Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Evaluation of Germplasm Resources and Construction of a Core Collection for Juglans mandshurica

qinhui zhang
Jilin Agricultural University
yitong chen
Jilin Agricultural University
taijin zhang
Sichuan University
chengyao wang
Jilin Agricultural University
xiyang zhao
xiyangz@jlau.edu.cn

Jilin Agricultural University

Research Article

Keywords: Juglans mandshurica, Phenotypic differentiation, Comprehensive evaluation, EST-SSR, Core collection

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11406961/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Evaluation of Germplasm Resources and Construction of a Core Collection for *Juglans mandshurica*

Qinhui Zhang · Yitong Chen · Taijin Zhang · Chengyao Wang · Xiyang Zhao

Abstract

Background The evaluation and conservation of germplasm resources provide the foundation for genetic improvement of forest trees. Early growth phenotypes of *Juglans mandshurica* Maxim. are highly susceptible to environmental influences, increasing the complexity of family evaluation. In addition, its limited seed dispersal capacity often results in the aggregation of closely related individuals within populations, leading to genetic redundancy in collected germplasm resources. In the present study, a total of 117 *J. mandshurica* families were evaluated through a comprehensive multi-trait assessment, and a core collection was constructed by integrating 11 phenotypic traits with genotypic information from 10 pairs of expressed sequence tag-simple sequence repeat (EST-SSR) molecular markers.

Results Phenotypic analyses revealed significant differences among families for all measured traits ($p < 0.01$). Nut and leaf traits, including nut weight and leaf area, showed stronger genetic control ($h^2 > 0.5$), whereas early growth traits, such as seedling height and ground diameter, exhibited greater environmental plasticity. Based on the variation patterns identified by principal component analysis, 11 superior families were selected using a comprehensive multi-trait evaluation approach. These selected families demonstrated relatively high genetic gains for highly heritable traits. Molecular marker analysis further showed that the 10 EST-SSR primer pairs exhibited strong discriminatory power among the tested materials with an average polymorphism information content (PIC) of 0.818. By comparing different

sampling strategies and proportions, the simulated annealing–genetic diversity maximization (SAGD) strategy combined with a 30% sampling proportion was identified as the optimal approach for constructing the core collection. Using this scheme, 35 core germplasm accessions were selected, retaining 99.5% of alleles. Both observed and expected heterozygosity were significantly higher in the core collection than in the original germplasm ($p < 0.05$).

Conclusions The increased heterozygosity indicates that this construction strategy effectively reduced the proportion of closely related individuals in the original germplasm, thereby minimizing genetic redundancy. Collectively, the superior families identified and the core collection established in this study provide a valuable foundation for targeted genetic improvement and scientific management of *J. mandshurica* resources.

Keywords *Juglans mandshurica* · Phenotypic differentiation · Comprehensive evaluation · EST-SSR · Core collection

Introduction

Forest tree germplasm resources provide the material foundation for elucidating species evolutionary mechanisms, breeding superior varieties, and maintaining ecosystem stability. Accurate evaluation of genetic variation and efficient resource management remain major scientific challenges in contemporary forest genetics and breeding [1, 2]. Historically, germplasm resource evaluation has primarily relied on phenotypic observations. Conventional phenotypic assessments mainly include growth-related traits (e.g. tree height and diameter at breast height) [3], leaf characteristics [4], and seed and fruit morphological traits [5]. These traits provide direct insights into the growth performance and environmental adaptability of different germplasm resources. Nevertheless, phenotypic trait expression is inherently shaped by long-

term interactions between genotype and environment [6]. Consequently, these traits are highly susceptible to environmental variation caused by habitat heterogeneity, resulting in pronounced phenotypic plasticity [7]. This plasticity makes it challenging to fully and accurately characterize the underlying genetic background of a species based solely on phenotypic assessments [8]. Therefore, analyses of population genetic characteristics using molecular markers, such as simple sequence repeats (SSRs), which exhibit high polymorphism, codominance, and independence from environmental influences, have become an essential complement to phenotypic evaluations [9, 10].

In recent years, an integrated "phenotypic–molecular" evaluation approach has increasingly become a standard strategy for the identification and characterization of forest tree germplasm resources [11]. However, as the scope of germplasm collection continues to expand, population sizes have increased substantially. The resulting genetic redundancy, together with rising conservation and maintenance costs, has become a major constraint limiting the in-depth exploration and efficient utilization of germplasm resources [12, 13]. The introduction of the "core collection" concept provides an effective solution to this challenge. Its primary objective is to maximize the retention of genetic diversity from the original population while using the minimum possible sample size [11, 13, 14]. Currently, core collection construction strategies have evolved from early random sampling and stratified sampling approaches to heuristic algorithm-based methods, such as simulated annealing–maximizing number of alleles (SANA) strategy and the simulated annealing–maximizing genetic diversity (SAGD) strategy [15]. Because population genetic structures and patterns of polymorphism distribution differ among species, identifying the optimal combination of sampling algorithms and sampling proportions to minimize the loss of rare alleles still requires targeted empirical evaluation [16, 17].

Juglans mandshurica Maxim. is a native broad-leaved tree species endemic to Northeast China. It

has significant comprehensive economic value, providing high-quality timber, edible nuts, and medicinal resources, while also playing an important role in maintaining regional ecological barrier functions [18, 19]. In recent years, driven by the combined effects of intensive commercial logging and habitat fragmentation, natural *J. mandshurica* populations have declined sharply, resulting in severe losses of valuable genetic resources. Therefore, systematic evaluation and conservation of these resources are urgently needed [20].

Although foundational studies on the germplasm resources of *J. mandshurica* have been conducted, two major limitations remain. First, existing phenotypic evaluations have primarily focused on single-dimensional morphological comparisons or assessments within specific geographical regions. Comprehensive analyses of the coordinated variation patterns among multiple organs (including seed/fruit, leaf, and growth traits) at the family level remain limited [20-22]. Second, although applicable SSR marker systems have been developed for molecular identification, few studies have integrated comprehensive multi-trait phenotypic evaluations with polymorphic molecular data to systematically investigate and optimize strategies for constructing a *J. mandshurica* core collection [23, 24]. Together, these limitations impede the accurate identification of superior genetic resources and the advancement of subsequent breeding efforts, including genome-wide association studies (GWAS).

To address these knowledge gaps, the present study investigated 117 *J. mandshurica* families collected from four representative provenances in Jilin Province. The specific objectives were to: (1) measure 11 representative phenotypic traits, including mean nut diameter, nut weight, specific leaf area, and seedling height, to characterize phenotypic variation patterns among different provenances and families; (2) evaluate genetic polymorphism in the tested materials using expressed sequence tag-simple sequence repeat (EST-SSR) markers to assess the discriminatory power of the marker system; and (3)

systematically compare the efficiency of random sampling and simulated annealing algorithms in retaining genetic parameters across different sampling proportions, thereby identifying the optimal algorithm and sampling proportion for constructing the *J. mandshurica* core collection. Overall, this study aims to provide a theoretical basis and material foundation for the evaluation, conservation, and selective breeding of *J. mandshurica* germplasm resources, while also offering valuable insights into the resource management of other native broad-leaved tree species.

Materials and Methods

Experimental Site and Plant Materials

In September 2022, *J. mandshurica* fruits were collected from individual maternal trees (representing distinct families) from four representative wild provenances in Jilin Province, China: Wangqing (WQ), Linjiang (LJ), Huinan (HN), and Hongshi (HS). Detailed geographical coordinates and climatic factors of the four provenances are provided in **Table 1**. A minimum spatial distance of 100 m was maintained between the sampled mother trees to prevent sampling closely related individuals. The collection of wild plant materials was conducted with the explicit permission and field assistance of the local forestry authorities, specifically the Jilin Changbai Mountain Forestry Industry Group and its subsidiaries (namely, Jilin Changbai Mountain Forestry Industry Wangqing Forestry Co., Ltd., Jilin Changbai Mountain Forestry Industry Linjiang Forestry Co., Ltd., Jilin Changbai Mountain Forestry Industry Huinan Forestry Co., Ltd., and Jilin Changbai Mountain Forestry Industry Hongshi Forestry Co., Ltd.).

Formal identification of the wild *J. mandshurica* plants was undertaken in the field by Qinhui Zhang (Jilin Agricultural University). Although traditional dried herbarium voucher specimens were not deposited, a portion of the collected fruits was transported to the laboratory for phenotypic measurements,

and the remaining fruits were sown to establish a *J. mandshurica* living germplasm resource repository. In May 2024, the seedlings were transplanted to a mountainous site to conserve the genetic resources of 117 families. These 117 families (designated as JM-1 to JM-117) have been explicitly cataloged and mapped to their original provenance collection codes, serving as living vouchers that are maintained by the authors for future reference and verification.

Table 1 Climatic and geographical factors of *J. mandshurica* from different provenances

Provenance	Code	Families	Altitude (m)	Longitude (°E)	Latitude (°N)	Frost-free period (d)	Mean annual temperature (℃)	Mean annual precipitation (mm)	Mean annual sunshine duration (h)
Wangqing	WQ	20	806	129.05	43.06	125	3.9	580	2700
Linjiang	LJ	45	338	126.53	41.49	125	3.0	875	1894
Huinan	HN	30	454	125.58	42.16	138	5.0	737	2296
Hongshi	HS	22	756	126.51	42.34	105	3.7	765	2560

Experimental Methods

(1) Measurement of Phenotypic Traits

The seed/fruit, leaf, and seedling growth traits of the 117 *J. mandshurica* families were evaluated, generating data for 24 phenotypic traits (**Supplementary material**).

① Measurement of seed and fruit traits: For each family, 30 healthy fruits free from pests and diseases were randomly selected. A vernier caliper was used to measure the longitudinal, transverse, and lateral diameters of both fruits (with green husks) and nuts (after de-husking). An electronic balance was used to determine the weights of fruits, nuts, and kernels. Based on these measurements, derived parameters, including mean nut diameter, fruit shape index, nut roundness index, and kernel percentage, were calculated [21].

② Measurement of leaf traits: In July 2025, during the vigorous seedling growth period, 15 pest- and disease-free leaves with uniform growth characteristics were randomly selected from each family.

After measuring leaf fresh weight, the leaves were vertically photographed using a digital camera. Morphological parameters, including leaf length, leaf width, and leaf area, were extracted using ImageJ image analysis software. The leaves were then oven-dried at 80 °C to a constant weight to determine leaf dry weight. Subsequently, derived traits, including leaf water content, leaf dry matter content, and specific leaf area, were calculated [25].

③ Measurement of growth traits: In November 2025, at the end of the growing season, 20 individual trees with uniform growth vigor were selected from each family. Seedling height and ground diameter were measured using a telescopic measuring rod and a vernier caliper, respectively.

(2) Selection of Evaluated Traits

Among the 24 initially measured phenotypic indicators, substantial multicollinearity was observed among some parameters. To avoid overlapping contributions in subsequent multivariate analyses, 11 key phenotypic traits were ultimately selected for statistical analysis after removing redundant variables. These traits represented three major categories: seed/fruit traits (mean nut diameter, nut roundness index, nut weight, kernel weight, and kernel percentage), leaf traits (leaf area, leaf dry weight, specific leaf area, and leaf dry matter content), and early growth traits (seedling height and ground diameter). Together, these traits comprehensively captured the phenotypic characteristics of the *J. mandshurica* families.

(3) DNA Extraction and SSR Amplification

Genomic DNA was extracted from the leaves of individual J. mandshurica plants using a modified cetyltrimethylammonium bromide (CTAB) method [26]. DNA quality and concentration were assessed by 1% agarose gel electrophoresis and a microvolume spectrophotometer, respectively. The DNA samples were then diluted to 25 ng/μL for subsequent analyses. Based on previously published SSR primer information for *J. mandshurica* [23], 10 pairs of EST-SSR primers were selected, with their

sequence information provided in **Table 2**. The M13 fluorescent primer labeling method was employed, in which a universal M13 sequence (5′-TGTAAAACGACGGCCAGT-3′) was attached to the 5′ end of the forward primers. Four fluorescent dyes, FAM, HEX, TAMRA, and ROX, were used for labeling and synthesis. The PCR reaction mixture had a total volume of 20 μL, containing 10 μL of 2× Super PCR Mix (Beijing Genomics Institution, Beijing, China), 0.2 μL of forward primer (10 μM), 0.8 μL of reverse primer (10 μM), 0.5 μL of fluorescently labeled M13 primer (10 μM, Sangon Biotech, Shanghai, China), 2 μL of genomic DNA template (approx. 35 ng/μL), and 6.5 μL of sterile ddH_2O. The amplification program consisted of an initial denaturation at 94 °C for 4 min; followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 60 °C for 30 s, and extension at 72 °C for 30 s; followed by another 5 cycles of denaturation at 94 °C for 30 s, annealing at 58 °C for 30 s, and extension at 72 °C for 30 s (to allow the fluorescent M13 primer to anneal); and a final extension at 72 °C for 10 min [27]. The PCR products were subsequently sent to Sangon Biotech (Shanghai) Co., Ltd., Shanghai, China for capillary electrophoresis and fragment size determination.

Table 2 Sequence information for the 10 EST-SSR primer pairs in *J. mandshurica*

Locus	Forward Primer	Reverse Primer	Motif	Tm (°C)	Size (bp)
NEJM-1	GGTTTTCTTGCATTTTGCTG	GCTTTGGTTTCAATTTGGCT	(AG)15	58.81	277
NEJM-2	TTCGCTGCTTCCTCTCTCTC	TCGTGTTGTCCGTTCTCATC	(CT)8	59.83	250
NEJM-3	CAACTCGTACGTGCGATCAT	AGTGAGAGCTTGGAGCTTGG	(CT)15	59.75	228
NEJM-4	GAGATGAGTGGGAGCTTTGC	GGCTTGTGATCCGAAATCAT	(TC)8	59.93	273
NEJM-5	GAGCGAAAACTGTGCCTTTC	TTGGCTAGAGGTTGAAGGGA	(CT)10	59.90	248
NEJM-7	AGACTCGAAGGCTCTGCTTG	TATGGCACCTTTGTGGTCAA	(TC)15	59.93	206
NEJM-8	TGGGGTGTTGATACACAAAGA	CCTCAGCAACAGTCCCTTTC	(TG)8	59.36	272
NEJM-12	GACCTCAAACCAGATCTTCCA	CCAGTTCCCAGGTGATGTCT	(TC)15	59.54	264
NEJM-14	TAACGCAATCCCAACACAAA	GTATTCGTCTGCTCCGCTTC	(AG)12	59.98	219
NEJM-15	CTAATGGTTGGTGTGCCAGA	GAACCACAAGGATTCCCAGA	(CT)12	59.73	182

Statistical Analysis

(1) Phenotypic Data Analysis

Data analysis and visualization of the 11 selected phenotypic traits were performed using Microsoft Excel 2024 (Microsoft Corp.), IBM SPSS Statistics version 26 (IBM Corp.), and R version 4.0.2 (R Core Team).

One-way analysis of variance (ANOVA) was conducted to assess the significance of differences in

phenotypic traits among families using the following linear model:

$$Y_{ij} = \mu + F_i + e_{ij} \quad (1)$$

where Y_{ij} is the observation value of the j-th individual in the i-th family, μ is the overall mean of the trait, F_i represents the random effect of the i-th family, and e_{ij} represents the random error.

To evaluate the relationships among different traits, Pearson correlation coefficients (representing phenotypic correlations) were calculated using the following formula [28]:

$$r_P(x,y) = \frac{Cov_P(x,y)}{\sqrt{\sigma_{P_x}^2 \cdot \sigma_{P_y}^2}} \quad (2)$$

where $r_P(x,y)$ is the phenotypic correlation coefficient between trait x and trait y, $Cov_P(x,y)$ is the phenotypic covariance between trait x and trait y, and $\sigma_{P_x}^2$ and $\sigma_{P_y}^2$ represent the phenotypic variance components for trait x and trait y, respectively.

To evaluate the genetic variation potential, the phenotypic coefficient of variation (PCV) and genotypic coefficient of variation (GCV) were calculated using the following formulas [29]:

$$PCV = \frac{\sqrt{\sigma_P^2}}{\bar{X}} \times 100\% \quad (3)$$

$$GCV = \frac{\sqrt{\sigma_g^2}}{\bar{X}} \times 100\% \quad (4)$$

where σ_p^2 is the phenotypic variance component, σ_g^2 is the genotypic (family) variance component, and $\bar{X}$ is the overall phenotypic mean of the trait.

The family heritability (h^2) was estimated to determine the degree of genetic control for each trait as follows [30]:

$$h^2 = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_e^2/n} \quad (5)$$

where σ_f^2 denotes the variance component for the family effect, σ_e^2 is the residual error variance, and n is the number of individuals measured per family.

Principal component analysis (PCA) was performed on standardized phenotypic data to reduce

dimensionality. Subsequently, a comprehensive multi-trait evaluation method was applied to calculate the comprehensive score (Q_i value) for each family using the following equation [31]:

$$Q_i = \sqrt{\sum_{j=1}^{n} a_{ij}} \quad (6)$$

where $a_{ij} = X_{ij}/X_{jmax}$; X_{ij} is the mean value of trait j for family i, X_{jmax} is the optimal (or maximum) value of trait j, and n is the total number of evaluated traits.

Based on the calculated Q_i scores, superior families were selected, and the genetic gain (ΔG) for each trait was estimated to evaluate the selection efficiency as follows [31]:

$$\Delta G = \frac{h^2 W}{\overline{X}} \times 100\% \quad (7)$$

where W is the selection differential (the difference between the mean of the selected superior families and the overall mean of the population), $\bar{X}$ is the overall mean, and h^2 is the family heritability of the given trait.

(2) Molecular Marker Data Analysis

The raw microsatellite data obtained from capillary electrophoresis were converted into GenePop format using MS_tools software [32]. GenAlEx 6.502 software [33] was used to calculate genetic diversity parameters, including the number of alleles (N_a), number of effective alleles (N_e), observed heterozygosity (H_o), expected heterozygosity (H_e), and Shannon's information index (I). In addition, PIC_CALC software [34] was used to calculate polymorphism information content (PIC) to evaluate primer polymorphism and the level of allelic variation in the tested materials.

(3) Construction and Evaluation of the Core Collection

The core collection was constructed using PowerMarker V3.25 software [35]. Five sampling proportions (10%, 15%, 20%, 25%, and 30%) were evaluated, and three sampling strategies were systematically compared: the simulated annealing-maximizing number of alleles (SANA) strategy, the simulated

annealing-maximizing genetic diversity (SAGD) strategy, and the random sampling (R) method [16, 36]. Student's t-tests were performed to determine whether significant differences existed in genetic parameters (N_a, N_e, I, H_o, H_e) between the core collection and the original germplasm. In addition, principal coordinate analysis (PCoA) was conducted to evaluate the spatial representativeness of the core collection in relation to the genetic distribution of the original germplasm [37].

Results

Evaluation of Phenotypic Traits

Analysis of Phenotypic Variance and Genetic Variation Parameters

The results of variance analyses and genetic variation parameter assessments for each phenotypic trait of *J. mandshurica* families are presented in **Table 3** and **Table 4**. ANOVA revealed significant differences among families for all 11 phenotypic traits ($p < 0.01$). In terms of variation magnitude, the PCV ranged from 8.76% to 56.46%, while the GCV ranged from 7.27% to 47.81%. Notably, the GCV values for nut weight, kernel weight, leaf area, leaf dry weight, specific leaf area, and seedling height all exceeded 20%. In terms of genetic control, the ratio of GCV to PCV among traits ranged from 34.41% to 85.66%. The family heritability (h^2) of nut- and leaf-related traits was generally higher than 0.5 (ranging from 0.505 to 0.734), whereas the family heritabilities of early growth traits, including seedling height and ground diameter, were relatively low (0.278 and 0.118, respectively).

Table 3 Variance analysis for 11 phenotypic traits in *J. mandshurica* families

Trait	Source of variation	SS	df	MS	*F*
Mean nut diameter (mm)	Family	20811.423	116	179.409	67.259**
Nut roundness index	Family	13.727	116	0.118	49.700**
Nut weight (g)	Family	20217.582	116	174.289	83.660**
Kernel weight (g)	Family	856.723	116	7.386	49.087**
Kernel percentage (%)	Family	37348.440	116	321.969	31.652**
Leaf area (cm^2)	Family	90189.457	116	777.495	22.469**
Leaf dry weight (g)	Family	4.504	116	0.039	30.644**
Specific leaf area (cm^2/g)	Family	6039760.800	116	52066.903	22.446**
Leaf dry matter content (%)	Family	52001.513	116	448.289	31.064**
Seedling height (cm)	Family	393341.964	116	3390.879	8.683**

Ground diameter (mm)	Family	1717.854	116	14.809	3.687**

Note: ** indicates significance at $p < 0.01$.

Table 4 Genetic variation parameters for 11 phenotypic traits in *J. mandshurica* families

Trait	Mean	Range	SD	GCV (%)	PCV (%)	h^2
Mean nut diameter (mm)	33.41	28.60~44.08	2.93	7.27	8.76	0.688
Nut roundness index	0.65	0.50~0.82	0.08	9.50	11.71	0.659
Nut weight (g)	11.51	7.18~19.27	2.80	20.82	24.30	0.734
Kernel weight (g)	2.11	0.90~3.59	0.63	23.23	29.59	0.617
Kernel percentage (%)	18.61	11.50~30.25	4.54	17.32	24.37	0.505
Leaf area (cm^2)	19.83	4.10~46.03	9.15	35.49	46.25	0.589
Leaf dry weight (g)	0.1053	0.0162~0.4354	0.0612	47.81	56.46	0.717
Specific leaf area (cm^2/g)	208.49	42.59~453.90	74.90	27.62	36.01	0.588
Leaf dry matter content (%)	35.50	25.38~51.03	6.57	15.15	18.55	0.667
Seedling height (cm)	47.6	23.5~78.4	23.2	25.72	48.82	0.278
Ground diameter (mm)	7.07	5.38~11.72	2.13	10.38	30.17	0.118

Correlation Analysis

The Pearson correlation analysis results for the 11 phenotypic traits of *J. mandshurica* families are presented in **Fig. 1**. The results revealed varying degrees of significant correlations among the traits. For stronger correlations ($r > 0.5$), significantly positive correlations were mainly observed among traits within the same organ categories. These included seed/fruit traits (mean nut diameter with nut weight and kernel weight; nut weight with kernel weight; and kernel weight with kernel percentage) and leaf traits (leaf area with leaf dry weight). The significantly negative correlations were primarily observed between specific leaf area and leaf dry weight, as well as between specific leaf area and leaf dry matter content. In addition, several traits showed significant but weak correlations ($r < 0.5$). Among these, nut roundness index was weakly negatively correlated with seedling height, and nut weight was weakly negatively correlated with kernel percentage. In contrast, weak positive correlations were observed between seedling height and ground diameter, kernel percentage and ground diameter, and leaf dry weight and leaf dry matter content.

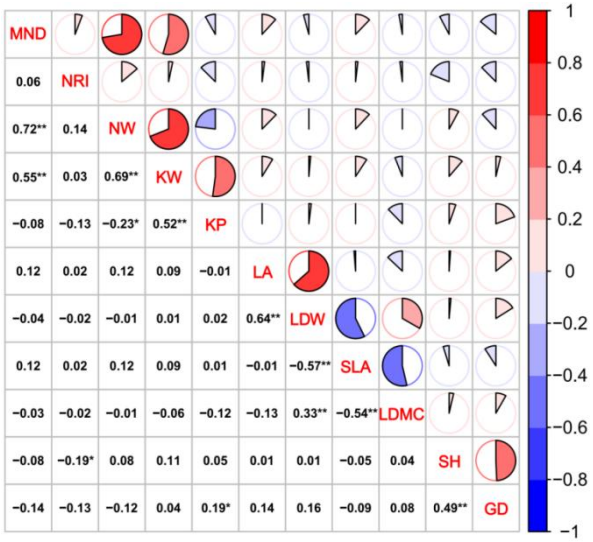

Fig. 1 Pearson correlations among 11 phenotypic traits of *J. mandshurica* families (MND: mean nut diameter; NRI: nut roundness index; NW: nut weight; KW: kernel weight; KP: kernel percentage; LA: leaf area; LDW: leaf dry weight; SLA: specific leaf area; LDMC: leaf dry matter content; SH: seedling height; GD: ground diameter)

Principal Component Analysis

The PCA results for the 11 phenotypic traits of *J. mandshurica* families are presented in **Table 5**. Based on the criterion of eigenvalues greater than 1, five principal components (PCs) were extracted, explaining a cumulative variance of 79.41%. The loading matrix indicated that these five PCs effectively reduced dimensionality while distinguishing traits from different organs. Principal Component I (variance contribution rate: 22.14%) showed high loadings for nut weight (0.950), mean nut diameter (0.859), and kernel weight (0.781). Principal Component II (19.30%) was mainly characterized by high loadings for leaf dry matter content (0.870) and specific leaf area (-0.865). Principal Component III (15.52%) was primarily associated with leaf area (0.948) and leaf dry weight (0.803). Principal Component IV (11.79%) exhibited high loadings for early growth traits, including seedling height (0.871) and ground diameter (0.798). Principal Component V (10.67%) showed the highest loading for kernel percentage (0.973).

Table 5 Principal component analysis (PCA) for 11 phenotypic traits in *J. mandshurica* families

Principal components	PC I	PC II	PC III	PC IV	PC V
Eigenvalue	2.435	2.123	1.707	1.296	1.174
Variance contribution rate (%)	22.14	19.30	15.52	11.79	10.67
Cumulative contribution rate (%)	22.14	41.44	56.95	68.74	79.41
Mean nut diameter (mm)	0.859	-0.046	0.028	-0.144	-0.037
Nut roundness index	0.136	-0.012	0.062	-0.414	-0.135
Nut weight (g)	0.950	-0.025	0.037	-0.038	-0.153

Kernel weight (g)	0.781	-0.029	0.021	0.079	0.582
Kernel percentage (%)	-0.056	-0.056	0.003	0.151	0.973
Leaf area (cm^2)	0.102	-0.129	0.948	0.032	-0.028
Leaf dry weight (g)	-0.029	0.521	0.803	0.031	0.050
Specific leaf area (cm^2/g)	0.118	-0.865	-0.180	-0.028	-0.045
Leaf dry matter content (%)	0.031	0.870	-0.110	0.052	-0.108
Seedling height (cm)	0.119	0.013	-0.038	0.871	-0.108
Ground diameter (mm)	-0.077	0.050	0.183	0.798	0.077

Evaluation and Selection of Superior Families

Using the 11 phenotypic traits as evaluation indicators, a comprehensive multi-trait evaluation method was applied to assess the 117 *J. mandshurica* families. The comprehensive scores (Q_i) for each family are presented in **Table 6**. Based on a 10% selection rate, 11 families, including JM-6 and JM-75, were selected. The genetic gain evaluation results (**Table 7**) showed that the selected families achieved higher genetic gains for seed/fruit and leaf traits, including leaf dry weight (38.03%), leaf area (27.78%), kernel weight (21.96%), and nut weight (13.99%). In contrast, the genetic gains for early growth traits, such as seedling height and ground diameter, were relatively low, at 8.08% and 1.79%, respectively.

Table 6 Comprehensive scores (Q_i) and rankings of different *J. mandshurica* families

Family	Q_i value	Rank	Family	Q_i value	Rank	Family	Q_i value	Rank	Family	Q_i value	Rank
JM-6	6.487	1	JM-74	5.309	31	JM-21	4.907	61	JM-3	4.618	91
JM-75	6.390	2	JM-18	5.288	32	JM-88	4.904	62	JM-94	4.606	92
JM-103	6.353	3	JM-92	5.277	33	JM-36	4.894	63	JM-86	4.600	93
JM-113	6.227	4	JM-16	5.263	34	JM-2	4.889	64	JM-77	4.599	94
JM-37	6.062	5	JM-15	5.233	35	JM-61	4.883	65	JM-111	4.580	95
JM-62	6.058	6	JM-8	5.219	36	JM-78	4.882	66	JM-51	4.574	96
JM-13	5.994	7	JM-59	5.198	37	JM-38	4.872	67	JM-104	4.568	97
JM-96	5.839	8	JM-39	5.195	38	JM-64	4.866	68	JM-27	4.521	98
JM-46	5.825	9	JM-87	5.184	39	JM-100	4.862	69	JM-71	4.509	99
JM-11	5.739	10	JM-102	5.153	40	JM-101	4.851	70	JM-106	4.501	100
JM-95	5.734	11	JM-68	5.103	41	JM-4	4.832	71	JM-48	4.497	101
JM-32	5.723	12	JM-1	5.102	42	JM-67	4.832	72	JM-10	4.480	102
JM-84	5.694	13	JM-14	5.091	43	JM-80	4.822	73	JM-33	4.468	103
JM-44	5.673	14	JM-112	5.090	44	JM-50	4.815	74	JM-65	4.402	104
JM-89	5.631	15	JM-117	5.089	45	JM-81	4.795	75	JM-79	4.384	105
JM-105	5.560	16	JM-35	5.059	46	JM-20	4.782	76	JM-63	4.373	106
JM-23	5.557	17	JM-30	5.036	47	JM-29	4.773	77	JM-114	4.373	107
JM-42	5.538	18	JM-110	5.029	48	JM-58	4.753	78	JM-109	4.349	108
JM-43	5.500	19	JM-60	5.026	49	JM-69	4.753	79	JM-54	4.322	109
JM-97	5.499	20	JM-98	5.016	50	JM-24	4.741	80	JM-99	4.298	110
JM-116	5.478	21	JM-25	5.012	51	JM-49	4.740	81	JM-90	4.265	111
JM-12	5.407	22	JM-28	4.993	52	JM-53	4.736	82	JM-56	4.248	112
JM-85	5.394	23	JM-66	4.990	53	JM-7	4.727	83	JM-26	4.194	113
JM-82	5.387	24	JM-19	4.953	54	JM-55	4.724	84	JM-83	4.186	114
JM-47	5.367	25	JM-115	4.951	55	JM-108	4.703	85	JM-91	4.129	115
JM-31	5.361	26	JM-70	4.933	56	JM-40	4.695	86	JM-93	4.104	116
JM-73	5.345	27	JM-52	4.931	57	JM-17	4.694	87	JM-76	3.790	117
JM-57	5.327	28	JM-9	4.925	58	JM-41	4.677	88			
JM-22	5.319	29	JM-34	4.913	59	JM-72	4.668	89			
JM-5	5.315	30	JM-107	4.908	60	JM-45	4.638	90			

Table 7 Genetic gains for phenotypic traits in the selected superior *J. mandshurica* families

Trait	Genetic gain (%)	Trait	Genetic gain (%)	Trait	Genetic gain (%)
Mean nut diameter (mm)	2.84	Kernel percentage (%)	8.74	Leaf dry matter content (%)	-1.62
Nut roundness index	1.58	Leaf area (cm^2)	27.78	Seedling height (cm)	8.08
Nut weight (g)	13.99	Leaf dry weight (g)	38.03	Ground diameter (mm)	1.79
Kernel weight (g)	21.96	Specific leaf area (cm^2/g)	-4.83		

Construction of the Core Collection

Primer Polymorphism Analysis

The results of the polymorphism parameter analysis for the 10 pairs of EST-SSR primers in the tested germplasm are presented in **Table 8**. A total of 172 alleles (N_a) were detected across the 10 loci, with the number of alleles per locus ranging from 11 to 24 and an average of 17.2 alleles per locus. The average number of effective alleles (N_e) was 5.252. The mean values of Shannon's information index (I), observed heterozygosity (H_o), and expected heterozygosity (H_e) were 1.810, 0.643, and 0.761, respectively. The PIC ranged from 0.634 to 0.909, with an average value of 0.818. These parameters demonstrate that the selected EST-SSR primer set has strong genetic discriminatory power for the tested *J. mandshurica* families.

Table 8 Polymorphism parameters for the 10 EST-SSR primer pairs in *J. mandshurica*

Locus	N_a	N_e	I	H_o	H_e	PIC
NEJM-1	16	4.268	1.709	0.564	0.724	0.806
NEJM-2	20	7.406	2.198	0.687	0.862	0.909
NEJM-3	24	7.534	2.218	0.753	0.852	0.904
NEJM-4	14	3.038	1.364	0.460	0.597	0.696
NEJM-5	14	4.803	1.754	0.684	0.783	0.824
NEJM-7	15	5.157	1.711	0.720	0.772	0.787
NEJM-8	18	5.906	1.996	0.779	0.815	0.852
NEJM-12	17	5.088	1.825	0.632	0.780	0.857
NEJM-14	11	2.906	1.331	0.509	0.618	0.634
NEJM-15	23	6.413	1.995	0.645	0.810	0.907
Mean	17.2	5.252	1.810	0.643	0.761	0.818

Construction and Validation of the Core Collection

Next, we compared the retention of genetic parameters across five sampling proportions (10%-30%) using the SANA strategy, the SAGD strategy, and the R method **(Table 9)**. The results showed that

genetic diversity indicators of the candidate core collections generally increased with increasing sampling proportions. Under the same sampling proportion, the retained values of N_e and I were generally higher under the SAGD strategy than under the SANA and R strategies. At low sampling proportions (10% and 15%), N_e showed a significant reduction under all three strategies. However, when the sampling proportion reached 30%, the genetic parameters retained by the SAGD strategy remained relatively stable.

Based on these findings, the scheme combining the SAGD strategy with a 30% sampling proportion was selected for core collection construction. Following this scheme, a total of 35 *J. mandshurica* core germplasm accessions were selected. Student's *t*-test results **(Table 9)** showed that the number of alleles (N_a = 6.49), number of effective alleles (N_e = 6.21), and Shannon's information index (I = 1.79) of the core collection did not differ significantly from those of the original germplasm ($p > 0.05$), with an allele retention rate of 99.5%. Conversely, the observed heterozygosity (H_o = 0.83) and expected heterozygosity (H_e = 0.82) of the core collection were significantly higher than those of the original germplasm ($p < 0.05$). PCoA (**Fig. 2**) further demonstrated that the 35 core germplasm accessions were relatively evenly distributed in the two-dimensional space defined by the first and second principal coordinates, effectively capturing the major patterns of genetic variation among the 117 original germplasm accessions.

Table 9 Comparison of genetic parameters among different construction strategies and sampling proportions

Method	Number of accessions	N_a	N_e	I	H_o	H_e
Original germplasm	117	6.52	6.18	1.79	0.78	0.81
SANA (10%)	12	6.38	5.18*	1.69*	0.82	0.79*
SANA (15%)	18	6.39	5.50*	1.72*	0.81	0.80
SANA (20%)	23	6.42	5.75*	1.74	0.80	0.80
SANA (25%)	29	6.44	5.85*	1.76	0.79	0.81
SANA (30%)	35	6.47	6.04	1.77	0.80	0.81
SAGD (10%)	12	6.39	5.81	1.75	0.78*	0.81*
SAGD (15%)	18	6.42	5.91	1.76	0.77	0.81
SAGD (20%)	23	6.43	6.09	1.78	0.80	0.81
SAGD (25%)	29	6.45	6.05	1.77	0.78	0.81
SAGD (30%)	35	6.49	6.21	1.79	0.83*	0.82*
R (10%)	12	6.22*	5.14*	1.68*	0.75	0.79*
R (15%)	18	6.34	5.49*	1.72*	0.75	0.80
R (20%)	23	6.39	5.70*	1.74	0.75	0.80
R (25%)	29	6.40	5.87*	1.76	0.76	0.81
R (30%)	35	6.44	5.92*	1.77	0.77	0.81

Note: SANA, simulated annealing-maximizing number of alleles; SAGD, simulated annealing-

maximizing genetic diversity; R, random sampling. * indicates a significant difference between the core collection and the original germplasm at $p < 0.05$.

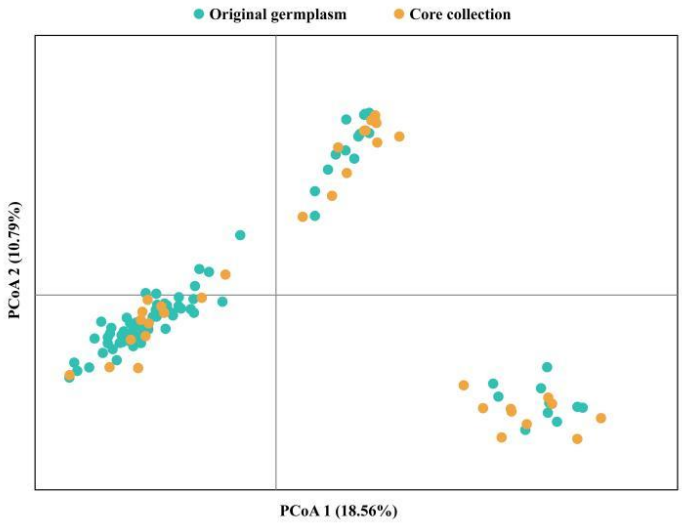

Fig. 2 Principal coordinate analysis (PCoA) of the original germplasm and the core collection

Discussion

Phenotypic variation provides the material basis for the evaluation of forest tree germplasm resources and the selective breeding of superior varieties. Understanding the degree of genetic control underlying different phenotypic traits is essential for developing effective breeding strategies [38, 39]. Herein, analyses of variance and genetic parameters revealed significant differences among families for all 11 phenotypic traits evaluated in the 117 *J. mandshurica* families ($p < 0.01$). This pronounced phenotypic differentiation indicates substantial diversity within the tested materials, consistent with the population variation patterns reported by Zhu et al. in their germplasm evaluation of *Camellia oleifera* [11]. In-depth analysis revealed differences in genetic expression patterns among traits associated with different organs. The GCV of seed/fruit traits (e.g. nut weight and kernel weight) and leaf functional traits (e.g. leaf dry weight and leaf area) were generally greater than 20%, and their family heritabilities (h^2) were consistently higher than 0.5. These traits are closely associated with reproductive resource allocation and

photosynthetic carbon sequestration capacity in trees [40, 41]. The high heritability of these traits suggests that reproductive organ development and the morphological characteristics of functional leaves in *J. mandshurica* are strongly regulated by genetic factors and are relatively less influenced by microenvironmental variation. This finding is consistent with Ren et al.'s assessment that leaf morphological traits in broad-leaved trees are under strong genetic control, and further agrees with the broader consensus that fruit and seed traits in woody oil and broad-leaved tree species generally exhibit moderate to high genetic stability [42-44]. Therefore, directional selection targeting traits with high genetic stability can achieve relatively stable genetic gains.

In contrast, although the phenotypic coefficients of variation for the two early growth traits, seedling height and ground diameter, were relatively high, their family heritabilities were low (0.278 and 0.118, respectively), and the proportion of genetic variation relative to phenotypic variation (GCV/PCV) was also limited. These findings indicate that vegetative growth of *J. mandshurica* during the seedling stage exhibits strong phenotypic plasticity [45, 46]. During the early developmental stages of forest trees, individual growth performance is often influenced by maternal effects (e.g. differences in initial seed size) and microhabitat environmental variation, including soil fertility and water competition, rather than being determined primarily by genetic factors alone [47-49]. Zhang et al. also reported similar environmentally sensitive characteristics in their investigation of early growth variation among *J. mandshurica* families [22]. These findings suggest that, when evaluating *J. mandshurica* germplasm resources at the seedling stage, relying solely on growth traits such as seedling height or ground diameter as selection criteria may compromise the accuracy of early selection due to environmental influences. Therefore, in breeding practice, a comprehensive multi-trait evaluation system centered on traits with high genetic stability should be established based on the genetic parameter characteristics of different

traits. Such an approach can reduce evaluation errors caused by environmental variation and substantially improve the reliability of forest tree germplasm resource evaluation.

Elucidating the correlations and variation patterns among traits is an essential prerequisite for establishing a comprehensive multi-trait evaluation system [50]. In the present study, mean nut diameter in *J. mandshurica* was significantly positively correlated with seed/fruit weight, and leaf area was significantly positively correlated with leaf dry weight. From a plant physiological perspective, these correlations reflect resource allocation patterns during organ development and represent the synergistic strategies of plants in light energy capture and biomass accumulation. This functional coordination is also consistent with the findings of Zhang et al. regarding leaf morphology and photosynthetic physiological adaptation in *J. mandshurica* [20]. PCA further revealed the underlying structure of phenotypic variation. The extracted principal components were primarily associated with seed/fruit characteristics and leaf morphology, indicating that biomass accumulation potential in reproductive organs and the photosynthetic physiological basis of vegetative organs are key drivers of phenotypic differentiation among families [51, 52].

Based on the characterization of phenotypic variation patterns, our study employed a comprehensive multi-trait evaluation method to systematically select superior families. The results showed that the selected families achieved higher genetic gains for high-heritability traits, such as nut weight, kernel weight, and leaf area. In contrast, genetic gains for early vegetative growth traits (seedling height and ground diameter), which are strongly influenced by microenvironmental factors, remained at moderate levels. This asymmetrical distribution of genetic gains highlights the advantages of combined selection: compared with conventional selection based on a single trait, comprehensive multi-trait evaluation enables a more balanced consideration of resource allocation between vegetative growth and

reproductive development in forest trees. Moreover, it reduces the limitations associated with single-trait selection and, to some extent, facilitates an optimal balance between overall genetic gain and environmental adaptability [53, 54]. The superior families identified in the present study provide valuable materials for the subsequent establishment of multi-site variety comparison trials and serve as important phenotypic references for constructing a core collection integrated with molecular marker information.

Phenotypic evaluation is highly susceptible to environmental influences, while molecular markers enable the identification of underlying genetic variation and serve as essential tools for reducing genetic redundancy and constructing core collections [55]. The 10 pairs of EST-SSR primers used in this study exhibited high polymorphism among the tested materials (average PIC = 0.818), demonstrating their strong discriminatory power for the 117 *J. mandshurica* families. Consistently, Wang et al. confirmed the applicability of EST-SSR markers in revealing genetic variation within *J. mandshurica* populations [24]. Because EST-SSR markers are derived from transcribed regions, germplasm screening based on these markers may facilitate the retention of functional genetic variation associated with species adaptation [56].

Regarding the biological characteristics of *J. mandshurica*, its large seeds are primarily dispersed by gravity, resulting in a highly limited natural dispersal distance. This characteristic facilitates the formation of spatial clusters dominated by half-sib or full-sib families within local microhabitats [57]. Similarly, Cui et al. reported that restricted seed dispersal contributes to the formation of family-based spatial clusters during natural regeneration of *J. mandshurica*, leading to a potentially high proportion of closely related individuals within collected germplasm [19]. These distinct population characteristics indicate that targeted sampling strategies are necessary for constructing a core collection to reduce the influence of closely related genetic structures and minimize genetic redundancy.

391 In the construction of core collections, the optimal combination of sampling proportion and
392 algorithm is crucial for preserving rare alleles and reducing genetic redundancy [58]. The traditional 10%
393 sampling strategy is primarily applicable to large-scale germplasm repositories. However, given the
394 relatively limited size of the original germplasm in the present study, low-proportion sampling (10% or
395 15%) resulted in a significant loss of effective alleles **(Table 8)**, making it difficult to capture rare genetic
396 variants generated through spatial isolation in *J. mandshurica*. For small-scale populations, moderately
397 increasing the sampling proportion combined with heuristic optimization algorithms represents an
398 effective approach to minimize genetic information loss [37]. Our results demonstrated that when the
399 sampling proportion was increased to 30%, the SAGD strategy achieved an optimal balance among
400 genetic parameters.

401 The resulting core collection exhibited significantly higher observed and expected heterozygosity
402 than the original germplasm ($p < 0.05$). This increase in heterozygosity indicates that the SAGD
403 algorithm effectively reduced the influence of closely related individual clustering caused by restricted
404 seed dispersal, thereby decreasing the representation of redundant individuals with similar genetic
405 backgrounds. Importantly, this finding is in line with previous related studies [16, 59]. The 35 selected
406 core germplasm accessions achieved excellent allele retention despite the small population size, forming
407 a high-quality germplasm pool characterized by high polymorphism and relatively low levels of
408 inbreeding. These core materials, validated through both phenotypic and molecular assessments, can
409 serve as the foundation for establishing a first-generation seed orchard of *J. mandshurica*. By reducing
410 conservation and management costs, they may also help mitigate the risk of inbreeding depression during
411 subsequent breeding processes, ultimately providing a robust material basis for the directional
412 improvement and scientific management of this species' genetic resources.

Conclusions

The comprehensive phenotypic evaluation of *J. mandshurica* families in this study demonstrated that family heritabilities for seed/fruit and leaf traits were generally high ($h^2 > 0.5$), whereas early-growth traits, including seedling height and ground diameter, exhibited pronounced phenotypic plasticity. Based on these variation patterns, the superior families selected through comprehensive multi-trait evaluation effectively reduced the influence of environmental variation during the seedling stage and achieved relatively stable genetic gains for high-heritability traits, such as nut weight and leaf area. Regarding molecular evaluation and core collection construction, the SAGD strategy combined with a 30% sampling proportion demonstrated excellent applicability. The 35 core germplasm accessions constructed using this optimal scheme retained 99.5% of the alleles present in the original germplasm, and both their observed and expected heterozygosities were significantly higher than those of the original germplasm ($p < 0.05$). This increase in heterozygosity indicates that the applied strategy effectively reduced the representation of closely related, redundant individuals within the originally collected germplasm, thereby maintaining high genetic diversity with a limited sample size. Ultimately, the superior families selected and the core collection constructed in our study provide a fundamental material basis for the directional improvement and subsequent scientific management of *J. mandshurica*.

Acknowledgements The authors express gratitude to Jilin Agricultural University for providing essential equipment and an experimental platform for this study, as well as to local forestry-related organizations for their invaluable assistance in collecting *Juglans mandshurica* fruits from various provenances.

Author contributions Qinhui Zhang: Conceptualization, Writing-original draft, Visualization. Yitong Chen: Writing-original draft, Methodology. Taijin Zhang: Formal analysis, Software. Chengyao Wang: Investigation, Data curation, Visualization. Xiyang Zhao: Conceptualization, Writing-review & editing.

Funding This work was supported by the Major Technological R&D Project for the Forestry and Grassland Industry of Jilin Province (JLLC-2026-51).

Data availability The datasets generated and/or analysed during the current study are available in the Zenodo repository, https://doi.org/10.5281/zenodo.23496595.

Declarations

Ethics approval and consent to participate The field collection of wild *J. mandshurica* seeds was permitted and supported by the Jilin Changbai Mountain Forestry Industry Group. The authors confirm that all experimental research on plants (either cultivated or wild), including the collection of plant material, complied with relevant institutional, national, and international guidelines and legislation.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Sang YP, Long ZQ, Dan XM, Feng JJ, Shi TT, Jia CF, et al. Genomic insights into local adaptation and future climate-induced vulnerability of a keystone forest tree in East Asia. Nat Commun. 2022; 13 (1), 6541. https://doi.org/10.1038/s41467-022-34206-8.
2. Archambeau J, Bianchi S, Buiteveld J, Callejas-Díaz M, Cavers S, Hallingbäck H, et al. Managing forest genetic resources for an uncertain future: findings and perspectives from an international conference. Tree Genet Genomes. 2023; 19 (3): 26. https://doi.org/10.1007/s11295-023-01603-z.
3. Dong ML, Zhou L, Yu N, Li RS, Wu SJ, Yang JC, et al. Genetic parameters and selection responses for important breeding traits in *Liquidambar formosana* based on a provenance-family trial. Forests. 2023; 14 (12), 2293. https://doi.org/10.3390/f14122293.
4. Hao SW, Zhang DL, Wen YF. Leaf phenotypic variation of Chuanqian crape myrtle in natural populations. HortScience. 2025; 60 (3): 399-403. https://doi.org/10.21273/HORTSCI18375-24.
5. Yan W, Xiang YQ, Gao M, Deng RY, Sun Y, Wan RP, et al. Phenotypic diversity and provenance variation of *Cupressus funebris*: a case study in the Sichuan Basin, China. PeerJ. 2024; 12, e18494. https://doi.org/10.7717/peerj.18494.
6. Li ZM, Gao CJ, Che FX, Li J, Wang L, Cui K. Trunk distortion weakens the tree productivity revealed by half-sib progeny determination of *Pinus yunnanensis*. BMC Plant Biol. 2024; 24 (1), 629. https://doi.org/10.1186/s12870-024-05350-8.
7. Uddin S, Jaskani MJ, Deng ZN, Maqbool R, Naqvi SA, Parajuli S, et al. Phenotypic and molecular-markers-based assessment of Jamun (*Syzygium cumini*) genotypes from Pakistan. Horticulturae. 2024; 10 (8), 879. https://doi.org/10.3390/horticulturae10080879.

8. Archambeau J, Benito Garzón M, Barraquand F, de Miguel Vega M, Plomion C, González-Martínez SC. Combining climatic and genomic data improves range-wide tree height growth prediction in a forest tree. Am Nat. 2022; 200 (4), E141-E159. https://doi.org/10.1086/720619.

9. Soriano JM, Romero C, Vilanova S, Llácer G, Badenes ML. Genetic diversity of loquat germplasm (*Eriobotrya japonica* (Thunb) Lindl) assessed by SSR markers. Genome, 2005; 48 (1), 108-114. https://doi.org/10.1139/g04-101

10. He ZL, Liu CX, Wang XN, Wang R, Chen YZ, Tian Y. Assessment of genetic diversity in *Camellia oleifera* Abel. accessions using morphological traits and simple sequence repeat (SSR) markers. Breed Sci. 2020; 70, 586-593. https://doi.org/10.1270/jsbbs.20066.

11. Zhu YZ, Liang DY, Song ZJ, Tan Y, Guo XL, Wang DL. Genetic diversity analysis and core germplasm collection construction of *Camellia oleifera* based on fruit phenotype and SSR data. Genes. 2022; 13 (12), 2351. https://doi.org/10.3390/genes13122351

12. dos Santos CC, de Andrade LRB, do Carmo CD, de Oliveira EJ. The development of thematic core collections in cassava based on yield, disease resistance, and root quality traits. Plants. 2023; 12 (19), 3474. https://doi.org/10.3390/plants12193474.

13. Wang WJ, Ma H, Qi YD, Ye JX, Lu M, Wei XP. Scientific evolution and translational horizons of plant core germplasm: a global bibliometric synthesis and strategic insights. Front Plant Sci. 2026; 17, 1771164. https://doi.org/10.3389/fpls.2026.1771164.

14.Brown AHD. Core collections: a practical approach to genetic resources management. Genome. 1989; 31 (2), 818-824. https://doi.org/10.1139/g89-144.

15. Gu R, Fan SH, Wei SP, Li JR, Zheng SH, Liu GL. Developments on core collections of plant genetic resources: do we know enough? Forests. 2023; 14 (5), 926. https://doi.org/10.3390/f14050926.

16. Wu HB, Duan AG, Wang XH, Chen ZY, Zhang X, He GP, et al. Construction of a core collection of germplasms from Chinese fir seed orchards. Forests. 2023; 14 (2), 305. https://doi.org/10.3390/f14020305.

17. Yang WH, Mao MY, Qin JQ, Nasar J, Pan ZD, Zhou LJ, et al. Genetic diversity and core collection construction of glutinous rice landraces in 'He' cultivation zone of Guizhou using SSR sequencing. PLoS One. 2026; 21 (3), e0343623. https://doi.org/10.1371/journal.pone.0343623.

18. Li X, Cai KW, Zhang QH, Pei XN, Chen S, Jiang LP, et al. The Manchurian walnut genome: insights into juglone and lipid biosynthesis. GigaScience. 2022; 11, giac057. https://doi.org/10.1093/gigascience/giac057.

19. Cui DH, Wang QC, Ma SJ, Zhang Y, Xu LQ, Yan LM. Natural regeneration patterns of *Juglans mandshurica* in different habitats. Forests. 2024; 15 (7), 1201. https://doi.org/10.3390/f15071201.

20. Zhang QH, Peng D, Pei XN, Chen YT, Yan ZH, Kang MR, et al. Unveiling high-light-use efficiency in *Juglans mandshurica*: Adaptive insights from leaf morphology and physiology. For Ecol Manage. 2025; 586, 122705. https://doi.org/10.1016/j.foreco.2025.122705.

21. Li ZX, Wang WH, Zhang HX, Liu JH, Shi BY, Dai WZ, et al. Diversity in fruit morphology and nutritional composition of *Juglans mandshurica* Maxim in Northeast China. Front Plant Sci. 2022; 13, 820457. https://doi.org/10.3389/fpls.2022.820457.

22. Zhang QH, Chen S, Qu GZ, Yang YC, Lu ZM, Wang J, et al. Provenance and family variations in early growth of Manchurian walnut (*Juglans mandshurica* Maxim.) and selection of superior families. PLoS One. 2024; 19 (3), e0298918. https://doi.org/10.1371/journal.pone.0298918.

23. Zhang QH, Zhang XX, Yang YC, Xu LF, Feng J, Wang JY, et al. Genetic diversity of *Juglans mandshurica* populations in Northeast China based on SSR markers. Front Plant Sci. 2022; 13,

931578. https://doi.org/10.3389/fpls.2022.931578

24. Wang ZH, Zhang H, Tong BQ, Han B, Liu D, Zhang P, et al. EST-SSR marker-based investigation on genetic diversity and genetic structure of *Juglans mandshurica* Maxim. in Shandong Province of China. Genet Resour Crop Evol. 2023; 70 (3), 981-991. https://doi.org/10.1007/s10722-022-01482-8.

25. Pérez-Harguindeguy N, Díaz S, Garnier E, Lavorel S, Poorter H, Jaureguiberry P, et al. Corrigendum to: New handbook for standardised measurement of plant functional traits worldwide. Aust J Bot. 2016; 64 (8), 715-716. https://doi.org/10.1071/BT12225_CO.

26. Zhao P, Woeste KE. DNA markers identify hybrids between butternut (*Juglans cinerea* L.) and Japanese walnut (*Juglans ailantifolia* Carr.). Tree Genet Genom. 2011; 7, 511-533. https://doi.org/10.1007/s11295-010-0352-4.

27. Li DY, Long C, Pang XM, Ning DL, Wu T, Dong ML, et al. The newly developed genomic-SSR markers uncover the genetic characteristics and relationships of olive accessions. PeerJ. 2020; 8, e8573. https://doi.org/10.7717/peerj.8573.

28. Han Q, Zhong CL, Zhang Y, Jiang QB, Chen Y, Chen Z, et al. Genetic variation and selection of *Casuarina junghuhniana* provenances at Lingao, Hainan. For Res. 2017; 30(4): 595-603. https://doi.org/10.13275/j.cnki.lykxyj.2017.04.009.

29. Metougui ML, Mokhtari M, Maughan PJ, Jellen EN, Benlhabib O. Morphological variability, heritability and correlation studies within an argan tree population (*Argania spinosa* (L.) Skeels) preserved in situ Int J Agric For. 2017; 7 (2), 42-51. https://doi.org/10.5923/j.ijaf.20170702.02.

30. Liu YX, Yan KY, He W, Pan HX. Genetic variation of fiber traits in *Populus deltoides* clones. J Nanjing For Univ (Nat Sci Ed). 2020; 44 (2), 67-74. https://doi.org/10.3969/j.issn.1000-

2006.201904054.

31. Kaviriri DK, Yang L, Zhao CL, Pei XN, Fan ZY, Wang JY, et al. Morphological growth performance and genetic parameters on Korean pine in Northeastern China. Silvae Genet. 2023; 72(1): 34-48. https://doi.org/10.2478/sg-2023-0004.

32. Scatigna AV, Oliveira FA, Mantello CC, Francisco PM, Souza AP, Simões AO. Microsatellite markers for studies with the carnivorous plant *Philcoxia minensis* (Plantaginaceae). Appl Plant Sci. 2015; 3 (8), 1500035. https://doi.org/10.3732/apps.1500035.

33. Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics. 2012; 28 (19), 2537-2539. https://doi.org/10.1093/bioinformatics/bts460.

34. Nagy S, Poczai P, Cernák I, Gorji AM, Hegedűs G, Taller J. PICcalc: an online program to calculate polymorphic information content for molecular genetic studies. Biochem. 2012; Genet 50 (9), 670-672. https://doi.org/10.1007/s10528-012-9509-1.

35. Liu KJ, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 2005; 21 (9), 2128-2129. https://doi.org/10.1093/bioinformatics/bti282.

36. Schoen DJ, Brown AHD. Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. Proc Natl Acad Sci U.S.A. 1993; 90 (22), 10623-10627. https://doi.org/10.1073/pnas.90.22.10623.

37. Odong TL, Jansen J, van Eeuwijk FA, van Hintum TJL. Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. Theor Appl Genet. 2013; 126 (2), 289-305. https://doi.org/10.1007/s00122-012-1971-y.

38. Marco de Lima B, Cappa EP, Silva-Junior OB, Garcia C, Mansfield SD, Grattapaglia D. Quantitative

567 genetic parameters for growth and wood properties in *Eucalyptus* "urograndis" hybrid using near-
568 infrared phenotyping and genome-wide SNP-based relationships. PLoS One. 2019; 14 (6),
569 e0218747. https://doi.org/10.1371/journal.pone.0218747.
570 39. Zheng Y, Feng MS, Li X, Huang XY, Chen G, Bai WY, et al. Phenotypic variation analysis and
571 excellent clone selection of *Alnus cremastogyne* from different provenances. Plants. 2023; 12 (18),
572 3259. https://doi.org/10.3390/plants12183259.
573 40. Wright IJ, Reich PB, Westoby M, Ackerly DD, Baruch Z, Bongers F, et al. The worldwide leaf
574 economics spectrum. Nature. 2004; 428 (6985), 821-827. https://doi.org/10.1038/nature02403.
575 41. Climent J, Alía R, Karkkainen K, Bastien C, Benito-Garzon M, Bouffier L, et al. Trade-offs and trait
576 integration in tree phenotypes: Consequences for the sustainable use of genetic resources. Current
577 Forestry Reports. 2024; 10, 196-222. https://doi.org/10.1007/s40725-024-00217-5.
578 42. Sudrajat DJ. Genetic variation of fruit, seed, and seedling characteristics among 11 populations of
579 white jabon in Indonesia. For Sci Technol. 2016; 12 (1), 9-
580 15. https://doi.org/10.1080/21580103.2015.1007896.
581 43. Ren JS, Ji XY, Wang CH, Hu JJ, Nervo G, Li JH. Variation and genetic parameters of leaf
582 morphological traits of eight families from *Populus simonii* × *P. nigra*. Forests. 2020; 11 (12),
583 1319. https://doi.org/10.3390/f11121319.
584 44. Xie WW, Yu YY, Xie YQ, Li Y, Huang Y, Lin WJ, et al. Genome-wide association study of fruit traits
585 using 109 germplasm accessions of *Camellia oleifera*. Biology. 2026; 15 (6),
586 483. https://doi.org/10.3390/biology15060483.
587 45. Khaliqi A, Rafii MY, Mazlan N, Jusoh M, Oladosu Y. Genetic analysis and selection criteria in
588 Bambara groundnut accessions based yield performance. Agronomy. 2021; 11 (8),

1634. https://doi.org/10.3390/agronomy11081634.

46. Sadras VO. Bread and hummus: Trait connectance and correlation pleiades in grain crops. J Exp Bot. 2024; 75 (22), 6903-6908. https://doi.org/10.1093/jxb/erae374.

47. Roach DA, Wulff RD. Maternal effects in plants. Annu Rev Ecol Syst. 1987; 18, 209-235. https://doi.org/10.1146/annurev.es.18.110187.001233.

48. Zas R, Sampedro L. Heritability of seed weight in Maritime pine, a relevant trait in the transmission of environmental maternal effects. Heredity. 2015; 114 (1), 116-124. https://doi.org/10.1038/hdy.2014.76.

49. Zeng XQ, Durka W, Welk E, Fischer M. Heritability of early growth traits and their plasticity in 14 woody species of Chinese subtropical forest. J Plant Ecol. 2017; 10 (1), 222-231. https://doi.org/10.1093/jpe/rtw086.

50. Missanjo E, Matsumura J. Multiple trait selection index for simultaneous improvement of wood properties and growth traits in *Pinus kesiya* Royle ex Gordon in Malawi. Forests. 2017; 8 (4), 96. https://doi.org/10.3390/f8040096.

51. Wang YL, Li Y. Genetic diversity analysis of phenotypic traits among 37 Xanthoceras sorbifolium elite germplasms. J For Res. 2022; 27 (2), 140-147. https://doi.org/10.1080/13416979.2021.2009094.

52. Yadav V, Mishra DS, Rane J, Apparao VV, Kumar P, Kumar P, et al. Uncovering key traits for breeding high-yield and nutrient-rich genotypes of wood apple (*Feronia limonia* (L.) Swingle). Int J Fruit Sci. 2026; 26 (1), 2691559. https://doi.org/10.1080/15538362.2026.2691559.

53. Climent J, Prada MA, Calama R, Chambel MR, Sánchez de Ron D, Alía R. To grow or to seed: Ecotypic variation in reproductive allocation and cone production by young female Aleppo pine

(*Pinus halepensis*, Pinaceae). Am J Bot, 2008; 95 (7), 833-842. https://doi.org/10.3732/ajb.2007354.

54. Lu XH, Liu JT, Lei FY, Zheng YX. Multi-trait comprehensive evaluation and elite germplasm screening of 43 celery germplasm resources. Horticulturae. 2026; 12 (6), 699. https://doi.org/10.3390/horticulturae12060699.

55. Billings GT, Jones MA, Rustgi S, Hulse-Kemp AM, Campbell BT. Population structure and genetic diversity of the Pee Dee cotton breeding program. G3: Genes Genomes Genetics. 2021; 11 (7), jkab145. https://doi.org/10.1093/g3journal/jkab145.

56. Yang YY, He RQ, Zheng J, Hu ZH, Wu J, Leng PS. Development of EST-SSR markers and association mapping with floral traits in Syringa oblata. BMC Plant Biol. 2020; 20 (1), 436. https://doi.org/10.1186/s12870-020-02652-5.

57. Bittencourt JVM, Sebbenn AM. Patterns of pollen and seed dispersal in a small, fragmented population of the wind-pollinated tree *Araucaria angustifolia* in southern Brazil. Heredity. 2007; 99 (6), 580-591. https://doi.org/10.1038/sj.hdy.6801019.

58. De Beukelaer H, Davenport GF, Fack V. Core Hunter 3: flexible core subset selection. BMC Bioinform. 2018; 19 (1), 203. https://doi.org/10.1186/s12859-018-2209-z.

59. Garcia-Lor A, Luro F, Ollitrault P, Navarro L. Comparative analysis of core collection sampling methods for mandarin germplasm based on molecular and phenotypic data. Ann Appl Biol. 2017; 171 (3): 327-339. https://doi.org/10.1111/aab.12376.

Authors and Affiliations

Qinhui Zhang[1] · Yitong Chen[2,†] · Taijin Zhang[3] · Chengyao Wang[1] · Xiyang Zhao[1,4]

✉ Xiyang Zhao

xiyangz@jlau.edu.cn

[1] Jilin Provincial Key Laboratory of Tree and Grass Genetics and Breeding, College of Forestry and Grassland Science, Jilin Agricultural University, Changchun 130118, China

[2] College of Horticulture, Jilin Agricultural University, Changchun 130118, China

[3] Key Laboratory of Bio-resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, 610065, China

[4] Jilin Changbai Mountain Forestry Industry Group, Yanji, 133000, China

† Co-first authors contributed equally to this paper.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 2026Supplementarymaterial.xlsx