

Integrating GWAS and Genomic Selection to Enhance Soybean Breeding

Jun Lei¹, Zhuwei Xu¹, Xiaowei Shao¹, Huan Jiang¹, Yumei Zhang² ✉

¹ Quzhou Academy of Agricultural and Forestry Sciences, Quzhou, 324000, Zhejiang, China

² Institute of Crop Sciences, Fujian Academy of Agricultural Sciences, Fujian Engineering Research Center for Characteristic Dry Crop Varieties Breeding, Fuzhou, 350013, Fujian, China

✉ Corresponding email: zym1122@126.com

Legume Genomics and Genetics, 2024 Vol.15, No.6 doi: [10.5376/lgg.2024.15.0026](https://doi.org/10.5376/lgg.2024.15.0026)

Received: 03 Nov., 2024

Accepted: 05 Dec., 2024

Published: 15 Dec., 2024

Copyright © 2024 Lei et al., 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:

Lei J., Xu Z.W., Shao X.W., Jiang H., and Zhang Y.M., 2024, Integrating GWAS and genomic selection to enhance soybean breeding, Legume Genomics and Genetics, 15(6): 270-279 (doi: [10.5376/lgg.2024.15.0026](https://doi.org/10.5376/lgg.2024.15.0026))

Abstract This study explores the integration of Genome-Wide Association Studies (GWAS) and Genomic Selection (GS) to enhance soybean breeding efficiency. By leveraging GWAS for genetic insights and GS for predictive selection, the study identifies key agronomic traits, including yield, disease resistance, and stress tolerance, that are essential to soybean crop improvement. Through case studies, it highlights the effectiveness of GWAS and GS in identifying high-performing genotypes and accelerating breeding cycles. The study further addresses challenges such as the resource demands of genomic technologies and potential solutions, including machine learning and high-throughput phenotyping. The findings underscore the transformative potential of combining GWAS and GS for breeding programs, aiming to meet global demands for high-yielding, resilient soybean varieties and to promote sustainable agricultural practices.

Keywords Soybean breeding; GWAS; Genomic selection; Yield improvement; Sustainable agriculture

1 Introduction

Soybean (*Glycine max* L.) is a globally significant crop, renowned for its versatile applications in food, feed, and industrial sectors. It is a primary source of plant-based protein and oil, contributing substantially to human diets, livestock feed, and various industrial products (Anderson et al., 2019; Singer et al., 2023). The crop's adaptability and high nutritional value have made it a staple in many regions, particularly in the Western Hemisphere, where the majority of the world's soybean cultivation occurs. Soybeans are also integral to the production of biodiesel and other industrial materials, underscoring their economic importance. The increasing global population and changing dietary preferences have escalated the demand for soybeans. Traditional breeding methods, while effective, are often time-consuming and may not keep pace with the rapid need for improved crop varieties (Anderson et al., 2019; Kumar et al., 2021). Enhanced breeding techniques are essential to develop soybean varieties that can meet the growing demand for higher yields, improved nutritional quality, and resistance to biotic and abiotic stresses (Miller et al., 2023). Advanced breeding technologies, such as genomic selection and genome editing, offer promising solutions to accelerate the development of superior soybean cultivars (Yao et al., 2023).

Genome-Wide Association Studies (GWAS) and Genomic Selection (GS) are cutting-edge techniques that have revolutionized plant breeding. GWAS involves scanning the genome of diverse germplasm collections to identify genetic variants associated with specific traits, thereby facilitating the discovery of quantitative trait loci (QTLs) and candidate genes (Kim et al., 2023). This method has been instrumental in dissecting the genetic basis of complex traits in soybeans, such as yield, protein content, and disease resistance (Wang et al., 2023). Genomic Selection (GS), on the other hand, uses genome-wide markers to predict the breeding values of individuals, enabling the selection of superior genotypes early in the breeding cycle. This approach has shown significant promise in improving traits like yield, protein, and oil content in soybean breeding programs (Miller et al., 2023). By integrating GWAS and GS, breeders can enhance the accuracy and efficiency of selecting desirable traits, thereby accelerating the development of high-performing soybean varieties (Rani et al., 2023).

This study aims to provide a comprehensive overview of the combined application of Genome-Wide Association Studies (GWAS) and Genomic Selection (GS) in soybean breeding. It will explore the advancements of these technologies, their application in the identification and selection of key agronomic traits, and the potential benefits of their integrated use in breeding programs. Additionally, the study will address the challenges and future directions in this field, offering insights on how to leverage these innovative approaches to meet the global demand for improved soybean varieties. The goal is to highlight the transformative potential of integrating GWAS and GS to enhance soybean breeding efforts.

2 Background on Soybean Breeding

2.1 Traditional breeding methods and their limitations

Traditional soybean breeding methods primarily involve phenotypic selection, where plants are selected based on observable traits such as yield, plant height, and disease resistance. These methods have been effective in improving soybean varieties over the years but come with several limitations. One major limitation is the long breeding cycle, which can span several years due to the time required for plants to grow and express the desired traits. Additionally, phenotypic selection is often influenced by environmental factors, making it challenging to accurately select for genetic potential. The complexity of quantitative traits, which are controlled by multiple genes, further complicates the breeding process, as traditional methods may not effectively capture the genetic variation underlying these traits (Ravelombola et al., 2021; Budhlakoti et al., 2022).

2.2 Achievements in soybean breeding over the decades

Despite the limitations of traditional breeding methods, significant achievements have been made in soybean breeding over the decades. Advances in breeding techniques have led to the development of high-yielding soybean varieties with improved resistance to diseases and pests. For instance, the identification and incorporation of specific genes responsible for disease resistance have resulted in varieties that are more resilient to common soybean pathogens. Additionally, breeding efforts have focused on improving agronomic traits such as seed composition, including protein and oil content, which are critical for both human consumption and industrial applications (Duhnen et al., 2017; Miller et al., 2023). The integration of molecular markers and genomic tools has further accelerated the breeding process, enabling more precise selection and faster development of superior soybean varieties (He et al., 2014; Sonah et al., 2015).

2.3 Role of molecular markers in improving breeding efficiency

The advent of molecular markers has revolutionized soybean breeding by providing tools for more accurate and efficient selection. Molecular markers, such as Single Nucleotide Polymorphisms (SNPs), allow breeders to identify and select for specific genetic variations associated with desirable traits. Marker-assisted selection (MAS) has been particularly effective in improving traits that are difficult to measure phenotypically or are influenced by multiple genes (He et al., 2014). The use of Genotyping-By-Sequencing (GBS) and Genome-Wide Association Studies (GWAS) has enabled the identification of numerous SNPs linked to important agronomic traits, such as yield, plant height, and seed composition (Sonah et al., 2015). These markers can be used in Genomic Selection (GS) models to predict the breeding value of individuals, thereby enhancing the accuracy and efficiency of the breeding process. The integration of GWAS and GS has shown promising results in increasing genetic gain and accelerating the development of improved soybean varieties (Ma et al., 2016; Ravelombola et al., 2021; Budhlakoti et al., 2022).

3 Principles of GWAS in Soybean

3.1 Explanation of GWAS and its application in crop genetics

Genome-Wide Association Studies (GWAS) are a powerful tool used to identify genetic variants associated with specific traits by scanning the genomes of many individuals. In crop genetics, GWAS helps in understanding the genetic basis of complex traits such as yield, disease resistance, and quality traits. By leveraging high-density Single Nucleotide Polymorphism (SNP) markers, GWAS can pinpoint loci that contribute to phenotypic variation, facilitating Marker-Assisted Selection (MAS) and Genomic Selection (GS) in breeding programs (Sonah et al., 2015; Ravelombola et al., 2021; Priyanatha et al., 2022).

3.2 Key findings from GWAS studies in soybean

Several GWAS studies have identified key loci associated with important agronomic traits in soybean. For instance, a study identified significant SNPs associated with maturity, plant height, seed weight, and yield, with some SNPs mapped to known loci such as E2, E4, and Dt1 (Ravelombola et al., 2021). Another study focusing on soybean germplasm derived from Canadian × Chinese crosses identified QTL regions controlling seed yield, oil, and protein content, highlighting the potential of Chinese cultivars in improving these traits (Priyanatha et al., 2022). Additionally, a haplotype-based GWAS identified stable haplotype associations for seed yield and seed weight across different environments, providing insights into the genetic determinants of these traits (Contreras-Soto et al., 2017).

3.3 Challenges and limitations of using GWAS in soybean breeding

Despite its advantages, GWAS in soybean breeding faces several challenges. One major limitation is the insufficient statistical power to detect QTL with small effects, especially in populations with narrow genetic bases (Yoosefzadeh-Najafabadi et al., 2021a). Additionally, the complex genetic architecture of traits, involving multiple loci with small effects and gene-environment interactions, complicates the identification of significant associations (Yoosefzadeh-Najafabadi et al., 2023). Another challenge is the need for high-density SNP markers and large, diverse populations to ensure robust and reproducible results (Sonah et al., 2015; Mandozai et al., 2021). Furthermore, integrating GWAS findings into practical breeding programs requires validation of identified markers and candidate genes, which can be resource-intensive and time-consuming.

4 Advances in Genomic Selection (GS)

4.1 Overview of gs and its application in crop improvement

Genomic Selection (GS) is a revolutionary approach in plant breeding that leverages genome-wide marker data to predict the breeding values of individuals within a population. Unlike traditional marker-assisted selection, which focuses on a few significant markers, GS incorporates all available marker information into the prediction model. This comprehensive approach allows for the capture of small-effect Quantitative Trait Loci (QTL) that contribute to complex traits, thereby improving the accuracy of selection and accelerating the breeding cycle (Jannink et al., 2010; Varshney et al., 2017; Merrick et al., 2022).

GS has been successfully applied in various crop improvement programs, including those for cereals like wheat, maize, and rice, as well as legumes such as soybeans. The method has shown promise in enhancing traits related to yield, quality, and stress tolerance, which are often controlled by multiple genes with small effects (Krishnappa et al., 2021; Xu et al., 2021; Budhlakoti et al., 2022). By integrating GS with other advanced technologies like high-throughput phenotyping and deep learning, breeding programs can achieve more rapid and cost-effective genetic gains (Krishnappa et al., 2021; Merrick et al., 2022).

4.2 Differences between gs and traditional marker-assisted selection

Traditional Marker-Assisted Selection (MAS) has been effective for traits controlled by a few major QTLs but has limitations when dealing with polygenic traits. MAS typically involves identifying and selecting for specific markers associated with large-effect QTLs, which can be a time-consuming and less efficient process for complex traits (Jannink et al., 2010; Varshney et al., 2017; Merrick et al., 2022). In contrast, GS uses genome-wide markers to estimate the effects of all loci simultaneously, providing a more holistic and accurate prediction of an individual's genetic potential. This method avoids the bias associated with selecting only significant markers and captures the cumulative effect of numerous small-effect QTLs. As a result, GS can accelerate the breeding cycle by allowing for earlier selection decisions based on Genomic Estimated Breeding Values (GEBVs) rather than waiting for phenotypic data (Wang et al., 2018). Additionally, GS can reduce the need for extensive phenotyping, thereby saving time and resources (Xu et al., 2021; Budhlakoti et al., 2022).

4.3 Successful cases of GS implementation in soybean breeding programs

Several soybean breeding programs have successfully implemented GS to enhance the selection of desirable traits. For instance, a study involving 483 elite soybean breeding lines demonstrated the potential of GS for improving

protein, oil, and yield traits. The study achieved high predictive abilities (rMP) for protein (0.81) and oil (0.71) using the RR-BLUP model, indicating the effectiveness of GS in soybean breeding (Stewart-Brown et al., 2019).

Another notable example is the integration of GS with Genome-Wide Association Studies (GWAS) to identify significant SNP markers associated with key agronomic traits such as maturity, plant height, and seed weight. This combined approach has led to the discovery of new loci and candidate genes that can be targeted for improving yield and other important traits in soybean (Ravelombola et al., 2021) (Figure 1).

Overall, the successful implementation of GS in soybean breeding programs highlights its potential to accelerate genetic gains and improve the efficiency of breeding processes. By leveraging comprehensive genomic data and advanced prediction models, breeders can make more informed selection decisions and develop superior soybean varieties more rapidly (Crossa et al., 2017; Stewart-Brown et al., 2019; Ravelombola et al., 2021).

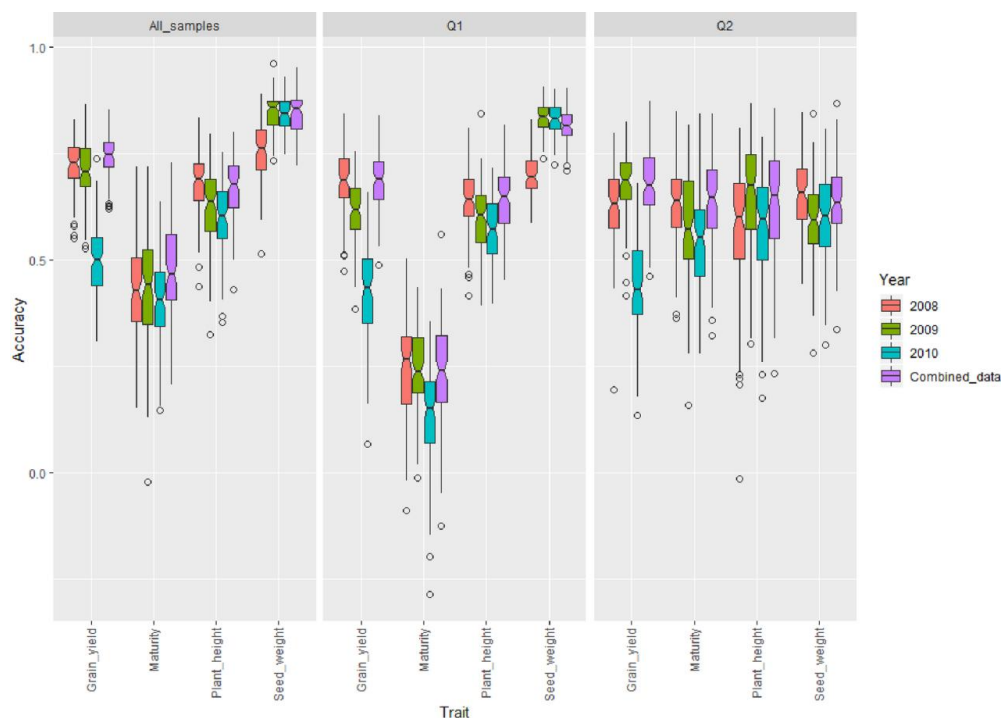


Figure 1 Genomic selection accuracy for yield, maturity, plant height, and seed weight using training/testing sets from all 250 soybean accessions (all samples), samples derived from Q1, and samples from the Q2 subpopulation (Adopted from Ravelombola et al., 2021)

5 Integration of GWAS and Genomic Selection

5.1 How GWAS can inform and enhance GS models

Genome-Wide Association Studies (GWAS) identify significant Single Nucleotide Polymorphisms (SNPs) associated with traits of interest, providing valuable insights into the genetic architecture of these traits. By pinpointing specific loci that contribute to phenotypic variation, GWAS can enhance Genomic Selection (GS) models by incorporating these significant SNPs, thereby improving the accuracy of trait prediction. For instance, the identification of SNPs associated with yield, maturity, plant height, and seed weight in soybean has been shown to enhance the predictive power of GS models (Ravelombola et al., 2021). Additionally, the integration of GWAS findings into GS models can help in selecting the most informative markers, which can reduce genotyping costs while maintaining or even improving prediction accuracy (Luo et al., 2021).

5.2 Methods for integrating GWAS findings into GS pipelines

Several methods can be employed to integrate GWAS findings into GS pipelines. One approach is to use a subset of SNPs identified by GWAS for GS, which has been shown to improve prediction accuracy compared to using all available SNPs (Luo et al., 2021). Another method involves the use of stepwise linear regression mixed models

(StepLMM), which combine GWAS and GS in a single statistical framework. This model first estimates the variance components using genomic best linear unbiased prediction (GBLUP) and then selects the most significant SNPs through a linear mixed model transformation, improving both mapping precision and prediction accuracy (Li et al., 2017). Additionally, Bayesian models such as Bayesian Lasso (BL) and ridge regression Best Linear Unbiased Prediction (rrBLUP) can be used to incorporate GWAS-identified SNPs, enhancing the efficiency of GS pipelines (Qin et al., 2022).

5.3 Benefits of combining GWAS and GS for trait prediction and selection accuracy

Combining GWAS and GS offers several benefits for trait prediction and selection accuracy. Firstly, it leverages the strengths of both approaches: GWAS provides high-resolution mapping of trait-associated loci, while GS offers robust prediction of breeding values using genome-wide markers. This synergy can lead to higher prediction accuracies for complex traits, as demonstrated in studies where the integration of GWAS findings into GS models resulted in improved prediction accuracy for traits such as yield, protein content, and disease resistance (Luo et al., 2021; Ravelombola et al., 2021; Qin et al., 2022). Moreover, the combined approach can reduce the rate of false positives in QTL mapping and enhance the precision of trait-associated loci identification, ultimately leading to more efficient and effective breeding programs (Li et al., 2017). The use of GWAS-informed GS models can also accelerate the breeding cycle by enabling earlier and more accurate selection of superior genotypes, thus increasing genetic gain over time (Matei et al., 2018; Stewart-Brown et al., 2019).

6 Case Studies

6.1 Yield improvement: integration of GWAS and GS to enhance yield traits

The integration of Genome-Wide Association Studies (GWAS) and Genomic Selection (GS) has shown significant promise in enhancing yield traits in soybean breeding. For instance, a study on soybean breeding lines demonstrated the potential of GS for selecting quantitative traits such as yield, where traditional marker-assisted selection has often been less effective. The study utilized 483 elite breeding lines and achieved predictive abilities (rMP) of 0.26 for yield, indicating the potential of GS to improve yield traits through more accurate predictions and selection (Stewart-Brown et al., 2019). Additionally, the application of GS in other crops like wheat has shown that leveraging genomic information can significantly increase prediction accuracies, thereby accelerating the development of high-yielding varieties (Budhlakoti et al., 2022).

6.2 Disease resistance: examples of breeding for resistance to soybean pathogens

Breeding for disease resistance in soybean has benefited from the integration of GWAS and GS. For example, the BREEDWHEAT project has successfully utilized these genomic tools to decipher traits of agronomical interest, including biotic resistance, which is crucial for developing disease-resistant varieties (Paux et al., 2022). Furthermore, the identification of specific loci associated with disease resistance through GWAS has enabled the development of soybean varieties with enhanced resistance to various pathogens. This approach not only helps in identifying candidate genes but also facilitates the implementation of genomic selection to improve disease resistance traits (Jannink et al., 2010).

6.3 Stress tolerance: breeding for abiotic stress (e.g., drought, salinity) using combined approaches

The combined use of GWAS and GS has been instrumental in breeding for abiotic stress tolerance in soybean. For instance, a study identified a major salt-tolerance locus controlled by the *E2* gene, which also influences flowering time and maturity. The loss of *E2* function not only enhanced salt tolerance but also shortened flowering time, demonstrating the effectiveness of integrating GWAS findings into breeding programs for stress tolerance (Dong et al., 2022). Additionally, research on drought tolerance in soybean has identified several quantitative trait loci (QTLs) and candidate genes associated with drought resistance. The over-expression of the *GmNFYB17* gene, identified through GWAS and linkage analysis, significantly improved drought resistance and yield accumulation in transgenic soybean plants (Sun et al., 2022) (Figure 2). These examples highlight the potential of integrating GWAS and GS to develop soybean varieties with improved tolerance to abiotic stresses such as drought and salinity (Valliyodan et al., 2016; Ouyang et al., 2022).

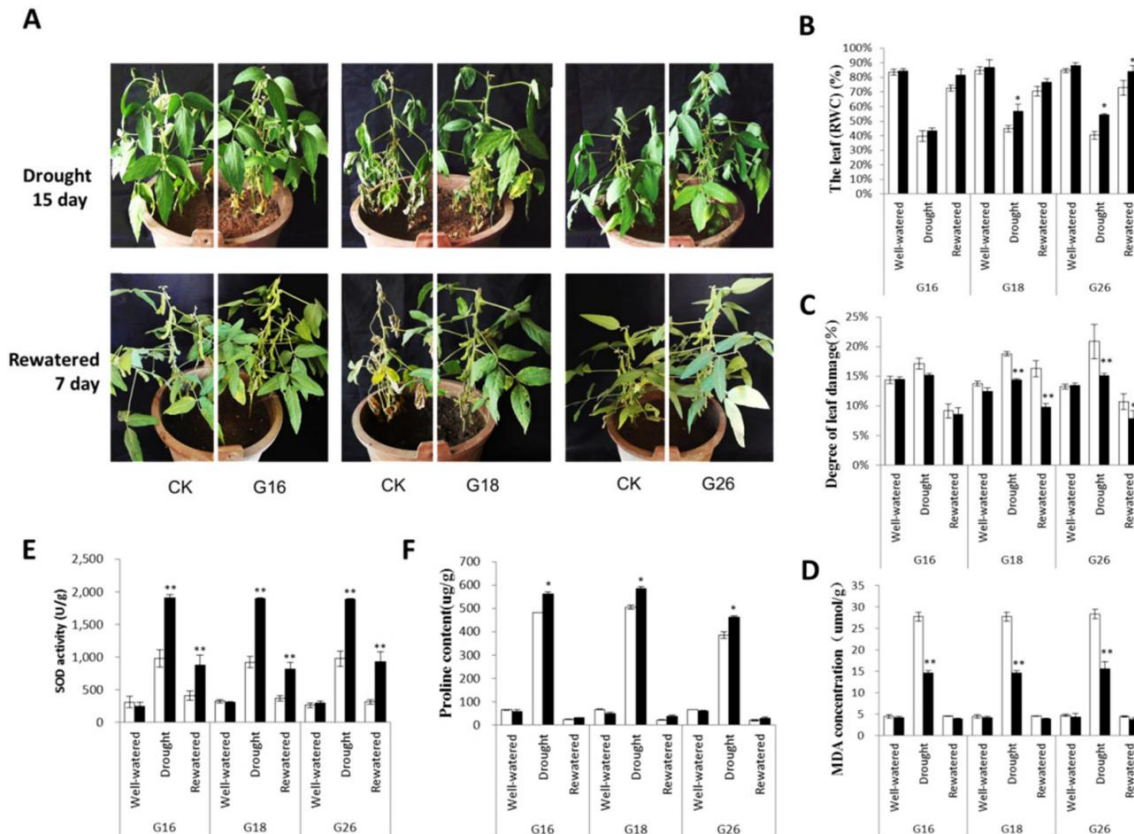


Figure 2 The GmNFYB17 transgenic soybean lines under drought treatment (Adopted from Sun et al., 2022)

Image caption: (A) Morphology of transgenic and non-transgenic plants under drought conditions. Water was withheld for 15 d, and then plants were re-watered for 7 d. G16, G18 and G26 are transgenic lines; CK is soybean DN50; (B,C)The leaf relative water content (RWC) and leaf damage of transgenic lines. G16, G18, G26 and non-transgenic control (CK) during the well-watered, drought and re-watered stage; (D-F) Comparison of physiological and biochemical indicators (MDA, SOD, Proline) between transgenic and non-transgenic plants. *: p-value ≤ 0.05 ; **: p-value ≤ 0.01 (Adopted from Sun et al., 2022)

7 Challenges and Future Prospects

7.1 Technical and computational challenges in integrating GWAS and GS

Integrating Genome-Wide Association Studies (GWAS) and Genomic Selection (GS) in soybean breeding presents several technical and computational challenges. One of the primary issues is the need for advanced statistical methods to accurately detect Quantitative Trait Loci (QTL) and their interactions. Traditional GWAS methods often focus on main effects and may miss significant interactions between QTL, which are crucial for complex traits like yield and seed quality (Yoosefzadeh-Najafabadi et al., 2023). Additionally, the computational power required to handle large datasets and perform complex analyses is substantial. Machine learning algorithms, such as Support Vector Regression (SVR) and Random Forest (RF), have shown promise in improving the accuracy of QTL detection, but they also demand significant computational resources and expertise (Yoosefzadeh-Najafabadi et al., 2021a; Yoosefzadeh-Najafabadi et al., 2021b).

7.2 Cost and resource considerations for breeding programs

The integration of GWAS and GS into breeding programs is resource-intensive. The costs associated with genotyping, phenotyping, and data analysis can be prohibitive, especially for smaller breeding programs. High-throughput phenotyping technologies and the generation of large-scale genomic data require substantial financial investment and technical infrastructure (Sandhu et al., 2022). Moreover, the need for continuous updates to computational tools and methods to keep pace with advancements in the field adds to the overall cost. Efficient allocation of resources and strategic planning are essential to maximize the return on investment in these technologies.

7.3 Potential solutions and future directions

To address these challenges, several potential solutions and future directions can be considered. Advances in machine learning and big data analytics offer promising avenues for enhancing the integration of GWAS and GS. Machine learning algorithms, such as SVR and RF, have demonstrated superior performance in identifying relevant QTL compared to traditional methods (Yoosefzadeh-Najafabadi et al., 2021a; Yoosefzadeh-Najafabadi et al., 2023). Additionally, the development of hierarchical data integration strategies, such as the hyperspectral wide association study (HypWAS), can improve the efficiency of phenome-genome association analyses and provide deeper insights into the genetic architecture of complex traits (Yoosefzadeh-Najafabadi et al., 2021b).

Furthermore, the adoption of phenomics and high-throughput phenotyping tools can alleviate some of the bottlenecks in data acquisition and analysis. These tools enable the collection of large-scale, high-quality phenotypic data, which is essential for accurate genomic predictions (Sandhu et al., 2022). The integration of phenomics with GS and machine learning can accelerate the breeding cycle and enhance genetic gains.

8 Concluding Remarks

This study has explored the integration of Genome-Wide Association Studies (GWAS) and Genomic Selection (GS) to enhance soybean breeding. GWAS has been instrumental in identifying genetic variants associated with complex traits by examining genome-wide genetic variants across diverse genetic materials. On the other hand, GS facilitates the rapid selection of superior genotypes and accelerates the breeding cycle by leveraging genomic-enabled prediction models. The integration of these two approaches can significantly enhance the efficiency and effectiveness of soybean breeding programs.

Integrating GWAS and GS holds immense potential for future soybean breeding. GWAS provides a robust framework for identifying genetic variants associated with important agronomic traits, which can then be used to inform GS models. This combined approach allows for more accurate prediction of phenotypic outcomes based on genetic data, thereby accelerating the breeding process and improving the selection of high-yield, disease-resistant soybean varieties. Additionally, the integration of these methods can help in understanding the genetic basis of complex traits, leading to more targeted and efficient breeding strategies.

To fully realize the benefits of integrating GWAS and GS in soybean breeding, continued research and development are essential. Future studies should focus on optimizing GWAS models to reduce false positives and improve statistical power. Additionally, advancements in genomic technologies, such as hyperspectral imaging and genotype imputation, should be leveraged to enhance the accuracy and efficiency of GS models. Collaborative efforts to curate and integrate high-quality GWAS data, as exemplified by resources like the GWAS Atlas, will also be crucial for advancing genetic research and breeding applications. By investing in these areas, we can unlock the full potential of GWAS and GS to drive innovation in soybean breeding and ensure sustainable agricultural practices.

Acknowledgments

The authors extend sincere thanks to two anonymous peer reviewers for their feedback on the manuscript of this study.

Funding

This work was supported by Major Science and Technology Special Project for New Variety Breeding in Agriculture of Zhejiang Province (2021C02064-5) and Science and Technology Innovative Team in Fujian Academy of Agricultural Sciences (CXTD2021011-2).

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

- Anderson E., Ali L., Beavis W., Chen P., Clemente T., Diers B., Graef G., Grassini P., Hyten D., McHale L., Nelson R., Parrott W., Patil G., Stupar R., and Tilmon K., 2019, Soybean [*Glycine max* (L.) Merr.] breeding: history, improvement, production and future opportunities, *Advances in Plant Breeding Strategies: Legumes*, 7: 431-516.
https://doi.org/10.1007/978-3-030-23400-3_12
- Budhlakoti N., Kushwaha A., Rai A., Chaturvedi K., Kumar A., Pradhan A., Kumar U., Kumar R., Juliana P., Mishra D., and Kumar S., 2022, Genomic selection: a tool for accelerating the efficiency of molecular breeding for development of climate-resilient crops, *Frontiers in Genetics*, 13: 832153.
<https://doi.org/10.3389/fgene.2022.832153>
- Contreras-Soto R., Mora F., Oliveira M., Higashi W., Scapim C., 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>
- Crossa J., Pérez-Rodríguez P., Cuevas J., Montesinos-López O., Jarquín D., Campos G., Burgueño J., González-Camacho J., Pérez-Elizalde S., Beyene Y., Dreisigacker S., Singh R., Zhang X., Gowda M., Roorkiwal M., Rutkoski J., and Varshney R., 2017, Genomic selection in plant breeding: methods, models, and perspectives, *Trends in Plant Science*, 22(11): 961-975.
<https://doi.org/10.1016/j.tplants.2017.08.011>
- Dong L., Hou Z., Li H., Li Z., Fang C., Kong L., Li Y., Du H., Li T., Wang L., He M., Zhao X., Cheng Q., Kong F., and Liu B., 2022, Agronomical selection on loss-of-function of GIGANTEA simultaneously facilitates soybean salt tolerance and early maturity, *Journal of Integrative Plant Biology*, 64(10): 1866-1882.
<https://doi.org/10.1111/jipb.13332>
- 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>
- He J., Zhao X., Laroche A., Lu Z., 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>
- Jannink J., Lorenz A., and Iwata H., 2010, Genomic selection in plant breeding: from theory to practice, *Briefings in Functional Genomics*, 9(2): 166-177.
<https://doi.org/10.1093/bfpg/elq001>
- Kim S., Tayade R., Kang B., Hahn B., Ha B., and Kim Y., 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>
- Krishnappa G., Savadi S., Tyagi B., Singh S., Masthigowda M., Kumar S., Mishra C., Khan H., Krishnappa G., Govindareddy U., Singh G., and Singh G., 2021, Integrated genomic selection for rapid improvement of crops, *Genomics*, 113(3): 1070-1086.
<https://doi.org/10.1016/j.ygeno.2021.02.007>
- Kumar V., Vats S., Kumawat S., Bisht A., Bhatt V., Shivaraj S., Padalkar G., Goyal V., Zargar S., Gupta S., Kumawat G., Chandra S., Chalam V., Ratnaparkhe M., Gill B., Jean M., Patil G., Vuong T., Rajcan I., Deshmukh R., Belzile F., Sharma T., Nguyen H., and Sonah H., 2021, Omics advances and integrative approaches for the simultaneous improvement of seed oil and protein content in soybean (*Glycine max* L.), *Critical Reviews in Plant Sciences*, 40: 398-421.
<https://doi.org/10.1080/07352689.2021.1954778>
- Li H., Su G., Jiang L., and Bao Z., 2017, An efficient unified model for genome-wide association studies and genomic selection, *Genetics, Selection, Evolution: GSE*, 49: 1-8.
<https://doi.org/10.1186/s12711-017-0338-x>
- Luo Z., Yu Y., Xiang J., and Li F., 2021, Genomic selection using a subset of SNPs identified by genome-wide association analysis for disease resistance traits in aquaculture species, *Aquaculture*, 539: 736620.
<https://doi.org/10.1016/J.AQUACULTURE.2021.736620>
- Ma Y., Reif J., Jiang Y., Wen Z., Wang D., Liu Z., Guo Y., Wei S., Wang S., Yang C., Wang H., Yang C., Lu W., Xu R., Zhou R., Wang R., Sun Z., Chen H., Zhang W., Wu J., Hu G., Liu C., Luan X., Fu Y., Guo T., Han T., Zhang M., Sun B., Zhang L., Chen W., Wu C., Sun S., Yuan B., Zhou X., Han D., Yan H., Li W., and Qiu L., 2016, Potential of marker selection to increase prediction accuracy of genomic selection in soybean (*Glycine max* L.), *Molecular Breeding*, 36: 1-10.
<https://doi.org/10.1007/s11032-016-0504-9>
- Mandozai A., Moussa 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>
- Matei G., Woyann L., Milioli A., Oliveira I., Zdziarski A., Zanella R., Coelho A., Finatto T., and Benin G., 2018, Genomic selection in soybean: accuracy and time gain in relation to phenotypic selection, *Molecular Breeding*, 38: 1-13.
<https://doi.org/10.1007/s11032-018-0872-4>
- Merrick L., Herr A., Sandhu K., Lozada D., and Carter A., 2022, Optimizing plant breeding programs for genomic selection, *Agronomy*, 12(3): 714.
<https://doi.org/10.20944/preprints202202.0048.v1>

- Miller M., Song Q., and Li Z., 2023, Genomic selection of soybean (*Glycine max*) for genetic improvement of yield and seed composition in a breeding context, *The Plant Genome*, 16(4): e20384.
<https://doi.org/10.1002/tpg2.20384>
- Ouyang W., Chen L., Ma J., Liu X., Chen H., Yang H., Guo W., Shan Z., Yang Z., Chen S., Zhan Y., Zhang H., Cao D., and Zhou X., 2022, Identification of quantitative trait locus and candidate genes for drought tolerance in a soybean recombinant inbred line population, *International Journal of Molecular Sciences*, 23(18): 10828.
<https://doi.org/10.3390/ijms231810828>
- Paux E., Lafarge S., Balfourier F., Derory J., Charmet G., Alaux M., Perchet G., Bondoux M., Baret F., Barillot R., Ravel C., Sourdille P., Gouis J., and Consortium O., 2022, Breeding for economically and environmentally sustainable wheat varieties: an integrated approach from genomics to selection, *Biology*, 11(1): 149.
<https://doi.org/10.3390/biology11010149>
- Priyanatha C., Torkamaneh D., and Rajcan I., 2022, Genome-wide association study of soybean germplasm derived from canadian × chinese crosses to mine for novel alleles to improve seed yield and seed quality traits, *Frontiers in Plant Science*, 13: 866300.
<https://doi.org/10.3389/fpls.2022.866300>
- 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>
- Rani R., Raza G., Ashfaq H., Rizwan M., Razzaq M., Waheed M., Shimelis H., Babar A., and Arif M., 2023, Genome-wide association study of soybean (*Glycine max* [L.] Merr.) germplasm for dissecting the quantitative trait nucleotides and candidate genes underlying yield-related traits, *Frontiers in Plant Science*, 14: 1229495.
<https://doi.org/10.3389/fpls.2023.1229495>
- 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>
- Sandhu K., Merrick L., Sankaran S., Zhang Z., and Carter A., 2022, Prospectus of genomic selection and phenomics in cereal, legume and oilseed breeding programs, *Frontiers in Genetics*, 12: 829131.
<https://doi.org/10.3389/fgene.2021.829131>
- Singer W., Lee Y., Shea Z., Vieira C., Lee D., Li X., Cunicelli M., Kadam S., Khan M., Shannon G., Mian M., Nguyen H., and Zhang B., 2023, Soybean genetics, genomics, and breeding for improving nutritional value and reducing antinutritional traits in food and feed, *The Plant Genome*, 16(4): e20415.
<https://doi.org/10.1002/tpg2.20415>
- Sonah H., O'Donoghue 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>
- Stewart-Brown B., Song Q., Vaughn J., and Li Z., 2019, Genomic selection for yield and seed composition traits within an applied soybean breeding program, *G3: Genes Genomes Genetics*, 9: 2253-2265.
<https://doi.org/10.1534/g3.118.200917>
- Sun M., Li Y., Zheng J., Wu D., Li C., Li Z., Zang Z., Zhang Y., Fang Q., Li W., Han Y., Zhao X., and Li Y., 2022, A nuclear factor Y-B transcription factor, GmNFYB17, regulates resistance to drought stress in soybean, *International Journal of Molecular Sciences*, 23(13): 7242.
<https://doi.org/10.3390/ijms23137242>
- 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>
- Varshney R., Roorkiwal M., and Sorrells M., 2017, Genomic selection for crop improvement, *Crop Science*, 49: 1-12.
<https://doi.org/10.1007/978-3-319-63170-7>
- Wang J., Feng H., Jia X., Ma S., Ma C., Wang Y., Pan S., Chen Q., Xin D., and Liu C., 2023, Identifications of QTLs and candidate genes associated with *Pseudomonas syringae* responses in cultivated soybean (*Glycine max*) and wild soybean (*Glycine soja*), *International Journal of Molecular Sciences*, 24(5): 4618.
<https://doi.org/10.3390/ijms24054618>
- Wang X., Wang X., Xu Y., Hu Z., and Xu C., 2018, Genomic selection methods for crop improvement: current status and prospects, *The Crop Journal*, 6(4): 330-340.
<https://doi.org/10.1016/J.CJ.2018.03.001>
- Xu Y., Ma K., Zhao Y., Wang X., Zhou K., Yu G., Li C., Li P., Yang Z., Xu C., and Xu S., 2021, Genomic selection: a breakthrough technology in rice breeding, *The Crop Journal*, 9(3): 669-677.
<https://doi.org/10.1016/J.CJ.2021.03.008>
- Yao D., Zhou J., Zhang A., Wang J., Liu Y., Wang L., Pi W., Li Z., Yue W., Cai J., Liu H., Hao W., and Qu X., 2023, Advances in CRISPR/Cas9-based research related to soybean [*Glycine max* (Linn.) Merr] molecular breeding, *Frontiers in Plant Science*, 14: 1247707.
<https://doi.org/10.3389/fpls.2023.1247707>

- Yoosefzadeh-Najafabadi M., Torabi S., Torkamaneh D., Tulpan D., Rajcan I., and Eskandari M., 2021a, Machine-learning-based genome-wide association studies for uncovering QTL underlying soybean yield and its components, *International Journal of Molecular Sciences*, 23(10): 5538.
<https://doi.org/10.3390/ijms23105538>
- Yoosefzadeh-Najafabadi M., Torabi S., Tulpan D., Rajcan I., and Eskandari M., 2023, Application of SVR-Mediated GWAS for identification of durable genetic regions associated with soybean seed quality traits, *Plants*, 12(14): 2659.
<https://doi.org/10.3390/plants12142659>
- Yoosefzadeh-Najafabadi M., Torabi S., Tulpan D., Rajcan I., and Eskandari M., 2021b, Genome-wide association studies of soybean yield-related hyperspectral reflectance bands using machine learning-mediated data integration methods, *Frontiers in Plant Science*, 12: 777028.
<https://doi.org/10.3389/fpls.2021.777028>



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.
