevisiae*. *J Bacteriol.* 1984; 157:475–483. <https://doi.org/10.1128/jb.157.2.475-483.1984> PMID: 6363386
60. Via S. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philos T R Soc B.* 2012; 367:451–460. <https://doi.org/10.1098/rstb.2011.0260> PMID: 22201174
61. Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, et al. Genetic structure and domestication history of the grape. *Proc Natl Acad Sci U S A.* 2011; 108:3530–3535. <https://doi.org/10.1073/pnas.1009363108> PMID: 21245334
62. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature.* 2011; 475:493–496. <https://doi.org/10.1038/nature10231> PMID: 21753753
63. Han F, Lamichhaney S, Grant BR, Grant PR, Andersson L, Webster MT. Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Res.* 2017; 27:1004–1015. <https://doi.org/10.1101/gr.212522.116> PMID: 28442558
64. Jin H, Do J, Moon D, Noh EW, Kim W, Kwon M. EST analysis of functional genes associated with cell wall biosynthesis and modification in the secondary xylem of the yellow poplar (*Liriodendron tulipifera*) stem during early stage of tension wood formation. *Planta.* 2011; 234:959–977.
65. Zhang L, Liu B, Zhang J, Hu J. Insights of molecular mechanism of xylem development in five black poplar cultivars. *Front Plant Sci.* 2020; 11:620. <https://doi.org/10.3389/fpls.2020.00620> PMID: 32547574
66. Lee SB, Suh MC. Cuticular wax biosynthesis is up-regulated by the MYB94 transcription factor in Arabidopsis. *Plant Cell Physiol.* 2015; 56:48–60. <https://doi.org/10.1093/pcp/pcu142> PMID: 25305760

67. Chen K, Song M, Guo Y, Liu L, Xue H, Dai H, Zhang Z. MdMYB46 could enhance salt and osmotic stress tolerance in apple by directly activating stress-responsive signals. *Plant Biotechnol J*. 2019; 17:2341–2355. <https://doi.org/10.1111/pbi.13151> PMID: 31077628
68. Vanholme R, De Meester B, Ralph J, Boerjan W. Lignin biosynthesis and its integration into metabolism. *Curr Opin Biotechnol*. 2019; 56:230–239. <https://doi.org/10.1016/j.copbio.2019.02.018> PMID: 30913460
69. Zhao P, Woeste K. DNA markers identify hybrids between butternut (*Juglans cinerea* L.), and Japanese walnut (*Juglans ailantifolia* Carr.). *Tree Genet Genomes*. 2011; 7:511–533.
70. Widayati A, Dewi S, Khasanah N M. Capacity-strengthening approach to vulnerability assessment (CaSAVA). *Worldagroforestry*. 2020;234–238.
71. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
72. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010; 20:1297–1303. <https://doi.org/10.1101/gr.107524.110> PMID: 20644199
73. Ji F, Ma Q, Zhang W, Liu J, Feng Y, Feng Y. et al. A genome variation map provides insights into the genetics of walnut adaptation and agronomic traits. *Genome Biol*. 2021; 22:300. <https://doi.org/10.1186/s13059-021-02517-6> PMID: 34706738
74. Retief JD. Phylogenetic analysis using PHYLIP. *Methods Mol Biol*. 2000; 132:243–258. <https://doi.org/10.1385/1-59259-192-2:243> PMID: 10547839
75. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81:559–575. <https://doi.org/10.1086/519795> PMID: 17701901
76. Patterson N, Price A.L, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2016; 2:e190.
77. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient admixture in human history. *Genetics* 2012; 192:1065–1093. <https://doi.org/10.1534/genetics.112.145037> PMID: 22960212
78. Montana G, Clive J.H. Statistical software for gene mapping by admixture linkage disequilibrium. *Brief Bioinform*. 2007; 8:393–395. <https://doi.org/10.1093/bib/bbm035> PMID: 17640923
79. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
80. Helen Wright, R. Wages: a means of testing their adequacy. *Social Service Review* 1931; 5:515–516.
81. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet*. 2012; 8:e1002967. <https://doi.org/10.1371/journal.pgen.1002967> PMID: 23166502
82. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, Depristo MA. et al. The variant call format and VCFtools. *Bioinformatics*. 2011; 27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330> PMID: 21653522
83. O'Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, et al. A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet*. 2014; 10:e1004234. <https://doi.org/10.1371/journal.pgen.1004234> PMID: 24743097
84. Bradbury PJ, Zhang Z, Kroon D. E, Casstevens TM, Buckler ES. Tassel: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007; 23:2633–2635. <https://doi.org/10.1093/bioinformatics/btm308> PMID: 17586829
85. Lipka A, Tian F, Wang QS, Peiffer JA, Li M, Bradbury PJ, et al. GAPIT: Genome association and prediction integrated tool. *Bioinformatics*. 2012; 28:2397–2399. <https://doi.org/10.1093/bioinformatics/bts444> PMID: 22796960
86. Barrett JC. Haploview: Visualization and analysis of SNP genotype data. *Cold Spring Harb Protoc*. 2009; 10:pdb.ip71. <https://doi.org/10.1101/pdb.ip71> PMID: 20147036
87. Dong SS, He WM, Ji JJ, Zhang C, Guo Y, Yang TL. LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Brief Bioinformatics*. 2021; 22:bbaa227. <https://doi.org/10.1093/bib/bbaa227> PMID: 33126247
88. Khan H, Yan F, Yan Y, Chen P, Xi R, Ulah I, et al. Genome-wide analysis of evolution and expression profiles of NAC transcription factor gene family in *Juglans regia* L. *Ann Forest Sci*. 2020; 77:82.
89. Hao F, Yang G, Zhou H, Yao J, Liu D, Zhao P, Zhang S. Genome-wide identification and transcriptional expression profiles of transcription factor *WRKY* in common walnut (*Juglans regia* L.). *Genes*. 2021; 12:1444.