frontiers | Frontiers in Plant Science

Genetic analysis of wild walnuts in Xinjiang based on whole-genome resequencing

Yuyu Liu, Ping Zhang*, Zhuli Wang, Jilong Fu, Jing Han, Yihang Gao and Zhibo Feng

College of Forestry and Landscape Architecture, Xinjiang Agricultural University, Urumqi, China

Introduction: As a treasured wild plant resource in the Tian shan Mountains, the genetics and evolutionary relationships of Xinjiang wild walnuts (*Juglans regia* L.) are of great interest for both walnut conservation and crop improvement.

Methods: In this study, a total of 200 walnut accessions, including a core germplasm collection of wild walnuts from Xinjiang and local walnut landraces and cultivars, were selected for whole-genome resequencing, with the final dataset supplemented with 24 other publicly available genomic datasets for other walnut taxa.

Results: Across all samples, there was evidence of four ancestral genetic populations, with three of these represented in the samples from Xinjiang. The Xinjiang wild walnuts form an independent evolutionary clade with low genetic diversity, which was further differentiated into six subgroups, and showed significant genetic differentiation from the cultivated accession. The walnut cultivars and landraces showed mixed ancestry, being assigned to two ancestral populations not represented in the wild walnuts. The Gongliu Wild Walnut Valley served as one of the refugia during the Last Glacial Maximum (LGM) for Tertiary relict species. The unique topography of the Ili River Valley in Xinjiang, along with the relatively isolated geographical location of the Walnut Valley, may have collectively facilitated the formation of a relatively isolated "genetic island" pattern in the Xinjiang wild walnuts. Selective sweep analysis identified 20 genes under selection, including *CYP450* genes closely associated with disease resistance and *NF-YB3* genes involved in cold stress and other adaptive responses.

Discussion: A new framework is needed to reconceptualize the genetic relationships of Xinjiang wild walnuts with other germplasms, clarifying their continuous role throughout the evolutionary continuum from glacial refugium to domestication and modern breeding.

1 Introduction

Walnuts (*Juglans regia* L.) are perennial deciduous trees in the Juglandaceae family, valued as an important economic crop. Notably, their seed kernels contain a high oil content of 60%–70%, earning them the designation "oil-rich tree (Zhao et al., 2018; Shi et al., 2022). Beyond their economic value, walnuts are rich in ω-3 fatty acids, antioxidants, and high-quality proteins, which have potential beneficial roles in health maintenance and disease prevention (Li, 2007). Xinjiang is an important walnut producing area in China, with a unique ecology and rich walnut genetic resources. As a widely cultivated nut crop, the evolutionary history and genetic diversity of cultivated walnuts have been studied in depth; however, few studies to date have systematically examined the population genetic structure of wild walnuts in Xinjiang, or how the wild walnuts are related to other walnut germplasm.

Xinjiang wild walnut (*Juglans regia* L.) is a valuable Tertiary-relict temperate broadleaf forest species (Wei et al., 2023) that is found primarily in the Wild Walnut Valley Nature Reserve in Gongliu County, Xinjiang, China (Zeng, 2005). During the last glacial period, wild walnuts survived in small refuges located across the Tianshan Mountains to Ferghana Ridge and Southern Kazakhstan, a region spanning 30°N to 45°N (Aradhya et al., 2017). Northern Xinjiang therefore likely contained walnut refugia, and extant wild walnut populations in Gongliu are also located within this geographical region. As such, the Xinjiang walnut populations likely contain unique germplasm and are of particular interest to researchers (Ye et al., 2024). To date, several phenotypic analyses (Wang et al., 1997, 1998; Zhang et al., 2011, 2012, 2013, 2014) and molecular marker-based genetic analyses (Li et al., 2010; Rui, 2010; Wang et al., 2015) have assessed the diversity of Xinjiang germplasm. The study of population genetic structure is essential for analyzing adaptive evolution and genetic relationships in different walnut populations (Gunn et al., 2010). However, very few published studies examining the evolutionary history of walnut have included Xinjiang wild walnut (Ding et al., 2022). In those that have, so few samples of wild Xinjiang materials were included that the studies are not representative of the wild population as a whole, conclusions based on this small sample must be considered tentative (Luo et al., 2022), and estimates of its genetic diversity may be inaccurate. Some studies have suggested that wild walnut in Xinjiang is the direct ancestor of cultivated walnut (Lin et al., 1984; Wang et al., 2015). In addition, Xinjiang wild walnut has been variously described as *Juglans regia* L (Han et al., 2024), *Juglans cathayensis* Dode (Zhang, 2016), and *Juglans fallax* Dode (Zeng, 2005). The taxonomic status of wild walnut in Xinjiang is therefore poorly resolved, and additional genetic studies of the Xinjiang germplasm, evaluating its relationship to other walnut taxa, are urgently needed. Until such studies are performed, the unique and valuable wild walnut resources in Xinjiang will remain poorly utilized and protected.

The study of genetic diversity and genetic structure of wild walnut is important for the conservation and utilization of walnut species (Li et al., 2024). In this study, 200 walnut accessions collected from the Ili, Aksu, Kashgar and Hotan regions of Xinjiang were subjected to whole-genome resequencing. The samples included 79 wild walnut accessions, 103 walnut landraces, and 18 cultivated walnut varieties; additional genomic data for 24 other walnut taxa were obtained for a larger comparative analysis. This study provides essential genomic data for Xinjiang walnut, allowing the elucidation of relationships among wild walnut populations within Xinjiang itself, as well as between wild walnuts and other walnut germplasm.

2 Materials and methods

2.1 Plant materials and DNA extraction

Plant tissue samples were collected from natural walnut forests in the Wild Walnut Valley Nature Reserve located in Gongliu County, Xinjiang, China, with the coordinates of 43°19'N-43°23'N, 82°15'E-82°17'E, and the altitude of 1250–1700 m (Figure 1). The total area of the reserve is about 10.19 km², and there are more than 5,500 mature trees of wild walnut. Building on previous work by our project team that used SSR and SRAP markers to analyze genetic diversity and establish a core germplasm collection for wild walnuts in Xinjiang, this study systematically assembled a representative set of 79 accessions, including accessions with cold and disease resistance (Yuan, 2012; Yuan et al., 2012; Zhang, 2015; Zhang et al., 2018; Wen et al., 2022, 2023). In addition, samples were collected from 103 individual trees belonging to ancient walnut landraces in Xinjiang, including 59 accessions from Hotan (HTMF, HTMY, HTPS, and HTX), 30 accessions from Kashgar (KSYC), and 14 accessions from Aksu (AK18-AK31), as well as from 18 cultivated walnut varieties local to Xinjiang (AK1-AK17 and AK32). Healthy, fresh young leaves were collected from each accession (n = 200 in total) and immediately placed in liquid nitrogen after collection.

The final dataset was supplemented by 24 publicly available walnut genomic datasets which were obtained to more comprehensively analyze phylogenetic relationships within walnut. The publicly available data comprised eight genomes published by Zhang et al. in 2019 (Zhang et al., 2019a) and Cao et al. in 2023 (Cao et al., 2023): *Juglans cathayensis* (n = 3), *Juglans mandshurica* (n = 2), *Juglans sigillata* (n = 2), and *Juglans hopeiensis* (n = 1). These were supplemented by 15 genomes published by Ji et al. in 2021 (Ji et al., 2021): JR2 (n = 9), JR3 (n = 4), and *Juglans sigillata* (n = 2) (see Supplementary Table S1 in the Supplementary Materials for an overview of all walnut samples included in the study). The geographic distribution of all sample locations was illustrated using ArcGIS, and elevation data for each location were downloaded from a geospatial data cloud website.

DNA was extracted from the leaf tissue samples using a CTAB-based method (Doyle and Doyle, 1987). The quality of the isolated genomic DNA was verified through both the use of agarose gels and a NanoDrop spectrophotometer, as follows: (1) DNA degradation and contamination were assessed on 1% agarose gels; and (2) DNA concentrations were measured with a ND-2000 (NanoDrop Technologies). Only high-quality DNA samples ($OD_{260/280}$ = 1.8-2.0, $OD_{260/230} \geq 2.0$) were used to construct sequencing libraries.
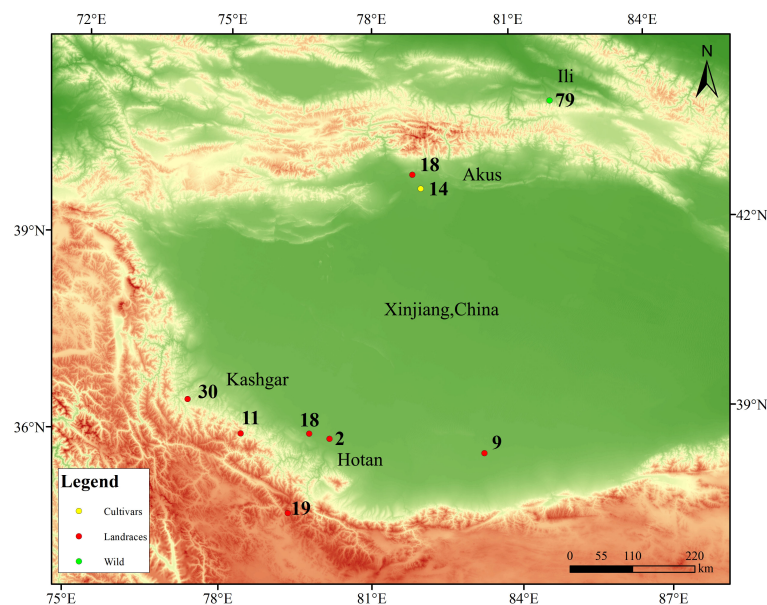
FIGURE 1
Sample site distribution map. Green circles represent wild walnuts, red circles represent landraces, and orange circles represent cultivars.

2.2 DNA sequencing and data quality controls

A total of 0.5 μg of DNA per sample was used as input material for DNA library preparation. Sequencing libraries were generated using a Truseq Nano DNA HT Sample Prep Kit (Illumina USA), with individual barcodes added to each sample. Briefly, genomic DNA samples were fragmented via sonication to a size of 350 bp. The resulting DNA fragments were end-polished, A-tailed, and ligated with full-length adapters for Illumina NovaSeq X Plus sequencing, followed by further PCR amplification. After purification of the PCR products [AMPure XP system (Petrova and Angel, 2024)], sequencing libraries were analyzed to assess their size distribution using an Agilent 2100 Bioanalyzer and quantified using real-time PCR (3nM). The final paired-end libraries were sequenced on an Illumina NovaSeq X Plus system by the Shanghai Majorbio Bio-Pharm Technology Co., Ltd.

The raw sequencing data contained both low-quality base calls and missing values, which would greatly interfere with subsequent data analyses. To obtain clean reads, the following filters were applied. Raw reads of low quality (mean phred score< 20), including reads containing adapter fragments and/or unrecognizable nucleotides (N base > 10), were trimmed or discarded using fastp (Chen et al., 2018).

2.3 SNP calling and filtering

To generate BAM files, quality-filtered reads were mapped to the walnut reference genome using *BWA-MEM* (Jung and Han, 2022) with default mapping parameters, the version of the reference genome used was *Juglans regia* Version 1.0 (Zhang et al., 2020). As a modified GATK Best Practice (McKenna et al., 2010), the alignment (BAM) files were

sorted using samtools (Li et al., 2009) and PCR duplicates marked with MarkDuplicates. After base quality recalibration, variant calling (for both SNPs and InDels) was performed across all samples using the Haplotyper and Gvcftyper algorithms in the Sentieon Genomics Tools package (Freed et al., 2017). Variants were filtered using standard hard filtering parameters following the GATK Best Practices pipeline (McKenna et al., 2010). Variant functional annotation was performed for the quality-filtered SNPs using *SnpEff* (Cingolani et al., 2012) in combination with gene predictions from the walnut reference genome. Both SNPs and InDels were categorized based on their chromosomal positions (i.e., intergenic regions [including 1-kb upstream and downstream], exons, introns, splice sites, and untranslated regions) and on their effects (i.e., missense, splicing, start codon gain/loss, and stop codon gain/loss mutations). Prior to further analysis, several SNP filtering steps were performed to reduce false positives for genotype calling: (i) SNPs with more than two alleles were removed; (ii) SNPs with a mean depth value over all samples of less than four were removed; (iii) SNPs with a minor allele frequency< 0.05 were removed; (iv) only SNPs that could be genotyped in at least 70% of the samples were retained; and (v) linkage disequilibrium (LD) pruning was performed in *PLINK* prior to population structure analyses (Purcell et al., 2007). Using the final SNP set, further analyses were carried out to assess walnut genetic diversity, kinship parameters, linkage disequilibrium, and population structure, among other investigations.

2.4 Phylogenetic tree and population structure analysis

Maximum likelihood (ML) and neighbor-joining (NJ) phylogenetic trees were constructed using *IQ-TREE 2* (model

GTR+I+G4; 1,000 bootstraps) (Minh et al., 2020) and *FastTree* (model -gtr -gamma; 1,000 bootstraps) (Price et al., 2010), respectively, based on the LD-pruned SNP set. An unsupervised ML clustering algorithm, as implemented in *ADMIXTURE* (Alexander et al., 2009) was used to estimate shared ancestry in the study walnut populations. Initial clustering was performed for K = 1 to K = 20 ancestral groups with default settings. To maximize the accuracy of the initial clustering, LD-pruned SNPs were used for all structure analyses.

2.5 Linkage disequilibrium and genetic diversity analyses, genetic differentiation analysis

To evaluate LD decay across the walnut genome, the squared correlation coefficients (r^2) between pairs of SNPs were calculated using *PopLDdecay* (Zhang et al., 2019b). The average r^2 value was calculated for each pair of SNPs within a 500 kb region and then averaged across the genome. To comprehensively assess the genetic diversity and population structure of the study walnut populations, Stacks (Catchen et al., 2011) was used to calculate population-based divergence indices. The genetic diversity of each locus was estimated based on parameters such as the expected heterozygosity (He), inbreeding coefficient (Fis), nucleotide diversity (π), observed heterozygosity (Ho), polymorphic information content (PIC), and Shannon diversity index (H). Together, these parameters were used to assess the structure of genetic diversity within and among the walnut populations included in the study, providing insight into the evolutionary history of the Xinjiang wild walnuts. To understand the degree of differentiation between populations, the differentiation index F_{ST} was calculated using a formula. The calculation formula for the F_{ST} of a single locus is as follows:

$$\hat{\theta} = \frac{s^2 - \frac{1}{2\bar{n}-1}\left[\bar{p}(1-\bar{p}) - \frac{r-1}{r}s^2\right]}{\left[1 - \frac{2\bar{n}C^2}{(2\bar{n}-1)r}\right]\bar{p}(1-\bar{p}) + \left[1 + \frac{2\bar{n}(r-1)C^2}{(2\bar{n}-1)r}\right]\frac{s^2}{r}}$$

Here, $\bar{n}$ represents the average sample size per population, C is the coefficient of variation in sample sizes across populations, $\bar{p}$ is the mean allele frequency across all populations, s^2 denotes the variance in allele frequencies among populations, and $\hat{\theta}$ denotes Weir & Cockerham's unbiased estimator of F_{ST}. The value of F_{ST} ranges between [0, 1], with values closer to 1 indicating greater genetic differentiation among populations.

2.6 Principal component analysis and genetic relatedness analysis, selective sweep analysis

To visualize the genetic relationships among samples, a principal component analysis (PCA) was performed based on the LD-pruned SNP set using *PLINK* (Purcell et al., 2007). *PLINK* was also used to estimate degrees of kinship (Manichaikul et al., 2010) between all study individuals based on pairwise SNP comparisons.

Using a sliding window approach, genetic diversity (π) and genetic differentiation indices (F_{ST} and d_{XY}) were calculated using PIXY (Korunes and Samuk, 2021) for windows of 10 kb; measures of selection (i.e., Tajima's *D*) were calculated using VCFtools (Danecek et al., 2011). Additionally, *ChiPlot* (https://www.chiplot.online/) (Xie et al., 2023) was used to create a heat map illustrating genetic differentiation (F_{ST}) among the various walnut populations. Only the window with average value above the threshold of the 95% confidence interval was considered as the candidate selected region.
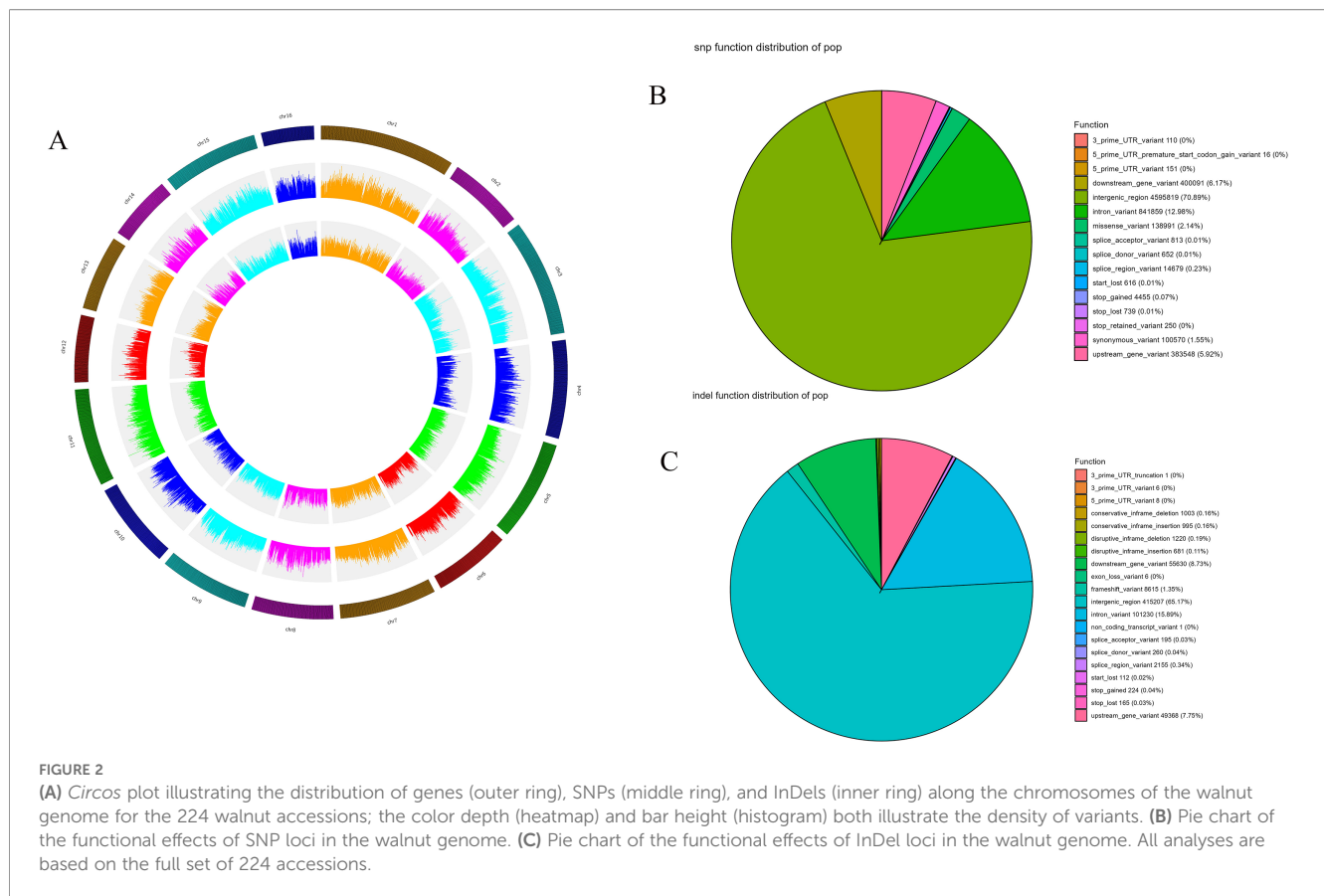
3 Results

3.1 Sequencing data overview and variant identification

To understand the genetic relationships among wild and cultivated walnuts in Xinjiang, China, at total of 79 wild walnut accessions, 103 walnut landraces, and 18 cultivated walnut varieties were selected for whole-genome resequencing (350 bp, paired-end reads). After filtering the raw sequencing data, a total of 1,676.33 Gb of high-quality, clean data was obtained. In terms of sequencing quality, the Q30 value averaged 96.22%, ranging from 85.25% to 100% for individual samples. The average GC content was 37.36%, ranging from 36.38% to 41.35% for individual samples. On a per sample basis, the amount of raw data ranged from a minimum of 5.68 Gb (AK15) to a maximum of 26.59 Gb (JC3), with an average efficiency of 99.05%. Following alignment to the reference genome, the average sequencing depth was 14.21 across all samples; on average, 94.26% of the genome was covered at 1X and 86.17% was covered at 4X. More than 95% of the samples had a match success rate of 98%. This rate reflects how similar the sample sequencing data are to the reference genome; the coverage and sequencing depth may also directly reflect homology with the reference sequence. Here, similarity to the reference genome was sufficient to allow the accurate identification of variants and meaningful assessments of the Xinjiang walnut population.

A total of 30,477,372 SNPs were identified across the walnut genome. After filtering, 5,679,611 high-quality SNPs were retained for analysis, as well as 5,905,456 insertion-deletion polymorphisms (InDels). Among the InDels, 3,135,605 were insertions and 2,769,851 were deletions. All variants were mapped using *Circos*, as illustrated in Figure 2, which visually presents the distributions of SNPs, InDels, and other variants across the chromosomes of the walnut genome. Variant-rich regions may contain important gene functions and/or be subject to natural selection.

3.2 Genetic diversity and linkage disequilibrium in genomes, genetic differentiation analysis

Nucleotide diversity (π) was calculated to understand patterns of genetic diversity in the walnut germplasm from Xinjiang. The

FIGURE 2
(A) *Circos* plot illustrating the distribution of genes (outer ring), SNPs (middle ring), and InDels (inner ring) along the chromosomes of the walnut genome for the 224 walnut accessions; the color depth (heatmap) and bar height (histogram) both illustrate the density of variants. **(B)** Pie chart of the functional effects of SNP loci in the walnut genome. **(C)** Pie chart of the functional effects of InDel loci in the walnut genome. All analyses are based on the full set of 224 accessions.
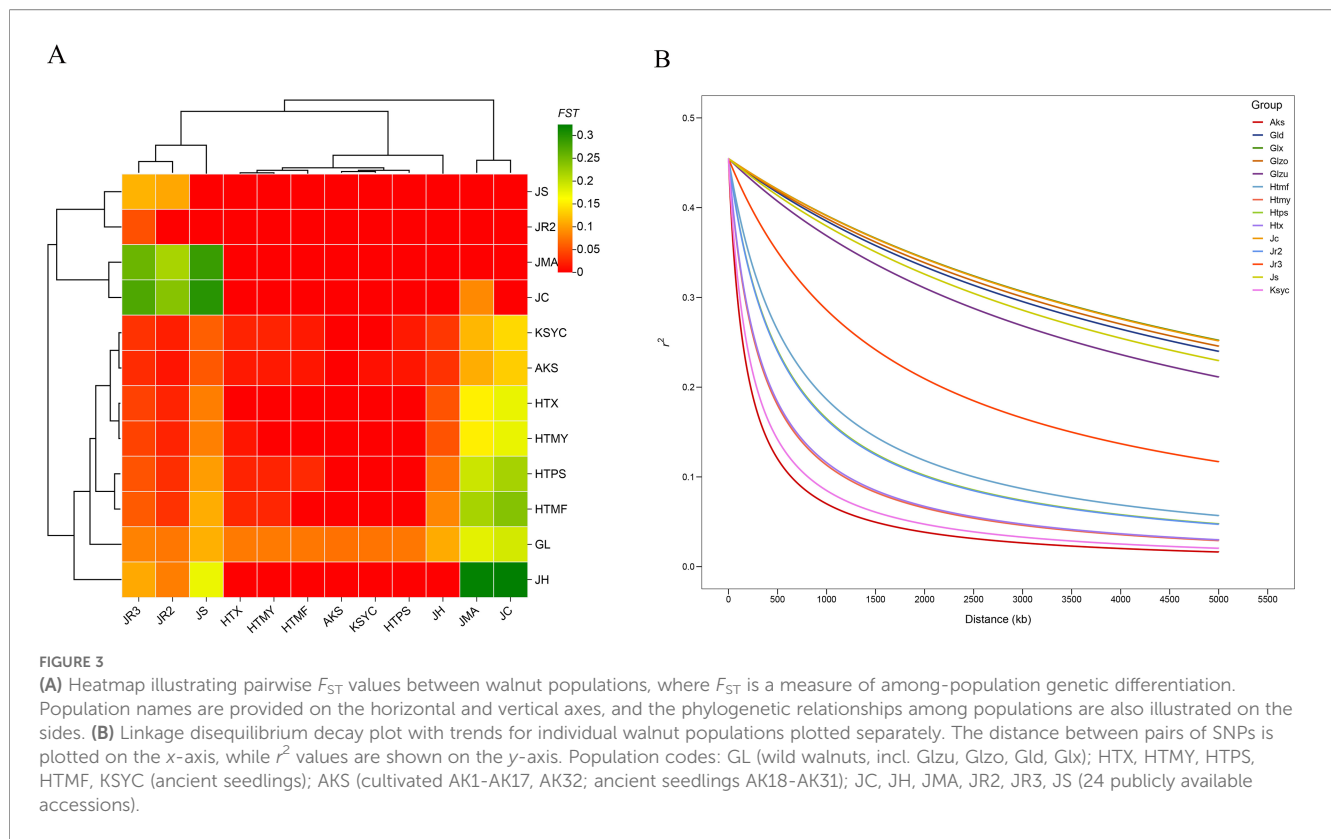
nucleotide diversity of all walnut samples measured 0.286 at the whole-genome level. The lowest nucleotide diversity was found in the wild walnut population from Xinjiang (0.252), which was significantly lower than that of walnut landraces (0.308) and cultivated walnut (0.305) (Table 1); the highest nucleotide diversity occurred in the Hotan landraces. Similarly, the Shannon diversity index was highest in cultivated walnut (0.457), followed by the landraces (0.447) and wild walnut population (0.371) (the genetic diversity indices of the populations are presented in Supplementary Table 2). Differences in SNP and InDel frequencies revealed genetic differentiation between wild and cultivated walnut samples from Xinjiang. The population segregation (F_{ST}) index measures among-population differentiation by assessing allele frequency differences between populations. Pairwise F_{ST} values between wild walnuts and landraces were higher, demonstrating greater genetic differentiation than that between cultivated walnuts and landraces, based on both SNP and InDel datasets (the F_{ST} values between clusters are provided in Supplementary Table 3). Looking more broadly across all walnut populations, the Xinjiang wild walnuts were the most genetically distinct, showing substantial differentiation from both landraces (average F_{ST} = 0.076) and cultivated walnuts (F_{ST} = 0.072). Conversely, walnut cultivars and landraces were the least differentiated (F_{ST} = 0.014), underscoring their close genetic affinity.

Linkage disequilibrium (LD) is a critical concept in population genetics and evolutionary biology, and the extent of genetic

differentiation among populations, as well as the evolutionary history, genetic diversity, and genome structure of individual populations, may all influence LD levels. In this study, a LD analysis was performed for all walnut populations. The wild population exhibited a slower LD decay rate than other populations (Figure 3B), which corresponded to a considerably larger LD half-decay distance in the wild population (approximately 1 Mb) compared to landraces (173.582 kb) or cultivated walnuts (91.67 kb). The slow decay of LD in Xinjiang wild walnuts might be explained by the geographical proximity of samples. The LD decay rates of cultivated varieties and landraces were similar. In AKS, HTMY, HTPS, HTX, and KSYC, LD decayed rapidly over short physical distances, suggesting that these populations had high recombination rates and genetic diversity. In contrast, in the wild group, LD remained high over long physical distances, suggesting a low recombination rate and lower genetic diversity, perhaps due to bottleneck events and/or geographic isolation.

3.3 Phylogenetic relationships and population structure among wild and cultivated walnuts in Xinjiang

To infer the phylogenetic relationships of wild walnuts in Xinjiang, we constructed a phylogenetic dendrogram based on genetic data from 79 samples (Figure 4A). In the dendrogram, six subpopulations could be distinguished within the wild walnut

FIGURE 3

(A) Heatmap illustrating pairwise F_{ST} values between walnut populations, where F_{ST} is a measure of among-population genetic differentiation. Population names are provided on the horizontal and vertical axes, and the phylogenetic relationships among populations are also illustrated on the sides. **(B)** Linkage disequilibrium decay plot with trends for individual walnut populations plotted separately. The distance between pairs of SNPs is plotted on the x-axis, while r^2 values are shown on the y-axis. Population codes: GL (wild walnuts, incl. Glzu, Glzo, Gld, Glx); HTX, HTMY, HTPS, HTMF, KSYC (ancient seedlings); AKS (cultivated AK1-AK17, AK32; ancient seedlings AK18-AK31); JC, JH, JMA, JR2, JR3, JS (24 publicly available accessions).

population, with samples from the east valley, the main valley, the middle valley, and the west valley forming four unique, independent subpopulations. The other two subpopulations were mostly comprised of samples from intersection zones between the valleys; one subpopulation included samples from the intersection of the east, main, and middle valleys, while the second included samples from the intersection of the east, main, and west valleys. Genetically, walnut trees located in the intersection zones were a mixture of the four valley subpopulations. Comparing subpopulations, the east valley subpopulation localized to the base of the dendrogram and was the first to diverge from the remaining subpopulations; the main, middle, and west valley subpopulations each formed three separate branches. The four core subpopulations likely evolved independently due to topographic barriers (i.e., steep slopes); however, post-glacial expansion led to gene flow within the intersection zones, forming two genetic "transition zones". Individuals within valleys were tightly clustered on the phylogenetic tree, while different valleys were clearly segregated from each other, with valleys acting as genetic isolation barriers at the local scale.

To further explore how Xinjiang wild walnuts are related to other walnut accessions, the neighbor-joining method was used to construct a phylogenetic tree based on the SNP set (5,679,611 high-quality SNPs) created for all 224 walnut accessions (Figure 4B). In the dendrogram, walnut cultivars and landraces clustered into one large group that was distinct from that formed by the wild walnuts. Walnut landraces from Kashgar were most closely related to AK10, AK19, AK30, and AK31. Most of the walnut cultivars were most closely related to one another, although a few cultivars clustered

with landraces from Hotan and Kashgar. Landraces from Hotan clustered into a single group which also contained some cultivars (AK1, AK4, AK6, AK7, and AK16). In the tree, the 24 accessions representing other walnut taxa were distantly related to the 200 accessions from Xinjiang, with the exception of the JR2 samples. For example, samples JR2-1, JR2-2, JR2-6, and JR2–9 were closely related to cultivated accessions from the Aksu region of Xinjiang, while JR2-3, JR2-4, JR2-5, JR2-7, and JR2-8 were closely related to walnut landraces from Hotan.

The maximum likelihood estimation of population structure identified four ancestral populations (i.e., K = 4) among the 224 walnut accessions (which comprised walnut cultivars [n = 18], landraces [n = 103], wild individuals [n = 79], and other germplasm [n = 24]) (Figure 4D). The genetic uniqueness of the Xinjiang wild walnuts was unequivocally supported by population structure analysis. At the optimal K = 4 (minimized cross-validation error; Figure 4C), they formed a distinct ancestral group, clearly differentiated from the admixed cultivars and landraces. This pattern of distinctness held true across multiple K-values (for K = 1 to K = 20; Supplementary Figure S2).

3.4 Principal component analysis and kinship analysis

Principal component and kinship analyses confirmed the genetic distinctness of the Xinjiang wild walnuts. In the PCA, the first three principal components collectively explained 20.9% of the genetic variance. The wild walnuts formed a tight, distinct cluster
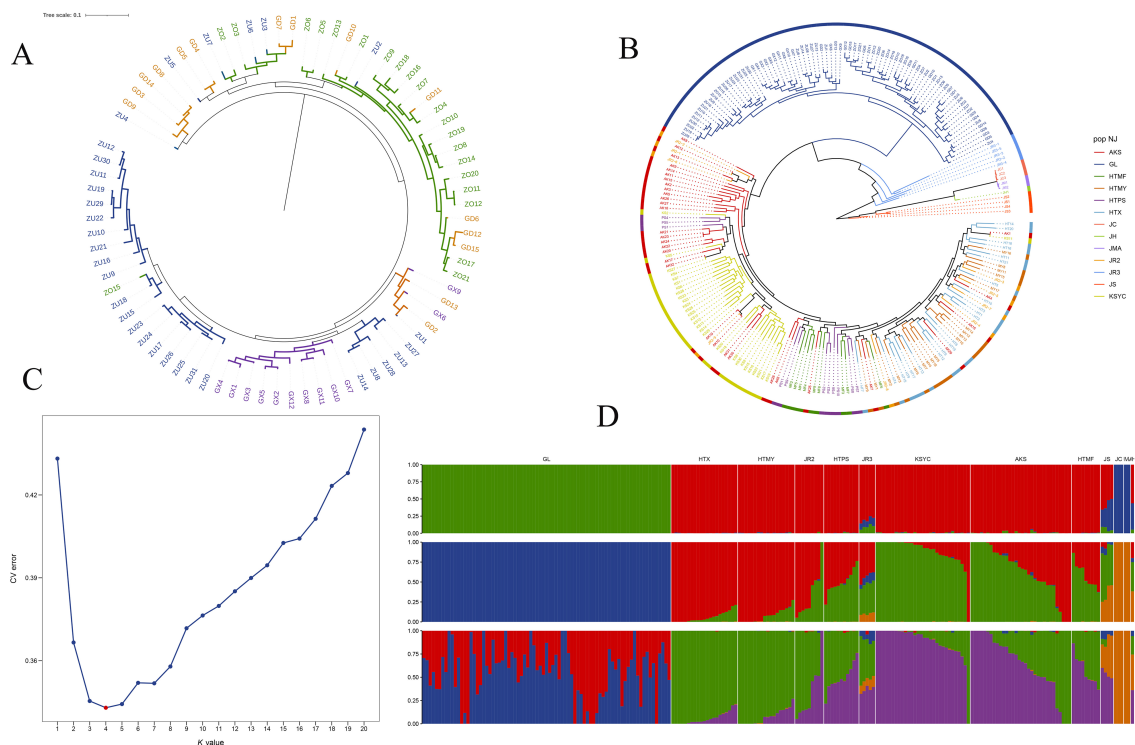
FIGURE 4

(A) Phylogenetic dendrogram of 79 wild walnut samples from Xinjiang. ZU represents the main valley, ZO represents the middle valley, GD represents the eastern valley, and GX represents the western valley. **(B)** Phylogenetic tree of 224 walnut samples constructed using the neighbor-joining method. Each branch represents a sample, with the junction of any two neighboring branches forming a node. The most basal node at the very base of tree is called the root node and represents the inferred ancestor of all extant lineages. Branches representing closely related species are labeled with the same color. **(C)** Cross-validation error for different K-values in a population structure analysis. The optimal K-value, with the smallest error, is marked in red. **(D)** Population structure diagrams (for K = 3, 4, and 5) created using an unsupervised maximum likelihood clustering algorithm in *ADMIXTURE*. Each color represents a distinct genetic cluster, with individuals represented by vertical bars that are subdivided to represent the proportion of an individual's ancestry from each of the K genetic clusters. The optimal K-value was assessed by testing K values of 1 to 20, with K = 4 minimizing the cross-validation error. Population codes: GL (wild walnuts, incl. Glzu, Glzo, Gld, Glx); HTX, HTMY, HTPS, HTMF, KSYC (ancient seedlings); AKS (cultivated AK1-AK17, AK32; ancient seedlings AK18-AK31); JC, JH, JMA, JR2, JR3, JS (24 publicly available accessions).

that was clearly separate from the cultivars, landraces, and the 24 additional accessions (Figure 5A). This clear separation was further supported by kinship analysis, which revealed strong genetic cohesion within the wild group and distant relationships with all other populations (Figure 5B).

3.5 Selective sweep analysis

A selective sweep analysis based on F_{ST} values was performed between the Xinjiang wild and the ancient seedling populations of walnut to identify genes under strong selection (Figure 6A). Using the top 5% F_{ST} as a cutoff, we identified 28 candidate genes that may have contributed to local adaptation. Notably, several core transcription factors and enzyme-encoding genes intimately involved in disease resistance were under selection, indicating an enhanced immune system in the Xinjiang wild population. These included: WRKY transcription factors (Eulgem and Somssich, 2007; Chen et al., 2017) such asWRKY41 (JreChr03G11438), WRKY48 (JreChr01G10474), WRKY29 (JreChr11G11287), WRKY69 (JreChr08G10036), WRKY3

(JreChr03G10701), and WRKY23 (JreChr08G10021), which act as central regulators of plant immunity, extensively participating in salicylic acid and jasmonic acid signaling pathways; MYB transcription factors (Ma and Constabel, 2019; Liu et al., 2024) including MYB1 (JreChr01G13373), MYB4 (JreChr01G12459), MYB39 (JreChr11G10859), andMYB83 (JreChr13G11560), which primarily regulate the phenylpropanoid pathway for the synthesis of lignin and antimicrobial compounds; furthermore, the key jasmonate signaling transcription factor bHLH42 (JreChr08G10317) (Zhang et al., 2024) and several cytochrome P450 genes (Schuler and Werck-Reichhart, 2003; Nelson and Werck-Reichhart, 2011) involved in the synthesis of defense compounds, such as CYP71D8 (JreChr14G10451), CYP72A219 (JreChr11G10191), CYP71AU50 (JreChr06G12173), and CYP71A6 (JreChr12G10259), were also selected.

Furthermore, the analysis also revealed the selection of genes associated with cold stress and abiotic stress response, consistent with adaptation to the native frigid climate. These primarily included: the cold and drought-regulated protein CORA-like (JreChr11G12006) (Jha et al., 2021); the nuclear transcription factor NF-YB3 (JreChr03G13037) (Lin et al., 2024), which
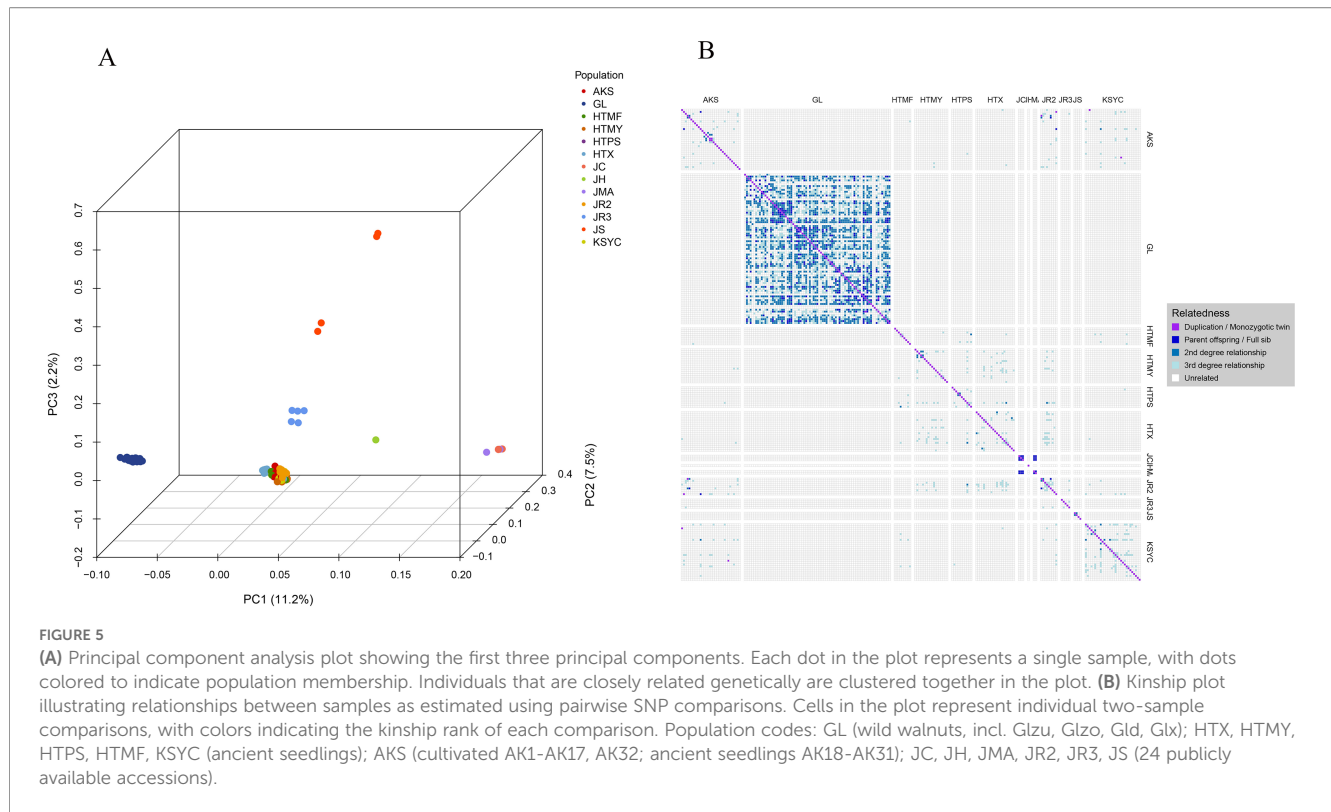
FIGURE 5

(A) Principal component analysis plot showing the first three principal components. Each dot in the plot represents a single sample, with dots colored to indicate population membership. Individuals that are closely related genetically are clustered together in the plot. **(B)** Kinship plot illustrating relationships between samples as estimated using pairwise SNP comparisons. Cells in the plot represent individual two-sample comparisons, with colors indicating the kinship rank of each comparison. Population codes: GL (wild walnuts, incl. Glzu, Glzo, Gld, Glx); HTX, HTMY, HTPS, HTMF, KSYC (ancient seedlings); AKS (cultivated AK1–AK17, AK32; ancient seedlings AK18–AK31); JC, JH, JMA, JR2, JR3, JS (24 publicly available accessions).

enhances tolerance to drought and cold; cytochrome P450 genes (He et al., 2023) involved in hormone homeostasis to fine-tune the growth-defense trade-off, such as CYP94A1 (JreChr13G10978) and CYP714A1 (JreChr08G10054); as well as genes contributing to physical barrier formation, like CYP86A22 (JreChr11G11327) (Han et al., 2010) involved in cuticular wax synthesis and CYP704B1 (JreChr05G10202) (Kobayashi et al., 2021) involved in pollen wall development. These genes collectively constitute a multi-layered protective mechanism for the Xinjiang wild walnut to cope with low-temperature stress.

Based on the F_{ST} analysis between the GL wild and HTX ancient seedling populations, GO enrichment analysis (Figure 6B) revealed significant enrichment in molecular functions, including catalytic activity, acting on a nucleic acid (GO: GO:0140640), ATP-dependent activity (GO:0140657), catalytic activity, acting on DNA (GO:0140097), nucleotidyltransferase activity (GO:0016779), ATP-dependent activity, acting on DNA (GO:0008094), exonuclease activity (GO:0004527), 3'-5' exonuclease activity (GO:0008408), DNA topoisomerase activity (GO:0003916).Among the cellular component category, significant enrichment was observed for the following terms: protein acetyltransferase complex (GO:0031248), acetyltransferase complex (GO:1902493), membrane-bounded organelle (GO:0043227), intracellular membrane-bounded organelle (GO:0043231), nucleus (GO:0005634), membrane-enclosed lumen (GO:0031974), organelle lumen (GO:0043233), intracellular organelle lumen (GO:0070013). These processes likely underlie the adaptive differentiation observed between the GL wild and

HTX ancient seedling populations (comprehensive GO enrichment results are provided in Supplementary Table 4).

To further elucidate the biological functions enriched among genes with significant genetic differentiation (F_{ST}) between the GL wild and HTX ancient seedling populations, KEGG pathway enrichment analysis (Figure 6C) revealed highly significant enrichment in pathways including galactose metabolism, starch and sucrose metabolism, the spliceosome, ubiquitin-mediated proteolysis, nucleocytoplasmic transport, glycolysis/gluconeogenesis, purine metabolism, monobactam biosynthesis, and selenocompound metabolism (detailed results of the KEGG enrichment analysis are provided in Supplementary Table 5).

4 Discussion

The publication of a high-quality walnut genome has facilitated research on related species of economic value (Zhang et al., 2020). Through genome annotation, genes associated with important agronomic traits have been identified in walnut, providing targets for molecular breeding programs aimed at developing varieties with enhanced disease resistance (Dai et al., 2025), cold tolerance (Han et al., 2024), and nut quality (Wang et al., 2022). Annotated genomic datasets have provided insights into gene family evolution and the molecular basis of pest/disease resistance in Juglandaceae species (Yan et al., 2021). As a rare wild plant resource in China, the Xinjiang wild walnut possesses significant ecological and genetic value. Its genome harbors a wealth of
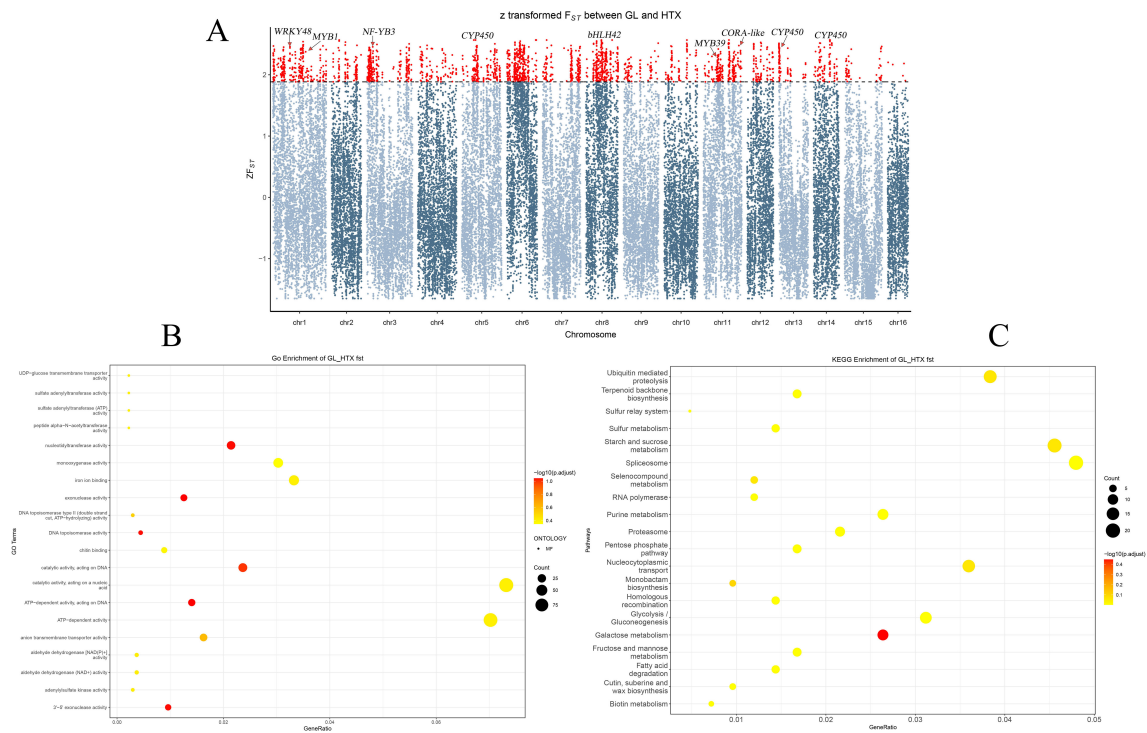
FIGURE 6

(A) Manhattan Plot of Z-Transformed F_{ST} Between Wild (GL) and Landraces (HTX) Walnut Populations. The x-axis represents the genomic coordinates across all 16 chromosomes. The y-axis shows the Z-transformed F_{ST} values, a measure of population genetic differentiation. The dashed horizontal line indicates the significance threshold (Default: top 5%). Red data points above this threshold represent genomic regions with significant selective sweep signals. Several candidate genes discussed in the study (e.g., GRP2, LEA16, NF-YA10, *CYP450*, COR) are highlighted near their respective genomic loci. The candidate genomic regions under positive selection may contribute to the stress resistance observed in the wild walnut (GL population). **(B)** GO enrichment of high F_{ST} Genes between GL and HTX Populations. The x-axis represents the gene ratio. The y-axis indicates the significantly enriched GO terms. Bubble color reflects the enrichment significance level, and bubble size corresponds to the number of candidate genes annotated in each GO term. **(C)** KEGG Enrichment of high F_{ST} Genes between GL and HTX Populations. The x-axis represents the gene ratio; the y-axis shows enriched KEGG pathways. Bubble color indicates enrichment significance, and bubble size corresponds to the number of genes in each pathway.

resistance gene resources (Xinjiang uygur autonomous region nature reserve research team, 1982), as exemplified by the identification of the cold-responsive *JfDREB1A* gene, which has been demonstrated to play a critical role in the species' response to low-temperature stress (Han et al., 2023, 2024). However, compared to cultivated walnut samples, the unique genetic resources contained within Xinjiang wild walnut remain poorly characterized, and further in-depth studies of wild walnut samples are needed. In this study, whole genome resequencing was performed for 200 accessions of Xinjiang wild walnut, and a total of 5,679,611 high-quality SNPs were identified; this dataset represents a valuable resource for future studies of genetic differentiation in walnut.

The Xinjiang wild walnut population has a relatively restricted range within the Tianshan Mountains of China, concentrated in the Wild Walnut Valley Nature Reserve in Gongliu (Wang, 2011). Wild walnut trees are typically not seen in other regions of China, except for a dozen or so plants distributed within Huocheng County. Due to the distance between the Xinjiang population and wild walnuts in Kazakhstan, gene flow is unlikely to occur under natural conditions. It has been hypothesized that the low genetic diversity observed in the Xinjiang wild walnuts may be due to habitat fragmentation,

resulting from environmental changes that led to isolation of the Tianshan Mountain populations. As a core element of biodiversity (Wei et al., 2021), genetic diversity directly determines a population's adaptive potential (Taberlet et al., 2012; Du, 2023) and extinction risk threshold (Wang et al., 2023) Genetic diversity may be affected by anthropogenic disturbance (Lu et al., 2020; Perrino and Wagensommer, 2022), environmental changes (Provan, 2013), founder effects (Pellissier et al., 2016), gene flow (Taberlet et al., 2012), natural selection (Chhatre and Rajora, 2014), and population size (Crutsinger et al., 2008). As a result, the Tianshan populations may have experienced high levels of genetic drift or population bottlenecks, leading to a loss of heterozygosity, increased inter-individual similarity, and reduced overall genetic diversity (Allendorf et al., 2024). Phylogenetic analysis revealed that the Xinjiang wild walnut samples comprised six subpopulations, each exhibiting generally low genetic diversity. This finding is consistent with previous reports (Zhang et al., 2018; Li et al., 2024; Ye et al., 2024). Here, LD decayed relatively slowly within the wild walnut genomes, and samples were found to be closely related. This suggests the occurrence of historical bottlenecks or persistent small population effects, both of which may lead to heightened kinship among individuals. Frequent inbreeding

TABLE 1 Genetic diversity metrics for 200 walnut accessions from the Xinjiang region in China.

Group	Number of samples	SNPs (N)	SNP/ kb	π	Shannon diversity index
Wild walnuts	79	3,517,987	6.82	0.252	0.371
Landraces	103	3,707,223	7.19	0.308	0.447
Cultivars	18	3,813,489	7.39	0.305	0.457

within wild populations may have also led to declines in genetic diversity and increased homozygosity. Both wind-pollination and asexual reproduction, two features of walnut biology, may increase the likelihood of inbreeding, potentially decreasing the heterozygosity of alleles within a given population (Wang and Dang, 1999; Zhang et al., 2019c).

Population structure and kinship analyses demonstrated that the Xinjiang wild walnut population constitutes a unique and independent genetic group, showing no shared ancestry with other populations. Concordance between structural analysis (based on allele frequencies) and kinship analysis (based on genetic correlations) provides strong support for the finding that the Xinjiang wild walnut population constitutes a distinct genetic lineage. This conclusion of genetic distinctiveness was further supported by the results of the principal component analysis (PCA). It is worth noting that the first three principal components of PCA explained only 20.9% of the total variance cumulatively, which is comparable to the recent high-dimensional resequencing study of walnut (Wang et al., 2022). The low level of inter-valley differentiation results from the dispersion of genetic variance across a large number of genome-wide SNPs, each with a minimal effect, thereby distributing the variation among numerous low-effect loci. The close phylogenetic relationship between JR2 and some Xinjiang cultivated and landrace walnuts suggests they may share a common genetic background. In contrast, the significant differentiation between all 24 publicly available accessions and the Xinjiang wild walnuts further demonstrates that this wild population represents a unique evolutionary lineage, likely maintained in genetic isolation due to geographic barriers and limited gene flow.

Selective sweep analysis highlights genetic adaptations in Xinjiang wild walnuts supporting environmental resilience (Kajino et al., 2022). Strong selection signals in disease-resistant transcription factors such as WRKY (Caarls et al., 2015) and MYB families (Jiao et al., 2025) (e.g., WRKY41 and MYB4) indicate enhanced immune signaling and phenylpropanoid metabolism, significantly improving resistance to pathogens. Meanwhile, genes such as NF-YB3 (Lin et al., 2024) and cytochrome P450 genes (Huang et al., 2024) were under selection, demonstrating molecular adaptations to low-temperature stress. These results underscore the population's unique genetic capacity for stress tolerance, highlighting its potential as a resource for resistance traits. The genetic differentiation observed in the Xinjiang wild walnut population was predominantly driven by biological

functions conferring adaptation to natural environments, while in the cultivated walnut population, it was largely shaped by agronomic traits aligned with human selection, a pattern consistent with their independent evolutionary histories.

In a phylogenetic dendrogram of the Xinjiang samples, the 200 accessions were classified into three groups. The clear genetic separation of the wild walnuts stands in sharp contrast to the significant introgression observed between the ancient seedling walnuts from the Kashgar and Hotan regions. Walnut seeds are primarily dispersed by rodents (Zhang et al., 2017; Lenda et al., 2018), although anthropogenic dispersal also occurs. In walnut, pollination and seed dispersal mechanisms facilitate intra- and inter-population gene flow, thereby reducing genetic differentiation among populations (Nybom, 2004). The lower genetic differentiation indices of ancient solid walnut in the Hotan region as well as in the Kashgar region are inextricably linked to these factors. Habitat fragmentation (due to anthropogenic disturbance) may reduce gene flow and increase inbreeding in many tree species (Manners et al., 2013), but reductions in natural gene flow may be offset by human-mediated dispersal (e.g., of landraces). The remarkable genetic differentiation of wild walnut, consistent with its independent clustering in phylogenetic trees, suggests a unique evolutionary history characterized by long periods of isolation. The differing levels of genetic differentiation observed among these taxa suggest that the previously proposed evolutionary relationships between wild walnuts and various cultivated germplasms may require further validation with additional data.

The walnut ancestor was widely distributed across Eurasia from the Miocene onward (Erdei and Magyari, 2011; Yao et al., 2011; Zhang and Sun, 2011; Ivanov et al., 2012; Shatilova et al., 2014). However, the Last Glacial Maximum (LGM) caused a severe range contraction, forcing the species to persist in fragmented refugia, including Western Europe, Xinjiang, and a broad zone spanning from northeastern to southwestern China (Aradhya et al., 2017; Pollegioni et al., 2017; Feng et al., 2018; Li et al., 2024). Notably, genetic diversity within glacial refugia often declines due to repeated bottlenecks and post-glacial isolation (de Lafontaine et al., 2013; Gao and Gao, 2016). The discontinuous distribution of Xinjiang wild walnut is similar to that of the LGM glacial refuges. Genetic evidence also suggests that the Wild Walnut Valley may have been a naturally occurring glacial refuge (Feng et al., 2018; Li et al., 2024). Post-glacial colonies are known to exhibit lower genetic diversity than refugial populations (Hewitt, 1996, 1999). As population expansion occurs, frontier individuals will be derived from neighboring areas rather than from more distant ones (Hewitt, 1993); this may be one explanation for the low genetic diversity of wild walnuts in Xinjiang.

As humans resided in Xinjiang before the last glacial period (Finlayson, 2010; Stringer, 2012; García-Rodríguez et al., 2024), the genetic structure of contemporary wild walnut populations in Xinjiang cannot be explained solely by geography; anthropogenic factors must also be considered. For example, the flourishing of the Silk Road (French, 1998; Christian, 2000) facilitated extensive trade in walnuts between Asia and Europe (Aubaile, 2012), accelerating

the spread of high-quality cultivated germplasm across the region. Archaeological evidence has dated widespread walnut remains along the Silk Road to the Han Dynasty (202 B.C.-A.D. 220) and earlier periods (Mishra, 2020), suggesting that the initial domestication of walnut predates the formation of transcontinental trade networks, such as the Silk Road. The center of domestication was likely located in western Central Asia or on the Iranian Plateau. Therefore, cultivated samples from Xinjiang (described here) may have been introduced via the Silk Road, with no direct relationship to local wild populations (not involved in domestication). Alternatively, wild walnuts in Xinjiang may represent de-domesticated ferals, originating from ancient cultivated varieties, with natural selection then enhancing cold- and drought-tolerance. Walnut species (and natural populations) have continued to evolve since the time of domestication, despite ongoing human activities (Miller and Gross, 2011).

Combining phylogenetic, population structure, and principal component analyses of 224 walnut samples, Xinjiang wild walnuts were found to be only distantly related to all Xinjiang cultivars and landraces, as well as other tested walnut germplasm. Wild Xinjiang populations have existed for a long time (surviving the last glacial maximum), and therefore may contain unique diversity potentially useful for crop improvement efforts, particularly for climate adaptation. The unique genetic identity of the Xinjiang wild walnut makes it a valuable wild genetic reservoir. Its genome may harbor resistance genes to cold or specific diseases that have been lost in cultivated varieties. This holds significant potential value for future genetic improvement of walnuts using strategies such as selection or hybridization. Thus, future conservation efforts should prioritize the protection of existing wild walnut forests, which serve as a germplasm resource for specific adaptive traits. Wild populations might be best protected by adopting *in situ* conservation measures and utilizing targeted relocation as needed (Ye et al., 2024).

5 Conclusions

In this study, the genetic structure of wild walnut populations and cultivated samples from Xinjiang, China, was analyzed using whole-genome resequencing. The study dataset revealed kinship relationships between Xinjiang wild walnuts and other walnut germplasm, and serves as an important resource for future walnut research and genomics-assisted breeding. In general, wild walnuts from Xinjiang had low genetic diversity and were significantly differentiated from other walnut samples, indicating a distant relationship to other walnut germplasm. However, the study included only a small number of samples of walnuts from outside Xinjiang. To more comprehensively characterize phylogenetic relationships among wild and cultivated walnuts, future studies should include more samples from areas surrounding Xinjiang. The construction of a cross-regional phylogenetic network would help elucidate the origins and evolutionary history of Xinjiang wild walnuts, helping to resolve the unique position of these samples within *Juglans*.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: SRR14888646, SRR5097507, SRR5097508, SRR9313661, SRR9313668, SRR9313665, SRR5192959, SRR5192958, SRR14430138, SRR14430165, SRR1443005, SRR14430040, SRR14430587, SRR14430584, SRR14430588, SRR14430059, SRR14430582, SRR14430540, SRR14430606, SRR14430345, SRR14430324, SRR14430515, SRR14430334, SRR14430326.

Author contributions

YL: Formal analysis, Investigation, Data curation, Software, Writing – review & editing, Methodology, Writing – original draft. PZ: Project administration, Resources, Data curation, Funding acquisition, Supervision, Writing – review & editing. ZW: Formal analysis, Writing – review & editing, Data curation, Investigation. JF: Investigation, Formal analysis, Data curation, Writing – review & editing. JH: Investigation, Writing – review & editing. YG: Investigation, Writing – review & editing. ZF: Investigation, Writing – review & editing.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2025.1645319/full#supplementary-material

References

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109

Allendorf, F. W., Hössjer, O., and Ryman, N. (2024). What does effective population size tell us about loss of allelic variation? *Evol. Appl.* 17:e13733. doi: 10.1111/eva.13733

Aradhya, M., Velasco, D., Ibrahimov, Z., Toktoraliev, B., Maghradze, D., Musayev, M., et al. (2017). Genetic and ecological insights into glacial refugia of walnut (*Juglans regia* L.). *PloS One* 12, e0185974. doi: 10.1371/journal.pone.0185974

Aubaile, F. (2012). Pathways of diffusion of some plants and animals between Asia and the Mediterranean region. *Rev. d'ethnoécologie*. 1. doi: 10.4000/ethnoecologie.714

Caarls, L., Pieterse, C. M. J., and Van Wees, S. C. M. (2015). How salicylic acid takes transcriptional control over jasmonic acid signaling. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00170

Cao, Y., Almeida-Silva, F., Zhang, W. P., Ding, Y. M., Bai, D., Bai, W. N., et al. (2023). Genomic Insights into Adaptation to Karst Limestone and Incipient Speciation in East Asian Platycarya spp. (Juglandaceae). *Mol. Biol. Evol.* 40, msad121. doi: 10.1093/molbev/msad121

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011). Stacks: Building and genotyping loci *de novo* from short-read sequences. *G3: Genes Genomes Genet.* 1, 171–182. doi: 10.1534/g3.111.000240

Chen, F., Hu, Y., Vannozzi, A., Wu, K., Cai, H., Qin, Y., et al. (2017). The WRKY transcription factor family in model plants and crops. *Crit. Rev. Plant Sci.* 36, 311–335. doi: 10.1080/07352689.2018.1441103

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Chhatre, V. E., and Rajora, O. P. (2014). Genetic divergence and signatures of natural selection in marginal populations of a keystone, long-lived conifer, eastern white pine (Pinus strobus) from Northern Ontario. *PloS One* 9, 1–13. doi: 10.1371/journal.pone.0097291

Christian, D. (2000). Silk roads or steppe roads? The silk roads in world history. *J. World History* 11, 1. doi: 10.1353/jwh.2000.0004

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 6, 80–92. doi: 10.4161/fly.19695

Crutsinger, G. M., Souza, L., and Sanders, N. J. (2008). Intraspecific diversity and dominant genotypes resist plant invasions. *Ecol. Lett.* 11, 16–23. doi: 10.1111/j.1461-0248.2007.01118.x

Dai, W., Li, Y., Chen, Z., He, F., Wang, H., Peng, J., et al. (2025). Gibberellin Regulates *LBD38-1* Responses to *Xanthomonas arboricola pv. juglandis* Infection in Walnut Bacterial Blight Pathogenesis. *BMC Genomics* 26, 370. doi: 10.1186/s12864-025-11518-9

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

De Lafontaine, G., Ducousso, A., Lefèvre, S., Magnanou, E., and Petit, R. J. (2013). Stronger spatial genetic structure in recolonized areas than in refugia in the European beech. *Mol. Ecol.* 22, 4397–4412. doi: 10.1111/mec.12403

Ding, Y.-M., Cao, Y., Zhang, W.-P., Chen, J., Liu, J., Li, P., et al. (2022). Population-genomic analyses reveal bottlenecks and asymmetric introgression from Persian into iron walnut during domestication. *Genome Biol.* 23, 145. doi: 10.1186/s13059-022-02720-z

Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical bulletin*.

Du, F. (2023). Conservation of genetic diversity, the fundamental for integrity and authenticity of terrestrial natural ecosystems in national parks. *Natl. Park* 1, 27–33. doi: 10.20152/j.np.2023.01.004

Erdei, B., and Magyari, E. (2011). Late Miocene plant remains from Bükkábrány, Hungary. *Stud. Bot. Hung* 42, 135–151.

Eulgem, T., and Somssich, I. E. (2007). Networks of WRKY transcription factors in defense signaling. *Studia Botanica Hungarica* 10, 366–371. doi: 10.1016/j.pbi.2007.04.020

Feng, X. J., Zhou, H. J., and Saman, Z. (2018). The phytogeographic history of common walnut in China. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01399

Finlayson, C. (2010). *The humans who went extinct: Why Neanderthals died out and we survived* (Oxford University Press).

Freed, D., Aldana, R., Weber, J. A., and Edwards, J. S. (2017). The Sentieon Genomics Tools - A fast and accurate solution to variant calling from next-generation sequence data. *BioRxiv*, 115717. doi: 10.1101/115717

French, D. (1998). Pre- and Early-Roman roads of Asia Minor: The persian royal road. *Iran* 36, 15–43. doi: 10.2307/4299973

Gao, L.-Z., and Gao, C.-W. (2016). Lowered Diversity and Increased Inbreeding Depression within Peripheral Populations of Wild Rice Oryza rufipogon. *PloS One* 11, e0150468. doi: 10.1371/journal.pone.0150468

Garcia-Rodríguez, O., Hardouin, E. A., Pedreschi, D., Richards, M. B., Stafford, R., Searle, J. B., et al. (2024). Contrasting patterns of genetic diversity in european mammals in the context of glacial refugia. *Diversity (Basel)* 16, 1–13. doi: 10.3390/d16100611

Gunn, B. F., Aradhya, M., Salick, J. M., Miller, A. J., Yongping, Y., Lin, L., et al. (2010). Genetic variation in walnuts (*Juglans regia* and *J. sigillata* ; Juglandaceae): Species distinctions, human impacts, and the conservation of agrobiodiversity in Yunnan, China. *Am. J. Bot.* 97, 660–671. doi: 10.3732/ajb.0900114

Han, J., Clement, J. M., Li, J., King, A., Ng, S., and Jaworski, J. G. (2010). The cytochrome P450 *CYP86A22* is a fatty acyl-coA ω-hydroxylase essential for estolide synthesis in the stigma of petunia hybrida. *J. Biol. Chem.* 285, 3986–3996. doi: 10.1074/jbc.M109.050765

Han, L., Luo, X., Zhao, Y., Li, N., Xu, Y., and Ma, K. (2024). A haplotype-resolved genome provides insight into allele-specific expression in wild walnut (*Juglans regia* L.). *Sci. Data* 11, 278. doi: 10.1038/s41597-024-03096-4

Han, L., Zhang, J., Zhao, Y., Mei, C., and Ma, K. (2023). Cloning and prokaryotic expression analysis of *jfDREB1A* gene from xinjiang wild walnut. *J. Northwest Agric.* 32, 1050–1057. doi: 10.7606/j.issn.1004-1389.2023.07.008

He, S., Liu, M., Chen, W., Bai, D., Liao, Y., Bai, L., et al. (2023). Eleusine indica cytochrome P450 and glutathione S-transferase are linked to high-level resistance to glufosinate. *J. Agric. Food Chem.* 71, 14243–14250. doi: 10.1021/acs.jafc.3c04325

Hewitt, G. M. (1993). Postglacial distribution and species substructure: lessons from pollen, insects and hybrid zones. *Evolutionary patterns processes* 14, 97–123.

Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciatio. *Biol. J. Linn. Soc.* 58, 247–276. doi: 10.1111/j.1095-8312.1996.tb01434.x

Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biol. J. Linn. Soc.* 68, 87–112. doi: 10.1111/j.1095-8312.1999.tb01160.x

Huang, H., Wang, Y., Yang, P., Zhao, H., Jenks, M. A., Lü, S., et al. (2024). The Arabidopsis cytochrome P450 enzyme *CYP96A4* is involved in the wound-induced biosynthesis of cuticular wax and cutin monomers. *Plant J.* 118, 1619–1634. doi: 10.1111/tpj.16701

Ivanov, D., Utescher, T., Ashraf, A. R., Mosbrugger, V., Bozukov, V., Djorgova, N., et al. (2012). Late miocene palaeoclimate and ecosystem dynamics in Southwestern

Bulgaria a study based on pollen data from the Gotse-delchev basin. *Turkish J. Earth Sci.* 21, 187–211. doi: 10.3906/yer-1004-45

Jha, R. K., Patel, J., Patel, M. K., Mishra, A., and Jha, B. (2021). Introgression of a novel cold and drought regulatory-protein encoding CORA-like gene, *SbCDR*, induced osmotic tolerance in transgenic tobacco. *Physiologia Plantarum* 172, 1170–1188. doi: 10.1111/ppl.13280

Ji, F., Ma, Q., Zhang, W., Liu, J., Feng, Y., Zhao, P., et al. (2021). A genome variation map provides insights into the genetics of walnut adaptation and agronomic traits. *Genome Biol.* 22, 1–22. doi: 10.1186/s13059-021-02517-6

Jiao, L., Yuan, S., Zhang, L., Shi, Z., Wang, X., Ma, M., et al. (2025). Methyl jasmonate-induced postharvest disease resistance in Agaricus bisporus is synergistically regulated by *AbbZIP2* and *AbMYB11* to facilitate ROS balance. *Postharvest Biol. Technol.* 228, 113667. doi: 10.1016/j.postharvbio.2025.113667

Jung, Y., and Han, D. (2022). BWA-MEME: BWA-MEM emulated with a machine learning approach. *Bioinformatics* 38, 2404–2413. doi: 10.1093/bioinformatics/btac137

Kajino, T., Yamaguchi, M., Oshima, Y., Nakamura, A., Narushima, J., Yaguchi, Y., et al. (2022). *KLU/CYP78A5*, a cytochrome P450 monooxygenase identified via fox hunting, contributes to cuticle biosynthesis and improves various abiotic stress tolerances. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.904121

Kobayashi, K., Akita, K., Suzuki, M., Ohta, D., and Nagata, N. (2021). Fertile Arabidopsis cyp704b1 mutant, defective in sporopollenin biosynthesis, has a normal pollen coat and lipidic organelles in the tapetum. *Plant Biotechnol.* 38, 109–116. doi: 10.5511/plantbiotechnology.20.1214b

Korunes, K. L., and Samuk, K. (2021). pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* 21, 1359–1368. doi: 10.1111/1755-0998.13326

Lenda, M., Knops, J. H., Skórka, P., Moroń, D., and Woyciechowski, M. (2018). Cascading effects of changes in land use on the invasion of the walnut *Juglans regia* in forest ecosystems. *J. Ecol.* 106, 671–686. doi: 10.1111/1365-2745.12827

Li, C., Luo, S. P., Qin, W., Zeng, B., and Shan, J. H. (2010). ISSR analysis of the kinship of walnut in Xinjiang. *Xinjiang Agric. Sci.* 47, 1722–1727.

Li, G. H. (2007). *Research on walnut germplasm resources* (Sichuan Agricultural University).

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, X., Wang, X., Zhang, D., Huang, J., Shi, W., and Wang, J. (2024). Historical spread routes of wild walnuts in Central Asia shaped by man-made and nature. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1394409

Lin, C., Lan, C., Li, X., Xie, W., Lin, F., Liang, Y., et al. (2024). A pair of nuclear factor Y transcription factors act as positive regulators in jasmonate signaling and disease resistance in Arabidopsis. *J. Integr. Plant Biol.* 66, 2042–2057. doi: 10.1111/jipb.13732

Lin, P., Lin, D., and Wang, L. (1984). Wild relatives of fruit trees in Xinjiang. *J. Bayi Agric. Coll.* 25–32. doi: 10.20088/j.cnki.jxau.1984.04.004

Liu, J., Wang, Z., Chen, B., Wang, G., Ke, H., Zhang, J., et al. (2024). Expression analysis of the *R2R3-MYB* gene family in upland cotton and functional study of *GhMYB3D5* in regulating Verticillium wilt resistance. *J. Integr. Agric.* 23, 3294–3310. doi: 10.1016/j.jia.2024.07.040

Lu, Y., Zhang, C., Li, X., Liang, Y., Wang, Y., and Li, W. (2020). Development of EST-SSR markers and their application in the analysis of the genetic diversity of Sophora japonica Linn. *Trees - Structure Funct.* 34, 1147–1156. doi: 10.1007/s00468-020-01985-w

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, 1–19. doi: 10.1371/journal.pgen.1010513

Ma, D., and Constabel, C. P. (2019). MYB repressors as regulators of phenylpropanoid metabolism in plants. *Trends Plant Sci.* 24, 275–289. doi: 10.1016/j.tplants.2018.12.003

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., and Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics* 26, 2867–2873. doi: 10.1093/bioinformatics/btq559

Manners, V., Kumaria, S., and Tandon, P. (2013). SPAR methods revealed high genetic diversity within populations and high gene flow of Vanda coerulea Griff ex Lindl (Blue Vanda), an endangered orchid species. *Gene* 519, 91–97. doi: 10.1016/j.gene.2013.01.037

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., and Cibulskis, K. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110.20

Miller, A. J., and Gross, B. L. (2011). From forest to field: Perennial fruit crop domestication. *Am. J. Bot.* 98, 1389–1414. doi: 10.3732/ajb.1000522

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015

Mishra, R. K. (2020). The 'Silk road': historical perspectives and modern constructions. *Indian Historical Rev.* 47, 21–39. doi: 10.1177/0376983620922431

Nelson, D., and Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. *Plant J.* 66, 194–211. doi: 10.1111/j.1365-313X.2011.04529.x

Nybom, H. (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol. Ecol.* 13, 1143–1155. doi: 10.1111/j.1365-294X.2004.02141.x

Pellissier, L., Eidesen, P. B., Ehrich, D., Descombes, P., Schönswetter, P., Tribsch, A., et al. (2016). Past climate-driven range shifts and population genetic diversity in arctic plants. *J. Biogeography* 43, 461–470. doi: 10.1111/jbi.12657

Perrino, E. V., and Wagensommer, R. P. (2022). Crop wild relatives (CWRs) threatened and endemic to Italy: urgent actions for protection and use. *Biol. (Basel)* 11, 1–32. doi: 10.3390/biology11020193

Petrova, E., and Angel, R. (2024). AMPureXP purification.

Pollegioni, P., Woeste, K., Chiocchini, F., Del Lungo, S., Ciolfi, M., Olimpieri, I., et al. (2017). Rethinking the history of common walnut (*Juglans regia* L.) in Europe: Its origins and human interactions. *PloS One* 12, e0172541. doi: 10.1371/journal.pone.0172541

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PloS One* 5, e9490. doi: 10.1371/journal.pone.0009490

Provan, J. (2013). The effects of past, present and future climate change on range - wide genetic diversity in northern North Atlantic marine species. *Front. Biogeography* 5. doi: 10.21425/F5FBG14732

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Rui, Z. (2010). *Genetic diversity and phylogenetic study of walnut resources in Xinjiang* (Huazhong Agricultural University).

Schuler, M. A., and Werck-Reichhart, D. (2003). Functional genomics of P450S. *Annu. Rev. Plant Biol.* 54, 629–667. doi: 10.1146/annurev.arplant.54.031902.134840

Shatilova, I., Rukhadze, L., and Kokolashvili, I. (2014). The history of genus Juglans L. on the territory of Georgia. *Bull. Georgian Natl. Acad. Sci.* 8, 109–115.

Shi, W., Zhang, D., and Ma, Z. (2022). Transcriptome analysis of genes involved in fatty acid and lipid biosynthesis in developing walnut (*Juglans regia* L.) seed kernels from Qinghai plateau. *Plants* 11, 3207. doi: 10.3390/plants11233207

Stringer, C. (2012). What makes a modern human. *Nature* 485, 33–35. doi: 10.1038/485033a

Taberlet, P., Zimmermann, N. E., Englisch, T., Tribsch, A., Holderegger, R., Alvarez, N., et al. (2012). Genetic diversity in widespread species is not congruent with species richness in alpine plant communities. *Ecol. Lett.* 15, 1439–1448. doi: 10.1111/ele.12004

Wang, Z. Y. (2011). *Analysis of wild walnut resources and genetic diversity in Xinjiang* (Xinjiang Agricultural University).

Wang, L., Cui, N. R., and Zhang, H. F. (1997). Research on wild walnut in Xinjiang. *Arid Zone Stud.* 17–27. doi: 10.13866/j.azr.1997.01.003

Wang, C. Y., and Dang, C. L. (1999). Mating systems and their evolutionary mechanisms in plants and population adaptation. *Wuhan Botanical Res.* 163–172.

Wang, L., Li, X., Yang, L., and Zhang, H. F. (1998). Quantitative classification of wild walnut germplasm resources in Xinjiang. *Northern Gardening.* 5–7.

Wang, G. A., Yahef, A., Zhang, Q., Huang, M. M., and Geng, W. J. (2015). Study on the origin of common walnut in Xinjiang. *Xinjiang Agric. Sci.* 52, 572–579. doi: 10.6048/j.issn.1001–4330.2015.03.028

Wang, J., Ye, H., Zhou, H., Chen, P., Liu, H., Xi, R., et al. (2022). Genome-wide association analysis of 101 accessions dissects the genetic basis of shell thickness for genetic improvement in Persian walnut (*Juglans regia* L.). *BMC Plant Biol.* 22, 436. doi: 10.1186/s12870-022-03824-1

Wang, Z., Zhang, H., Tong, B., Han, B., and Liu, D. (2023). EST - SSR marker - based investigation on genetic diversity and genetic structure of Juglans mandshurica Maxim. in Shandong Province of China. *Genet. Resour. Crop Evol.* 70, 981–991. doi: 10.1007/s10722-022-01482-8

Wei, H. Y., Li, Y. Y., Shang, T. C., and Zhang, W. (2023). Population structure and dynamics of branching wild walnut in Xinjiang Wild Walnut Nature Reserve with different slope directions. *J. Ecol.* 42, 266–273. doi: 10.13292/j.1000-4890.202302.012

Wei, F. W., Ma, T. X., and Hu, Y. B. (2021). Progress and prospects of conservation genetics of endangered veterinary species in China. *J. Veterinary Sci.* 41, 571–580. doi: 10.16829/j.slxb.150517

Wen, J. K., Ma, R., Wang, D. F., and Zhang, P. (2022). Evaluation of wild walnut germplasm resources in Xinjiang for resistance to walnut rot disease. *J. Fruit Trees* 39, 1469–1478. doi: 10.13925/j.cnki.gsxb.20210569

Wen, J. K., Zhang, P., Wang, R., Xing, C. J., Tang, X. X., and Zhu, X. H. (2023). Relationship between branch organization and resistance to walnut rot disease in wild walnut in Xinjiang. *Southwest J. Agric.* 36, 1686–1693. doi: 10.16213/j.cnki.scjas.2023.8.013

Xie, J., Chen, Y., Cai, G., Cai, R., Hu, Z., and Wang, H. (2023). Tree Visualization by One Table (tvBOT): A web application for visualizing, modifying and annotating phylogenetic trees. *Nucleic Acids Res.* 51, W587–W592. doi: 10.1093/nar/gkad359

Xinjiang uygur autonomous region nature reserve research team (1982). Xinjiang wild walnut forest. *For. Sci. Technol.* 15–18. doi: 10.13456/j.cnki.lykt.1982.10.009

Yan, F., Xi, R. M., She, R. X., Chen, P. P., Yan, Y. J., Yang, G., et al. (2021). Improved *de novo* chromosome-level genome assembly of the vulnerable walnut tree *Juglans mandshurica* reveals gene family evolution and possible genome basis of resistance to lesion nematode. *Mol. Ecol. Resour.* 21, 2063–2076. doi: 10.1111/1755-0998.13394

Ye, L., Shavvon, R. H., Qi, H., Wu, H., Fan, P., Shalizi, M. N., et al. (2024). Population genetic insights into the conservation of common walnut (*Juglans regia*) in Central Asia. *Plant Divers.* 46, 600–610. doi: 10.1016/j.pld.2024.06.001

Yao, Y. F., Bruch, A. A., Mosbrugger, V., and Li, C. S. (2011). Quantitative reconstruction of Miocene climate patterns and evolution in Southern China based on plant fossils. *Palaeogeogr Palaeoclimatol Palaeoecol* 304, 291–307. doi: 10.1016/j.palaeo.2010.04.012

Yuan, H. T. (2012). *Construction of a basic database of wild walnut germplasm resources and core germplasm construction method in Xinjiang* (Xinjiang Agricultural University).

Yuan, H. T., Dong, Y. Z., and Wang, Z. Y. (2012). Construction of wild walnut core germplasm by minimum distance stepwise sampling method. *Zhejiang Agric. Sci.* 972–974. doi: 10.16178/j.issn.0528-9017.2012.07.044

Zeng, B. (2005). Status and development of wild walnut resources in Xinjiang. *Northern Fruit Trees.* 1–3. doi: 10.16376/j.cnki.bfgs.2005.04.001

Zhang, W. (2016). *Basic research on the conservation biology of wild walnut, an endangered plant in Tianshan Gorge,Xinjiang* (Northeast Normal University).

Zhang, W., Ba, Y., Yang, X. R., and Yang, Y. F. (2014). Biomass plasticity of wild walnut seeds in Xinjiang and its growth modeling. *Ecol. Lett.* 33, 991–997. doi: 10.14108/j.cnki.1008-8873.2014.05.027

Zhang, H., Chu, W., and Zhang, Z. (2017). Cultivated walnut trees showed earlier but not final advantage over its wild relatives in competing for seed dispersers. *Integr. Zoology* 12, 12–25. doi: 10.1111/1749-4877.12242

Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., and Yang, T. L. (2019b). PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35, 1786–1788. doi: 10.1093/bioinformatics/bty875

Zhang, W., Luo, X. Z., Zhang, N., and Yang, Y. F. (2013). Phenotypic variation and growth characteristics of wild walnut seeds in Xinjiang, China. *J. Biol.* 32, 2281–2288. doi: 10.13292/j.1000-4890.2013.0351

Zhang, W., Ren, Y. L., Zhao, Y., Hai, Y., Yang, Y. F., and Li, J. D. (2012). Biomass plasticity and distribution pattern of different leaflet number compound leaf components in Xinjiang wild walnut. *J. Northeast Forestry Univ.* 40, 37–40+67. doi: 10.13759/j.cnki.dlxb.2012.07.024

Zhang, Z., and Sun, J. (2011). Palynological evidence for Neogene environmental change in the foreland basin of the southern Tianshan range, northwestern China. *Glob Planet Change* 75, 56–66. doi: 10.1016/j.gloplacha.2010.10.006

Zhang, D., Sun, L., Xi, D., Li, X., Gao, L., Miao, L., et al. (2024). Methyl jasmonate-induced bHLH42 mediates tissue-specific accumulation of anthocyanins via regulating flavonoid metabolism-related pathways in Caitai. *Physiologia Plantarum.* 176:4. doi: 10.1111/ppl.14434

Zhang, X. X., Wang, X., Hu, Y., Zhou, W., Chen, X. Y., and Hu, X. S. (2019c). Progress in the study of genetic diversity in marginal populations of plants. *J. Plant Ecol.* 43, 383–395. doi: 10.17521/cjpe.2018.0252

Zhang, B. W., Xu, L. L., Li, N., Yan, P. C., Jiang, X. H., Woeste, K. E., et al. (2019a). Phylogenomics reveals an ancient hybrid origin of the persian walnut. *Mol. Biol. Evol.* 36, 2451–2461. doi: 10.1093/molbev/msz112

Zhang, J., Zhang, W., Ji, F., Qiu, J., Song, X., Bu, D., et al. (2020). A high-quality walnut genome assembly reveals extensive gene expression divergences after whole-genome duplication. *Plant Biotechnol. J.* 18, 1848–1850. doi: 10.1111/pbi.13350

Zhang, J. (2015). Analysis of SRAP genetic diversity and construction of core germplasm of wild walnut in Xinjiang. *Xinjiang Agricultural University.* doi: 10.1128/AAC.03728-14

Zhang, J., Zhang, P., and Li, Q. X. (2018). Construction of core germplasm of wild walnut in Xinjiang. *J. Fruit Trees* 35, 168–176. doi: 10.13925/j.cnki.gsxb.20170342

Zhang, W., Zhao, Y., Zhang, X. F., Yang, Y. F., and Li, J. D. (2011). Phenotypic variation and growth pattern of compound leaves of wild walnut in Ili, Xinjiang. *J. Northeast Normal Univ. (Natural Sci. Edition)* 43, 113–117. doi: 10.16163/j.cnki.22-1123/n.2011.01.015

Zhao, S., Zhang, X., Su, Y., Chen, Y., Liu, Y., Sun, M., et al. (2018). Transcriptome analysis reveals dynamic fat accumulation in the walnut kernel. *Int. J. Genomics.* 2018, 1–13. doi: 10.1155/2018/8931651

Zhao, P., Zhou, H. J., Liu, Z. L., Woeste, K. E., Hu, D. F., Dang, M., et al. (2014). Progress in molecular phylogenetic and biogeographic studies of Juglans. *Forestry Sci.* 50, 147–157. doi: 10.11707/j.1001-7488.20141120