Legume Genomics and Genetics 2026, Vol.17, No.1, 49-67
http://cropscipublisher.com/index.php/lgg

Review Article Open Access

Genomic Diversity and Population Structure Analysis of Global Soybean Germplasm Using SNP Markers

Xingzhu Feng
Hainan Institute of Biotechnology, Haikou, 570206, Hainan, China
Corresponding email: xingzhu.feng@hitar.org

Legume Genomics and Genetics, 2026 Vol.17, No.1 doi: 10.5376/lgg.2026.17.0004
Received: 13 Feb., 2026
Accepted: 17 Feb., 2026
Published: 27 Mar., 2026

Preferred citation for this article:
Feng X.Z., 2026, Genomic diversity and population structure analysis of global soybean germplasm using SNP markers, Legume Genomics and Genetics, 17(1): 49-67 (doi: 10.5376/lgg.2026.17.0004)

Abstract Soybean (*Glycine max* L.) is one of the most important oilseed and protein crops worldwide, and the effective utilization of global soybean germplasm resources plays a critical role in genetic improvement and sustainable agricultural development. With the rapid advancement of high-throughput sequencing technologies, single nucleotide polymorphism (SNP) markers have become powerful tools for studying genomic diversity and population structure in crop species. This study reviews the current progress in genomic diversity analysis of global soybean germplasm based on SNP markers. First, the distribution and conservation status of global soybean germplasm resources and the main methods used in genetic diversity research are summarized. Subsequently, the development and screening of SNP markers, evaluation metrics for genomic diversity, and commonly used bioinformatics analysis approaches are discussed. Furthermore, the population structure of soybean germplasm from different geographic regions is analyzed, and its relationship with genetic diversity and important agronomic traits is explored. A case study focusing on the population structure of global soybean core germplasm is also presented to illustrate the application of SNP-based analysis in germplasm evaluation and molecular breeding. Finally, the prospects for applying SNP markers in soybean genetic improvement, including marker-assisted selection, genomic selection, and multi-omics integration, are discussed. This review provides a theoretical reference for the efficient utilization of soybean germplasm resources and the development of improved soybean varieties.

Keywords Soybean (*Glycine max* L.); SNP markers; Genomic diversity; Population structure; Germplasm resources

1 Introduction

Genetic diversity within crop species underpins long-term gains in yield, resilience, and quality, and soybean is a prime example of a crop whose global importance is tightly linked to the breadth and structure of its germplasm resources. As a major source of plant protein and oil for food, feed, and industrial uses, soybean (*Glycine max* (L.) Merr.) contributes substantially to global food security, yet modern breeding has often relied on a relatively narrow subset of the available genetic pool (Duan et al., 2025; Viana et al., 2022). Historical domestication from wild soybean (*Glycine soja*) and subsequent breeding in geographically isolated programs have produced strong genetic bottlenecks and regionally distinct allelic compositions, especially in North American, South American, and East Asian cultivars (Viana et al., 2022; Song et al., 2015). At the same time, emerging production regions such as sub-Saharan Africa, Southern Africa, and Central Asia are expanding soybean cultivation, often with germplasm of limited diversity adapted to local environments (sindi et al., 2023; Zatybekov et al., 2025). Comprehensive characterization of global soybean germplasm—encompassing landraces, modern cultivars, and wild relatives—is therefore essential to identify unique alleles, diagnose redundancy, understand population structure, and design effective strategies for broadening the genetic base and improving adaptation.

Systematic germplasm research provides the framework for targeted introgression and informed parental selection in breeding programs. Studies in the USDA Soybean Germplasm Collection and other large panels have demonstrated that detailed molecular “fingerprinting” can reveal hidden population structure, delineate domestication and improvement sweeps, and identify private alleles maintained in specific gene pools or breeding programs (Kofsky et al., 2018; Song et al., 2015). Analyses of African and tropical soybean collections have similarly shown that, while some elite lines possess broad within-population diversity, many regional germplasm pools exhibit low molecular diversity and strong relatedness, with most variation residing within rather than

among populations (Obua et al., 2020). Such findings have direct implications for breeding strategies, highlighting the need for pre-breeding, expansion of geographic sources of introductions, and more deliberate use of wild and exotic accessions to counteract the erosion of genetic variation (Duan et al., 2025). In this context, global, SNP-based assessments of diversity and population structure are a critical step toward rational utilization, conservation, and deployment of soybean genetic resources.

Molecular marker technologies have transformed crop genetic research by enabling robust, high-throughput assessment of diversity, relatedness, and genome–trait associations independent of environmental noise. Traditional approaches based on morphology or biochemical markers are limited by genotype-by-environment interactions, developmental stage specificity, and the small number of traits that can be scored reliably (Rani et al., 2023). DNA markers overcome these constraints by directly assaying heritable variation at the nucleotide level, and have been widely used in soybeans and other crops for diversity analysis, QTL mapping, marker-assisted selection (MAS), genomic selection, and cultivar identification (Bunjkar et al., 2024). A broad suite of marker systems—including RFLP, RAPD, AFLP, SSR, EST-SSR, ISSR, and SNPs—has been deployed in legume genetics, each with specific advantages in terms of polymorphism, cost, throughput, and ease of scoring (Bunjkar et al., 2024). In soybean, SSR and EST-SSR markers have played a central role in early diversity and population structure studies, revealing high polymorphism and enabling differentiation of germplasm from diverse geographic origins (Zatybekov et al., 2023).

However, continuing advances in genotyping technologies and next-generation sequencing have shifted the focus toward sequence-based markers, especially single nucleotide polymorphisms (SNPs), which now dominate large-scale diversity and association studies. Modern SNP platforms such as genotyping-by-sequencing (GBS), diversity array technology (DArTseq), medium-density panels, and fixed arrays like SoySNP50K and its nested derivatives have greatly reduced the cost per data point and enabled genome-wide coverage in large germplasm collections (Song et al., 2024). These markers feed directly into multivariate and model-based analytical frameworks—principal component analysis (PCA), principal coordinate analysis (PCoA), hierarchical clustering, and Bayesian or likelihood-based STRUCTURE-type models—allowing fine-scale dissection of population structure, admixture, and genetic differentiation (Bunjkar et al., 2024; Zatybekov et al., 2025). Integration of SNP data with phenotypic and environmental information further supports genome-wide association studies and genomic prediction, accelerating the identification and deployment of favorable alleles in breeding pipelines (Chander et al., 2021).

Within this spectrum of marker systems, SNPs offer particular advantages for soybean genome research and for global germplasm diversity and structure analyses. SNPs are the most abundant form of genetic variation in eukaryotic genomes, broadly and evenly distributed across coding and non-coding regions, and exhibit low recurrent mutation rates that make them evolutionarily stable and well suited for tracing haplotypes and demographic history (Rani et al., 2023; Bunjkar et al., 2024). High-throughput SNP assays, including SoySNP50K, Axiom® SoyaSNP, and reduced panels such as SoySNP6K, SoySNP3K and SoySNP1K, combine high marker density with automation, low error rates, and scalability from hundreds to thousands of accessions (Kofsky et al., 2018). These platforms have enabled comprehensive genotyping of entire national and international germplasm collections, facilitating the detection of redundant accessions, construction of haplotype block maps, and precise estimation of linkage disequilibrium patterns across wild, landrace, and elite populations (Song et al., 2024). In soybean, SNP-based studies have successfully resolved population structure at global and regional scales, distinguishing wild from cultivated accessions, identifying transitional genotypes, and quantifying genetic similarity between local germplasm and foreign cultivars (Tsindi et al., 2023).

SNP markers are also particularly powerful for integrated analyses that link diversity patterns to breeding history and future improvement prospects. Population structure analyses using SNP panels have revealed low genetic differentiation among some regional collections, reflecting extensive germplasm exchange, but also identified unique clusters and private alleles in underexploited gene pools that can serve as reservoirs of novel variation for

stress tolerance and yield (Viana et al., 2022). Whole-genome resequencing and large-scale SNP discovery in global soybean panels have provided millions of variants that can be mined for signatures of selection, adaptive introgression, and domestication sweeps, as well as for fine mapping of quantitative trait loci underlying agronomic traits (Song et al., 2015). For emerging production regions such as Kazakhstan and sub-Saharan Africa, SNP-based characterization of local germplasm in the context of global collections offers actionable guidance on whether to prioritize introgression of exotic and wild alleles, or to intensify selection within existing adapted gene pools (Zatybekov et al., 2025). Against this backdrop, a comprehensive genomic diversity and population structure analysis of global soybean germplasm using SNP markers is both timely and necessary to support strategic conservation and accelerate the development of high-performing, resilient cultivars for diverse agroecological zones.

2 Current Status of Global Soybean Germplasm Resources and Genetic Diversity Research

2.1 Distribution and conservation of global soybean germplasm resources

Soybean germplasm is conserved in large ex situ collections as well as in situ in traditional farming systems and natural habitats of wild relatives. Major global repositories, such as the USDA Soybean Germplasm Collection, Asian national gene banks, and international centers, collectively maintain tens of thousands of accessions representing cultivated soybean (*Glycine max*), its wild progenitor (*Glycine soja*), and breeding lines from diverse agroecological zones (Nawaz et al., 2020). Regional collections, including those in Africa, South America, Central and Eastern Europe, and Central Asia, increasingly capture germplasm adapted to local environments and emerging production regions (Shaibu et al., 2021; Zatybekov et al., 2023). Wild soybean populations remain especially important reservoirs of adaptive variation for biotic and abiotic stress tolerance, and targeted collections in centers of diversity such as East Asia are recognized as priorities for long-term soybean improvement (Nawaz et al., 2020).

Conservation strategies emphasize both safeguarding genetic resources and generating characterization data that enable efficient use. Large collections often show substantial redundancy and uneven representation of geographic regions, maturity groups, and end-use types, underscoring the need for systematic molecular characterization to rationalize holdings and identify gaps (Rani et al., 2023). Recent SNP- and SSR-based surveys in Africa, India, Kazakhstan and other regions highlight contrasting patterns: some collections exhibit relatively broad diversity (e.g., TGx lines in sub-Saharan Africa), while others display narrow genetic bases linked to repeated use of a few elite parents Zatybekov et al., 2023). Genomic data are therefore being used not only to inform core and mini-core set development and to flag duplicate accessions, but also to guide targeted introgression of wild and exotic germplasm into locally adapted backgrounds to counteract genetic erosion and enhance resilience (Rani et al., 2023).

2.2 Main methods for studying soybean genetic diversity

Research on soybean genetic diversity has evolved from reliance on phenotypic descriptors to extensive use of DNA marker technologies. Early studies used morphological and agronomic traits (e.g., plant height, maturity, seed size, yield) to estimate diversity and relationships among accessions, but these traits are strongly influenced by the environment and often provide limited resolution (Rani et al., 2023). Biochemical markers and multi-environment field evaluations have helped to refine phenotypic clustering, yet environmental noise and the small number of measurable characters constrained their utility for detailed population structure analysis and germplasm management (Ullah et al., 2021). Consequently, morphological data are now typically combined with molecular information to capture both adaptive differentiation and underlying genomic variation (Perić et al., 2025).

DNA-based markers have become central tools for characterizing soybean diversity and population structure. A wide range of marker systems—including RAPD, AFLP, ISSR, SSR, EST-SSR, DArT and SNPs—has been applied to differentiate cultivars, landraces, and wild accessions, estimate allelic richness, and dissect within- and among-population variation (Wibisono et al., 2025). SSR markers, in particular, have been extensively used

because of their high polymorphic information content, codominant inheritance, and relatively uniform distribution across the genome, enabling reliable assessment of relatedness, clustering of accessions, and alignment with pedigree information. More recently, high-density SNP genotyping and genotyping-by-sequencing platforms have allowed genome-wide diversity analysis in large panels, supporting robust population structure inference (STRUCTURE, PCA, PCoA, DAPC), AMOVA, and identification of genetically distinct or redundant accessions for breeding and conservation (Chander et al., 2021).

2.3 Progress in the application of SNP technology in soybean diversity research

The application of SNP technology has markedly advanced the resolution and scale of soybean diversity and population structure studies. Genotyping-by-sequencing (GBS) and diversity array technology (DArTseq) have been used to generate tens of thousands of SNPs across the 20 soybean chromosomes, producing high-density datasets for panels ranging from fewer than 100 to several hundred accessions (Fu et al., 2021). These approaches have revealed that most genetic variation typically resides within rather than among soybean populations, even in regionally focused collections, and have enabled the detection of distinct genetic clusters associated with breeding histories, adaptation zones and, in some cases, seed quality traits such as seed longevity (Shaibu et al., 2021). GBS-based SNP datasets also feed into genome-wide association and QTL mapping efforts, linking diversity patterns to complex traits, while highlighting the potential of selected accessions as donors of favorable alleles.

Parallel development of fixed SNP arrays and nested marker panels has further expanded SNP applications in germplasm research. Arrays such as Axiom® SoyaSNP, SoySNP50K, SoySNP6K, and higher-density platforms like SoySNP618K provide reproducible, genome-wide SNP coverage suitable for evaluating entire national and international collections, detecting redundant accessions, and constructing high-resolution haplotype maps (Zatybekov et al., 2025). Reduced, cost-effective panels (SoySNP3K, SoySNP1K, and targeted GBTS panels of 10–40K SNPs) are now widely used for routine germplasm characterization, diversity analysis, and parent selection in breeding programs (Song et al., 2024). Whole-genome resequencing of thousands of accessions has also yielded comprehensive SNP datasets that distinguish wild and cultivated gene pools, identify large-effect mutations in agronomically important genes, and support development of diagnostic marker sets tailored for germplasm evaluation and reverse genetics (Zatybekov et al., 2025). Collectively, these advances have made SNP technology the backbone of contemporary soybean diversity research, enabling integrative analyses that connect global germplasm structure with breeding history and future improvement strategies (Figure 1).

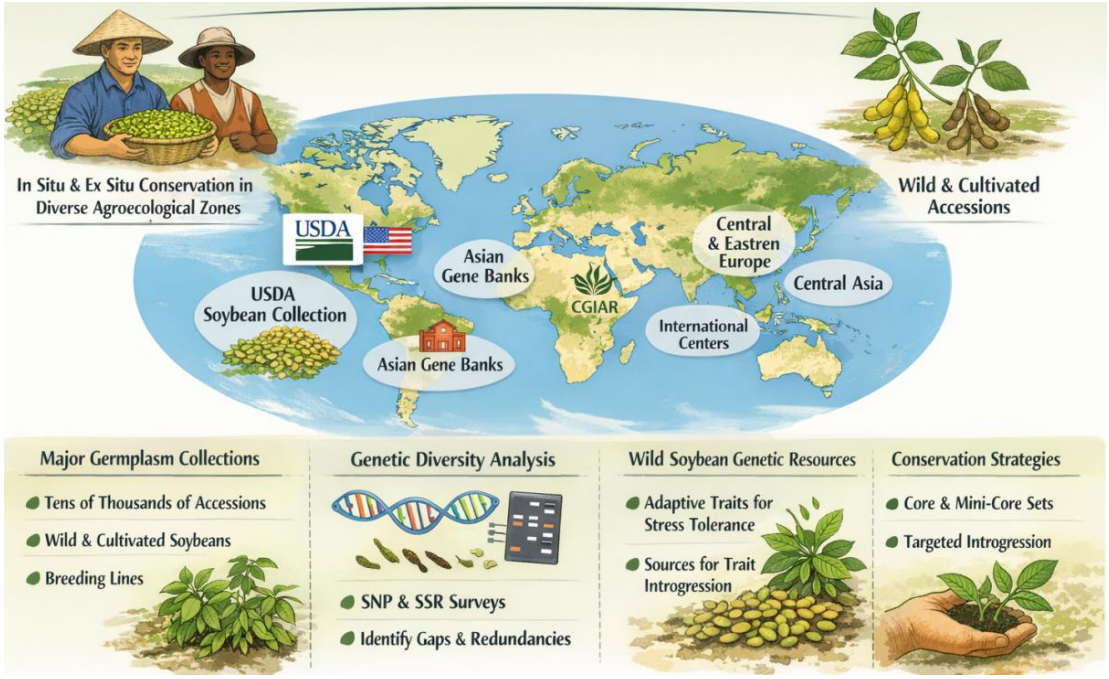

Figure 1 Global soybean germplasm resources and diversity conservation

3 Methods for Analyzing Soybean Genomic Diversity Based on SNP Markers

3.1 Development and screening of SNP markers

High-quality SNP marker sets were developed from whole-genome resequencing or genotyping-by-sequencing (GBS) of large soybean panels spanning wild and cultivated germplasm. Resequencing thousands of accessions enables discovery of millions of raw SNPs, which are then filtered for read depth, base quality, biallelic status, and low missing data to obtain a robust variant catalogue distributed across all 20 chromosomes (Niu et al., 2024; Valliyodan et al., 2021). Targeted genotyping arrays, such as the Axiom SoyaSNP array with ~180K SNPs and the widely used SoySNP50K platform, were designed from these catalogs by prioritizing markers in gene-rich regions, evenly spaced along chromosomes, and with intermediate minor allele frequencies (MAF) to maximize information content in diversity analyses (Lee et al., 2015; Valliyodan et al., 2021). More recently, nested SNP assay series (SoySNP50K/6K/3K/1K) and reduced GBTS panels (40K/20K/10K) were assembled as subsets of high-density arrays, allowing researchers to match marker density and cost to specific germplasm characterization or breeding applications while retaining compatibility with legacy data sets (Song et al., 2024).

For global germplasm diversity studies, additional criteria were applied to ensure that the SNP panel discriminates both between wild and cultivated soybean and within each group. Panels were refined by removing monomorphic loci, markers with high missing data, and those with extreme allele frequency skews, and by retaining sites showing differentiation between *Glycine soja* and *Glycine max* and among cultivated subgroups (Niu et al., 2024). Quality control steps typically included excluding accessions with excessive missing data, applying MAF thresholds (e.g. ≥0.05), and checking marker performance through concordance with resequencing genotypes or replicate assays (Chander et al., 2021). The resulting datasets often contain tens of thousands of high-quality SNPs with low error rates and broad genomic coverage, suitable for downstream estimation of genomic diversity, identification of large-effect variants, and construction of mutant gene libraries linked to agronomic traits (Niu et al., 2024).

3.2 Metrics for assessing genomic diversity

Genomic diversity based on SNP data was quantified using standard population-genetic metrics computed at both locus and genome levels. Per-marker statistics included polymorphic information content (PIC), gene diversity (expected heterozygosity), observed heterozygosity, major allele frequency, and MAF, which together describe the informativeness and allele frequency spectrum of the SNP set (Chander et al., 2021). In soybean panels genotyped with high-throughput SNP arrays, average gene diversity values around 0.41–0.42 and PIC values near 0.32–0.33 have been reported, with a substantial proportion of markers exhibiting PIC >0.35, indicating adequate polymorphism despite the bi-allelic nature of SNPs and the relatively narrow genetic base of cultivated soybean (Abebe et al., 2021). Shannon's diversity index and unbiased diversity estimates were also used at the population level to compare diversity among geographic or breeding groups (Shaibu et al., 2021).

To assess differentiation and structure across global germplasm, fixation indices (F_ST) and analysis of molecular variance (AMOVA) partitioned genetic variation within and among predefined groups. These analyses consistently showed that the majority of variation (often >90%) resides within populations, with a smaller but significant fraction among regions or breeding pools (Rani et al., 2023). Pairwise genetic distances, Nei's diversity, and clustering-based measures summarized relationships among accessions and populations, while indices such as the number of private alleles provided additional insight into unique diversity that may be under-utilized in breeding (Figure 2) (Abebe et al., 2021). In large resequenced panels, linkage disequilibrium (LD) decay and the distribution of conserved versus highly polymorphic genomic regions were further examined to infer domestication bottlenecks, selection sweeps, and the effective resolution for association mapping and haplotype-based analyses of soybean diversity (Valliyodan et al., 2021).

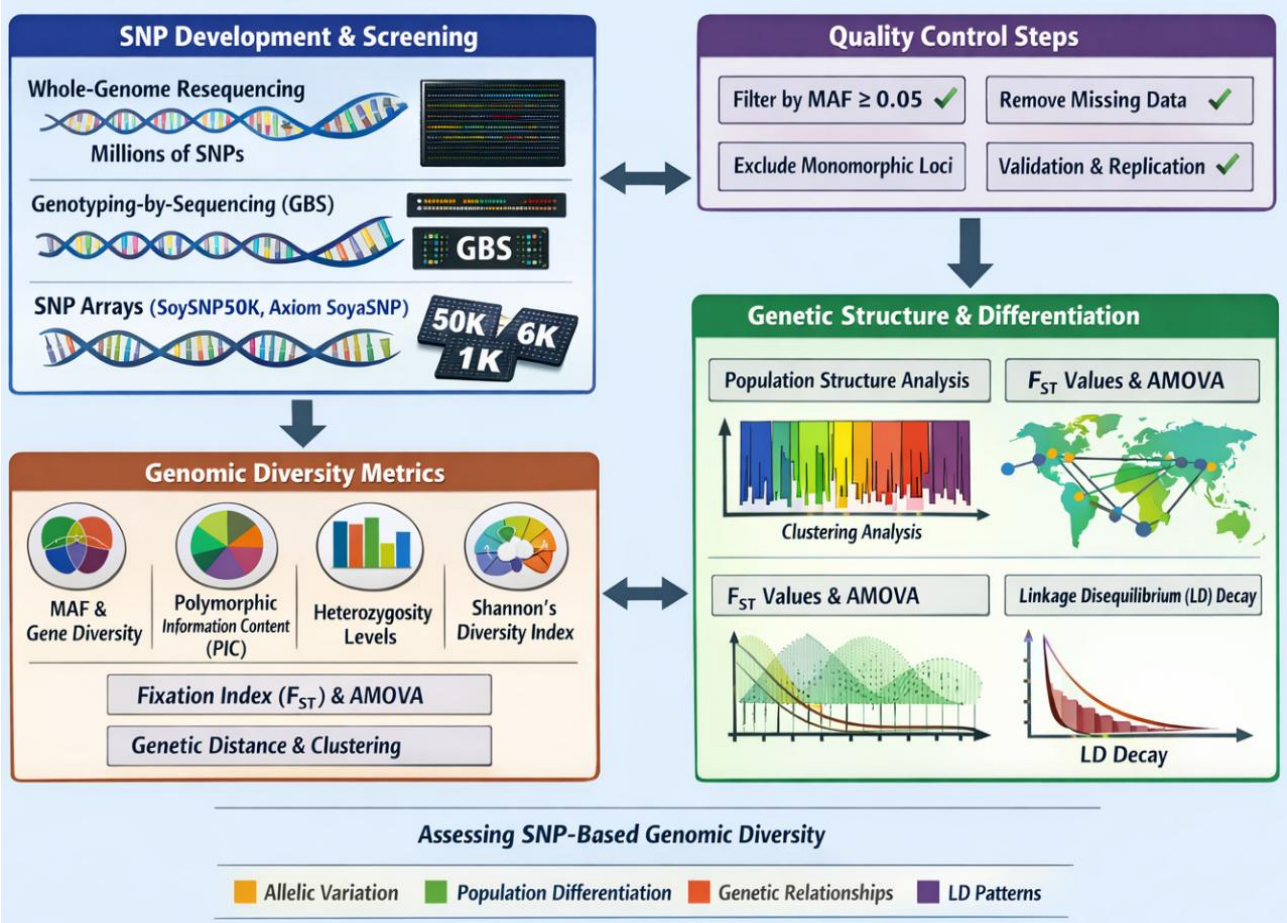

Figure 2 Methods for analyzing soybean genomic diversity using SNP markers

3.3 Data analysis and bioinformatics methods

SNP datasets generated from arrays, GBTS, or GBS platforms were processed through standardized bioinformatics pipelines prior to population-genetic analyses. Initial steps included SNP calling against the reference genome, filtering for missingness and MAF thresholds, and removal of redundant or poorly mapped markers to yield a high-quality matrix of individuals by loci (Niu et al., 2024). Genetic diversity parameters and marker statistics (PIC, gene diversity, heterozygosity, allele frequencies) were calculated using specialized software such as PowerMarker, POPGENE, or comparable population-genetic packages (Abebe et al., 2021). To explore population structure, Bayesian clustering using STRUCTURE and likelihood-based determination of the optimal number of subpopulations (K) via the Evanno ΔK method were commonly applied under admixture models, often with thousands of burn-in and MCMC iterations to ensure convergence.

Complementary multivariate methods, including principal component analysis (PCA) or principal coordinate analysis (PCoA), and discriminant analysis of principal components (DAPC), were used to visualize genetic relationships and validate STRUCTURE-defined clusters (Chander et al., 2021). Hierarchical clustering (e.g. UPGMA or neighbor-joining based on genetic distance matrices) and phylogenetic tree construction provided additional perspectives on the grouping of accessions by geography, improvement status, or pedigree (Rani et al., 2023). AMOVA was implemented to quantify variance components and F-statistics among and within groups, while software such as STRUCTURE HARVESTER and R packages (e.g. adegenet) facilitated model selection and graphical output (Shaibu et al., 2021). For very large resequencing-based SNP resources, further analyses included genome-wide LD estimation, identification of conserved and highly variable genomic intervals, and functional annotation of large-effect SNPs and InDels using gene ontology and pathway databases to link diversity patterns with candidate genes underlying key agronomic traits (Valliyodan et al., 2021).

4 Population Structure Analysis of Global Soybean Germplasm Resources

4.1 Methods for analyzing population structure

Population structure analysis of soybean germplasm relies on multilocus genotyping combined with multivariate and model-based statistical approaches. High-throughput SNP platforms (arrays, DArTseq, GBS, and whole-genome resequencing) now provide thousands to tens of thousands of markers distributed across all 20 chromosomes, enabling robust inference of subpopulations and admixture in large panels of cultivated and wild accessions (Valliyodan et al., 2021; Zatybekov et al., 2025). Common analytical workflows begin with estimation of basic diversity indices (allele frequencies, expected heterozygosity, polymorphism information content), followed by clustering using principal component analysis (PCA), principal coordinate analysis (PCoA), and distance-based phylogenetic trees such as UPGMA or neighbor-joining (Andrijanić et al., 2023). These multivariate methods offer an initial visualization of genetic relationships and can reveal major divisions between wild and cultivated gene pools, between regions, or among breeding groups. Model-based Bayesian clustering, implemented in programs such as STRUCTURE and related admixture models, is then used to infer the most likely number of genetic clusters (K), assign membership coefficients to each accession, and quantify admixture proportions (Chander et al., 2021).

Complementary statistics deepen insight into the organization of diversity. Analysis of molecular variance (AMOVA) partitions total genetic variation within and among predefined groups (e.g., regions, maturity groups, breeding programs), clarifying whether diversity is primarily within or between populations (Da Silva et al., 2025;). Fixation indices (F_ST) quantify genetic differentiation between pairs of populations and are widely used to classify divergence as negligible, moderate, or strong, guiding the choice of contrasting parents for crossing (Tsindi et al., 2023). Linkage disequilibrium (LD) decay analyses, often based on genome-wide SNPs, inform on historical recombination and selection, and help define the resolution of association mapping in each subpopulation. Recent studies also integrate haplotype-based analyses and genome-wide scans for selection (e.g., BayeScan, EigenGWAS) to identify genomic regions whose allele frequency differentiation aligns with population structure and local adaptation (Kim et al., 2025). Together, these methods provide a coherent framework for dissecting population structure, controlling for stratification in GWAS, and designing efficient germplasm utilization strategies.

4.2 Genetic differentiation among soybean populations from different geographic origins

Comparative SNP-based studies consistently show that geographic origin and domestication status are primary drivers of soybean population structure. Large-scale analyses of global germplasm, including the USDA collection and broad Korean–Chinese–Japanese panels, clearly separate wild (*Glycine soja*) from cultivated (*Glycine max*) accessions, with wild populations further partitioned into multiple lineages that track their East Asian collection zones (Kaga et al., 2012; Li et al., 2024). Within cultivated soybean, Asian, North American, South American, and European gene pools typically form distinct but partially overlapping clusters, reflecting historical patterns of germplasm exchange, founder effects, and regional selection (Potapova et al., 2023). For example, Japanese and Korean accessions are relatively homogeneous and distinct from Chinese accessions, while American cultivars derive their ancestry largely from a subset of Chinese subpopulations (Jeong et al., 2018). European cultivars cluster into two main groups with substructure corresponding to country of origin and maturity group; American introductions show the lowest differentiation from European material, whereas Swiss lines and some Eastern European cultivars are more distinct.

Regional studies further highlight variable levels of differentiation and diversity among emerging production areas. In sub-Saharan Africa, elite TGx lines and cultivars adapted to African environments form several SNP-defined clusters, but overall show a broad genetic base compared with some temperate breeding pools (Tsindi et al., 2023). Southern African collections combining temperate and tropical material exhibit very low F_ST (~0.06) between subgroups, indicating weak genetic differentiation and extensive germplasm sharing across programs (Tsindi et al., 2023). Brazilian germplasm, in contrast, often displays a narrow genetic base and strong signatures of selection, with structure shaped by region, company, and relative maturity group; Asian accessions are consistently the most

differentiated and genetically diverse reference group. Newer soybean regions such as Kazakhstan and West Siberia show accessions most similar to European and North American cultivars, with low within-group diversity in Kazakhstan pointing to a particularly narrow local base (Potapova et al., 2023). Studies of wild soybean at continental scale also reveal deep north–south differentiation and distinct lineages in Korea, Japan, and different parts of China, underscoring the importance of geographic structure in the ancestral gene pool (Meng et al., 2023).

4.3 The relationship between population structure and genetic diversity

Population structure and genetic diversity are tightly coupled, with structure both reflecting historical changes in diversity and influencing how existing variation can be used in breeding. AMOVA and F_ST estimates across multiple SNP-based studies show that the majority of variation in soybean is usually found within, rather than among, populations, even when clear geographic or breeding-program clusters are present (Shaibu et al., 2021; Rani et al., 2023). For example, analyses of European cultivars, African germplasm, and IITA accessions all report 90%–98% of variance within populations and only a small fraction attributed to differences among countries, maturity groups, or STRUCTURE-defined clusters (Lukanda et al., 2023). Similarly, Brazilian and African collections often exhibit low to moderate F_ST between subpopulations despite discernible clustering by origin, maturity, or company, indicating substantial shared allelic backgrounds and extensive germplasm exchange . This pattern implies that carefully chosen parents from within a region can still capture meaningful diversity, but that crossing between more differentiated geographic or wild–cultivated groups is necessary to introduce novel alleles.

At the same time, strong population structure can signal both reservoirs of unique variation and zones of genetic erosion. Wild soybean lineages generally harbor higher nucleotide diversity and stronger geographic differentiation than cultivated pools, confirming their value as sources of private alleles for stress tolerance and adaptive traits (Li et al., 2024). In contrast, historical breeding and domestication have produced monophyletic or weakly structured cultivated groups with conserved haplotypes in genomic regions under selection for yield, maturity, or seed composition (Valliyodan et al., 2021). Studies of Brazilian cultivars, European germplasm, and large resequenced panels identify large fixed or low-diversity segments associated with key agronomic QTL, alongside more diverse genomic regions that still retain useful variability (Andrijanić et al., 2023; Kim et al., 2025). In emerging regions such as Kazakhstan, the combination of clear clustering with temperate germplasm and very low within-group diversity indicates a narrow, vulnerable genetic base, motivating targeted introgression from diverse foreign and wild accessions. Thus, integrating population-structure analysis with diversity metrics helps breeders balance immediate adaptation needs—by exploiting existing structured variation—with long-term goals of broadening the genetic base through informed use of differentiated, high-diversity gene pools.

5 Associations Between Soybean Genomic Diversity and Important Agronomic Traits

5.1 The relationship between genetic diversity and yield traits

Genomic diversity underpins phenotypic variation in key yield components such as seed yield per plant, number of pods and seeds, plant height, and 100-seed weight. Classical quantitative genetic studies across diverse soybean panels consistently report significant genotypic variance, high heritability, and substantial genetic advance for grain yield and its components, indicating abundant additive genetic variation that can be exploited through selection (Mitiku et al., 2025). Correlation and path analyses show that traits including number of seeds per plant, number of pods per plant, plant height, 100-seed weight, biological yield, and harvest index are positively and often strongly associated with seed yield, and frequently exert high positive direct effects, identifying them as efficient indirect selection targets. Morphological assessments combined with molecular markers (e.g., SSRs) further reveal that genotypes grouped as genetically distant often carry complementary yield-enhancing alleles, and crosses between such divergent parents tend to maximize transgressive segregation for yield (Ferreira et al., 2025).

SNP-based association studies refine these relationships by linking diversity at specific loci and haplotypes to yield and yield components. Nested association mapping and diversity panels genotyped with high-density SNP arrays or GBS have identified dozens of loci and haplotypes affecting yield, maturity, plant height, lodging, seed

mass, and related traits across multiple environments (Ravelombola et al., 2021). In several studies, stable QTL or SNP-based haplotypes co-regulate seed yield and component traits such as 100-seed weight or seeds per plant, reflecting pleiotropy or tight linkage and clarifying trade-offs among traits. High-diversity association panels encompassing landraces, elite cultivars, and exotic accessions enhance the power to detect such loci and confirm that exotic and wild-derived alleles can increase yield or specific components in elite backgrounds (Diers et al., 2018). Thus, genomic diversity, when captured with dense SNP markers, directly translates into exploitable allelic variation for yield improvement and guides the design of heterotic and complementary crossing schemes.

5.2 The relationship between genetic diversity and stress tolerance traits

Genomic diversity is also crucial for buffering soybean against biotic and abiotic stresses, with SNP-based GWAS increasingly clarifying the genetic bases of stress tolerance traits. High-density SNP genotyping of diverse germplasm panels has identified loci and candidate genes associated with root system architecture, which underlies nutrient uptake efficiency and tolerance to drought and other climate-related stresses (Kim et al., 2023). For example, GWAS using 180K or SLAF-seq SNP datasets in landraces and spring soybean panels detected over 100 significant loci for root and shoot traits, and prioritized candidate genes whose expression levels correlate with root branching and early seedling vigor, traits strongly linked to resilience under low-input or stressful environments. Similarly, genome-wide analyses of seed flooding tolerance at germination identified SNPs and hub genes associated with electrical conductivity, germination rate, root length, and shoot length under flooding, revealing allelic variants that confer enhanced stress tolerance and can be pyramided by marker-assisted breeding (Sharmin et al., 2021).

Resistance to major diseases such as soybean mosaic virus (SMV) also depends on standing genomic variation at resistance loci. GWAS of global or regional panels challenged with SMV strains uncovered multiple resistance loci across chromosomes and pinpointed candidate genes such as Glyma.04G086700, encoding an LRR protein kinase involved in pathogen recognition, with distinct haplotypes explaining differential resistance responses among accessions (Zhao et al., 2025). Population structure analyses in these panels indicate that specific resistance alleles or haplotypes are often enriched in particular geographic or breeding subgroups, emphasizing the need to sample broadly to capture the full spectrum of stress-related diversity (Sharmin et al., 2021). Collectively, these findings demonstrate that maintaining and utilizing genomic diversity—particularly in landraces, wild relatives, and regionally adapted varieties—provides the allelic reservoir necessary for breeding soybean cultivars resilient to current and emerging stresses.

5.3 Applications of SNP markers in genome-wide association studies (GWAS)

SNP markers form the backbone of modern GWAS in soybean and have transformed understanding of the genetic architecture of yield, domestication, and adaptive traits. High-density arrays (e.g., 50K–180K SoyaSNP) and GBS or SLAF-seq platforms routinely generate tens to hundreds of thousands of polymorphic SNPs across germplasm panels, providing sufficient marker density for genome-wide coverage and fine mapping of loci through linkage disequilibrium (Ravelombola et al., 2021). GWAS using mixed-model and multilocus approaches (e.g., MLM, MLMM, FarmCPU, BLINK) have identified numerous SNPs and quantitative trait nucleotides for seed yield, maturity, plant height, seed weight, pod and seed number, root traits, domestication-related traits, and stress responses (Mandozai et al., 2021). Many significant SNPs co-localize with known QTL or cloned genes (e.g., E-loci for maturity, Dt1 for plant height, pod-shattering genes), while others represent novel regions underlying complex trait variation or domestication signatures (Sonah et al., 2015).

Beyond single-marker tests, haplotype-based GWAS and integrative models extend the utility of SNP datasets. Haplotype analyses refine association signals, identify stable multi-SNP blocks with strong effects across environments, and reveal pleiotropic haplotypes affecting multiple agronomic traits (Bhat et al., 2022). Structural equation model-based GWAS further decomposes SNP effects into direct and indirect components along causal trait networks, clarifying how genomic regions simultaneously influence yield components such as pod number, grain number, and seed weight (Suela et al., 2025). Many GWAS also couple SNP discovery with genomic

prediction analyses, demonstrating that SNP-based models such as rrBLUP and other machine-learning approaches can achieve moderate to high predictive accuracies for yield, maturity, plant height, seed weight, disease resistance, and stress tolerance, though performance is trait-and environment-dependent (Ravelombola et al., 2021). The SNPs, haplotypes, and candidate genes revealed by these GWAS now underpin marker-assisted selection, development of diagnostic assays, and genomic selection pipelines, directly linking genomic diversity and population structure to practical improvement of soybean agronomic performance.

6 Case Study: Population Structure Analysis of Global Soybean Core Germplasm

6.1 Principles for constructing core germplasm collections

Core germplasm collections are designed to capture the maximum genetic and phenotypic diversity of an entire genebank in a substantially reduced number of accessions, thereby making evaluation and utilization more efficient. Conceptually, a soybean core collection usually represents about 2%–5% of the entire collection, with mini-core sets often reduced to ~1% while still maintaining major patterns of geographic and agro-morphological variation. Construction typically begins with stratification of the base collection by ecoregion, maturity group, improvement status (wild, landrace, cultivar), and key phenotypes such as seed size or growth habit, followed by sampling within strata using proportional, logarithmic, or multivariate strategies (Oliveira et al., 2010). In soybean, core development has progressively integrated molecular marker data—initially SSR and later SNP genotypes—together with multi-environment phenotypic data to ensure that rare alleles and extreme trait values are retained at acceptable frequencies (Lijuan et al., 2009).

Several methodological frameworks have been proposed to operationalize these principles in soybean. The Chinese national platform developed a three-tier system of core, mini-core, and integrated applied core collections, combining SSR-based diversity with extensive phenotyping to select accessions that represent broad gene pools as well as trait-specific subsets for stress tolerance and quality (Guo et al., 2014). The USDA Soybean Germplasm Collection applied stratified and multivariate sampling of passport, phenotypic, and later SNP data to assemble cores that maximize variability while preserving quantitative trait distributions of the entire collection (Satyawan and Tasma, 2021). More recently, algorithmic approaches such as Core Hunter have been applied directly to high-density SNP datasets from >20,000 accessions to identify a few hundred entries with maximal pairwise genetic distance, while maintaining original allele frequencies and phenotypic variation for key traits like yield and branching (Song et al., 2015). Collectively, these experiences establish that effective soybean core sets must balance representativeness, low redundancy, manageability, and explicit conservation of molecular diversity.

6.2 Elucidation of population structure based on SNP markers

High-density SNP genotyping has transformed population-structure analysis in soybean core and global collections by enabling genome-wide characterization of relationships among thousands of accessions. Using tens of thousands of SNPs from whole-genome resequencing or SoySNP50K-type arrays, Bayesian clustering, ADMIXTURE/STRUCTURE analysis, and principal component analysis (PCA) routinely distinguish wild (*Glycine soja*) from cultivated (*G. max*) groups and reveal transitional accessions with mixed ancestry (Zatybekov et al., 2025). Within cultivated soybean, SNP-based analyses consistently identify geographically and ecologically coherent subpopulations, such as distinct clusters for Chinese, Japanese, Korean, and American germplasm, as well as separations among tropical, temperate, and high-latitude maturity groups (Tsindi et al., 2023). For example, a 14,000-accession SNP survey of the USDA collection resolved five major ancestral clusters and demonstrated that most North American cultivars trace their ancestry to a limited subset of Chinese landrace gene pools (Bandillo et al., 2015).

Case studies using SNP-genotyped regional cores highlight both the power and limitations of global germplasm. In a WGRS-based analysis of 694 accessions including Kazakh, European, North American, and wild soybeans, PCA and phylogenetic trees showed Kazakh cultivars clustering closely with European and North American material, while maintaining clear separation from G. soja; however, Kazakh accessions exhibited the lowest within-group diversity, underscoring a narrow genetic base for this emerging production region (Zatybekov et al.,

2025). Similarly, SNP-based studies in Southern Africa and Central Europe revealed low F_ST among introduced temperate lines and modest overall molecular diversity, reflecting intensive germplasm exchange but also high redundancy and shared pedigrees (Haupt and Schmid, 2019; Tsindi et al., 2023). When core collections are derived from such global datasets, population-structure analysis is essential not only for defining clusters and admixture patterns, but also for guiding sampling so that both major groups and admixed genotypes are proportionally represented. In SNP-driven USDA core construction, for instance, structure analysis showed that optimization algorithms tend to favor admixed accessions, which can efficiently capture allelic variation from multiple ancestral groups in a limited number of entries (Figure 3) (Satyawan and Tasma, 2021).

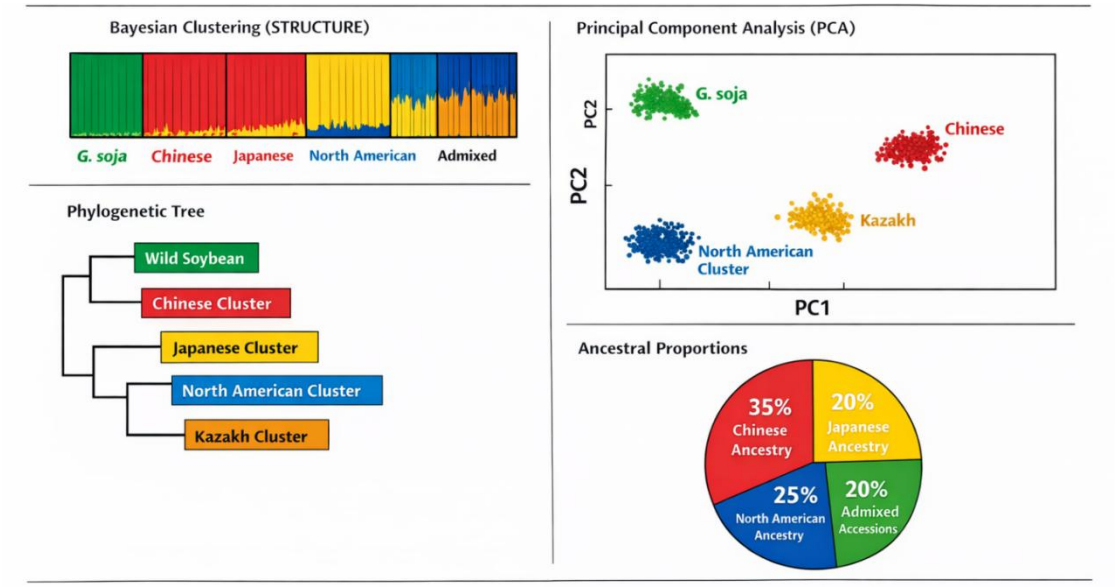

Figure 3 Population structure analysis of global soybean core germplasm (Adopted from Satyawan and Tasma, 2021)

6.3 Applications of core germplasm resources in molecular breeding

Core and mini-core collections genotyped with dense SNP markers serve as powerful diversity panels for molecular breeding, enabling efficient trait discovery, allele mining, and pre-breeding. Because they retain most of the allelic richness and phenotypic range of the base collection, these panels are ideal for genome-wide association studies (GWAS) targeting complex traits such as seed protein and oil content, flowering time, and stress tolerance (Jo et al., 2023). The SoySNP50K-genotyped USDA collection has already supported large-scale GWAS that identified and refined major loci controlling seed composition traits, including narrowing a chromosome-20 region for protein/oil to a handful of candidate genes and confirming many previously mapped QTL (Bandillo et al., 2015). Similarly, wild soybean core subsets have been used to map loci for days to flowering and maturity, revealing allelic variation at E-genes that can be introgressed into cultivars to broaden adaptation (Jo et al., 2023). Trait-specific “integrated applied core collections” composed of accessions with documented resistance to cold, drought, salinity, soybean cyst nematode, and viral diseases provide ready-to-use donor sets for marker-assisted backcrossing and pyramiding of multiple resistance genes (Li et al., 2023).

Beyond gene discovery, SNP-anchored core collections help breeders rationalize crossing schemes and widen genetic bases in targeted environments. Objective-driven cores assembled for Central European or Southern African conditions, for example, combine environmentally pre-adapted accessions with maximum molecular diversity, creating tailored panels for phenotyping under local climates and for identifying parental combinations that optimize heterogeneity while avoiding close relatedness (Tsindi et al., 2023). In countries with emerging soybean industries such as Kazakhstan, structure and diversity analyses of local versus global germplasm suggest dual strategies: introgressing novel alleles from wild and exotic sources to broaden the base, while simultaneously increasing the frequency of favorable alleles already present in adapted lines (Zatybekov et al., 2025). At a global scale, SNP-based core populations that maintain haplotype diversity and LD structure are also central to genomic

selection pipelines, where genomic estimated breeding values can be trained on core panels and applied to larger breeding populations. As genebank-scale SNP datasets become ubiquitous, integrating well-designed core collections with high-throughput phenotyping and omics will be critical for converting conserved diversity into elite, climate-resilient soybean cultivars.

7 Prospects for the Application of SNP Markers in Soybean Genetic Improvement

7.1 Marker-assisted selection (MAS) breeding

SNP markers are now central to marker-assisted selection in soybean because they are abundant, codominant, and amenable to high-throughput, low-cost genotyping. A series of SNP arrays and targeted panels (e.g., BARCSoySNP6K, SoySNP3K, SoySNP1K, and GBTS-based 10K–40K panels) have been specifically optimized for breeding, providing genome-wide coverage with 1,000–6,000 informative markers that are sufficient for most MAS and genomic applications while avoiding redundant information and unnecessary costs (Niu et al., 2024). These panels show high genotyping accuracy, with >98% concordance to resequencing data and high minor allele frequencies in elite germplasm, ensuring that selected markers are polymorphic and reliable across breeding populations (Yang et al., 2023). Such platforms support routine tasks in breeding pipelines, including germplasm fingerprinting, pedigree verification, rapid backcross recovery, and early-generation selection of lines carrying desirable alleles for key loci.

Trait-linked SNPs and KASP/TaqMan assays derived from GWAS, QTL mapping, and candidate gene studies are increasingly deployed to target specific agronomic and quality traits. For example, tightly linked SNP assays to the salt-tolerance gene GmCHX1 accurately distinguish tolerant and sensitive genotypes in diverse panels and biparental populations (>91%–98% classification accuracy), greatly facilitating the introgression of salinity tolerance into elite backgrounds (Patil et al., 2016). Similarly, a diagnostic TaqMan SNP test for pod-shattering resistance (KSS-SNP5) achieved 92%–96% prediction accuracy across F2:3 and advanced breeding lines, demonstrating that single-locus MAS can efficiently enrich resistant genotypes and reduce costly phenotyping for difficult traits (Kim et al., 2020). GWAS-identified SNPs and haplotypes controlling yield components, seed protein, sucrose, and other seed composition traits are also being integrated into MAS schemes, where pyramiding favorable alleles at multiple minor-effect loci can substantially increase phenotypic variation explained for target traits (Ravelombola et al., 2021; Qin et al., 2022; Rïaz et al., 2023). As more trait-diagnostic SNPs are validated across environments and genetic backgrounds, MAS based on robust SNP assays will remain a practical, cost-effective strategy, especially for major genes and moderate-effect loci with clear, stable effects.

7.2 Genomic selection (GS) breeding

Beyond locus-specific MAS, genome-wide SNP datasets enable genomic selection, which estimates genomic breeding values from all markers simultaneously and is particularly powerful for complex, polygenic traits such as yield and seed composition. Multiple studies in soybean demonstrate that GS can achieve moderate to high predictive accuracies for protein, oil, seed weight, maturity, and related traits using ridge regression BLUP, GBLUP, Bayesian models, and selected machine-learning approaches (Jiahao et al., 2025). For instance, GS accuracies of ~0.75–0.87 for seed weight and ~0.81 and 0.71 for protein and oil have been reported in elite breeding populations, clearly outperforming traditional MAS for these quantitative traits (Ravelombola et al., 2021; Qin et al., 2022). Even for grain yield, where prediction remains more challenging, GS routinely achieves useful accuracies (0.26–0.4) that can accelerate selection cycles when combined with optimized training sets and appropriate statistical models (Ćeran et al., 2024).

Efficient GS pipelines depend critically on SNP density, marker quality, training population composition, and the selection of informative marker subsets. Empirical evaluations in soybean suggest that approximately 1,000–2,000 genome-wide, well-distributed SNPs are sufficient to reach a plateau in prediction accuracy; further increases in marker number add cost but little additional information (Qin et al., 2022; Song et al., 2024). Marker sets derived from GWAS—i.e., SNPs significantly associated with the target trait—can further improve prediction efficiency and allow high accuracies at relatively low marker densities (~5K), especially when combined with Bayesian models (Rïaz et al., 2023). Selective genotyping and phenotyping schemes that maintain the genetic diversity of

the initial population while reducing the number of genotypes or markers offer additional cost savings without compromising accuracy, particularly when model-based strategies are used to choose training individuals and markers (Ćeran et al., 2024). With these methodological and technological advances, SNP-enabled GS is poised to become a routine component of soybean breeding programs, supporting rapid recycling of parents, optimal cross prediction, and multi-trait selection for yield, quality, and adaptation (Jiahao et al., 2025).

7.3 Multi-omics technologies and soybean genetic improvement

The expanding use of SNP markers is increasingly integrated with other omics layers—transcriptomics, metabolomics, and phenomics—to dissect complex traits and drive more precise soybean improvement. High-density SNParrays and resequencing provide the foundational genomic variation, which can be linked to expression (eQTL) data, metabolite profiles, and detailed phenotypes to identify causal genes and pathways underlying yield, stress tolerance, and seed quality (Miller et al., 2023; Gai et al., 2025). For example, combining GWAS with transcriptomic data has helped prioritize candidate genes within QTL regions for traits such as flowering and maturity, while metabolite-associated SNPs refine the genetic control of nutritional components like sucrose, isoflavones, tocopherols, and amino acids (Jiahao et al., 2025). Multi-omics data also reveal pleiotropic effects and gene networks, providing systems-level targets for breeding and genome editing rather than focusing solely on single loci.

In parallel, advances in high-throughput genotyping (e.g., GBTS panels, GBS) and phenotyping, together with machine-learning and artificial-intelligence approaches, are reshaping how breeders exploit SNP-based diversity in a multi-omics context. Integrative frameworks now couple SNP-based genomic prediction with environmental, physiological, and management data to model genotype-by-environment interactions and support climate-resilient cultivar development (Miller et al., 2023). Multi-omics-informed GS models, which incorporate SNPs alongside expression or metabolite markers, are being explored to improve prediction for complex traits and to design ideotypes optimized for both productivity and sustainability. At the same time, large resequencing datasets and SNP resources spanning thousands of accessions facilitate the construction of mutant gene libraries and enable rapid identification of natural alleles suitable for CRISPR/Cas-based editing or allele replacement. Together, these developments indicate that SNP markers will remain the genomic backbone of soybean improvement, increasingly embedded within holistic, multi-omics breeding strategies that integrate MAS, GS, and genome editing to accelerate genetic gain and broaden the adaptive potential of global soybean germplasm.

8 Summary and Outlook

Over the past decade, SNP genotyping and resequencing have transformed understanding of global soybean genetic diversity and population structure. Large-scale efforts have characterized hundreds to thousands of accessions spanning wild relatives, landraces, and elite cultivars, revealing millions to tens of millions of SNPs and providing a high-resolution view of genome-wide variation. These studies consistently distinguish wild from cultivated groups, identify transitional or hybrid genotypes, and show that domestication and modern breeding have dramatically reduced diversity and reshaped linkage disequilibrium (LD) through selective sweeps. At regional scales, population-wide SNP analyses have clarified how breeding history and geography structured germplasm in Brazil, Southern Africa, Kazakhstan, and Korea, often revealing a small number of genetic clusters and strong within-population variation with relatively low differentiation among groups. Such insights have directly informed strategies to broaden the genetic base and guide parental selection for adaptation to tropical, temperate, or stress-prone environments

Parallel advances in SNP array and GBS platforms (e.g., SoySNP50K, Axiom SoyaSNP, DArT-SNP, and custom low-to medium-density panels) have delivered robust, evenly distributed, and cost-effective marker sets that are now widely used for diversity analysis, fingerprinting, and association mapping. Consolidation of resequencing data for 1.5K and 3,661+ accessions into unified variant resources with 30–32 million SNPs has created versatile public datasets for post-genomic research, facilitating in-silico genotyping, high-resolution GWAS, and cross-collection comparisons. These genomic resources underpin discovery of loci for domestication traits, yield components, quality, and stress tolerance, and enable identification of region-specific or rare favorable alleles in

landraces and wild soybean. Collectively, SNP-based diversity and structure analyses have shifted soybean from a sparsely characterized crop to one with dense global genomic coverage, providing a foundation for modern molecular breeding.

Despite this progress, several challenges limit full exploitation of genomic diversity for soybean improvement. Many regional germplasm pools, including Brazilian, Kazakh, and Southern African collections, show narrow genetic bases and low molecular diversity, often reflecting heavy reliance on a small set of ancestors and extensive germplasm sharing. Such bottlenecks constrain long-term genetic gain, reduce resilience to emerging stresses, and increase vulnerability to climate change. Even where diversity exists globally, it is unevenly represented in working breeding pools, and pre-breeding to introgress favorable alleles from wild and exotic sources remains limited and slow. In addition, most diversity studies report that the overwhelming proportion of variation resides within rather than among populations, emphasizing that effective use of diversity requires careful within-pool sampling and crossing strategies, not just inter-population contrasts.

Methodological and translational gaps also persist. Although high-density SNP resources are abundant, their integration with deep, standardized multi-environment phenotyping remains incomplete, leading to underpowered GWAS for complex traits and limited validation of candidate genes. Many association signals are population-or environment-specific, and genotype-by-environment interactions reduce the robustness of marker–trait relationships for direct deployment in breeding. In emerging production regions, infrastructure for high-throughput genotyping and data analysis is often insufficient, slowing adoption of genomic tools. Finally, even where strong genomic resources exist (e.g., USDA and Asian collections), regulatory, logistical, and data-sharing barriers can impede the exchange and utilization of diverse germplasm and associated genomic data at a truly global scale.

Future work on soybean genomic diversity should prioritize systematic broadening of breeding germplasm using global SNP and resequencing resources as guides. Whole-genome analyses already demonstrate that wild soybean, Asian landraces, and regionally adapted landraces harbor substantially higher diversity and distinct haplotypes than many modern cultivars. Strategic identification of complementary parental combinations—such as crossings between narrow-based regional pools and diverse foreign or wild accessions—can be optimized using genome-wide similarity matrices, haplotype block maps, and FST scans to maximize novel allelic recombination while preserving local adaptation. Pre-breeding pipelines that couple genomic prediction with targeted introgression of domestication and adaptation alleles from underutilized germplasm will be essential to generate broadly adapted, yet genetically rich, base populations for both temperate and tropical regions.

At the same time, integrating SNP-based diversity analyses with multi-omics and advanced computational approaches will deepen understanding of how genetic variation translates into phenotype. Combining genome-wide SNP data with transcriptomics, metabolomics, epigenomics, and high-throughput phenotyping, together with AI-driven modeling, can resolve causal genes and networks underlying yield, stress resilience, and quality traits and refine models of genotype-by-environment interaction. Large consolidated variant resources and curated mutant libraries derived from millions of SNPs and InDels offer powerful starting points for reverse genetics and genome editing to validate candidate alleles and engineer ideal haplotypes. Continued development of low-cost, breeder-friendly SNP panels tailored to regional germplasm, combined with training and infrastructure for MAS, GS, and GWAS in under-resourced programs, will help translate genomic diversity knowledge into practical genetic gain worldwide. Together, these directions point toward a future in which global soybean improvement is driven by coordinated, data-rich exploitation of the full spectrum of genomic diversity.

Acknowledgments

Thanks to the reviewers for providing detailed comments and guidance on the manuscript of this study. The reviewers' keen insights into the issues and attention to detail have greatly benefited the authors.

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

Abebe A., Kolawole A., Unachukwu N., Chigeza G., Tefera H., and Gedil M., 2021, Assessment of diversity in tropical soybean (*Glycine max* (L.) Merr.) varieties and elite breeding lines using single nucleotide polymorphism markers, Plant Genetic Resources: Characterization and Utilization, 19(1): 20-28. https://doi.org/10.1017/S1479262121000034

Andrijanić Z., Nazzicari N., Šarčević H., Sudarić A., Annicchiarico P., and Pejić I., 2023, Genetic diversity and population structure of european soybean germplasm revealed by single nucleotide polymorphism, Plants, 12(9): 1837. https://doi.org/10.3390/plants12091837

Bandillo N., Jarquín D., Song Q., Nelson R., Cregan P., Specht J., and Lorenz A., 2015, A population structure and genome-wide association analysis on the USDA soybean germplasm collection, The Plant Genome, 8(2): plantgenome2015.04.0024. https://doi.org/10.3835/plantgenome2015.04.0024

Bhat J.A., Adeboye K.A., Ganie S.A., Barmukh R., Hu D., Varshney R.K., and Yu D., 2022, Genome-wide association study, haplotype analysis, and genomic prediction reveal the genetic basis of yield-related traits in soybean (*Glycine max* L.), Frontiers in Genetics, 13: 953833. https://doi.org/10.3389/fgene.2022.953833

Bunjkar A., Walia P., and Sandal S., 2024, Unlocking genetic diversity and germplasm characterization with molecular markers: strategies for crop improvement, Journal of Advances in Biology and Biotechnology, 27(6): 873. https://doi.org/10.9734/jabb/2024/v27i6873

Ćeran M., Đorđević V., Miladinović J., Vasiljević M., Đukić V., Ranđelović P., and Jaćimović S., 2024, Selective genotyping and phenotyping for optimization of genomic prediction models for populations with different diversity, Plants, 13(7): 975. https://doi.org/10.3390/plants13070975

Chander S., Garcia-Oliveira A., Gedil M., Shah T., Otusanya G., Asiedu R., and Chigeza G., 2021, Genetic diversity and population structure of soybean lines adapted to sub-saharan africa using single nucleotide polymorphism (SNP) markers, Agronomy, 11(3): 604. https://doi.org/10.3390/agronomy11030604

Contreras-Soto R.I., Mora F., de Oliveira M.A.R., Higashi W., Scapim C.A., and Schuster I., 2017, A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype analysis, PLOS ONE, 12(2): e0171105. https://doi.org/10.1371/journal.pone.0171105

da Silva A.J., da Silva D.C.G., Ferreira E.G., Abdelnoor R.V., Borém A., Arias C.A.A., and Marcelino-Guimarães F.C., 2025, Genetic diversity, population structure in a historical panel of Brazilian soybean cultivars, PLOS ONE, 20(1): e0313151. https://doi.org/10.1371/journal.pone.0313151

Diers B.W., Specht J., Rainey K.M., Cregan P., Song Q., Ramasubramanian V., Graef G., Nelson R., Schapaugh W., Wang D., Shannon G., McHale L., Kantartzi S., Xavier A., Mian R., Stupar R., Michno J., An Y., Goettel W., Ward R., Fox C., Lipka A., Hyten D., Cary T., and Beavis W.D., 2018, Genetic Architecture of Soybean Yield and Agronomic Traits, G3: Genes|Genomes|Genetics, 8(10): 3367-3375. https://doi.org/10.1534/g3.118.200332

Dong Q., Cheng Y., Li Y., Tong Y., Liu D., Yu J., Zhao N., Liu B., Ding X., and Xu C., 2025, Genome-wide association study and genomic prediction of essential agronomic traits in diversity panel of soybean varieties, Agronomy, 15(5): 1181. https://doi.org/10.3390/agronomy15051181

Duan Z., Xu L., Zhou G., Zhu Z., Wang X., Shen Y., Tian Z., and Fang C., 2025, Unlocking soybean potential: genetic resources and omics for breeding, Journal of Genetics and Genomics, 52(4): 100997. https://doi.org/10.1016/j.jgg.2025.02.004

Ferreira S.C., Dias P.P., Rezende A.A., Gomes B.F., Bonetti A.M., and Nogueira A.P.O., 2025, Analysis of genetic diversity in soybean based on agronomic traits and microsatellite markers, Ciência e Agrotecnologia, 49: e017424. https://doi.org/10.1590/1413-7054202549017424

Fu Y.B., Cober E.R., Morrison M.J., Marsolais F., Peterson G.W., and Horbach C., 2021, Patterns of genetic variation in a soybean germplasm collection as characterized with genotyping-by-sequencing, Plants, 10(8): 1611. https://doi.org/10.3390/plants10081611

Gai Y., Liu S., Zhang Z., Wei J., Wang H., Liu L., Bai Q., Qin Q., Zhao C., Zhang S., Xiang N., and Zhang X., 2025, Integrative approaches to soybean resilience, productivity, and utility: a review of genomics, computational modeling, and economic viability, Plants, 14(5): 671. https://doi.org/10.3390/plants14050671

Guo Y., Li Y., Hong H., and Qiu L., 2014, Establishment of the integrated applied core collection and its comparison with mini core collection in soybean (*Glycine max*), The Crop Journal, 2(1): 38-45.

Hasan N., Choudhary S., Naaz N., Sharma N., and Laskar R.A., 2021, Recent advancements in molecular marker-assisted selection and applications in plant breeding programmes, Journal of Genetic Engineering and Biotechnology, 19(1): 128. https://doi.org/10.1186/s43141-021-00231-1

Haupt M., and Schmid K.J., 2019, Combining focused identification of germplasm and core collection strategies to identify genebank accessions for central European soybean breeding, bioRxiv, preprint: 848978.
https://doi.org/10.1101/848978

He J., Zhao X., Laroche A., Lu Z.X., Liu H., and Li Z., 2014, Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding, Frontiers in Plant Science, 5: 484.
https://doi.org/10.3389/fpls.2014.00484

Jeong S.C., Moon J.K., Park S.K., Kim M.S., Lee K., Lee S.R., Jeong N., Choi M.S., Kim N., Kang S.T., and Park E., 2019, Genetic diversity patterns and domestication origin of soybean, Theoretical and Applied Genetics, 132(4): 1179-1193.
https://doi.org/10.1007/s00122-018-3271-7

Jo H., Ha B.K., Park S.Y., Jeong S.C., Lee J.D., and Moon J.K., 2023, Genetic diversity of korean wild soybean core collections and genome-wide association study for days to flowering, Plants, 12(6): 1305.
https://doi.org/10.3390/plants12061305

Kaga A., Shimizu T., Watanabe S., Tsubokura Y., Katayose Y., Harada K., Vaughan D.A., and Tomooka N., 2012, Evaluation of soybean germplasm conserved in NIAS genebank and development of mini core collections, Breeding Science, 61(5): 566-592.
https://doi.org/10.1270/jsbbs.61.566

Kim E., Shin M., Wang X., Choi Y., Lee G., Yoo E., Lee J., Lee S., Desta K.T., Kim M., Oh H., and Yi J., 2025, Integrative genome-wide association and haplotype-based analyses reveal genetic structure and local adaptation in Korean landrace soybeans, BMC Plant Biology, 25(1): 79.
https://doi.org/10.1186/s12870-025-07479-6

Kim J.H., Kim K.S., Jung J.H., Kang B.K., Lee J.D., Ha B.K., and Kang S.T., 2020, Validation of marker-assisted selection in soybean breeding program for pod shattering resistance, Euphytica, 216(12): 183.
https://doi.org/10.1007/s10681-020-02703-w

Kim J., Lee J., Kim D., Lyu J., Baek J., Ha B.K., and Kwon S.J., 2025, Image-based GWAS identifies the genetic architecture of seed-related traits in a soybean mutant population, Molecular Breeding, 45(4): 84.
https://doi.org/10.1007/s11032-025-01584-y

Kim S., Tayade R., Kang B.K., Hahn B.S., Ha B.K., and Kim Y.J., 2023, Genome-wide association studies of seven root traits in soybean (*Glycine max* L.) landraces, International Journal of Molecular Sciences, 24(1): 873.
https://doi.org/10.3390/ijms24010873

Kim W., Kang B.K., Moon C., Kang S.T., Shin S.O., Chowdhury S., Jeong S.C., Choi M.S., Park S.Y., Moon J.K., and Ha B.K., 2023, Genome-wide association study for agronomic traits in wild soybean (*Glycine soja*), Agronomy, 13(3): 739.
https://doi.org/10.3390/agronomy13030739

Kofsky J., Zhang H., and Song B.H., 2018, The untapped genetic reservoir: the past, current, and future applications of the wild soybean (Glycine soja), Frontiers in Plant Science, 9: 949.
https://doi.org/10.3389/fpls.2018.00949

Kumar R., Das S., Choudhury B., Kumar A., Prakash N., Verma R.L., Chakraborti M., Devi A.P., Bhattacharjee B., Das R., Das B., Devi H.L., Das B., Rawat S., and Mishra V.K., 2024, Advances in genomic tools for plant breeding: harnessing DNA molecular markers, genomic selection, and genome editing, Biological Research, 57(1): 62.
https://doi.org/10.1186/s40659-024-00562-6

Kumar S., Susmita C., Sripathy K.V., Agarwal D.K., Pal G., Singh A.K., Kumar S., Rai A., and Simal-Gándara J., 2022, Molecular characterization and genetic diversity studies of Indian soybean (*Glycine max* (L.) Merr.) cultivars using SSR markers, Molecular Biology Reports, 49(3): 2129-2140.
https://doi.org/10.1007/s11033-021-07030-4

Lee Y.G., Jeong N., Kim J.H., Lee K., Kim K.H., Pirani A., Ha B.K., Kang S.T., Park B.S., Moon J.K., Kim N., and Jeong S.C., 2015, Development, validation and genetic analysis of a large soybean SNP genotyping array, The Plant Journal, 81(4): 625-636.
https://doi.org/10.1111/tpj.12755

Li D., Zhang Z., Gao X., Zhang H., Bai D., Wang Q., Zheng T., Li Y., and Qiu L., 2023, The elite variations in germplasms for soybean breeding, Molecular Breeding, 43(3): 18.
https://doi.org/10.1007/s11032-023-01378-0

Li F., Sayama T., Yokota Y., Hiraga S., Hashiguchi M., Tanaka H., Akashi R., and Ishimoto M., 2024, Assessing genetic diversity and geographical differentiation in a global collection of wild soybean (Glycine soja Sieb. et Zucc.) and assigning a mini-core collection, DNA Research, 31(2): dsae009.
https://doi.org/10.1093/dnares/dsae009

Li Y., Li Y., Su S., Reif J.C., Qi Z., Wang X., Wang X., Tian Y., Li D., Sun R., Liu Z., Xu Z., Fu G., Ji Y., Chen Q., Liu J., and Qiu L., 2021, The SoySNP618K array: A high-resolution SNP platform as a valuable genomic resource for soybean genetics and breeding, Journal of Integrative Plant Biology, 64(2): 440-454.
https://doi.org/10.1111/jipb.13202

Qiu L.J., Yinghui L., Gao R.X., Zhangxiong L., Lixia W., and Ru-Zhen C., 2009, Establishment, representative testing and research progress of soybean core collection and mini core collection, Acta Agronomica Sinica, 35(3): 571-579.

Liu Z., Li H., Wen Z., Fan X., Li Y., Guan R., Guo Y., Wang S., Wang D., and Qiu L., 2017, Comparison of genetic diversity between chinese and american soybean (Glycine max (L.) accessions revealed by high-density SNPs, Frontiers in Plant Science, 8: 2014.
https://doi.org/10.3389/fpls.2017.02014

Lukanda M., Dramadri I.O., Adjei E.A., Arusei P., Gitonga H., Wasswa P., Edema R., Ssemakula M., Tukamuhabwa P., and Tusiime G., 2023, Genetic diversity and population structure of ugandan soybean (*Glycine max* L.) germplasm based on DArTseq, Plant Molecular Biology Reporter, 42(1): 1-16.
https://doi.org/10.21203/rs.3.rs-1689218/v1

Mandozai A., Moussa A.A., Zhang Q., Qu J., Du Y., Anwari G., Amin N., and Wang P., 2021, Genome-wide association study of root and shoot related traits in spring soybean (*Glycine max* L.) at seedling stages using SLAF-Seq, Frontiers in Plant Science, 12: 568995.
https://doi.org/10.3389/fpls.2021.568995

Meng J., Yang G., Li X., Zhao Y., and He S., 2023, Population structure of wild soybean (Glycine soja) based on SLAF-seq have implications for its conservation, PeerJ, 11: e16415.
https://doi.org/10.7717/peerj.16415

Miller M.C., Song Q., Fallen B., and Li Z., 2023, Genomic prediction of optimal cross combinations to accelerate genetic improvement of soybean (*Glycine max*), Frontiers in Plant Science, 14: 1171135.
https://doi.org/10.3389/fpls.2023.1171135

Mitiku A., Gudina G., Yadeta B., and Abdu M., 2025, Comprehensive assessment of genetic variability, association analysis, and elucidation of direct and indirect effects of yield and yield contributing traits from diverse exogenous soybean (*Glycine max* L.) genotypes in jimma district, Southwest Ethiopia, Advances in Agriculture, 2025: 2865503.
https://doi.org/10.1155/aia/2865503

Nair R.M., Yan M.R., Vemula A., Rathore A., van Zonneveld M., and Schafleitner R., 2022, Development of core collections in soybean on the basis of seed size, Legume Science, 5(1): e158.
https://doi.org/10.1002/leg3.158

Nawaz M.A., Lin X., Chan T.F., Ham J.H., Shin T.S., Ercişli S., Golokhvast K.S., Lam H.M., and Chung G., 2020, Korean wild soybeans (Glycine soja Sieb and Zucc.): Geographic distribution and germplasm conservation, Agronomy, 10(2): 214.
https://doi.org/10.3390/agronomy10020214

Niu Y., Yung W.S., Sze C.C., Wong F.L., Li M.W., Chung G., and Lam H.M., 2024, Developing an SNP dataset for efficiently evaluating soybean germplasm resources using the genome sequencing data of 3,661 soybean accessions, BMC Genomics, 25(1): 328.
https://doi.org/10.1186/s12864-024-10382-3

Obua T., Sserumaga J., Opiyo S.O., Tukamuhabwa P., Odong T.L., Mutuku J.M., and Yao N., 2020, Genetic diversity and population structure analysis of tropical soybean (*Glycine max* (L.) Merrill) using single nucleotide polymorphic markers, Global Journal of Science Frontier Research, 20(6): 35-43.
https://doi.org/10.34257/GJSFRDVOL20IS6PG35

Oliveira M.F., Nelson R.L., Geraldi I.O., Cruz C.D., and Toledo J.F.F., 2010, Establishing a soybean germplasm core collection, Field Crops Research, 119(2-3): 277-289.
https://doi.org/10.1016/j.fcr.2010.07.021

Patil A.B., Oak M.D., Gijare S.J., Gobade A.N., Jaybhay S.A., Surve V.H., Salunkhe D.K., Waghmare B.M., Idhol B.B., Patil R.M., and Pawar D.P., 2025, Genome-wide exploration of soybean domestication traits: integrating association mapping and SNP×SNP interaction analyses, Plant Molecular Biology, 115(1-2): 83-99.
https://doi.org/10.1007/s11103-025-01583-9

Patil G., Tuyen D.D., Vuong T.D., Valliyodan B., Lee J.D., Chaudhary J., Shannon J.G., and Nguyen H.T., 2016, Genomic-assisted haplotype analysis and the development of high-throughput SNP markers for salinity tolerance in soybean, Scientific Reports, 6: 19199.
https://doi.org/10.1038/srep19199

Perić V., Kravić N., Tabaković M., Drinić M., Nikolić V., Simić M., and Nikolić A., 2025, Depicting sOYBEAN dIVERSITY VIA cOMPLEMENTARY aPPLICATION OF tHREE mARKER tYPes, Plants, 14(2): 201.
https://doi.org/10.3390/plants14020201

Potapova N.V., Zlobin A.E., Perfil'ev R.I., Vasiliev G.V., Salina E.A., and Tsepilov Y.A., 2023, Population structure and genetic diversity of the 175 soybean breeding lines and varieties cultivated in west siberia and other regions of Russia, Plants, 12(19): 3490.
https://doi.org/10.3390/plants12193490

Qin J., Wang F., Zhao Q., Shi A., Zhao T., Song Q., Ravelombola W., An H., Yan L., Yang C., and Zhang M., 2022, Identification of candidate genes and genomic selection for seed protein in soybean breeding pipeline, Frontiers in Plant Science, 13: 882732.
https://doi.org/10.3389/fpls.2022.882732

Qiu L.J., Xing L.L., Guo Y., Wang J., Jackson S.A., and Chang R.Z., 2013, A platform for soybean molecular breeding: the utilization of core collections for food security, Plant Molecular Biology, 83(1-2): 41-50.
https://doi.org/10.1007/s11103-013-0076-6

Ramesh P., Mallikarjuna G., Sameena S., Kumar A., Gurulakshmi K., Reddy B.O., Reddy P.C.O., and Sekhar A.C., 2020, Advancements in molecular marker technologies and their applications in diversity studies, Journal of Biosciences, 45: 128.
https://doi.org/10.1007/s12038-020-00089-4

Rani R., Raza G., Tung M.H., Rizwan M., Ashfaq H., Shimelis H., Razzaq M.K., and Arif M., 2023, Genetic diversity and population structure analysis in cultivated soybean (*Glycine max* [L.] Merr.) using SSR and EST-SSR markers, PLOS ONE, 18(6): e0286099.
https://doi.org/10.1371/journal.pone.0286099

Ravelombola W., Qin J., Shi A., Song Q., Yuan J., Wang F., Chen P., Yan L., Feng Y., Zhao T., Meng Y., Guan K., Yang C., and Zhang M., 2021, Genome-wide association study and genomic selection for yield and related traits in soybean, PLOS ONE, 16(8): e0255761.
https://doi.org/10.1371/journal.pone.0255761

Riaz A., Raza Q., Kumar A., Dean D., Chiwina K., Phiri T., Thomas J., and Shi A., 2023, GWAS and genomic selection for marker-assisted development of sucrose enriched soybean cultivars, Euphytica, 219(8): 117.
https://doi.org/10.1007/s10681-023-03224-y

S. O., T. A., D. A., and C. A., 2025, Studies on Character Expression for yield Components in Soybean, Environmental Reports, 7(2): 159.
https://doi.org/10.51470/ER.2025.7.2.159

Satyawan D., and Tasma I.M., 2021, Identification of prospective soybean accessions for the creation of a genebank core collection based on high density DNA marker data, IOP Conference Series: Earth and Environmental Science, 762(1): 012069.
https://doi.org/10.1088/1755-1315/762/1/012069

Shaibu A.S., Ibrahim H.Y., Miko S., Mohammed I.S., Mohammed S.G., Yusuf H., Kamara A.Y., Omoigui L.O., and Karikari B., 2022, Assessment of the genetic structure and diversity of soybean (*Glycine max* L.) germplasm using diversity array technology and single nucleotide polymorphism markers, Plants, 11(1): 68.
https://doi.org/10.3390/plants11010068

Sharmin R., Karikari B., Chang F., Amin M.N.G., Bhuiyan M.S.R., Hina A., Lv W., Zhao C., Begum N., and Zhao T., 2021, Genome-wide association study uncovers major genetic loci associated with seed flooding tolerance in soybean, BMC Plant Biology, 21(1): 497.
https://doi.org/10.1186/s12870-021-03268-z

Sonah H., O'Donoughue L., Cober E., Rajcan I., and Belzile F., 2015, Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean, Plant Biotechnology Journal, 13(2): 211-221.
https://doi.org/10.1111/pbi.12249

Song Q., Hyten D.L., Jia G., Quigley C.V., Fickus E.W., Nelson R.L., and Cregan P.B., 2015, Fingerprinting soybean germplasm and its utility in genomic research, G3: Genes Genomes Genetics, 5(10): 1999-2006.
https://doi.org/10.1534/g3.115.019000

Song Q., Quigley C., He R., Wang D., Nguyen H.T., Miranda C., and Li Z., 2024, Development and implementation of nested single-nucleotide polymorphism (SNP) assays for breeding and genetic research applications, The Plant Genome, 17(2): e20491.
https://doi.org/10.1002/tpg2.20491

Song Q., Yan L., Quigley C., Fickus E., Wei H., Chen L., Dong F., Araya S., Liu J., Hyten D., Pantalone V., and Nelson R., 2020, Soybean BARCSoySNP6K: An assay for soybean genetics and breeding research, The Plant Journal, 104(3): 800-811.
https://doi.org/10.1111/tpj.14960

Stewart-Brown B.B., Song Q., Vaughn J.N., and Li Z., 2019, Genomic selection for yield and seed composition traits within an applied soybean breeding program, G3: Genes Genomes Genetics, 9(7): 2253-2265.
https://doi.org/10.1534/g3.118.200917

Suela M., Azevedo C.F., Nascimento A.C.C., Morota G., da Silva F.F., Malone G., Giasson N., and Nascimento M., 2025, Using structural equation models to interpret genome-wide association studies for morphological and productive traits in soybean [*Glycine max* (L.) Merr.], Plants, 14(19): 3015.
https://doi.org/10.3390/plants14193015

T. N., S. R., and L. R., 2022, Population structure and genetic diversity characterization of soybean for seed longevity, PLOS ONE, 17(12): e0278631.
https://doi.org/10.1371/journal.pone.0278631

Tsindi A., Eleblu J.S.Y., Gasura E., Mushoriwa H., Tongoona P., Danquah E.Y., Mwadzingeni L., Zikhali M., Ziramba E., Mabuyaye G.T., and Derera J., 2023, Analysis of population structure and genetic diversity in a Southern African soybean collection based on single nucleotide polymorphism markers, CABI Agriculture and Bioscience, 4(1): 58.
https://doi.org/10.1186/s43170-023-00158-2

Ullah A., Akram Z., Malik S.I., and Khan K.S., 2021, Assessment of phenotypic and molecular diversity in soybean [*Glycine max* (L.) Merr.] germplasm using morpho-biochemical attributes and SSR markers, Genetic Resources and Crop Evolution, 68(7): 2827-2847.
https://doi.org/10.1007/s10722-021-01157-w

Valliyodan B., Brown A.V., Wang J., Patil G., Liu Y., Otyama P.I., Nelson R.T., Vuong T.D., Song Q., Musket T.A., Wagner R., Marri P.R., Reddy S.K., Sessions A., Wu X., Grant D., Bayer P.E., Roorkiwal M., Varshney R.K., Liu X., Edwards D., Xu D., Joshi T., Cannon S.B., and Nguyen H.T., 2021, Genetic variation among 481 diverse soybean accessions, inferred from genomic re-sequencing, Scientific Data, 8(1): 50.
https://doi.org/10.1038/s41597-021-00834-w

Viana J.P.G., Fang Y., Avalos A.E., Song Q., Nelson R.L., and Hudson M.E., 2022, Impact of multiple selective breeding programs on genetic diversity in soybean germplasm, Theoretical and Applied Genetics, 135(5): 1591-1602.
https://doi.org/10.1007/s00122-022-04056-5

Wibisono K., Dyah R., Utari R., Suparjo S., Umar U., Rijzaani H., Hakim L., Suhendar A., Purwanto O., Satyawan D., Witjaksono W., Mastur M., Lestari P., and Tasma I.M., 2025, Genetic diversity and DNA barcoding construction of tropical soybean advanced lines based on SSR markers, Jurnal Ilmu Pertanian Indonesia, 30(2): 293-302.
https://doi.org/10.18343/jipi.30.2.293

Xue Y., Tang X., Zhu X., Zhang R., Yao Y., Cao D., He W., Liu Q., Luan X., Shu Y., and Liu X., 2025, Leveraging GWAS-identified markers in combination with bayesian and machine learning models to improve genomic selection in soybean, International Journal of Molecular Sciences, 26(19): 9586.
https://doi.org/10.3390/ijms26199586

Yang Q., Zhang J., Shi X., Chen L., Qin J., Zhang M., Yang C., Song Q., and Yan L., 2023, Development of SNP marker panels for genotyping by target sequencing (GBTS) and its application in soybean, Molecular Breeding, 43(2): 26.
https://doi.org/10.1007/s11032-023-01372-6

Zatybekov A., Genievskaya Y., Fang C., Abugalieva S., and Turuspekov Y., 2025, Uncovering the genetic landscape of soybean accessions from Kazakhstan in comparison with global germplasm using whole genome resequencing, BMC Genomics, 26(1): 248.
https://doi.org/10.1186/s12864-025-12024-8

Zatybekov A., Yermagambetova M., Genievskaya Y., Didorenko S., and Abugalieva S., 2023, Genetic diversity analysis of soybean collection using simple sequence repeat markers, Plants, 12(19): 3445.
https://doi.org/10.3390/plants12193445

Zhang J., Song Q., Cregan P.B., and Jiang G.L., 2016, Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*), Theoretical and Applied Genetics, 129(1): 117-130.
https://doi.org/10.1007/s00122-015-2614-x

Zhao T., Wang F., Qi J., Chen Q., Zhu L., Liu L., Yan L., Chen Y., Yang C., and Qin J., 2025, Genome-wide association analysis study and genomic prediction for resistance to soybean mosaic virus in soybean population, BMC Plant Biology, 25(1): 775.
https://doi.org/10.1186/s12870-025-06775-5