

Mining Key Agronomic Traits through GWAS and Integrating Breeding Strategies for Soybean

Chunxia Wu, Qishan Chen ✉

Modern Agricultural Research Center, Cuixi Academy of Biotechnology, Zhuji, 311800, Zhejiang, China

✉ Corresponding email: qishan.chen@cuixi.org

Legume Genomics and Genetics, 2025 Vol.16, No.1 doi: [10.5376/lgg.2025.16.0001](https://doi.org/10.5376/lgg.2025.16.0001)

Received: 11 Nov., 2024

Accepted: 21 Dec., 2024

Published: 05 Jan., 2025

Copyright © 2025 Wu and Chen, 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:

Wu C.X., and Chen Q.S., 2025, Mining key agronomic traits through GWAS and integrating breeding strategies for soybean, Legume Genomics and Genetics, 16(1): 1-10 (doi: [10.5376/lgg.2025.16.0001](https://doi.org/10.5376/lgg.2025.16.0001))

Abstract Soybean (*Glycine max*) is a very important crop and is cultivated in many places around the world. It is rich in protein and oil and is a common and important part of agriculture and food production. With the continuous progress of genomic research, scientists have also found more ways to improve soybeans. Among them, GWAS (Genome-wide Association Study) is a commonly used technique that can be employed to identify the locations of genes related to agronomic traits. This study introduces the basic concepts and main methods of GWAS, such as how to conduct genotyping, how to collect phenotypic data, and common statistical analysis approaches. All these contents are closely related to soybean research. Nowadays, GWAS has been used to discover many genes related to soybean yield, disease resistance and stress tolerance. These discoveries have provided many references for breeding and also accelerated the progress of breeding. However, there are also many challenges in the practical application of GWAS. For instance, there are significant differences in genetic background among different soybean varieties, and their phenotypes are also easily influenced by the environment. In addition, the interaction between genes (superiority) also makes the analysis more complex. This study also takes the disease resistance of soybeans as an example, focusing on introducing the achievements of GWAS in improving disease-resistant varieties, especially its application in genetically modified soybeans and the benefits it brings. Looking to the future, there are still many areas where GWAS can be improved. For instance, multiple omics data can be combined, more powerful computing tools can be used, and the technology for trait collection can also be improved. All these practices will make GWAS more useful in soybean research. GWAS plays a very important role in soybean breeding. This research has laid a foundation for future genetic studies and provided technical support for the screening and improvement of high-quality soybean varieties.

Keywords Soybean; Genome-wide association studies; Crop improvement; Genetic loci; Breeding programs

1 Introduction

Soybean (*Glycine max* L.) is a major crop that is grown all over the world. Because it is rich in both protein and vegetable oil, it is widely used in many places. Soybeans are an important source of feed for livestock and aquaculture. They are also often used to extract oil and can be processed into various foods. Many people consume them in their daily diet. Soybeans were first grown in China and East Asia and have now spread all over the world. At present, the soybeans produced by countries in the Western Hemisphere account for 80% to 85% of the global total output (Anderson et al., 2019). Because soybeans have strong adaptability and high returns, researchers and farmers in many countries are striving to improve the varieties. Nowadays, soybeans have become one of the oil crops with the largest planting area in the world.

In recent years, genomics has developed rapidly, and scientists have gained more and more understanding of the genes of soybeans. Techniques such as genome-wide association analysis (GWAS) have helped researchers identify many gene loci related to agronomic traits (Jiang, 2024). For instance, the appearance of soybean roots, the high yield, and the proportion of protein and oil in seeds are all related to certain specific genes (Kim et al., 2023a; Rani et al., 2023). These studies have identified many SNP (single nucleotide polymorphism) markers and some potential candidate genes. This information is very useful for subsequent precision breeding. Now, the research team is combining these genomic data with traditional breeding methods, hoping to breed new drought-tolerant and high-yield soybean varieties more quickly (Almeida-Silva et al., 2020).

Nowadays, GWAS has become a commonly used method in breeding research. It is particularly suitable for studying complex traits determined by many genes. In soybean research, scientists have identified many key gene loci through GWAS. These loci are related to some important traits, such as flowering time, maturity period, plant height, and yield performance in different environments (Zhang et al., 2015; Li et al., 2023). The locations of these SNPs are all quite clear. Breeders can select plants with ideal traits more quickly based on these markers, thereby accelerating the breeding process (Bhat et al., 2022). In addition, GWAS can also help us better understand how these traits are inherited. These research results also provide a considerable amount of useful data support for future soybean breeding (Kim et al., 2023b).

The purpose of this study is to summarize some current research achievements on soybean GWAS and look forward to the future development direction. We focused on introducing some gene loci that have been discovered so far and are related to important traits. We also discussed how GWAS and traditional breeding methods can be combined, and finally explained the practical role of these achievements in variety improvement. Through this summary, we hope to make more people recognize the importance of genomic research in crop breeding and also provide some reference directions for future research and breeding work.

2 Principles of GWAS

2.1 GWAS methodology

Genome-wide association studies (GWAS) are a very common genetic analysis method nowadays, mainly used to identify whether there is a relationship between gene variations and traits. Its approach is to search for SNP (single nucleotide polymorphism) variations throughout the entire genome. Then compare individuals with a certain trait and those without it to see if they have any particularly common SNPs. If a certain SNP appears more frequently in individuals with a trait, then this SNP may be related to that trait. To make the results more reliable, researchers later introduced a hybrid model. This method can reduce false positives, that is, false positives, making the analysis more reliable (Cortes et al., 2021). Nowadays, GWAS can not only be used to study common traits such as plant height and yield, but also begins to be applied to analyze more detailed molecular traits like metabolites and enzyme activity. All of these can help us identify key genes more quickly, which is very helpful for breeding and can accelerate the process of cultivating superior varieties.

2.2 Genotyping and phenotyping for GWAS

For GWAS analysis to be conducted well, both genotyping and phenotypic data must be accurate. Genotyping is to identify the variations at different locations in the genome. There are currently two commonly used methods: one is SNP chip, and the other is whole genome resequencing (Korte and Farlow, 2013). The way of obtaining phenotypic data is also improving. In the past, it mainly relied on manual observation and scoring. Now, many people have begun to use image recognition and deep learning (DL) to automatically extract trait information. This approach saves both effort and time, and the results are more stable (Rairdin et al., 2022). For instance, some studies have used deep learning to assess the severity of soybean diseases and identified SNP loci that might be related to disease resistance.

2.3 Statistical approaches in GWAS

In GWAS, the choice of statistical methods is crucial as it determines whether truly useful gene loci can be identified. The most widely used model nowadays is the hybrid linear model (MLM). It can simultaneously take into account the structure among populations and the kinship among individuals, effectively reducing false positives. However, in some crops, such as soybeans with a relatively single genetic background, traditional methods are sometimes not strong enough (Yoosefzadeh-Najafabadi et al., 2021). To address this issue, some studies have introduced machine learning methods, such as support vector regression (SVR) and Random Forest (RF). These methods perform better and have higher accuracy in identifying QTLs (quantitative trait loci) (Yoosefzadeh-Najafabadi et al., 2023). In addition, some new methods are also constantly evolving. For example,

extreme phenotype GWAS (XP-GWAS), as well as the method of analysis using k-mer sequence features. These new technologies can more accurately identify variant sites and candidate genes related to traits (Yang et al., 2015; Lemay et al., 2023).

3 Applications of GWAS in Soybean

3.1 Identification of trait-associated loci

Nowadays, many scientists use GWAS (Genome-wide Association Studies) to identify gene regions related to traits in soybeans. This method is particularly suitable for studying locations related to agronomic traits. For instance, one study utilized GBS (Genotyping sequencing) technology to identify several related regions in the soybean genome. These regions are associated with eight important traits, such as maturity time, plant height, seed size, oil content and protein content (Sonah et al., 2015). Another study combined the results of 73 independent GWAS and conducted a meta-GWAS (Meta-Analysis). A total of 393 gene regions were identified, among which 483 QTL (quantitative trait loci) were found to be related to yield, disease resistance, plant height and other traits (Shook et al., 2021). These research results indicate that GWAS is a very practical tool. It can help us identify the gene regions related to important traits more quickly and also provide many useful references for soybean breeding.

3.2 Agronomically important traits

In soybeans, traits such as yield, plant height and seed weight are generally not controlled by a single gene, but are regulated by many genes together. GWAS is particularly suitable for studying such complex traits. Several SNP haplotypes have been identified in studies, and these gene loci have a significant relationship with major agronomic traits (Contreras-Soto et al., 2017). Another study employed a support vector regression (SVR) model to identify stable gene regions related to soy protein content and oil content (Yoosefzadeh-Najafabadi et al., 2023). These achievements are very useful because they enable us to have a clearer understanding from a genetic perspective which regions affect the quality of soybeans, which is conducive to increasing yield and market competitiveness (Priyanatha et al., 2022).

3.3 Contributions to breeding programs

GWAS can not only identify key genes but also provide some genetic markers that can be directly used in breeding. For instance, these markers can be used in marker-assisted selection (MAS) and genomic selection (GS) to help breeders make decisions more quickly. Some studies have identified SNP markers related to yield, maturity period, plant height and seed weight, all of which can be applied in actual breeding to enhance work efficiency (Ravelombola et al., 2021). Some researchers have also analyzed soybean mutant resources using GWAS, identifying some mutation hotspots and key loci that affect agronomic traits, providing new ideas for breeding (Kim et al., 2022). Judging from these achievements, GWAS has played a significant role in the breeding of new soybean varieties. It can help us precisely pick out those plants with good traits, accelerate the breeding speed and make the improvement work more efficient.

4 Challenges and Limitations of GWAS in Soybean

4.1 Population structure and genetic diversity

When conducting GWAS on soybeans, some problems are often encountered. One of the main reasons is that the genetic differences among different varieties are too great. The germplasm resources of soybeans come from many places, and the genetic backgrounds of different groups are also different. If these group differences are not taken into account during the analysis, it is very easy to achieve the result of "false association". For instance, a study analyzed approximately 14 000 soybean samples and found that soybeans from China, Japan and South Korea vary greatly in genetic composition. Most of the soybean varieties in the United States can be traced back to two subgroups in China (Bandillo et al., 2015). Therefore, when conducting GWAS, it is essential to first clarify the source of the materials and their genetic background. Only in this way can errors be reduced and the gene regions related to traits be identified more accurately.

4.2 Phenotypic and environmental variability

Some important traits of soybeans, such as yield per unit area, plant height and nutrient content in seeds, are often affected by the environment. The same variety may perform very differently when grown in different places. This will make GWAS more difficult to conduct, as sometimes we cannot find those gene markers that remain stable in various environments. For instance, a study tested the same batch of soybean varieties in four places in southern Brazil. It was found that although some SNPS could recur at multiple locations, most markers only exhibited character-related relationships in certain specific environments (Contreras-Soto et al., 2017). This indicates that to obtain more reliable results, it is necessary to verify multiple times in different environments and use more rigorous statistical methods.

4.3 Complex traits and epistasis

Many important traits of soybeans, such as disease resistance, maturity time and yield, are controlled by multiple genes together. Such traits are called "complex traits". Moreover, these genes can also influence each other, and this phenomenon is called "superiority". The traditional GWAS method analyzes each SNP one by one. This approach makes it difficult to identify the interactions between genes, and many complex genetic relationships may be missed. Now, researchers have begun to use some new methods to solve this problem. For instance, GWAS based on haplotypes can analyze multiple adjacent mutation points at once, and sometimes even find results that traditional methods cannot. In addition, some subsequent analyses have also begun to integrate more types of data, such as gene expression, protein interaction information, etc. These practices can help us understand more comprehensively how traits are regulated and also clearly see how different genes interact with each other (Mortezaei and Tavallaee, 2021). However, these new methods are still not mature enough at present, especially when analyzing the issues of "one gene influencing multiple traits" (pleiotropy) and "the interaction of multiple genes", there are still many technical challenges to be solved.

5 Case Study: GWAS in Soybean Disease Resistance

5.1 Background and objectives

Soybean (*Glycine max*) is one of the most widely grown food crops in the world and also an important cash crop. However, during the planting process, soybeans are often infected by various bacteria. These diseases will lead to a decrease in output and a deterioration in quality. To reduce these losses, an effective approach is to take advantage of the inherent disease resistance of soybeans. Nowadays, many scientists use a method called GWAS (Genome-wide Association Study) to identify key gene regions related to disease resistance. This section mainly introduces some achievements made by GWAS in the study of several common soybean diseases. These diseases include leaf spot disease caused by Cardamom, brown rust disease resulting from cardamom, and sclerotinia (SSR), which can cause the stems of soybeans to rot. Through these studies, scientists hope to figure out the genetic mechanism of soybeans' disease resistance. They also hope that these discoveries can provide useful assistance for future breeding.

5.2 Methodology and findings

Nowadays, GWAS has been widely used to study the disease resistance of soybeans. For instance, there was a study that analyzed 246 samples of soybean materials using the SNP50K gene chip, with the target being *Corynebacterium carinii*. This study identified 14 related SNPS and 33 loci, which were associated with the resistance of the two strains. Six loci were consistent in both strains. In addition, researchers also combined RNA-Seq technology to analyze the changes in gene expression and identified a total of 238 genes that were significantly altered in the disease resistance response. These genes may be involved in the immune mechanism of soybeans (Figure 1).

Another study specifically focused on soybean brown rust (SBR), a disease caused by a pathogen called cardamom. Researchers analyzed 3 082 samples of soybean materials and identified many significant SNPS related to disease resistance (Table 1). Some SNP loci are located near known disease-resistant genes, such as Rpp1, Rpp2, Rpp3 and Rpp4. Some new locations were also discovered in the research, which might be related to

new disease-resistant genes. These SNPS also performed well in genomic prediction, with a higher accuracy rate than before (Xiong et al., 2023).

In the study of sclerotinosis (SSR), a GWAS combined with an upper-level analysis based on 466 soybean samples identified 58 major loci and 24 groups of gene interaction signals. All of these are related to disease resistance. The research also identified some candidate genes, which may be involved in processes such as cell wall regulation, hormone conduction, and sugar distribution. This also indicates that the resistance mechanism of SSR is rather complex (Moellers et al., 2017).

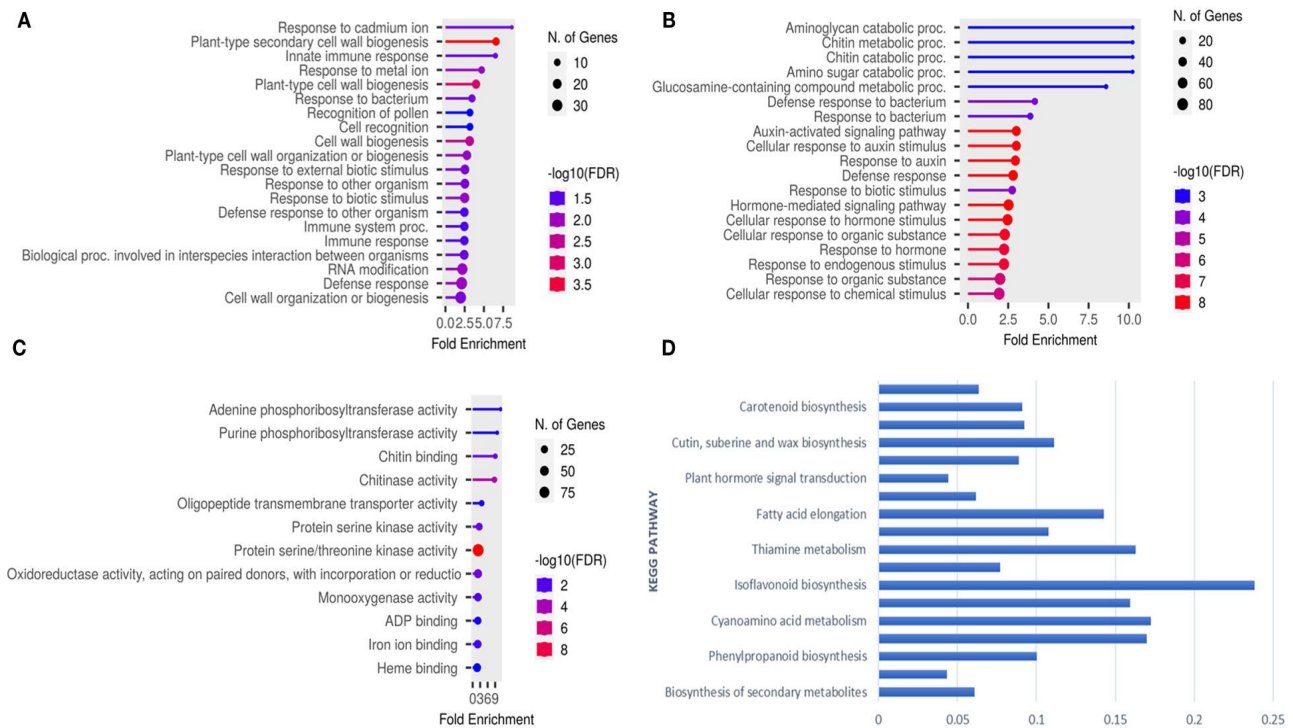


Figure 1 Go term analysis conducted on upregulated genes within resistance genotypes yielded plots illustrating (Adopted from Patel et al., 2023)

Image caption: (A) Biological processes for Bedford (B) Biological processes for Council, (C) Molecular processes for Council, and (D) Represent significantly enriched KEGG pathways identified for differentially expressed genes (DEGs) that were shared across all four genotypes (Adopted from Patel et al., 2023)

5.3 Implications for breeding

The research results of these GWAS are very helpful for soybean breeding. As long as SNPS or genes related to disease resistance are identified, breeders can use them for marker-assisted selection (MAS) or genomic selection (GS). In this way, the breeding efficiency can be significantly improved (Huang, 2024). For instance, when studying *Corynebacterium carinii*, scientists analyzed the data from GWAS and RNA-Seq together. They used these two methods together to study the disease resistance traits of soybeans. This approach enabled them to have a more comprehensive understanding of the genetic basis of soybeans and also provided a clear direction for breeding work (Patel et al., 2023). For instance, when studying soybean brown rust (SBR), scientists discovered some SNPS that were close to known disease-resistant genes. Breeders can use these SNPS to screen out resistant materials more quickly and accurately. Sclerotinia (SSR) is even more complicated. Its genetic structure is rather unique, so scientists use GWAS in combination with upper-level analysis to study it. Not only did they identify the key loci, but they also discovered many interactions between genes. These newly discovered candidate genes can also serve as key targets for subsequent breeding. They may help breed new soybean varieties that are resistant to multiple diseases and have stable resistance.

Table 1 The genes within 50 kb genomic region of the top 28 significant SBR-associated SNPs with functional annotations (Adopted from Xiong et al., 2023)

SNP	GWAS model (Ranking)	LOD	Allele Type	Gene name	Functional annotations
Gm02_7235181	SUPER(1), CMLM(5), MLMM(10)	FarmCPU,7.91	T/C	Glyma.02G083500 Glyma.02G083300	LRR; RCC1; response to bacterial origin; defense response;
Gm02_7234594	SUPER(2), MLMM(11)	7.61	C/T	Glyma.02G084100	structural constituent of cell wall
Gm02_7315227	SUPER(3), GLM, Blink(5), MLMM(6)	MLM,7.52	G/A	Glyma.02G084100 Glyma.02G084900	RCC1 repeat; Ankyrin repeat family protein/domain
Gm03_38913029	GLM, MLM, Blink (2), FarmCPU, CMLM(3), MLMM(7), GLM	6.85	T/C	Glyma.03G175800 Glyma.03G177400 Glyma.03G175300	Response to aluminum ion; cell wall; ABC transporter
Gm04_45884688	MLM, Blink(7), MLMM(16), CMLM(26)	SUPER(15),6.23	T/C	Glyma.04g188000	LRR
Gm04_46003059	SUPER(20), MLMM(24)	6.03	G/A	Glyma.04G189300, Glyma.04g189500	Membrane; Cytochrome P450
Gm04_46295839	SUPER(16), MLMM(18)	6.08	C/T	Glyma.04G192300	Cell wall organization; cellular membrane fusion;
Gm04_46389651	SUPER(22), MLMM(27)	5.94	C/T		NAC domain
Gm04_47132429	MLMM(4), FarmCPU, CMLM(6), GLM, MLM, Blink(13), SUPER(25)	5.78	T/C	Glyma.04G211100, Glyma.04G212000	
Gm06_36808946	SUPER(6), GLM, Blink(9), CMLM(34)	MLM,6.73	G/A	Glyma.06G232500	Response to molecule of bacterial origin
Gm08_43955878	FarmCPU, CMLM(19), SUPER(32), MLMM(33)	5.61	A/C	Glyma.08g319300, Glyma.08G321700	LRR; response to abscisic acid stimulus/cold/water deprivation
Gm09_1944730	MLMM(2), SUPER(27)	5.77	C/A	Glyma.09G024700	LRR-RLKs
Gm09_1943831	MLMM(3), SUPER(28)	5.73	G/A		
Gm09_1951644	FarmCPU, CMLM, MLMM(1), GLM, MLM, Blink(4), SUPER(18)	10.07	T/G		
Gm10_5573877	SUPER(5), MLMM(12), MLM, Blink(14)	GLM,6.73	C/T	Glyma.10G060100, Glyma.10G060200, Glyma.10G060600	Respiratory burst involved in defense response, response to bacterium/chitin; cell wall organization
Gm10_5573007	SUPER(7), MLMM(15)	6.58	C/T		
Gm10_5559592	SUPER(9), MLMM(20)	6.48	C/A		
Gm10_5541691	SUPER(33), MLMM(44)	5.60	C/T		
Gm10_5578693	SUPER(23), MLMM(32)	5.93	G/A		
Gm10_39142024	GLM, MLM, Blink(1), MLMM(8), FarmCPU, CMLM(14)	7.12	C/T	Glyma.10g157500	LRR-RLKs, regulation of plant immunity
Gm10_39147121	MLMM(9), SUPER(21)	6.02	T/G		
Gm12_28136735	SUPER(4), GLM, Blink(8), MLMM(39)	MLM,7.03	G/A	Glyma.12G160100, Glyma.12G160400	NAC domain protein; Cytochrome P450
Gm13_16350701	FarmCPU, CMLM(16), MLM, Blink(23), SUPER(29)	GLM,5.63	T/C	Glyma.13G064500	F-box and WD40 domain protein, disease resistance protein
Gm14_2492139	GLM, MLM, Blink(6), SUPER(13), CMLM(25), MLMM(26)	6.26	A/C	Glyma.14G034200, Glyma.14G040000	RCC1 family protein; LRR-RLKs
Gm14_6185611	MLMM(28), GLM, MLM, Blink(46)	SUPER(36),5.51	C/T	Glyma.14g073300, Glyma.14G073800	F-box domain; regulation of defense response

Continuing Table1

SNP	GWAS model (Ranking)	LOD	Allele Type	Gene name	Functional annotations
Gm16_4935328	GLM, MLM, Blink(10),5.61 MLMM(22), SUPER(31), FarmCPU, CMLM(32)		T/G	Glyma.16G051800, Glyma.16G052200	NAC domain protein; LRR- RLKs
Gm19_44734953	GLM, MLM, Blink(3),6.02 FarmCPU, CMLM(4), MLMM(25)		G/A	Glyma.19G189900, Glyma.19G190200, Glyma.19G190800	Defense response to bacterium; LRR- RLKs; plant-type cell wall
Gm20_36724867	FarmCPU, CMLM(2)	6.54	C/T	Glyma.20G124700	QSOX1 regulates plant immunity

6 Future Directions for GWAS in Soybean

6.1 Integrating multi-omics data

If different types of data can be grouped together, such as the genome (DNA), transcriptome (RNA expression), proteome and metabolome (metabolites), the results of GWAS will be more convincing. This method can help us understand how complex traits come about from multiple perspectives and also narrow down some overly large associated regions. Sometimes, chain disequilibrium can make the range of results very wide. In the research of rapeseed, scientists used multi-omics analysis methods to overlap and compare QTLs found by different omics, successfully locating important genes related to nutritional metabolism and growth (Knoch et al., 2023). In soybean research, if a similar approach is also adopted, more useful candidate genes and signaling pathways may be discovered. In this way, the judgment of traits during breeding will be more accurate, and the efficiency of selection and breeding can also be improved.

6.2 Advancements in computational tools

Nowadays, technology is becoming increasingly advanced, and there are already many new tools available for analyzing GWAS data. Especially some methods with machine learning (ML) capabilities perform very well when dealing with complex data. Previous GWAS methods did not perform well in analysis when there was an extremely large amount of data or when the genetic background of crops was relatively simple (such as soybeans). However, machine learning algorithms such as support vector regression (SVR) and random forest (RF) perform better in finding quantitative trait loci (QTL) (Yoosefzadeh-Najafabadi et al., 2021; 2023). These methods can also handle large volumes of data and make it easier to identify the genetic patterns behind complex traits. These tools also provide very practical technical support for research on genomic breeding.

6.3 Improving phenotyping techniques

GWAS not only requires genetic data, but trait (phenotypic) data is equally important. Nowadays, an increasing number of studies are employing high-throughput phenotypic techniques to collect trait information from samples. Compared with traditional manual scoring, the new technology is more efficient and has less error. High-throughput phenotypes generally do not damage plants and can be continuously monitored, facilitating the observation of changes in traits over time (Xiao et al., 2021). For instance, scientists have employed deep learning (DL) technology to identify soybean disease conditions, with an accuracy rate even higher than that of manual judgment (Rairdin et al., 2022). If these advanced phenotypic methods are combined with GWAS, not only can the accuracy of the analysis results be improved, but also the traits with good performance can be selected more quickly, which is particularly helpful for the breeding of new soybean varieties.

7 Concluding Remarks

Nowadays, when scientists study the genetic mechanism of soybeans, they often use a method called GWAS (Genome-wide Association Study). This method has been of great help. Through GWAS, researchers have identified many quantitative trait loci (QTLs) and also discovered some candidate genes related to yield, quality, plant height and disease resistance. For instance, one study utilized soybean varieties from Canada and China. They used GWAS to identify five gene regions related to yield, protein content and oil content. These findings are very helpful for soybean breeding. Another study combined the results of 73 GWAS for a pooled analysis, which

is known as a "meta-analysis" (meta-GWAS). A total of 393 genomic regions were identified this time, among which 483 QTLs were discovered. This indicates that when the results of multiple studies are analyzed together, more useful information can be identified. In addition to these traditional practices, nowadays some people have begun to use GWAS based on haplotypes. It is somewhat different from the method of analyzing only a single SNP and is more suitable for studying the genetic background related to complex traits. In addition, some studies have also introduced machine learning methods, such as using artificial intelligence to find QTLs. This method not only improves the accuracy of the analysis, but also can locate key genes more quickly. With the continuous development of computing technology, the role of GWAS in soybean genetic research will definitely become increasingly significant.

Next, there are still many areas where the research on soybean GWAS can be further advanced: First, try more advanced statistical models. For instance, the new model 3VmrMLM not only analyzes genes but also takes environmental factors into account. This can better explain the manifestation of traits in different environments. Second, incorporate structural variation and k-mer analysis. Structural variations include gene insertions or deletions, and K-mers are some very short DNA fragments. By using these methods, some new functional regions may be discovered, and the genetic characteristics of soybeans can also be understood more comprehensively. Third, encourage more meta-GWAS to be conducted. It is to integrate and analyze the data from different teams and experiments. The result obtained in this way is more stable and reliable. Finally, and most importantly: These research results should be applied to actual breeding. Combining the results of GWAS with genomic selection (GS) or marker-assisted selection (MAS) can more quickly screen out soybean materials with high yield and strong adaptability. Overall, technology is still advancing. There is still much room for GWAS in the study of soybean genetics and the improvement of breeding efficiency. In the future, it will remain an important tool for promoting the improvement of soybean varieties.

Acknowledgments

The authors thank anonymous reviewers for their suggestions and comments that were useful for improving our paper's presentation.

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

- Almeida-Silva F., Moharana K., Machado F., and Venancio T., 2020, Exploring the complexity of soybean (*Glycine max*) transcriptional regulation using global gene co-expression networks, *Planta*, 252(6):104.
<https://doi.org/10.1007/s00425-020-03499-8>
- 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*, 12: 431-516.
https://doi.org/10.1007/978-3-030-23400-3_12
- Bandillo N., Jarquín D., Song Q., Nelson R., Cregan P., Specht J., and Lorenz A., 2015, A population structure and genome-wide association analysis on the usda soybean germplasm collection, *The Plant Genome*, 8(3): 24.
<https://doi.org/10.3835/plantgenome2015.04.0024>
- Bhat J., Adeboye K., Ganie S., Barmukh R., Hu D., Varshney R., and Yu D., 2022, Genome-wide association study, haplotype analysis, and genomic prediction reveal the genetic basis of yield-related traits in soybean (*Glycine max* L.), *Frontiers in Genetics*, 13: 953833.
<https://doi.org/10.3389/fgene.2022.953833>
- 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>
- Cortes L., Zhang Z., and Yu J., 2021, Status and prospects of genome-wide association studies in plants, *The Plant Genome*, 14(1): e20077.
<https://doi.org/10.1002/tpg2.20077>
- Huang W.Z., 2024, The current situation and future of using GWAS strategies to accelerate the improvement of crop stress resistance traits, *Molecular Plant Breeding*, 15(2): 52-62.
<http://dx.doi.org/10.5376/mpb.2024.15.0007>

- Jiang C., 2024, Genetic mechanisms of crop disease resistance: new advances in GWAS, *Plant Gene and Trait*, 15(1): 15-22.
<http://dx.doi.org/10.5376/pgt.2024.15.0003>
- Kim D., Lyu J., Kim J., Seo J., Choi H., Jo Y., Kim S., Eom S., Ahn J., Bae C., and Kwon S., 2022, Identification of loci governing agronomic traits and mutation hotspots via a GBS-based genome-wide association study in a soybean mutant diversity pool, *International Journal of Molecular Sciences*, 23(18): 10441.
<https://doi.org/10.3390/ijms231810441>
- Kim S., Tayade R., Kang B., Hahn B., Ha B., and Kim Y., 2023a, Genome-wide association studies of seven root traits in soybean (*Glycine max* L.) landraces, *International Journal of Molecular Sciences*, 24(1): 873.
<https://doi.org/10.3390/ijms24010873>
- Kim W., Kang B., Kang S., Shin S., Chowdhury S., Jeong S., Choi M., Park S., Moon J., Ryu J., and Ha B., 2023b, A genome-wide association study of protein, oil, and amino acid content in wild soybean (*Glycine soja*), *Plants*, 12(8): 1665.
<https://doi.org/10.3390/plants12081665>
- Knoch D., Meyer R., Heuermann M., Riewe D., Peleke F., Szymański J., Abbadi A., Snowdon R., and Altmann T., 2023, Integrated multi-omics analyses and genome-wide association studies reveal prime candidate genes of metabolic and vegetative growth variation in canola, *The Plant Journal*, 117(3): 713-728.
<https://doi.org/10.1111/tpj.16524>
- Korte A., and Farlow A., 2013, The advantages and limitations of trait analysis with GWAS: a review, *Plant Methods*, 9: 29.
<https://doi.org/10.1186/1746-4811-9-29>
- Lemay M., Ronne M., Bélanger R., and Belzile F., 2023, *k*-mer-based GWAS enhances the discovery of causal variants and candidate genes in soybean, *The Plant Genome*, 16(4): e20374.
<https://doi.org/10.1002/tpg2.20374>
- Li S., Cao Y., Wang C., Yan C., Sun X., Zhang L., Wang W., and Song S., 2023, Genome-wide association mapping for yield-related traits in soybean (*Glycine max*) under well-watered and drought-stressed conditions, *Frontiers in Plant Science*, 14: 1265574.
<https://doi.org/10.3389/fpls.2023.1265574>
- Moellers T., Singh A., Zhang J., Brungardt J., Kabbage M., Mueller D., Grau C., Ranjan A., Smith D., Chowda-Reddy R., and Singh A., 2017, Main and epistatic loci studies in soybean for *Sclerotinia sclerotiorum* resistance reveal multiple modes of resistance in multi-environments, *Scientific Reports*, 7: 3554.
<https://doi.org/10.1038/s41598-017-03695-9>
- Mortezaei Z., and Tavallaee M., 2021, Recent innovations and in-depth aspects of post-genome wide association study (Post-GWAS) to understand the genetic basis of complex phenotypes, *Heredity*, 127: 485-497.
<https://doi.org/10.1038/s41437-021-00479-w>
- Patel S., Patel J., Bowen K., and Kobernick J., 2023, Deciphering the genetic architecture of resistance to *Corynespora cassiicola* in soybean (*Glycine max* L.) by integrating genome-wide association mapping and RNA-Seq analysis, *Frontiers in Plant Science*, 14: 1255763.
<https://doi.org/10.3389/fpls.2023.1255763>
- 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>
- Rairdin A., Fotouhi F., Zhang J., Mueller D., Ganapathysubramanian B., Singh A., Dutta S., Sarkar S., and Singh A., 2022, Deep learning-based phenotyping for genome wide association studies of sudden death syndrome in soybean, *Frontiers in Plant Science*, 13: 966244.
<https://doi.org/10.3389/fpls.2022.966244>
- 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>
- Shook J., Zhang J., Jones S., Singh A., Diers B., and Singh A., 2021, Meta-GWAS for quantitative trait loci identification in soybean, *G3: Genes|Genomes|Genetics*, 11(7): jkab117.
<https://doi.org/10.1093/g3journal/jkab117>
- 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>
- Xiao Q., Bai X., Zhang C., and He Y., 2021, Advanced high-throughput plant phenotyping techniques for genome-wide association studies: a review, *Journal of Advanced Research*, 35: 215-230.
<https://doi.org/10.1016/j.jare.2021.05.002>
- Xiong H., Chen Y., Pan Y., Wang J., Lu W., and Shi A., 2023, A genome-wide association study and genomic prediction for *Phakopsora pachyrhizi* resistance in soybean, *Frontiers in Plant Science*, 14: 1179357.
<https://doi.org/10.3389/fpls.2023.1179357>

- Yang J., Jiang H., Yeh C., Yu J., Jeddeloh J., Nettleton D., and Schnable P., 2015, Extreme-phenotype genome-wide association study (XP-GWAS): a method for identifying trait-associated variants by sequencing pools of individuals selected from a diversity panel, *The Plant Journal*, 84(3): 587-596.
<https://doi.org/10.1111/tpi.13029>
- Yoosefzadeh-Najafabadi M., Torabi S., Torkamaneh D., Tulpan D., Rajcan I., and Eskandari M., 2021, 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>
- Zhang J., Song Q., Cregan P., Nelson R., Wang X., Wu J., and Jiang G., 2015, Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm, *BMC Genomics*, 16: 217.
<https://doi.org/10.1186/s12864-015-1441-4>



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.
