Volume 3, Number 3, 2026 Pages 1595-1613

Bioscience Insights

ISSN: 3080-3489 (Print)
ISSN: 3080-3497 (Online)

BioSciInsights.com

# GENOME-WIDE ASSOCIATION STUDIES ACROSS CROPS FOR PREDICTIVE BREEDING UNDER CHANGING CLIMATES

**Muhammad Mehraj Riasat[*1], Muhammad Haseeb[2], Sara Faridoon[3], Muhammad Waleed[4], Muhammad Muddasir[5], Uzair Ahmad[6], Ghanzanfar Hameed Awan[7], Shafaq Munir[8], Atika Liaqat[9], Shoaib Khan[10], Farhan Tahir[11]**

*[1,2,3,6,7,8,9,10,11]*Department of Agriculture, Hazara University, Mansehra, Pakistan*
[4]*Department of Agriculture, Lincoln University College, Petaling Jaya 47301, Malaysia*
[5]*Department of Plant Breeding and Genetics, Gomal University, Dera Ismail Khan, Pakistan*

[*1]mehrajhum30278@gmail.com

https://doi.org/10.5281/zenodo.24106611

**Article Information**

***Article History***

| | |
|---|---|
| Received: | June 19, 2026 |
| Revised: | August 25, 2026 |
| Accepted: | September 07, 2026 |
| Available Online: | October 31, 2026 |

**Keywords:**
*Crop improvement; Climate-resilient crops; Stress tolerance; Yield stability; Genomic selection; High-throughput phenotyping; Pangenome; G×E interaction*

***Copyright @Author***
**Corresponding Author: ***
**Muhammad Mehraj Riasat**

**Abstract**

*Genome-wide association studies (GWAS) have advanced the understanding of complex agronomic traits; however, their translation into practical crop improvement remains limited. This review provides a cross-crop synthesis of GWAS applications to identify conserved and crop-specific genetic architectures underlying stress resilience and yield stability. Moving beyond descriptive analysis, we propose an integrative framework that connects GWAS outputs with predictive breeding by incorporating phenomics, pangenomics, and environmental interactions. We highlight how conserved pathways enable transferable genetic insights across species, while crop-specific adaptations support targeted breeding strategies. Key challenges, including genotype-by-environment interactions, phenotyping limitations, and population structure, are critically evaluated in the context of predictive accuracy. By bridging gene discovery with breeding application, this work offers a conceptual roadmap for next-generation, data-driven crop improvement under changing environmental conditions.*

## INTRODUCTION

Modern agricultural systems are entering an era of unprecedented challenges shaped by climate instability, persistent food insecurity, and stagnating yield gains in major crops. The combined effects of increasing temperatures, irregular precipitation, soil degradation, and frequent extreme climatic events are steadily compromising the productivity and sustainability of crop production worldwide (Lesk et al., 2016; Wheeler & von Braun, 2013; UNICEF, 2021). Simultaneously, accelerating population growth and shifts in dietary preferences are expected to raise global food requirements, thereby placing additional pressure on agroecosystems (Springmann et al., 2018; Ray et al., 2013). Traditional breeding has historically contributed to crop improvement. Still, it is now limited in its ability to address complex, quantitative, and environmentally sensitive traits required for climate resilience, including drought

tolerance, heat-stress adaptation, and durable disease resistance (Tester & Langridge, 2010). These limitations underscore the need for next-generation crop breeding approaches that can rapidly enhance genetic gain while effectively addressing the complex dynamics of stress resilience in modern agricultural landscapes.

In this framework, genome-wide association studies (GWAS) have gained recognition as an essential tool for identifying the genetic determinants of complex agronomic traits. Unlike studies focused on single-gene analyses or biparental populations, GWAS exploits natural genetic diversity in large populations to associate genome-wide variation with phenotypic performance (Huang & Han, 2014; Tibbs Cortes et al., 2021). Over the past decade, advances in high-throughput genotyping, whole-genome sequencing, and phenotyping have markedly increased the resolution of GWAS, enabling the identification of genomic loci associated with yield stability, abiotic stress tolerance, and disease resistance across diverse environments (Huang, 2024; Yang et al., 2020; D'Agostino & Tripodi, 2017). Importantly, GWAS has shifted crop genetics from single-trait analyses toward a systems-level perspective on resilience, in which multiple loci, molecular pathways, and gene–environment interactions jointly shape adaptive responses (Korte & Farlow, 2013; Crossa et al., 2017). Consequently, GWAS has evolved from a gene-discovery tool into a core component of predictive, data-driven breeding programs.

GWAS-based studies, which are expanding rapidly, focus primarily on single crops or narrowly defined trait categories, limiting their capacity to capture broader genetic principles of resilience. Although crop-focused syntheses are informative, they often neglect conserved pathways and shared adaptive mechanisms that operate beyond taxonomic limits (Varshney et al., 2021; Wallace et al., 2018). As climate stressors increasingly affect agroecosystems rather than individual species, fragmented analytical approaches limit the effective translation of genomic knowledge into robust breeding solutions. Mounting evidence indicates that essential stress-response networks, such as osmotic regulation, hormonal signalling, and immune defence, are partially conserved across cereals, legumes, and industrial crops (Bailey-Serres et al., 2019; Daszkowska-Golec, 2022). Accordingly, an integrative, multi-crop synthesis of GWAS findings is urgently needed to move isolated discoveries toward a unified model of genetic resilience that can guide broad-scale advanced breeding innovations.

This review aims to provide an integrated and comprehensive overview of GWAS in advancing next-generation crop breeding. Specifically, we synthesise GWAS across major crop groups to characterise the genetic architectures underlying resilience traits, both conserved and crop-specific. Through comparative analyses of cereals, legumes, industrial, and horticultural crops, we uncover common biological pathways and divergent adaptive strategies relevant to molecular breeding and genomic selection. Furthermore, we examine the translation of GWAS findings into practical breeding applications while critically addressing existing limitations and emerging opportunities. A cross-crop, forward-looking framework is intended to facilitate the development of integrated, resilient breeding pipelines suited to the challenges of an evolving global food system.

## THE GENOMIC REVOLUTION IN CROP BREEDING

Crop breeding has transitioned from phenotype-based selection to molecularly informed and genome-enabled strategies. Conventional breeding methods play a central role in agricultural improvement, depending heavily on observable traits and prolonged field evaluations, which reduces their efficiency for complex traits sensitive to environmental variations, such as yield stability and drought tolerance. The introduction of molecular markers represented an important advancement by facilitating indirect selection of genomic regions via quantitative trait locus (QTL) mapping and marker-assisted selection (MAS) (Collard & Mackill, 2008). Nevertheless, these approaches were constrained by low mapping resolution, dependence on biparental populations, and an inability to represent natural genetic diversity. Breeding objectives become more complex as climate variability increases. Thus, it became evident that

incremental enhancements to conventional breeding were insufficient to address emerging agricultural challenges (Bernardo, 2016; Xu et al., 2017; Moose & Mumm, 2008).

Genome-wide association studies (GWAS) represented a decisive break from traditional genetic mapping frameworks by exploiting historical recombination in diverse germplasm resources to provide high-resolution mapping of complex traits. Unlike QTL mapping, which depends on structured populations, GWAS harnesses population-scale genetic diversity, enabling the detection of multiple loci underlying quantitative traits across wide genetic backgrounds (Yu et al., 2006; Brachi et al., 2011; Nordborg & Weigel, 2008). Seminal studies in major crops, including rice and maize, demonstrated the ability of GWAS to identify loci associated with flowering time, yield components, and stress responses, firmly establishing its role as a foundational genomic discovery tool (Yano et al., 2016; Li et al., 2012; Buckler et al., 2009). These foundational studies enabled rapid expansion of GWAS across diverse crop species, supported by declining sequencing costs and the broad availability of high-density single-nucleotide polymorphism (SNP) arrays (Poland et al., 2012; Elshire et al., 2011; Varshney et al., 2005). Notably, GWAS transcended the limitations of marker-assisted selection by uncovering complex polygenic architectures and allelic series, which cannot be resolved within single-locus models.

In recent years, the role of GWAS in crop breeding has evolved from isolated gene discovery to trait prediction and system-level insights. Advances in statistical modelling, multi-locus association techniques, and the integration of phenotypic and environmental data have enabled GWAS to capture the biological complexity of adaptive traits. Instead of highlighting individual major genes, modern GWAS emphasises identifying interconnected loci and biological pathways as key contributors to resilience across fluctuating environmental conditions (Resende Jr et al., 2012; Gianola, 2013; Jarquín et al., 2014). This shift reflects a broader evolution in breeding philosophy, from focusing on single-trait improvement to optimising performance across diverse environments and multiple stressors. Consequently, GWAS bridges genomic discovery and predictive breeding by supporting genomic selection, candidate gene prioritisation, and resilience-oriented crop improvement (Mackay et al., 2009; Boyle et al., 2017; Xu, 2016). This genomic transformation has redefined crop breeding as a data-driven, integrative discipline, positioning GWAS as a core component of next-generation efforts to build sustainable, climate-resilient agricultural systems.

## PRINCIPLES AND FRAMEWORK OF GWAS

Genome-wide association studies are more accurately described as a dynamic analytical framework rather than a static mapping approach. Modern GWAS pipelines have been designed to address the complexity of the genome, structural variation, and population-specific diversity that were previously inaccessible through reference–genome–based methods. Advances in population-scale whole-genome resequencing have enabled direct interrogation of rare variants, copy-number variation, and haplotype diversity. Pangenome resources now support GWAS analyses across multiple genomic representations, extending beyond a single linear reference (Bayer et al., 2020; Danilevicz et al., 2020; Jayakodi et al., 2020). This transition has redefined the GWAS input layer into a genome-inclusive framework capable of detecting adaptive variation spanning core and dispensable genomic regions. Consequently, GWAS is now better aligned with the biological complexity of crop genomes shaped by domestication, selection, and local adaptation.

At the analytical level, GWAS has moved away from isolated marker–trait testing toward models that explicitly address trait complexity and interaction structures. Contemporary multi-locus GWAS approaches jointly estimate the contributions of multiple genomic regions, enhancing the detection of small-effect loci while preserving statistical power across genetically diverse populations (Liu et al., 2016; Asan et al., 2016). Concurrently, machine learning–based association models are increasingly employed to capture non-linear genotype–phenotype relationships and higher-order interactions that are not well addressed by traditional parametric methods (Libbrecht & Noble, 2015; Montesinos-López et al., 2018).

Together, these methodological advances reposition GWAS as a predictive modelling tool rather than a purely inferential framework, particularly when supported by cross-validation and model transferability evaluations. The resulting framework aligns well with trait architectures driven by distributed genetic control, commonly observed in resilience-related traits under changing environmental conditions.

A defining characteristic of modern GWAS is its integration with phenomics, enviromics, and multi-omics data, which enables biologically interpretable associations and actionable outcomes. High-resolution phenotyping platforms, including image-based field systems and temporal trait profiling, now generate data at scales compatible with genome-wide analyses, thereby reducing noise and improving reproducibility (Shivaiah et al., 2020; Araus et al., 2018). When environmental variables are explicitly incorporated into GWAS models, allowing them to partition genetic effects across environmental gradients, this enhances their relevance for real-world breeding contexts (Houmanat et al., 2021). In addition, integrating GWAS signals with transcriptomic, metabolomic, and chromatin-level data enables more precise prioritisation of candidate genes and regulatory elements within associated regions (Kim et al., 2008; Chen et al., 2020). As part of a comprehensive genomic–breeding workflow, GWAS functions as a central decision point that links genomic variation to candidate gene identification, functional validation, and breeding implementation (Figure 1).

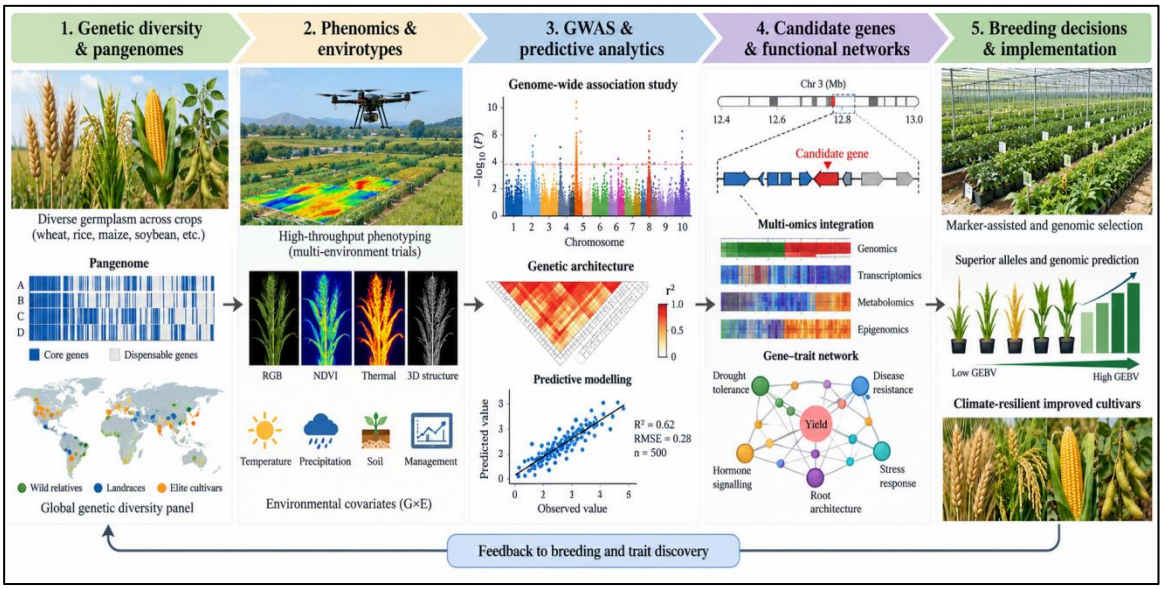



**Figure 1. The modern integrated GWAS-to-breeding pipeline for resilience-focused crop improvement.**

## A MULTI-CROP PANORAMA OF GWAS APPLICATIONS

Genome-wide association studies have evolved from single-crop, trait-centred analyses into a comparative framework for understanding how distinct plant clades leverage both shared and unique genetic mechanisms to achieve resilience, yield, and quality. This section reviews GWAS results across crop groups, focusing on (a) the types of genetic signals recovered in each group, (b) their implications for practical breeding pipeline implementation, and (c) opportunities for translating insights across crops. Where informative, examples illustrate how GWAS outputs have been refined through haplotype resolution and pangenome-aware analyses, and how they have been leveraged in breeding or functional follow-up (Tables 1–4).

## CEREAL CROPS (WHEAT, RICE, MAIZE, AND BARLEY)

Among crop groups, cereal crops represent the most informative and analytically powerful systems for genome-wide association studies, reflecting their economic importance, extensively sampled germplasm collections, and long-standing networks for multi-environment phenotyping (Morrell et al., 2012; Xiao et al., 2022; Araus et al., 2018). These conditions have enabled GWAS applications in rice, maize, wheat, and barley to transition from early-stage locus discovery to resolving complex, environment-sensitive genetic architectures directly relevant to breeding (Yano et al., 2016; Wallace et al., 2018; Cooper et al., 2014). High-resolution GWAS in rice and maize consistently reveal regulatory variants, haplotypes, and rare alleles affecting phenology, grain size, and abiotic stress responses, while whole-genome resequencing supports the identification of loci expressed only under specific stress scenarios such as drought during reproductive development (Yano et al., 2016; Cabrera-Bosquet et al., 2012; Wei et al., 2021; Gage et al., 2018).

**Table 1. Representative genomic and association studies in cereal crops.**

| Crop | Population and design | Genomic data | Main outputs | Reference |
|---|---|---|---|---|
| **Rice** | 176 Japanese cultivars, including 86 sake-brewing cultivars | Whole-genome resequencing; 493,881 polymorphisms | Gene-based GWAS identified four new genes affecting agronomic traits; known flowering genes Hd1, Hd2 and Hd6 also located | Yano et al. (2016) |
| **Maize** | NAM population: 25 founder lines crossed to B73, about 5,000 RILs; about 1 million plants in 8 environments | Nested association mapping design (25 families) | No large-effect QTL for flowering time; 29–56 small-effect additive QTL | Buckler et al. (2009) |
| **Maize** | NAM population, about 5,000 RILs; five traits (leaf size and angle, flowering time) | 1.01 million SNPs from RNA-seq and HapMap | Trait-associated SNPs explained 44–59% of phenotypic variation; genic and 5-kb upstream regions (13% of the genome) held 71% of associations | Li et al. (2012) |
| **Wheat** | Bread wheat lines under drought and heat stress | Genomic prediction with high-throughput phenotyping | Cross-validated accuracy 0.50 (drought) and 0.51 (heat); 0.18 and 0.23 when another nursery or year was predicted | Juliana et al. (2019) |
| **Barley** | 20 genotypes assembled at chromosome scale; 300 gene-bank accessions shotgun-sequenced | Pan-genome; presence–absence variants | 63% of genic regions shared among the 20 genotypes, 37% variable; large inversions, one linked to mutation breeding | Jayakodi et al. (2020) |

GWAS, genome-wide association study; NAM, nested association mapping; QTL, quantitative trait locus; RIL, recombinant inbred line. Values are as reported in each study.

In wheat and barley, progress in genome assembly quality, genotype imputation, and multi-locus association modelling has substantially mitigated challenges associated with large and polyploid genomes, enabling robust identification of loci underlying disease resistance, root system traits, and tolerance to heat and drought stress (Juliana et al., 2019; Herrmann et al., 2021; Pasam et al., 2012). Collectively,

findings from cereal GWAS demonstrate that yield stability is primarily shaped by polygenic networks composed of loci with small to moderate effects with strong genotype-by-environment interactions. This architecture reduces the effectiveness of single-marker approaches and underscores the need for genomic prediction frameworks that integrate biologically relevant GWAS signals and targeted allele mining from diverse germplasm resources (Crossa et al., 2017; Cooper et al., 2019; Wei et al., 2021; Cooper et al., 2014).

Large-scale GWAS coupled with whole-genome sequencing in rice has identified loci associated with flowering time regulation, ionomic traits, and drought response, supporting haplotype-based allele mining for breeding improvement (Araus et al., 2018; Yano et al., 2016). Moreover, large association panels and nested association mapping in maize have shown that GWAS signals influence inflorescence structure and yield components, with image-derived phenotypes enhancing trait resolution (Gage et al., 2018). The combination of high-throughput phenotyping, genomic prediction, and GWAS has begun to yield drought-adapted selections despite the complexity of polyploid genomes (Juliana et al., 2019; Herrmann et al., 2021). Representative cereal studies are summarised in Table 1.

## LEGUME CROPS (SOYBEAN, CHICKPEA, LENTIL, AND RELATED SPECIES)

Legume crops provide a unique biological landscape for genome-wide association studies due to integration of agronomic performance with essential ecosystem services, most notably symbiotic nitrogen fixation (Wang et al., 2022; Graham and Vance, 2003). Accordingly, legume GWAS expands the scope of legumes beyond yield and stress tolerance to include seed composition, symbiotic efficiency, and nitrogen-use efficiency, highlighting their importance for both nutritional security and sustainable production systems (Varshney et al., 2019; Bao et al., 2021). Across legume species, association mapping consistently identifies complex and pleiotropic genetic architectures, where loci affecting stress responses, developmental timing, and seed traits challenge direct selection, while affording opportunities for coordinated, systems-oriented breeding approaches (Liu et al., 2020; Tibbs Cortes et al., 2021; Zhang et al., 2021).

In soybean, repeated GWAS analyses identified loci with antagonistic and pleiotropic effects on seed protein and oil concentration, indicating the regulatory complexity of key biosynthetic pathways governing seed composition (Shingote et al., 2022; Fang et al., 2017b; Zhou et al., 2015). The combination of GWAS with soybean pangenome resources has revealed substantial structural variation and gene presence–absence polymorphisms associated with seed composition and stress-responsive traits, allowing haplotype-based allele mining with better resolution and breeding effectiveness (Zhou et al., 2025; Li et al., 2023a). The concentration of these loci within regulatory and structurally variable regions points to the limitations of single-reference genomes in capturing functional variation (Liu et al., 2020). In chickpea and lentil, GWAS has successfully dissected the genetic basis of heat and drought tolerance, flowering phenology, and reproductive resilience traits essential for adaptation in water-limited and terminal-stress environments (Varshney et al., 2019; Li et al., 2018). Studies using diverse global panels have identified loci conferring reproductive-stage heat tolerance and adaptive phenology, directly supporting breeding approaches to reduce terminal heat stress (Istanbuli et al., 2024; Panigrahi et al., 2024). Importantly, stress-adaptive alleles with phenological traits reinforce the need for correlated trait responses in legume breeding (Thudi et al., 2014).

A defining advance in legume GWAS has been the incorporation of symbiosis-related traits into association analyses. GWAS has identified loci controlling nodulation efficiency, rhizobial compatibility, and nitrogen fixation performance, directly linking genetic variation to diminished reliance on fertilisers and climate-smart production systems (Ratnaparkhe et al., 2022; Goyal et al., 2021; Varshney et al., 2019). Unlike cereals, where resilience traits are largely decoupled from nutrient acquisition, legumes show a unique genetic context in which stress responses, root architecture, and symbiotic regulation are interdependent (Sulieman and Tran, 2013).

From a breeding perspective, legume GWAS is most impactful when combined with detailed physiological characterisation, including nodulation behaviour, root traits, and multi-environment testing, which captures variation in soil nitrogen and stress conditions (Prior et al., 2017; Varshney et al., 2021). The emergence of the pangenome, particularly in soybean, has enhanced the discovery of structural variation and dispensable genomic regions influencing seed composition and stress tolerance, enabling haplotype-level allele mining with direct relevance to improvement programs (Zhou et al., 2025; Bayer et al., 2020). Collectively, these advances establish GWAS in legumes as a principal interface between genomic discovery and sustainability-oriented crop improvement. Representative legume studies are summarised in Table 2.

**Table 2. Representative genomic and association studies in legume crops.**

| Crop | Population and design | Genomic data | Main outputs | Reference |
|---|---|---|---|---|
| **Soybean** | 302 wild, landrace and improved accessions | Resequencing at >11× depth | 230 selective sweeps and 162 selected CNVs; 10 selected regions linked to 9 domestication or improvement traits; 13 new loci (oil content, plant height, pubescence) | Zhou et al. (2015) |
| **Soybean** | 809 accessions; 84 traits; 3 locations over 2 years | WGS at 8.3×; 10.4 million SNPs, 1.0 million indels | 245 significant loci (95 interacting); 23 pleiotropic loci including Dt1, E1 and E2 | Fang et al. (2017b) |
| **Soybean** | 26 de novo assemblies chosen from 2,898 deeply sequenced accessions | Graph-based pan-genome | Over 55,000 structural variants genotyped; over 1,000 core-gene SVs linked to altered expression; iron-deficiency chlorosis QTL clarified | Liu et al. (2020) |
| **Chickpea** | 429 lines from 45 countries | Resequencing; 4.97 million SNPs | 122 selected regions (204 genes); GWAS identified 262 markers for 13 traits | Varshney et al. (2019) |
| **Chickpea** | Reference set of 300 accessions (211 mini-core) | 1,872 markers (DArT, SNP, gene-based SNP, SSR) | 312 significant marker–trait associations (70 for 100-seed weight); 18 SNPs in ERECTA, ASR, DREB, CAP2 and AMDH | Thudi et al. (2014) |

CNV, copy number variant; DArT, diversity arrays technology; GWAS, genome-wide association study; QTL, quantitative trait locus; SSR, simple sequence repeat; SV, structural variant; WGS, whole-genome sequencing.

## INDUSTRIAL & CASH CROPS (COTTON, SUGARCANE, SUNFLOWER, OILSEED RAPE)

Industrial and cash crops are strong targets for genetic improvement, particularly for traits such as fibre characteristics, sugar content, oil quality, and disease resistance. However, these crops are often genomically complicated due to large genome size, extensive structural variation, and frequent polyploidy, which directly affects GWAS design, mapping resolution, and analytical approaches. As a result, GWAS

in cash crops has developed under constraints distinct from those in cereals and legumes, emphasising methodologically flexible and integrative analytical approaches (Ma et al., 2018; Sone et al., 2019; Chalhoub et al., 2014).

Across industrial crops, GWAS consistently reveals polygenic architectures of quality traits strongly influenced by environmental modulation. In cotton, association studies have identified numerous loci associated with fibre length and strength, many of which display regulatory and pleiotropic functions with yield-associated regions, revealing strong genetic coupling between productivity and quality traits (Fang et al., 2017a; Ma et al., 2018). Multi-environment GWAS have further identified fibre-quality loci with stable effects throughout environments, highlighting candidate genes being validated in the breeding pipeline (Sun et al., 2017; Ma et al., 2018). These data demonstrate that fibre improvement cannot be based on single-marker introgression and that multi-trait, multi-environment selection approaches are required (Crossa et al., 2017; Cooper et al., 2019).

**Table 3. Representative genomic and association studies in industrial and cash crops (cotton).**

| Crop | Population and design | Genomic data | Main outputs | Reference |
|---|---|---|---|---|
| **Cotton** | 318 upland accessions | Resequencing | 119 SNP loci: 71 for fibre yield, 45 for fibre quality, 3 for Verticillium wilt resistance | Fang et al. (2017a) |
| **Cotton** | 419 upland accessions (core collection); 13 fibre traits; 12 environments | Resequencing at 6.55×; about 3.66 million SNPs | 7,383 SNPs associated, within or near 4,820 genes; more loci for fibre quality than yield; more fibre genes in the D subgenome | Ma et al. (2018) |
| **Cotton** | 719 upland accessions; 8 environments | CottonSNP63K array; 10,511 polymorphic SNPs | 46 significant SNPs for five fibre quality traits | Sun et al. (2017) |

SNP, single-nucleotide polymorphism. Values are as reported in each study.

In sugarcane, the combination of extreme polyploidy and a highly complex genome structure has consistently limited GWAS resolution. Nonetheless, studies that combine targeted marker systems with transcriptomic and metabolomic data have begun to link sugar content, biomass yield, and stress tolerance with specific genomic regions, demonstrating that informative GWAS can be extracted even from highly complex genomes (Chalhoub et al., 2014; Varshney et al., 2021). In sunflower and oilseed rape (canola), the expansion of pangenome resources and extensive germplasm sampling enables the identification of loci controlling oil composition, flowering time, and disease resistance. In sunflower, pangenome-based analyses uncovered introgressed genomic regions enriched for resistance genes, acting as a valuable resource for GWAS-based discovery of transferable alleles (Sone et al., 2019). At the same time, in oilseed rape, association mapping has identified oil-quality and resistance loci suitable for marker-assisted deployment (Chalhoub et al., 2014; Bayer et al., 2020).

From a breeding standpoint, GWAS achieves maximum value in industrial crops when association signals are integrated into genomic prediction models enriched for trait-specific markers and are supported by metabolic and biochemical phenotyping for quality attributes. Breeding success in these systems depends on combining GWAS-derived loci into multi-trait selection strategies rather than pursuing single-trait improvement (Crossa et al., 2017; Cooper et al., 2019). Collectively, these outcomes illustrate how GWAS, when customised to genomic complexity and market-oriented priorities, underpins quality- and resilience-focused industrial crop improvement. Representative studies in cotton are summarised in Table 3.

## HORTICULTURAL & MEDICINAL CROPS

GWAS in horticultural and medicinal crops necessitate a high level of methodological flexibility due to limited genetic diversity, frequent clonal propagation, strong population structure, and complex secondary metabolism. Although these challenges exist, GWAS has been effective in these systems when analytical systems are adapted to biological context, such as metabolite-based association mapping, temporal phenotyping, and multi-omics integration (Fernie and Tohge, 2017; Tibbs Cortes et al., 2021). Across fruit and vegetable crops, quality traits such as sugar levels, organic acids, and volatile compounds are governed by complex polygenic and developmentally regulated genetic architectures. The combined use of metabolite-based GWAS with transcriptomic data has been particularly successful in explaining the genetic basis of traits related to flavour and dietary quality in tomato, grape, and apple, and reflects regulatory variation within metabolic pathways (Migicovsky et al., 2021; Fernie and Tohge, 2017). In tomato, mGWAS coupled with gene expression identifies loci controlling sugar and volatile profiles for flavour-oriented breeding (Tieman et al., 2017). Similarly, in grapevine GWAS of berry composition and ripening traits leveraging long-term phenomic datasets, loci associated with both quality and stress resilience had been identified, illustrating the importance of multi-year phenotyping in perennial crops (Guo et al., 2019).

In medicinal plants, GWAS primarily targets variation in biosynthetic pathways governing bioactive compound production. GWAS of secondary metabolite traits have identified candidate genes regulating key enzymatic and regulatory steps in specialised metabolism, supporting metabolic breeding and quality standardisation in pharmaceutical and nutraceutical systems (Fernie and Bauwe, 2020). Across horticultural and medicinal crops, constraints including small population sizes, clonal propagation, and environmental heterogeneity require adapted analytical approaches, such as family-based GWAS, multi-site trials, and intensive multi-omics integration (Guo et al., 2019; Migicovsky et al., 2021). Under these conditions, GWAS has identified biologically interpretable loci for flavour, disease resistance, and phytochemical accumulation with direct breeding relevance. Representative studies are summarised in Table 4.

**Table 4. Representative genomic and association studies in horticultural crops.**

| Crop | Population and design | Genomic data | Main outputs | Reference |
|---|---|---|---|---|
| **Tomato** | 398 modern, heirloom and wild accessions; consumer taste panels | WGS and targeted metabolomics (sugars, acids, volatiles) | 251 associated variants for 4 non-volatile and 15 volatile flavour chemicals | Tieman et al. (2017) |
| **Apple** | Over 1,000 USDA germplasm accessions | Over 30,000 SNPs | Population genomics of domestication and improvement; signatures of selection | Migicovsky et al. (2021) |

SNP, single-nucleotide polymorphism; WGS, whole-genome sequencing. Values are as reported in each study.

## CROSS-CROP SYNERGIES FOR GENETIC RESILIENCE

Comparative analyses of genome-wide association studies across multiple crops indicate that resilience-related traits are governed by partially conserved genetic architectures, despite major differences in genome structure and domestication histories. In cereals, legumes, and industrial crops, GWAS commonly identifies overlapping genomic regions linked to drought tolerance, thermal stress response, flowering time modulation, and disease resistance, suggesting that environmental adaptation is mediated by a restricted set of fundamental biological processes (Uffelmann et al., 2021; Loreto & D'Auria, 2022). These shared signals more often reflect regulatory or pathway convergence than identical genes, enabling

discoveries in one crop to inform breeding priorities in others. This cross-crop view serves as the basis for translational genomics approaches that bridge species boundaries while maintaining strong agricultural relevance. Mechanistically, GWAS convergence across crops is most evident within stress response networks that integrate hormonal signalling, redox homeostasis, and growth modulation. Association studies across multiple species consistently identify variation in pathways related to abscisic acid–dependent signalling, cellular defence mechanisms, and stress-responsive transcriptional control (Gupta et al., 2020; Zhang et al., 2018b). Importantly, these findings increasingly indicate that stress tolerance differences are shaped primarily by allelic diversity within regulatory nodes that adjust pathway activity across environments. From a breeding standpoint, this reinforces a transition toward optimising allele combinations within conserved molecular networks to achieve resilience across variable environments.

Comparative GWAS integration across crops highlights a graded spectrum of genetic responses from broadly conserved mechanisms to highly crop-specific adaptations. Semi-conserved responses commonly involve pathways that are modified by lineage-specific regulatory divergence or gene family expansion, as frequently observed in disease resistance and secondary metabolism traits (Xu & Zhu, 2020; Li et al., 2023b). By contrast, crop-specific associations often arise from unique ecological niches, domestication trajectories, or genome structures. Polyploid crops such as wheat and cotton exhibit resilience loci influenced by subgenome interactions and dosage effects, whereas legumes integrate stress adaptation with nitrogen fixation and root symbiosis processes that are absent in cereals (Jiang et al., 2021; Juliana et al., 2019). Accounting for this structured diversity is necessary to prevent overgeneralisation while still leveraging transferable genetic principles. Building on these patterns, Figure 2 summarises the cross-crop framework, linking diverse crops and genomic resources, comparative genomic analysis, and shared and crop-specific genetic principles to applications in resilience breeding. Explicitly distinguishing transferability across conserved, partially conserved, and crop-specific layers enables practical prioritisation of GWAS outputs by breeding relevance, supporting a shift toward predictive comparative genomics (Table 5).

**Table 5. Conserved, partially conserved and crop-specific layers of resilience genetics across crops.**

| Layer | Biological basis | Examples in this review | Breeding use |
|---|---|---|---|
| **Conserved** | Hormonal (ABA-dependent) signalling, redox homeostasis, growth modulation and stress-responsive transcriptional control | Overlapping genomic regions for drought tolerance, thermal stress response, flowering time and disease resistance in cereals, legumes and industrial crops | Highly transferable; optimise allele combinations within conserved networks |
| **Partially conserved** | Shared pathways modified by lineage-specific regulatory divergence or gene-family expansion | Disease resistance and secondary metabolism traits | Partly transferable; prioritise by breeding relevance |
| **Crop-specific** | Unique ecological niches, domestication trajectories or genome structures | Subgenome interactions and dosage effects in wheat and cotton; nitrogen fixation and root symbiosis in legumes | Not assumed to transfer; targeted, crop-specific breeding |

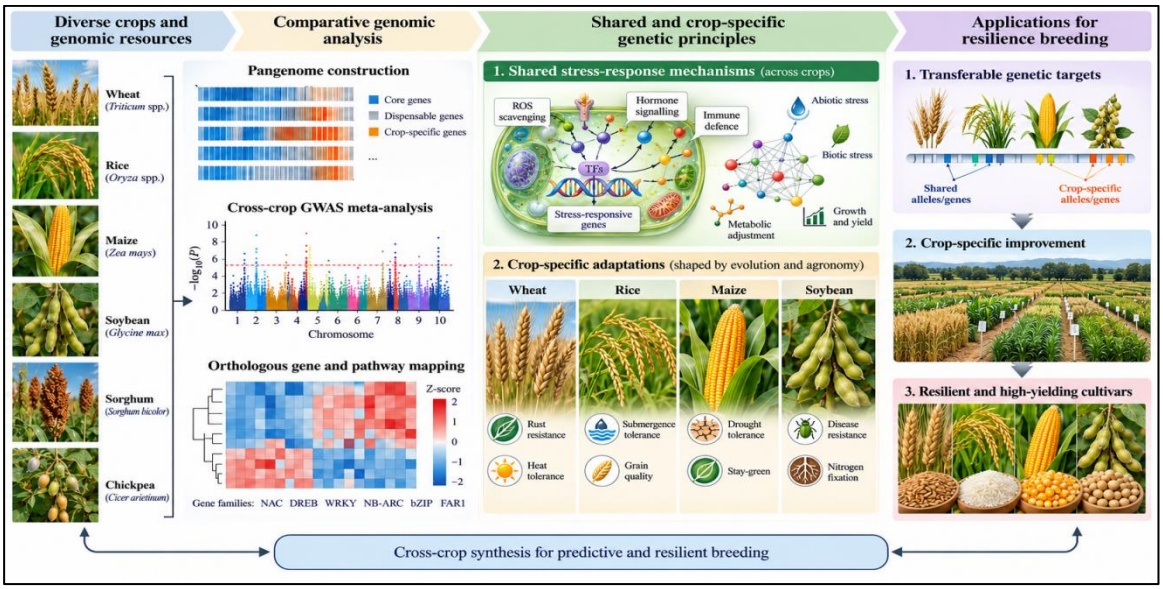



**Figure 2. Cross-crop comparative genomic principles for stress-resilience breeding.**

## TRANSLATING GWAS DISCOVERIES INTO BREEDING INNOVATIONS

The utility of GWAS is most evident when association results are translated into actionable breeding decisions that enhance selection efficiency under field conditions. Instead of functioning as a conventional selection approach, GWAS increasingly serves to refine breeding pipelines by identifying genomic regions relevant to target traits, training population design, and selection intensity. In practical breeding programs, loci exhibiting stable effects across environments are preferentially incorporated into trait-focused marker panels, while genome-wide signals inform prediction models for complex traits with polygenic architectures (Cooper et al., 2019; Loreto & D'Auria, 2022). This strategic use of GWAS output reduces uncertainty in early-generation selection and supports aligning genomic tools with breeding objectives under variable stress conditions.

A major advancement in GWAS translation has been the integration of association mapping with functional validation workflows, accelerating causal inference. Combining high-resolution GWAS with haplotype information and expression evidence now enables more reliable prioritisation of candidate genes within associated loci (Danilevicz et al., 2020). Increasingly, these candidates are subjected to genome-editing approaches, enabling allele function to be tested in elite breeding backgrounds rather than in model systems. Notably, recent studies position GWAS-guided editing as a validation step in breeding pipelines, in which edited alleles are rapidly evaluated for agronomic performance and stability (Zhang et al., 2018a; Gao, 2021). This coordinated approach has shifted functional genomics toward outcome-oriented validation, enhancing the application of GWAS in breeding programs.

Evidence of GWAS-driven innovation is emerging across crop species through measurable breeding outcomes. In rice, extensive diversity panels have been used to identify haplotypes associated with yield and stress tolerance, which were subsequently incorporated into elite breeding lines through targeted selection, thereby enhancing performance under drought and heat stress (Wei et al., 2021; Yano et al., 2016). In maize, haplotype selection has been used to optimise flowering synchrony and yield stability across diverse environments, informing hybrid development decisions (Gage et al., 2018). In legumes, association mapping has facilitated the deployment of stress-resilient alleles related to phenology, seed quality, and disease resistance, particularly in chickpea and soybean programs targeting marginal environments (Bao et al., 2021; Liu et al., 2020). Collectively, these cases indicate that GWAS is most effective when it informs breeding decisions rather than when it is used solely for gene discovery.

Embedding GWAS outputs directly into the development of resilient cultivars is therefore essential for consistent performance under real-world production constraints.

## CHALLENGES, LIMITATIONS, AND GAPS

Despite major advances in genome-wide association studies, several enduring limitations continue to reduce their impact on crop improvement (Table 6). Population structure and cryptic relatedness remain particularly problematic in breeding populations shaped by intensive selection and narrow genetic bases. Even with advanced mixed-model corrections, confounding effects can limit the detection of rare alleles and environmentally responsive loci that play a disproportionate role in complex agronomic traits (Yu et al., 2006; Morrell et al., 2012). In parallel, GWAS accounts for only a fraction of phenotypic variance in traits such as yield stability and stress tolerance. Increasing evidence suggests that this missing heritability arises not only from epistasis and gene–environment effects, but also from limitations in current association models that fail to capture and account for dynamic trait expression across environments (Uffelmann et al., 2021). Consequently, even statistically robust signals may still provide an incomplete representation of the genetic architecture required for effective breeding decisions.

**Table 6. Main limitations of GWAS for resilience breeding and the responses proposed.**

| Limitation | Effect on GWAS | Response proposed |
|---|---|---|
| **Population structure and cryptic relatedness** | Confounding that limits detection of rare alleles and environmentally responsive loci, even with mixed-model corrections | Broader germplasm, including landraces and wild relatives |
| **Missing heritability** | Only a fraction of variance in yield stability and stress tolerance explained; epistasis, gene–environment effects and dynamic trait expression | Association models that capture dynamic trait expression across environments |
| **Phenotyping quality and access** | Poor resolution for traits with strong temporal dynamics or non-linear responses; uneven adoption owing to cost, infrastructure and trained personnel | Long-term phenotyping networks and wider access to high-throughput platforms |
| **Environmental heterogeneity** | Genotype performance varies across locations and seasons, complicating interpretation and validation | Enviromics and multi-environment modelling with long-term, harmonised datasets |
| **Under-represented crops and germplasm** | Orphan cereals, minor legumes and indigenous horticultural species poorly covered; landraces and wild relatives under-sampled | Inclusive germplasm development |
| **Data sharing** | Inconsistent data standards, inadequate metadata and restricted access limit comparative analyses | Open-data systems and harmonised metadata standards |

Technical and logistical constraints continue to limit the translational impact of GWAS in crop improvement. Phenotyping quality represents an additional barrier, particularly for traits with strong temporal dynamics or that respond nonlinearly to environmental variation. Although high-throughput phenotyping platforms and remote sensing technologies have enhanced data resolution, their adoption remains uneven across regions and crop systems due to costs, infrastructure requirements, and shortages of trained personnel (Araus et al., 2018; Yang et al., 2020). Environmental heterogeneity introduces additional challenges, as genotype performance varies across locations and seasons, complicating interpretation and validation. Although enviromics and multi-environment modelling offer promising avenues, their effective implementation depends on long-term, harmonised datasets that remain scarce

in many breeding programs. Consequently, technological capabilities have exceeded the institutional capacity needed for effective large-scale application.

In addition to methodological and technical limitations, enduring structural knowledge gaps continue to limit the scope and effectiveness of GWAS. Numerous crops central to regional food security, including orphan cereals, minor legumes, and indigenous horticultural species, are still poorly represented in association studies owing to limited genomic resources and small, fragmented diversity panels (Bohra et al., 2022). Even within major crops, GWAS panels frequently underrepresent landraces and wild relatives, constraining access to adaptive alleles shaped by local environments. Data-sharing limitations across studies restrict comparative analyses, as inconsistent data standards, inadequate metadata, and restricted access impede interoperability and reuse (Leonelli et al., 2017).

Importantly, these gaps arise less from analytical shortcomings than from systemic constraints related to funding priorities, governance structures, and disparities in research capacity. Overcoming these barriers will require coordinated investment in inclusive germplasm development, open-data systems, and long-term phenotyping networks. Without such structural interventions, the promise of GWAS for resilience-focused breeding will remain only partially fulfilled.

## THE NEXT DECADE OF GWAS IN CROP BREEDING

Over the next decade, genome-wide association studies will move beyond their traditional analytical role to become a central component of predictive, decision-focused crop breeding systems. A key evidence-based pathway underlying this evolution is the integration of pangenomic frameworks into GWAS. By explicitly accounting for structural variation, gene presence–absence diversity, and lineage-specific genomic regions, pangenome-enabled GWAS overcomes the limitations of single-reference genome analyses and fully captures functional diversity within crop populations. Early applications in major crops already show that many adaptive and resilience-associated variations are located outside core reference genomes, suggesting that future GWAS pipelines will increasingly rely on multi-reference architectures. As pangenome resources expand and standardisation improves, their integration with GWAS is likely to redefine both the resolution and interpretability of association mapping.

A second, equally transformative development will be the routine incorporation of high-resolution phenotyping and environmental characterisation into GWAS analyses. Progress in field-based phenomics, temporal trait tracking, and sensor-enabled data collection now enable the quantification of dynamic traits at scales compatible with population genomic studies. When combined with explicit environmental descriptors, GWAS can extend beyond static associations to model genetic effects across environmental gradients. In the coming decade, phenotype- and environment-aware GWAS approaches are expected to become essential for dissecting yield stability and stress resilience under realistic agricultural conditions. This shift is crucial for aligning GWAS outputs with breeding priorities informed by climate variability, rather than controlled experimental conditions.

In addition to these empirical trends, a transformative opportunity is emerging from the convergence of GWAS with genomic prediction and genome-editing approaches. GWAS is increasingly integrated into predictive models to identify biologically relevant genomic regions, improve model structure, and inform feature selection. Looking forward, association mapping, genomic selection, and genome editing are expected to function as interdependent components of iterative breeding pipelines rather than as standalone tools. From a forward-looking perspective, the most effective breeding systems will operate as closed-loop frameworks in which GWAS identifies candidate regions, genomic prediction optimises allele combinations across environments, and genome editing enables rapid validation or refinement of key variants in elite germplasm. While such fully integrated systems are not yet widely implemented, the necessary conceptual and technological foundations are now well established.

The inclusivity and integration of global research efforts will determine the ultimate impact of GWAS-based crop improvement. Despite substantial technological progress, GWAS datasets remain disproportionately centred on a limited number of intensively bred crops and elite germplasm resources. Expanding association studies to include landraces, wild relatives, and underutilised species will be essential for uncovering adaptive variation relevant to marginal environments and regions most vulnerable to climate change. At the same time, the development of interoperable data infrastructures, harmonised metadata standards, and policies that encourage open data sharing enables meaningful cross-study synthesis and comparative genomics. The forward-looking perspective of GWAS with pangenomics, phenomics, and AI-enabled breeding pipelines is conceptualised in Figure 3.

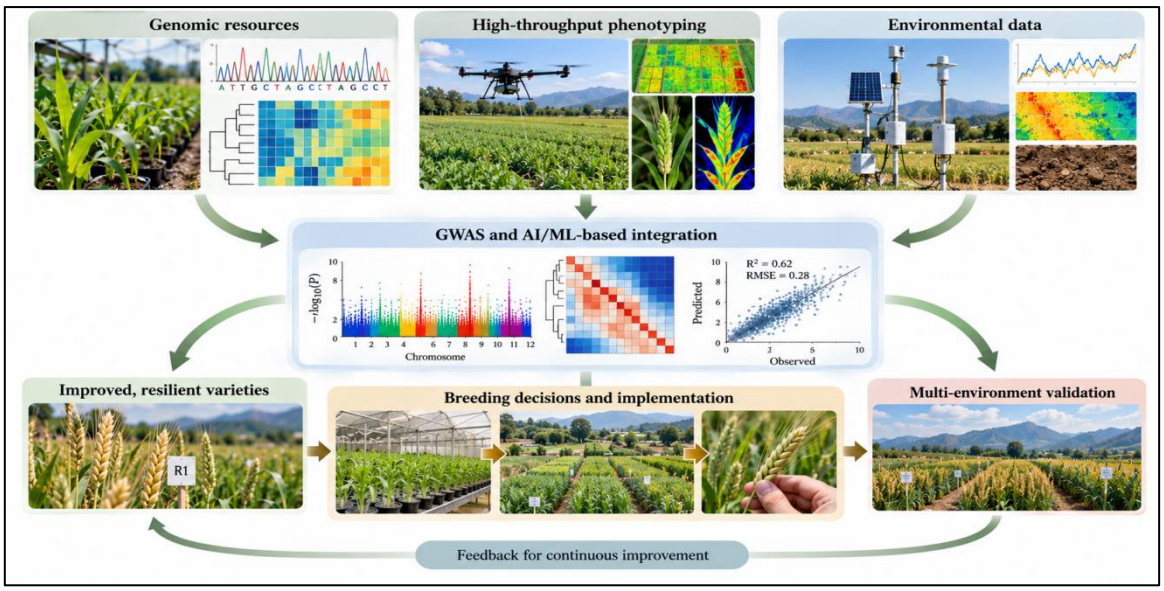



**Figure 3. A vision for an AI-enabled, predictive, next-generation GWAS–breeding framework.**

## CONCLUSION

Genome-wide association studies have transformed crop improvement by enabling high-resolution analysis of complex, environmentally responsive traits. Across diverse crops, GWAS reveals that yield stability, stress tolerance, and disease resistance are polygenic traits strongly shaped by genotype-by-environment interactions. The integration of GWAS with pangenomics, phenomics, and multi-omics has shifted GWAS toward predictive breeding, with successful translation into genomic selection, marker-assisted breeding, and genome editing for climate-resilient cultivars. The full potential of GWAS depends on its integration into closed-loop, data-driven breeding systems that combine diverse germplasm, advanced analytics, and functional validation. Cross-crop synthesis provides a framework for strategic use of conserved resilience mechanisms while maintaining crop-specific adaptations. In this way, GWAS will underpin sustainable and adaptive agriculture in a rapidly changing world.

## ACKNOWLEDGEMENTS

I gratefully acknowledge the valuable contributions of the researchers whose work is cited in this review. I also thank my colleagues for their constructive discussions and insights that helped improve the quality of this manuscript.

## AUTHOR CONTRIBUTION

MH, SF, and UA conceived and designed the present research. MW and GHA analyzed the data. AL wrote the manuscript. SM, SK, MM, and FT contributed to writing and reviewing the manuscript. MMR supervised the study and provided technical guidance. All authors contributed and approved the final version of the manuscript.

**Funding:** Not applicable.

**Data availability:** Not applicable.
**Declarations Conflict of interest**: The authors have declared that no competing interests exist.
**Ethics approval and consent to participate:** Not applicable.
**Research involving human participants/animals:** Not applicable.
**Consent for publication:** Not applicable.

**REFERENCES**

Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., & Cairns, J. E. (2018). Translating highthroughput phenotyping into genetic gain. *Trends in plant science*, *23*(5), 451-466.

Asan, A., Raiders, S. A., & Priess, J. R. (2016). Morphogenesis of the C. elegans intestine involves axon guidance genes. *PLoS genetics*, *12*(4), e1005950.

Bohra, A., Kilian, B., Sivasankar, S., Caccamo, M., Mba, C., McCouch, S. R., & Varshney, R. K. (2022). Reap the crop wild relatives for breeding future crops. *Trends in Biotechnology*, *40*(4), 412-431.

Bayer, P. E., Golicz, A. A., Scheben, A., Batley, J., & Edwards, D. (2020). Plant pan-genomes are the new reference. *Nature plants*, *6*(8), 914-920.

Bernardo, R. (2016). Bandwagons I, too, have known. *Theoretical and applied genetics*, *129*(12), 2323-2332.

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: from polygenic to omnigenic. *Cell*, *169*(7), 1177-1186.

Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., & McMullen, M. D. (2009). The genetic architecture of maize flowering time. *Science*, *325*(5941), 714-718.

Brachi, B., Morris, G. P., & Borevitz, J. O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome biology*, *12*(10), 232.

Bailey-Serres, J., Parker, J. E., Ainsworth, E. A., Oldroyd, G. E., & Schroeder, J. I. (2019). Genetic strategies for improving crop yields. *Nature*, *575*(7781), 109-118.

Bao, W., Li, A., Zhang, Y., Diao, P., Zhao, Q., Yan, T., & Wuriyanghan, H. (2021). Improvement of host-induced gene silencing efficiency via polycistronic-tRNA-amiR expression for multiple target genes and characterization of RNAi mechanism in Mythimna separata. *Plant Biotechnology Journal*, *19*(7), 1370-1385.

Cabrera-Bosquet, L., Crossa, J., von Zitzewitz, J., Serret, M. D., & Luis Araus, J. (2012). Highthroughput phenotyping and genomic selection: The frontiers of crop breeding converge F. *Journal of integrative plant biology*, *54*(5), 312-320.

Collard, B. C., & Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1491), 557-572.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., De Los Campos, G., & Varshney, R. K. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends in plant science*, *22*(11), 961-975.

Chen, Y., Banerjee, D., Mukhopadhyay, A., & Petzold, C. J. (2020). Systems and synthetic biology tools for advanced bioproduction hosts. *Current opinion in biotechnology*, *64*, 101-109.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., & Wincker, P. (2014). Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. *science*, *345*(6199), 950-953.

Cooper, B. S., Vanderpool, D., Conner, W. R., Matute, D. R., & Turelli, M. (2019). Wolbachia acquisition by Drosophila yakuba-clade hosts and transfer of incompatibility loci between distantly related Wolbachia. *Genetics*, *212*(4), 1399-1419.

Cooper, M., Messina, C. D., Podlich, D., Totir, L. R., Baumgarten, A., Hausmann, N. J., & Graham, G. (2014). Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. *Crop and Pasture Science*, *65*(4), 311-336.

Danilevicz, M. F., Fernandez, C. G. T., Marsh, J. I., Bayer, P. E., & Edwards, D. (2020). Plant pangenomics: approaches, applications and advancements. *Current opinion in plant biology*, *54*, 18-25.

D'Agostino, N., & Tripodi, P. (2017). NGS-based genotyping, high-throughput phenotyping and genome-wide association studies laid the foundations for next-generation breeding in horticultural crops. *Diversity*, *9*(3), 38.

Daszkowska-Golec, A. (2022). ABA is important not only under stress–revealed by the discovery of new ABA transporters. *Trends in plant science*, *27*(5), 423-425.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*, *6*(5), e19379.

Fernie, A. R., & Bauwe, H. (2020). Wasteful, essential, evolutionary stepping stone? The multiple personalities of the photorespiratory pathway. *The Plant Journal*, *102*(4), 666-677.

Fernie, A. R., & Tohge, T. (2017). The genetics of plant metabolism. *Annual review of genetics*, *51*(1), 287-310.

Fang, L., Wang, Q., Hu, Y., Jia, Y., Chen, J., Liu, B., & Zhang, T. (2017a). Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nature genetics*, *49*(7), 1089-1098.

Fang, C., Ma, Y., Wu, S., Liu, Z., Wang, Z., Yang, R., & Tian, Z. (2017b). Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome biology*, *18*(1), 161.

Gage, J. L., White, M. R., Edwards, J. W., Kaeppler, S., & de Leon, N. (2018). Selection signatures underlying dramatic male inflorescence transformation during modern hybrid maize breeding. *Genetics*, *210*(3), 1125-1138.

Gao, C. (2021). Genome engineering for crop improvement and future agriculture. *Cell*, *184*(6), 16211635.

Gianola, D. (2013). Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics*, *194*(3), 573-596.

Goyal, R. K., Mattoo, A. K., & Schmidt, M. A. (2021). Rhizobial–host interactions and symbiotic nitrogen fixation in legume crops toward agriculture sustainability. *Frontiers in Microbiology*, *12*, 669404.

Gupta, A., Rico-Medina, A., & Caño-Delgado, A. I. (2020). The physiology of plant responses to drought. *Science*, *368*(6488), 266-269.

Guo, D. L., Zhao, H. L., Li, Q., Zhang, G. H., Jiang, J. F., Liu, C. H., & Yu, Y. H. (2019). Genomewide association study of berry-related traits in grape [Vitis vinifera L.] based on genotypingby-sequencing markers. *Horticulture research*, *6*.

Graham, P. H., & Vance, C. P. (2003). Legumes: importance and constraints to greater use. *Plant physiology*, *131*(3), 872-877.

Herrmann, A., Livanos, P., Zimmermann, S., Berendzen, K., Rohr, L., Lipka, E., & Müller, S. (2021). KINESIN-12E regulates metaphase spindle flux and helps control spindle size in Arabidopsis. *The Plant Cell*, *33*(1), 27-43.

Huang, W. (2024). The current situation and future of using GWAS strategies to accelerate the improvement of crop stress resistance traits. *Molecular Plant Breeding*, *15*.

Houmanat, K., Douaik, A., Charafi, J., Hssaini, L., El Fechtali, M., & Nabloussi, A. (2021). Appropriate statistical methods for analysis of safflower genetic diversity using agglomerative hierarchical cluster analysis through combination of phenotypic traits and molecular markers. *Crop Science*, *61*(6), 4164-4180.

Huang, X., & Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Annual review of plant biology*, *65*(1), 531-551.

Istanbuli, T., Nassar, A. E., Abd El-Maksoud, M. M., Tawkaz, S., Alsamman, A. M., & Hamwieh, A. (2024). Genome-wide association study reveals SNP markers controlling drought tolerance and related agronomic traits in chickpea across multiple environments. *Frontiers in plant science*, *15*, 1260690.

Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., & de los Campos, G. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and applied genetics*, *127*(3), 595-607.

Jiang, Y., An, X., Li, Z., Yan, T., Zhu, T., Xie, K., & Wan, X. (2021). CRISPR/Cas9-based discovery of maize transcription factors regulating male sterility and their functional conservation in plants. *Plant Biotechnology Journal*, *19*(9), 1769-1784.

Juliana, P., Montesinos-López, O. A., Crossa, J., Mondal, S., González Pérez, L., Poland, J., & Singh, R. P. (2019). Integrating genomic-enabled prediction and high-throughput phenotyping in breeding for climate-resilient bread wheat. *Theoretical and Applied Genetics*, *132*(1), 177-194.

Jayakodi, M., Padmarasu, S., Haberer, G., Bonthala, V. S., Gundlach, H., Monat, C., & Stein, N. (2020). The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature*, *588*(7837), 284-289.

Kim, D. H., Yamaguchi, S., Lim, S., Oh, E., Park, J., Hanada, A., & Choi, G. (2008). SOMNUS, a CCCH-type zinc finger protein in Arabidopsis, negatively regulates light-dependent seed germination downstream of PIL5. *The Plant Cell*, *20*(5), 1260-1277.

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant methods*, *9*(1), 29.

Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., & Tian, Z. (2020). Pan-genome of wild and cultivated soybeans. *Cell*, *182*(1), 162-176.

Li, D., Zhang, Z., Gao, X., Zhang, H., Bai, D., Wang, Q., & Qiu, L. J. (2023a). The elite variations in germplasms for soybean breeding. *Molecular Breeding*, *43*(5), 37.

Li, Y., Ruperao, P., Batley, J., Edwards, D., Khan, T., Colmer, T. D., & Sutton, T. (2018). Investigating drought tolerance in chickpea using genome-wide association mapping and genomic selection based on whole-genome resequencing data. *Frontiers in Plant Science*, *9*, 190.

Li, Y., Si, Z., Wang, G., Shi, Z., Chen, J., Qi, G., & Zhang, T. (2023b). Genomic insights into the genetic basis of cotton breeding in China. *Molecular Plant*, *16*(4), 662-677.

Li, X., Zhu, C., Yeh, C. T., Wu, W., Takacs, E. M., Petsch, K. A., & Yu, J. (2012). Genic and nongenic contributions to natural variation of quantitative traits in maize. *Genome research*, *22*(12), 2436-2444.

Lesk, C., Rowhani, P., & Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. *Nature*, *529*(7584), 84-87.

Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, 16, 321–332. https://doi.org/10.1038/nrg3920

Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS genetics*, *12*(2), e1005767.

Loreto, F., & D'Auria, S. (2022). How do plants sense volatiles sent by other plants?. *Trends in plant science*, *27*(1), 29-38.

Leonelli, S., Davey, R. P., Arnaud, E., Parry, G., & Bastow, R. (2017). Data management and best practice for plant science. *Nature Plants*, *3*(6), 1-4.

Montesinos-López, A., Montesinos-López, O. A., Gianola, D., Crossa, J., & Hernández-Suárez, C. M. (2018). Multi-environment genomic prediction of plant traits using deep learners with dense architecture. *G3: Genes, Genomes, Genetics*, *8*(12), 3813-3828.

Migicovsky, Z., Gardner, K. M., Richards, C., Thomas Chao, C., Schwaninger, H. R., Fazio, G., ... & Myles, S. (2021). Genomic consequences of apple improvement. *Horticulture Research*, *8*.

Ma, Z., He, S., Wang, X., Sun, J., Zhang, Y., Zhang, G., & Du, X. (2018). Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nature genetics*, *50*(6), 803-813.

Morrell, P. L., Buckler, E. S., & Ross-Ibarra, J. (2012). Crop genomics: advances and applications. *Nature Reviews Genetics*, *13*(2), 85-96.

Moose, S. P., & Mumm, R. H. (2008). Molecular plant breeding as the foundation for 21st century crop improvement. *Plant physiology*, *147*(3), 969-977.

Mackay, T. F., Stone, E. A., & Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics*, *10*(8), 565-577.

Nordborg, M., & Weigel, D. (2008). Next-generation genetics in plants. *Nature*, *456*(7223), 720-723.

Panigrahi, S., Kumar, U., Swami, S., Singh, Y., Balyan, P., Singh, K. P., & Mir, R. R. (2024). Meta QTL analysis for dissecting abiotic stress tolerance in chickpea. *BMC genomics*, *25*(1), 439.

Prior, H., Jawad, A. K., MacConnachie, L., & Beg, A. A. (2017). Highly efficient, rapid and CoCRISPR-independent genome editing in Caenorhabditis elegans. *G3: Genes, Genomes, Genetics*, *7*(11), 3693-3698.

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PloS one*, *7*(2), e32253.

Ratnaparkhe, M. B., Satpute, G. K., Kumawat, G., Chandra, S., Kamble, V. G., Kavishwar, R., & Nguyen, H. T. (2022). Genomic designing for abiotic stress tolerant soybean. In *Genomic designing for abiotic stress resistant oilseed crops* (pp. 1-73). Cham: Springer International Publishing.

Resende Jr, M. F. R., Munoz, P., Resende, M. D., Garrick, D. J., Fernando, R. L., Davis, J. M., & Kirst, M. (2012). Accuracy of genomic selection methods in a standard data set of loblolly pine (Pinus taeda L.). *Genetics*, *190*(4), 1503-1510.

Ray, D. K., Mueller, N. D., West, P. C., & Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PloS one*, *8*(6), e66428.

Sun, Z., Wang, X., Liu, Z., Gu, Q., Zhang, Y., Li, Z., & Ma, Z. (2017). Genome-wide association study discovered genetic variation and candidate genes of fibre quality traits in Gossypium hirsutum L. *Plant biotechnology journal*, *15*(8), 982-996.

Shingote, P. R., Gotarkar, D. N., Kale, R. R., Limbalkar, O. M., & Wasule, D. L. (2022). Recent advances and applicability of GBS, GWAS, and GS in soybean. *Genotyping by Sequencing for Crop Improvement*, 218-249.

Sone, J., Mitsuhashi, S., Fujita, A., Mizuguchi, T., Hamanaka, K., Mori, K., & Sobue, G. (2019). Longread sequencing identifies GGC repeat expansions in NOTCH2NLC associated with neuronal intranuclear inclusion disease. *Nature genetics*, *51*(8), 1215-1221.

Sulieman, S., & Tran, L. S. P. (2013). Asparagine: an amide of particular distinction in the regulation of symbiotic nitrogen fixation of legumes. *Critical Reviews in Biotechnology*, *33*(3), 309-327.

Springmann, M., Clark, M., Mason-D'Croz, D., Wiebe, K., Bodirsky, B. L., Lassaletta, L., & Willett, W. (2018). Options for keeping the food system within environmental limits. *Nature*, *562*(7728), 519-525.

Shivaiah, K. K., Ding, G., Upton, B., & Nikolau, B. J. (2020). Non-catalytic subunits facilitate quaternary organization of plastidic acetyl-CoA carboxylase. *Plant Physiology*, *182*(2), 756775.

Tester, M., & Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science*, *327*(5967), 818-822.

Thudi, M., Upadhyaya, H. D., Rathore, A., Gaur, P. M., Krishnamurthy, L., Roorkiwal, M., & Varshney, R. K. (2014). Genetic dissection of drought and heat tolerance in chickpea through genome-wide and candidate gene-based association mapping approaches. *Plos one*, *9*(5), e96758.

Tieman, D., Zhu, G., Resende Jr, M. F., Lin, T., Nguyen, C., Bies, D., ... & Klee, H. (2017). A chemical genetic roadmap to improved tomato flavor. *Science*, *355*(6323), 391-394.

Tibbs Cortes, L., Zhang, Z., & Yu, J. (2021). Status and prospects of genome-wide association studies in plants. *The plant genome*, *14*(1), e20077.

Uffelmann, E., Huang, Q. Q., Munung, N. S., De Vries, J., Okada, Y., Martin, A. R., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1), 59.

UNICEF. (2021). *The state of food security and nutrition in the world 2021*. Food and Agriculture Organization of the United Nations.

Varshney, R. K., Thudi, M., Roorkiwal, M., He, W., Upadhyaya, H. D., Yang, W., & Liu, X. (2019). Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nature genetics*, *51*(5), 857-864.

Varshney, R. K., Graner, A., & Sorrells, M. E. (2005). Genomics-assisted breeding for crop improvement. *Trends in plant science*, *10*(12), 621-630.

Varshney, R. K., Bohra, A., Yu, J., Graner, A., Zhang, Q., & Sorrells, M. E. (2021). Designing future crops: genomics-assisted breeding comes of age. *Trends in plant science*, *26*(6), 631-649.

Wei, Z., Abdelrahman, M., Gao, Y., Ji, Z., Mishra, R., Sun, H., & Zhao, K. (2021). Engineering broadspectrum resistance to bacterial blight by CRISPR-Cas9-mediated precise homology directed repair in rice. *Molecular Plant*, *14*(8), 1215-1218.

Wallace, J. G., Rodgers-Melnick, E., & Buckler, E. S. (2018). On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics. *Annual review of genetics*, *52*(1), 421-444.

Wheeler, T., & Von Braun, J. (2013). Climate change impacts on global food security. *Science*, *341*(6145), 508-513.

Wang, Q., Smith, S. M., & Huang, J. (2022). Origins of strigolactone and karrikin signaling in plants. *Trends in Plant Science*, *27*(5), 450-459.

Xiao, B., Li, P., & Wu, Q. (2022). Weak allele versus null allele: which one to select?. *Trends in Genetics*, *38*(10), 989-990.

Xu, Y., & Zhu, Z. (2020). UV-B response: when UVR8 meets MYBs. *Trends Plant Sci*, *25*(6), 515-517.

Xu, Y. (2016). Envirotyping for deciphering environmental impacts on crop plants. *Theoretical and Applied Genetics*, *129*(4), 653-673.

Xu YunBi, X. Y., Li Ping, L. P., Zou Cheng, Z. C., Lu YanLi, L. Y., Xie ChuanXiao, X. C., Zhang XueCai, Z. X., & Olsen, M. S. (2017). Enhancing genetic gain in the era of molecular breeding.

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P. C., Hu, L., & Matsuoka, M. (2016). Genomewide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nature genetics*, *48*(8), 927-934.

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., & Yan, J. (2020). Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Molecular plant*, *13*(2), 187-214.

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature genetics*, *38*(2), 203-208.

Zhang, Y., Massel, K., Godwin, I. D., & Gao, C. (2018a). Applications and potential of genome editing in crop improvement. *Genome biology*, *19*(1), 210.

Zhou, C., Wang, H., Zhu, X., Li, Y., Zhang, B., Tadege, M., & Xia, Z. (2025). Functional genomics: From soybean to legume. *International Journal of Molecular Sciences*, *26*(13), 6323.

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., & Tian, Z. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature biotechnology*, *33*(4), 408-414.

Zhang, S., Ghatak, A., Bazargani, M. M., Bajaj, P., Varshney, R. K., Chaturvedi, P., & Weckwerth, W. (2021). Spatial distribution of proteins and metabolites in developing wheat grain and their differential regulatory response during the grain filling process. *The Plant Journal*, *107*(3), 669-687.

Zhang, H., Lang, Z., & Zhu, J. K. (2018b). Dynamics and function of DNA methylation in plants. *Nature reviews Molecular cell biology*, *19*(8), 489-506.