Genetic Mechanisms of Crop Disease Resistance: New Advances in GWAS

Jiang Cheng ✉

Institute of Life Sciences, Jiyang College of Zhejiang A&F University, Zhuji, 311800, China

✉ Corresponding author email: 2397383131@qq.com

Abstract The genetic mechanism of crop disease resistance is one of the key research contents in the field of agricultural biotechnology. In recent years, genome-wide association study (GWAS), as an effective genetic analysis tool, has made remarkable progress in revealing the genetic basis of crop disease resistance. This study summarized the application of GWAS in crop disease resistance research, including the principle and method of association analysis, the discovery of disease-resistance-related genes, and discussed the application of GWAS in disease resistance breeding, as well as the challenges and directions of future research. By summarizing these research results, this study can further understand the genetic mechanism of crop disease resistance and provide scientific basis for crop disease resistance breeding.

Keywords Crop disease resistance; Genome-wide association study; Genetic mechanism; Resistance gene; Breeding technology

Crops are important food and cash crops for mankind, supporting the livelihoods and livelihood needs of hundreds of millions of people around the world. However, crop growth and development are often threatened by bacteria, fungi, viruses and other pathogens, which can lead to crop loss, quality degradation and even death. It seriously affects the stability and sustainability of agriculture (Fones et al., 2020). Crop diseases are estimated to cause tens of billions of dollars in economic losses globally each year, posing serious challenges to food security and farmers' livelihoods.

In order to effectively deal with crop diseases, scientists have long been committed to exploring the genetic mechanism of crop disease resistance. Disease resistance is the result of the interaction between crops and pathogens, and has a complex genetic basis. Traditionally, some disease-resistant varieties have been bred through selection, hybridization, etc., but this method has some problems such as slow genetic progress, time-consuming and labor-consuming, and unstable disease resistance (Deng et al., 2020). With the rapid development of molecular biology and genomics technology, researchers began to explore the use of molecular genetics and genomics to reveal the genetic mechanism of crop disease resistance (Figure 1), in order to improve the resistance of crops more effectively.

Genome-wide association study (GWAS), as a high-throughput genetic analysis method, has been widely used in crop disease resistance research in recent years. GWAS has numerous advantages over traditional breeding methods, such as its ability to reveal the genetic basis of complex traits by simultaneously detecting associations between thousands of natural variation loci and traits (Uffelmann et al., 2021). GWAS is not limited by genome structure and can be carried out without relying on reference varieties. GWAS can also uncover new associations between genotypes and phenotypes, providing new insights into the molecular mechanisms of crop disease resistance.

This study systematically summarized the application progress of GWAS in crop disease resistance research in recent years, including the method principle, case studies and achievements, discussed the advantages and limitations of GWAS in revealing the genetic mechanism of crop disease resistance, and prospected the development trend of GWAS in crop disease resistance research in the future. It is hoped that this study can provide reference and inspiration for relevant researchers, promote the research of crop disease resistance to make new breakthroughs, and contribute to food security and sustainable agricultural development.
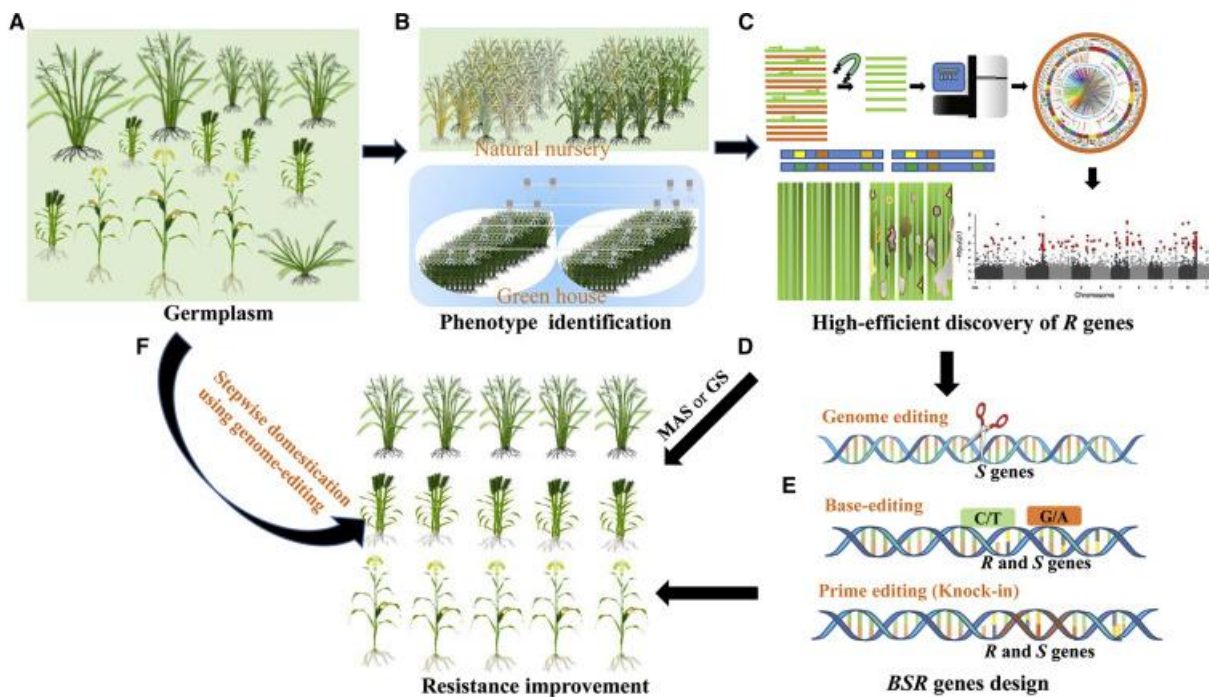
Figure 1 Breeding strategies for improving disease resistance in crops (Deng et al., 2020)

1 Principles and Methods of GWAS Research

1.1 The fundamentals of GWAS

Genome-wide association study (GWAS) is an important method to reveal the genetic basis of complex traits. Its core principle is to search for genetic variation loci (SNPS, single nucleotide polymorphisms) associated with target traits by comparing genotype differences between individuals or populations with different phenotypic characteristics (Tam et al., 2019). GWAS typically involves the collection of large-scale genotypic and phenotypic data and the analysis of these data using statistical methods to determine associations. Through GWAS studies, researchers can discover candidate genes associated with a target trait and gain insight into the genetic mechanisms of that trait.

The basic GWAS process consists of the following steps: Researchers select a representative population, including individuals exhibiting different traits, and then collect genomic and phenotypic data of these individuals, using high-throughput sequencing technology or gene chip technology to obtain genotype data of individuals, often involving millions of SNPS (Uffelmann et al., 2021). After data preparation is complete, the association between genotype and phenotype is analyzed using statistical methods to determine which genetic variants are associated with the target trait, and finally the results are validated and interpreted by the researchers, and further functional studies may be conducted to reveal the underlying biological mechanisms.

1.2 Statistical methods of association study

Association study is a core step in genome-wide association study that is used to determine the association between genotype and phenotype (Figure 2), and when conducting association study, researchers typically use a variety of statistical methods to assess the relationship between genetic variation sites (such as SNPS) and target traits. Commonly used statistical methods include Chi-square test, linear regression, Logistic regression, mixed model, etc. (Zeng et al., 2015).

Chi-square test is a commonly used non-parametric test method to compare whether the difference between the observed frequency and the expected frequency is significant. Linear regression is suitable for analyzing the association between continuous traits and SNPS, and evaluating the relationship between genotype and phenotype by fitting linear models (Zeng et al., 2015). Logistic regression is often used to analyze bivariate disaggregated data, such as the occurrence or absence of disease, to assess the effect of SNP genotypes on disease risk. The

hybrid model combines fixed effects and random effects, and can take into account the influence of factors such as population structure and kinship, which improves the accuracy of association analysis. When conducting large-scale association analysis, multiple comparison correction is required to control the error rate.
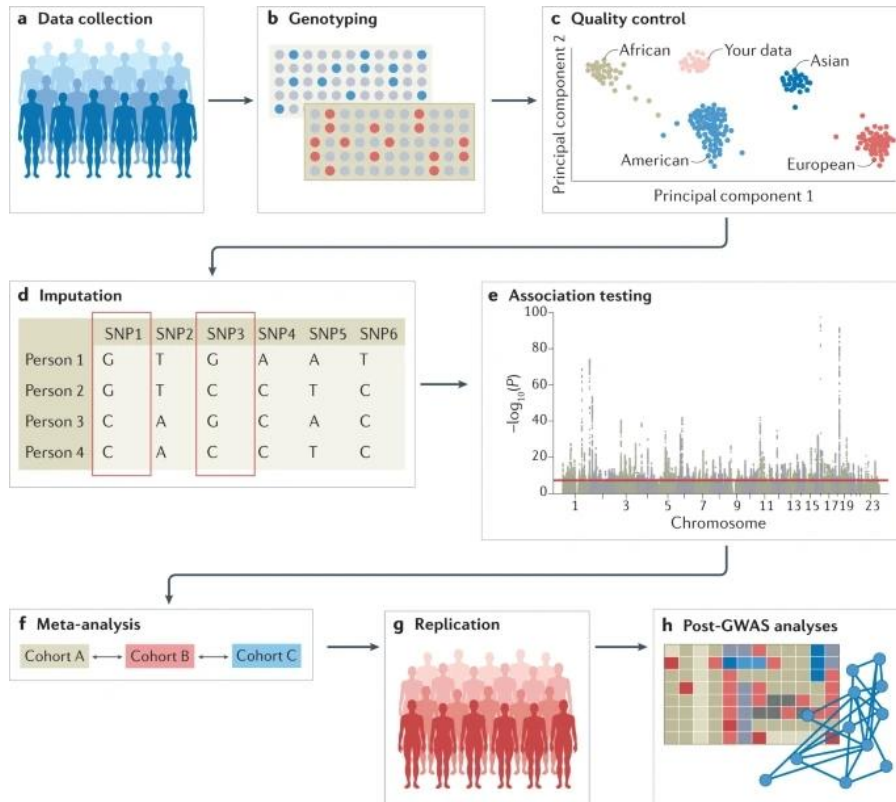

Figure 2 The basic flow of GWAS (Uffelmann et al., 2021)

1.3 Identification and validation of candidate genes

The identification and validation of candidate genes is a crucial step in genome-wide association study, which helps researchers identify potentially functional genes associated with target traits. In GWAS, SNPS identified by association study are typically located in the genome and may involve multiple genes. In order to identify specific functional genes, researchers need to conduct further identification and validation (Pang et al., 2020).

The identification of candidate genes usually includes the following steps: Firstly, SNPS highly correlated with target traits are identified based on the results of association study, and then candidate genes around these SNPS are identified by genome annotation and other methods (Pang et al., 2020). Subsequently, bioinformatics tools and experimental validation techniques are used to perform functional analysis and validation of these candidate genes to determine their association with target traits.

In the verification process of candidate genes, researchers usually use a variety of methods to verify the relationship between candidate genes and target traits, including functional experiments (such as gene knockout, expression analysis, etc.), transgenic model construction and analysis, and population and family studies. Through these methods, researchers can verify the impact of candidate genes on target traits and further understand their mechanisms of action in biological processes.

2 Genetic Basis of Crop Disease Resistance

2.1 Discovery of disease-resistance related genes

The genetic basis of crop disease resistance is a complex and diverse systematic trait, which is influenced by multiple genes and environmental factors. The discovery of disease-resistant genes is of great significance for understanding the genetic basis of crop disease resistance (van Esse et al., 2020). Through genome-wide association study (GWAS) and other methods, researchers have identified many genes associated with crop

disease resistance. These genes can be divided into two types: direct resistance genes and genes regulating resistance response. Direct resistance genes encode resistance proteins and can interact with pathogens to trigger plant resistance response, while genes regulating resistance response affect plant recognition and signal transduction of pathogens and regulate plant immune response.

For example, in rice, a series of important disease-resistance related genes have been identified, such as *Xa21*, *Xa4*, *Xa5*, etc. These genes encode disease-resistant proteins that are effective against common rice pathogens, such as Fusarium oryzae (Zhang et al., 2019). Similarly, many genes related to disease resistance have been found in wheat (*Triticum aestivum*), maize (*Zea mays*), soybean (*Glycine max*) and other crops, providing important molecular markers and candidate genes for crop resistance breeding. The genetic basis of crop disease resistance is a complex system, which is co-regulated by several genes. The discovery of disease-resistance-related genes provides important clues for further understanding the molecular mechanism of crop disease resistance, and provides important genetic resources and molecular markers for crop disease resistance breeding.

2.2 Function and mechanism of resistance genes

Resistance genes are genes that produce resistance response against pathogen infection in plants. These genes are involved in regulating plant disease resistance by encoding specific proteins or regulating gene expression. The function and mechanism of resistance genes can be divided into direct resistance and regulation of resistance response (Garrett et al., 2017). Proteins encoded by direct disease resistance genes usually have specific disease resistance effects. For example, proteins encoded by some genes can bind to specific components of pathogens, trigger plant immune responses, such as activating signal transduction pathways, producing toxins, etc., and ultimately lead to the death of pathogens or inhibit their infection. These proteins may be directly involved in the physical blocking or chemical attack of a pathogen-infected site to ensure the health of the plant.

Another class of resistance genes regulates gene expression and activation of resistance response. These genes may encode proteins in signal transduction pathways (Garrett et al., 2017), such as kinases or transcription factors, that can influence a plant's disease resistance response by regulating the expression of other genes. For example, some genes can activate the synthesis of antibacterial substances, enhance the strength of cell walls, promote cell death, and so on, thereby enhancing the plant's resistance to pathogens.

2.3 Genetic diversity and evolution of resistance genes

The genetic diversity and evolution of resistance genes is an important part of crop disease resistance research, which is of great significance for understanding the evolution and resistance mechanism of resistance genes. Genetic diversity of resistance genes refers to differences in the types and amounts of resistance genes present in different crop varieties or germplasm resources (Andersen et al., 2016). Because different crop varieties or germplasm resources are affected by natural selection, artificial selection and other factors, there are certain differences in the composition of resistance genes, which is reflected in the diversity of the types of resistance genes, genotype frequency and other aspects. The existence of genetic diversity provides rich genetic resources for crop disease resistance breeding and helps to breed new varieties with more disease resistance.

The evolution of resistance genes refers to the origin, differentiation and propagation of resistance genes during the evolution of species (Andersen et al., 2016). Resistance genes can be produced in many ways, such as gene mutation, gene recombination, gene transfer, etc. In the long-term adaptive evolution process of species, resistance genes continue to mutate and accumulate, forming different genotypes. These genotypes are subject to selection pressures in the environment, resulting in changes in the frequency of different genotypes that affect the genetic diversity and evolution of resistance genes.

3 New Progress in GWAS Research

3.1 GWAS case studies of resistance to different crop diseases

In 2020, Voichek and Weigel's research team published a paper in *Nature Genetics* entitled 'Identifying genetic variants underlying phenotypic variation in plants without complete genomes'. The study explored the genetic variation behind phenotypic variation in plants by using GWAS without reference to the genome, and the results

demonstrated the ability to perform GWAS before linking sequence reads to specific genomic regions. This allows the detection of a wider range of genetic variants that cause phenotypic variation (Voichek and Weigel, 2020). The method can be applied not only to species with complete genome sequences, but also to those that have not been fully sequenced, providing a new way to identify the genetic basis of important agronomic traits such as disease resistance.

In 2017, Posthuma's research team published a study in *Nature Communications* using FUMA for functional mapping and annotation of genetic associations. FUMA can take GWAS summary statistics as input, prioritize functional SNPS and genes, and map genes based on function. This process includes localization mapping, eQTL mapping, and chromatin interaction mapping to identify key genes and biological processes associated with diseases or traits (Watanabe et al., 2017), and the application of this tool could improve researchers' understanding of the genetic basis of disease resistance in crops such as rice.

3.2 Biological interpretation and functional validation of GWAS results

Genome-wide association study (GWAS) is a powerful tool to help identify genetic variants associated with complex traits such as crop disease resistance, but associations identified by GWAS alone do not fully explain the underlying biological mechanisms, so biological interpretation and functional validation of GWAS results are essential.

Biological interpretation usually involves functional annotation and pathway analysis of GWAS-associated genes to understand the roles and interrelationships of these genes in biological processes, This can be achieved through Gene ontology analysis, pathway enrichment analysis, and protein interaction network analysis (Ritchie and Steen, 2018), which can reveal biological pathways and key regulators associated with crop disease resistance.

Functional verification is a key step to confirm GWAS results, which usually includes gene knockout, gene expression analysis, protein interaction experiments, etc. (Gallagher and Chen-Plotkin, 2018). For example, gene editing technology (such as CRISPR-Cas9) can be used to knock out candidate genes identified by GWAS, observe the impact on crop disease resistance, and verify the expression pattern and level of candidate genes in the process of disease resistance through gene expression analysis.

3.3 Application of GWAS research in disease resistance breeding

Genome-wide association study (GWAS) has made significant progress in disease resistance breeding, and by analyzing large-scale genotypic and phenotypic data, GWAS can help identify candidate genes and molecular markers associated with disease resistance. These genes and markers can be used as important genetic resources for crop disease resistance breeding (Gyawali et al., 2018), accelerating the process of crop disease resistance breeding. Traditional crop disease resistance breeding usually requires a long time of observation and selection, while GWAS can quickly identify disease-resistation-related genes, shorten the breeding cycle and improve breeding efficiency.

GWAS can also improve the precision of breeding, helping researchers to more precisely understand the genetic basis of crop resistance and select more effective breeding methods and strategies. The disease-resistance-related molecular markers identified by GWAS can be used in marker-assisted selection (MAS) to help breeders more effectively select crop varieties with disease resistance genotypes and speed up the breeding process. It has a wide application prospect in disease resistance breeding, provides important scientific basis and technical support for crop disease resistance improvement, helps to breed new varieties with more disease resistance, and promotes the sustainable development of agricultural production.

4 Challenges and Future Directions

4.1 Limitations and challenges of GWAS research

Although genome-wide association study has made remarkable progress in revealing the genetic basis of crop disease resistance, there are some limitations and challenges. GWAS requires the support of a large sample size to obtain statistically significant results. For complex traits such as crop disease resistance, larger samples may be

needed to ensure the reliability of the results. GWAS results are also susceptible to factors such as population structure and kinship, and appropriate correction is needed to reduce false positive results (Uffelmann et al., 2021).

The GWAS results require further validation and functional studies to identify genes associated with resistance. GWAS can usually identify only a few candidate genes or regions, and further molecular biology experiments are required to verify the functional association between these genes and resistance, which requires significant time and resources and may face technical and methodological challenges (Uffelmann et al., 2021). GWAS is also limited by the quality of genotype data and the collection of phenotypic data, the quality and coverage of genotype data directly affect the reliability of analysis results, and the accuracy and consistency of phenotypic data are also key factors to ensure the success of research.

4.2 The importance of multi-omics data integration
Multi-omics data integration plays an important role in crop disease resistance research. With the development of high-throughput techniques such as genome-wide association study, researchers have access to a large amount of genotype and phenotypic data, but analyzing these data alone may not fully reveal the genetic mechanisms of crop disease resistance. Multi-omics data integration can comprehensively consider data at different levels (Choi, 2019), such as genome, transcriptome, proteome, metabolome, etc., so as to more comprehensively understand the formation and regulation mechanism of crop disease resistance.

By integrating multiple omics data, researchers can discover correlations between data at different levels and identify key genes and regulatory networks that influence crop disease resistance (Choi, 2019). For example, a gene may not be significantly associated at the genomic level, but may show a distinct difference at the transcriptome or proteome level, and this association may have important implications for crop resistance. Multi-omics data integration can also help researchers discover new biomarkers and provide more accurate markers for molecular breeding of crop disease resistance.

4.3 Application prospect of gene editing technology in disease resistance improvement
As a revolutionary genome engineering tool, gene editing technology provides a new approach and possibility for the improvement of crop disease resistance. Through gene editing technology, researchers can accurately edit genes related to disease resistance in crop genome (Zaidi et al., 2018), so as to achieve rapid and accurate improvement of crop disease resistance.

One of the biggest advantages of gene editing technology is that it can achieve highly accurate genome modification, which can be precisely edited for specific gene loci, avoiding the time and labor costs of multi-generation mating in traditional breeding methods. By editing genes known to be associated with disease resistance, new varieties with strong resistance can be rapidly bred (Ahmar et al., 2021).

Gene editing technology can also help researchers study and understand the molecular mechanisms of crop disease resistance. By editing different gene loci, we can verify the exact contribution of these genes to crop disease resistance and reveal the molecular mechanism of disease resistance. As gene-editing technology continues to develop and improve, researchers believe it will play an increasingly important role in improving crop disease resistance. Through gene editing technology, researchers can more quickly and accurately breed new crop varieties that are more resistant to disease and have higher yields, providing strong support for solving problems such as food security and sustainable agricultural development.

5 Outlook
With the development and application of genome-wide association study (GWAS), many new advances have been made in the study of genetic mechanisms of crop disease resistance. In the future, as the technology continues to innovate and progress, researchers can expect more GWAS to be carried out in different crops and disease resistance to reveal more resistance-related genes and regulatory networks. With the application of new technologies such as single-cell sequencing and spatial transcriptomics, researchers can better understand the

molecular mechanisms of crop resistance, including the expression and regulation of genes at the level of individual cells.

With the continuous advancement of bioinformatics and artificial intelligence technologies, researchers can better process and analyze large-scale GWAS data, discover more robust and reliable gene-phenotype associations, and accelerate the identification and verification process of disease resistance genes. Functional verification and mechanism analysis of GWAS research results are also important directions for future research, and the function of disease resistance genes is verified by gene editing and other technologies, and the mechanism of action in crop disease resistance is deeply explored.

GWAS research also needs to pay more attention to the combination with actual breeding, quickly transform the discovered disease resistance genes into breeding applications, and speed up the breeding process of new varieties. Researchers also need to pay attention to the differences and commonalities of genetic mechanisms of disease resistance between different crops, so as to provide more comprehensive theoretical support for disease resistance breeding of different crops. GWAS research has great significance and potential in revealing the genetic mechanism of crop disease resistance. Future studies will pay more attention to the integration of multi-omics data, the application of new technologies and the combination with breeding practice, so as to provide more scientific basis and methodological support for crop disease resistance breeding and help the sustainable development of food production.

References

Ahmar S., Mahmood T., Fiaz S., Mora-Poblete F., Shafique M.S., Chattha M.S., and Jung K.H., 2021, Advantage of nanotechnology-based genome editing system and its application in crop improvement, Front. Plant Sci., 12: 663849.
https://doi.org/10.3389/fpls.2021.663849
PMid:34122485 PMCid:PMC8194497

Andersen E.J., Ali S., Reese R.N., Yen Y., Neupane S., and Nepal M.P., 2016, Diversity and evolution of disease resistance genes in barley (Hordeum vulgare L.), Evolutionary Bioinformatics, 12: 99-108.
https://doi.org/10.4137/EBO.S38085
PMid:27168720 PMCid:PMC4857794

Choi H.K., 2019, Translational genomics and multi-omics integrated approaches as a useful strategy for crop breeding, Genes & Genomics, 41: 133-146.
https://doi.org/10.1007/s13258-018-0751-8
PMid:30353370 PMCid:PMC6394800

Deng Y., Ning Y., Yang D.L., Zhai K., Wang G.L., and He Z., 2020, Molecular basis of disease resistance and perspectives on breeding strategies for resistance improvement in crops, Molecular Plant, 13(10): 1402-1419.
https://doi.org/10.1016/j.molp.2020.09.018
PMid:32979566

Fones H.N., Bebber D.P., Chaloner T.M., Kay W.T., Steinberg G., and Gurr S.J., 2020, Threats to global food security from emerging fungal and oomycete crop pathogens, Nat. Food, 1(6): 332-342.
https://doi.org/10.1038/s43016-020-0075-0
PMid:37128085

Gallagher M.D., and Chen-Plotkin A.S., 2018, The post-GWAS era: From association to function, Am. J. Hum. Genet., 102(5): 717-730.
https://doi.org/10.1016/j.ajhg.2018.04.002
PMid:29727686 PMCid:PMC5986732

Garrett K.A., Andersen K.F., Asche F., Bowden R.L., Forbes G.A., Kulakow P.A., and Zhou B., 2017, Resistance genes in global crop breeding networks, Phytopathology, 107(10): 1268-1278.
https://doi.org/10.1094/PHYTO-03-17-0082-FI
PMid:28742460

Gyawali S., Chao S., Vaish S.S., Singh S.P., Rehman S., Vishwakarma S.R., and Verma R.P.S., 2018, Genome wide association studies (GWAS) of spot blotch resistance at the seedling and the adult plant stages in a collection of spring barley, Molecular Breeding, 38: 62.
https://doi.org/10.1007/s11032-018-0815-0

Pang Y., Liu C., Wang D., Amand P.S., Bernardo A., Li W., He F., Li L., Wang L., Yuan X., Dong L., Su Y., Zhang H., Zhao M., Liang Y., Jia H., Shen X., Lu Y., Jiang H., Wu Y., Li A., Wang H., Kong L., Bai G., and Liu S., 2020, High-resolution genome-wide association study identifies genomic regions and candidate genes for important agronomic traits in wheat, Mol. Plant, 13(9): 1311-1327.
https://doi.org/10.1016/j.molp.2020.07.008
PMid:32702458

Ritchie M.D., and Steen K.V., 2018, The search for gene-gene interactions in genome-wide association studies: challenges in abundance of methods, practical considerations, and biological interpretation, Ann. Transl. Med., 6(8): 157.

https://doi.org/10.21037/atm.2018.04.05

PMid:29862246 PMCid:PMC5952010

Tam V., Patel N., Turcotte M., Bossé Y., Paré G., and Meyre D., 2019, Benefits and limitations of genome-wide association studies, Nature Reviews Genetics, 20: 467-484.

https://doi.org/10.1038/s41576-019-0127-1

PMid:31068683

Uffelmann E., Huang Q.Q., Munung N.S., de Vries J., Okada Y., Martin A.R., Martin H.C., Lappalainen T., and Posthuma D., 2021, Genome-wide association studies, Nature Reviews Methods Primers, 1: 59.

https://doi.org/10.1038/s43586-021-00056-9

van Esse H.P., Reuber T.L., and van der Does D., 2020, Genetic modification to improve disease resistance in crops, New Phytologist, 225(1): 70-86.

https://doi.org/10.1111/nph.15967

PMid:31135961 PMCid:PMC6916320

Voichek Y., and Weigel D., 2020, Identifying genetic variants underlying phenotypic variation in plants without complete genomes, Nature Genetics, 52: 534-540.

https://doi.org/10.1038/s41588-020-0612-7

PMid:32284578 PMCid:PMC7610390

Watanabe K., Taskesen E., van Bochoven A., and Posthuma D., 2017, Functional mapping and annotation of genetic associations with FUMA, Nature Communications, 8: 1826.

https://doi.org/10.1038/s41467-017-01261-5

PMid:29184056 PMCid:PMC5705698

Zaidi S.S., Mukhtar M.S., and Mansoor S., 2018, Genome editing: Targeting susceptibility genes for plant disease resistance, Trends Biotechnol., 36(9): 898-906.

https://doi.org/10.1016/j.tibtech.2018.04.005

PMid:29752192

Zeng P., Zhao Y., Qian C., Zhang L., Zhang R., Gou J., Liu J., Liu L., and Chen F., 2015, Statistical analysis for genome-wide association study, J. Biomed. Res., 29(4): 285-297.

https://doi.org/10.7555/JBR.29.20140007

PMid:26243515 PMCid:PMC4547377

Zhang M., Wang S., and Yuan M., 2019, An update on molecular mechanism of disease resistance genes and their application for genetic improvement of rice, Molecular Breeding, 39: 154.

https://doi.org/10.1007/s11032-019-1056-6