Legume Genomics and Genetics 2024, Vol.15, No.3, 126-139
http://cropscipublisher.com/index.php/lgg

Review and Progress Open Access

Genomic Tools in Soybean Breeding: Innovations and Impacts

Xiaoxi Zhou, Tianxia Guo
Institute of Life Science, Jiyang College of Zhejiang A&F University, Zhuji, 311800, Zhejiang, China
Corresponding email: tianxia.guo@cuixi.org

Legume Genomics and Genetics, 2024 Vol.15, No.3 doi: 10.5376/lgg.2024.15.0014
Received: 08 May, 2024
Accepted: 10 Jun., 2024
Published: 21 Jun., 2024

Copyright © 2024 Zhou and Guo, 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:
Zhou X.X., and Guo T.X., 2024, Genomic tools in soybean breeding: innovations and impacts, Legume Genomics and Genetics, 15(3): 126-139 (doi: 10.5376/lgg.2024.15.0014)

Abstract This study explores the application and progress of genomic technologies in soybean breeding. As a crucial source of protein and oil globally, soybean breeding methods have gradually shifted from traditional phenotypic selection and hybridization techniques to reliance on genomic technologies. Modern genomic tools, such as marker-assisted selection (MAS), genomic selection (GS), and CRISPR/Cas9 gene editing, have significantly improved breeding efficiency and accuracy. These tools accelerate the development of superior cultivars by predicting the genetic potential of breeding lines and utilizing a broader genetic base to introduce more beneficial traits. The study reviews the historical development of soybean breeding, highlighting the limitations of traditional methods, such as a narrow genetic base and slow breeding cycles. Genomic tools show great potential in enhancing yield, quality, disease resistance, and stress tolerance. For example, genomic selection predicts traits using genome-wide molecular markers, reducing dependence on phenotypic evaluation. Marker-assisted selection uses specific DNA markers for precise trait selection, and CRISPR/Cas9 gene editing allows for precise modifications of specific genes, enhancing soybean disease resistance and stress tolerance.

Keywords Genomic tools; Soybean; Breeding; Marker-assisted selection (MAS); CRISPR/Cas9

1 Introduction

Soybean (*Glycine max*) is a crucial crop globally, serving as a primary source of protein and oil. It is instrumental in various industries, including food, animal feed, and biofuel production. Traditional soybean breeding has relied on phenotypic selection and crossbreeding techniques, aiming to improve yield, disease resistance, and environmental adaptability. Despite its success, traditional breeding is time-consuming and limited by the genetic variation present in the parental lines. Advancements in genetic and genomic technologies have revolutionized soybean breeding. These innovations have enabled breeders to harness a broader genetic base, improve selection accuracy, and accelerate the development of superior cultivars. Modern genomic tools, such as marker-assisted selection (MAS), genomic selection (GS), and genome editing, have emerged as powerful methodologies in enhancing soybean breeding efficiency and effectiveness (Du et al., 2022).

The integration of genomic tools in agriculture marks a paradigm shift, offering unprecedented opportunities to address global food security challenges. Genomic tools facilitate the identification of key genetic traits, allowing for precise manipulation and enhancement of crop characteristics. In soybean breeding, these tools have been pivotal in achieving significant improvements in yield, quality, and resistance to biotic and abiotic stresses. Genomic tools provide several advantages. They increase breeding efficiency by accelerating the breeding cycle through predicting the performance of breeding lines based on their genetic makeup rather than waiting for phenotypic expression. Enhanced precision is achieved as molecular markers and genome editing technologies enable targeted modifications, reducing the uncertainty associated with traditional breeding (Valliyodan et al., 2016). Furthermore, greater genetic diversity can be exploited within and between wild and cultivated soybean species, introducing beneficial traits into the breeding pool. Additionally, by improving resistance to pests, diseases, and environmental stresses, genomic tools contribute to more sustainable agricultural practices, reducing the reliance on chemical inputs.

This study aims to provide a comprehensive overview of the innovations and impacts of genomic tools in soybean breeding. The specific objective is to examine the historical context and evolution of soybean breeding practices, highlighting the transition from traditional methods to the adoption of genomic technologies; To explore the various genomic tools employed in soybean breeding, including marker-assisted selection, genomic selection, and genome editing, and their specific applications; To assess the impact of these genomic innovations on soybean breeding outcomes, such as yield improvement, disease resistance, and environmental adaptability; To discuss future prospects and challenges in the integration of genomic tools in soybean breeding, identifying potential areas for further research and development. By addressing these objectives, the study aims to elucidate the transformative role of genomic tools in soybean breeding and their broader implications for modern agriculture.

2 Historical Background of Soybean Breeding

2.1 Traditional breeding methods

Traditional soybean breeding methods have their roots in China, where soybean (*Glycine max*) originated over 5 000 years ago. The early domestication and artificial selection practices focused on selecting plants with desirable traits such as higher yield, disease resistance, and better adaptability to local climates. These methods predominantly involved mass selection and simple crossbreeding techniques to improve soybean varieties. The breeding process was labor-intensive and time-consuming, relying heavily on phenotypic selection and natural variation within the soybean population.

The use of hybridization techniques, where two different parent plants are crossbred to combine desirable traits, became more refined over time. The Single Seed Descent (SSD) method emerged as a popular technique, where only one seed from each plant is selected for the next generation. This method allowed for more controlled and efficient breeding processes, reducing space and labor requirements while maintaining genetic diversity. Despite these advancements, traditional breeding methods had limitations. For instance, they often resulted in a narrow genetic base, making crops more susceptible to diseases and environmental changes. The slow pace of breeding cycles also hindered rapid improvements and adaptability (Zhang et al., 2021). Furthermore, traditional methods were less effective in incorporating complex traits such as stress resistance and nutritional improvements.

2.2 Milestones in soybean breeding

Soybean breeding has achieved significant milestones that have shaped its global prominence. Early domestication and dissemination of soybean began in China and spread to Korea, Japan, and later to the rest of the world, setting the foundation for modern soybean cultivation. Early efforts focused on adapting the crop to different climatic conditions. Soybean was introduced to North America in the 18th century, with significant cultivation beginning in the early 20th century. This period saw the development of varieties suited to the American Midwest, leveraging the region's growing conditions. The completion of the soybean genome sequencing in 2010 marked a new era in soybean breeding. This milestone facilitated advances in genomic selection, marker-assisted breeding, and the development of transgenic soybean varieties with enhanced traits such as herbicide tolerance and pest resistance.

2.3 Limitations of conventional breeding techniques

Despite the progress made through traditional breeding methods, several limitations persist. Conventional breeding is inherently slow, often requiring several generations to achieve desirable traits due to the reliance on natural genetic variation and phenotypic selection. Traditional breeding often suffers from a narrow genetic base, which can lead to reduced genetic diversity and increased vulnerability to diseases and pests. This limitation has been particularly noted in the genetic bases of soybean cultivars in both the U.S. and China. Conventional methods are less adaptable to rapidly changing environmental conditions, such as climate change, which can adversely affect crop yields and resilience. Additionally, traditional breeding requires significant manual effort in selecting and crossbreeding plants, which can be resource-intensive and impractical for large-scale breeding programs (Silva et al., 2017).

3 Genomic Tools and Technologies in Soybean Breeding

3.1 Genomic selection (GS)

Genomic Selection (GS) is an advanced breeding technique that employs genome-wide molecular markers to predict the genetic potential of plants for various traits, thus accelerating the breeding process. This method is particularly valuable for improving complex traits such as yield, disease resistance, and seed quality, which are controlled by multiple genes. In soybean breeding, GS has been shown to significantly enhance the accuracy of selection and the speed of breeding cycles. For example, a study demonstrated that using GS with a medium density of markers and a genomic best linear unbiased prediction (GBLUP) model could predict up to 39% of the phenotypic variation in yield among soybean lines (Duhnen et al., 2017).

Another study highlighted the potential of GS to achieve predictive accuracies of 0.81 for protein content, 0.71 for oil content, and 0.26 for yield, demonstrating its effectiveness in soybean breeding programs (Stewart-Brown et al., 2019). The use of GS reduces the reliance on extensive phenotypic evaluations, which are time-consuming and labor-intensive, thereby streamlining the breeding process and enabling the rapid development of superior soybean varieties.

3.2 Marker-assisted selection (MAS)

Marker-Assisted Selection (MAS) uses specific DNA markers linked to desirable traits to guide the selection process in breeding programs. This technique is particularly effective for traits that are difficult to assess through traditional phenotypic methods, such as resistance to diseases and pests, as well as tolerance to environmental stresses. In soybean breeding, MAS has been instrumental in improving resistance to key diseases like Phytophthora Root Rot and Soybean Cyst Nematode. For instance, MAS has been shown to be highly efficient in selecting soybean lines with improved resistance to these diseases, achieving high prediction accuracies and reducing the need for labor-intensive phenotypic evaluations.

Additionally, MAS has been utilized to enhance pod shattering resistance in soybeans, with a study demonstrating prediction accuracies of up to 96% using specific markers (Kim et al., 2020). The integration of MAS into soybean breeding programs has led to the development of more robust and resilient soybean varieties, capable of withstanding various biotic and abiotic stresses, thereby contributing to higher yields and improved crop performance.

3.3 Genome-wide association studies (GWAS)

Genome-Wide Association Studies (GWAS) are a powerful tool used to identify associations between genetic variations and traits across the entire genome. GWAS has been extensively applied in soybean breeding to uncover the genetic basis of complex traits such as yield, maturity, plant height, and seed composition. For example, a study involving 250 soybean accessions identified significant single nucleotide polymorphisms (SNPs) associated with traits like grain yield, plant height, and seed weight. This study used a Bayesian Information and Linkage Disequilibrium Iteratively Nested Keyway (BLINK) model to perform GWAS and found numerous SNPs linked to these important agronomic traits (Figure 1) (Ravelombola et al., 2020).

The identification of these SNPs allows breeders to implement marker-assisted selection and genomic selection strategies more effectively. Moreover, GWAS has facilitated the discovery of novel alleles and the refinement of previously known loci, which are crucial for improving soybean traits. The integration of GWAS findings into breeding programs accelerates the development of high-yielding, disease-resistant, and stress-tolerant soybean varieties, enhancing the overall efficiency and effectiveness of soybean breeding efforts.

3.4 Quantitative trait loci (QTL) mapping

Quantitative Trait Loci (QTL) mapping is a technique used to identify regions of the genome that are associated with specific quantitative traits. In soybean breeding, QTL mapping has been instrumental in identifying genomic regions linked to important traits such as yield, seed size, and disease resistance. For instance, a study using chromosome segment substitution lines (CSSLs) developed from a cross between wild and cultivated soybeans

identified several QTLs associated with 100-seed weight, a critical factor for yield improvement. The study mapped 12 main effect QTLs and analyzed epistatic interactions between these loci, providing valuable insights for marker-assisted selection and breeding strategies. QTL mapping not only helps in understanding the genetic architecture of complex traits but also aids in the development of molecular markers for use in MAS. By incorporating QTL mapping results, breeders can more precisely target and select for beneficial traits, leading to the development of superior soybean varieties with enhanced performance and resilience.

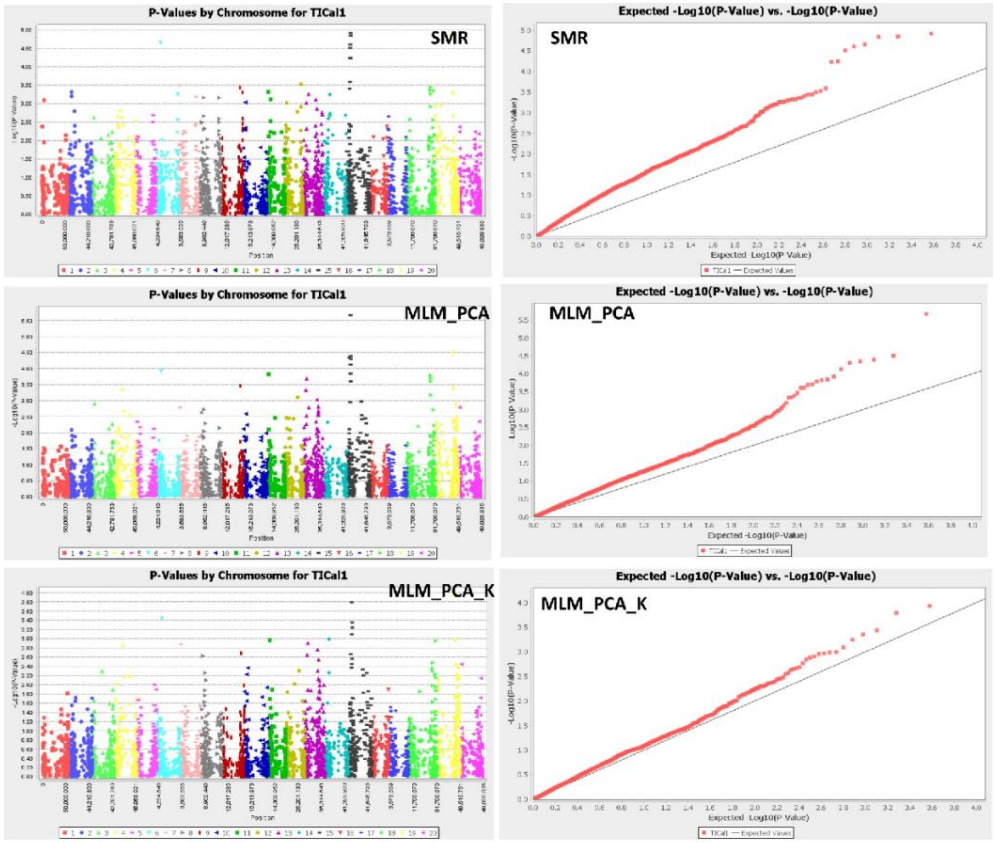

Figure 1 Manhattan plots and QQ-plots for tolerance indexes based on biomass reduction under SCN infestation (Adopted from Ravelombola et al., 2020)

Image caption: Figure 1 displays Manhattan plots and QQ plots for tolerance indexes based on biomass reduction under SCN (soybean cyst nematode) infestation, using different statistical models (SMR, MLM_PCA, and MLM_PCA_K). In each Manhattan plot, the x-axis represents chromosome numbers, and the y-axis denotes the LOD value (-log10 (p-value)). Different colors represent different chromosomes. In the QQ plots, the x-axis represents the expected -log10 (p-value), and the y-axis shows the observed -log10 (p-value). Plot A shows the results from the single marker regression model (SMR), Plot B shows the results from the generalized linear model (GLM(PCA)), and Plot C shows the results from the mixed linear model (MLM (PCA+K)). The QQ plots are used to assess the deviation of p-value distributions, with a linear distribution indicating a better model fit (Adapted from Ravelombola et al., 2020)

3.5 Genotyping-by-sequencing (GBS)

Genotyping-by-Sequencing (GBS) is a cost-effective and high-throughput method for generating large amounts of genetic data. This technique involves sequencing a subset of the genome, allowing for the identification of numerous genetic markers across the genome. GBS has been widely used in soybean breeding to identify SNPs and other genetic variations associated with important traits. For example, a study using GBS on a soybean population identified over 10 000 high-quality SNPs and mapped several genomic regions associated with traits such as yield, maturity, and seed weight. The study demonstrated the utility of GBS in enhancing the accuracy of genomic selection and marker-assisted selection in soybean breeding programs (Ravelombola et al., 2021). The high density of markers generated by GBS enables detailed genetic analyses and the identification of marker-trait associations, facilitating the rapid and efficient selection of superior soybean lines.

3.6 CRISPR/Cas9 and other gene editing technologies

CRISPR/Cas9 and other gene editing technologies have revolutionized plant breeding by enabling precise modifications to the genome. In soybean breeding, CRISPR/Cas9 has been used to introduce targeted mutations and edit genes associated with important traits such as yield, disease resistance, and stress tolerance. For instance, a study demonstrated the successful use of CRISPR/Cas9 to generate soybean lines with targeted insertions and deletions, achieving mutation rates of up to 90% at the target site. This technology allows for the rapid development of new soybean varieties with enhanced traits without the need for traditional breeding cycles. The ability to precisely edit specific genes makes CRISPR/Cas9 a powerful tool for soybean breeding, enabling the development of varieties that can meet the demands of changing environmental conditions and market needs.

4 Applications of Genomic Tools in Soybean Breeding

4.1 Enhancing yield and quality

Genomic tools have revolutionized soybean breeding by enabling more precise and efficient enhancement of yield and quality traits. Techniques like genomic selection (GS), genome-wide association studies (GWAS), and marker-assisted selection (MAS) have significantly improved the selection process for high-yielding and high-quality soybean varieties. For instance, GS has shown high predictive accuracies for yield and seed protein content, making it a valuable tool for soybean breeders. Studies have demonstrated that GS can explain up to 39% of the phenotypic variation in yield and 32% in seed protein content, providing a robust framework for selecting superior genotypes (Figure 2) (Duhnen et al., 2017).

Furthermore, the use of CRISPR/Cas9 has opened new avenues for improving soybean traits by allowing precise gene editing to enhance yield and quality traits. This technology has been used to develop soybean mutants with better yield and quality, demonstrating its potential to revolutionize soybean breeding.

4.2 Improving resistance to pests and diseases

Improving resistance to pests and diseases is a critical application of genomic tools in soybean breeding. Genomic approaches like GWAS, QTL mapping, and genome editing have been employed to identify and utilize resistance genes. For instance, the integration of genetic analysis, molecular biology, and genomic approaches has drastically enhanced our understanding of genetic control of nematode resistance. This has led to the identification of major resistance loci, such as Rhg1 and Rhg4, which are crucial for breeding nematode-resistant soybean varieties (Kim et al., 2016). Additionally, genome editing tools like CRISPR/Cas9 have been employed to develop soybean varieties with enhanced resistance to pathogens by targeting specific genes involved in disease susceptibility and resistance (Yin and Qiu, 2019).

4.3 Abiotic stress tolerance (drought, heat, salinity)

Abiotic stress tolerance is a critical focus in soybean breeding due to the adverse effects of environmental stresses such as drought, heat, and salinity on crop yield. Genomic tools have greatly enhanced our ability to identify and incorporate genes associated with stress tolerance into soybean breeding programs. For instance, quantitative trait loci (QTL) mapping and genome-wide association studies (GWAS) have identified key genomic regions linked to drought tolerance, enabling the development of soybean varieties that can thrive under water-limited conditions. Studies have identified several transcription factors (TFs) like DREBs, ERFs, ZIP, WRKY, and MYB that play significant roles in conferring tolerance to drought, heat, and salinity. The HD-Zip gene family, for example, has been found to impart tolerance to both drought and saline environments in soybeans.

CRISPR/Cas9 genome editing has been utilized to create mutations in genes such as *GmAITR*, resulting in enhanced salinity tolerance. Mutant soybean lines showed improved performance under saline conditions without detrimental fitness costs (Wang et al., 2021). Advances in genomic technologies and high-throughput phenotyping have facilitated the identification of drought-tolerant traits such as root system architecture and canopy wilting, contributing to the breeding of resilient soybean varieties (Valliyodan et al., 2016).

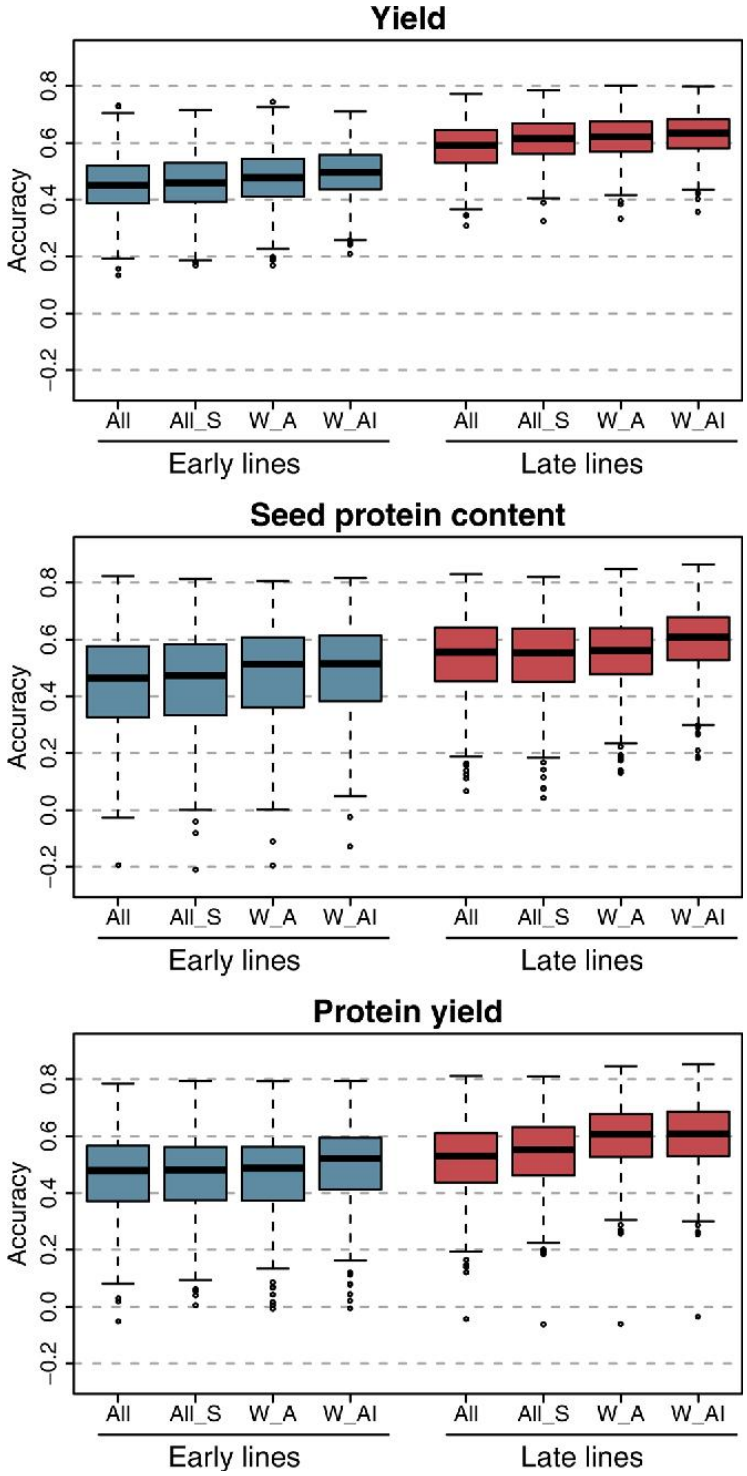

Figure 2 Boxplots of single-sample prediction accuracy of different genomic estimated best linear unbiased prediction (GBLUP) models over 500 samplings of test sets (Adopted from Duhnen et al., 2017)

Image caption: Figure 2 displays boxplots of single-sample prediction accuracy of different genomic estimated best linear unbiased prediction (GBLUP) models over 500 samplings of test sets. Each sampling consists of a random split of each subpopulation (early lines or late lines) into a 90% training set and a 10% test set. The models include: All and All_S, which are additive models calibrated on both training sets (early and late), assuming a homogeneous (All) or structured (All_S) population; W_A and W_AI, which are models calibrated on the training set of the subpopulation to be predicted, including additive effects alone (W_A) or together with additive-by-additive epistatic effects (W_AI). As a measure of single-sample prediction accuracy, we used Pearson's correlation between predicted genetic values and adjusted means. The plots show the prediction accuracy for three traits: yield, seed protein content, and protein yield, in both early and late lines (Adapted from Duhnen et al., 2017)

4.4 Nutritional and industrial trait improvement

Improving the nutritional and industrial traits of soybeans is essential for meeting the increasing global demand for high-quality food and feed. Genomic tools have been instrumental in enhancing traits such as oil content, protein quality, and fatty acid composition. For instance, genome-wide association studies (GWAS) have identified SNPs linked to high oil and protein content, facilitating the development of soybean varieties with enhanced nutritional profiles. Advances in omics technologies, including genomics, transcriptomics, and metabolomics, have provided comprehensive insights into the genetic basis of these traits. The integration of these technologies has enabled the identification of key regulatory genes and metabolic pathways involved in seed composition. For example, the GmFBL144 gene has been identified to interact with small heat shock proteins, playing a role in regulating drought stress tolerance and potentially influencing seed composition (Xu et al., 2022).

CRISPR/Cas9 has also been employed to create specific mutations that enhance oil quality and protein content in soybeans, demonstrating the potential of gene editing in nutritional improvement (Nagamine and Ezura, 2022). Additionally, high-throughput genotyping and phenotyping platforms have improved the precision and efficiency of selecting for desirable traits, further enhancing the potential for nutritional and industrial trait improvement in soybean breeding programs (Bhat and Yu, 2021).

4.5 Enhancing biodiversity and genetic variation

Enhancing biodiversity and genetic variation is fundamental for the sustainability and adaptability of soybean breeding programs. Genomic tools have enabled the identification and incorporation of novel alleles from diverse germplasm collections into cultivated soybean varieties. Pangenome analysis and GWAS have provided insights into the genetic diversity of soybean, revealing significant structural variations and uncovering beneficial alleles from wild soybean relatives. For example, a study involving the resequencing of 302 wild and cultivated soybean accessions identified genes related to domestication and improvement, highlighting the importance of genetic variation for future breeding efforts (Zhou et al., 2015).

The development of comprehensive genomic resources, such as SNP datasets and pangenomes, has facilitated the identification of alleles that can enhance traits like yield, disease resistance, and stress tolerance. By integrating these resources into breeding programs, breeders can develop soybean varieties with enhanced genetic diversity, ensuring their resilience and adaptability to changing environmental conditions (Petereit et al., 2022). Furthermore, the use of genomic selection and marker-assisted selection has accelerated the incorporation of diverse genetic material into breeding lines, promoting biodiversity and improving the overall genetic health of soybean crops.

5 Case Studies and Success Stories

5.1 Successful implementation of genomic selection (GS) in soybean breeding programs

Genomic selection (GS) has proven to be a powerful tool in enhancing the efficiency of soybean breeding programs, particularly for complex traits like yield and seed composition. A prominent example is the study conducted by Stewart-Brown et al. (2019), which utilized GS to improve yield and seed composition traits within an applied soybean breeding program. The study involved 483 elite breeding lines genotyped with BARCSoySNP6K iSelect BeadChips. Cross-validation using RR-BLUP revealed high predictive abilities for protein (0.81), oil (0.71), and yield (0.26) at the largest tested training set size. This demonstrates the potential of GS to enhance genetic gains in soybean breeding programs (Figure 3) (Stewart-Brown et al., 2019).

Genomic selection (GS) has been successfully implemented in soybean breeding programs, leading to significant improvements in genetic gain and efficiency. For example, a study on the implementation of GS in soybean demonstrated its potential to enhance breeding efficiency by predicting the genetic merit of soybean lines more accurately than traditional methods (Jin et al., 2023). Another study confirmed that GS could reduce the breeding cycle time, accelerating the development of high-yield and disease-resistant soybean varieties (Bernardo, 2016).

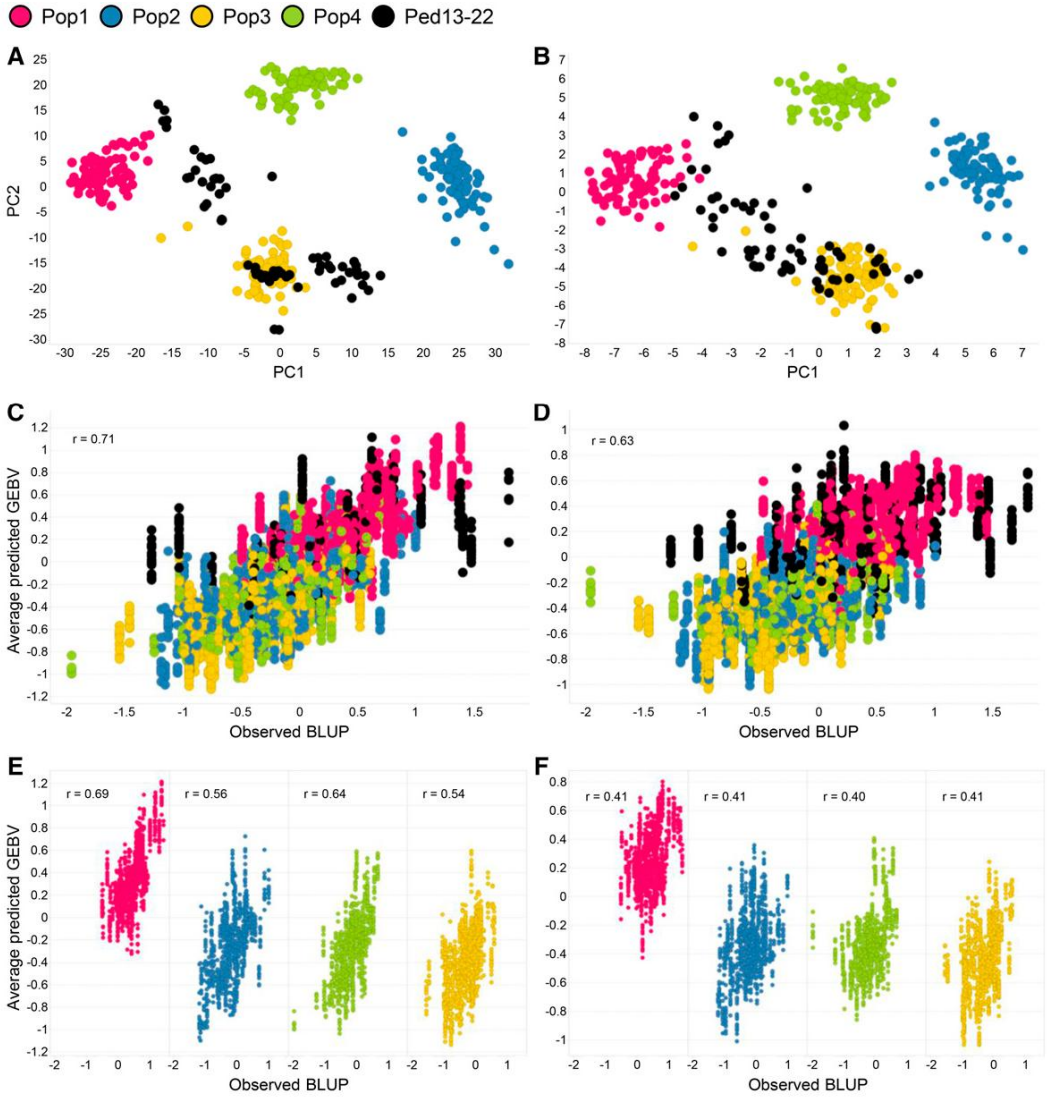

Figure 3 Effects of population structure on prediction of oil content when utilizing the entire genomic selection dataset (EGSD) method (Adopted from Stewart-Brown et al., 2019)

Image caption: Figure 3 shows the effects of population structure on the prediction of oil content when utilizing the entire genomic selection dataset (EGSD) method. Panel A displays the principal component analysis (PCA) of the genomic prediction population using all SNPs, while panel B shows the PCA using the 8th tag SNPs. Panel C presents the relationship between the average predicted genomic estimated breeding values (GEBV) and the observed best linear unbiased prediction (BLUP) values when using all SNPs, and panel D shows the similar relationship when using the 8th tag SNPs. Panel E illustrates the relationship between the average predicted GEBV and the observed BLUP values within Pop1-4 when using all SNPs, whereas panel F shows this relationship within Pop1-4 when using the 8th tag SNPs. The scatterplots in panels C-F display the correlation coefficients (Adapted from Stewart-Brown et al., 2019)

5.2 Achievements in marker-assisted selection (MAS) for disease resistance

Marker-assisted selection (MAS) has played a crucial role in developing disease-resistant soybean varieties. Research has identified several quantitative trait loci (QTLs) associated with resistance to various soybean diseases, enabling breeders to incorporate these resistance genes more effectively. For instance, MAS has been used to introgress resistance genes for soybean cyst nematode (SCN) and Phytophthora root rot, significantly improving soybean resilience against these pathogens (Huang et al., 2021). Additionally, MAS has facilitated the pyramiding of multiple resistance genes, providing durable resistance to soybean rust and other major diseases (Liu et al., 2017).

In combating the soybean cyst nematode (SCN), Espindola et al. (2016) utilized microsatellite markers Sat_141 and Sat_168 to identify SCN-resistant genotypes. This study demonstrated the effectiveness of MAS in selecting resistant plants, achieving high accuracy in genotypic versus phenotypic evaluations (Espindola et al., 2016). Dubiela et al. (2019) developed SNP markers associated with resistance to *Meloidogyne incognita*, a root-knot nematode. These markers were validated and used in MAS protocols to accurately identify resistant plants, showcasing the practical application of molecular markers in breeding programs.

5.3 Notable GWAS findings and their implications

Genome-wide association studies (GWAS) have identified numerous genetic loci associated with key agronomic traits in soybean, such as yield, oil content, and stress tolerance. For example, a GWAS identified loci associated with seed protein and oil content, which are crucial for soybean quality improvement (Cao et al., 2017). Another study highlighted the discovery of loci related to drought tolerance, providing insights into the genetic mechanisms underlying stress responses in soybean (Li et al., 2017). These findings have significant implications for soybean breeding, enabling the development of varieties with enhanced nutritional value and environmental resilience.

Che et al. (2020) conducted a GWAS for resistance to Soybean Mosaic Virus (SMV) and identified novel loci and candidate genes on chromosomes 8 and 20. These findings provided new genetic sources for MAS breeding programs aimed at developing SMV-resistant soybean varieties. Another notable study by Ravelombola et al. (2021) combined GWAS and GS to dissect the genetic basis of yield-related traits. They identified significant SNPs associated with yield, plant height, and seed weight, contributing valuable markers for breeding programs focused on yield improvement.

5.4 Advances in QTL mapping and utilization

Quantitative Trait Loci (QTL) mapping has been instrumental in identifying genetic regions associated with important traits in soybean. Pham et al. (2015) fine-mapped resistance genes to *Cercospora sojina*, the causal agent of frogeye leaf spot. This study narrowed down the resistance loci to specific genomic regions, facilitating the development of resistant cultivars through MAS.

Another significant advancement was made by Chandra et al. (2022), who reviewed the progress in genetic mapping for resistance to Phytophthora root and stem rot (PRSR). This review highlighted the identification of multiple Rps genes and partial resistance loci, contributing to the genetic improvement of soybean against PRSR. Moreover, Zhong et al. (2018) identified and fine-mapped a novel Phytophthora resistance gene, RpsHC18, using next-generation sequencing. This study developed diagnostic markers for MAS, enhancing the selection process for Phytophthora-resistant soybean cultivars.

5.5 Breakthroughs with gene editing in soybean

Gene editing technologies, particularly CRISPR/Cas9, have revolutionized soybean breeding by enabling precise modifications of the genome. Successful examples include the development of high oleic acid soybean varieties through CRISPR/Cas9-mediated editing of the fatty acid desaturase genes (Do et al., 2019). Additionally, gene editing has been used to enhance resistance to herbicides, facilitating more effective weed management in soybean cultivation (Wei et al., 2023). These breakthroughs demonstrate the potential of gene editing to accelerate the development of soybean varieties with improved traits and adaptability.

Li et al. (2017), who used CRISPR/Cas9 to enhance oleic acid content in soybean oil. This study showcased the potential of gene editing to improve the nutritional quality of soybean, aligning with health and market demands. In addition, Nagamine et al. (2020) successfully edited multiple genes simultaneously using CRISPR/Cas9, demonstrating the feasibility of multiplex gene editing in soybean. This approach significantly accelerates the development of varieties with stacked traits, enhancing overall crop performance.

6 Challenges and Limitations of Genomic Tools

6.1 Technical challenges

The implementation of genomic tools in soybean breeding faces several technical challenges. One significant issue is the complexity of the soybean genome, which includes numerous genes and regulatory elements that interact in intricate networks. The high genetic variability within soybean populations further complicates the accurate prediction of phenotypic traits from genotypic data. For example, the development of the SoyDNGP model highlighted the challenges associated with accurately predicting complex traits due to the high parameter volume and complexity of the genetic structure.

Another technical challenge is the need for high-quality and high-density marker data. The efficiency of genomic selection and marker-assisted selection depends on the availability of comprehensive SNP datasets and high-throughput genotyping platforms. However, generating such datasets can be resource-intensive and technically demanding. Additionally, integrating different types of omics data, such as genomics, transcriptomics, and proteomics, poses significant computational challenges. The analysis and interpretation of this vast amount of data require advanced bioinformatics tools and expertise (Li et al., 2016).

6.2 Economic and resource constraints

The economic and resource constraints associated with the use of genomic tools in soybean breeding are significant barriers to their widespread adoption. The initial costs of setting up high-throughput genotyping and phenotyping facilities are substantial, and ongoing expenses for reagents, equipment maintenance, and skilled personnel can be prohibitive for many breeding programs. Furthermore, the cost of whole-genome sequencing and genotyping by sequencing, although decreasing, remains high, especially when large populations need to be analyzed (Bhat and Yu, 2021).

Another economic challenge is the need for substantial investment in computational infrastructure and bioinformatics expertise to handle and analyze large genomic datasets. Many breeding programs, particularly in developing countries, lack access to the necessary computational resources and trained personnel. This disparity can lead to uneven progress in the adoption and implementation of genomic tools across different regions and breeding programs (Belzile et al., 2022).

6.3 Regulatory and ethical considerations

The use of genomic tools, especially gene editing technologies like CRISPR/Cas9, raises several regulatory and ethical considerations. The regulatory landscape for genetically modified organisms (GMOs) varies widely across different countries, with some having stringent regulations that can hinder the development and commercialization of genetically edited soybean varieties. Navigating these regulatory frameworks can be complex and time-consuming, requiring significant legal and administrative expertise (Nagamine and Ezura, 2022).

Ethical concerns also arise regarding the potential environmental impacts of releasing genetically edited crops and the broader implications for biodiversity. The possibility of unintended off-target effects and gene flow to wild relatives are major concerns that need to be addressed through rigorous risk assessment and management strategies. Additionally, there is ongoing debate about the societal acceptance of GMOs and gene-edited crops, which can influence regulatory policies and market acceptance.

6.4 Integration with traditional breeding programs

Integrating genomic tools with traditional breeding programs presents both logistical and methodological challenges. One of the main issues is the need to align the objectives and timelines of genomic and traditional breeding approaches. Traditional breeding cycles are often lengthy, and incorporating genomic tools requires careful planning and coordination to ensure that genomic data can be effectively used to inform selection decisions at appropriate stages of the breeding cycle.

Another challenge is the need for capacity building and training of breeders in the use of genomic tools. Many breeders are accustomed to conventional phenotypic selection methods and may lack the expertise required to

implement genomic selection, GWAS, and other advanced genomic techniques. This necessitates ongoing training and education programs to equip breeders with the necessary skills and knowledge. Moreover, the integration process requires the development of user-friendly bioinformatics tools and platforms that can facilitate the seamless incorporation of genomic data into breeding decisions (Du et al., 2022).

7 Future Prospects and Innovations

7.1 Emerging genomic technologies

The future of soybean breeding is set to be transformed by a range of emerging genomic technologies. Among these, CRISPR/Cas9 has been particularly influential, providing a precise, efficient, and versatile tool for genome editing. This technology has enabled the targeted modification of genes to improve traits such as yield, disease resistance, and stress tolerance. For instance, the development of high oleic acid soybean oil through CRISPR/Cas9 without the need for GMO regulations marks a significant milestone in precision breeding.

Advancements in high-throughput sequencing and genotyping technologies, such as the SoySNP618K array, have facilitated the generation of detailed genomic data. This data is crucial for identifying SNPs and QTLs associated with desirable traits, enabling more accurate selection in breeding programs (Li et al., 2017). The integration of artificial intelligence (AI) and machine learning (ML) in genomic research is also opening new frontiers. These technologies can analyze large datasets to identify patterns and predict outcomes, enhancing the efficiency and precision of breeding programs. For example, the use of ML algorithms in phenomic-assisted breeding has shown promise in optimizing breeding decisions by accurately predicting traits based on phenomic data (Parmley et al., 2019).

7.2 Integration of multi-omics approaches

The integration of multi-omics approaches-combining genomics, transcriptomics, proteomics, metabolomics, and phenomics, is expected to revolutionize soybean breeding. These approaches provide a comprehensive understanding of the molecular mechanisms underlying complex traits, facilitating more precise and targeted breeding strategies. For example, advances in metabolomics have enabled the identification of key metabolic pathways and biomarkers associated with stress tolerance and yield improvement (Chaudhary et al., 2018).

The integration of omics data allows for the holistic analysis of biological processes, leading to the identification of candidate genes and regulatory networks that control important traits. This integrated approach has been used to enhance soybean seed composition traits, such as oil and protein content, by identifying and selecting for beneficial alleles (Zhang et al., 2021). Moreover, the combination of genomic and phenomic data through high-throughput phenotyping platforms has improved the accuracy of trait prediction and selection. This approach has been particularly effective in identifying QTLs and developing molecular markers for complex traits, thereby accelerating the breeding process (Cao et al., 2022).

7.3 Precision breeding and digital agriculture

Precision breeding, supported by digital agriculture technologies, is poised to enhance the efficiency and effectiveness of soybean breeding programs. Precision breeding involves the use of genomic tools to make precise genetic improvements, while digital agriculture employs advanced technologies such as drones, sensors, and data analytics to monitor and manage crop production.

The use of digital tools for high-throughput phenotyping allows for the collection of large-scale, accurate phenotypic data, which can be integrated with genomic information to improve breeding decisions. For instance, the deployment of sensor technologies and AI in field trials enables real-time monitoring of crop performance, facilitating the identification of superior genotypes under various environmental conditions (Parmley et al., 2019).

Moreover, precision agriculture techniques, such as variable rate technology (VRT) and automated machinery, optimize the application of inputs like water, fertilizers, and pesticides, reducing waste and enhancing crop productivity. These technologies are essential for implementing sustainable farming practices and improving the environmental footprint of soybean production (Du et al., 2022).

7.4 Global collaboration and data sharing

Global collaboration and data sharing are crucial for advancing soybean breeding. The exchange of genetic resources, data, and knowledge among international researchers accelerates the discovery and utilization of beneficial alleles and breeding techniques. Collaborative efforts, such as the development of soybean pangenomes and global germplasm collections, provide valuable resources for identifying genetic diversity and improving breeding programs (Petereit et al., 2022).

Initiatives like the Soybean Genome Initiative (SGI) and the Soybean Research and Development Consortium (SRDC) facilitate the sharing of genomic data and breeding materials, fostering innovation and progress in soybean breeding. These collaborations ensure that breeding advancements benefit a wide range of regions and farming systems, contributing to global food security and agricultural sustainability (Li et al., 2017).

Open-access databases and bioinformatics tools, such as SoyBase and the Soybean Breeder's Toolbox, provide platforms for researchers to access and analyze genomic and phenotypic data. These resources enhance the capacity for genomic research and breeding, supporting the development of improved soybean varieties worldwide.

8 Concluding Remarks

The introduction of genomic tools in soybean breeding has brought significant advancements and efficiencies to the field. Techniques such as genomic selection (GS), marker-assisted selection (MAS), genome-wide association studies (GWAS), and CRISPR/Cas9 gene editing have revolutionized the way breeders enhance desirable traits. GS has demonstrated high predictive accuracy for complex traits, improving yield and seed quality. MAS has been successful in developing disease-resistant varieties, while GWAS has identified numerous loci associated with essential agronomic traits. Additionally, CRISPR/Cas9 technology has enabled precise gene modifications, leading to the creation of soybean varieties with enhanced stress tolerance and nutritional profiles. These genomic tools have not only accelerated the breeding cycle but have also provided a deeper understanding of the genetic basis of important traits, facilitating more targeted and effective breeding strategies. The integration of these tools has led to the development of soybean varieties that are more productive, resilient, and nutritionally rich, contributing to food security and agricultural sustainability.

Genomic tools play a crucial role in promoting sustainable agriculture by enhancing crop productivity and resilience while minimizing environmental impact. The precision offered by genomic selection and gene editing allows for the development of crops that require fewer inputs, such as water, fertilizers, and pesticides, thereby reducing the agricultural footprint. For example, genomic interventions have been pivotal in developing climate-smart soybean varieties that can withstand abiotic stresses like drought and salinity, which are becoming increasingly prevalent due to climate change. Genomic tools contribute to biodiversity conservation by enabling the utilization of a broader genetic base in breeding programs. This diversification is critical for maintaining ecosystem stability and ensuring long-term agricultural sustainability. Advances in multi-omics approaches integrate genomic, transcriptomic, and metabolomic data to provide a comprehensive understanding of plant biology, further supporting the development of resilient and sustainable crop varieties.

The future of soybean breeding and agriculture, in general, is set to be shaped by continuous innovations in genomic technologies. The integration of artificial intelligence (AI) and machine learning (ML) with genomic data holds promise for further enhancing the efficiency and precision of breeding programs. These technologies can analyze vast amounts of data to predict trait outcomes and optimize breeding decisions, accelerating the development of superior soybean varieties. Global collaboration and data sharing will be essential to leverage these advancements fully. Initiatives such as the Soybean Genome Initiative (SGI) and the development of global germplasm collections provide valuable resources for breeders worldwide. Open-access databases and bioinformatics platforms will play a pivotal role in facilitating research and innovation, ensuring that the benefits of genomic advancements are widely accessible. In conclusion, the integration of genomic tools into soybean breeding has the potential to revolutionize agriculture by enhancing productivity, resilience, and sustainability.

Continued innovation, collaboration, and data sharing will be key to realizing the full potential of these technologies and addressing the global challenges of food security and environmental sustainability.

Acknowledgments

The authors sincerely thank the two anonymous peer reviewers for their valuable comments and suggestions on 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

Belzile F., Jean M., Torkamaneh D., Tardivel A., Lemay M.A., Boudhrioua C., Arsenault-Labrecque G., Dussault-Benoit C., Lebreton, A., de Ronne M., Tremblay V., Labbé C., O'Donoughue L., Boucher St-Amour V.T., Copley T.R., Fortier E., Ste-Croix D.T., Mimee B., Cober E., Rajcan I., Warkentin T., Gagnon É., Legay S., Auclair J., and Bélanger R., 2022, The soyagen project: putting genomics to work for soybean breeders, Frontiers in Plant Science, 13: 887553.
https://doi.org/10.3389/fpls.2022.887553

Bernardo R., 2016, Optimizing genomic selection in recurrent selection programs: insights from maize, Genetics, 202(2): 639-647.

Bhat J., and Yu D., 2021, High-throughput NGS-based genotyping and phenotyping: role in genomics-assisted breeding for soybean improvement, Legume Science, 3(3): e81.
https://doi.org/10.1002/leg3.81

Chandra S., Choudhary M., Bagaria P., Nataraj V., Kumawat G., Choudhary J., Sonah H., Gupta S., Wani S.H., and Ratnaparkhe M., 2022, Progress and prospectus in genetics and genomics of phytophthora root and stem rot resistance in soybean (*Glycine max* L.), Frontiers in Genetics, 13: 939182.
https://doi.org/10.3389/fgene.2022.939182

Chaudhary J., Deshmukh R., Mir Z., and Bhat J., 2018, Metabolomics: an emerging technology for soybean improvement, EcoProduction, 2019: 175-186.
https://doi.org/10.1007/978-3-319-92399-4_12

Che Z., Yan H., Liu H., Yang H., Du H., Yang Y., and Liu B., 2020, Genome-wide association study for soybean mosaic virus SC3 resistance in soybean, Molecular Breeding, 40: 1-14.
https://doi.org/10.1007/s11032-020-01149-1

Do P., Nguyen C.X., Bui H., Tran L.T.N., Stacey G., Gillman J., Zhang Z.J., and Stacey M., 2019, Demonstration of highly efficient dual gRNA CRISPR/Cas9 editing of the homeologous GmFAD2-1A and GmFAD2-1B genes to yield a high oleic, low linoleic and α-linolenic acid phenotype in soybean, BMC Plant Biology, 19: 1906.
https://doi.org/10.1186/s12870-019-1906-8

Du H., Fang C., Li Y., Kong F., and Liu B., 2022, Understandings and future challenges in soybean functional genomics and molecular breeding, Journal of Integrative Plant Biology, 65(2): 468-495.
https://doi.org/10.1111/jipb.13433

Dubiela C.R., Montecelli T.D., Lazzari F., Souto E.R., and Schuster I., 2019, Development and validation of SNP assays for the selection of resistance to *Meloidogyne incognita* in soybean, Crop Breeding and Applied Biotechnology, 19(1): 102-109.
https://doi.org/10.1590/1984-70332019v19n1a14

Duhnen A., Gras A., Teyssedre S., Romestant M., Claustres B., Daydé J., and Mangin B., 2017, Genomic selection for yield and seed protein content in soybean: a study of breeding program data and assessment of prediction accuracy, Crop Science, 57: 1325-1337.
https://doi.org/10.2135/cropsci2016.06.0496

Espindola S., Hamawaki O.T., Oliveira A.P., Hamawaki C., Hamawaki R.L., and Takahashi L.M., 2016, Selecting soybean resistant to the cyst nematode *Heterodera glycines* using simple sequence repeat (microssatellite) markers, Genetics and Molecular Research, 15(1): 1-7.
https://doi.org/10.4238/gmr.15016850

Jin Z., Zhang X., Shen S., Wang X., and Liu W., 2023, Digital breeding: Intelligent acceleration of breeding process and innovative soybean breeding program using genomics, Plant Science, 316: 111-120.

Kim J.M., Kim K., Jung J., Kang B., Lee J., Ha B., and Kang S., 2020, Validation of marker-assisted selection in soybean breeding program for pod shattering resistance, Euphytica, 216: 1-9.
https://doi.org/10.1007/s10681-020-02703-w

Kim K., Vuong T., Qiu D., Robbins R., Shannon J.G., Li Z., and Nguyen H., 2016, Advancements in breeding, genetics, and genomics for resistance to three nematode species in soybean, Theoretical and Applied Genetics, 129: 2295-2311.
https://doi.org/10.1007/s00122-016-2816-x

Li Z., Stewart-Brown B.B., Steketee C.J., and Vaughn J.N., 2017, Impact of genomic research on soybean breeding, The Soybean Genome, 2017: 111-129.
https://doi.org/10.1007/978-3-319-64198-0_8

Nagamine A., and Ezura H., 2022, Genome editing for improving crop nutrition, Frontiers in Genome Editing, 4: 850104.
https://doi.org/10.3389/fgeed.2022.850104

Parmley K.A., Nagasubramanian K., Sarkar S., Ganapathysubramanian B., and Singh A.K., 2019, Development of optimized phenomic predictors for efficient plant breeding decisions using phenomic-assisted selection in soybean, Plant Phenomics, 2019: 1-15.
https://doi.org/10.34133/2019/5809404

Petereit J., Marsh J.I., Bayer P., Danilevicz M.F., Thomas W.J.W., Batley J., and Edwards D., 2022, Genetic and genomic resources for soybean breeding research, Plants, 11(9): 1181.
https://doi.org/10.3390/plants11091181

Pham A., Harris D.K., Buck J., Hoskins A., Serrano J., Abdel-Haleem H., Cregan P., Song Q., Boerma H., and Li Z., 2015, Fine mapping and characterization of candidate genes that control resistance to *Cercospora sojina* K. Hara in two soybean germplasm accessions, PLoS One, 10(5): e0126753.
https://doi.org/10.1371/journal.pone.0126753

Ravelombola W., Qin J., Shi A., Nice L., Bao Y., Lorenz A.J., Orf J.H., Young N.D., and Chen S., 2020, Genome-wide association study and genomic selection for tolerance of soybean biomass to soybean cyst nematode infestation, PLoS One, 15(7): e0235089.
https://doi.org/10.1371/journal.pone.0235089

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

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

Valliyodan B., Ye H., Song L., Murphy M., Shannon J., and Nguyen H., 2016, Genetic diversity and genomic strategies for improving drought and waterlogging tolerance in soybeans, Journal of Experimental Botany, 68: 1835-1849.
https://doi.org/10.1093/jxb/erw433

Wang T., Xun H., Wang W., Ding X., Tian H., Hussain S., Dong Q., Li Y., Cheng Y., Wang C., Lin R., Li G., Qian X., Pang J., Feng X., Dong Y., and Liu B., 2021, Mutation of *GmAITR* genes by CRISPR/Cas9 genome editing results in enhanced salinity stress tolerance in soybean, Frontiers in Plant Science, 12: 779598.
https://doi.org/10.3389/fpls.2021.779598

Wei T., Jiang L., You X., Ma P., Xi Z., and Wang N., 2023, Generation of herbicide-resistant soybean by base editing, Biology, 12(5): 741.
https://doi.org/10.3390/biology12050741

Xu K., Zhao Y., Zhao Y., Feng C., Zhang Y., Wang F., Li X., Gao H., Liu W., Jing Y., Saxena R., Feng X., Zhou Y., and Li H., 2022, Soybean F-box-like protein GmFBL144 interacts with small heat shock protein and negatively regulates plant drought stress tolerance, Frontiers in Plant Science, 13: 823529.
https://doi.org/10.3389/fpls.2022.823529

Yin K., and Qiu J., 2019, Genome editing for plant disease resistance: applications and perspectives, Philosophical Transactions of the Royal Society B, 374(1767), 20180322.
https://doi.org/10.1098/rstb.2018.0322

Zhang M., Liu S., Wang Z., Yuan Y., Zhang Z., Liang Q., Yang X., Duan Z., Liu Y., Kong F., Liu B., and Ren B., 2021, Progress in soybean functional genomics over the past decade, Plant Biotechnology Journal, 20: 256-282.
https://doi.org/10.1111/pbi.13682

Zhong C., Sun S., Li Y., Duan C., and Zhu Z., 2018, Next-generation sequencing to identify candidate genes and develop diagnostic markers for a novel phytophthora resistance gene, RpsHC18, in soybean, Theoretical and Applied Genetics, 131: 525-538.
https://doi.org/10.1007/s00122-017-3016-z

Zhou Z., Jiang Y., Wang Z., Gou Z., Lyu J., Li W., Yu Y., Shu L., Zhao Y., Ma Y., Fang C., Shen Y., Liu T., Li C., Li Q., Wu M., Wang M., Wu Y., Dong Y., Wan W., Wang X., Ding Z., Gao Y., Xiang H., Zhu B., Lee S.H., Wang W., and Tian Z., 2015, Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean, Nature Biotechnology, 33: 408-414.
https://doi.org/10.1038/nbt.3096

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.