Research Insight **Open Access**

Key Loci Identified by GWAS for Agronomic Traits in Soybean

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

Legume Genomics and Genetics, 2025 Vol.16, No.1 doi: 10.5376/lgg.2025.16.0005
Received: 01 Jan., 2025
Accepted: 12 Feb., 2025
Published: 27 Feb., 2025

Preferred citation for this article:
Zhou X.X., and Guo T.X., 2025, Key loci identified by GWAS for agronomic traits in soybean, Legume Genomics and Genetics, 16(1): 44-53 (doi: 10.5376/lgg.2025.16.0005)

Abstract Soybean (*Glycine max* [L.] Merr.) holds an important position worldwide due to its high protein and oil content, and is a key source of human consumption and animal feed. However, soybean cultivation is confronted with the challenges of climate change and the need to increase yield and stress resistance. Genome-wide association studies (GWAS) are of great value in identifying key genetic loci associated with complex agronomic traits, including yield, stress resistance, nutritional quality and disease resistance. This review summarizes the progress made in soybean genomics through GWAS and elaborates on the loci and candidate genes that affect traits such as seed composition, plant height, and root development. Integrating the findings of GWAS into molecular breeding strategies such as marker-assisted selection (MAS) and genomic selection (GS) can promote the development of high-yield and climate-adapted soybean varieties. Furthermore, the combination of GWAS with advanced genomic tools and computational methods provides insights for future research. These research findings contribute to the sustainable improvement of soybean productivity to address the urgent need for global food sec。urity under environmental challenges

Keywords Soybean; GWAS; Agronomic traits; Molecular Breeding; Stress tolerance

1 Introduction

When it comes to soybeans (*Glycine max* [L.] Merr.), this thing really feeds a lot of people. Despite its small size, it contains a considerable amount of protein and fat. It is indispensable from edible oil to feed (Sonah et al., 2015; Kim et al., 2023). In fact, in places like Southeast Asia and Africa, soybeans have long been a traditional food and are widely used in industry. However, nowadays there are more and more places around the world that need soybeans (Rani et al., 2023), and the original output alone may not be sufficient. When it comes to this, increasing production and quality becomes particularly important - although exactly how to do it still depends on the actual situation.

Growing soybeans is not that simple nowadays. The weather is getting more and more unpredictable, with droughts and floods alternating, yet the global demand for soybeans is still on the rise. When it comes to solutions, the key actually lies in the soybeans themselves - such as the components in the seeds, how tall the plants can grow, and how deep the roots are (Van et al., 2017; Kim et al., 2023). If these traits are improved well, perhaps more durable and nutrient-rich soybeans can be grown. Of course, relying solely on old methods for gradual breeding is definitely not enough (Rani et al., 2023). Nowadays, molecular breeding techniques are emphasized, although they may not be so easy to operate in practice.

Nowadays, there is an interesting method for conducting soybean research called genome-wide Association studies (GWAS). To put it simply, it is to find the patterns of genetic variations in a large number of soybean samples. This technology is particularly good at discovering the small details that affect the growth and yield of soybeans, such as the gene loci that determine the protein level of seeds or the development degree of root systems (Sonah et al., 2015). However, GWAS alone may not be accurate enough. Therefore, researchers often combine it with other techniques - such as genotypic sequencing (GBS) or SNP chips (Almeida-Silva et al., 2020). Although the operation is rather complicated, it can indeed help the breeding work avoid many detours.

How to identify the key loci that control important characteristics from soybean genes. We plan to use the method of genome-wide association studies (GWAS) to compare the genetic data of various soybean varieties with their actual growth performance - with a focus on tangible traits such as seed composition, plant height and root system. Of course, merely identifying suspicious SNP markers is not enough (although these markers are now quite convenient to measure). It is also necessary to draw a QTL graph to verify them again, and at the same time, see if they can match the previously reported QTLS (after all, previous research cannot be in vain). In fact, to put it bluntly, it's all about figuring out which genes control these traits, so that we can avoid some detours when breeding in the future. If it really works out, it might be possible to cultivate soybeans that are both high-yielding and can withstand extreme weather. These days, climate change is so rampant. It's always a good thing to be able to grow more grain.

2 Agronomic Traits of Interest in Soybean

2.1 Yield and yield-related traits

When it comes to soybean yields, it's really not something that can be solved merely by fertilizing and watering. When researchers used GWAS to dig through genetic data, they found that some SNP markers were particularly interesting - such as those that simultaneously affected seed weight, plant height and final yield (Sonah et al., 2015; Shook et al., 2021). Interestingly, these connections can all hold true in different planting environments. Even more amazing is that someone has identified the gene locus that directly controls the 100-grain weight (Copley et al., 2018), which is a real key factor affecting the harvest. However, to be honest, although many related sites have been found now, how to combine them specifically to grow high-yield soybeans still needs further consideration.

2.2 Stress tolerance traits

Nowadays, the most troublesome thing for growing soybeans is that the weather is too dry. These days, rain is becoming less and less reliable. Researchers rummaged through the soybean gene bank and actually found some tricks - for example, the gene called *GmNFYB17* (Sun et al., 2022) can make the root system of soybeans more developed when water is scarce, and the yield naturally increases. It's quite interesting to say that after they crossed drought-tolerant and drought-intolerant varieties, a more drought-resistant strain did emerge in the offspring (Ouyang et al., 2022), which indicates that drought resistance can indeed be inherited. However, on the other hand, compared with drought resistance research, people's understanding of the heat tolerance of soybeans is still far behind. But since the drought-resistant genes can be identified, it should also work to study heat tolerance in this way. The weather is so abnormal now. If we could cultivate soybeans that are both drought-resistant and heat-tolerant, it would really be a great help to the farmers. In fact, the most practical use of these discoveries is that they can help breeders avoid detours and develop new varieties that can adapt to harsh environments earlier - after all, food security is no small matter.

2.3 Nutritional quality traits

Whether soybeans taste good and are nutritious mainly depends on the oil and protein in the seeds. When researchers scanned genes using GWAS, they found that some loci were particularly interesting - they directly determined the lipid and protein contents (Sonah et al., 2015). More detailed research also uncovered genes that control the composition of fatty acids and amino acids (Shook et al., 2021), which is like finding the switch to adjust the nutritional formula of soybeans. To be honest, although we now know which genes are at play, to truly cultivate an ideal variety, we still need to fully understand the combination of these genes. After all, the matter of fat and protein content is not something that can be settled by just one or two genes.

2.4 Disease resistance traits

Whether soybeans can grow healthily or not, their disease resistance is very crucial. Recently, I did a GWAS scan and found quite a few tricks - one study identified 33 candidate genes related to disease resistance in one go (Shook et al., 2021), which shows that improving disease resistance at the genetic level is completely feasible. In fact, it's not just disease resistance. In recent years, using GWAS to find genes has been quite effective. From yield to stress resistance and then to nutritional quality, the important loci that should be found have basically been identified (Yoosefzadeh-Najafabadi et al., 2021). However, to be honest, knowing which genes are useful is only the first step. How to truly apply these discoveries to breeding practice might be the tough nut to crack next.

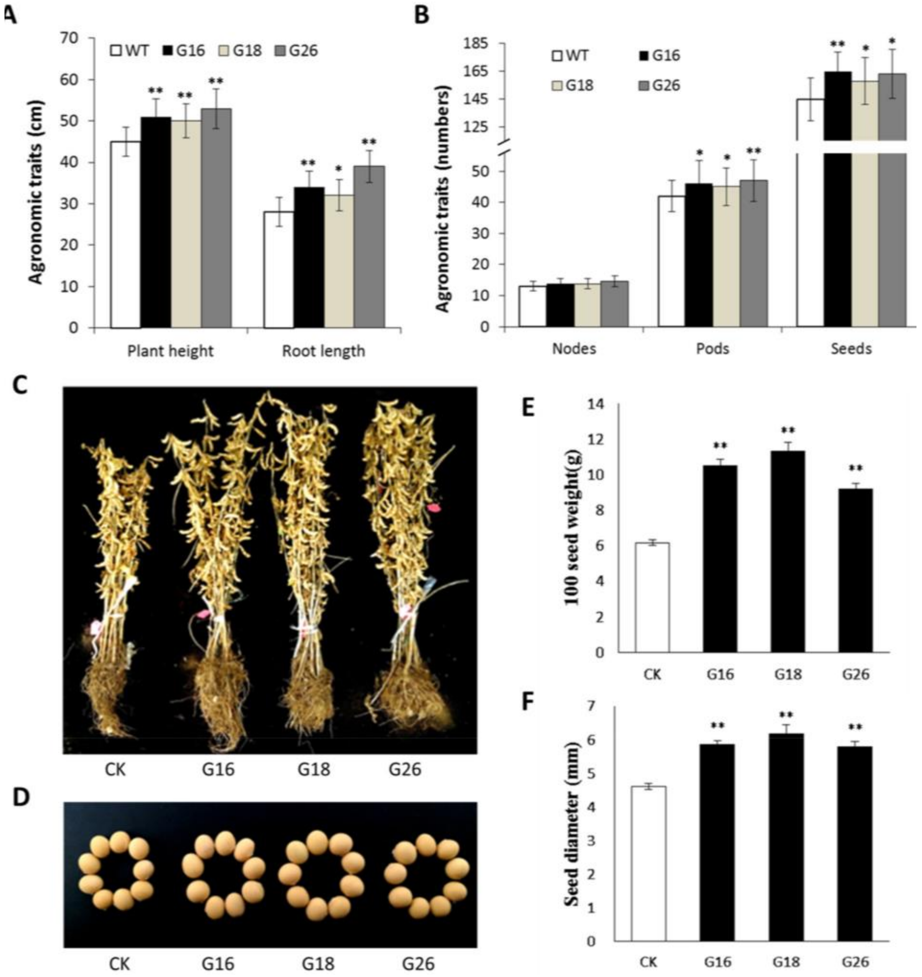

Figure 1 Investigation of plant and agronomic traits of transgenic lines (Adopted from Sun et al., 2022)

Image caption: *: p-value ≤ 0.05; **: p-value ≤ 0.01; (A) Plant height and root length of G16, G18, G26 and CK; (B) The number of branches, nodes, pods and seeds in the transgenic lines and non-transgenic plants; (C) The phenotype of *GmNFYB17* after harvest; (D) Comparison of seed size among *GmNFYB17* lines and CK; (E) The 100-seed weight of the transgenic lines and non-transgenic plants; (F) The diameter of transgenic and non-transgenic seeds (Adopted from Sun et al., 2022)

3 Methodology of GWAS in Soybean

3.1 Overview of GWAS and its relevance to soybean research

When it comes to the study of soybean genes, GWAS has really been of great help nowadays. In essence, this technology involves identifying genetic differences among various soybean varieties to see which gene segments can affect the actual planting performance. For instance, some genes control when soybeans mature, while others are linked to the strength of disease resistance (Contreras-Soto et al., 2017). Interestingly, the same method can also be used to study yielt-related traits (Copley et al., 2018), which is much better than blind hybridization breeding in the past. In fact, as early as 2015, studies proved that GWAS could accurately locate the gene regions that control important agronomic traits of soybeans (Sonah et al., 2015). However, to be honest, although technology has advanced now, merely relying on GWAS to locate genes is not enough to truly cultivate ideal varieties. The subsequent verification work is what really requires effort.

3.2 GWAS workflow

When it comes to conducting GWAS research on soybeans, choosing the right population type is crucial. Take the diversity group for example. This method of studying various genotypes together (Sonah et al., 2015) can indeed discover many interesting genetic variations. However, sometimes in order to verify the results, it is necessary to use the parent population for assistance - that is, to specifically hybridize two parents with particularly large differences (Copley et al., 2018), so that the gene localization is more precise.

When it comes to genetic testing technology, the two main types currently used are GBS and SNP chips. The GBS technology is quite powerful. It can scan the entire genome (Kim et al., 2022), and is particularly suitable for finding genes of those complex traits. Sure, if cost is taken into account, the SoySNP50K chip (Zhang et al., 2015) is also an affordable option, but the detection sites are not as comprehensive.

Finally, it is necessary to mention the statistical methods. The commonly used MLM model nowadays (Zhang et al., 2015) can indeed solve the interference caused by the group structure. However, researchers are still constantly improving. New methods such as FarmCPU (Priyanatha et al., 2022) and BLINK (Ravelombola et al., 2021) are faster in calculation and more accurate in finding gene loci. To put it bluntly, GWAS research requires mastering the three key techniques of sample size, technology and statistics.

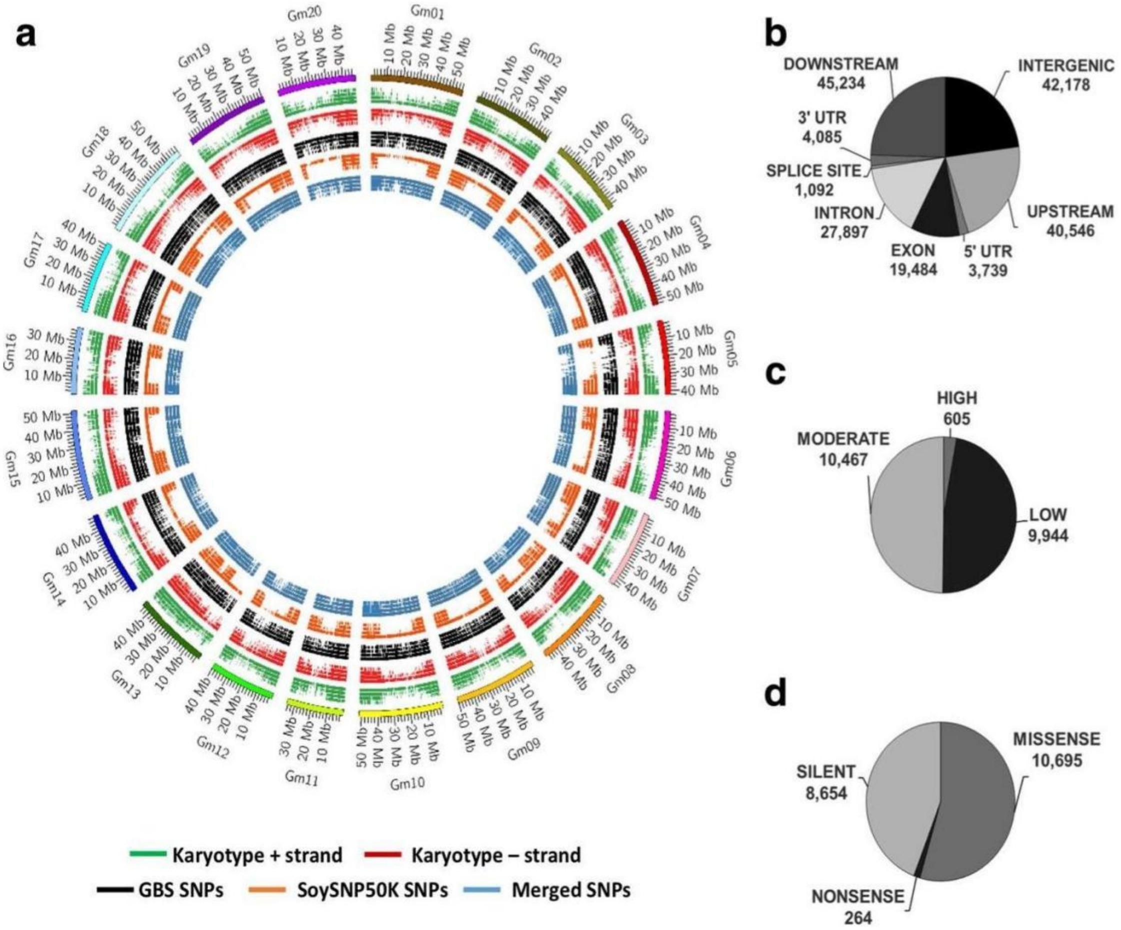

Figure 2 SNP distributions across the soybean genome (v2) and SNP effects within the population of plant introduction genotypes (Adopted from Copley et al., 2018)

Image caption: a Gene and SNP distributions used for genotyping across the soybean chromosomes. From the outer to inner circle: Soybean chromosomes 1 to 20; gene locations on the positive and negative chromosome strands; and GBS, SoySNP50K microarray and the merged data set SNP locations. b Distribution of SNPs based on genomic region within the merged data set. c Predicted SNP effects based on degree of impact within the merged data set. d Predicted SNP effects based on function class for SNPs located within coding regions within the merged data set (Adopted from Copley et al., 2018)

3.3 Challenges in GWAS for soybean (population structure, false positives)

GWAS is indeed useful in soybean research, but it is not omnipotent. The most headache-inducing issue is the population structure problem - simply put, different subgroups may be mixed in the sample, which can lead to deviation in the analysis results. However, there are now solutions, such as using PCA analysis or kinship matrix for correction (Zhang et al., 2015), and the recently released FarmCPU model (Priyanatha et al., 2022) is also quite good at handling such problems. Another common pitfall is false positives, where unreliable results keep popping up. At this point, the significance threshold needs to be adjusted more strictly, or re-validated in another population (Sonah et al., 2015). If conditions permit, double confirmation of QTL localization should also be conducted (Kim et al., 2022). Although there are many problems, as long as all these pitfalls are eliminated, GWAS can still help us dig out a lot of useful genetic information, which is of great help to breeding work. To put it bluntly, no matter how good a tool is, it depends on how people use it. The key is to do every link well.

4 Genetic Loci Underpinning Yield, Stress Tolerance, Nutritional Value, and Disease Resistance in Soybean

4.1 Loci associated with yield-related traits

GWAS has indeed been of great help in studying the size of soybean seeds and the number of pods. The genome was scanned using GBS technology and the results were quite interesting - 1 to 8 key loci were discovered. These sites not only controlled seed weight and plant height (Sonah et al., 2015), but also overlapped with the previously reported QTL regions, indicating that the right location was found. Even better, the same method also identified the gene loci that affect the flowering time and maturity period. To ensure there were no mistakes, the researchers deliberately used two-parent hybrid populations for verification, and the results showed that these loci were indeed reliable. However, to be fair, although these key loci have been identified, how exactly these traits are regulated may still require further exploration.

4.2 Loci linked to abiotic stress tolerance

Although there is currently a lack of direct GWAS evidence of drought resistance in soybeans, research on other crops may provide clues. For instance, the rice field is quite interesting. Someone identified 82 meta-Qtls related to drought resistance through meta-analysis (Yang et al., 2020). This research approach might be applicable to soybeans. When it comes to stress resistance, there are some findings in soybean salt resistance studies themselves - there is a major locus on chromosome 3 that is particularly prominent (Do et al., 2019), but there are also some genes on chromosomes 1, 8 and 18 that affect the salt stress response, such as those that control the degree of leaf damage and chlorophyll content. These results indicate that salt resistance is a rather complex issue that cannot be resolved by a single gene, but at least it points out a clear path for molecular marker-assisted breeding. If the research on salt resistance and drought resistance can be combined, perhaps some common stress resistance mechanisms can be discovered, which should be helpful for breeding new varieties that adapt to climate change.

4.3 Loci associated with nutritional traits

The GBS-GWAS technology has indeed found a treasure in improving the quality of soybeans. When analyzing the seed protein content, researchers not only identified key sites (Sonah et al., 2015), but more excitingly, these sites were highly consistent with the previously reported QTL regions - indicating that we may have grasped the core genes that determine the protein content. Coincidentally, a similar pattern was also discovered when analyzing the fat content (Wang et al., 2020). These overlapping genetic loci are simply natural quality control switches. However, interestingly, although the genetic loci for both proteins and fats have been identified, in actual breeding, one still needs to pay attention to the possible waxing and waning relationship between them. After all, from a molecular perspective, the metabolic pathways of these two traits are very likely to influence each other. The most practical value of these findings lies in the fact that in the future, when cultivating high-protein or high-oil soybean varieties, we can directly target these key sites for precise selection and breeding.

4.4 Loci contributing to disease resistance

The GBS-GWAS method is also quite reliable for identifying disease-resistant genes. Although the data at hand does not specifically state where the resistance sites of soybean cyst nematodes are, following this train of thought is definitely correct. Just think about it, as long as the markers of the whole genome are densely dotted and the disease phenotypic data are carefully compared (Abdelraheem et al., 2020), even a tough bone like Phytophthora root rot can find resistance sites (Shook et al., 2021). To put it bluntly, the key lies in two points: one is that there should be a sufficient number of SNP markers, and the other is that QTL positioning should be used for further verification. With such a combination of measures in place, no matter how deeply hidden the disease-resistant genes are, they can be exposed.

In fact, it's not just disease resistance. The approach of GWAS combined with QTL mapping has proven effective time and again in soybean research. From yield to stress resistance, from nutritional quality to disease resistance, almost all the important sites that need to be identified have been thoroughly investigated. The most practical use of these discoveries is that they have drawn a precise "treasure map" for molecular breeding - in the future, if you want to improve a certain trait, you can simply follow the map, avoiding the need to search for a needle in a haystack in the vast genome.

5 Applications of GWAS Results in Soybean Breeding

5.1 Marker-assisted selection (MAS) using identified loci

When it comes to soybean breeding nowadays, it is getting more and more precise. Remember that in the past, seed selection relied entirely on experience. Now, with the technology of marker-assisted selection (MAS), one can directly look at the genetic markers - such as those loci that control protein and lipid content (Sonah et al., 2015; Huang, 2024). In fact, GWAS has already helped us find many useful markers, not only seed components, but also traits such as plant height and seed size have corresponding loci (Shook et al., 2021). Although there are still some troubles in actual operation, applying these markers to MAS is indeed effective. At least it can help breeding work avoid detours and the speed of new varieties coming out is much faster. However, to be fair, having marks alone is not enough. In the end, it also depends on whether the field performance can truly achieve the expected results.

5.2 Genomic selection (GS) and its integration with GWAS

Nowadays, there is an interesting technique in soybean breeding called genomic selection (GS). To put it simply, it involves scanning the entire genome to predict which soybeans are more worthy of cultivation. However, using GS alone might not be accurate enough, so the researchers came up with a brilliant idea - adding the important loci found in GWAS. For instance, some people incorporated SNP markers that control protein content into the GS model (Qin et al., 2022), and as a result, the efficiency of selecting high-protein varieties increased significantly. Interestingly, this combination of measures is much more reliable than randomly selecting some genetic markers, indicating that the combination of GWAS and GS technologies can indeed produce an effect greater than the sum of its parts. Although various parameters still need to be debugged in the actual operation, at least it is now known that this approach is feasible. In the future, when cultivating new varieties, we should be able to avoid some detours.

5.3 Development of climate-resilient varieties

It's getting harder and harder to grow soybeans nowadays - the weather is either too dry or too hot, which makes the yield very unstable. Fortunately, GWAS technology has helped us identify some key genes, such as those loci that control flowering time (Kim et al., 2022), and these findings might come in handy. Although it still needs to be explored exactly how to use them, applying these stress resistance gene markers to breeding can at least increase the chances of new varieties winning in bad weather (Ravelombola et al., 2021). Ultimately, nowadays, in breeding, it is not only necessary to pursue high yields, but also to find ways to enable soybeans to withstand increasingly abnormal weather conditions; otherwise, even having enough to eat will be a problem in the future.

5.4 Enhancing nutritional quality through identified genes

When it comes to soybeans, in the final analysis, it all depends on their nutritional value - after all, so many people around the world rely on them to supplement protein and oil. In recent years, GWAS studies have indeed unearthed many valuable genes, such as those sites that directly affect protein content and amino acid composition (Shook et al., 2021), and new discoveries were added last year (Yoosefzadeh-Najafabadi et al., 2023). However, interestingly, although so many key loci have been identified, it still depends on how these genes are combined to truly cultivate high-quality varieties. Just like someone specifically focuses on high-protein sites for breeding (Qin et al., 2022), the soybeans cultivated in this way are more suitable for use as feed or processed foods.

Breeding nowadays is much smarter than before. Applying these markers found in GWAS to MAS or GS (Stewart-Brown et al., 2019) can not only address climate change but also improve nutritional quality. But to be honest, although these technologies are powerful, ultimately it still depends on how the soybeans grow in the fields are (McLeod et al., 2023). After all, no matter how good the genes are, they must stand the test of actual cultivation.

6 Limitations and Future Directions

6.1 Limitations of GWAS (population size, environmental interactions)

GWAS studies on soybean genes have indeed achieved considerable success over the years, but there are also many problems in actual operation. Take the sample size for example. The population size used in many studies is really too small - for instance, one study found some key sites using 304 soybean strains (Sonah et al., 2015), but the researchers themselves admitted that the sample size would have to be increased to make the results more reliable (Shook et al., 2021). This is like conducting a public opinion poll. If too few people are asked, the result is bound to be unreliable. What's more troublesome is that the important traits of soybeans, such as yield and plant height, are constantly affected by the environment. It's often hard to tell whether it's the genes at play or the weather causing trouble. Fortunately, there is now a meta-GWAS method that can analyze data from multiple studies together. The conclusions drawn in this way are indeed more convincing. Ultimately, no matter how powerful GWAS is, it is still a tool. The key lies in how people use it.

6.2 Integration with other genomic tools (QTL mapping, CRISPR)

Although GWAS is useful, it is really not enough to rely on it alone. Nowadays, when conducting research, a combination of measures is emphasized - for instance, incorporating QTL mapping as an assistant. Just like those loci found in previous GWAS, when re-validated in the parent population using QTL (Sonah et al., 2015), the results were indeed much more reliable. However, the most remarkable one is the CRISPR technology that has become popular in the past two years. It is simply a "microscope" for genetic research (Contreras-Soto et al., 2017). Want to confirm which gene is useful? Just edit it out and see the effect (Kim et al., 2022). Although each of these technologies has its own tricks, when used together, they can indeed complement each other's shortcomings. To put it bluntly, nowadays, when studying genes, one not only needs to be able to "find" them but also "verify" them, so as to truly contribute to breeding work.

6.3 Opportunities for multi-trait GWAS and meta-analysis

There is a new trend in the study of soybean genes now - the joint efforts of multi-trait GWAS and meta-analysis. This trick is quite interesting. It can identify the "versatile" genes that affect multiple traits at once. For instance, one study aggregated data from 73 independent experiments (Shook et al., 2021), and as a result, 483 QTLS were identified, many of which were associated with several agronomic traits simultaneously. More practical is the meta-analysis method, which involves analyzing research data from different teams and in different environments together (Hu et al., 2021). The gene loci found in this way are particularly reliable. Key sites like Joukhadar et al. (2021) that can affect yield and plant height regardless of weather changes were identified in this way. However, to be honest, although these new methods are very powerful, they can be quite mentally challenging to analyze when the volume of data grows large.

6.4 Future trends in soybean genomics research

The research on soybean genes may be going to play some new tricks next. Those algorithms in machine learning are quite popular now, such as support vector regression (SVR) and random forest (RF) (Yoosefzadeh-Najafabadi et al., 2021) It is indeed more accurate and faster to find QTL in GWAS (Yoosefzadeh-Najafabadi et al., 2023). Sequencing technology is also becoming increasingly advanced. It is estimated that in the future, when analyzing the soybean genome, we will be able to see it more thoroughly. However, the most urgent matter might still be the study of climate adaptability - after all, the weather is getting more and more strange now, and it is necessary to quickly figure out the genetic code of drought and heat resistance (Joukhadar et al., 2021). Although these new technologies are not easy to use, if we can truly understand them, breeding will be much easier in the future. Ultimately, no matter how advanced the research methods are, it still depends on whether higher-quality and higher-yield soybeans can be grown.

7 Concluding Remarks

The research on soybean genes may be going to play some new tricks next. Those algorithms in machine learning are quite popular now, such as support vector regression (SVR) and random forest (RF). It is indeed more accurate and faster to find QTL in GWAS. Sequencing technology is also becoming increasingly advanced. It is estimated that in the future, when analyzing the soybean genome, we will be able to see it more thoroughly. However, the most urgent matter might still be the study of climate adaptability - after all, the weather is getting more and more strange now, and it is necessary to quickly figure out the genetic code of drought and heat resistance. Although these new technologies are not easy to use, if we can truly understand them, breeding will be much easier in the future. Ultimately, no matter how advanced the research methods are, it still depends on whether higher-quality and higher-yield soybeans can be grown.

When it comes to soybean breeding, it's actually quite interesting. Previously, people might not have paid much attention to it, but recently some particularly crucial genetic loci have been discovered (both MAS and GS technologies are applicable). However, to be fair, merely finding the site is not enough; it also depends on how it is actually used. Take seed yield and weight for example. Some haplotypes can remain stable in different environments, which is indeed helpful for cultivating more adaptable soybeans. Now, what's even more impressive is that machine learning has also been incorporated, making GWAS analysis more accurate and it's much easier to find QTLS. Although technology is becoming increasingly advanced, in the final analysis, it is still for one goal: to make breeding more efficient, to grow soybeans better and with higher yields (this is indeed quite important for the sustainability of global soybean cultivation). However, it must be admitted that various unexpected situations still occur in actual operation. After all, agriculture has never been smooth sailing.

How will GWAS research on soybeans develop in the future? To be honest, it's really hard to say. Although genotyping technology is constantly advancing (and machine learning methods are becoming increasingly useful), there are still many variables in practical application. The integration of multi-environment data can indeed enhance the accuracy of analysis. However, given the significant differences in climate and soil among various regions, it remains questionable whether the results can be universally applied. Interestingly, it has recently been discovered that there may be certain interactions between different loci that we have not yet fully understood, and these may have a more significant impact on agronomic traits than individual loci. Of course, collaborative research and meta-analysis are indispensable. After all, it is necessary to verify the reliability of these findings in different environments (otherwise, breeders would not dare to use them casually). Overall, the outlook is quite optimistic, but to fully predict the genetic structure of complex traits, it may still take some time to explore.

Acknowledgments

CropSci Publisher sincerely thanks the two anonymous peer reviewers for their suggestions.

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

Abdelraheem A., Thyssen G., Fang D., Jenkins J., McCarty J., Wedegaertner T., and Zhang J., 2020, GWAS reveals consistent QTL for drought and salt tolerance in a MAGIC population of 550 lines derived from intermating of 11 upland cotton (*Gossypium hirsutum*) parents, Molecular Genetics and Genomics, 296: 119-129.
https://doi.org/10.1007/s00438-020-01733-2

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: 104.
https://doi.org/10.1007/s00425-020-03499-8

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

Copley T., Duceppe M., and O'Donoughue L., 2018, Identification of novel loci associated with maturity and yield traits in early maturity soybean plant introduction lines, BMC Genomics, 19: 167.
https://doi.org/10.1186/s12864-018-4558-4

Do T., Vuong T., Dunn D., Clubb M., Valliyodan B., Patil G., Chen P., Xu D., Nguyen H., and Shannon J., 2019, Identification of new loci for salt tolerance in soybean by high-resolution genome-wide association mapping, BMC Genomics, 20: 318.
https://doi.org/10.1186/s12864-019-5662-9

Hu B., Li Y., Wu H., Zhai H., Xu K., Gao Y., Zhu J., Li Y., and Xia Z., 2021, Identification of quantitative trait loci underlying five major agronomic traits of soybean in three biparental populations by specific length amplified fragment sequencing (SLAF-seq), PeerJ, 9: e12416.
https://doi.org/10.7717/peerj.12416

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.
https://doi.org/10.5376/mpb.2024.15.0007

Joukhadar R., Thistlethwaite R., Trethowan R., Keeble-Gagnère G., Hayden M., Ullah S., and Daetwyler H., 2021, Meta-analysis of genome-wide association studies reveal common loci controlling agronomic and quality traits in a wide range of normal and heat stressed environments, Theoretical and Applied Genetics, 134: 2113-2127.
https://doi.org/10.1007/s00122-021-03809-y

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, *I*nternational 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., 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

McLeod L., Barchi L., Tumino G., Tripodi P., Salinier J., Gros C., Boyaci H., Ozalp R., Borovsky Y., Schafleitner R., Barchenger D., Finkers R., Brouwer M., Stein N., Rabanus-Wallace M., Giuliano G., Voorrips R., Paran I., and Lefebvre V., 2023, Multi-environment association study highlights candidate genes for robust agronomic quantitative trait loci in a novel worldwide *Capsicum* core collection, The Plant Journal, 116(5): 1508-1528.
https://doi.org/10.1111/tpj.16425

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

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

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

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

Van K., and McHale L., 2017, Meta-analyses of QTLs associated with protein and oil contents and compositions in soybean [*Glycine max* (L.) Merr.] seed, International Journal of Molecular Sciences, 18(6): 1180.
https://doi.org/10.3390/ijms18061180

Wang W., Zhou B., He J., Zhao J., Liu C., Chen X., Xing G., Chen S., Xing H., and Gai J., 2020, Comprehensive identification of drought tolerance QTL-allele and candidate gene systems in Chinese cultivated soybean population, International Journal of Molecular Sciences, 21(14): 4830.
https://doi.org/10.3390/ijms21144830

Yang L., Lei L., Liu H., Wang J., Zheng H., and Zou D., 2020, Whole-genome mining of abiotic stress gene loci in rice, Planta, 252: 85.
https://doi.org/10.1007/s00425-020-03488-x

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

Zhao X., Liu Z., Li H., Zhang Y., Yu L., Qi X., Gao H., Li Y., and Qiu L., 2020, Identification of drought-tolerance genes in the germination stage of soybean, Biology, 11(12): 1812.
https://doi.org/10.3390/biology11121812