

## Application of Genome-wide Association Study in Crop Disease Resistance Breeding

Fu Cheng ✉

Hainan Key Laboratory of Crop Molecular Breeding, Sanya, 572000, China

✉ Corresponding author email: [2397383131@qq.com](mailto:2397383131@qq.com)

Field Crop, 2024, Vol.7, No.1 doi: [10.5376/fc.2024.07.0001](https://doi.org/10.5376/fc.2024.07.0001)

Received: 05 Dec., 2023

Accepted: 08 Jan., 2024

Published: 25 Jan., 2024

**Copyright © 2024** Cheng. This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Preferred citation for this article:**

Cheng F., 2024, Application of genome-wide association study in crop disease resistance breeding, Field Crop, 7(1): 1-8 (doi: [10.5376/fc.2024.07.0001](https://doi.org/10.5376/fc.2024.07.0001))

**Abstract** Genome-wide association study (GWAS), as an effective genetic research tool, has been widely used in crop disease resistance breeding, which can identify genetic markers and genes related to disease resistance in the whole genome and provide molecular basis for breeding. This study introduced the basic principles and methods of GWAS, demonstrated the application of GWAS in crop disease resistance breeding through specific application examples, then discussed the advantages and limitations of GWAS in crop disease resistance breeding, and prospected the future development direction of GWAS in crop disease resistance breeding. This includes applications that combine high-throughput sequencing techniques, multi-omics data integration, and precision breeding techniques. GWAS provides a new research idea and method for crop disease resistance breeding, which is expected to promote the rapid cultivation of disease-resistant varieties and the sustainable development of agricultural production.

**Keywords** Genome-wide association study; Crop disease resistance; Breeding; Genetic marker; Precision breeding

The importance of crop disease resistance breeding should not be ignored, especially in the current globalized agricultural production environment, where diseases not only lead to direct loss of crop yield, but also may lead to decline in quality and market value, and even lead to food security crisis in serious cases (Mores et al., 2021). Therefore, improving crop disease resistance can effectively reduce the use of pesticides, reduce environmental pollution, and improve the economic efficiency and sustainability of agricultural production.

Although traditional disease resistance breeding methods have made certain achievements in history, with the rapid evolution of pathogens and changes in ecological environment, these methods have been difficult to meet the current breeding needs, and traditional breeding methods often rely on long-term phenotypic selection, and are limited by the diversity and availability of genetic resources. Traditional methods are also limited in resolving the genetic basis of complex traits, and it is difficult to accurately locate and utilize key genes related to disease resistance (Lamichhane and Thapa, 2022).

The rise of genome-wide association study (GWAS) has brought new opportunities for crop disease resistance breeding, which can explore the association between genetic variation and trait phenotype in the whole genome, and provide a powerful tool for revealing the genetic mechanism of disease resistance. Through GWAS, researchers can quickly identify genetic markers and candidate genes associated with disease resistance (Osorio-Guarin et al., 2020), and this information is of great value in guiding molecular marker-assisted breeding (MAS) and gene directed editing.

This study summarized and analyzed the application and significance of genome-wide association study in crop disease resistance breeding, and discussed its application cases in different crop disease resistance breeding from the principle and method of GWAS, and analyzed its advantages and limitations. This study also looks forward to the future development of GWAS in crop resistance breeding, including the combination of multi-omics data, the use of high-throughput sequencing technology, and the integration of gene editing technology. Through this study, we hope to provide new ideas and strategies for crop disease resistance breeding, promote scientific and precise crop disease management, and improve the sustainability and stress resistance of agricultural production.

# 1 Principles and Methods of GWAS

## 1.1 Definition and principle of GWAS

Genome-wide association study (GWAS) is a genetic method used to study the relationship between genetic variation and phenotypic traits. Its principle is to explore the association between genetic variation (such as single nucleotide polymorphisms, SNPs) and specific traits through statistical analysis based on genotype and phenotypic data of a large number of individuals (Uffelmann et al., 2021). This analysis can reveal the role of genetic variation in the expression of traits and provide clues for understanding the genetic basis of traits.

The key of GWAS is to conduct high density genotyping of a large number of samples to cover the variation information of the whole genome, and to evaluate the correlation between each genetic variation site and the trait phenotype through statistical tests (such as linear regression analysis). If the frequency of a mutation site is significantly different in different phenotype populations, it is considered that the site is associated with the trait phenotype. GWAS can fully explore the genome without relying on prior knowledge, reveal the polygenic genetic basis of complex traits, and provide molecular markers for breeding.

## 1.2 The main steps of GWAS

Genome-wide association study (GWAS) is a widely used genetic tool in crop disease resistance breeding, which reveals genetic markers and genes associated with traits by analyzing the association between genetic variation and phenotypic traits. The main steps of GWAS include sample collection and genotyping, phenotypic data collection, association analysis, result validation, and biological interpretation (Figure 1). It requires collecting a sufficient number of samples and conducting genotyping to obtain genome-wide genetic variation information. The phenotypes of the samples were then recorded in detail (Belzile and Torkamaneh, 2022). These data will be used for subsequent association analysis, which will then use statistical methods to analyze the association between genotype and phenotypic data to identify genetic markers associated with specific traits.

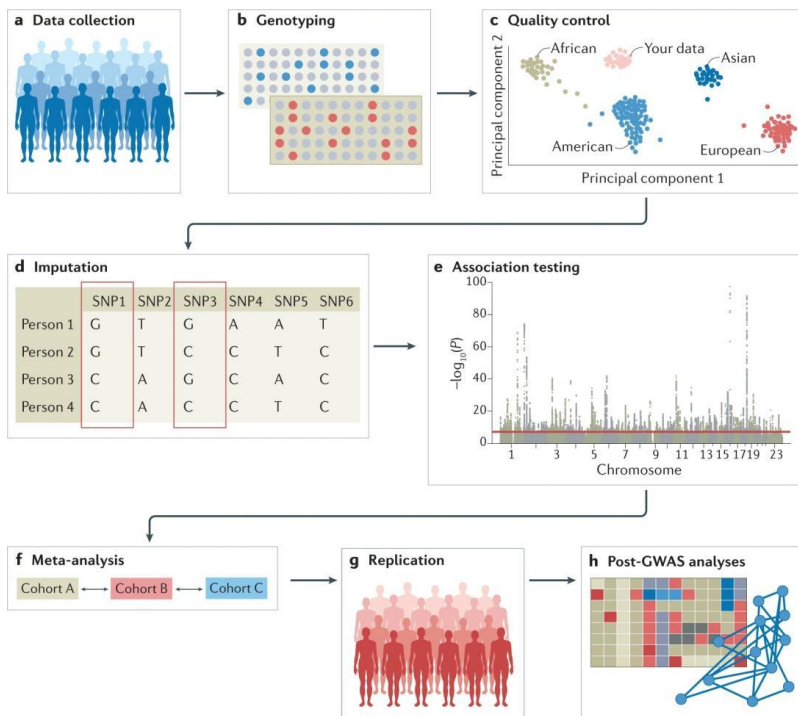


Figure 1 The steps for conducting GWAS (Uffelmann et al., 2021)

The identified associated sites or candidate genes need to be verified through independent sample sets or functional verification experiments. Finally, the biological interpretation of the identified associated sites or candidate genes is carried out to explore their roles and mechanisms in phenotype formation. Through these steps, GWAS can provide valuable genetic information for crop disease resistance breeding.

### 1.3 Data and technical requirements required by GWAS

As a powerful genetic research tool, the application of genome-wide association study (GWAS) in crop disease resistance breeding requires certain data and technical requirements. GWAS requires large amounts of genetic variation data, typically from high-density single nucleotide polymorphism (SNP) markers or whole genome sequencing, which are used to construct genetic linkage maps (Marees et al., 2018) to identify genetic loci associated with trait phenotypes.

GWAS also requires accurate phenotypic data. In crop disease resistance studies, this usually involves detailed assessment of disease responses of plants of different genotypes, and the accuracy of phenotypic data directly affects the reliability of GWAS results. GWAS also requires strong statistical analysis capabilities, association analysis requires processing large amounts of data, and mathematical control of false positive rates, the use of appropriate statistical models and correction methods (such as Bonferroni correction or false discovery rate control) is necessary (Marees et al., 2018). With the advancement of technology, the data and technical requirements of GWAS are also increasing, for example, with the reduction of sequencing costs, whole genome sequencing has gradually become an important data source for GWAS, and the development of bioinformatics tools has also provided more possibilities for the processing and analysis of GWAS data.

## 2 Application of GWAS in Crop Disease Resistance Breeding (A Case Study of Soybean)

Soybean (*Glycine max*) is an important crop in the legume family, which is widely favored worldwide for its high protein and oil content. Soybean is not only an important source of human food, such as tofu, soybean milk and soybean oil, but also contains a variety of bioactive substances that are beneficial to the human body, such as isoflavones and lecithin. These ingredients have been shown to have positive effects on cardiovascular health, bone health and menopausal symptoms, and are a major component of many animal feeds, having an irreplaceable impact on the global agriculture and food industry.

The origin of soybeans can be traced back to China, it has thousands of years of cultivation history, has become one of the most widely cultivated crops in the world, in agricultural production, soybeans are not only an important cash crop, but also a key component of sustainable agriculture. Major challenges include sudden death syndrome (SDS) caused by *Fusarium virguliforme*, one of the key diseases limiting its production. At present, the genetic mechanism of soybean resistance to SDS, especially the epistatic role between genes, is still not fully understood. Mueller and Singh's team published a paper in *The Plant Journal* in 2015 titled 'Genome-wide association and epistasis studies unravel the genetic architecture of sudden death syndrome resistance in soybean'. By analyzing the genetic data of 214 soybean varieties and 31 914 single nucleotide polymorphism (SNP) markers, this study conducted a comprehensive genomic association analysis and epistatic role study (Figure 2), aiming to further explore the genetic background of soybean resistance to SDS (Zhang et al., 2015).

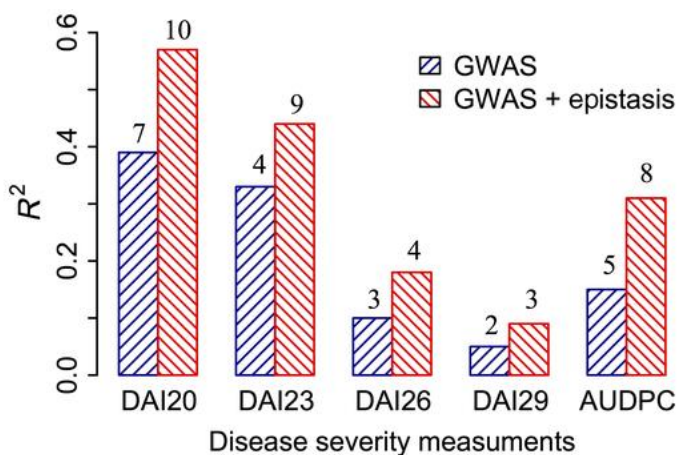


Figure 2 Contributions of identified sudden death syndrome loci via genome-wide association studies (GWAS) and epistatic analysis to the phenotypic variance of each disease severity measurement (Zhang et al., 2015)

Twelve key sites associated with SDS resistance and 12 interactions between SNPS and SNPS were identified. The additive and epistatic effects of these loci together contributed 24% to 52% of phenotypic variation. In the vicinity of these key SNPS, genes associated with disease resistance, pathogenesis, chitin response, and wound healing were also identified, in particular a trait associated SNP-locked-stress-induced receptor-like kinase gene 1 (SIK1) encoding a protein rich in leucine repeats. This study emphasizes that epistatic effects must be taken into account in breeding for SDS resistance in soybean to improve the explanation of phenotypic variation. Accordingly, the researchers also constructed a soybean root model for SDS pathogen defense (Figure 3). The findings of this study not only reveal the molecular mechanism of soybean resistance to SDS, but also provide a scientific basis for future anti-SDS breeding strategies based on genetic epistasis.

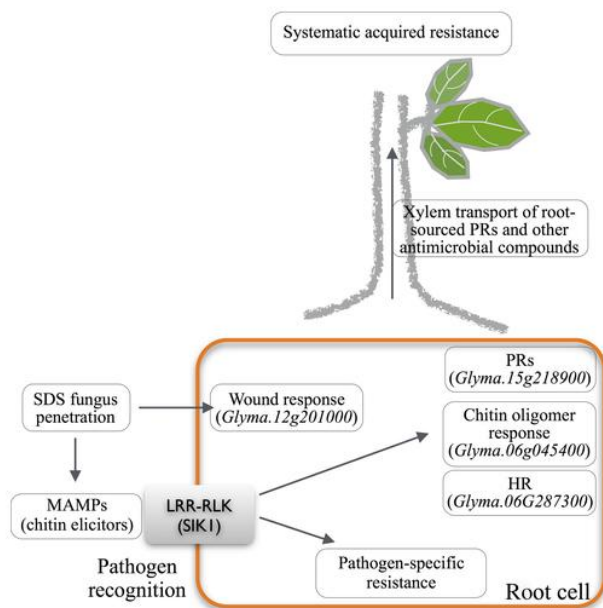


Figure 3 Putative model for soybean defense against sudden death syndrome (SDS) based on the results of genome-wide association and epistasis studies (Zhang et al., 2015)

### 3 The Advantages and Limitations of GWAS in Crop Disease Resistance Breeding

#### 3.1 The advantages of GWAS compared with traditional breeding methods

Genome-wide association study (GWAS) has obvious advantages over traditional breeding methods in crop disease resistance breeding. GWAS has high efficiency, it can quickly identify the genetic markers and genes related to disease resistance within the whole genome in a short time, and accelerate the breeding process. There is no need for specific genetic background, and it can make use of existing natural population and variety resources for analysis (Tam et al., 2019), without the need to build specific genetic populations, reducing the cost and time of research.

GWAS can also reveal the genetic structure of complex traits and identify multi-genes and inter-gene interactions that control complex traits, which is of great significance for understanding the genetic mechanism of crop disease resistance and guiding molecular marker-assisted breeding (MAS) (Tam et al., 2019). GWAS can also utilize existing natural population and variety resources without the need to construct large-scale genetic populations, thus reducing the cost and time of research, which makes GWAS an efficient breeding tool, especially suitable for research environments with limited resources.

#### 3.2 Limitations and challenges of GWAS

Although genome-wide association study (GWAS) has made remarkable progress in crop disease resistance breeding, it still faces some limitations and challenges. Population structure and linkage imbalance (LD) are major challenges for GWAS. Population structure refers to differences in genetic background in the sample, which can lead to false positive associations, and linkage imbalance refers to non-random associations between different loci,

which can affect the accurate localization of related trait loci. When designing GWAS, researchers need to adopt appropriate statistical methods and correction measures to reduce the impact of these factors (Heeney, 2021). GWAS has a limited ability to detect rare variants. In crops, some important resistance traits may be controlled by lower-frequency variants, and GWAS is limited in its ability to detect these rare variants, which requires larger sample sizes and higher density of genotype data to improve detection capabilities.

The biological interpretation and functional verification of GWAS results is also one of its challenges. GWAS can identify genetic markers associated with traits, but these markers may not be directly involved in the regulation of traits, which requires subsequent biological experiments to verify the function of these associated sites and reveal their mechanism of action in disease resistance (Ciochetti et al., 2023). Large amount of data and computational complexity are another difficulty in GWAS research. With the development of sequencing technology, the amount of genetic data generated is increasing, which puts higher requirements on data storage and analysis, and more efficient computational methods and software tools need to be developed to deal with these large-scale data.

### **3.3 Strategies to improve the efficiency and accuracy of GWAS research**

Improving the efficiency and accuracy of genome-wide association study (GWAS) is essential to reveal the genetic basis of complex traits such as crop disease resistance and is an important topic in current genetic research. To improve the efficiency and accuracy of GWAS research, it is necessary to comprehensively consider many factors, such as sample size, gene chip density, phenotypic data quality, population structure control, multi-omics data integration, functional validation, meta-analysis and repeated validation. By adopting these strategies, genetic variants associated with important traits such as crop disease resistance can be found more effectively, providing powerful molecular tools for crop breeding.

Increasing the sample size can improve the statistical power of GWAS and help detect more genetic variation related to traits, and a large sample size can also help reduce the incidence of false positives (Spencer et al., 2009). The use of high-density gene chips can improve the coverage of genetic variation and increase the chance of detecting genetic markers associated with traits. Accurate and reliable phenotypic data are key to the success of GWAS, and improved methods for collecting and measuring phenotypic data, as well as the use of standardized phenotypic evaluation systems, can improve the accuracy of studies.

Population structure and kinship can influence the results of GWAS, and the use of appropriate statistical models or methods (such as mixed linear models) to control for these factors can reduce the occurrence of false positives. Integrating multi-omics data such as transcriptomics, proteomics, and metabolomics can provide a more comprehensive biological context to help interpret GWAS results and uncover potential functional genetic variants. Functional verification of candidate genes identified by GWAS through gene editing techniques such as CRISPR/Cas9 (Laurie et al., 2010) can ensure that these genes are indeed associated with traits, thus improving the accuracy of studies. By meta-analysis of GWAS results from different studies, the reliability of the results can be improved. At the same time, repeated validation is also an important step to ensure that genetic variants of biological significance are found.

## **4 Future Development Direction and Prospect**

### **4.1 Development of high-throughput sequencing techniques and bioinformatics tools**

The development of high-throughput sequencing techniques and bioinformatics tools has had a profound impact on the application of genome-wide association study (GWAS) in crop disease resistance breeding. High-throughput sequencing technologies, including second-generation sequencing (such as Illumina) and third-generation sequencing (such as PacBio and Oxford Nanopore), allow researchers to access vast amounts of genetic information with unprecedented speed and precision. These techniques can not only provide high density single nucleotide polymorphism (SNP) markers, but also reveal structural and rare variants, providing rich genetic variation data for GWAS (Xiao et al., 2022). High-throughput sequencing technology can also be used in transcriptomics, epigenetics and genome resequencing studies, providing a new perspective for the molecular mechanism of crop disease resistance.



The development of bioinformatics tools supports the processing and analysis of high-throughput sequencing data. With the advent of the big data era, how to effectively manage, analyze and interpret massive biological information has become a challenge. Bioinformatics tools, including software and algorithms for sequence alignment, variation detection, gene annotation, and association analysis (Normand and Yanai, 2013), enable researchers to extract useful information from complex data. The application of machine learning and artificial intelligence technology also provides new methods for bioinformatics analysis, helping to reveal the complex genetic laws of crop disease resistance.

#### **4.2 The application of multi-omics data integration and systems biology methods**

The application of multi-omics data integration and systems biology methods is an important development direction of genome-wide association study (GWAS) in crop disease resistance breeding in the future. With the progress of biotechnology, researchers can obtain the genome, transcriptomics, proteomics, metabolomics and other multi-omics data of crops, which provides comprehensive information on the physiological and molecular level of crops, and helps to deeply understand the complex mechanism of crop disease resistance.

Multi-omics data integration refers to the comprehensive analysis of different levels of biological data to reveal a comprehensive picture of crop disease resistance. By integrating genomic and transcriptomic data, researchers can identify genes whose expression changes significantly during disease infection (Subramanian et al., 2020) and explore the role of these genes in disease resistance response. The integration of proteomic and metabolomic data helps to reveal changes in proteins and metabolites associated with disease resistance, providing deeper insight.

The systems biology approach refers to the use of mathematical and computational models to analyze and explain complex biological systems. In crop disease resistance studies, systems biology approaches can be used to construct network models of disease resistance responses and reveal interactions between different genes, proteins, and metabolites (Pazhamala et al., 2021). This network model helps to identify the key regulatory factors and signaling pathways of disease resistance and provide targeted strategies for breeding.

#### **4.3 The combination of precision breeding and gene editing**

The combination of precision breeding and gene editing technology is an important trend in contemporary crop improvement. Precision breeding, also known as molecular breeding, relies on molecular markers and genomic information to precisely select and aggregate genes associated with target traits through methods such as molecular marker-assisted selection (MAS). Gene editing technologies, such as the CRISPR/Cas9 system, can achieve precise modifications at specific sites in the crop genome to directly change the genetic characteristics of the crop (Nerkar et al., 2022), and the combination of these two technologies provides new strategies for crop disease resistance breeding.

After using GWAS and other methods to identify the key genes or genetic markers related to disease resistance, these favorable genes can be aggregated into a variety through precision breeding technology to improve the disease resistance of crops. For some resistance traits that are difficult to obtain through traditional breeding methods, gene editing technology can directly introduce or modify specific resistance genes into the crop genome, thereby rapidly breeding new varieties with strong resistance to disease (Scheben and Edwards, 2017).

The breeding strategy combined with precision breeding and gene editing technology can not only improve the efficiency and accuracy of breeding, but also expand the possibility of breeding, providing more choices and flexibility for crop disease resistance breeding. With the continuous development and improvement of these technologies, crop disease resistance breeding will be more efficient and accurate in the future, and is expected to make greater contributions to the sustainable development of agricultural production.

### **5 Summary**

The application of genome-wide association study (GWAS) in crop disease resistance breeding has made remarkable achievements. Through GWAS, researchers can quickly and accurately identify genetic markers and genes related to disease resistance in the whole genome, providing a powerful molecular tool for crop disease

resistance breeding. These results not only deepen our understanding of the genetic mechanism of crop resistance, but also provide a reliable basis for molecular marker-assisted breeding (MAS) and gene directed improvement.

However, the application of GWAS in crop disease resistance breeding also faces some challenges. Due to the influence of population structure and linkage imbalance, GWAS may produce false positive results, affecting the accuracy of the results. Generally, GWAS can only identify large effect genes related to traits, while some small effect genes may be ignored. The biological interpretation and functional verification of GWAS results is also a complex process, which requires in-depth analysis with transcriptomic and proteomic data.

With the continuous development of high-throughput sequencing technology and bioinformatics tools, the application of GWAS in crop disease resistance breeding will be more extensive and in-depth. Combined with multi-omics data and systems biology methods, GWAS will be able to reveal the complex genetic network of crop disease resistance in a more comprehensive way, and the combination of gene editing technology will also enable the disease resistance genes identified by GWAS to be quickly and accurately applied in breeding practice, improving the efficiency and accuracy of breeding.

As a powerful genetic analysis tool, GWAS is playing an increasingly important role in crop disease resistance breeding. Future studies will further optimize GWAS analysis methods, combine multi-omics data and gene editing technology, and provide more accurate and efficient solutions for crop disease resistance breeding to meet the challenges faced by global agricultural production.

## References

- Belzile F., and Torkamaneh D., 2022, Designing a genome-wide association study: Main steps and critical decisions, In: Torkamaneh D., and Belzile F. (eds.), *Genome-wide association studies, Methods in Molecular Biology*, Humana, New York, USA, pp.3-12.  
[https://doi.org/10.1007/978-1-0716-2237-7\\_1](https://doi.org/10.1007/978-1-0716-2237-7_1)
- Ciochetti N.P., Lugli-Moraes B., da Silva B.S., and Rovaris D.L., 2023, Genome-wide association studies: utility and limitations for research in physiology, *The Journal of Physiology*, 601(14): 2771-2799.  
<https://doi.org/10.1113/JP284241>
- Heeney C., 2021, Problems and promises: How to tell the story of a genome wide association study? *Stud. Hist. Philos. Sci.*, 89: 1-10.  
<https://doi.org/10.1016/j.shpsa.2021.06.003>
- Lamichhane S., and Thapa S., 2022, Advances from conventional to modern plant breeding methodologies, *Plant Breed. Biotech.*, 10: 1-14.  
<https://doi.org/10.9787/PBB.2022.10.1.1>
- Laurie C.C., Doheny K.F., Mirel D.B., Pugh E.W., Bierut L.J., Bhargale T., Boehm F., Caporaso N.E., Cornelis M.C., Edenberg H.J., Gabriel S.B., Harris E.L., Hu F.B., Jacobs K.B., Kraft P., Landi M.T., Lumley T., Manolio T.A., McHugh C., Painter I., Paschall J., Rice J.P., Rice K.M., Zheng X., Weir B.S., and GENEVA Investigators, 2010, Quality control and quality assurance in genotypic data for genome-wide association studies, *Genet. Epidemiol.*, 34(6): 591-602.  
<https://doi.org/10.1002/gepi.20516>
- Marees A.T., de Kluiver H., Stringer S., Vorspan F., Curis E., Marie-Claire C., and Derks E.M., 2018, A tutorial on conducting genome-wide association studies: Quality control and statistical analysis, *Int. J. Methods Psychiatr. Res.*, 27(2): e1608.  
<https://doi.org/10.1002/mpr.1608>
- Mores A., Borrelli G.M., Laidò G., Petruzzino G., Pecchioni N., Amoroso L.G.M., Desiderio F., Mazzucotelli E., Mastrangelo A.M., and Marone D., 2021, Genomic approaches to identify molecular bases of crop resistance to diseases and to develop future breeding strategies, *Int. J. Mol. Sci.*, 22(11): 5423.  
<https://doi.org/10.3390/ijms22115423>
- Nerkar G., Devarumath S., Purankar M., Kumar A., Valarmathi R., Devarumath R., and Appunu C., 2022, Advances in crop breeding through precision genome editing, *Front. Genet.*, 13: 880195.  
<https://doi.org/10.3389/fgene.2022.880195>
- Normand R., and Yanai I., 2013, An introduction to high-throughput sequencing experiments: Design and bioinformatics analysis, In: Shomron N. (ed.), *Deep sequencing data analysis, Methods in molecular biology*, Humana Press, Totowa, USA, pp.1-26.  
[https://doi.org/10.1007/978-1-62703-514-9\\_1](https://doi.org/10.1007/978-1-62703-514-9_1)
- Osorio-Guarín J.A., Berdugo-Cely J.A., Coronado-Silva R.A., Baez E., Jaimes Y., and Yockteng R., 2020, Genome-wide association study reveals novel candidate genes associated with productivity and disease resistance to *Moniliophthora* spp. in cacao (*Theobroma cacao* L.), *G3 Genes|Genomes|Genetics*, 10(5): 1713-1725.  
<https://doi.org/10.1534/g3.120.401153>
- Pazhamala L.T., Kudapa H., Weckwerth W., Millar A.H., and Varshney R.K., 2021, Systems biology for crop improvement, *The Plant Genome*, 14(2): e20098.  
<https://doi.org/10.1002/tpg2.20098>

- Scheben A., and Edwards D., 2017, Genome editors take on crops, *Science*, 355(6330): 1122-1123.  
<https://doi.org/10.1126/science.aal4680>
- Spencer C.C.A., Su Z., Donnelly P., and Marchini J., 2009, Designing genome-wide association studies: Sample size, power, imputation, and the choice of genotyping chip, *PLoS Genet.*, 5(5): e1000477.  
<https://doi.org/10.1371/journal.pgen.1000477>
- Subramanian I., Verma S., Kumar S., Jere A., and Anamika K., 2020, Multi-omics data integration, interpretation, and its application, *Bioinform. Biol. Insights*, 14: 1177932219899051.  
<https://doi.org/10.1177/1177932219899051>
- 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>
- 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>
- Xiao Q., Bai X., Zhang C., and He Y., 2022, Advanced high-throughput plant phenotyping techniques for genome-wide association studies: A review, *Journal of Advanced Research*, 35: 215-230.  
<https://doi.org/10.1016/j.jare.2021.05.002>
- Zhang J., Singh A., Mueller D.S., and Singh A.K., 2015, Genome-wide association and epistasis studies unravel the genetic architecture of sudden death syndrome resistance in soybean, *The Plant Journal*, 84(6): 1124-1136.  
<https://doi.org/10.1111/tpj.13069>