Legume Genomics and Genetics 2025, Vol.16, No.1, 11-22
http://cropscipublisher.com/index.php/lgg

Research Report **Open Access**

Genome-Wide Association Mapping of Drought Resistance Traits in Soybean

Dandan Huang
Hainan Institute of Biotechnology, Haikou, 570206, Hainan, China
Corresponding email: dandan.huang@hibio.org

Legume Genomics and Genetics, 2025 Vol.16, No.1 doi: 10.5376/lgg.2025.16.0002
Received: 22 Nov., 2024
Accepted: 03 Jan., 2025
Published: 18 Jan., 2025

Preferred citation for this article:
Huang D.D., 2025, Genome-wide association mapping of drought resistance traits in soybean, Legume Genomics and Genetics, 16(1): 11-22 (doi: 10.5376/lgg.2025.16.0002)

Abstract Genome wide association studies (GWAS), as a powerful genomic tool, have been widely used to analyze the genetic basis of drought resistance traits in soybean. By mining quantitative trait loci (QTLs) related to drought resistance, they provide important molecular markers for drought resistance breeding. This study introduces the application of GWAS in the research of drought resistance traits in soybeans, with a focus on analyzing the mapping of drought resistance QTLs, the mining of candidate genes, and their application in drought resistance breeding. At the same time, combining GWAS with other molecular breeding techniques such as marker assisted selection (MAS) and genome selection (GS), we have promoted the improvement of drought resistance traits and explored the potential of gene editing technology in enhancing soybean drought resistance. Research has found that GWAS has made significant progress in the study of soybean drought resistance, identifying multiple key QTLs that affect root development, water use efficiency (WUE), and metabolic pathways, and revealing the impact of gene environment interactions on drought resistance traits. Through gene functional analysis, candidate genes for drought resistance and their regulatory networks have been identified, providing a new direction for molecular breeding of drought resistant traits. GWAS has demonstrated strong potential in the study of drought resistant traits in soybeans, not only revealing complex genetic regulatory networks, but also providing valuable molecular tools for drought resistant breeding. In the future, by integrating new technologies such as big data, machine learning, and gene editing, precision breeding of drought resistant traits will be further optimized and promoted, providing more adaptable varieties for global soybean production.

Keywords Soybean; Drought resistance; Genome-wide association study (GWAS); Genomic selection; Gene editing

1 Introduction

When it comes to soybeans (*Glycine max* L.), despite its unremarkable appearance, it is a tough player in global agriculture (Suo et al., 2022). Southeast Asia, Africa and the Americas all point to it. After all, the protein and fat content is there (Kim et al., 2023b). However, on the other hand, although soybeans are now an important source of animal feed and human food, the cultivation conditions vary greatly from place to place (Rani et al., 2023). The market demand keeps rising, leaving researchers struggling with how to cultivate high-yield varieties adapted to different environments. To be honest, it's not without reason that this crop has reached its current status.

Nowadays, global warming is becoming more and more serious, and the problem of drought is also becoming more and more frequent (Cao et al., 2020; Kim et al., 2023a). When it comes to soybeans, this thing is most afraid of water shortage. Once there is drought, the output drops sharply. (Kim et al., 2023c; Li et al., 2023). In fact, drought resistance is particularly important for soybeans. After all, climate change is so unstable now. However, it is not easy to cultivate drought-resistant varieties. The genetic mechanism behind them must be clarified first (Xiong et al., 2020). Although soybeans can grow normally, they still cannot withstand drought. Therefore, breeding experts are all working hard to study this.

Nowadays, those engaged in research on drought resistance of soybeans are all using GWAS (Genome-wide Association Study), and this method is indeed quite effective. Although traditional breeding methods have not been completely phased out, GWAS can directly identify those gene loci (QTLs) and candidate genes related to drought resistance (Kim et al., 2023b; Rani et al., 2023). To put it simply, it is to conduct experiments on different soybean varieties to see which genetic markers are linked to drought resistance performance (Kim et al., 2023c).

Especially those SNPs (Single Nucleotide Polymorphisms), which can reflect the genetic characteristics of yield-related traits under both normal and drought conditions (Li et al., 2023). However, on the other hand, this technology is not omnipotent, but it does help researchers understand many genetic mechanisms of drought resistance (Huang et al., 2024).

This research aims to thoroughly explore the secrets behind soybeans' drought resistance. To be honest, people engaged in breeding nowadays all know that relying solely on traditional methods is a bit insufficient. We are going to use the latest genotyping technology to identify all the key gene loci related to drought resistance of soybeans. Although the workload is not small, if a few useful candidate genes can really be found, it will be of great significance. Just think, nowadays the climate is getting more and more extreme. Drought-resistant soybean varieties are simply treasures. If these genetic mechanisms can be understood, perhaps good soybeans can be grown in arid areas in the future. Of course, this matter cannot be rushed. It has to be done step by step. However, it will definitely be helpful for ensuring food security.

2 Genetic Basis of Soybean Drought Resistance Traits

2.1 Complexity of drought resistance traits

When it comes to soybean drought resistance, it is not just one or two genes the final say. Researchers have found that just one GWAS study identified 11 SNP loci and 22 QTLs, particularly the transcription factor GmNFYB17, which can both resist drought and increase yield (Sun et al., 2020). However, to be honest, this is just the tip of the iceberg-a study has discovered 75 and 64 drought resistance related QTLs in one go, and these loci explain many phenotypic differences (Wang et al., 2020). Even more exaggeratedly, someone compiled 73 studies and found 483 QTLs distributed in 393 different locations (Shook et al., 2020). You see, drought resistance is so complex, it all relies on a bunch of genes working together to create an expression regulation network. Although we have found many clues now, we are still far from fully understanding them.

2.2 Phenotypic characteristics of drought resistance traits

To determine the drought resistance of soybeans, three key aspects need to be considered. Firstly, the root system characteristics-plants overexpressing GmNFYB17 are particularly interesting, with faster root development, significantly increased number of lateral roots, and significantly improved root shoot ratio (Sun et al., 2020). Of course, looking at the root system alone is not enough. The water retention capacity of leaves is also crucial, such as relative water content (RWC), SOD activity, and proline content, which vary significantly under drought conditions (Figure 1). When measuring these traits, it is usually necessary to set different water treatments and combine them with SNP markers for genetic analysis (Dhungana et al., 2021; Ouyang et al., 2022). When it comes to water use efficiency, although often overlooked, it is indeed one of the important indicators for evaluating drought resistance.

2.3 Genetic diversity of soybean drought resistance traits

Let's do drought-resistant breeding for soybeans. It's actually quite interesting. Look at those local old varieties. Compared with the varieties promoted now, the drought resistance characteristics at the genetic level are really quite different (Wang et al., 2020). Although the exact details still need to be studied in detail, this difference might just be a breakthrough-by recombining genes from different sources, it might be possible to create new varieties that are more drought-resistant. When it comes to this, the soybean population in China is quite valuable for reference, as it contains many QTLS and candidate genes related to drought resistance. However, then again, it is not enough to just focus on a single study. The QTL data from different teams need to be integrated and viewed together (Hwang and Lee, 2019; Shook et al., 2020), in order to figure out the genetic pattern of drought resistance. Of course, this is easier said than done, but if it can be truly understood, it will definitely be a great benefit for cultivating drought-resistant soybeans.

3 Overview of Genome-Wide Association Studies (GWAS)

3.1 Principles and methods of GWAS

When it comes to GWAS method, it is currently being used by breeding researchers. The principle is actually not complicated, it is to compare the genomic data of different individuals together to see which gene variations are associated with specific traits. However, the actual operation is not that simple. Just collecting enough samples for genotyping is enough to make a fuss (Korte and Farlow, 2013). Fortunately, with the help of molecular markers such as SNPs, statistical analysis has become much more convenient. In crops like soybeans, key loci controlling complex traits were identified using GWAS in 2016 (Chang et al., 2016). But on the other hand, this method also depends on the specific situation. A paper in 2021 pointed out that the results obtained under different experimental conditions may have significant differences (Cortes et al., 2021).

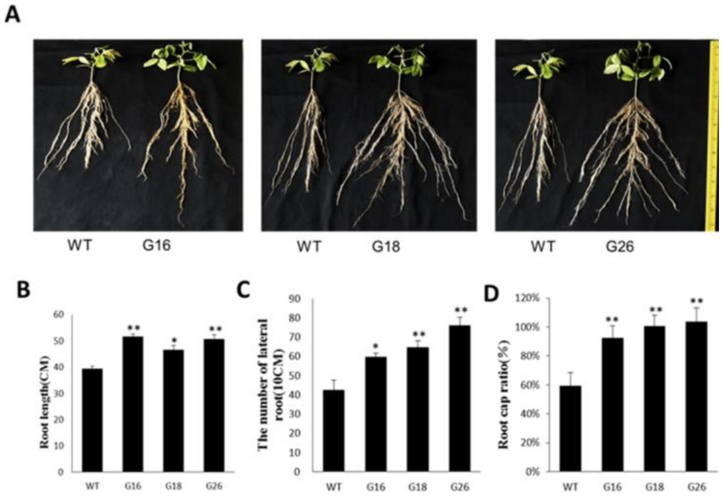

Figure 1 Analysis of root growth and root-to-shoot ratio in transgenic soybean (Adapted from Sun et al., 2020)

Image caption: A: Morphological differences in the roots of transgenic soybean lines G16, G18, G26, and wild-type (WT) soybean after a 6-week growth period; B: Differences in root length between each line and WT; C: Number of lateral roots within a 10 cm root segment; D: Comparison of root-to-shoot ratio differences between transgenic lines and WT; Statistical data indicate that transgenic lines G16, G18, and G26 exhibit significantly greater root length, lateral root number, and root-to-shoot ratio compared to WT, with these differences showing statistical significance across multiple transgenic lines (*: $p \leq 0.05$; **: $p \leq 0.01$) (Adapted from Sun et al., 2020)

3.2 Current applications of GWAS in soybean research

Nowadays, GWAS is widely used in soybean research, especially in identifying those gene loci related to drought resistance. It is quite interesting to remember that there was a study in 2023. They discovered some SNP markers, which were associated with yield traits such as pod number and biomass (Li et al., 2023), and could be detected under both normal watering and drought conditions. In fact, as early as 2015, researchers used the GBS method for genome-wide labeling (Sonah et al., 2015), and extracted the gene loci of traits such as maturity period and plant height. But the most impressive one was the later analysis that combined 73 studies (Shook et al., 2020), which put data from different sources together, and the accuracy of finding drought-resistant genes improved all at once. These studies have indeed deepened our understanding of the drought resistance of soybeans significantly.

3.3 GWAS data analysis tools and methods

When it comes to the data analysis of GWAS, there are actually quite a few methods, each with its own uses. Look at that hybrid model framework, which is quite interesting (Figure 2) (Cortes et al., 2021). It can take into account population structure and kinship, so there can be much fewer false positive results. However, this alone is not enough. Specialized software like BLINK and rrBLUP are also quite crucial (Ravelombola et al., 2021). One is used for GWAS analysis, and the other for homosexuals etiome prediction. They are particularly useful in marker-assisted selection. If we talk about the most convenient method, it is still XP-GWAS (Yang et al., 2015). It specifically selects individuals with extreme phenotypes for sequencing and is particularly friendly to those species with limited genotyping resources. Although these tools may seem diverse, they are ultimately all aimed at better analyzing the genetic mechanism of drought resistance traits in soybeans, so that we can cultivate more drought-resistant varieties.

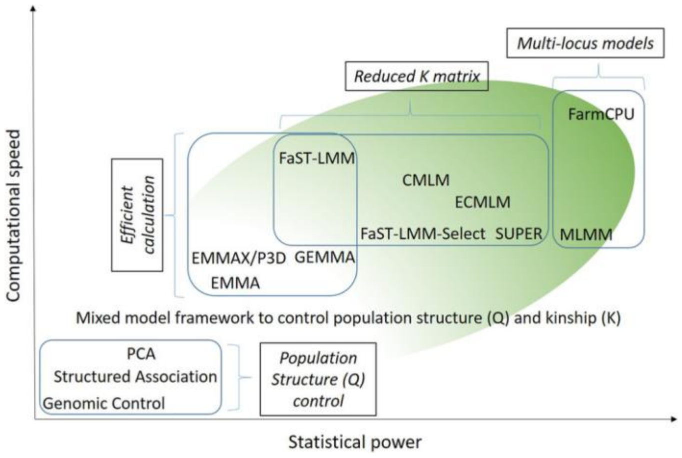

Figure 2 Computational speed and statistical power improvement of genome-wide association study methods (Adapted from Cortes et al., 2021)

Image caption: This figure displays the performance of various genome-wide association study (GWAS) methods in terms of improvements in computational speed and statistical power. The methods are categorized into several types, including multi-locus models, strategies for reducing the K matrix, and mixed-model frameworks that control for population structure (Q) and kinship (K). Among them, multi-locus models like FarmCPU and MLMM show higher statistical power, while mixed-model-based strategies such as CMLM and ECMLM demonstrate improvements in computational speed (Adapted from Cortes et al., 2021)

4 GWAS of Soybean Drought Resistance Traits

4.1 Mapping of drought-related QTLs and candidate gene discovery

GWAS has been a great help in researching the drought resistance of soybeans. There was an interesting study in 2020 (Wang et al., 2020), which found 75 QTLs affecting plant weight and 64 QTLs affecting plant height in cultivated soybeans in China. These loci together can explain more than half of the phenotypic variation. Interestingly, drought resistance is not determined by a single gene, but rather by the collective action of a large number of QTL allele genes. Later in 2023, further research was conducted (Li et al., 2023), and 39 yield related SNP loci were identified under both normal watering and drought conditions, distributed across 26 genomic regions. Of particular note are genes *Glyma.19G211300* and *Glyma.17G057100*, which are directly involved in drought stress response. These findings have further deepened people's understanding of the drought resistance mechanism of soybeans.

When it comes to the localization of drought resistant genes in soybeans, there have been many new discoveries in recent research. In 2022, there was an interesting study (Sun et al., 2022) that combined GWAS with linkage analysis and found 11 SNPs and 22 QTLs in one go. Of particular note is the qGI10-1 locus, which has been validated through both methods, and it has been found that the *GmNFYB17* gene not only regulates drought resistance but also affects root growth. However, when it comes to gene function, another study using RIL population is more detailed (Ouyang et al., 2022), where they identified 5 drought resistant QTLs and identified 9 candidate genes, including some NAC transport factors. These proteins may play a key role in drought resistance, although the specific mechanism still needs further research.

4.2 Gene function analysis and drought regulatory networks

Bioinformatics analysis can be of great help in understanding how these drought resistant genes work specifically. Taking the 2020 study by Wang et al. (2020) as an example, candidate genes in Chinese cultivated soybeans can be classified into nine major categories, with ABA and stress response factors accounting for the majority. This indicates that they are particularly important in drought resistance. However, in terms of practical applications, the *GmNFYB17* gene is more interesting (Sun et al., 2022). After overexpressing it in soybeans, not only does it enhance drought resistance, but it also increases yield. Especially under water deficient conditions, these genetically modified plants have better root growth and significantly improved yield related traits, although further research is needed on how to achieve this.

There have been some new discoveries recently regarding the study of soybean drought resistance. There is an interesting paper in 2023 (Li et al., 2023), which found that genes *BAH1* and *WRKY11* are crucial, especially the E3 ubiquitin protein ligase and transcription factor, which seem to play a significant role in plants' response to drought stress. However, if there are many types of genes, we still need to look at the 2022 study (Ouyang et al., 2022), which identified nine candidate genes alone. Among them, NAC and GATA transport factors are particularly noteworthy, as they seem to regulate gene expression under drought conditions. To be honest, it is particularly important to understand the functions of these genes, after all, the cultivation of drought resistant soybean varieties now relies on these discoveries, although the specific application still needs to be explored.

4.3 Gene-environment interactions on drought resistance traits

When it comes to the application of GWAS, it was recently discovered that it can also study the "tug-of-war" between genes and the environment. There was an interesting study in 2022 (Yu et al., 2022), which used a model called 3VmrMLM to analyze tocopherols in soybean seeds. As a result, 57 QTLS were identified. Even more interestingly, 13 QTLS were found to "interact" with the environment-genes close to these loci were particularly susceptible to environmental influences. This doesn't mean that when conducting GWAS analysis, it's really not advisable to ignore environmental factors. However, there are exceptions. A study using the RIL population in 2021 found (Dhungana et al., 2021) that the positions of many drought-resistant QTLS completely coincided with those previously reported, indicating that these sites are particularly "stubborn" and remain stable regardless of how the environment changes.

The recent research is quite interesting. Whether under normal watering or drought conditions (Li et al., 2023), some QTLs are particularly "stubborn" and can be detected in both environments. This indicates that these loci may be inherently stable and not affected by the environment. However, on the other hand, QTLs and genes that interact with the environment are more worthy of attention, as soybeans in reality have to face various complex situations. By understanding these, perhaps a new variety that is both drought resistant and high-yielding can be cultivated, although it is easier said than done.

5 Successful GWAS Cases in Soybean Drought Resistance Traits

5.1 Mapping of root development-related QTLs and breeding applications

When it comes to drought resistance of soybeans, the characteristics of their root systems are truly a major discovery. Look at that *GmNFYB17* gene (Sun et al., 2022). After it was transferred into soybeans and overexpressed, the effect was particularly obvious-the root system grew fast and abundant, the lateral roots were dense, and the drought resistance was immediately enhanced. This gene is quite interesting. It was discovered through GWAS combined with linkage analysis, which can be said to have utilized the advantages of both methods. In fact, as early as 2019, studies noticed the importance of root systems (Khan et al., 2019), and they found 46 new QTLS in a semi-identical population at one go, all of which controlled plant length and dry weight. After organizing these QTLS into an allele matrix, it is particularly helpful for drought-resistant breeding, although the specific application still depends on the actual situation. Ultimately, to cultivate drought-resistant soybeans, merely focusing on the above-ground parts is not enough; the underground root system is the key.

5.2 Discovery and validation of water use efficiency (WUE)-related genes

When it comes to the water use efficiency of soybeans, GWAS has been a great help. A study in 2023 was quite interesting (Li et al., 2023), which identified several particularly stable QTLs that can affect yield traits such as pod number and biomass under both normal watering and drought conditions. More importantly, *Glyma.19G211300* genes were discovered, which encode proteins directly related to stress response-although the specific mechanism still needs further research. These findings provide new ideas for cultivating water-saving and drought resistant varieties. In fact, as early as 2020, researchers used SLAF sequencing to create an ultra-high density genetic map, mapping QTLs that control plant height and grain weight (Ren et al., 2020). Now breeding experts can directly use these linkage markers for assisted selection, greatly accelerating the breeding progress of drought resistant varieties, but the actual operation still depends on specific environmental conditions.

5.3 Discovery of drought-related metabolic pathways

When it comes to drought resistance of soybeans, GWAS studies have indeed unearthed a lot of valuable insights. Take the *GmNFYB17* gene for example (Sun et al., 2022). It not only enables the root system to grow stronger, but also controls a host of physiological indicators for drought resistance-such as the degree of leaf damage, MDA content, etc. (Figure 3). Interestingly, it can also regulate SOD activity and proline content. When these changes are combined, the drought resistance of the plants increases. However, for a more comprehensive discovery, the study using SLAF sequencing in 2020 has to be mentioned (Ren et al., 2020). The QTL they identified clarified the genetic basis of drought resistance metabolic pathways. Although the specific operation still needs to be further explored, these findings do provide new ideas for cultivating drought-resistant soybeans, such as designing breeding strategies based on these metabolic pathways.

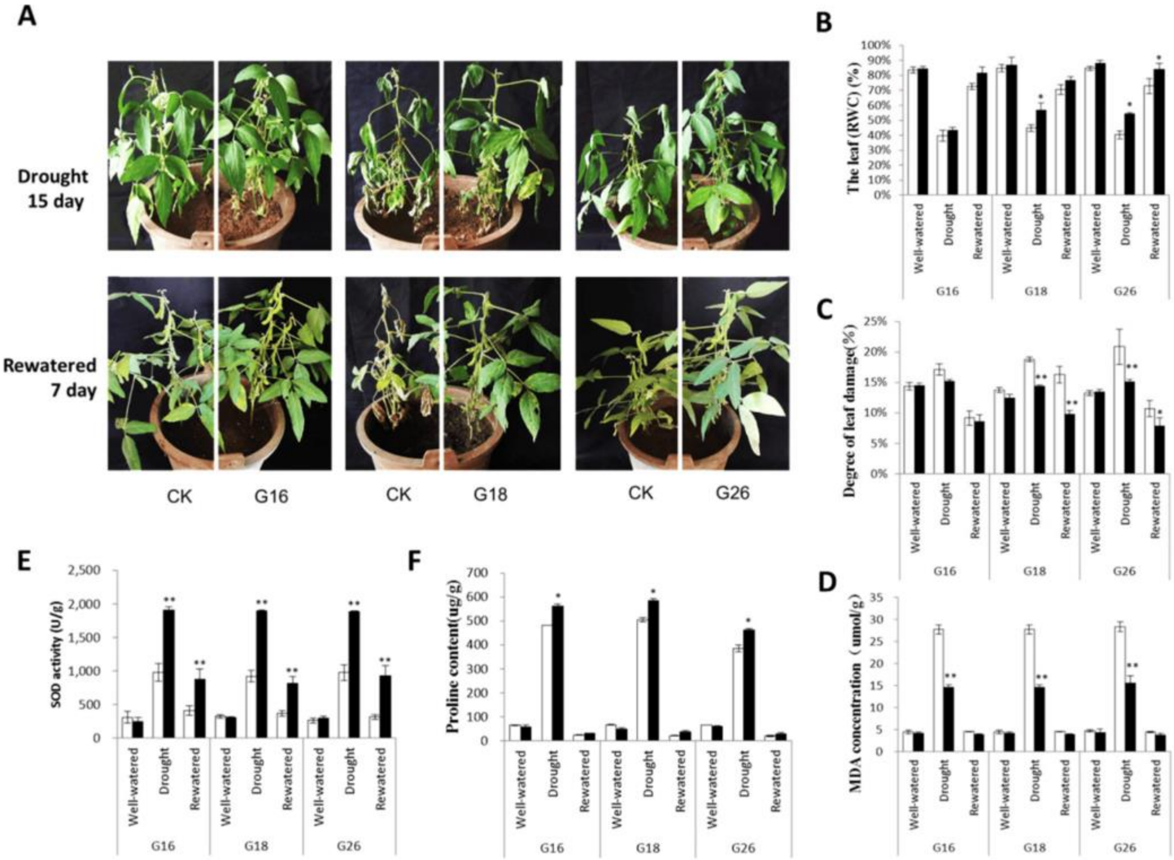

Figure 3 Performance of *GmNFYB17* transgenic soybean under drought stress (Adapted from Sun et al., 2022)

Image caption: A: Morphological differences between transgenic soybeans (G16, G18, G26) and non-transgenic control (CK) after 15 days of drought stress followed by 7 days of re-watering; B; C: Relative water content (RWC) of leaves and degree of leaf damage at different treatment stages; D~F: Comparison of physiological and biochemical indicators under drought conditions between transgenic and non-transgenic soybeans, including malondialdehyde (MDA) concentration, superoxide dismutase (SOD) activity, and proline content (Adapted from Sun et al., 2022)

The study on the *GmNFYB17* gene in 2022 was quite interesting (Sun et al., 2022). They conducted a drought experiment by putting genetically modified soybeans and ordinary soybeans together and found that the genetically modified ones were indeed more resistant to drought-the water content in the leaves was maintained better and the damage was much less. More importantly, a bunch of physiological indicators were tested: the activity of SOD increased and the content of MDA decreased, indicating that the cell membrane is more stable. The proline content has also increased. This thing can help regulate osmotic pressure. Although further research is needed on exactly how it works, at least it has been proven that this gene does work. Nowadays, colleagues engaged in drought-resistant breeding have a new option. However, if it is to be applied to variety selection, it still needs to go through the test of field trials.

6 Application of Genome-Wide Association Mapping in Soybean Drought Resistance Breeding

6.1 Marker-assisted selection (MAS) and improvement of drought resistance traits

Nowadays, when it comes to drought-resistant breeding of soybeans, marker-assisted selection (MAS) has really been of great help. I remember that the study in 2020 was quite practical (Ren et al., 2020). They used SLAF-Seq technology to create a high-density genetic map, clearly marking all the QTLS related to drought resistance. This technology features high sequencing efficiency and relatively controllable costs, making it particularly suitable for breeding applications. With these precise molecular markers, breeding experts can screen out superior genotypes at the seedling stage, no longer having to wait until flowering and fruiting to make judgments as before, saving both time and effort. Although problems such as recombination rate and environmental interaction may still be encountered in actual operation, compared with traditional breeding methods, MAS has indeed accelerated the progress of drought-resistant variety selection and breeding significantly.

The mark-assisted selection (MAS) technique works not only well on soybeans but also on other crops. As early as the 1990s, studies had already reaped benefits on common legumes (Schneider and Kelly, 1997), proving that this method was indeed effective in dealing with complex traits such as drought resistance. Later, around 2005, researchers found that the combination of MAS and traditional breeding was more effective (Francia et al., 2005; (Miklas et al., 2006), whether under normal or drought conditions, the output can be increased. Although there may be some differences among various crops, the molecular markers provided by GWAS have indeed saved breeding efforts a lot of detours. Ultimately, without the assistance of these markers, conducting drought-resistant breeding nowadays would be like walking in the dark, and the efficiency would be in a completely different league.

6.2 Application of genomic selection (GS) in drought resistance breeding

Genomic selection (GS) has indeed gained popularity in the breeding circle in recent years, especially in dealing with complex traits such as drought resistance. It's quite interesting to say that it can be regarded as a golden pair with GWAS-by feeding the labeled data found by GWAS to the GS model, the prediction accuracy can improve significantly (He et al., 2014). The most remarkable aspect of this technology lies in the fact that it can estimate the breeding value based on whole-genome markers without waiting for the plants to grow up and exhibit traits. This trick was used in soybean breeding and the results were quite good. However, in actual operation, the debugging of model parameters and the marking density and other details are quite meticulous. But compared with traditional methods, it is indeed much more time-saving and labor-saving. Nowadays, more and more breeding projects are trying this approach. Although they are still in the exploratory stage, the prospects are indeed quite attractive.

The combination of genomic selection (GS) and genome-wide association analysis (GWAS) has indeed sparked quite a few ideas. Especially when studying the drought resistance of soybeans, this combination approach is particularly effective-GWAS is responsible for extracting the key QTLS, and GS uses these data to build predictive models. The entire process is like an assembly line (Ren et al., 2020). Nowadays, with those high-throughput genotyping platforms and increasingly intelligent statistical models, the efficiency of finding drought-resistant loci has improved significantly. A study in 2021 proved this point (Hasan et al., 2021). Using GS to select drought-resistant genotypes is not only fast but also reliable in accuracy. Although the model training stage still requires some effort, once it runs smoothly, the breeding efficiency will increase visibly to the naked eye. Ultimately, this approach of integrated analysis has put molecular breeding on a fast track.

6.3 Potential of gene editing technology

Gene editing technology has truly brought revolutionary changes to breeding work in recent years, especially CRISPR-Cas9, the "gene scissors", which are particularly convenient to use. In the past, when doing drought-resistant breeding, it took a lot of effort just to find the key genes. Now it's better. GWAS screens out the candidate genes first (Manavalan et al., 2009), and CRISPR can precisely operate on these genes. For instance, for genes related to drought resistance, if you want to enhance their expression, you can do so; if you want to knock them out, you can knock them out, just like playing with building blocks. However, in actual operation, it is not that simple. Issues such as delivery efficiency and off-target effects still need to be gradually optimized. But it cannot be denied that this technology has ushered in a new era for crop breeding-not only can it quickly introduce beneficial mutations, but also repair those defective genes that affect drought resistance. Although its application on soybeans is still in the laboratory stage at present, who wouldn't say it's amazing when they see this potential?

Gene editing technology has indeed played a new role in crop breeding in recent years, especially the CRISPR-Cas9 "molecular scissors". In addition to its impressive performance on soybeans, staple crops such as rice and wheat have also tasted the sweetness (Manavalan et al., 2009). Interestingly, combining this new technology with traditional breeding has yielded surprisingly good results-GWAS first locates drought resistant genes, CRISPR then precisely edits them, and finally screens them using conventional breeding methods. With the entire process, the breeding cycle can be shortened by several years. A study in 2021 demonstrated the power of this combination punch (Hasan et al., 2021), which not only improved drought resistance but also decreased yield. However, it is still too early to completely replace traditional methods, as practical issues such as field performance and regulatory approval need to be addressed step by step. But it cannot be denied that this technology provides new ideas for addressing the food security challenges brought by climate change, although there are still many obstacles to overcome on the road to promotion.

7 Challenges and Future Research Directions

7.1 Challenges in phenotypic data collection

Measuring the drought resistance of soybeans is not that simple in practice. Those key indicators-such as root development, leaf wilting degree, etc.-are particularly troublesome to measure, and data from different plots and different years often do not match (Manavalan et al., 2009). Take the root system for example. Although everyone knows that it is particularly important for drought resistance, every time a measurement is made, the plant has to be pulled out by the roots, making it impossible to continuously track the changes of the same plant. The traditional manual measurement method is slow and prone to errors. Often, after working hard for a long time, the data obtained may not be reliable. Although there are some new technologies that can offer assistance now, to truly establish a large-scale and high-quality phenotypic database, many difficulties still need to be overcome.

Now there are finally some new techniques for measuring soybean drought resistance, and those high-tech phenotype platforms have indeed been of great help (Kim et al., 2023a). For example, using cameras to automatically capture the growth process of plants, or installing sensors to monitor water conditions in real time, these non-destructive methods are much stronger than before-there is no need to uproot plants, and the same batch of materials can be tracked from seedling stage to maturity stage. However, to be honest, these new technologies need to be further adjusted when applied to drought resistance research in soybeans, as soybeans are quite unique and have different responses to drought at different growth stages. Recently, research has been attempting to combine these automated phenotype data with GWAS analysis. Although it is still in the exploratory stage, if it can be standardized, it will be much more convenient for drought resistant breeding in the future. Of course, practical issues such as equipment costs and data analysis algorithms also need to be considered, but at least we have seen the direction now.

7.2 Complexity of genetic regulatory networks for drought resistance traits

The matter of soybeans' drought resistance is simple, but it's actually very complicated. A study in 2022 found (Ouyang et al., 2022) that merely finding a few QTLS is far from enough. These loci are interlinked to form a complex genetic network. Look, some genes are responsible for allowing cells to accumulate more proline to retain water, some control the root system to grow deeper, and others regulate the level of ABA hormone-these mechanisms each have their own responsibilities, but they have to work in coordination with each other. Just like a symphony orchestra, it is not enough to merely understand the sheet music of each musician; the key is to figure out how they play in coordination. Although many drought-resistant genes have been identified now, scientists are still scratching their heads over how these genes interact with each other. To completely crack this genetic code, it is estimated that it will still take some more years of effort.

When it comes to drought resistance of soybeans, nowadays research has to adopt a "combination punch" (Wang et al., 2020). Just looking at the genome is not enough; it is also necessary to analyze the data of the transcriptome, proteome, and metabolome together-just like having to read a score, listen to a performance, and study the movements of each musician. Although this approach is troublesome, it can indeed reveal quite a few tricks: for instance, when will a certain drought-resistant gene be activated? Which proteins will be driven to work? What metabolites were finally produced? Only when these clues are strung together can a complete drought resistance mechanism be pieced together. To be honest, with so much omics data piled up together, just processing it is enough to make a big deal out of it. The existing analysis tools often get stuck. What is most needed now is smarter algorithms that can automatically correlate data at different levels and, ideally, predict the regulatory relationships between genes. Although there is still some way to go before the drought-resistant network is completely cracked, at least the path is becoming clearer and clearer as we go along.

7.3 In-depth exploration of gene-environment interaction studies

When it comes to soybeans' drought resistance, it's not something that can be decided by genes alone-environmental factors are the real "masterminds" behind the scenes. The research in 2022 was particularly interesting (Sun et al., 2022). The performance of the same drought-resistant genes in sandy soil and yellow clay can be worlds apart. As soon as the soil moisture changes, the gene expression pattern seems to undergo a magic trick. The temperature level is directly related to the effectiveness of drought resistance. Sometimes, genes that work in high temperatures can be counterproductive in low temperatures. The most infuriating thing is that nutritional status also gets involved. Different levels of nitrogen, phosphorus and potassium result in different "work enthusiasm" of genes. Although QTL mapping is doing very well now, we still know too little about how these genes "tailor their treatment to the individual" in different environments. Just like the same actor can perform completely different effects in different film crews, the ins and outs of this are enough for researchers to ponder over for another ten or eight years.

Indeed, to truly crack the code for soybeans' drought resistance, merely tinkering in the laboratory is not enough. Recently, some scholars have proposed (Wang and Li., 2024) that the same materials should be planted all over the country, from the black soil of Northeast China to the arid areas of Northwest China, so as to screen out those reliable QTLS that "shine brightly with a little sunlight". It's easier said than done. Such large-scale experiments are both costly and time-consuming. But there's no way around it. After all, genes are just like people; a different environment might completely change their appearance. In 2016, a study pointed out (Chang et al., 2016) that the current gene-environment interaction model is still too "rigid" and a more intelligent prediction system needs to be developed. Imagine if, like weather forecasts, parameters such as soil moisture and temperature could be input to predict the expression of a certain genotype in a specific area. Then, breeders would have a much easier time. However, to be fair, the way soybeans interact with the environment is just too complex. If this system is really to be developed, it will probably take several more years of hard work.

7.4 Application of big data and machine learning in GWAS research

Big data and machine learning have now become the "golden pair" in GWAS research-this is particularly evident when analyzing the drought resistance traits of soybeans. Just think about it. GWAS often generates massive amounts of genotype and phenotype data (Zhang et al., 2021), and traditional statistical methods handle them as if calculating space orbits. But machine learning is different, especially deep learning, which can dig out hidden patterns from these data that the human eye cannot detect at all. For instance, in 2021, a team successfully experimented with other crops. The predictive model they developed could even capture those nonlinear gene interactions. To be honest, however, there are still many problems to be solved when applying these techniques to the research on soybean drought resistance: the data quality varies greatly, the model is prone to overfitting, and the most troublesome issue is the explainability of the results. However, in any case, this approach has indeed given researchers new hope. At least they no longer have to stare blankly at piles of data.

Research on drought resistance of soybeans can now be regarded as entering the "big data era" (Bao et al., 2015). Imagine throwing genetic data, field phenotypic data and meteorological, soil and environmental data all into an intelligent analysis platform, allowing machine learning algorithms to find patterns on their own-this is much faster than manual analysis in the past. Especially for those key genes with pleiotropy, traditional methods often miss them. Now, through multi-dimensional association analysis, none of them can escape. But to be honest, building such a platform is no easy task: first of all, the issue of data standardization must be addressed, after all, the data collection methods in different test fields are all over the place. Secondly, the algorithm must be proficient enough to distinguish between genuine gene-environment interactions and merely data noise. As early as 2015, teams attempted to develop such tools. Although they were still in the refinement stage, they could already help breeders identify drought-resistant candidate genes more quickly. If in the future these intelligent prediction systems can be directly connected to the breeding process, from gene discovery to variety selection and breeding can be completed in one go, that would be truly exciting!

8 Concluding Remarks

Research on the drought resistance of soybeans has made breakthrough progress in recent years-and GWAS deserves much credit for this. In the past, finding drought-resistant genes was like looking for a needle in a haystack. Now, with high-density SNP markers, it's just like having opened the "Heavenly Eye". Look, just for controlling yield and plant height, dozens of QTLS have been found, not to mention those key sites that affect root development. It is particularly worth mentioning the star gene *GmNFYB17*. It is truly amazing. It can not only help soybeans resist drought but also increase their yield. It is simply the dream "cause" for breeders. But then again, drought resistance is too complex to be solved by just one or two genes. It's like a jigsaw puzzle, and now there are finally a few key pieces. Although the complete genetic map is still being drawn, these discoveries have already saved drought-resistant breeding from many detours. At least now we know in which direction to focus our efforts.

GWAS has become increasingly proficient in the research on drought resistance of soybeans and is sure to achieve great success in the future. Now, with those high-tech means-like GBS sequencing, SNP chips and the like-finding drought-resistant genes is like having a cheat code. When these technologies work together, they can identify all the drought-resistant alleles hidden in soybean germplasm resources, and perhaps even discover some new variations that have never been seen before. Even more amazing is the use of meta-analysis, which packages and processes the data from various studies. The statistical power is directly maximized, and the reliability of the QTLS found also increases rapidly. In this way, marker-assisted selection becomes much more precise, and breeders no longer have to rely on luck as they did before. Although it is still impossible to achieve 100% accuracy at present, if we continue along this path, it will surely become easier and easier to cultivate drought-resistant soybean varieties.

Nowadays, in soybean drought resistant breeding, GWAS, GS, and gene editing are advancing together, and the effect is really 1+1+1>3. GWAS first acts as a scout, identifying drought resistance related QTLs and candidate genes one by one; Then GS comes on stage and uses these labeled data to build a prediction model, which can predict which plants are more drought resistant during the seedling stage; Finally, CRISPR emerged and edited target genes like a precision surgical knife. The most wonderful thing about this combination boxing style is that it retains the advantages of traditional breeding while incorporating precise molecular level operations. For example, the *GmNFYB17* gene was first discovered by GWAS, and then gene edited to enhance expression. Finally, the GS model proved that it can indeed improve yield stability under drought conditions. Although this process is still being optimized, it has already shortened the breeding cycle for drought resistant varieties by several years. As climate change becomes increasingly challenging, this multi technology integration strategy may be the key to ensuring soybean production in the future.

Acknowledgments

Thanks to the anonymous peer review for providing targeted revision suggestions for the manuscript.

Conflict of Interest Disclosure

The author affirms that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Bao Y., Kurle J., Anderson G., and Young N., 2015, Association mapping and genomic prediction for resistance to sudden death syndrome in early maturing soybean germplasm, Molecular Breeding, 35: 128.
https://doi.org/10.1007/s11032-015-0324-3

Cao L., Jin X.J., Zhang Y.X., Zhang M.C., and Wang Y.H., 2020, Transcriptomic and metabolomic profiling of melatonin treated soybean (*Glycine max* L.) under drought stress during grain filling period through regulation of secondary metabolite biosynthesis pathways, PLOS ONE, 15(10): e0239701.
https://doi.org/10.1371/JOURNAL.PONE.0239701

Chang H., Lipka A., Domier L., and Hartman G., 2016, Characterization of disease resistance loci in the USDA soybean germplasm collection using genome-wide association studies, Phytopathology, 106(10): 1139-1151.
https://doi.org/10.1094/PHYTO-01-16-0042-FI

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

Dhungana S., Park J., Oh J., Kang B., Seo J., Sung J., Kim H., Shin S., Baek I., and Jung C., 2021, Quantitative trait locus mapping for drought tolerance in soybean recombinant inbred line population, Plants, 10(9): 1816.
https://doi.org/10.3390/plants10091816

Francia E., Tacconi G., Crosatti C., Barabaschi D., Bulgarelli D., Dall'Aglio E., and Valè G., 2005, Marker assisted selection in crop plants, Plant Cell, Tissue and Organ Culture, 82: 317-342.
https://doi.org/10.1007/s11240-005-2387-z

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

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

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

Hwang S., and Lee T., 2019, Integration of lodging resistance QTL in soybean, Scientific Reports, 9: 6540.
https://doi.org/10.1038/s41598-019-42965-6

Kim J., Lee C., Park J., Kim N., Kim S., Baek J., Chung Y., and Kim K., 2023a, Comparison of various drought resistance traits in soybean (*Glycine max* L.) based on image analysis for precision agriculture, Plants, 12(12): 2331.
https://doi.org/10.3390/plants12122331

Kim S., Tayade R., Kang B., Hahn B., Ha B., and Kim Y., 2023b, 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., 2023c, 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

Khan M., Tong F., Wang W., He J., Zhao T., and Gai J., 2019, Using the RTM-GWAS procedure to detect the drought tolerance QTL-allele system at the seedling stage under sand culture in a half-sib population of soybean [*Glycine max* (L.) Merr.], Canadian Journal of Plant Science, 99: 801-814.
https://doi.org/10.1139/cjps-2018-0309

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

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

Manavalan L., Guttikonda S., Tran L., and Nguyen H., 2009, Physiological and molecular approaches to improve drought resistance in soybean, Plant & Cell Physiology, 50(7): 1260-1276.
https://doi.org/10.1093/pcp/pcp082

Miklas P., Kelly J., Beebe S., and Blair M., 2006, Common bean breeding for resistance against biotic and abiotic stresses: from classical to MAS breeding, Euphytica, 147: 105-131.
https://doi.org/10.1007/s10681-006-4600-5.

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

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

Ren H., Jianan H., Wang X., Zhang B., Yu L., Gao H., Huilong H., Rujian S., Tian Y., Qi X., Liu Z., Wu X., and Qiu L., 2020, QTL mapping of drought tolerance traits in soybean with SLAF sequencing, Crop Journal, 8: 977-989.
https://doi.org/10.1016/j.cj.2020.04.004

Schneider K., and Kelly J., 1997, Marker-assisted selection to improve drought resistance in common bean, Crop Science, 37: 51-60.
https://doi.org/10.2135/CROPSCI1997.0011183X003700010008X

Shook J., Zhang J., Jones S., Singh A., Diers B., and Singh A., 2020, 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

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

Suo R., Sandhu K., Wang M., You F., Conner R., Cober E., and Hou A., 2022, Soybean (*Glycine max* L.) seed germination in response to waterlogging and cold climate: a review on the genetics and molecular mechanisms of resistance to the abiotic stress, Canadian Journal of Plant Science, 103: 13-28.
https://doi.org/10.1139/cjps-2022-0111.

Wang H.P., and Li H.M., 2024, Application of molecular marker assisted selection in wheat stress resistance breeding, Triticeae Genomics and Genetics, 15(1): 1-9.
https://doi.org/10.5376/tgg.2024.15.0001

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

Xiong R., Liu S., Considine M., Siddique K., Lam H., and Chen Y., 2020, Root system architecture, physiological and transcriptional traits of soybean (*Glycine max* L.) in response to water deficit: a review, Physiologia Plantarum, 172(2): 405-418.
https://doi.org/10.1111/ppl.13201

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/tpj.13029

Yu K., Miao H., Liu H., Zhou J., Sui M., Zhan Y., Xia N., Zhao X., and Han Y., 2022, Genome-wide association studies reveal novel QTLs, QTL-by-environment interactions and their candidate genes for tocopherol content in soybean seed, Frontiers in Plant Science, 13: 1026581.
https://doi.org/10.3389/fpls.2022.1026581

Zhang Y., Wang Y., Zhou W., Zheng S., and Ye R., 2021, Detection of candidate gene networks involved in resistance to *Sclerotinia sclerotiorum* in soybean, Journal of Applied Genetics, 63: 1-14.
https://doi.org/10.1007/s13353-021-00654-z