

Study on the Correlation Between Phenotypic Traits and Genotypic Diversity of *Sapindus* Germplasm Resources

Jie Zhang¹, Yincheng Zhao² ✉

¹ Institute of Life Science, Jiyang College of Zhejiang A&F University, Zhuji, 311800, Zhejiang, China

² Tropical Medicinal Plant Research Center, Hainan Institute of Tropical Agricultural Resources, Sanya, 572025, Hainan, China

✉ Corresponding author: jie.zhang@jicat.org

Medicinal Plant Research, 2025, Vol.15, No.3 doi: [10.5376/mpr.2025.15.0011](https://doi.org/10.5376/mpr.2025.15.0011)

Received: 01 Mar., 2025

Accepted: 10 Apr., 2025

Published: 05 May, 2025

Copyright © 2025 Zhang and Zhao, This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:

Zhang J., and Zhao Y.C., 2025, Study on the correlation between phenotypic traits and genotypic diversity of *Sapindus* germplasm resources, Medicinal Plant Research, 15(3): 99-109 (doi: [10.5376/mpr.2025.15.0011](https://doi.org/10.5376/mpr.2025.15.0011))

Abstract *Sapindus* has both economic and ecological value, is renowned for its saponin and oil utilization, and is widely distributed. This study integrates agronomic-morphological assessment with molecular markers such as SSR/ISSR to analyze the phenotype and genetic diversity of *Sapindus* germplasm and their correlations. The results showed that the oil content, saponin content and fruit size of *Sapindus* varied significantly. The population structure was consistent with the ecological adaptation zone. In some materials, saponins were negatively correlated with oil content, and candidate genes, like *CYP716A*, *CAMTA*, *HD-ZIP*, *FATA*, *FAD2* and *DGAT3*, were identified. Research shows that, the integrated framework of phenotype, genotype and environment can support the construction of core germplasm and marker/genome selection, serving precision breeding and resource conservation. This study provides a reference for *Sapindus* marker-assisted/genomic selection, precision breeding, and resource conservation.

Keywords *Sapindus*; Phenotypic diversity; SSR/ISSR; Correlation analysis; Core germplasm; Mark-assisted selection

1 Introduction

Sapindus, belonging to the Sapindaceae family, is a multi-purpose tree species with economic, medicinal and ecological values. It is widely used in biodiesel production, biomedicine, and is valued as a source of natural surfactants for its high saponin content (Sun et al., 2018; Liu et al., 2021; 2022; 2025). The fruits and seeds of *S. saponaria* and its related species are particularly important, and its vegetable oils and saponins have extensive uses in industry and medicine (Wang et al., 2020; Liu et al., 2021). In addition to economic value, the *Sapindus* also plays an important role in ecological restoration and sustainable land management, and is an important resource for balancing industrial development, and environmental protection (Liu et al., 2022; 2025).

Sapindus saponaria and its closely related species are widely cultivated, because of their diverse utilization values. *Saponins* extracted from fruit peels are highly efficient natural surfactants, while seed oil is regarded as a promising raw material for biodiesel (Liu et al., 2021; 2022). These characteristics enable *Sapindus* to be widely used in detergent, cosmetic, pharmaceutical production and ecological restoration projects (Sun et al., 2017; Wang et al., 2020; Liu et al., 2025). Also, the *Sapindus* genus has a strong adaptability to various environmental conditions, further enhancing its value in sustainable agriculture and forestry (Sun et al., 2018).

Sapindus plants are distributed in tropical and subtropical regions. China and other regions in Asia and America have important genetic resources (Sun et al., 2019; Wang et al., 2020). But, the conservation and effective utilization of these germplasm resources are confronted with challenges, such as habitat loss, overexploitation, and limited breeding programs (Liu et al., 2022). Recent studies have emphasized the necessity of systematic assessment, and conservation strategies to maintain genetic diversity and ensure the long-term sustainability of *Sapindus* resources (Sun et al., 2017; 2018; Liu et al., 2022).

Phenotypic traits, like fruit size, saponin content and oil content, are key indicators for measuring economic value and adaptability in *Sapindus* breeding programs (Wang et al., 2020; Liu et al., 2021). The assessment of agronomic morphological diversity helps to screen out superior materials with excellent traits, which can be used as parents to breed new varieties with high yield, high quality and strong stress resistance (Sun et al., 2017; 2018).

Correlation analysis between phenotypic traits and environmental factors, can also reveal the adaptability and potential of different germplasms under variable environmental conditions (Wang et al., 2020).

Genotype diversity is the basis for the evolutionary potential and adaptability of *Sapindus* populations (Liu et al., 2025). Molecular marker analyses (such as SSR and ISSR) reveal rich genetic variations within and outside the *Sapindus* genus, which are important for establishing core germplasm banks and identifying genetic factors related to important traits (Sun et al., 2019; Liu et al., 2022). High levels of genotype diversity help discover new alleles, and support marker-assisted selection, accelerating the breeding of superior cultivated varieties.

This study attempts to reveal the correlation, between phenotypic traits and genotype diversity, in *Sapindus* germplasm resources. By integrating agronomic morphological assessment with molecular marker analysis, the study will describe the range of phenotypic and genotypic variations, screen out dominant germplasm materials suitable for breeding, and reveal the genetic structure and adaptive potential of the *Sapindus* population, providing a theoretical basis for the protection, and sustainable utilization of *Sapindus* germplasm resources. This study will provide references for breeding strategies, enhance the level of resource management, and promote the innovation and sustainable development of the *Sapindus* related industry.

2 Distribution and Characteristics of *Sapindus* Germplasm Resources

2.1 Geographical distribution and ecological adaptability

Plants of the *Saponus* genus, including *S. saponaria*, are widely distributed in tropical and subtropical regions around the world. The main populations are concentrated in East and Southeast Asia, North and South America, as well as some areas of Africa and Australia (Liu et al., 2021; 2022). In Asia, *S. mukorossi*, *S. delavayi* and *S. rarak* are mainly distributed in China, India and neighboring countries, while *S. saponaria* is more common in the Americas (Liu et al., 2021). Due to their economic and ecological value, these species have also been introduced to new regions.

The environmental adaptability of *Sapindus* germplasm resources is relatively strong, and its suitable growth environment usually meets the following conditions: The minimum temperature of the coldest month is $-20\text{ }^{\circ}\text{C}$ - $20\text{ }^{\circ}\text{C}$, the soil moisture content is 40-140 mm, the average temperature of the driest season is $20\text{ }^{\circ}\text{C}$ - $25\text{ }^{\circ}\text{C}$, and the soil pH value is between 5.6-7.6 (Liu et al., 2021; Li et al., 2021; 2022). For instance, *S. mukorossi* is suitable for warm and humid climates and low-altitude areas, while *S. delavayi* is more sensitive to the minimum temperature and higher altitude (Wang et al., 2020; Li et al., 2022; Liu et al., 2022). Climate change predictions indicate that its suitable habitat may shift to higher latitudes, while the distribution range in lower latitudes will shrink.

2.2 Resource collection and germplasm classification

Now, a large number of field investigations and germplasm collections have been carried out through distribution records in national and global databases, especially in China (Liu et al., 2021; Liu et al., 2022). Conservation strategies emphasize that, in-situ and ex-situ conservation should be given priority in areas, with a higher risk of habitat shrinkage, especially in regions with rapid environmental changes (Liu et al., 2022). Establishing a core germplasm bank that can represent important genetic diversity is of great significance for the conservation and breeding of *Sapindus* (Sun et al., 2019).

Morphological evaluation revealed that there were differences in fruit, seed and leaf traits in *Sapindus* germplasm (Sun et al., 2017; Wang et al., 2020; Liu et al., 2021). Through cluster analysis and principal component analysis, materials can be classified based on traits such as fruit size, saponin content and oil content, thereby screening out excellent materials suitable for breeding and resource management (Liu et al., 2021). Wang et al. (2020) demonstrated that intraspecific variation often exceeds interspecific differences, highlighting the significance of local adaptability.

2.3 Current research on germplasm resources

Recent advancements include genome-wide association studies, high-precision habitat modeling, and the development of molecular markers for trait selection (Liu et al., 2022; Xue et al., 2022). Domestic and

international studies have jointly enhanced the understanding of the genetic structure, ecological adaptability and climate change impact of *Sapindus* (Xue et al., 2022; Liu et al., 2025). The establishment of the core germplasm bank and the identification of reverse resistance genotypes are important achievements among them (Sun et al., 2019; Liu et al., 2022).

Despite the progress made, there are still challenges in the comprehensive collection of wild germplasm, long-term preservation in the context of climate change, and the integration of phenotypic and genotype data in breeding (Liu et al., 2025). Insufficient research on some species and regions, as well as weak conservation strategies, remain gaps in this field (Liu et al., 2021; 2022).

3 Investigation and Analysis of Phenotypic Traits

3.1 Determination methods for phenotypic indicators

The morphological traits of *Sapindus* and its related species are usually evaluated through standardized field measurement methods. During the mature stage, the height and crown width of the plants are measured using a tape measure or a laser rangefinder, while the growth rate is calculated by continuously tracking the annual increase in height or diameter over many years. These traits are important indicators for evaluating the adaptability, growth vigor and structural diversity of germplasm resources, and are often incorporated into comprehensive agronomic-morphological surveys (Sun et al., 2018; Liu et al., 2022).

Economic traits are particularly crucial for breeding and commercial utilization. The size of fruits is usually measured by dimensions, including weight, length, width, and thickness, using calipers and electronic scales. The color of the fruit can be evaluated by visual inspection with the naked eye or a colorimeter, providing a basis for market selection and processing utilization. Saponin content, as one of the core economic traits, is generally quantified through chemical extraction and chromatographic analysis, like HPLC, to determine the concentration of its active components (Figure 1). These methods can identify excellent materials with high oil content or high saponin yield, providing support for directed breeding (Sun et al., 2018; 2019; Zhou et al., 2024).

3.2 Statistical analysis of phenotypic data

Phenotypic data are typically quantified for diversity among and within different materials through range and coefficient of variation (CV) analysis. Previous studies have shown that there are variations in both morphological and economic traits, and the CV values of fruit size, saponin content and kernel oil content are relatively high. This diversity reflects the great potential of selection and improvement in breeding (Sun et al., 2019; Liu et al., 2022). For instance, the economic trait variations of *S. mukorossi* and *S. rarak*, were greater than those of *S. delavayi* (Liu et al., 2022).

Principal component analysis (PCA) and cluster analysis, are important methods for identifying the sources of major phenotypic variations and grouping germplasm materials with similar trait combinations. These multivariate statistical techniques can reveal the association patterns among traits, and promote the screening of superior germplasms. For example, PCA can identify materials with high oil content or high saponin content, while cluster analysis can group materials based on comprehensive traits, improving the efficiency of resource management and breeding (Sun et al., 2018; 2019). Correlation analysis further revealed the trade-off relationships among traits, such as the possibility, that an increase in saponin content might lead to a decrease in kernel oil content (Liu et al., 2022).

3.3 Relationship between phenotypic traits and environmental factors

Environmental factors (e.g., temperature, altitude and soil type, etc.) have a significant impact on the phenotypic traits of *Sapindus*. For example, under conditions of moisture, warmth and low altitude, the morphology of leaves and fruits is usually larger, longer and thicker (Wang et al., 2020). The weight and content of kernels are positively correlated with altitude and negatively correlated with longitude. However, an increase in sunshine duration may reduce the weight of seeds and the thickness of seed shells (Sun et al., 2018; Wang et al., 2020). These results emphasize that to achieve the best trait performance, it is necessary to match germplasm materials with suitable environments.

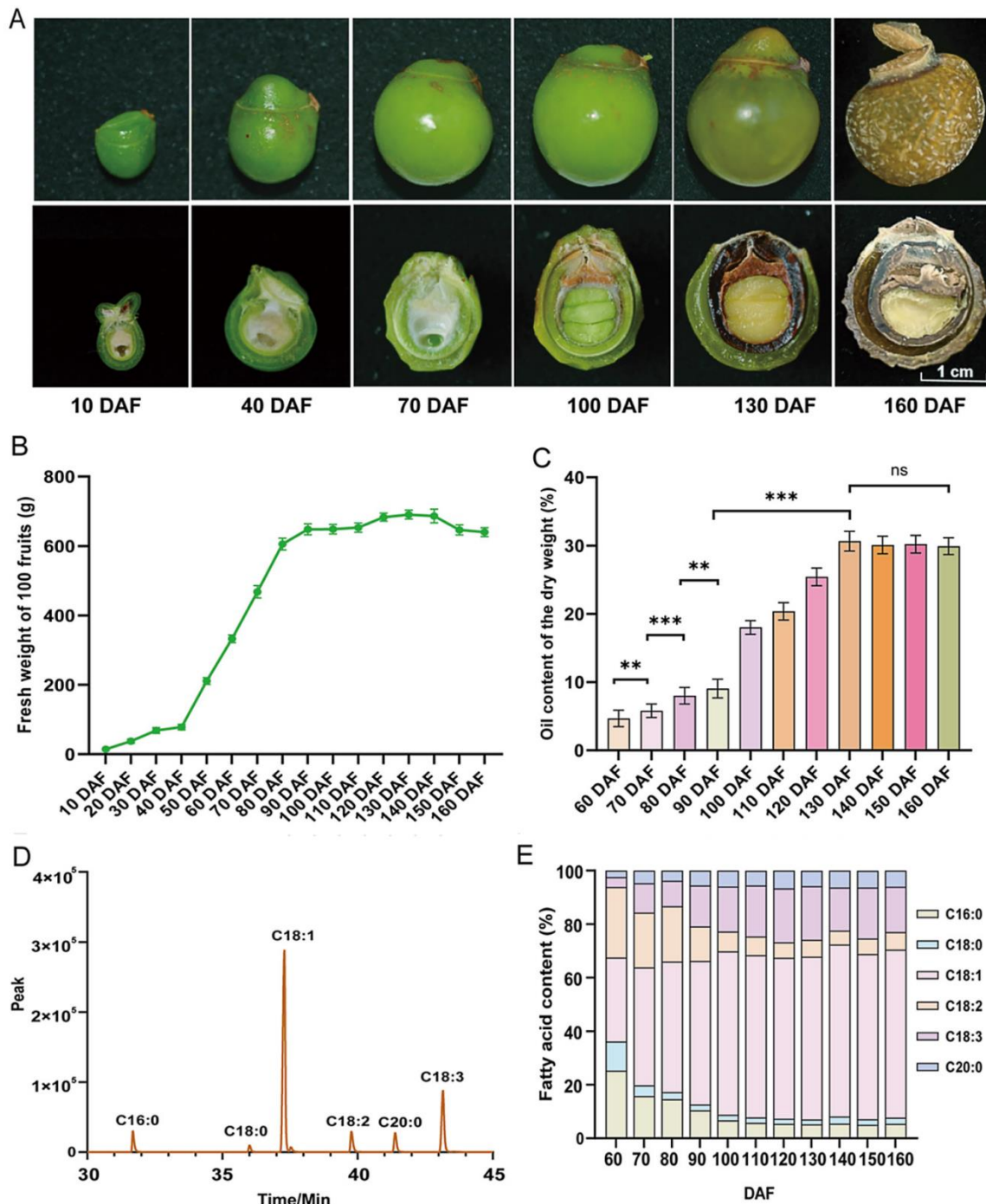


Figure 1 Phenotypic characteristics, oil content, and fatty acid dynamics of *S. saponaria* fruit at different developmental stages: (A) phenotypic characteristics of *S. saponaria* fruits; (B) changes in fresh weight of 100 fruits of *S. saponaria*; (C) changes in oil content in *S. saponaria* kernels; (D) the gas chromatography peak plot of *S. saponaria* kernel oil in 130 DAF; and (E) the fatty acid content of *S. saponaria* kernels. Note: ** means oil content with significant difference, *** means oil content with very significant difference, ns means oil content with insignificant differences. C18:1: oleic acid, C18:3: linolenic acid, C18:2: linoleic acid, C20:0: arachidic acid, C16:0: palmitic acid, C18:0: stearic acid (Adopted from Zhou et al., 2024)

Adaptive phenotypes (like larger kernels and higher oil content), are more common in materials under specific environmental conditions - lower longitude and higher altitude often facilitate kernel development, and increase in oil content (Sun et al., 2018). Intraspecific variation usually exceeds interspecific variation, indicating that local adaptability plays an important role in shaping phenotypic diversity (Wang et al., 2020).

4 Genotypic Diversity Analysis Based on Molecular Markers

4.1 Selection of molecular marker types

In the genetic diversity analysis of the *Sapindus* genus, SSRs, ISSRs, and SNPs, are the most widely used types of molecular markers. Studies have shown that the EST-SSR marker is highly effective in revealing the high genetic diversity and population structure of *Sapindus sapindus* germplasm (Liu et al., 2022). ISSR markers have also been successfully used to detect genetic variations and can link specific loci with important traits of fruits (kernel oil content, saponin content, etc.) (Sun et al., 2019). Although the application of SNP markers in *Sapindus* research is not as common as the former two, they have gradually been widely used in plant genomics due to their high resolution and abundant quantity (Nam et al., 2021; Xue et al., 2022).

Both SSR and ISSR markers show advantages such as high polymorphism and large information volume, which are suitable for differentiating related germplasm materials and providing support for the construction of core germplasm banks (Liu et al., 2022). Among them, SSR markers are valued for their good repeatability and co-dominant genetic characteristics, while ISSR markers are widely adopted because they can generate a large number of loci under the condition of fewer DNA samples (Sun et al., 2019; Nam et al., 2021).

4.2 DNA extraction and genotyping protocols

Samples are usually collected from the leaf tissues of healthy and mature plants, and then stored at low temperatures to ensure the integrity of DNA. DNA extraction is generally carried out using standard plant genomic DNA extraction kits or CTAB-based methods, and high-quality DNA suitable for PCR labeling analysis can be obtained (Sun et al., 2019; Nam et al., 2021).

PCR amplification is carried out using specific primers targeting the SSR or ISSR regions. The amplification products were then separated by agarose gel or polyacrylamide gel electrophoresis, and the band patterns were revealed using staining agents. The scoring method of ISSR markers is to record the presence or absence of bands, while SSR markers are judged by allele size, thereby generating a data matrix of binary or polyallelic genotypes to provide a basis for subsequent analysis (Liu et al., 2022).

4.3 Genetic diversity and population structure analysis

Genetic distance and similarity coefficient can be calculated by using labeled data. Commonly used indicators include Nei genetic distance and Jaccard similarity coefficient, which are used to quantify the genetic relationship between different materials (Sun et al., 2019; Liu et al., 2022). Cluster analysis methods, such as UPGMA and principal coordinate analysis (PCoA), are often used to group materials based on genetic similarity.

The population structure was further analyzed through model inference methods. The results indicated that there was significant differentiation within different species and their populations (Liu et al., 2022). For instance, *S. mukorossi* and *S. delavayi* can be clearly classified into different genetic populations, and establishing a core germplasm bank can cover genetic diversity to the greatest extent, providing a solid foundation for conservation and breeding (Sun et al., 2019).

5 Correlation Analysis Between Phenotypic Traits and Genotypic Diversity

5.1 Methodologies for correlation analysis

Canonical correlation analysis (CCA) is a multivariate statistical method, which is used to explore the relationship between two groups of variables. In the research of *Sapindus* and related plant germplasm, CCA has been applied to link environmental and phenotypic data, to reveal how multiple traits change cooperatively with genetic background and environmental gradient (Sun et al., 2017). For instance, CCA can identify which environmental factors have the greatest impact on the expression of key fruit and seed traits, thereby providing an overall perspective on the genotype-phenotypic-environment interaction.

The Mantel test is widely used to evaluate the correlations, between genetic distance matrices (based on molecular markers) and phenotypic distance matrices (based on trait data), etc. This method is mainly used to help determine that, whether genetic diversity can be reflected in observable trait variations (Sun et al., 2018; 2019; Verma et al.,

2019). Multiple regression models, including the Mixed linear Model MLM, are also used to identify marker-trait associations while controlling for population structure and kinship (Sun et al., 2019). These models are particularly effective in genome-wide association studies (GWAS), and have been used to locate gene loci related to important economic traits of *Sapindus* and other crops (Bahjat et al., 2025; Liu et al., 2025).

5.2 Association between key traits and genetic loci

Association analysis using ISSR and SSR markers has identified specific loci associated with the fruit size and related economic traits of *Sapindus*. For example, 18 ISSR loci were identified to be associated with 13 fruit traits, including kernel oil content, fruit saponin content and fruit quality (Sun et al., 2019). These marker-trait associations provide genetic evidence for the selection of superior germplasms and marker-assisted breeding, which helps to improve target traits more efficiently.

Genetic STRUCTURE analysis (STRUCTURE, principal component analysis PCA, etc.), revealed significant differentiation within different species and populations of the genus *Sapindus* (Sun et al., 2019; Liu et al., 2022). This group structure can affect the performance of key economic traits, so it must be taken into account when conducting association studies to avoid confounding effects. For instance, GWAS of *Sapindus* has identified candidate genes related to saponin biosynthesis and fatty acid quality of seed kernel, providing support for the genetic improvement of these traits (Liu et al., 2025). Population stratification, can also explain why certain marker-trait associations are only more obvious in specific subpopulations.

5.3 Implications for breeding and trait improvement

The integration of phenotypic data and genotype data is conducive to the efficient screening of germplasm with superior character-genotype combinations. By establishing a core germplasm bank with both high genetic diversity and superior trait expression, the protection and utilization of superior germplasm can be promoted, and rich materials can be provided for the *Sapindus* breeding program (Sun et al., 2019; Liu et al., 2022). This method can accelerate the screening of breeding materials that perform excellently in oil content, saponin content and other economic traits.

The identification of genetic markers associated with superior phenotypes (e.g., high oil content or high saponin content), provides conditions for the implementation of marker-assisted selection (MAS) strategies. This not only accelerates the breeding process, but enhances the potential of genetic improvement of *Sapindus* by targeting specific sites for selection (Sun et al., 2019; Bahjat et al., 2025; Liu et al., 2025). The combination of GWAS and core germplasm bank construction provides a robust framework for analyzing the genetic basis of complex traits and, applying this knowledge to actual breeding.

6 Case Studies on *Sapindus* Germplasm Correlation Analysis

6.1 Genetic diversity and adaptive evolution of *Sapindus* plants in China

Sapindus is an important economic forest tree species in China, and has application value in biodiesel, biomedicine, biochemistry and forestry. To understand the genetic basis of its genomic evolution and important bioeconomic traits, Liu et al. (2025) conducted whole-genome resequencing on 100 core germplasm resources of *Sapindus* from different geographical sources in China, revealing the genetic structure and evolutionary history of this genus of plants.

The results show that *Sapindus* is divided into six different populations in China. They are *S. delavayi*, *S. rarak*, North group, Southeast group, Guizhou group and hybrids group, respectively. These populations vary in species and geographical distribution (Figure 2). Phylogenetic analysis indicates that, the southeast population might be the ancestral population of *Sapindus* in China. Subsequent natural selection, gene flow and bottleneck effects led to the emergence of other populations.

It also found that different *Sapindus* populations show differences in environmental adaptation. For instance, the population in Guizhou might have undergone the selection of drought and low-temperature stress, while the population in the north might have acquired cold resistance. GWAS revealed several homologues of fatty acid

biosynthesis genes related to the quality of kernel fatty acids. The study also detected genes of different groups under selection pressure, like *CAMTA*, *GH3*, *NLRC3* and *R1*, through methods like XP-CLR.

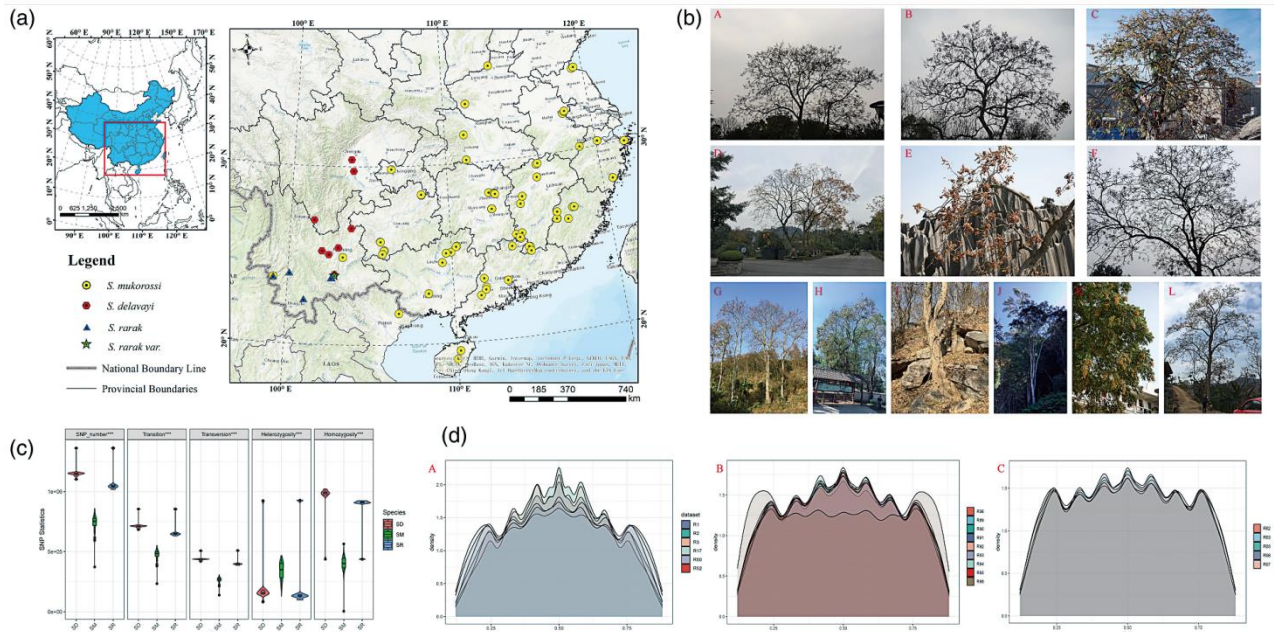


Figure 2 Basic information and descriptive statistics of the resequenced *Sapindus* accessions in China. (a) Geographical distribution of the *Sapindus* accessions. (Generated by ArcGIS 10.5). (b) Representative photos of resequenced *Sapindus* accessions. (A) *S. mukorossi* from Dejiang county, Guizhou province. (B) *S. mukorossi* from Dingnan county, Jiangxi province. (C) *S. rarak* from Lincang city, Yunnan province. (D) *S. delavayi* from Xinjin county, Sichuan province. (E) *S. delavayi* from Shilin county, Yunnan province. (F) *S. mukorossi* from Changsha city, Hunan province. (G) *S. mukorossi* from Zhenfeng county, Guizhou province. (H) *S. rarak* from Lincang city, Yunnan province. (I) *S. mukorossi* from Lianyungang city, Jiangsu province. (J) *S. rarak* from Menghan county, Yunnan province. (K) *S. mukorossi* from Jianning county, Fujian province. (L) *S. mukorossi* from Zhenfeng county, Guizhou province. (c) Genome-wide SNP descriptive statistics of resequenced *S. mukorossi* (SM), *S. delavayi* (SD) and *S. rarak* (SR) accessions. *, **, *** indicates statistically significant different (Kruskal test, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). (d) Density plot of alternative allele frequency of *S. mukorossi* (A), *S. delavayi* (B) and *S. rarak* (C) (Adopted from Liu et al., 2025)

6.2 Chromosome-level assembly and population diversity analysis of *Sapindus mukorossi*

Sapindus mukorossi is a *Sapindus* tree species of significant economic value. Its fruit is rich in saponins and oils and is widely used in fields, like biomedicine, biodiesel and biochemistry (Sun et al., 2018; Xue et al., 2022). To better understand the genetic basis and evolutionary history of *Sapindus mukorossi*, Xue et al. (2022) conducted an in-depth analysis of its genome, and reported the chromosome-level genome assembly of the first *Sapindus mukorossi*, with a genome size of approximately 391 Mb and a scaffold N50 of 24.66 Mb.

By resequencing 104 *Sapindus mukorossi* germplasm resources, researchers analyzed the genetic structure and evolution of *Sapindus mukorossi*. Population genetic analysis indicates that the genetic diversity in the southwest distribution area is relatively higher than that in the northeast distribution area. The gene flow event indicates that the species in the southwest region might be the donor population in the distribution area of China.

Genome-wide selective scanning analysis indicates that, a large number of genes are involved in defense responses, growth and development, including *SmRPS2*, *SmRPS4*, *SmRPS7*, *SmNAC2*, *SmNAC23*, *SmNAC102*, *SmWRKY6*, *SmWRKY26* and *SmWRKY33* (Figure 3). The study also identified candidate genes controlling six agronomic traits through genome-wide association studies, containing *SmPCBP2*, *SmbHLH1*, *SmCSLD1*, *SmPP2C*, *SmLRR-RKs* and *SmAHP*.

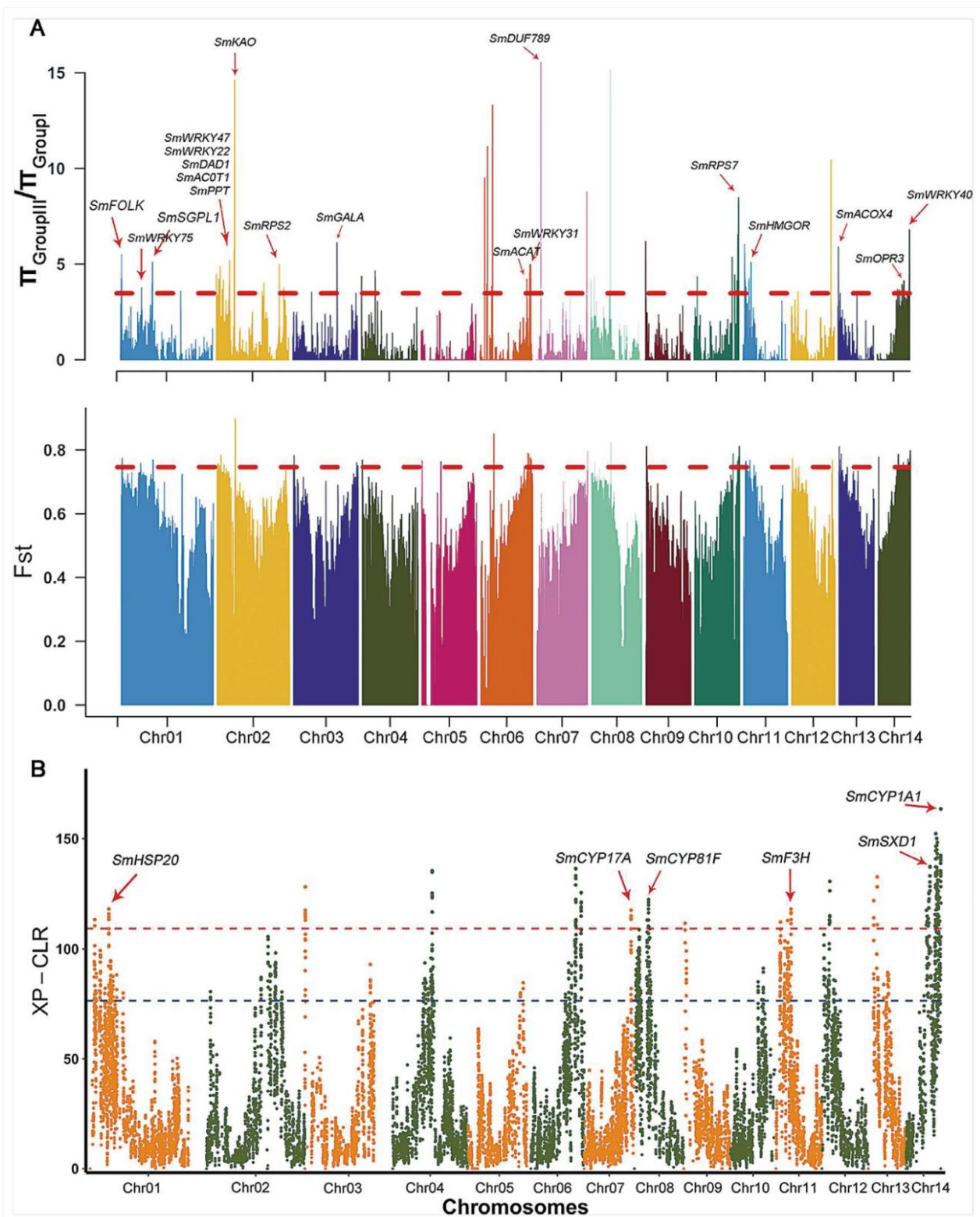


Figure 3 Genome-wide detection and annotation of selective sweeps in 104 *S. mukorossi* accessions. a Selective sweeps were identified by π GroupIII/ π GroupI and FST. The red dashed line corresponds to the thresholds for very high significance in the top 5% of the highest values. b Manhattan plots of the CLR values among the Group I and Group III *S. mukorossi* populations. Red and blue dashed lines indicate the threshold for the top 1% and top 5% of CLR values, respectively. Genes located within the significant CLR peaks and corresponding annotations are denoted (Adopted from Xue et al., 2022)

Image caption: The figure reveals the genomic regions under selective pressure in different populations, which may be related to adaptive evolution and phenotypic differences. The identification of these selective sweep regions helps in understanding the adaptation mechanisms of *S. mukorossi* in different environments (Adapted from Xue et al., 2022)

7 Germplasm Innovation and Breeding Strategy Development

7.1 Core germplasm screening and construction

The core germplasm screening of *Sapindus* is based on the dual evaluation of genetic diversity and agronomic-morphological trait expression. Selection criteria typically include: high genetic variability revealed by EST-SSR or ISSR markers, broad representativeness of population structure, and excellent performance in key economic traits such as saponin content, oil content, and fruit quality (Sun et al., 2019; Liu et al., 2021; 2022). The evaluation indicators cover genetic diversity indices (Nei index, observed heterozygosity etc.), principal component analysis (PCA) based on trait data, and the representativeness of the material in the full diversity spectrum. For instance, a study on 149 *Sapindus mukorossi* germplasms screened out dominant materials, using 19 agronomic-morphological traits and PCA, highlighting the importance of trait diversity and genetic background in the construction of core germplasms (Liu et al., 2021).

A well-designed core germplasm bank should not only cover genotype diversity but also take into account phenotypic trait diversity. In the *Sapindus mukorossi* study, the core set constructed from 161 germplasms contained 40% of the total genetic variation and included representative materials of all major species and trait types (Liu et al., 2022). Among them, materials with high saponin content or high oil content are given priority for inclusion to ensure their application value in breeding and conservation. The integration of molecular and morphological data ensures that the core germplasm has both broad genetic representativeness and outstanding agronomic value, thereby promoting efficient resource management and directed breeding (Sun et al., 2019; Liu et al., 2021; 2022).

7.2 Trait-oriented breeding strategies

The breeding program for high saponin content combines phenotypic screening with marker-assisted selection. Genomic studies have identified selective clearance regions and candidate genes (such as *CYP716A*, *CAMTA*, *HD-ZIP*) related to triterpene saponin biosynthesis, providing molecular targets for the improvement of saponin content (Liu et al., 2025). Dominant materials with natural high saponin levels are often used as donor parents in hybridization schemes, and ISSR/EST-SSR markers related to saponin traits make efficient selection in breeding populations possible (Sun et al., 2019; Liu et al., 2021; 2025).

The multi-trait aggregation strategy emphasizes the joint introduction of materials with complementary advantages, like high oil content, high saponin content and excellent fruit quality, into the breeding population. Association analysis and core germplasm strategy are helpful for aggregating ideal alleles, while PCA and cluster analysis can help screen out material, with the best combination of multiple traits (Sun et al., 2019; Liu et al., 2021). This method can help develop cultivated varieties suitable for diverse industrial applications, and take into account potential trade-offs as well, such as the negative correlation between saponins and oil content in some materials (Liu et al., 2021).

7.3 Sustainable utilization of genetic resources

The sustainable utilization of germplasm resources requires the combination of in-situ conservation and ex-situ conservation. The core germplasm bank can provide genetic reserves for ex-situ conservation, while in-situ conservation can maintain the evolutionary process and local adaptability (Liu et al., 2021; 2022). The combination of the two ensures that genetic resources can still achieve long-term preservation and availability in the context of climate change and habitat loss (Liu et al., 2021).

participatory breeding, which involves the participation of local communities and stakeholders, is conducive to improving the promotion and application effect of improved *Sapindus* varieties. The application of community-based germplasm sharing and focused identification strategies (like FIGS) can accelerate the discovery and utilization of new alleles on target traits, making breeding programs more inclusive and responsive to local needs (Stenberg and Ortiz, 2020; Sunitha et al., 2023).

8 Concluding Remarks

In the distribution area, *Sapindus* exhibited obvious phenotypic and genotypic diversity. There are significant differences among different populations in economic traits such as oil content, saponin content, fruit size and

environmental adaptability. Existing population structure analysis and core germplasm bank construction show that interspecific genetic variation is generally higher than intraspecific variation. Some of these species and populations, especially the *Sapindus mukorossi*, contain extremely rich genetic resources.

Correlation and association analysis have identified several key genetic loci and candidate genes. Such as *CYP716A*, *CAMTA*, *HD-ZIP* related to saponin biosynthesis, as well as *FATA*, *FAD2*, *DGAT3* related to oil-bearing traits, are all considered to determine important phenotypic characteristics. These achievements have laid a molecular foundation for trait selection. But, it also reveals certain trade-offs. For instance, in some materials, there is a negative correlation between saponin content and oil content. In other words, improving one trait may mean sacrificing another.

It should be noted that, environmental heterogeneity and insufficient sampling in some areas can easily interfere with the determination of genotype-phenotypic associations, making it more complex to distinguish genetic effects from environmental influences. In addition, the scattered distribution and fragmented population of *Sapindus* pose challenges to resource collection and protection. Although EST-SSR and ISSR markers have been used for diversity assessment and trait mapping, compared with high-throughput genomics techniques, these methods have limited resolution and are difficult to capture rare alleles, fine population structures, and weak genotype-phenotypic associations.

Future research should rely more on methods such as whole-genome resequencing, whole-genome association analysis (GWAS), and transcriptomics to locate gene loci related to traits with higher resolution and analyze the mechanisms of adaptive evolution. These measures can help discover new alleles and also accelerate the development of precision breeding tools.

By integrating high-throughput genotyping, phenomics and environmental data, it is expected to establish a precise breeding platform for *Sapindus*. This platform can support marker-assisted selection and genomic selection, promote the aggregation of multiple traits, and facilitate the rapid breeding of superior varieties, that adapt to diverse environments and different utilization requirements.

Acknowledgments

The authors sincerely thank Dr. Wang for reviewing the manuscript and providing valuable suggestions, which contributed to its improvement. Additionally, heartfelt gratitude is extended to the two anonymous peer reviewers for their comprehensive evaluation of the manuscript.

Conflict of Interest Disclosure

The authors affirm that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Bahjat N., Yıldız M., Nadeem M., Morales A., Wohlfeiler J., Baloch F., Tunçtürk M., Koçak M., Chung Y., Grzebelus D., Sadık G., Kuzğun C., and Cavagnaro P., 2025, Population structure, genetic diversity, and GWAS analyses with GBS-derived SNPs and silicodart markers unveil genetic potential for breeding and candidate genes for agronomic and root quality traits in an international sugar beet germplasm collection, *BMC Plant Biology*, 25: 65.
<https://doi.org/10.1186/s12870-025-06525-7>
- Li Y., Shao W., and Jiang J., 2021, Predicting the potential global distribution of *Sapindus mukorossi* under climate change based on MaxEnt modelling, *Environmental Science and Pollution Research*, 29(15): 21751-21768.
<https://doi.org/10.1007/s11356-021-17294-9>
- Li Y., Shao W., Huang S., Zhang Y., Fang H., and Jiang J., 2022, Prediction of suitable habitats for *Sapindus delavayi* based on the MaxEnt model, *Forests*, 13(10): 1611.
<https://doi.org/10.3390/f13101611>
- Liu J., Gao S., Xu Y., Wang M., Ngiam J., Wen N., Yi J., Weng X., Jia L., and Salojärvi J., 2022, Genetic diversity analysis of *Sapindus* in China and extraction of a core germplasm collection using EST-SSR markers, *Frontiers in Plant Science*, 13: 857993.
<https://doi.org/10.3389/fpls.2022.857993>
- Liu J., Wang L., Sun C., Xi B., Li D., Chen Z., He Q., Weng X., and Jia L., 2021, Global distribution of soapberries (*Sapindus L.*) habitats under current and future climate scenarios, *Scientific Reports*, 11(1): 19740.
<https://doi.org/10.1038/s41598-021-98389-8>

- Liu J., Zhao G., Zheng Y., Xu Y., Wang M., Li L., Sun C., He Q., Apuli R., Jong J., Ngiam J., Vaulin A., Tham R., Jia L., Chen Z., and Salojärvi J., 2025, Genetic diversity and adaptive evolutionary history of *Sapindus* in China: insights from whole-genome resequencing of 100 representative individuals, *Plant Biotechnology Journal*, 23: 2485-2500.
<https://doi.org/10.1111/pbi.70058>
- Nam V., Hăng P., Linh N., Ly L., Hue H., Ha N., Hãnh H., and Hien L., 2021, Molecular markers for analysis of plant genetic diversity, *Vietnam Journal of Science and Technology*, 18(4): 589-608.
<https://doi.org/10.15625/1811-4989/18/4/15326>
- Sobral M., and Sampedro L., 2022, Phenotypic, epigenetic, and fitness diversity within plant genotypes, *Trends in Plant Science*, 27(11): 1102-1115.
<https://doi.org/10.1016/j.tplants.2022.06.008>
- Stenberg J., and Ortiz R., 2020, Focused identification of germplasm strategy (FIGS): Polishing a rough diamond, *Current Opinion in Insect Science*, 40: 77-81.
<https://doi.org/10.1016/j.cois.2020.11.001>
- Sun C., Jia L., Xi B., Liu J., Wang L., and Weng X., 2019, Genetic diversity and association analyses of fruit traits with microsatellite ISSRs in *Sapindus*, *Journal of Forestry Research*, 30(1): 193-203.
<https://doi.org/10.1007/s11676-017-0580-7>
- Sun C., Wang J., Duan J., Zhao G., Weng X., and Jia L., 2017, Association of fruit and seed traits of *Sapindus mukorossi* germplasm with environmental factors in southern China, *Forests*, 8(12): 491.
<https://doi.org/10.3390/f8120491>
- Sun C., Wang L., Liu J., Zhao G., Gao S., Xi B., Duan J., Weng X., and Jia L., 2018, Genetic structure and biogeographic divergence among *Sapindus* species: An inter-simple sequence repeat-based study of germplasms in China, *Industrial Crops and Products*, 118: 1-10.
<https://doi.org/10.1016/j.indcrop.2018.03.029>
- Sunitha N., Prathibha M., Thiribhuvan R., Lokeshkumar B., Basavaraj P., Lohithaswa H., and Anilkumar C., 2023, Focused identification of germplasm strategy (FIGS): a strategic approach for trait-enhanced pre-breeding, *Genetic Resources and Crop Evolution*, 71: 1-16.
<https://doi.org/10.1007/s10722-023-01669-7>
- Verma H., Borah J., and Sarma R., 2019, Variability assessment for root and drought tolerance traits and genetic diversity analysis of rice germplasm using SSR markers, *Scientific Reports*, 9: 16716.
<https://doi.org/10.1038/s41598-019-52884-1>
- Wang X., Liu J., Rui X., Xu Y., Zhao G., Wang L., Weng X., Chen Z., and Jia L., 2020, Biogeographic divergence in leaf traits of *Sapindus mukorossi* and *Sapindus delavayi* and its relation to climate, *Journal of Forestry Research*, 32: 1445-1456.
<https://doi.org/10.1007/s11676-020-01206-7>
- Weckwerth W., Ghatak A., Bellaire A., Chaturvedi P., and Varshney R., 2020, PANOMICS meets germplasm, *Plant Biotechnology Journal*, 18: 1507-1525.
<https://doi.org/10.1111/pbi.13372>
- Xue T., Chen D., Zhang T., Chen Y., Fan H., Huang Y., Zhong Q., and Li B., 2022, Chromosome-scale assembly and population diversity analyses provide insights into the evolution of *Sapindus mukorossi*, *Horticulture Research*, 9: 12.
<https://doi.org/10.1093/hr/uhac012>
- Zhou X., Jiang L., Li P., Chen J., Chen Y., Yang Y., Zhang L.H., Ji Y.N., Xiao Z.H., Sheng K.Z., Sheng X.Q., Yao H., Liu Q., and Li C.Z., 2024, The biosynthesis pattern and transcriptome analysis of *Sapindus saponaria* oil, *Plants*, 13(13): 1781.
<https://doi.org/10.3390/plants13131781>

Disclaimer/Publisher's Note

The statements, opinions, and data contained in all publications are solely those of the individual authors and contributors and do not represent the views of the publishing house and/or its editors. The publisher and/or its editors disclaim all responsibility for any harm or damage to persons or property that may result from the application of ideas, methods, instructions, or products discussed in the content. Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.