

The genetic diversity and population structure of Kudouzi (*Sophora alopecuroides*) population were revealed by using SNP markers combined with seed phenotypic traits

Cunkai lu

Agricultural College, Shihezi University

Fanyan Ma

Agricultural College, Shihezi University

Panxin Niu

Agricultural College, Shihezi University

Zhao Zhang

Bozhou Water Conservancy Irrigation Test Station

Weiting Chen

Agricultural College, Shihezi University

Ping Jiang

Agricultural College, Shihezi University

Mei Wang

wangm1205@shzu.edu.cn

Agricultural College, Shihezi University

Guangming Chu

Agricultural College, Shihezi University

Xiang Huang

Agricultural College, Shihezi University

Research Article

Keywords: *Sophora alopecuroides*, Single nucleotide polymorphism, Genetic diversity, Seed phenotypic traits, Population structure

Posted Date: April 29th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-6061561/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background: *Sophora alopecuroides* L. is a perennial herb of Leguminosae. It is mainly distributed in northwest China and has important medicinal, feeding and ecological restoration values. However, in recent years, due to the intensification of human activities, its wild population resources have plummeted and genetic diversity has continued to decline. In order to fully reveal the genetic diversity and population structure characteristics of *S. alopecuroides* in the natural distribution area. In this study, Single nucleotide polymorphism (SNP) molecular markers combined with seed phenotypic traits were used to systematically study 65 wild samples of *S. alopecuroides* in northwest China.

Results: The results showed that the coefficient of variation (CV) of 8 phenotypic traits of *S. alopecuroides* seeds ranged from 2.87 % to 7.94 %, and the diversity index (H) ranged from 1.639 to 1.767. There was a significant correlation between phenotypic traits ($P < 0.01$), indicating that the phenotypic diversity of *S. alopecuroides* seeds was rich. Cluster analysis based on phenotypic traits divided the *S. alopecuroides* population into two groups. At the same time, based on SNP molecular markers, the genetic diversity of *S. alopecuroides* was relatively low. The average values of expected heterozygosity (H_e), observed heterozygosity (H_o) and single nucleotide diversity index (P_i) were 0.22, 0.17 and 0.19, respectively. The results of molecular variance analysis (AMOVA) showed that the level of variation between individuals (132.83%) was higher than that between populations and within populations. The Pairwise population differentiation (F_{st}) was between 0.00 and 0.04, which confirmed that there was no obvious differentiation among the populations. Population structure analysis, principal component analysis (PCA) and phylogenetic tree analysis roughly divided all populations into two clusters, which was consistent with the phenotypic clustering results. It is worth noting that the distribution patterns of the samples in the Southern Tianshan Mountains (TSNL) and the Altai Mountains (ARTS) are more complex. In addition, redundancy analysis (RDA) showed that the cumulative interpretation rates of environmental factors on phenotypic traits and genetic diversity were 99.75 % and 67.89 %, respectively. Among them, the mean temperature of the driest quarter (MTD) and annual mean wind speed (YWS) were identified as the primary factors influencing phenotypic traits, while the precipitation of the coldest quarter (PC), isothermality (ISO), and precipitation of the wettest quarter (PWE) have an important impact on genetic diversity.

Conclusions: This study provides a foundation for the genetic evaluation and conservation of *S. alopecuroides* genetic resources in China and offers important insights for its breeding programs.

Background

Sophora alopecuroides L. is a perennial herb of the Leguminosae family that is widely distributed across arid and semi-arid regions of West and Central Asia. In China, it is predominantly found in Xinjiang, Gansu, Ningxia, and Inner Mongolia [1, 2]. Studies have shown that *S. alopecuroides* exhibits strong root development, drought resistance, salinity-alkalinity tolerance, and nitrogen fixation ability [1, 3]. As such, it has become a crucial pioneer plant for environmental protection in northwest China, owing to its

robust drought resistance, alkalinity tolerance, and wind-sand resistance. In addition to its ecological significance, *S. alopecuroides* possesses substantial medicinal and agricultural value. Its seeds, aerial parts, and roots contain bioactive compounds such as alkaloids, flavonoids, volatile oils, steroids, polysaccharides, and free fatty acids [4–7]. These compounds underpin its use in traditional medicine and potential application in biological pesticides [8], green manure [9], and high-quality forage [10].

Genetic diversity and population structure play a key role in species conservation, utilization and breeding [11]. Gao et al. [12] reported that the concentrations of key medicinal compounds, such as oxymatrine and total flavonoids, vary significantly in *S. alopecuroides* seeds from different geographical regions. Similarly, An et al. [13] found that the contents of genistein and sophoridine in seeds from northern Xinjiang were generally higher than those from southern Xinjiang. In addition, gene flow between different *S. alopecuroides* populations is limited, resulting in high genetic differentiation between populations [14]. Despite its ecological and medicinal importance, few fundamental studies on the large-scale population resources and genetic diversity of *S. alopecuroides* have been conducted.

Seed traits are crucial for population reproduction and preservation, serving as important indicators of species inheritance [15, 16]. Phenotypic plasticity, the ability of a single genotype to produce different phenotypes in varying environments, is largely a result of genetic selection and is closely related to phenotypic diversity [17–19]. Significant differences in seed traits, such as size and thousand-grain weight, among different wild populations of *S. alopecuroides* suggest rich ecological clusters and complex genetic backgrounds [20]. However, the correlation between seed phenotypic traits and genetic diversity in *S. alopecuroides* has not been thoroughly investigated.

Genotyping-by-Sequencing (GBS) is a high-throughput, next-generation sequencing method that reduces genomic complexity using restriction enzymes. This technique generates high-density single nucleotide polymorphism (SNP) markers by tagging random genomic deoxyribonucleic acid (DNA) fragments common across samples with unique short DNA sequences (barcodes), pooling samples into a single sequencing run to detect and label mutations at a relatively low cost [21]. Currently, GBS serves as a powerful genotyping tool, widely applied in genetic mapping, marker-assisted selection [22, 23], diversity analysis, germplasm and species identification [24–25]. Among molecular markers, SNP are key tools in population and quantitative genetics. SNP represent single-base differences among individuals of the same species and are the most abundant form of genetic variation [26]. As third-generation markers, SNP offer advantages such as biallelic nature, high density, genetic stability, easy detection, and genome-wide distribution [27, 28]. With advances in sequencing and bioinformatics, SNP have been extensively used to study genetic diversity, population structure, and differentiation [29–33]. However, their application in the genetic diversity analysis of *S. alopecuroides* remains unreported.

In this study, seeds were collected from 65 wild populations of *S. alopecuroides* across diverse geographic regions. Phenotypic traits were measured, and SNP markers were employed to assess the genetic structure and diversity within these populations. The main objectives of this study are as follows: (1) what is the level of genetic diversity and population structure of *S. alopecuroides* in these different

populations? (2) What are the effects of geographical and environmental factors on the genetic diversity of *S. alopecuroides*? Based on the findings, this research aims to provide scientific basis for the effective conservation, management, and utilization of *S. alopecuroides* germplasm, as well as for the development of high-quality cultivars to meet market demands.

Methods

Plant Materials

The field collection of *S. alopecuroides* samples was approved by the Chinese government and carried out in compliance with all relevant national regulations. Species identification was confirmed by Professor Zhuowen Zhang (College of Horticulture and Forestry, Huazhong Agricultural University), and voucher specimens were deposited in the Department of Forestry, College of Agriculture, Shihezi University (Accession Nos: SHAF20230801–SHAF20230865).

Based on the main distribution range of *S. alopecuroides* in China [34] and the maturity stage of the plants, a total of 65 samples were collected between August and October 2023 from Xinjiang, Gansu, Ningxia, Inner Mongolia, and Shaanxi provinces (Fig. 1, Table 1 and Table S1). Sampling sites were spaced at least 20 km apart. At each site, 10–15 healthy and mature individuals were randomly selected, maintaining a minimum distance of 10 meters between sampled plants. From each individual, 80–120 mature seeds were collected from multiple pods. Seeds from all individuals within each site were pooled to form a composite sample representing the genetic diversity of that population. After field collection and natural air-drying, seeds were stored in a cool and dry environment [35].

Table 1 Geographic information of *Sophora alopecuroides* population

Population	Location	Sample	Number
YLHG	Yili, Xinjiang, China	G1, G2, G3, G4, G5	5
TSBL	Bortala, Xinjiang, China	G6, G7, G15, G16, G17, G18, G19, G20, G21, G22	10
TCDQ	Tacheng, Xinjiang, China	G8, G9, G10, G11	4
ARTS	Altay, Xinjiang, China	G12, G13, G14	3
KLSM	Bayingolin; Hotan; Kashgar, Xinjiang, China	G23, G24, G28, G29, G30, G37, G38, G39, G40, G41, G42, G43, G44, G46	14
TSNL	Aksu, Xinjiang, China	G25, G26, G27, G31, G32, G33, G34, G35, G36, G45	10
THPD	Turpan; Hami; Xinjiang, China	G47, G48, G49, G50, G51	5
QLSM	Gansu, China	G52, G53, G54, G55	4
HTGY	Ningxia; Shaanxi, China	G56, G57, G58, G59, G60, G65	6
NMGY	Inner Mongolia, China	G61, G62, G63, G64	4

Based on their geographic origin, the 65 samples were grouped into 10 populations. Global Satellite Positioning System (GPS) coordinates for each sampling location were recorded using handheld GPS devices, and bioclimatic variables were obtained from the WorldCLIM global high resolution climate database (<http://www.worldclim.org/>) and local meteorological stations (Table S2 and S3). A subset of seeds from each composite sample was randomly selected and sent to Beijing Nuohe Zhiyuan Bioinformatics Technology Co., Ltd. for GBS analysis.

Measurement of seed phenotypic traits

We used a Wanshen SC-G automatic seed testing instrument (Hangzhou Wanshen Detection Technology Co., Ltd., China) to measure the length, width, diameter, roundness, area, perimeter and shape index data of 65 samples of *S. alopecuroides* seeds. Each index was measured three times, 200 seeds were randomly selected for each repetition, and the mean value was taken. The thousand-grain weight was measured by a one-thousandth electronic balance, and 1,000 seeds were randomly selected each time. Three measurement replicates were set up, and the average value was taken [20, 36].

Statistics and analysis of phenotypic traits

Microsoft Excel 2016 was used to analyze the experimental data, Origin Pro 2021 software was used for plotting, SPSS 27.0 software was used to test the homogeneity of variance, and variance analysis was performed on the seed trait data. The least significant test (LSD) method was based on one-way analysis

of variance (one-way ANOVA), which was performed $\pm$ the mean value of the trait. The coefficient of variation (CV) and phenotypic genetic Shannon diversity index (H) were calculated in Excel as follows:

$$CV = (\text{Standard deviation} \times \text{Mean}) \times 100\% \quad (1)$$

The observed values of each trait were divided into 10 levels according to the mean value (X_i) and standard deviation (SD). Each 0.5 SD was 1 level, from the first level ($X_i < X - 2 \text{ SD}$) to the 10th level ($X_i \geq X + 2 \text{ SD}$). The relative frequency of each group was used to calculate H [42].

$$H = - \sum P_i \ln P_i \quad (2)$$

In Eq. (2), i is the classification of a trait, and P_i is the percentage of the number of materials in the total number of materials in the i th grade of the trait.

On the basis of the mean values of the seed phenotypic traits, cluster analysis was carried out via OriginPro 2021 software. By controlling the data quality, the differences in the quantitative phenotypic traits of the 65 samples in this study were visually displayed. The cluster analysis of the phenotypic traits of 65 samples was carried out via the longest distance method of square Euclidean distance. To eliminate the influence of dimension on the clustering relationship, the seed trait data were standardized by the z value and then clustered [46]. The z value standardized calculation formula is as follows:

$$X^* = \frac{x - \bar{x}}{\sigma} \quad (3)$$

In Eq. (3), X^* is the normalized value of z value, x is the original value, $\bar{x}$ is the sample mean, and σ is the sample standard deviation.

DNA extraction, GBS library construction and sequencing (dup: abstract ?)

The DNA of 65 *S. alopecuroides* seeds from different provenances was extracted via the CTAB method [37]. The integrity of the extracted DNA was detected via 1% agarose gel electrophoresis, and the concentration and quality of the extracted DNA (OD260/280 ratio) were measured via a Nanodrop UV–visible spectrophotometer. After the quality test was performed, the extracted DNA stock solution was stored at -20°C until use.

The construction and sequencing of the GBS library begins with the digestion of genomic DNA. First, 0.1–1 μg of genomic DNA was extracted and digested with restriction endonucleases to obtain a suitable marker density. The two ends of the digested DNA fragment are connected with P1 and P2

adaptor sequences (which can be complementary to the digested DNA gap). The tag sequences containing P1 and P2 adapters at both ends were amplified by polymerase chain reaction (PCR). The amplified DNA fragments were subsequently separated and recovered via the DNA fragment mixing pool, and the DNA fragments in the expected size range were selected via agarose gel electrophoresis. After library construction was completed, a Qubit 2.0 instrument was used for preliminary quantification, and the library was diluted to 1 ng/μl. Then, an Agilent 2100 instrument was used to determine the size of the insert in the library. After the insert size was in line with expectations, the fluorescence polymerase chain reaction (q-PCR) method was used to accurately quantify the effective concentration of the library (the effective concentration of the library > 2 nM) to ensure the quality of the library. After the library inspection was performed, Illumina HiSeq PE150 sequencing was performed after pooling different libraries according to the effective concentration and the target offline data volume.

DNA extraction, GBS library construction and sequencing

The original image data files obtained via Illumina HiSeq sequencing were converted into sequenced reads via base-calling analysis, and the results were stored in FASTQ file format. The raw data obtained by sequencing were filtered to filter out the reads containing the adaptor sequence; when the N content in the single-ended sequencing read exceeded 10% of the length ratio of the read, the paired reads needed to be removed. When the number of low-quality (≤ 5) bases contained in the single-end sequencing read exceeds 50% of the read length ratio, this pair of paired reads needs to be removed. Finally, high-quality, clean data were obtained for subsequent analysis.

Among the 65 samples, the sample labeled G60—which had the highest number of sequence tags—was selected to perform de novo stack clustering and construct a quasi-reference sequence for alignment. The clean reads of all samples were aligned to this quasi-reference using the BWA software [38] (parameters: mem -t 4 -k 32 -M). The alignment files were then sorted using SAMTOOLS [39] (sort function). SNPs across the population were detected using SAMTOOLS and related tools, yielding a total of 58,448 raw SNPs. After filtering with a call rate threshold of 0.8 and a minimum allele frequency (MAF) of 0.03, a final set of 10,584 high-quality SNP loci was retained for subsequent analyses.

Population structure and diversity analysis

The samples were selected on the basis of the filtered SNP loci, and the distance matrix was calculated via TreeBest software. A phylogenetic tree was subsequently constructed via the neighbor-joining method [40]. The feature vectors and eigenvalues were calculated via GCTA software, and the principal component analysis (PCA) distribution map was drawn R v. 4.4.3. Admixture was used to analyze the population structure [41]. First, the input file-Ped file of PLINK was created, and then the population genetic structure and population pedigree information were constructed via admixture software. The software Arlequin [42] was used to calculate the genetic diversity indices of each population: observed heterozygosity (H_o), expected heterozygosity (H_e), molecular variance analysis (AMOVA) and F_{st} between groups to measure possible differences between different groups. The method of Nei & Li [43]

was used to analyze the population nucleotide diversity (π) and calculate the single nucleotide diversity index (Pi).

Climatic association analysis

To estimate the degree to which genomic variation is influenced by environmental variables, Redundancy analysis (RDA) was performed to examine the correlations between phenotypic characteristics, genetic diversity parameters and environmental factors with the package *vegan* in the R v. 4.4.3 statistical environment [44]. RDA involves a multiple linear regression followed by a PCA on the matrix of regression-fitted values. A dependent matrix of minor allele frequencies for each population and an independent matrix of environmental variables were included. To avoid high collinearity, we excluded those with a VIF over 20 [45].

Results

Diversity analysis of seed phenotypic traits

We measured 8 phenotypic traits of 65 *S. alopecuroides* samples and calculated their CV and H values (Table 2). The results revealed that the CV of the 8 phenotypic traits ranged from 2.87% to 7.94%, and the average CV was 4.48%. Among them, the CV of thousand-grain weight was the largest, whereas the CV of width was the smallest. The diversity indices of the 8 phenotypic traits ranged from 1.639 to 1.767, with an average of 1.713. All indicators were normally distributed (Fig. 2). This result indicated that the diversity of the 8 phenotypic traits in the 65 *S. alopecuroides* samples in this study was higher.

Table 2 Diversity of seed phenotypic traits of *S. alopecuroides* samples

parameter	Length (mm)	Width (mm)	Diameter (mm)	Roundness	Area (mm ²)	Perimeter (mm)	Shape	Weight (g)
Maximum	4.271	3.023	3.626	0.840	1.434	10.356	12.136	29.060
Minimum	3.461	2.641	3.074	0.700	1.200	7.464	10.129	20.056
Mean	3.728	2.838	3.302	0.764	1.317	8.607	10.878	23.776
SD	0.161	0.081	0.105	0.029	0.049	0.553	0.388	1.888
CV	4.32%	2.87%	3.19%	3.80%	6.43%	3.56%	3.69%	7.94%
H	1.654	1.767	1.725	1.736	1.721	1.639	1.737	1.724

SD: standard deviation; CV: coefficient of variation; H: Shannon 's diversity index; Shape: shape index; Weight: thousand-grain weight

Correlation analysis of seed phenotypic traits

Correlation analysis was performed on the phenotypic traits of all the tested *S. alopecuroides* seeds (Fig. 3). The results revealed that length was significantly positively correlated with width, diameter, shape index, area, perimeter and thousand-grain weight ($P < 0.01$) but negatively correlated with roundness ($P < 0.01$). Width was significantly positively correlated with diameter, area, perimeter and thousand-grain weight ($P < 0.01$) but weakly correlated with roundness and the shape index. There was a significant positive correlation between diameter and shape index, area, perimeter and thousand-grain weight ($P < 0.01$) and a significant negative correlation with roundness ($P < 0.01$). The roundness was significantly negatively correlated with the shape index, area and perimeter ($P < 0.01$), and the correlation with thousand-grain weight was weak. The shape index was significantly positively correlated with area and perimeter ($P < 0.01$) and weakly negatively correlated with thousand-grain weight. Area was significantly positively correlated with perimeter and thousand-grain weight ($P < 0.01$), and thousand-grain weight was also significantly positively correlated with perimeter ($P < 0.01$).

Cluster analysis of seed phenotypic traits

Systematic cluster analysis of the seed phenotypic traits of 65 *S. alopecuroides* samples revealed that the samples were divided into two groups (Fig. 4). Among them, group I is shown in blue in the figure and includes 19 samples from the YLHG, TSBL, KLSM, TSNL, HTGY, and NMGY populations. The samples of this group are characterized by high roundness but small other traits, indicating that the seeds of this group are small, the seed weight is light, and the comprehensive traits are poor. Group II, which included 46 samples, is shown in green in the figure, and all regions were included. The seeds in this group are generally larger in size, heavier in weight, and exhibit better performance in terms of seed length, width, thickness, and thousand-grain weight. These characteristics suggest superior seed quality and potential for improved yield.

Sequencing Data Analysis

A total of 65 *S. alopecuroides* samples were used for sequencing analysis. The total sequencing data volume was 61.20257 Gb, with an average of 941.578 Mb per sample. The results revealed that the sequencing quality was high ($Q20 \geq 93.97\%$, $Q30 \geq 84.85\%$), with a normal GC distribution (35.61%–39.78%) (Table 3).

Table 3 Statistics of sequencing data

Parameter	Raw Base(bp)	Clean Base(bp)	Effective Rate (%)	Error Rate (%)	Q20(%)	Q30(%)	GC Content (%)
Maximum	1,591,091,424	1,556,618,112	99.07	0.04	97.73	93.25	39.78
Minimum	533,957,184	480,835,584	84.08	0.03	93.97	84.85	35.61
Mean	827,187,193	775,839,775	93.31	0.03	96.51	90.24	38.06
Total	53,767,167,552	50,429,585,376					

No samples were contaminated by adapter sequences, indicating successful library construction. The sequencing data of 65 *S. alopecuroides* were subsequently mapped to the reference genome. The average mapping rate of the population samples was 81.15% to 94.47%, and the average sequencing depth of the genome was 12.87-34.92x, with a 1 × coverage rate (at least one base coverage) for more than 0.10% of the samples (Table 4).

Table 4 Statistics of sequencing depth and coverage

Parameter	Clean Reads	Mapped Reads	Mapping Rate (%)	Average Depth	Coverage at Least 1X (%)	Coverage at Least 4X (%)
Maximum	10,809,848	9,812,988	94.47%	34.92	0.10%	12.38%
Minimum	3,339,136	2,709,853	81.15%	12.87	35.27%	57.47%
Mean	5,387,776	4,732,293	87.45%	19.56	25.82%	48.67%

Next, we detected a total of 58,448 SNP sites via SAMTOOLS, which were filtered to obtain high-quality SNPs. A total of 10,584 SNPs were obtained for subsequent analysis.

Population genetic diversity analysis

The *S. alopecuroides* samples were divided into 10 clusters according to geographical source: YLHG, TSBL, TCDQ, ARTS, KLSM, TSNL, THPD, QLSM, and NMGY, for a total of 65 samples. Moreover, the heterozygosity of each population was quite different (Table 5). The H_o values of the *S. alopecuroides* populations ranged from 0.20037 to 0.23511, with an average of 0.22213. The H_e ranged from 0.14387 to 0.18852, with an average of 0.16861, and the P_i ranged from 0.16442 to 0.19803, with an average of 0.18647. Among them, the H_o in the TCDQ was the highest, the H_e in the KLSM was the highest, and the H_o , H_e and P_i in the NMGY were the lowest. The genetic diversity was the highest in the TCDQ, and the NMGY had a relatively low level of genetic diversity.

Table 5 Genetic diversity level of ten subpopulations

Population	Sample Size	Observed Heterozygosity (Ho)	Expected Heterozygosity (He)	Nucleotide Diversity Index (Pi)
YLHG	5	0.204 22	0.152 39	0.169 32
TSBL	10	0.223 56	0.181 86	0.191 43
TCDQ	4	0.235 11	0.173 28	0.198 03
ARTS	3	0.230 11	0.159 74	0.191 69
KLSM	14	0.230 89	0.188 52	0.195 50
TSNL	10	0.214 18	0.174 53	0.183 72
THPD	5	0.216 48	0.162 17	0.180 19
QLSM	4	0.234 44	0.170 96	0.195 38
HTGY	6	0.231 97	0.178 76	0.195 01
NMGY	4	0.200 37	0.143 87	0.164 42

Population structure analysis and principal component analysis

In order to further understand the genetic background relationships of *S. alopecuroides* in different regions, Admixture software was used to analyze the population structure of 65 samples. The results revealed that when $K = 2$, the CV error was the smallest, indicating that the optimal grouping of the genetic structure of the 65 *S. alopecuroides* samples was two clusters (Fig. 5a and Fig. 5b). Moreover, to supplement the results of the population structure analysis, we used GCTA for PCA. According to the degree of SNP difference between individuals, PCA (Fig.5c) showed that 65 *S. alopecuroides* samples could not be effectively divided into two groups, and the distribution of some populations gradually overlapped, which was similar to the results of population structure analysis.

Phylogenetic tree analysis

The 10584 filtered SNPs were used to analyze the phylogenetic tree of 65 samples from 10 populations of *S. alopecuroides* via the neighbor-joining method (Fig. 6). The results revealed that the 65 samples could be divided into two large clusters. In general, the samples belonging to the same geographical area have a relative aggregation phenomenon in the two large clusters and are not completely merged into one place. Some samples were distributed in both clusters. Among them, cluster I mainly included *S. alopecuroides* samples from the TSBL, QLSM, HTGY, NMGY and THPD, whereas the samples from the YLHG, TCDQ and KLSM were clustered into cluster II. The distribution of *S. alopecuroides* samples in ARTS and TSNL was more disordered in the two clusters. These findings indicate that there is geographical isolation between the germplasm resources of *S. alopecuroides* from different geographical sources.

Analysis of molecular variance and Pairwise population differentiation

The F_{st} based on SNP data ranged from 0.00 to 0.04 among the 10 populations (Fig. 7). The highest F_{st} value was observed between TSNL and ARTS, with relatively high values also found between YLHG, TSNL, and other populations. AMOVA results indicated that 1.94% of genetic variation was attributed to differences among populations, while 34.77% and 132.83% originated from among individuals and within individuals, respectively, and all differences were statistically significant ($P < 0.05$). These results suggest that most genetic variation in *S. alopecuroides* is derived from within-population differences. Notably, the occurrence of negative variance components may reflect limitations of the AMOVA model under conditions of small sample size and weak genetic structure.

Table 6 Analysis of molecular variance analysis (AMOVA) of population

Source of variation	df	Sum of squares	Variance components	Percentage of variation (%)
Among populations	9	879.598	2.16256Va	1.94
Among individuals	55	3876.463	-38.71712Vb	-34.77
Within individuals	65	9614.500	147.91538Vc	132.83
Total	129	14370.562	111.36083	100

Redundancy analysis

RDA revealed the relationship between phenotypic traits, genetic diversity, and environmental factors in *S. alopecuroides*. As shown in Fig. 8a and 8b, the cumulative explanatory power of environmental factors for phenotypic traits and genetic diversity was 99.75% and 67.89%, respectively, suggesting that the first two RDA axes captured most of the variation. The length of the environmental vector indicates the strength of its influence, while the proximity between a sample point and an environmental vector reflects the degree of impact. In Fig. 8a, the key environmental factors influencing seed phenotypic traits included MTD (Mean Temperature of Driest Quarter), YWS (Annual Mean Wind Speed), Alt (Altitude), MAE (Mean Annual Evaporation), MTWE (Mean Temperature of Wettest Quarter), and MAS (Mean Annual Sunshine Time), with MTD, Alt, MAS, and YWS exerting the strongest effects. In Fig. 8b, genetic diversity was mainly affected by PWE (Precipitation of Wettest Quarter), ISO (Isothermality), and PC (Precipitation of Coldest Quarter). Specifically, PC was most associated with YLHG, TCDQ, ARTS, and TSBL populations; ISO influenced KLSM, QLSM, and TSNL; and PWE had the greatest impact on HTGY, THPD, and NMGY populations.

Discussion

Genetic diversity

Genetic diversity analysis is of great significance in the evaluation, utilization, origin and evolution of plant germplasm resources [47, 48]. Phenotypic diversity is a comprehensive reflection of genetic

diversity and environmental diversity [18]. Over long-term natural selection, different populations may undergo significant genetic variations, leading to relatively stable phenotypic traits [49]. In this study, the phenotypic traits of 65 *S. alopecuroides* seeds from various provenances were analyzed (Table 2), revealing a relatively rich phenotypic diversity. Among these, seed length, area, thousand-grain weight, and other traits exhibited greater variation, consistent with the findings of Yang et al. [20] and Wang et al. [36]. In addition, the *H* ranged from 1.639 to 1.767, and all traits were normally distributed (Fig. 2), which further verified that the phenotypic trait diversity was rich and statistically uniform. At the same time, seed phenotypic traits are mostly linked, so the correlation analysis between traits is very important in the study of seed phenotypic diversity [50, 51]. The correlation analysis of phenotypic traits in *S. alopecuroides* seeds revealed significant correlations between most traits ($P < 0.01$) (Fig. 3), suggesting that these traits exhibit a synergistic pattern in phenotypic expression, providing a basis for further exploration of the genetic mechanisms underlying phenotypic traits.

Meanwhile, the application of SNP molecular markers has become an important tool for the identification and conservation of plant germplasm resources [52, 53]. In this study, the genetic diversity analysis based on SNP molecular markers showed that the genetic diversity of *S. alopecuroides* was low, H_o , H_e and $P_i = 0.19$ were 0.22, 0.17 and 0.19 (Table 5), respectively, which was similar to the results of previous genetic diversity analysis of *S. alopecuroides* [54-56]. Furthermore, it showed relatively low genetic diversity compared with other *Sophora* species [57, 58]. Additionally, most species of the genus *Sophora* are monoecious, insect pollinated and self-compatible [36]; *S. alopecuroides* is no exception. These factors will lead to a low level of genetic diversity.

The AMOVA based on SNP markers showed that the genetic variation among individuals was -34.77%, while the genetic variation within individuals was as high as 132.83% (Table 6). This shows that there is a high genetic similarity within the population, resulting in a negative difference between individuals. At the same time, the F_{st} only between 0.00 and 0.04 (Fig. 7), which was also lower than that of other species of *Sophora* [59, 60]. In addition, small differences between populations and low genetic differentiation also indicate that these samples may be the same ancestral group and subsequently experienced geographical isolation. The occurrence of negative genetic variation may reflect the limitations of AMOVA when applied to poorly differentiated populations. Additionally, population size plays a critical role, as smaller populations reduce the evolutionary potential of wild species, thereby diminishing genetic diversity [61, 62]. Therefore, it is not excluded that the small number of samples collected in this study may lead to aberrant. Moreover, due to the small populations size of *S. alopecuroides* studied previously and the limited sample sizes, understanding of the genetic structure and variation of *S. alopecuroides* remains constrained [54, 56]. In contrast, we collected a larger number of germplasm resources from 65 regions across five provinces in northwest China. Although the limitations of previous studies have been overcome to some extent, the results still show that it is difficult to fully reveal the genetic diversity and population structure of *S. alopecuroides* populations.

Population structure

Population structure analysis in this study indicated that $K = 2$ is the optimal number of groups (Fig. 5). PCA and phylogenetic tree analysis further confirmed the division of the populations into two main clusters (Fig. 6). However, some populations were still divided into different clusters, which may be due to the different calculation focuses of the three analysis methods and the insufficient sample size, resulting in some populations with complex genetic relationships being divided into different clusters. Because geographical isolation also has a significant impact on the genetic structure of populations, it is a major pioneer in the formation of exotic species [63, 64]. The distribution of many *S. alopecuroides* populations across mountains and plateaus limits pollen and seed dispersal, creating barriers between populations [65]. Through the above analysis, we found that according to the characteristics of geographical, ecological and climatic conditions, *S. alopecuroides* can be divided into two clusters: north and south, with the Tianshan Mountains and the Qilian Mountains as the boundaries. The north-south geographical division is consistent with the population structure classification of the sample. However, the TSNL and ARTS populations showed more complex patterns, with higher genetic differentiation coefficients than other populations. At the same time, due to human activities, the suitable habitat area of *S. alopecuroides* has been greatly reduced, showing a fragmented pattern [34]. We speculate that the two populations were originally an ancestral population, and then experienced geographical isolation. Human factors also exacerbated the isolation effect, and TSBL may be a transitional group between the two populations. The results of phylogenetic tree analysis also verify the possibility of this situation.

Bioclimatic redundancy analysis

Environmental factors are the main limiting factors affecting the distribution and growth of plants on a large geographical scale. The differences in temperature, precipitation and other factors in different regions lead to significant differentiation in the growth adaptability of different populations to their habitats [66, 67]. To further assess the impact of environmental factors on genetic variation, we performed RDA. The phenotypic traits are the most direct indicators of the impact of environmental factors on plants. We found that the effects of environmental factors on the phenotype of *S. alopecuroides* seeds are manifested in external conditions such as average temperature in dry and wet seasons, average annual wind speed, evaporation, altitude and light. This is also an important factor affecting the distribution of *S. alopecuroides* [68, 69]. Rong et al. [34] also expressed that the future expansion trend of *S. alopecuroides* population was affected, and the main environmental factors were temperature and rainfall. The RDA based on the molecular level shows that the PC, ISO and PWE. PC had a strong influence on the genetic diversity of *S. alopecuroides* populations. This reflects that the seasonal variation of precipitation plays a key role in the gene flow and adaptability of *S. alopecuroides* population. ISO also has a significant effect on the genetic diversity of some populations, particularly in KLSM, QLMS and TSNL populations. Due to the relatively gentle climate change in areas with high ISO, it is conducive to the stable growth of plants for a long time, promoting gene flow and the accumulation of adaptive genetics [70]. This is also the main factor affecting the low genetic differentiation of populations in these areas.

Conclusion

This study utilized 10,584 SNP markers obtained via GBS technology, combined with seed phenotypic traits, to analyze the genetic diversity of 65 *S. alopecuroides* samples from 10 populations in northwest China. The results showed that the genetic diversity of *S. alopecuroides* populations was low and there was no obvious differentiation. The genetic variation mainly came from individuals. The population structure is mainly divided into two main clusters, which are related to environmental factors such as geographical distribution and climatic conditions. These findings underscore the need for further genetic protection and breeding efforts based on the geographical and genetic distance between populations. Therefore, on the basis of studying the genetic characteristics of *S. alopecuroides* populations, it is necessary to develop in situ conservation and regeneration strategies suitable for *S. alopecuroides* populations, such as giving priority to populations with relatively high genetic diversity and strengthening introduction measures to maximize the genetic diversity of *S. alopecuroides* populations. In future research, it is necessary to increase the size of the research population, which can be combined with population dynamics analysis, screening of adaptive genes, and interaction between environment and genetics. This study provides a foundation for the genetic evaluation and conservation of *S. alopecuroides* genetic resources in China and offers important insights for its breeding programs.

Abbreviations

GBS: genotyping-by-sequencing; SNP: single nucleotide polymorphism; PCA: principal component analysis; GPS: global satellite positioning system; DNA: deoxyribonucleic acid; PCR: polymerase chain reaction; Q-PCR: fluorescence polymerase chain reaction; He: expected heterozygosity; Ho: observed heterozygosity; AMOVA: molecular variance analysis; Fst: pairwise population differentiation; RDA: redundancy analysis; MTD: mean temperature of the driest quarter; YWS: annual mean wind; PC: precipitation of the coldest quarter; ISO: isothermality; PWE: precipitation of the wettest quarter; Pi: single nucleotide diversity index; LSD: least significant test; One-Way ANOVA: one-way analysis of variance; CV: coefficient of variation; H: Shannon's diversity index; SD: standard deviation; Alt: Altitude; MAE: Mean Annual Evaporation; MTWE :Mean Temperature of Wettest Quarter; MAS: Mean Annual Sunshine Time

Declarations

Acknowledgements

We acknowledge to Prof. Z.W Z. for specimens' identification, and appreciate the assistance of Mr. X.X.Z., Mr. Z.Y.T., Mr. W.W.R., Mr. J.W.S., and Mr. R.Y. in sample collections.

Authors' contributions

C.L., M.W., and X.H. conceptualized the topic. C.L. and F.N. analyzed the data. C.L., F.N., and P.N. set up the methodology. Z.Z., W.C., and G.M. provided resources. P.N. and Z.Z. conducted the investigation. M.W., G.C., and X.H. supervised the study. C.L. and F.M. wrote the main manuscript text. M.W., and X.H. reviewed and edited the manuscript. P.N. and P.J. contributed to visualization. M.W., P.J., and W.C. acquired funding. All authors have read and approved the final manuscript.

Funding

This research was funded by Corps Guiding Science and Technology Plan Project (Grant No. 2023ZD088), New Variety Cultivation Project of Shihezi University (Grant No. YZZX202303), College Students' Innovation and Entrepreneurship Training Program (Grant No. SRP2024232).

Data availability

The raw sequencing data have been deposited at the National Center for Biotechnology Information database under BioProject PRJNA1225430. The dataset associated with this study will be made publicly available under the accession numbers SRP564835, SRR32391254~SRR32391318 after the release date.

Ethics approval and consent to participate

This study including the collection on plants material complies with relevant institutional, national, and international guidelines and legislation.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹ Agricultural College, Shihezi University, Shihezi 832003, China

² Xinjiang Yuli National Positioning Observation and Research Station for Desert Ecosystems, Yuli 841500, China

³ Bozhou Water Conservancy Irrigation Test Station, Bole 833400, China

References

1. Wang RZ, Deng XX, Gao QX, Wu XL, Han L, Gao XJ, et al. *Sophora alopecuroides* L.: An ethnopharmacological, phytochemical, and pharmacological review. *J Ethnopharmacol.* 2020;248.

- doi:10.1016/j.jep.2019.112172.
2. Rong ZJ, Hu GS, Lin SY, Yan T, Li N, Zhao Y, et al. Constituents from the seeds of *Sophora alopecuroides* L. *Molecules*. 2020;25(2). doi:10.3390/molecules25020411.
 3. Li RS, Gao QQ, Mao LP, Liu WY, Sun LL, Zhang P, et al. Molecular mechanism of saline-alkali stress tolerance in the green manure crop *Sophora alopecuroides*. *Environ Exp Bot*. 2023;210. doi:10.1016/j.envexpbot.2023.105321.
 4. Huang YX, Wang G, Zhu JS, Zhang R, Zhang J. Traditional uses, phytochemistry, and pharmacological properties of *Sophora alopecuroides* L. *Eur J Inflam*. 2016;14(2):128-132. doi:10.1177/17217272x16642779.
 5. Huang X, Rong WW, Zhang XX, Gao YD, Zhou YS, Su JJ, et al. Transcriptome and metabolome analysis reveal the dynamic changes and biosynthesis pathways of alkaloids in *Sophora alopecuroides* L. under drought stress. *Ind Crops Prod*. 2024;212. doi:10.1016/j.indcrop.2024.118365.
 6. Gao HY, Li GY, Wang JH. A new alkaloid from the seeds of *Sophora alopecuroides* L. *Helv Chim Acta*. 2012;95(7):1108-1113. doi:10.1002/hlca.201100454.
 7. Aly SH, Elissawy AM, Eldahshan OA, Elshanawany MA, Efferth T, Singab ANB. The pharmacology of the genus *Sophora* (Fabaceae): An updated review. *Phytomedicine*. 2019;64. doi:10.1016/j.phymed.2019.153070.
 8. Qi XC, Zhao K, Ye ZP, Wang T, Chen B. Effects of *Sophora alopecuroides* green manure on growth and nutrient characteristics of melon. *J Agric Sci Technol*. 2018.
 9. Ma T, Yan H, Shi XL, Liu BT, Ma ZQ, Zhang X. Comprehensive evaluation of effective constituents in total alkaloids from *Sophora alopecuroides* L. and their joint action against aphids by laboratory toxicity and field efficacy. *Indl Crops Prod*. 2018;111:149-157. doi:10.1016/j.indcrop.2017.10.021.
 10. Yan MA, Sun Y, Cai HY, Dong Y, Zang SJ. Development and application of *Sophora alopecuroides* L. in animal husbandry. *Feed Res*. 2023;46(5). doi:10.13557/j.cnki.issn1002-2813.2023.05.029.
 11. Caroline T, Segatto ALA, Geraldo M, Rodrigues DM, Bonatto SL, Freitas LB. High levels of genetic diversity and population structure in an endemic and rare species: implications for conservation. *AoB Plants*. 2016;8:1-13. doi:10.1093/aobpla/plw002.
 12. Gao HN. comparative study on contents of oxymatrine and general flavone in *Sophora flavescens* from different habitats. *Chinese Journal of Experimental Traditional Medical Formulae* 2013;23(04):848-852.
 13. An JL Li Y, Dong C, Liu BB Shuang SM. Termination and analysis of alkaloid content in *Sophora alopecuroides* seeds from different producing areas in Xinjiang. *Journal of Shanxi University (Natural Science Edition)*. 2023;46(05):1188-1195. doi:10.13451/j.sxu.ns.2023009.
 14. Yang C, Liu P, Liu JR, Guo M. Genetic diversity analysis of ISSR markers in *Sophora alopecuroides*. *Chinese herbal medicine*. 2013;44(10):1323-1327.
 15. Zhang XX, Zhang YL, Niu LX, Sun JY, Li LH, Zhang J, et al. Chemometric classification of different tree peony species native to China based on the assessment of major fatty acids of seed oil and

- phenotypic characteristics of the seeds. *Chem Biodivers.* 2017;14(1). doi:10.1002/cbdv.201600111.
16. Andrea V, Věroslava H, Vigdis V, Zuzana M. Maternal effects strengthen interactions of temperature and precipitation, determining seed germination of dominant alpine grass species. *Am J Bot.* 2021;108(5):798-810. doi:10.1002/ajb2.1657.
 17. Pfennig DW, Wund MA, Snell-Rood EC, Cruickshank T, Schlichting CD, Moczek AP. Phenotypic plasticity's impacts on diversification and speciation. *Trends Ecol Evol.* 2010;25(8):459-467. doi:10.1016/j.tree.2010.05.006.
 18. Nicotra AB, Atkin OK, Bonser SP, Davidson AM, Finnegan EJ, Mathesius U, et al. Plant phenotypic plasticity in a changing climate. *Trends Plant Sci.* 2010;15(12):684-692. doi:10.1016/j.tplants.2010.09.008.
 19. Hu P, Zhong CL, Zhang Y, Jiang QB, Chen Y, Chen Z. Variation analysis on morphological traits of *Casuarina equisetifolia* at seedling stage from different regions. *Acta Bot Boreali-Occident Sin.* 2015;35(05):1013-1020.
 20. Yang C, Yang F, Ma H, Liu P. Distribution and morphological variation of germplasm resource of *Sophora alopecuroides*. *China Journal of Chinese Materia Medica.* 2010;35(7):817-820. doi:10.4268/cjcmm20100702.
 21. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *Plos One.* 2011;6(5). doi:10.1371/journal.pone.0019379.
 22. Geleta M, Gustafsson C, Glaubitz JC, Ortiz R. Hgh-density genetic linkage mapping of *lepidium* based on genotyping-by-sequencing SNPs and segregating contig tag haplotypes. *Front Plant Sci.* 2020;11:448. doi:10.3389/fpls.2020.00448.
 23. Lu L, Su RC, Yong PL, Kang SY, Yi SY. A GBS-based genetic linkage map and quantitative trait loci (QTL) associated with resistance to *Xanthomonas campestris* pv. *campestris* race 1 identified in *Brassica oleracea*. *Front Plant Sci.* 2023;14:1205681-1205681. doi:10.3389/fpls.2023.1205681.
 24. Lee KJ, Lee JR, Sebastin R, Shin MJ, Kim SH, Cho GT, et al. Genetic diversity assessed by genotyping by sequencing (GBS) in watermelon germplasm. *Genes.* 2019;10(10). doi:10.3390/genes10100822.
 25. Al SF, Habibi N, Al AS, Al DB. Genetic diversity of *Rhanterium eppaposum* Oliv. populations in Kuwait as revealed by GBS. *Plants.* 2022;11(11):1435. doi:10.3390/plants11111435.
 26. Uddin N, Shinde H, Ali N, Zhang Q, Ullah R, Iqbal Z, et al. Systematic exploration of *Ziziphus* species collection and gene identification influencing physiological traits using genotype-by-sequencing technology. *Physiol Plant.* 2024;176(4):e14426. doi:10.1111/ppl.14426.
 27. Rayda BA, Fabienne M, Hajer BH, Ahmed R, Sezai E, Narendra K, et al. SNP discovery and structural insights into *OeFAD2* unravelling high oleic/linoleic ratio in olive oil. *Comput Struct Biotec.* 2022;20:1229-1243. doi:10.1016/j.csbj.2022.02.028.
 28. Liu Y, Li S, Wang Y, Liu P, Han W. De novo assembly of the seed transcriptome and search for potential EST-SSR markers for an endangered, economically important tree species: *Elaeagnus mollis* Diels. *J Forestry Res.* 2019;31(3):1-9.

29. Li Y, Zhang J, Chang SX, Jiang P, Zhou G, Shen Z, et al. Converting native shrub forests to Chinese chestnut plantations and subsequent intensive management affected soil C and N pools. *For Ecol Manage.* 2014;312:161-169. doi:10.1016/j.foreco.2013.10.008.
30. Basheer-Salimia R, Camilli B, Scacchi S, Noli E, Awad M. Assessment of genetic diversity in lentils (*Lens culinaris* Medik.) based on SNPs. *Genet Mol Res: GMR.* 2015;14(142):5870-5878. doi:10.4238/2015.June.1.4.
31. Niu S, Song Q, Koiwa H, Qiao D, Zhao D, Chen Z, et al. Genetic diversity, linkage disequilibrium, and population structure analysis of the tea plant (*Camellia sinensis*) from an origin center, Guizhou plateau, using genome-wide SNPs developed by genotyping-by-sequencing. *BMC Plant Biol.* 2019;19(1):1-12. doi:10.1186/s12870-019-1917-5.
32. Muli JK, Neondo JO, Kamau PK, Michuki GN, Odari E, Budambula NLM. Genetic diversity and population structure of wild and cultivated *Crotalaria* species based on genotyping-by-sequencing. *Plos One.* 2022;17(9):e0272955. doi:10.1371/journal.pone.0272955.
33. Wu X, Blair MW. Diversity in grain amaranths and relatives distinguished by genotyping by sequencing (GBS). *Front Plant Sci.* 2017;8:1960. doi:10.3389/fpls.2017.01960.
34. Rong W, Huang X, Hu S, Zhang X, Jiang P, Niu P, et al. Impacts of climate change on the habitat suitability and natural product accumulation of the medicinal plant *Sophora alopecuroides* L. based on the MaxEnt model. *Plants.* 2024;13(11). doi:10.3390/plants13111424.
35. Huang X, Luo C, Zhang X, Liu L, Zhou X, Niu P, et al. Screening of ecotypes and construction of evaluation system for drought resistance during seed germination in Kudouzi (*Sophora alopecuroides*). *Agronomy.* 2024;14(10):2295. doi:10.3390/agronomy14102295.
36. Wang Y, Zhou TT, Li DH, Zhang XH, Yu WW, Cai JF, et al. The genetic diversity and population structure of (*Faboideae*) as determined by microsatellite markers developed from transcriptome. *Plos One.* 2019;14(12). doi:10.1371/journal.pone.0226100.
37. Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. *Nucl Acids Res.* 1980.
38. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-60. doi:10.1093/bioinformatics/btp324.
39. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-9. doi:10.1093/bioinformatics/btp352.
40. Eickmeyer K, Huggins P, Pachter L, Yoshida R. On the optimality of the neighbor-joining algorithm. *Algorithms Mol Biol.* 2008;3:5. doi:10.1186/1748-7188-3-5.
41. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19(9):1655-64. doi:10.1101/gr.094052.109.
42. Schneider S, Roessli D, Excoffier L, Roeslli D. Arlequin: a software for population genetics data analysis, Version 2.000. In: *Lilog*: 2000.
43. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A.* 1979;76(10):5269-5273. doi: 10.1073/pnas.76.10.5269.

44. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. Vegan: Community Ecology Package. R Package Version 2.6–10. Available at: [https:// CRAN.R- proje ct. org/ packa ge= vegan](https://CRAN.R-project.org/package=vegan) (Published: January 29, 2025).
45. Borcard D, Gillet F, Legendre P. Spatial analysis of ecological data. Numeri-cal ecology with R. Springer, Cham. 2018. p. 299–367.
46. Di GS, Yan LX, Ren JN, Wu LZ, Wang ZY, Ma CY. Genetic diversity and association analysis of *Anemarrhena asphodeloides* based on phenotypic traits and SSR markers. *Journal of Plant Genetic Resources* 2024;1-11. doi:10.13430/j.cnki.jpgr.20240527001.
47. Huang L, Xie WH, Lu LX, Gao HJ, Liu XY, Zhao LL. Genetic diversity analysis and evaluation of seed phenotypic traits of 59 alfalfa germplasms. *Shandong Agricultural Sciences*. 2024;56(10):34-42. doi:10.14083/j.issn.1001-4942.2024.10.005.
48. Lavanya GR, Srivastava J, Ranade SA. Molecular assessment of genetic diversity in mung bean germplasm. *J Genet*. 2008;87(1):65-74. doi:10.1007/s12041-008-0009-3.
49. Wu YH, Fan ZL, Li J, Guo JH, Guo YK, Wang YL. Phenotypic diversity of seeds and fruits of natural populations of *Acer ginnala*. *Plants of Guangxi*, 2018;38(06):795-803.
50. Lu C, Gao MB, Jiao XZ, He F, Li XJ, Wang H. Combining ability evaluation and heritability analysis of main agronomic traits of several wheat parents. *Acta Triticeae Crops*. 2010;30(06):1023-1028.
51. Xu Y, Liu S, Finnegan PM, Liu F, Ali I, Zhang H, Yang M. Geographical variation and genetic diversity of *Parashorea chinensis* germplasm resources. *Front Plant Sci*. 2024;15:1452521. doi:10.3389/fpls.2024.1452521.
52. Sun J, Yuan J, Wang B, Pan J, Zhang D. Development and characterization of 10 microsatellite loci in *Paeonia lactiflora* (Paeoniaceae). *Am J Bot*. 2011;98(9):e242-3. doi:10.3732/ajb.1100083.
53. Bendhifi M, Baraket G, Zourgui L, Souid S, Salhi-Hannachi A. Assessment of genetic diversity of Tunisian Barbary fig (*Opuntia ficus indica*) cultivars by RAPD markers and morphological traits. *Sci Hortic*. 2013;158:1-7. doi:10.1016/J.SCIENTA.2013.04.015.
54. Wang Y, Yu WW, Zhou K, Wang GB, Cao FL. Analysis and comprehensive evaluation of phenotypic diversity of *Sophora alopecuroides* L. *Plants in Guangxi*. 2019;39(09):1147-1158.
55. Liu P, Ma GZ, Ma L. Genetic diversity of allozymes in *Sophora alopecuroides*. *Chinese Journal of Pharmacy*. 2013;48(10):818-822.
56. Li WJ, Liu JR, Zhou YM, Hong YS, Liu P. SSR marker association analysis of phenotypic traits of main alkaloid content in *Sophora alopecuroides*. *Chinese Journal of Traditional Chinese Medicine*. 2020;45(13):3104-3111. doi:10.19540/j.cnki.cjcmm.20200329.106.
57. Shu WJ, Tang JM, Chen ZY, Jiang YS, Wang ZF, Wei X. Analysis of genetic diversity and population structure in *Sophora japonica* Linn. in China with newly developed SSR markers. *Plant Mol Biol Rep*. 2019;37(1-2):87-97. doi:10.1007/s11105-019-01141-w.
58. Heenan P, Mitchell C, Houlston G. Genetic variation and hybridisation among eight species of kōwhai (*Sophora*: Fabaceae) from new zealand revealed by microsatellite markers. *Genes*. 2018;9(2):111. doi:10.3390/genes9020111.

59. Fu Y, Li S, Ma B, Liu C, Qi Y, Pang C. Genetic Diversity analysis and core collection construction of ancient *Sophora japonica* L. Using SSR Markers. *Int J Mol Sci.* 2024;25(23):12776. doi:10.3390/ijms252312776.
60. Liu ZM, Zhao AM, Kang XY, Zhou SL, López-Pujol J. Genetic diversity, population structure, and conservation of *Sophora moorcroftiana* (Fabaceae), a shrub endemic to the Tibetan Plateau. *Plant Biol (Stuttg).* 2006;8(1):81-92. doi:10.1055/s-2005-872889.
61. Frankham R. Challenges and opportunities of genetic approaches to biological conservation. *Biol Con-serv.* 2010;143:1919–1927.
62. Mondini L, Noorani A, Pagnotta MA. Assessing plant genetic diversity by molecular tools. *Diversity.* 2009;1(1):19-35. doi:10.3390/d1010019.
63. Yang X, Li H, Yu H, Chai L, Xu Q, Deng X. Molecular phylogeography and population evolution analysis of *Citrus ichangensis* (Rutaceae). *Tree Genet Genom.* 2017;13(1):1-16. doi:10.1007/s11295-017-1113-4.
64. Hoskin CJ, Higgie M, McDonald KR, Moritz C. Reinforcement drives rapid allopatric speciation. *Nature.* 2005;437(7063):1353-6. doi:10.1038/nature04004.
65. Ye JW, Wu HY, Fu MJ, Zhang P, Tian B. insights into the significance of the chinense loess plateau for preserving biodiversity from the phylogeography of *Speranskia tuberculata* (Euphorbiaceae). *Front Plant Sci.* 2021;12:604251. doi: 10.3389/fpls.2021.604251.
66. Yuan HJ, Cheng XR, Yu MK, Wang YD, Tai JW, Zhang CX. Geographical variation of growth traits among different *Quercus acutissima* provenances. *Chin. J. Appl. Ecol.* 2021;32(8):2791-2799. doi: 10.13287/j.1001-9332.202108.007.
67. Ji R, Yu X, Ren T, Chang Y, Li Z, Xia X, et al. Genetic diversity and population structure of *Caryopteris mongholica* revealed by reduced representation sequencing. *BMC Plant Biol.* 2022;22(1):297. doi:10.1186/s12870-022-03681-y.
68. Wang, YN. Phylogeography of *Sophora alopecuroides* (Fabaceae) in Inner Mongolia Plateau and its adjacent areas. China: Master's Thesis, Qinghai Normal University; 2023.
69. Yao, Z.Y.; Han, Q.F.; Lin, B. Prediction of the distribution area of major noxious weeds in Xinjiang based on maximum entropy model. *J. Ecol.* 2023;43:5096–5109. doi:10.5846/stxb202205061252.
70. Bansal S, Harrington CA, Gould PJ, St Clair JB. Climate-related genetic variation in drought-resistance of Douglas-fir (*Pseudotsuga menziesii*). *Glob Chang Biol.* 2015;21(2):947-58. doi:10.1111/gcb.12719.

Figures

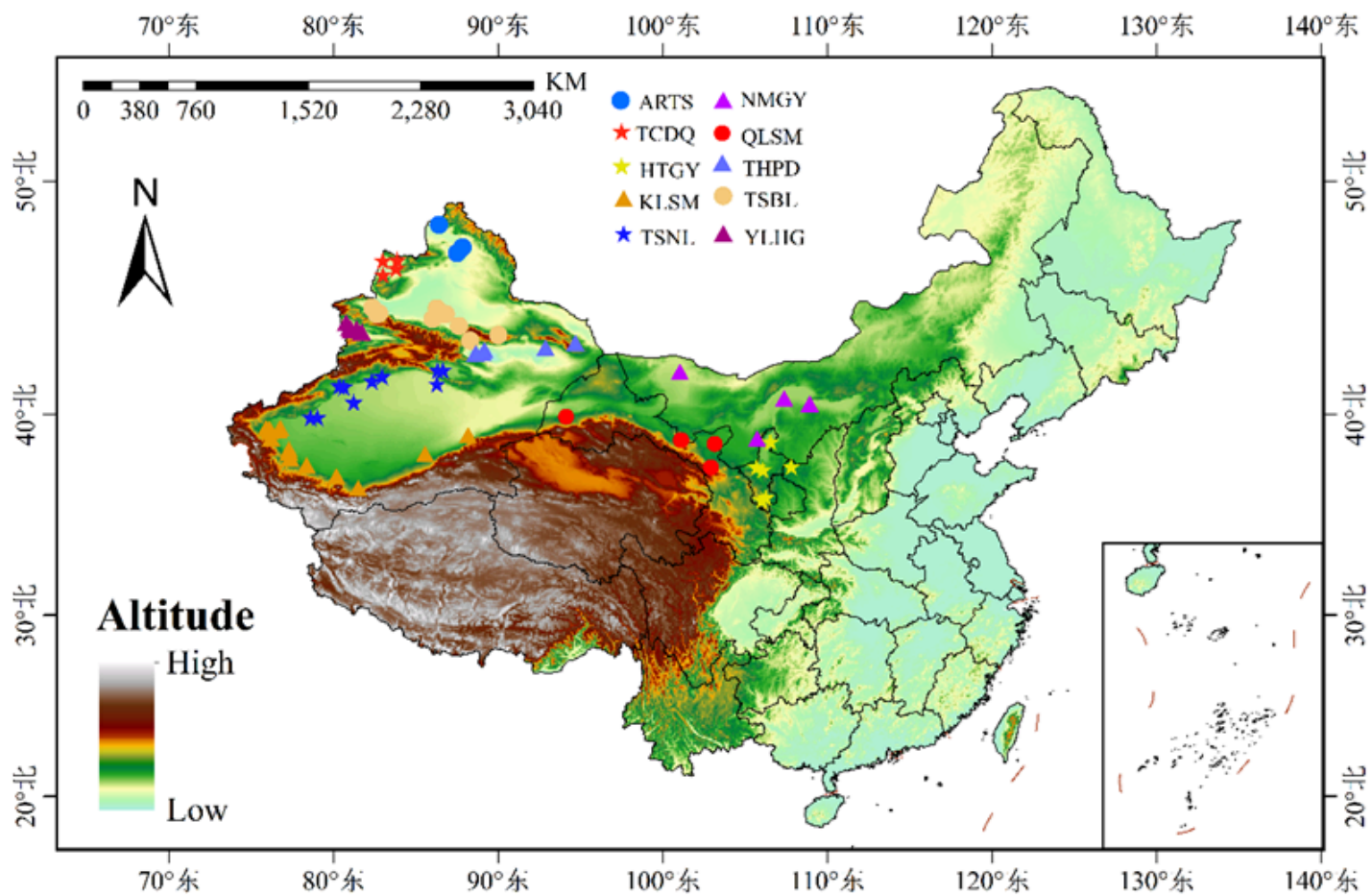


Figure 1

Sampling point distribution map of 10 *Sophora alopecuroides* populations

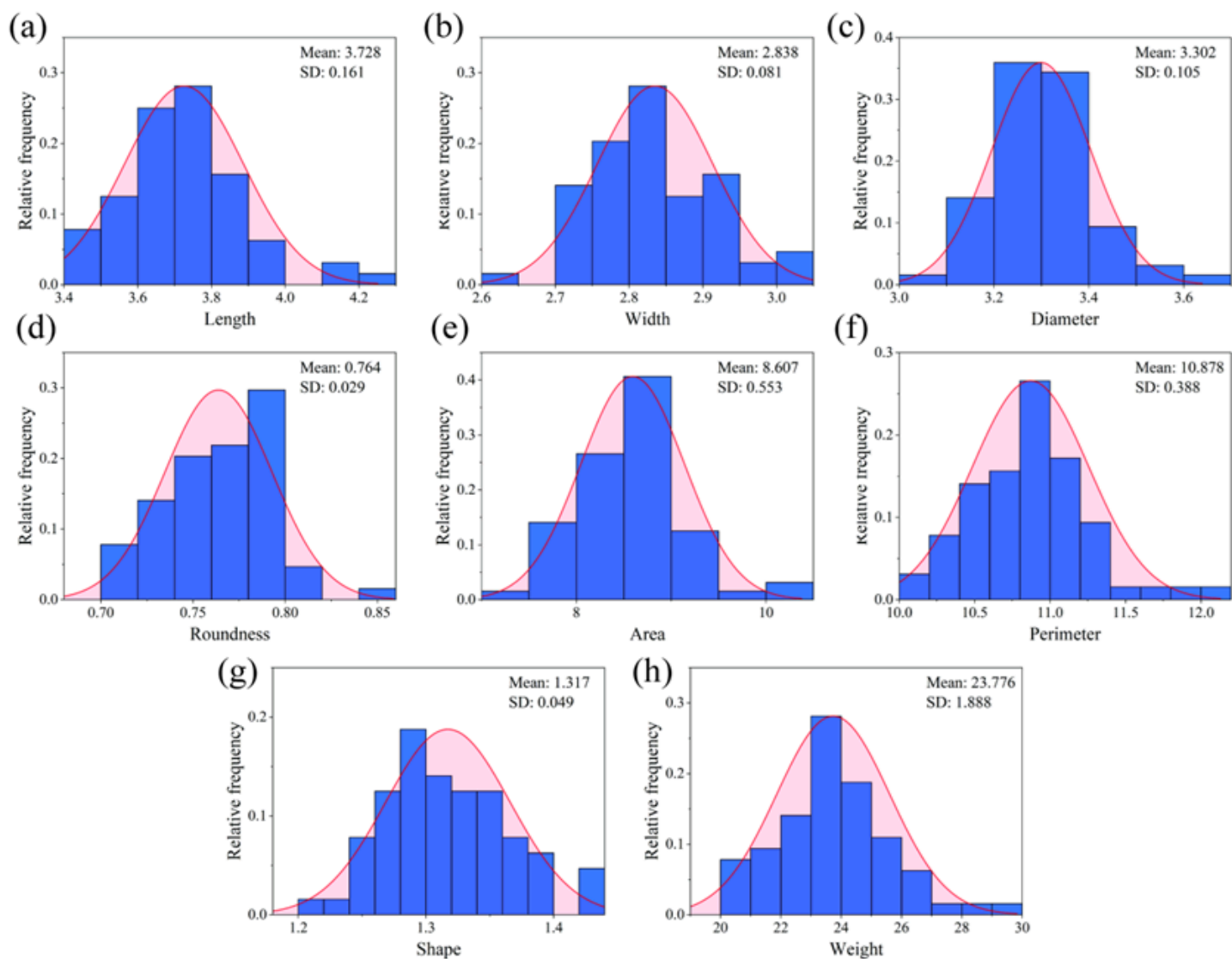


Figure 2

Histogram of frequency distribution of seeds phenotypic traits of *S. alopecuroides*. a: length; b: width; c: diameter; d: roundness; e: shape index; f: perimeter; g: Area; h: thousand-grain weight

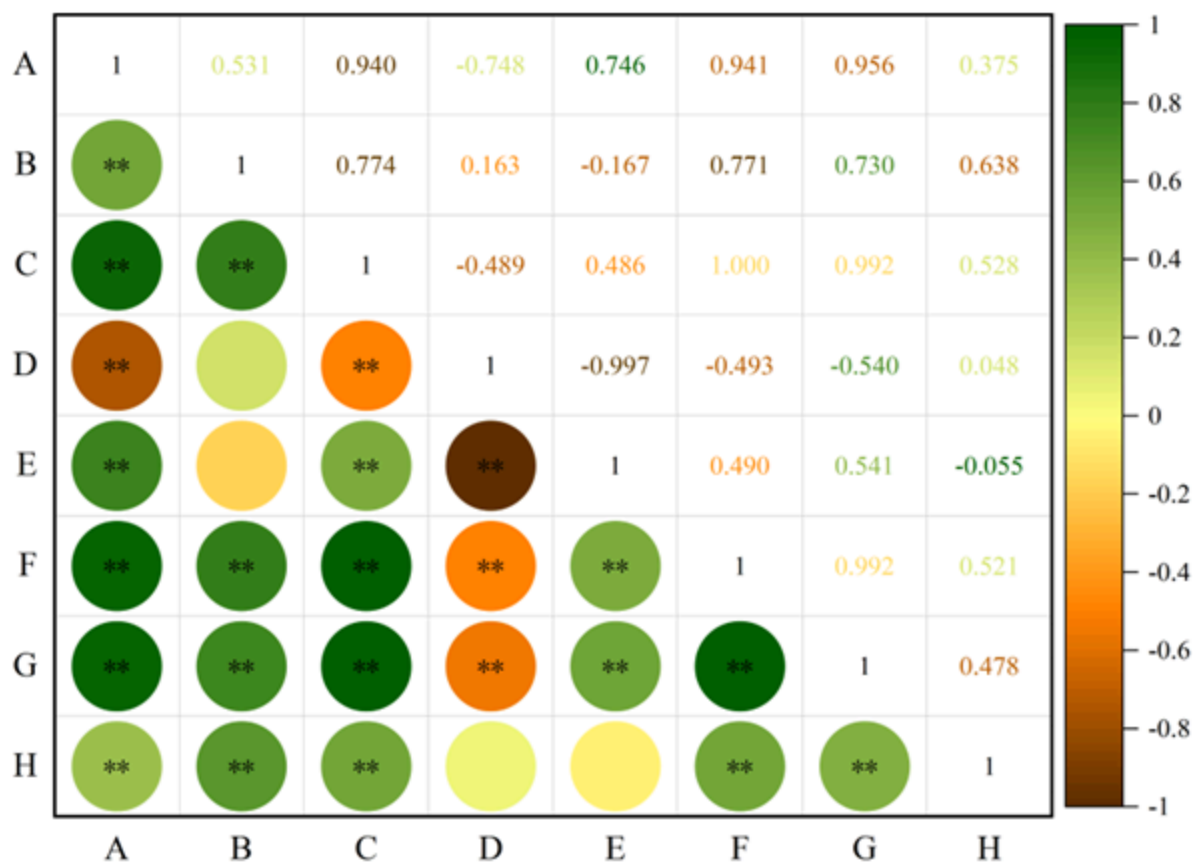


Figure 3

Correlation analysis of seeds phenotypic traits of *S. alopecuroides*. “**” indicates extremely significant correlation ($P < 0.01$); A: length, B: width, C: diameter, D: roundness, E: shape index, F: perimeter, G: area, H: thousand-grain weight.

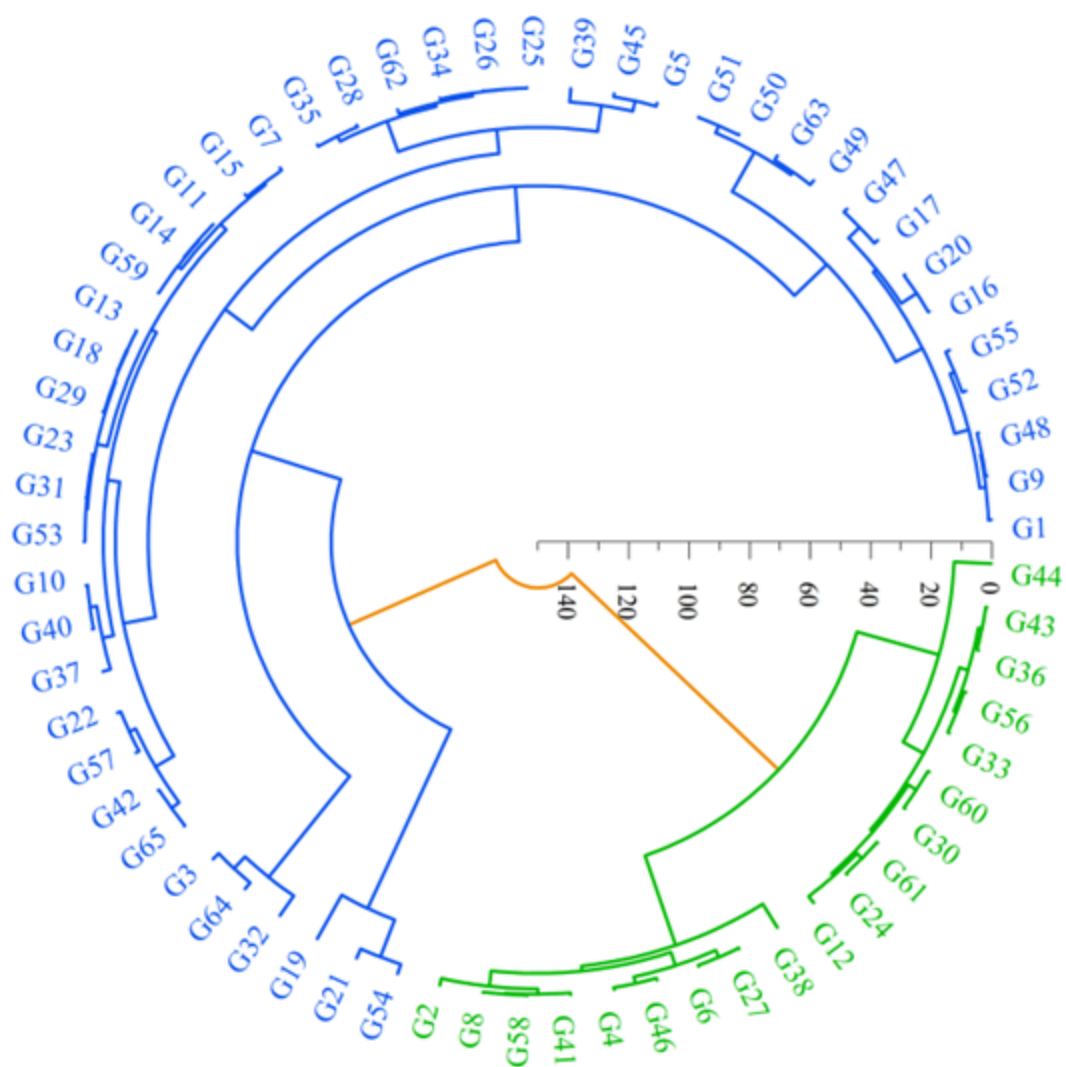


Figure 4

Cluster analysis of phenotypic traits of *S. alopecuroides* seeds

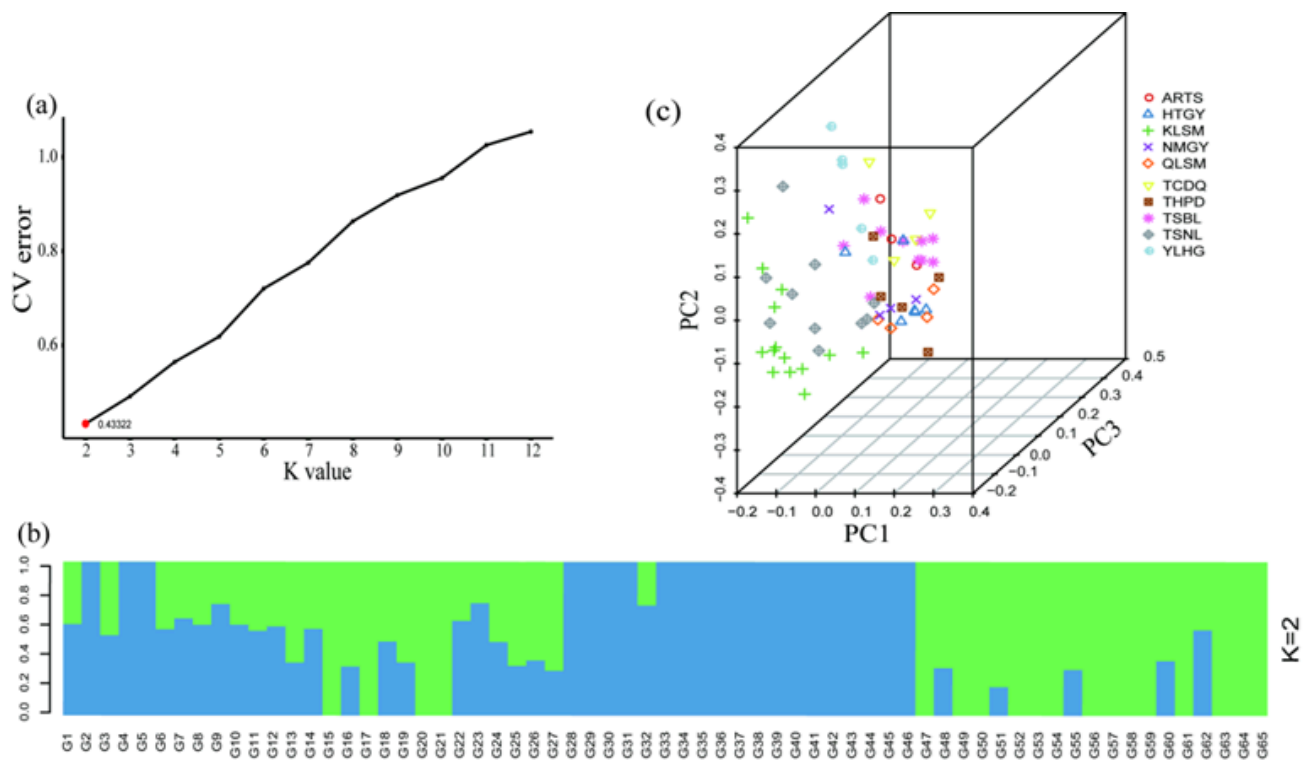


Figure 5

Population structure analysis and PCA of 65 samples of *S. alopecuroides*

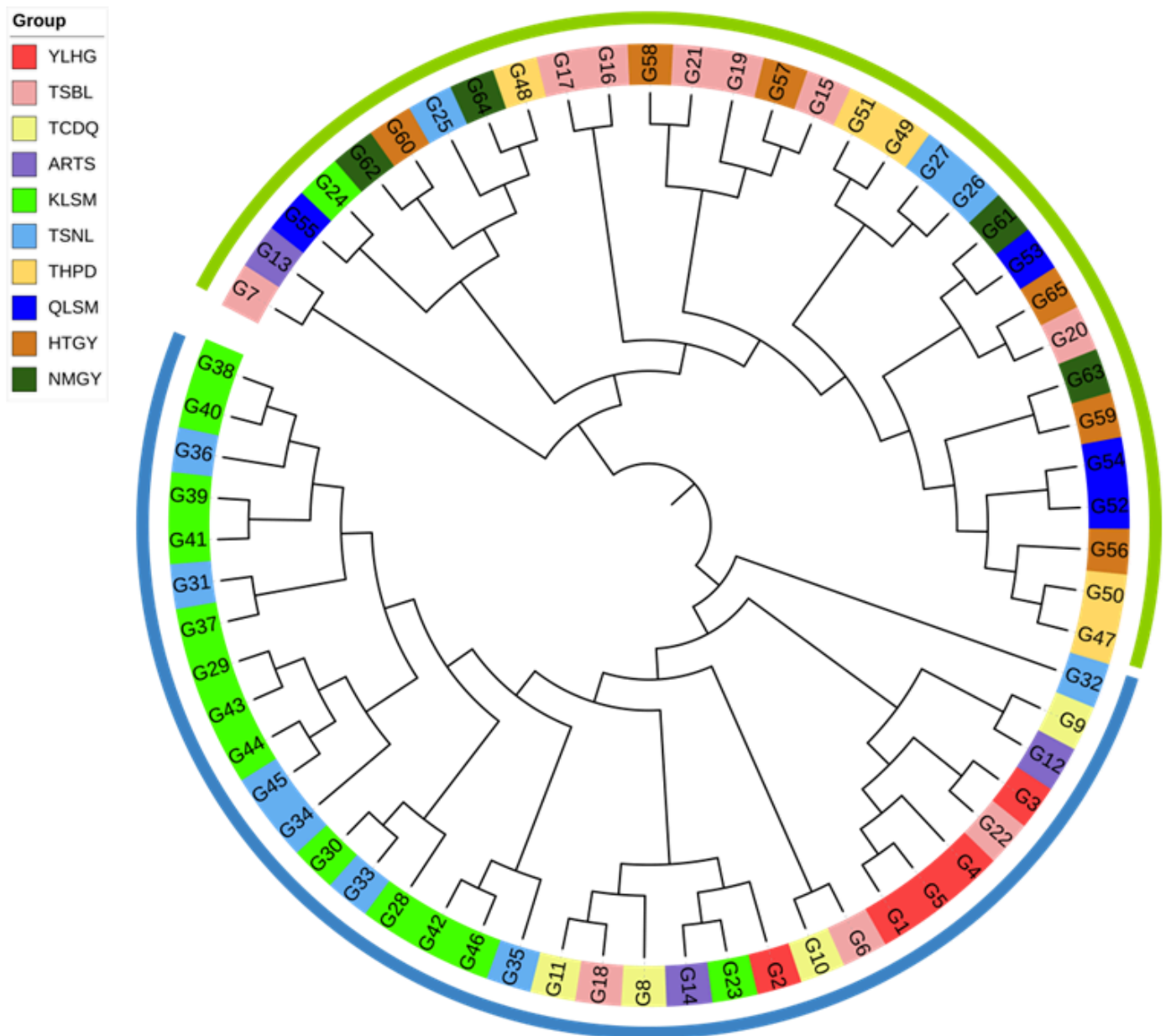


Figure 6

Phylogenetic tree of ten subpopulations of *S. alopecuroides*

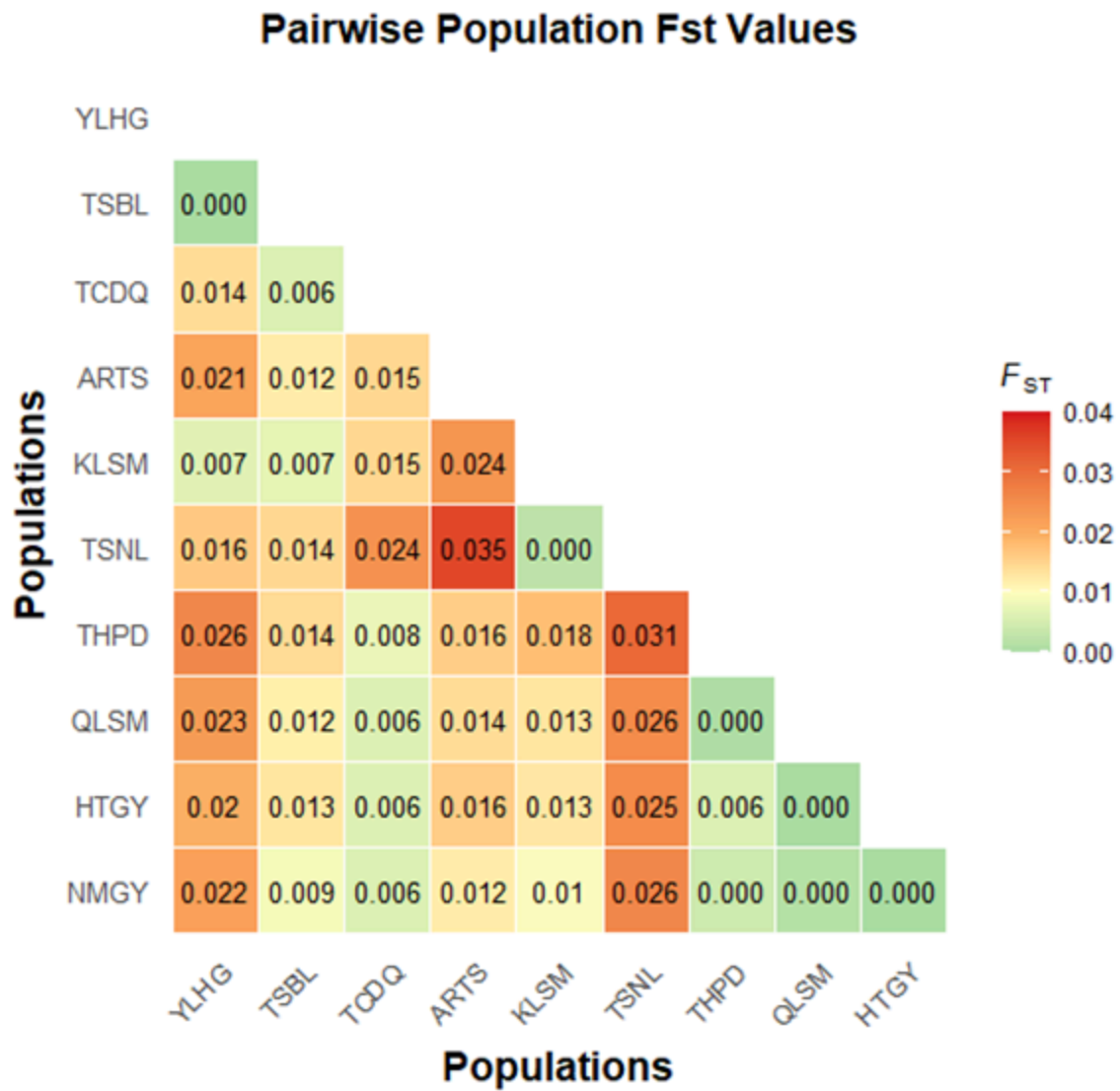


Figure 7

Fst among 10 *S. alopecuroides* populations

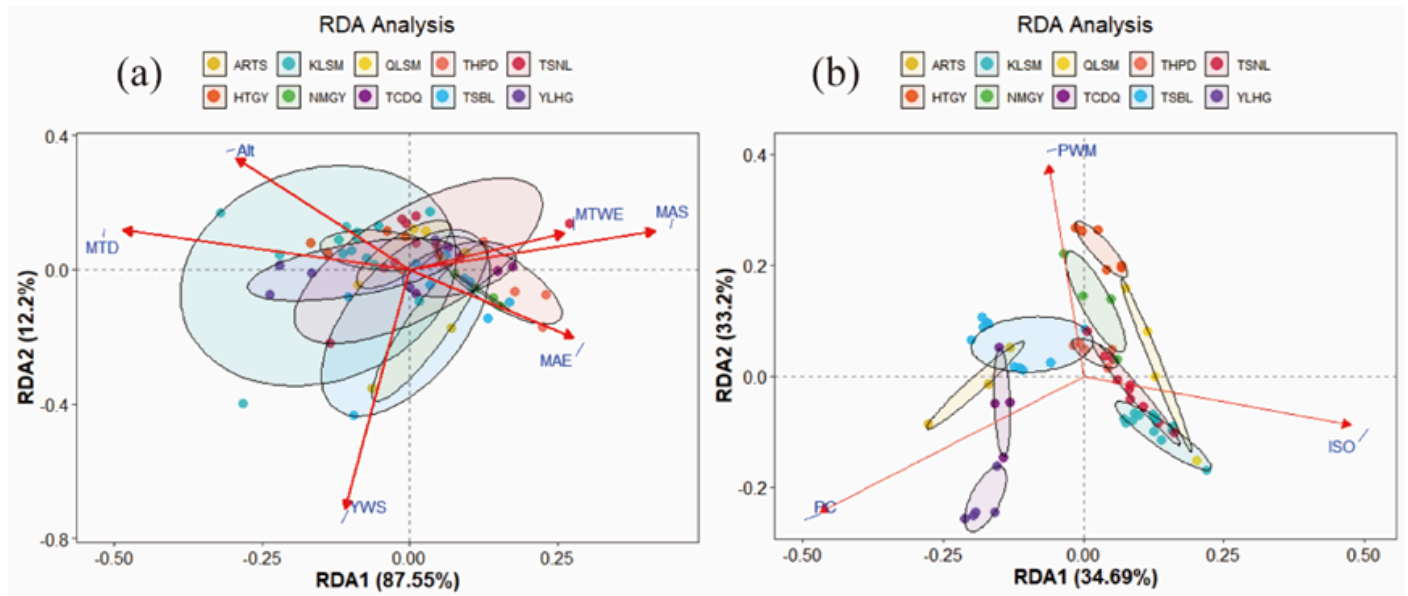


Figure 8

Redundancy analysis (RDA) of the relationship between seed phenotypic traits (a), genetic diversity (b) and environmental factors. The red arrows represent environmental factors, and different color circles represent sample points and populations

Supplementary Files

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

- [TableS1S3.docx](#)