



OPEN ACCESS

EDITED BY

Yuri Shavrukov,
Flinders University, Australia

REVIEWED BY

Dinesh Yadav,
Deen Dayal Upadhyay Gorakhpur University,
India
Jana Žiarovská,
Slovak University of Agriculture, Slovakia
Mohamed A. M. Atia,
Agricultural Research Center, Egypt

*CORRESPONDENCE

Tae-Chun Park
✉ tcpark@iastate.edu
M. Paul Scott
✉ pscott@iastate.edu

RECEIVED 29 May 2025

ACCEPTED 11 July 2025

PUBLISHED 05 August 2025

CITATION

Park T-C, Silva PC, Lübberstedt T and
Scott MP (2025) Beyond the genome:
the role of functional markers in
contemporary plant breeding.
Front. Plant Sci. 16:1637299.
doi: 10.3389/fpls.2025.1637299

COPYRIGHT

© 2025 Park, Silva, Lübberstedt and Scott. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Beyond the genome: the role of functional markers in contemporary plant breeding

Tae-Chun Park^{1,2*}, Pransiskudura Chamara Silva^{1,3},
Thomas Lübberstedt^{1,2} and M. Paul Scott^{1,2*}

¹Department of Agronomy, Iowa State University, Ames, IA, United States, ²Interdepartmental Plant Biology, Iowa State University, Ames, IA, United States, ³Crop Development Center (CDC), University of Saskatchewan, Saskatoon, SK, Canada

Functional markers (FMs) are derived from polymorphisms that confer phenotypic trait variation, making them powerful tools in plant breeding. Unlike random markers, for which trait associations are unknown, or at best established via linkage or quantitative trait locus (QTL) analysis, FMs are associated with causative polymorphisms, providing precise and reliable information for trait selection. Since the concept of FMs was first proposed in 2003, the emergence and adoption of technologies that were not available at the time have significantly advanced FM discovery and application by enhancing the ability to precisely identify causal variants underlying complex traits, which is a critical prerequisite for FM development. Novel technologies such as high-throughput sequencing, multi-omics, gene editing, and advanced computational tools have enabled the precise identification and functional validation of DNA polymorphisms associated with trait variation. FMs can be used in genomic selection (GS) and modern plant breeding programs by improving selection efficiency and accuracy. While FMs provide numerous benefits, challenges still remain regarding their stability and transferability, and innovative approaches to overcome these limitations are continually being explored. The role of FMs in plant breeding is expected to grow as functional annotation of genomes improves and technologies like genome editing become more accessible. These developments will enable breeders to effectively integrate FMs into breeding pipelines for accelerating genetic gains and addressing global agricultural challenges.

KEYWORDS

functional markers (FMs), random markers (RDMs), sequencing, genomic selection (GS), polymorphism

1 Introduction

Functional markers (FMs) are based on sequences that have been functionally characterized (Andersen and Lübberstedt, 2003; Salgotra and Neal Stewart, 2020). FMs can be developed from any type of DNA polymorphism. To qualify as FMs, polymorphisms between different alleles of genes must cause trait variation. FMs originate from quantitative or qualitative trait polymorphisms (QTPs), which include quantitative or qualitative trait nucleotides (QTNs), based on single nucleotide polymorphisms (SNPs), or indel polymorphisms (QTINDELs). Causes of allelic differences in trait expression include loss-of-function of mutations, changes in gene expression levels, or alterations in gene product structure. FMs are also known as perfect, precision, or diagnostic markers (Salgotra and Neal Stewart, 2020; Salgotra et al., 2014).

Random DNA markers (RDMs) report the state of polymorphisms in randomly selected positions in the genome. RDMs are distributed throughout the genome and are used to assess overall genetic diversity. They serve as tools in genetic mapping and diversity studies. These markers are relatively easy to construct and are effective for characterizing the genetic structure of diverse populations. However, RDMs lack a direct causal relationship with specific gene functions, which can limit their predictive power in marker-assisted selection (MAS). Due to recombination, the association between RDMs and target alleles weakens over successive generations. Despite these limitations, RDMs play a critical role in initial QTL mapping and genomic diversity analyses, providing essential baseline information for further genetic studies (Andersen and Lübberstedt, 2003; Bagge et al., 2007). While RDMs and FMs are indistinguishable from a technical perspective based on marker assays, their distinction lies in the association with traits (Andersen and Lübberstedt, 2003; Bagge et al., 2007). In applications such as MAS, where the goal is to efficiently transfer target traits into different genetic backgrounds, FMs provide a distinct advantage. These applications include marker-assisted backcrossing (MABC) (e.g., Frisch and Melchinger, 2001), F₂ enrichment (Bonnett et al., 2005), and MAS (Lande and Thompson, 1990). The key advantage of FMs lies in their perfect association with target traits, which reduces the risk of false positives due to recombination and improves the accuracy of marker-trait associations (Guo et al., 2010). Therefore, FMs are preferable over RDMs when they are available for tracking specific genes in breeding programs (Hasan et al., 2021).

Although FMs are defined based on polymorphisms with a clearly demonstrated causal relationship to phenotypic variation, not all markers are functionally characterized from the outset. In particular, when the FM concept was first introduced in the early 2000s, limitations in genomic and functional genomics technologies often made it difficult to draw a clear distinction between RDMs and FMs. However, with technological advances, an increasing number of markers initially used as RDMs have since been experimentally validated and reclassified as FMs. For example, in maize, the

opaque2 (*o2*) gene was identified in 1964 as a key regulator of lysine content in the endosperm. It was not until the 1980s that simple sequence repeat (SSR) markers such as *phi057*, *umc1066*, and *phi112*, located within the *o2* gene, were developed by Pioneer and the University of Missouri. These markers were initially used as associative or linked markers to facilitate MAS in quality protein maize (QPM) breeding programs. However, later studies revealed that some of these polymorphisms are closely linked to, or even causative of, the trait by directly affecting gene function. Consequently, markers that were originally used as indirect indicators, or RMD, were later reclassified as gene-based or FMs after functional validation confirmed their biological effects on lysine content (Chand et al., 2024). In rice, a SNP marker within the *BADH2* gene, which was initially a simple association marker, became a FM when the gene was identified as the genetic determinant of 2-acetyl-1-pyrroline, the major compound responsible for aroma (Singh et al., 2022). In wheat, polymorphisms in the *Ppd-D1* gene were shown to play a major role in controlling flowering time, and markers targeting this gene are now widely used as FMs (Beales et al., 2007). These examples illustrate that the boundary between RDMs and FMs is not fixed, and marker classification can evolve with the emergence of new empirical evidence. This continuum is a critical aspect for understanding the historical development of FMs and evaluating their utility in practical crop breeding.

The agricultural sector is currently facing a range of complex challenges, including climate change, emerging pests and diseases, soil degradation, and the urgent need to sustainably feed a growing global population (Hickey et al., 2019; Farooq et al., 2022). These pressures have significantly increased the demand for more precise, rapid, and efficient crop improvement strategies. In this context, FMs which directly target causal variants underlying phenotypic traits, offer high potential. Since the concept was first introduced in 2003 (Andersen and Lübberstedt, 2003), advancements in various technologies have greatly expanded the possibilities for FM development and application. High-throughput sequencing (Sun et al., 2022), genomic resources, and bioinformatics tools (Marsh et al., 2021) now enable the precise and efficient identification of causal variants for target traits (Figure 1). Furthermore, gene editing tools have provided experimental means to functionally validate candidate markers (Ahmar et al., 2020). Beyond traditional genome-wide marker effect-based GS, FMs are gaining attention as tools capable of directly capturing trait-associated variation, thus opening new possibilities for their integration into GS pipelines (Zhang et al., 2023). As such, FMs are no longer just molecular markers but are emerging as core components in breeding strategies aimed at dissecting and harnessing key genetic determinants of agronomic traits.

This paper provides a critical review of the progress in FM development and application in plant breeding since the publication of the FM concept in 2003. It includes, (i) the discovery and development of FMs using forward and reverse genetics, (ii) strategies for FM validation, (iii) the application of FMs in MAS and GS, (iv) recent expansions of FM applications

including genome construction and gene editing-based breeding strategies, and (v) current challenges and future prospects for FM-based breeding.

2 Functional marker discovery and development

2.1 Forward genetics for gene and QTP identification

Forward genetics begins with an observable phenotype and aims to uncover the genes and genetic polymorphisms responsible for trait variation (Sahu et al., 2020). Over the past two decades, advances in mapping technologies, in particular the introduction of genome-wide association studies (GWAS), have greatly increased the precision of forward genetics. While QTL mapping locates larger genome regions affecting traits of interest, GWAS enables fine-scale detection of genetic variants. Map-based cloning is a strategy to confirm gene-trait associations rather than systematically identifying functional variants (Yu et al., 2021). This approach begins with a genetic mapping procedure such as QTL mapping, or other methods to localize genomic regions associated with the trait of interest (Shen et al., 2004). Once a target region has been identified, recombinant individuals are used to progressively narrow down the interval containing the causal gene. The population sizes required to identify causal polymorphisms at high resolution would need to be extremely large, likely in the tens of thousands or more. Thus, while map-based cloning is an established strategy to systematically clone genes for qualitative and quantitative traits, this approach is not feasible for QTP detection (Ganal et al., 2009; Watanabe et al., 2009).

Development of high-throughput and low-cost marker systems, such as Genotyping-by-Sequencing (GBS) were prerequisite for high-resolution GWAS (Thomson et al., 2022). These marker systems dramatically reduced the cost per sample, allowing for large populations to be analyzed. GWAS leverage high-density genotyping and populations with rapid linkage disequilibrium (LD) decay to fine-map candidate genes at high resolution. LD refers to the non-random association of alleles at different loci in a population. In plant breeding, LD is a crucial concept because it affects how genetic markers are associated with traits of interest (Flint-Garcia et al., 2003), and allows to efficiently pinpoint causal genetic variants. In QTL mapping populations, where LD decay is slow, large genomic regions spanning several mega-bases are identified, making it difficult to identify causal genes (Yan et al., 2009; Wallace et al., 2014). By exploiting low LD associations between specific polymorphisms and target traits in GWAS populations, candidate QTPs can be identified (Lu et al., 2010; Wallace et al., 2014). In maize, GWAS has revealed key genetic loci associated with agronomic traits. For example, SNPs associated with plant height have been linked to genes such as *Zm00001d018617* and *Zm00001d023659*, which are involved in gibberellin and auxin signaling pathways (Zhang et al., 2019). Similarly, husk number was associated with variation in the 3'

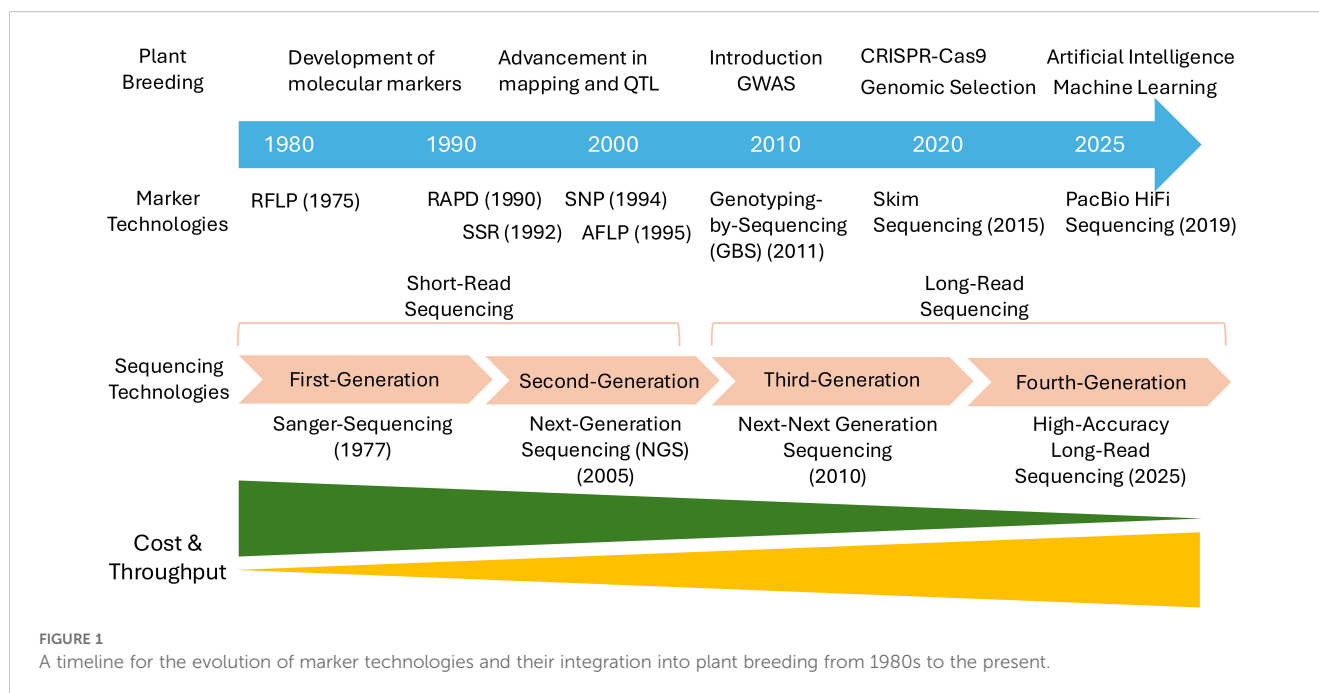
UTR of the ZMET2 gene, a DNA methyltransferase (Wang et al., 2025c). However, distinguishing true causal variants from those merely linked through LD remains a limitation. In maize, LD typically decays within 1–5 kb, depending on population structure and local genomic context (Flint-Garcia et al., 2005; Yan et al., 2011), underscoring the need for further functional validation of candidate QTPs.

Some notable examples highlight the success of forward genetics in QTP discovery. For instance, the *Sub1A* gene in rice was identified by fine mapping of the *Sub1* QTL on chromosome 9 (Figure 2). Expression and functional analyses confirmed that *Sub1A-1*, a specific allele present in tolerant varieties, regulates submergence tolerance. A SNP in *Sub1A-1* was identified as a QTP associated with enhanced survival under flooding. FMs derived from *Sub1A-1* have been successfully implemented in MAS, leading to the development of submergence-tolerant rice varieties such as *Swarna-Sub1* (Xu et al., 2006). Similarly, the *MATRILINEAL* (*MTL*) gene in maize, encoding a patatin-like phospholipase, was identified as a key determinant of maternal haploid induction through GWAS and QTL mapping. Additional QTL, including a significant locus on chromosome 10 harboring a *Kokopelli* ortholog, were identified, confirming the polygenic nature of haploid induction rate (HIR) regulation. A frameshift mutation caused by a 4-bp insertion within *MTL* was validated as a QTP by demonstrating a strong association with increased HIR across diverse maize lines. FMs developed based on this QTP have been successfully used to enhance the efficiency of haploid inducer breeding (Trentin et al., 2023).

2.2 Reverse genetics for candidate gene and QTP identification

Reverse genetics enables researchers to directly verify the function of known genes by deliberately modifying them and observing the resulting phenotypic changes. In contrast to forward genetics, reverse genetics starts with a gene sequence and uses genomic sequencing, functional genomics, and gene editing technologies (Slade et al., 2005; Barkley and Wang, 2008) to assess its role in trait expression. By altering target genes through knockout, knockdown, or precise editing, researchers can evaluate the impact of specific allelic variants and establish causal links between gene function and phenotype.

In 2003, only *Arabidopsis* and rice had fully sequenced genomes (Yu et al., 2002; The Arabidopsis Genome Initiative, 2000). Today, most major crop species, including maize, have multiple high-quality genome assemblies available in public databases. For example, over 40 maize genomes are accessible in MaizeGDB (Portwood et al., 2019). This expansion enables researchers not only to work from a single reference genome, but also to explore allelic variation, presence/absence variation (PAV), and structural genome changes across diverse lines. Pan-genome resources capture genomic diversity beyond a single reference and enable development of broadly applicable markers. PAV offers particularly promising opportunities for designing highly selective markers that target lineage-specific or unique trait-associated genes (Hirsch et al., 2014). These advances



have deepened our understanding of core and orphan genes and made it increasingly feasible to link polymorphisms with biological function (Gao et al., 2019). Nevertheless, while access to extensive genome sequence resources greatly enhances candidate gene discovery, it does not replace the necessity of validating QTPs through experimental studies.

Candidate genes identified by GWAS often serve as starting point for reverse genetics studies. Once loci associated with traits are discovered, targeted mutagenesis or gene editing can be applied to validate gene functions or those of specific polymorphisms. Mutagenesis arises from spontaneous errors during DNA replication and can be artificially induced using a variety of methods, such as chemical mutagenesis (e.g., ethyl methanesulfonate, EMS), physical treatments (e.g., ultraviolet radiation, UV or gamma irradiation), and biological approaches such as transposon insertions and targeted gene-editing technologies like CRISPR/Cas9 (Miao et al., 2013; Udage, 2021; Chen et al., 2023a). Mutagenesis enables validation of the function of specific polymorphisms. By deliberately introducing mutations, scientists can observe the resulting changes in phenotype, thereby separate causal from linked variants to elucidate gene function (Oladosu et al., 2016).

Transposons are mobile genetic elements that are integrated into the genome. They can disrupt gene sequences and regulatory regions. Transposon insertion and excision events can lead to loss-of-function or gain-of-function mutations. Insertion events caused by transposons can lead to loss-of-function or gain-of-function mutations. Transposon tagging has advanced significantly through the integration of NGS, enabling high-throughput identification of insertion sites. Large-scale mutant libraries with mapped or sequenced insertions have been developed in model and crop species, supporting genome-wide reverse genetics studies (Cain et al., 2020; Johnson et al., 2021). One example of using transposons for reverse genetic identification in maize is the Ac/

Ds system applied to the *teosinte branched1 (tb1)* gene. In this approach, Ds transposons were mobilized in the presence of the autonomous Ac element and inserted randomly into the genome. Through high-throughput sequencing of Ds insertion sites, researchers identified a Ds-tagged allele disrupting the *tb1* gene, which was previously known to control plant architecture. The tagged mutants displayed altered tillering phenotypes, validating the gene's function. This insertional mutant population, with sequenced Ds locations, allowed rapid identification of candidate genes and their associated stocks, providing a valuable resource for reverse genetics studies in maize (Studer et al., 2011).

Targeting Induced Local Lesions In Genomics (TILLING) uses chemical or physical mutagens to generate genetic variation and discover beneficial or novel alleles (Barkley and Wang, 2008). TILLING can be applied as reverse-genetics tool in any plant species. It facilitates the speedy and inexpensive generation of induced point mutations (G/C to A/T transition distributed randomly in the genomes) as well as the study of the functions of specific genes in mutants (McCallum et al., 2000). A public reverse genetics TILLING platform is available at the UC Davis Genome Center (<https://genomecenter.ucdavis.edu/>), providing EMS-induced mutant populations and sequencing-based screening for rice, wheat, Arabidopsis, and tomato (Comai and Henikoff, 2006; Studer et al., 2011). Despite the randomness of the induced mutations, systematically screening large mutant populations enables to detect mutations in genes of interest (Gilchrist and Haughn, 2005). In wheat, for example, TILLING has enabled the identification of novel alleles in functionally relevant genes such as *Wx-A1*, *Wx-D1*, and *Ppd-D1*, which are associated with starch composition and flowering time, respectively (Dong et al., 2009; Chen et al., 2012). In contrast, EcoTILLING uncovers natural genetic variation in specific genes, eliminating the need for artificial mutagenesis (Garvin and Gharett, 2007; Backes, 2013).

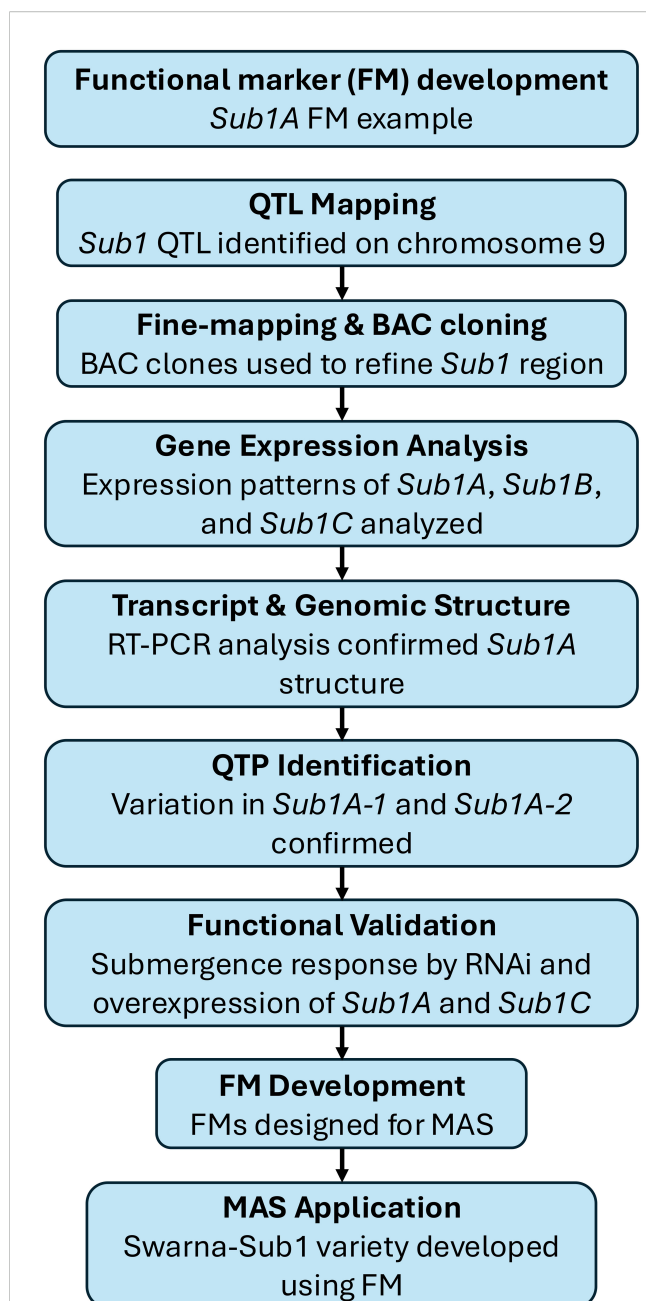


FIGURE 2

The example of functional marker (FM) development process for *Sub1A* in Rice (Xu et al., 2006). This diagram outlines the development of a FM for *Sub1A*, a key gene conferring submergence tolerance in rice. The process began with QTL mapping and fine-mapping using bacterial artificial chromosome (BAC) clones, leading to the identification of *Sub1A*. Gene expression analysis and reverse transcription-polymerase chain reaction (RT-PCR) confirmed its structure and differential expression. Quantitative trait polymorphism (QTP) identification revealed sequence variation between *Sub1A-1* (tolerant) and *Sub1A-2* (intolerant), followed by functional validation through RNA interference (RNAi) and overexpression studies. The FMs were then developed and applied in MAS, resulting in the *Swarna-Sub1* variety with improved flooding tolerance.

SequeTILLING has been proposed as an extension of TILLING with the help of NGS techniques (Weil and Monde, 2009; Backes, 2013). One disadvantage of TILLING is the presence of background mutations that can affect the phenotype and, hence, impede gene function analysis (Szurman-Zubrzycka et al., 2023). Backcrossing may be needed, which is time-consuming. TILLING has inherent limitations. To induce mutations at every nucleotide position within a target gene, an extremely large mutant population would be required which is not feasible, both experimentally and logistically. Thus, the probability of accurately targeting and identifying a specific QTP by TILLING is low. While TILLING is useful for identifying candidate genes, it has limited utility for the precise validation of specific QTPs (Tsai et al., 2011).

Finally, the increasing application of artificial intelligence (AI) and machine learning (ML) models has expanded the toolkit for QTP validation. Deep learning models trained on genomic, transcriptomic, and epigenomic datasets have been used to predict regulatory elements and prioritize functional variants. Notably, AI-guided prioritization of candidate genes has been applied in maize and soybean to narrow down GWAS signals and select variants for functional assays (Washburn et al., 2019; Zhang et al., 2023). As such, AI-assisted validation strategies are expected to play an increasingly important role in bridging large-scale association data with FM development (Washburn et al., 2019; Yoosefzadeh-Najafabadi et al., 2022) (Figure 3).

3 Validation of FMs

The primary goal of validation is to firmly establish a causal relationship between a specific genetic variant and trait expression. Validation is necessary when candidate QTPs have been identified through statistical associations. In some cases, validation methods not only confirm gene function but lead to discovery of new genes or QTPs. For example, gene editing approaches such as CRISPR/Cas9 can be used to create targeted mutations, revealing novel insights into gene function beyond initial candidate predictions. Thus, validation serves both as a confirmation tool and a discovery mechanism (Gaj et al., 2013; Arora and Narula, 2017). Recent advances in gene editing over the past decade have enabled precise, targeted mutagenesis, providing a more efficient and broadly applicable approach.

Gene editing begins with careful selection of a target gene and the design of specific mutations based on predicted functional domains or regulatory elements thought to be critical for gene activity. Subsequently, isogenic mutant and wildtype genotypes are subjected to comprehensive phenotypic analyses. These comparative studies reveal, how the specific alteration impacts the biological processes and trait expression, thereby establishing a direct link between the engineered mutation and its phenotypic effect. Moreover, by systematically correlating these changes with gene function, this approach not only validates candidate genes but

also aids in validation of QTPs (Gilchrist and Haughn, 2010; Saika et al., 2011).

CRISPR/Cas9 targets specific genomic sites by generating double-strand breaks (DSBs) at user-defined sequences using a single guide RNA (sgRNA) that directs the Cas9 nuclease to its target (Ferreira and Choupina, 2022). Once the DSB is induced, the cell predominantly employs the non-homologous end joining (NHEJ) repair pathway. NHEJ is one of the primary pathways for repairing DSBs in DNA, playing a crucial role in cellular DNA damage repair mechanisms (Chang et al., 2017). NHEJ functions by directly ligating the broken DNA ends, often introducing small Indels in the process. Due to its error-prone nature, NHEJ is more useful for intentionally disrupting genes to assess whether a gene is causally linked to loss-of-function phenotypes, rather than for validation of specific QTPs. In contrast, targeted genome editing tools such as TALENs CRISPR-Cas with homology-directed repair (HDR), or base editing offer higher precision and are more suitable for validating QTPs (Zhang et al., 2017; Chen et al., 2019; Ferreira and Choupina, 2022). However, HDR efficiency is generally low in plant cells due to the strong preference for NHEJ over HDR for repairing double-strand breaks. This limitation is particularly evident in somatic tissues, making the practical application of HDR in plant systems more challenging. To overcome these constraints, various strategies have been developed, including the use of single-stranded oligodeoxynucleotides (ssODNs) as repair templates (Schubert et al., 2021), synchronization of the cell cycle to favor HDR activity (Lin et al., 2014), and the application of NHEJ inhibitors (Yang et al., 2020). Despite its lower efficiency, HDR remains a critical tool for precise genome editing, especially in validating candidate FMs.

Base editing is an innovative CRISPR-derived technology that enables the precise conversion of one nucleotide to another without inducing DSBs (Gaj et al., 2013; Kantor et al., 2020; Xu et al., 2021). This method employs a fusion of a catalytically impaired Cas9 (or Cas9 nickase) with a deaminase enzyme, allowing targeted C-to-T or A-to-G substitutions. Its precision is particularly beneficial for creating SNPs (Salgotra and Neal Stewart, 2020). By avoiding DSBs, base editing minimizes the risk of undesired insertions or deletions and off-target effects, enhancing its reliability for QTP validation. Recent studies in rice have demonstrated the potential of base editing for multiplexed nucleotide modifications, making it a valuable tool for precise allele correction and crop improvement (Zong et al., 2017; Kantor et al., 2020; Xu et al., 2021). Prime editing represents a next-generation approach that extends the capabilities of precise genome modifications by allowing all 12 types of base substitutions, as well as small insertions and deletions, without the need for DSBs or donor DNA templates (Anzalone et al., 2019; Thomson et al., 2022). This technique utilizes a fusion protein combining a Cas9 nickase with a reverse transcriptase, guided by a prime editing guide RNA (pegRNA) that encodes the desired edit. Although still emerging, prime editing holds significant promise for accurately delivering beneficial natural mutations (Kantor et al., 2020; Thomson et al., 2022).

While gene editing enables precise genome modifications at the single-nucleotide level, increasing attention has been directed

toward technologies capable of generating or validating larger Indels and structural variations at the chromosomal scale. Notably, CRISPR-Cas9 was used to induce a pericentric inversion spanning 75.5 Mb on chromosome 2 in maize, demonstrating the feasibility of engineering large-scale genomic rearrangements previously unattainable with conventional editing tools (Schwartz et al., 2020). Some studies have further extended this capacity using Cas3-based systems, achieving targeted genomic deletions exceeding 200 kb (Csörgő et al., 2020). Together, these examples highlight the growing potential of advanced genome editing platforms for functional gene analysis and QTP validation, expanding the scope of FM development beyond simple SNP-level variation.

The recently developed non-editing approach Fast Identification of Nucleotide variants by droplet DigITal PCR (FIND-IT) offers a promising alternative for QTP validation. FIND-IT combines large-scale mutagenesis with systematic large-scale pooled genotyping to identify rare mutants, and subsequent genomic and phenomics characterization to validate candidate QTPs. This approach is particularly valuable in species where transformation and editing remain technically challenging or time-consuming (Knudsen et al., 2022). FIND-IT enables high-throughput mutation-to-phenotype mapping through pooled genotyping and droplet digital PCR (ddPCR), offering greater efficiency than traditional TILLING without requiring transgenesis or targeted editing (Huang et al., 2023). The FIND-IT technology was used to screen a mutant library consisting of approximately 500,000 barley individuals. As a result, more than 125 functional alleles were successfully identified. This demonstrates that combining large-scale EMS-induced mutant populations with high-throughput phenotyping is an effective strategy for validating.

TILLING was employed to generate powdery mildew-resistant hexaploidy bread wheat by targeting *TaMlo* genes, orthologues of the barley *Mlo* gene, which confers durable resistance to *Blumeria graminis* f. sp. *Tritici* (*Bgt*). By high-resolution melting (HRM) analysis, 16 missense mutations were identified in *TaMlo-A1*, *TaMlo-B1*, and *TaMlo-D1*, with functional validation in a barley transient expression assay confirming that specific mutations conferred reduced *Mlo* function and increased resistance. Homozygous triple mutants (*tamlo-aabdd*) exhibited strong resistance to *Bgt* without pleiotropic effects such as early leaf senescence. This study demonstrates the effectiveness of non-transgenic gene editing approaches in improving disease resistance in wheat (Acevedo-Garcia et al., 2017). CRISPR-based cytosine base editing (CBE) was used to fine-tune amylose content in rice by precisely modifying the *Wx* gene, which encodes granule-bound starch synthase I (GBSSI), a key enzyme in amylose biosynthesis. Three sgRNAs were designed to introduce specific base substitutions in the N-terminal domain of *Wx*, generating rice lines with amylose levels ranging from 1.4% to 11.9%, enabling precise control over grain quality. Genotypic analysis confirmed stable inheritance of the mutations, with no detectable off-target effects. The study highlights how CRISPR base editing can be leveraged to modify starch composition, improving eating and cooking quality in rice (Xu et al., 2021). CRISPR-Cas9 genome

editing was used to modify the *ARGOS8* gene in maize, improving drought tolerance and grain yield. The *ARGOS8* promoter was replaced with the *GOS2* promoter or inserted into the 5'-UTR, leading to increased expression. qRT-PCR analysis confirmed enhanced transcript levels, validating the modification. Field trials demonstrated a 5-bushel per acre yield increase under drought conditions without yield loss in well-watered environments. This study highlights the potential of precise genome editing for improving complex agronomic traits in maize (Shi et al., 2017). In soybean, CRISPR-Cas9-mediated editing of the *FAD2-2* gene significantly increased oleic acid content by disrupting the microsomal omega-6 desaturase enzyme, which converts oleic acid to linoleic acid. The targeted mutations, introduced using a single-guide RNA (sgRNA), resulted in up to a 65.9% oleic acid content in mutant lines, as confirmed by Near-Infrared Spectroscopy (NIR) analysis. No off-target effects were detected, demonstrating the specificity and efficiency of CRISPR-based genome editing in improving soybean oil quality (Al Amin et al., 2019) (Table 1).

4 Application of FMs in plant breeding

4.1 FMs for marker-assisted backcrossing and selection

The following examples were selected to illustrate the diverse applications of FMs across major crops, including wheat, rice, and maize. These cases represent both widely adopted and more specific uses of FMs in breeding programs, ranging from disease resistance to grain quality traits. In total, six representative cases are highlighted, demonstrating how functionally validated polymorphisms have been successfully integrated into MAS strategies. FMs should be based on functionally validated QTPs, where they are derived from two homozygous genotypes with identical genetic backgrounds but differing QTP alleles, showing clear phenotypic differences for the target trait. Consequently, fully validated QTPs in plants remain still rare.

MAS has been instrumental in addressing simple and oligogenic inherited traits, particularly through methods such as MABC, gene pyramiding, and F₂ enrichment (Cobb et al., 2019). These

approaches have been utilized for more than 20 years to efficiently tackle challenges like linkage drag, where undesirable genes are inherited along with target traits. Progress in marker technologies has significantly improved the efficiency and precision of these methods, enabling breeders to better address these challenges in complex breeding scenarios. In MABC, FMs help track the introgression of beneficial traits from donor parents into recipient lines by identifying alleles associated with those traits. This is particularly important when transferring beneficial traits such as disease resistance or stress tolerance. By targeting specific functional loci, FMs ensure that the desired traits are retained. In gene pyramiding, FMs enable the simultaneous selection of multiple favorable alleles from different parents, combining them into a single genotype with improved performance. This approach is critical for developing crop varieties with stacked traits, such as combined resistance to multiple pathogens. For F₂ enrichment, FMs allow for the early identification of individuals carrying favorable alleles in segregating populations, streamlining the breeding process by focusing resources on the most promising candidates (Collard and Mackill, 2008; Salgotra and Neal Stewart, 2020). These precise applications of FMs significantly enhance the effectiveness of MAS for simple and oligogenic traits (Kumar et al., 2019).

The *Lr34/Yr18/Pm38* locus provides durable, non-race-specific resistance to multiple fungal pathogens, including leaf rust (*Puccinia triticina*), stripe rust (*Puccinia striiformis*), and powdery mildew (*Blumeria graminis*). This resistance is controlled by a single gene encoding an ATP-binding cassette (ABC) transporter, with resistant and susceptible alleles differing by only three sequence polymorphisms (Lagudah et al., 2009). The gene originates from certain wheat landraces and has been introgressed into modern wheat cultivars through MAS. To facilitate its use in breeding, FMs (*cssfr1-*cssfr6**) were developed, enabling precise selection of resistant genotypes. These markers were validated in a diverse set of wheat cultivars, improving breeding efficiency by allowing early-generation screening for durable resistance without the need for pathogen exposure. Similarly, the *Yr36* (*WKS1*) gene, which confers non-race-specific stripe rust resistance at elevated temperatures, was originally identified in wild wheat but was largely absent in modern bread and pasta wheat varieties due to domestication bottlenecks (Fu et al., 2009). FMs for *Yr36* enabled its targeted reintroduction into

TABLE 1 Comparison of three functional marker (FM) validation methods addressed in this paper.

	TILLING (Targeting Induced Local Lesions IN Genomes)	CRISPR-based validation	FIND-IT (Fast Identification of Nucleotide variants by droplet DigITal PCR)
Mutation Source	Induced mutations	Targeted gene editing	Natural or induced variants
Specificity	Low (random mutations)	Very high (precise targeting)	Moderate to high (known variant-based)
Detection Method	Mismatch cleavage, sequencing	Genotyping of edited lines (e.g., sequencing)	Droplet digital PCR (ddPCR)
Speed & Efficiency	Slow, labor-intensive	Moderate to high	Fast, high-throughput
Functional Confirmation	Phenotypic analysis of mutants	Direct gene function validation	Screening rare beneficial alleles
Typical Application	Loss of function screening	QTP validation	Mining useful variants in large pools

The table summarized key features of Targeting Induced Local Lesions IN Genomes (TILLING), CRISPR-based validation, and Fast Identification of Nucleotide variants by droplet DigITal PCR (FIND-IT) in terms of detection mechanism, resolution, speed, and efficiency.

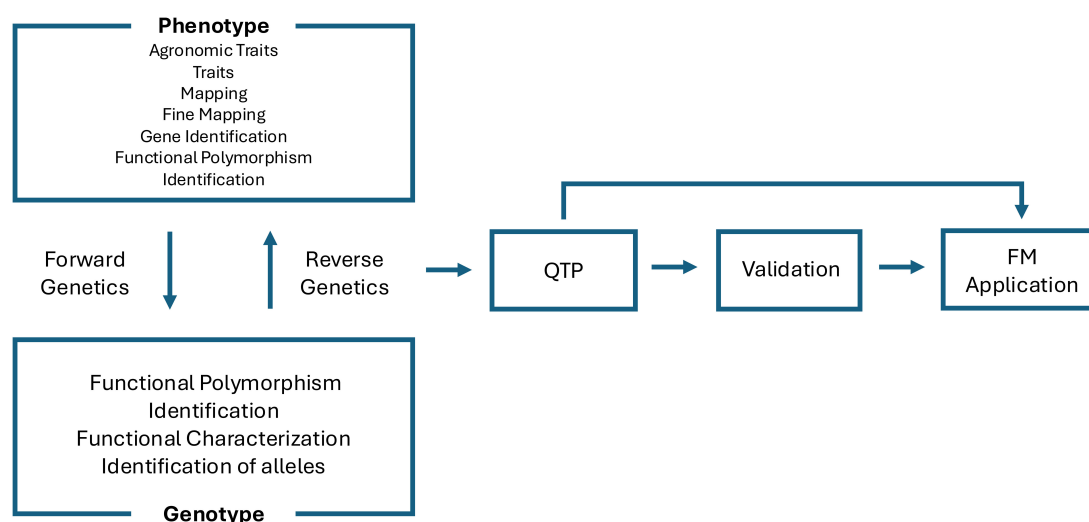


FIGURE 3

A schematic overview illustrating how forward and reverse genetic approaches link various phenotypic traits to genotypic information. Quantitative trait polymorphism (QTP) identification and validation concludes in Functional Marker (FM) application.

elite wheat germplasm via MAS, ensuring the incorporation of resistance without disrupting agronomically favorable traits (Fu et al., 2009; Lagudah et al., 2009). Amylose content (AC) is a key determinant of rice grain cooking and eating quality, with the *Waxy* (*Wx*) gene playing a major role in its regulation. The *Wx* gene encodes *granule-bound starch synthase 1* (*GBSS1*), an enzyme critical for amylose biosynthesis. A SNP in exon 6 (*Ex6A/C*), which results in an amino acid substitution from serine to tyrosine, is associated with intermediate amylose content (*Wx-in*). To facilitate selection for this trait in breeding programs, a FM was developed using polymerase chain reaction with confronting two-pair primers (PCR-CTPP) in a single-tube assay. This marker allows for efficient and cost-effective genotyping of *Wx-in* alleles. The marker was validated in a Chinese mini core collection (*Oryza sativa* L.) and a breeding population, demonstrating high specificity and applicability. By enabling precise selection of genotypes with intermediate amylose content, this marker streamlines the development of rice varieties tailored for consumer preferences in different markets. Its use in breeding programs ensures accurate trait selection without reliance on labor-intensive biochemical assays (Zhou et al., 2018). Moreover, *Fusarium head blight1* (*FHB1*) is a major fungal disease affecting wheat and barley, causing significant yield losses and grain contamination. *Fhb1* is the most widely recognized QTL for *FHB1* resistance, playing a crucial role in reducing disease severity. FMs were developed based on a critical sequence deletion in the *TaHRC* gene within the *Fhb1* region and were validated in a global wheat collection. Comparative genomic analysis between near-isogenic lines (NILs) with contrasting *Fhb1* alleles enabled the identification of two diagnostic markers. Haplotype and sequence analyses across multiple genetic diversity panels confirmed their effectiveness, demonstrating higher selection accuracy than previously used markers. These markers provide a precise and efficient tool for MAS in wheat breeding, facilitating the development of *FHB1*-resistant cultivars and improving disease management strategies (Su et al., 2018).

The development of maize hybrids with improved tolerance to drought and low nitrogen, along with higher provitamin A (PVA) levels in sub-Saharan Africa, can be facilitated by refining and confirming the function of PVA-associated genes in regionally adapted inbred lines (Obeng-Bio et al., 2020). This study aimed to evaluate drought and low-N tolerance and PVA concentrations in early-maturing PVA-quality protein maize (QPM) inbred lines and to identify inbred lines carrying the *crtRB1* and *LcyE* genes as potential sources of favorable PVA alleles. Seventy early-maturing PVA-QPM inbred lines were evaluated under drought, low-N, and optimal conditions in Nigeria over two years. PVA levels were quantified, and allele-specific PCR markers were used to detect the presence of PVA-associated genes. The inbred lines exhibited moderate variation in PVA content; however, the TZEIORQ 55 line demonstrated both high PVA concentration and tolerance to drought and low-N stress. Analysis using the *crtRB1*-3'TE primer and the KASP SNP marker (snpZM0015) consistently identified nine inbred lines, including TZEIORQ 55, harboring the favorable *crtRB1* allele. These inbred lines represent valuable genetic resources for PVA biofortification in maize breeding programs (Obeng-Bio et al., 2020).

4.2 Marker-assisted approaches with FMs including genomic selection

While MAS has been effective for simple or oligogenic traits, it has proven less effective for quantitative traits, which involve polygenic inheritance and complex genetic interactions. Moreover, MAS is advantageous for selecting simple traits, but for complex traits, GS, which is based on information throughout the genome, is more effective. Therefore, efforts to integrate FMs into GS models have emerged (Bhat et al., 2021; Sivabharathi et al., 2024). By decoupling selection from the need for extensive phenotyping, GS improves breeding efficiency and scalability for modern breeding

programs. Although GS does not require FMs, recent advancements have shown that incorporating fixed effects for known genes can further improve prediction accuracies. In particular, GS models incorporating known genes can significantly enhance the prediction accuracy of complex traits (Jeong et al., 2020). For example, integrating FMs associated with specific disease resistance genes into GS models has improved resistance prediction across various crops (Zhang et al., 2021). This approach combines the strengths of traditional MAS with advanced genomic prediction methodologies. Spindel et al. (2016) showed that FMs enhance GS accuracy by incorporating significant markers identified through GWAS as fixed effects in GS models. This integration provides population-specific insights into genetic architecture, improving the reliability of genome estimated breeding value (GEBV) predictions and supporting efficient breeding designs. Ultimately, FMs contribute to maximizing genetic gains while maintaining genetic diversity. Chen et al. (2025), moreover, proposed a model multi-trait ridge regression BLUP (rrBLUP) and *de novo* GWAS to improve genomic prediction accuracy for agronomic traits in maize. FMs identified through *de novo* GWAS were incorporated as fixed effects, significantly enhancing prediction accuracy for low-heritability traits. The multi-trait models further leveraged genetic correlations among traits to improve performance. Bayesian and multi-trait models outperformed rrBLUP for certain traits, particularly when known FMs were included as fixed effects. This suggests that integrating FMs into GS models can enhance prediction accuracy, thereby improving the efficiency of selection in breeding programs.

GS models that incorporate known genes have significantly advanced the field of plant breeding by enhancing the accuracy and efficiency of trait prediction. These models could leverage FMs that are directly associated with specific genes known to influence desirable traits, such as disease resistance, yield, and abiotic stress tolerance. By integrating these known genes into GS models, breeders can more precisely predict the performance of breeding lines and make more informed selection decisions (Jeong et al., 2020; Salgotra and Neal Stewart, 2020; Chen et al., 2023c). For instance, in maize breeding, integrating FMs associated with disease resistance genes such as the *Ht1* gene for northern corn leaf blight resistance has improved the accuracy of resistance prediction (Technow et al., 2013; Hurni et al., 2015). This integration allows breeders to not only select for high-yielding varieties but also ensure that these varieties possess robust disease resistance. Similarly, in wheat, incorporating markers linked to the *Rht-B1* and *Rht-D1* genes, which control plant height and lodging resistance, into GS models has facilitated the development of semi-dwarf varieties that are less lodging and higher yield potential (Zanke et al., 2014). Moreover, the use of known genes in GS models can help address the “large p , small n ” problem, where the number of markers (p) greatly exceeds the number of phenotypic records (n). By focusing on a subset of markers with known effects, the model complexity is reduced, leading to more stable and reliable predictions (Robertson et al., 2019). Additionally, the incorporation of these known FMs can improve the transferability of GS models across different breeding populations and environments, as the effects of these markers are more consistent and well-understood. This approach not only enhances the prediction accuracy but also accelerates the breeding cycle by enabling

earlier selection of superior genotypes (Cerrudo et al., 2018). Overall, the integration of known genes into GS models enhances the efficiency and effectiveness of plant breeding programs.

4.3 Expansion of FM applications

FMs are not limited to their traditional use such as MAS and backcrossing but also have a much broader scope of application in plant breeding, particularly with the advancement of new technologies. In modern plant breeding programs, FMs can enhance the utilization in a way of innovative approaches such as haploid inducer-edit (HI-Edit), genome construction, and promotion of alleles by genome editing (PAGE). HI-Edit is an approach that leverages functionally validated QTPs to directly introduce specific alleles into target genotypes through precise gene editing. Unlike traditional MAS, which relies on marker-based selection, HI-edit directly modifies the genome at the QTP sites, enabling rapid validation and application of functional polymorphisms (Delzer et al., 2024). For instance, by directly introducing a beneficial QTP allele into an elite line, researchers can efficiently test and confirm its phenotypic effect, providing a more accurate understanding of gene-trait relationships. PAGE is an advanced method that directly increases the frequency of favorable alleles through genome editing, significantly enhancing genetic gain compared to traditional GS (Hickey et al., 2016). By directly targeting specific genes or QTPs, PAGE provides a unique opportunity to leverage FMs more efficiently. Based on this background, a strategy can be proposed that integrates PAGE and FMs for application in plant breeding. Specifically, functionally validated QTPs identified through FMs can be used to detect favorable alleles, which can then be directly edited using PAGE to accelerate genetic improvement. This approach is expected to be particularly useful for crops where precise enhancement of complex traits is required. Genome construction in plant breeding is a strategy that involves designing, assembling, and optimizing genetic configurations within a plant genome to achieve desired traits (Varshney et al., 2021). Unlike traditional breeding methods, which rely on random genetic recombination, genome construction is a targeted approach that systematically combines beneficial genes and QTPs to create superior plant varieties. FMs enhance genome construction by enabling precise selection of functionally validated QTPs, ensuring that only beneficial alleles are included in the final genome design. This accelerates trait improvement, minimizes genetic drag, and increases the reliability of selected traits.

5 Challenges and future perspectives

5.1 Limitations in FM development and application

FMs are increasingly valuable tools in plant breeding due to their ability to directly associate genetic variation with phenotypic traits, offering precision in selecting desirable traits and accelerating the breeding process (Kage et al., 2016; Salgotra and Neal Stewart, 2020). However, the application of FMs is not without limitations.

Key challenges include genetic background differences, environmental influences (E), and genotype-by-environment (G×E) interactions, all of which can affect the stability and transferability of FMs. Additionally, epistasis, which manifests as genotype-by-genotype (G×G) interactions, can complicate the predictability of FMs by altering the effects of individual loci depending on the genetic background. Furthermore, incomplete penetrance can result in variability in trait expression even when the causal allele is present. These factors necessitate thorough validation and careful consideration in breeding schemes. In some cases, a single gene may contain redundant multiple QTPs that each contribute to the same phenotype. Traditional FM approaches often focus on a single SNP or specific QTP to predict a trait, potentially overlooking other polymorphisms that also influence the phenotype. This redundancy means that the information provided by an individual QTP may be duplicated by adjacent QTPs, and a marker based only on one may fail to capture the cumulative effect of the entire gene. In contrast, haplotype-based approaches combine multiple SNPs within a gene to represent the optimal haplotype that reflects all functional variations. Such an approach accounts for both additive and epistatic interactions among QTPs, thereby offering a more comprehensive view of the genotype–phenotype relationship. Integrating this strategy into FM development enhances the precision and stability of MAS, ensuring that the selected alleles truly represent the optimal genetic configuration. Recent studies advocate for the use of haplotype-based methods to overcome the limitations inherent in single-QTP markers (Bhat et al., 2021; Sivabharathi et al., 2024). To further enhance the utility of haplotype-based FMs, recent advances in computational tools such as Beagle and SHAPEIT4 have enabled accurate haplotype phasing and imputation, even in complex genomic regions (Browning and Browning, 2007; Delaneau et al., 2019). These tools facilitate the identification of functionally relevant haplotypes by resolving linkage patterns and structural variations. Integration with transcriptomic and epigenomic data, including expression haplotypes (eHaps) and expression quantitative trait locus (eQTL), informed haplotypes, improves the biological resolution of genotype, phenotype associations (Chien et al., 2023). In rice and wheat, for example, haplotype analyses at key loci such as *GW3*, *GW5*, and *FHB1* have enabled the selection of elite alleles for grain shape, and disease resistance (Radecka-Janusik et al., 2022; Xia et al., 2025). Moreover, incorporating haplotype-based markers into GS models has shown improved prediction accuracy, especially for traits influenced by multiple interacting loci (Alemu et al., 2023). This makes haplotype-based FM development a promising strategy for precision breeding under complex trait architectures.

One obstacle is the presence of repetitive sequences in plant genomes, which can complicate primer design and limit the specificity of marker assays. This is particularly problematic in large, complex, or polyploid genomes such as those of wheat or sugarcane, where homoeologous and paralogous sequences share high similarity across subgenomes or gene families. Such redundancy makes it difficult to design markers that uniquely target a single locus, increasing the risk of cross-amplification or false positives in genotyping assays (Claros

et al., 2012). Another significant limitation is gene functional redundancy. In many plant species, important agronomic traits are controlled not by single genes but by gene families with overlapping or compensatory functions. When one gene is mutated, other family members may mask the phenotypic effect, thereby reducing the likelihood of detecting a clear genotype–phenotype relationship. This makes it challenging to identify causative polymorphisms suitable for FM development, especially in cases where loss-of-function alleles do not lead to observable trait variation (Peng, 2019). To mitigate such challenges in practical breeding, several crop-specific strategies have been implemented. For instance, in rice, multi-environment QTL mapping has been used to identify stable loci for drought tolerance, which are less affected by G×E interactions (Dixit et al., 2014). In maize, environment-specific genomic prediction models incorporating FM information have improved trait predictability under varying stress conditions (Rogers and Holland, 2022). These examples demonstrate that integrating FMs with tailored breeding strategies can enhance marker robustness and increase their utility across diverse genetic and environmental contexts.

Genetic background effects influence the stability of FMs, often preventing a marker from performing consistently across diverse genetic backgrounds. These effects arise due to epistasis, where interactions between different loci modify the expression of a target allele. For instance, a FM designed based on a specific genetic variation may work effectively in one genetic background but fail to produce the expected phenotype in another due to differences in interactions with other background genes. This occurs because the phenotypic effect of a given allele is not solely determined by its presence but also by how it interacts with other alleles in the genome. If these interacting loci vary between backgrounds, the expected effect of the FM may not be observed. These interactions can amplify or suppress the marker's effects, leading to phenotypic variability even among individuals carrying the same marker. To ensure the stability of FMs, thorough validation across diverse genetic backgrounds and environments is necessary, along with strategies to minimize the influence of background effects (Guo et al., 2010; Salgotra and Neal Stewart, 2020). The environment (E) plays a crucial role in the stability and effectiveness of FMs. Environmental factors regulate gene expression, altering the phenotypic expression of traits targeted by specific markers, leading to phenotypic variation and instability in marker effects. G×E interaction further complicate this dynamic, as specific genetic variations may express differently across various environments, causing the same marker to have varying effects depending on environmental conditions. For example, an FM associated with drought tolerance may be effective under drought conditions but less so under combined stress factors or other environmental contexts. Due to these environmental influences, markers validated in one environment may not guarantee the same results in other settings. Consequently, breeding programs using FMs must rigorously test marker stability and effectiveness across diverse environmental conditions (Xu et al., 2012; Salgotra and Neal Stewart, 2020). Transferability, another critical factor, involves the ability to apply FMs identified in one population or environment to another. This is particularly important for global breeding

programs aiming to develop cultivars suited to diverse environments. To address these challenges, additional criteria beyond genetic linkage should be considered when defining a polymorphism as “functional.” These include the stability of marker effects across diverse contexts and consistent predictability of the target trait (Sargent et al., 2007). Thorough validation across multiple populations and environments is essential to confirm their reliability and ensure their intended roles in breeding programs. While FMs hold significant potential for improving plant breeding efficiency, their effectiveness depends on careful application and ongoing research to refine their use and address these limitations.

5.2 Future prospects for FMs

Over the next 20 years, FMs are expected to undergo significant advancements, driven by innovations in gene-editing, sequencing technologies, and multi-omics approaches, as well as increasing integration with artificial intelligence (AI) and machine learning (ML). One of the most transformative areas will be the creation of FMs with gene-editing technologies (Arora and Narula, 2017) by directly introducing specific mutations into genes of interest. Furthermore, as NGS costs continue to decrease, breeders will increasingly use whole-genome sequencing routinely, leading to the discovery of high-resolution polymorphisms in both coding and non-coding regions of the genome. This will include markers for regulatory elements, such as enhancers and promoters, which play important roles in gene expression (Salgotra et al., 2014). Looking forward, the application of ML-based gene–trait association models is expected to further enhance the predictive power of forward genetic approaches. By integrating vast datasets from genomic, phenotypic, and environmental sources, these models can identify complex and subtle gene–trait relationships, including non-linear interactions, that might otherwise be overlooked. This will ultimately accelerate genetic gain and facilitate the development of more resilient crop varieties (Negus et al., 2024).

The use of AI and ML will revolutionize the discovery and utilization of FMs and will have a profound impact on future breeding programs. In the phenotyping stage for FM development, large-scale visual data on target traits such as growth status, chlorophyll content, biomass, and disease resistance can be captured using various high-resolution imaging platforms including drones, satellites, ground-based sensors, and robots (Gill et al., 2022; Pinto et al., 2023; DeBruin et al., 2025). Image-based phenotyping enables accurate and consistent trait evaluation across large populations, offering significant efficiency over traditional manual scoring methods that can significantly reduce labor. At the core of this process is the convolutional neural network (CNN), a deep learning architecture specialized for image recognition and feature extraction (Fukushima, 1980). CNNs automatically learn to detect complex visual patterns and are widely used for tasks such as disease diagnosis, classification of healthy versus stressed plants, flowering time prediction, and automated leaf area estimation (Jiang and Li, 2020). Advanced CNN variants (e.g., ResNet, U-Net) also facilitate temporal image analysis, enabling dynamic

monitoring of crop development and high-resolution extraction of phenotypic states at specific time points (Tausen et al., 2020; Malagol et al., 2025). These AI-driven phenotypic data can then be integrated with genomic information to support QTP discovery and FM development.

Recent advances in AI, ML and multi-omics technologies are transforming how causal variants are detected and have strong potential for FM development. ML algorithms are increasingly applied to large-scale genomic data to distinguish causal polymorphisms from background noise (Farooq et al., 2024). For example, models such as eXtreme Gradient Boosting (XGBoost) or deep learning (DL)-based tools like DeepVariant can effectively prioritize candidate SNPs that are likely to affect gene function or regulation, streamlining downstream validation (Poplin et al., 2018; Wang et al., 2025a). These tools improve the accuracy of variant calling by reducing false positives and enhancing the detection of rare or structurally complex polymorphisms that conventional methods often miss (Chen et al., 2023b). In addition, AI-assisted primer design platforms, some of which incorporate ML layers into existing tools help researchers select optimal primers for experimental validation, thereby improving amplification success and efficiency (Dwivedi-Yu et al., 2023; Ghorbani et al., 2025).

Integration of multi-omics data, such as transcriptomics, epigenomics, and chromatin accessibility, provides a more comprehensive view of how genetic variants influence gene expression and phenotypes. By correlating sequence variants with expression profiles, regulatory modifications, and chromatin state across diverse tissues or developmental stages, researchers can prioritize variants that are more likely to be functionally relevant (Mahmood et al., 2022; Liu et al., 2025). Similarly, single-cell genomics is expected to advance, facilitating the identification of FMs (Cuperus, 2022). Single-cell genomics offers unprecedented resolution in understanding how individual cells respond to environmental stimuli or regulate complex traits. Unlike bulk sequencing, which averages signals across heterogeneous tissues, single-cell genomics enables the dissection of cell-type-specific gene expression and regulatory mechanisms. In plants, this approach has been successfully applied to uncover developmental trajectories in Arabidopsis root cells (Denyer et al., 2019) and to map stress-responsive transcriptional programs in rice and maize (Li et al., 2024; Wang et al., 2025b). These studies provide valuable insights into how specific cell types contribute to key agronomic traits such as drought tolerance, nutrient uptake, and disease resistance. Incorporating single-cell data into FM development enables the identification of variants that function in a cell-specific manner, thereby improving the precision and biological relevance of marker selection. This integrative approach enhances the resolution of causal variant detection and increases the biological confidence of selected FMs, ultimately improving the precision of marker development and downstream breeding applications. More importantly, this approach offers a more efficient strategy for FM development, significantly accelerating the overall process of FM development. As such, the convergence of AI, multi-omics, and advanced genomics is poised to become a cornerstone of precision breeding in the coming decades.

In response to environmental challenges, future breeding programs will focus on developing resilient crops using FMs for traits such as abiotic stress tolerance. FMs for genes controlling root architecture, water-use efficiency, and photosynthetic efficiency will become essential for developing varieties that can thrive under increasingly harsh conditions. Additionally, polygenic trait selection will become more refined through the use of multiple FMs, improving the ability to select for complex traits that enhance crop adaptation to different environments (Marsh et al., 2021). The next 20 years will also see the integration of FMs with multi-omics approaches, combining genomics with transcriptomics, metabolomics, and epigenomics. This view of gene-trait relationships will allow breeders to develop more precise FMs that reflect not only genetic variation but also its effects on gene expression and metabolic pathways (Cuperus, 2022).

Finally, due to the importance of global food security, international collaboration in plant breeding programs will become increasingly important. FMs will serve as a key tool for sharing genetic information across borders, ensuring that breeding efforts are aligned to address diverse environmental and agricultural challenges. Public genomic databases will play a significant role in facilitating this collaboration. These platforms will enable researchers and breeders from different areas to access and contribute to a shared pool of genetic data, including FMs and genomic variants. This will accelerate the discovery of new markers and their application in breeding programs, fostering the development of crop varieties that can thrive in various conditions while contributing to sustainable agricultural systems worldwide.

Author contributions

TP: Writing – review & editing, Writing – original draft. PC: Writing – original draft. TL: Supervision, Writing – review & editing, Funding acquisition, Validation, Conceptualization, Project administration. MS: Supervision, Conceptualization, Writing – review & editing, Funding acquisition, Validation, Project administration.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was

supported in part by the U.S. Department of Agriculture, Agricultural Research Service (project number 5030-21000-073-000-D) and by grants from the USDA-NIFA-OREI program, award numbers 2020-51300-32180 and 2021-51300-34896.

Acknowledgments

The findings and conclusions in this publication are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

Generative AI was used to assist with language editing and improving the clarity and conciseness of English expressions in the manuscript. AI was used only for grammar checking and did not contribute to the scientific content or analysis of the manuscript. The author(s) declare that Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Acevedo-Garcia, J., Spencer, D., Thieron, H., Reinstädler, A., Hammond-Kosack, K., Phillips, A. L., et al. (2017). mlo-based powdery mildew resistance in hexaploid bread wheat generated by a non-transgenic TILLING approach. *Plant Biotechnol. J.* 15, 367–378. doi: 10.1111/pbi.12631
- Ahmar, S., Saeed, S., Khan, M. H. U., Khan, S. U., Mora-Poblete, F., Kamran, M., et al. (2020). A revolution toward gene-editing technology and its application to crop improvement. *Int. J. Mol. Sci.* 21, 1–28. doi: 10.3390/ijms21165665
- Al Amin, N., Ahmad, N., Wu, N., Pu, X., Ma, T., Du, Y., et al. (2019). CRISPR-Cas9 mediated targeted disruption of FAD2-2 microsomal omega-6 desaturase in soybean (*Glycine max.*L.). *BMC Biotechnol.* 19, 9. doi: 10.1186/s12896-019-0501-2
- Alemu, A., Batista, L., Singh, P. K., Ceplitis, A., and Chawade, A. (2023). Haplotype-tagged SNPs improve genomic prediction accuracy for Fusarium head blight resistance and yield-related traits in wheat. *Theor. Appl. Genet.* 136, 92. doi: 10.1007/s00122-023-04352-8
- Andersen, J. R., and Lübberstedt, T. (2003). Functional markers in plants. *Trends Plant Sci.* 8, 554–560. doi: 10.1016/j.tplants.2003.09.010
- Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157. doi: 10.1038/s41586-019-1711-4
- Arora, L., and Narula, A. (2017). Gene editing and crop improvement using CRISPR-cas9 system. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01932

- Backes, G. (2013). TILLING and EcoTILLING. In: Lübberstedt, T., and Varshney, R. (eds) *Diagnostics in Plant Breeding*. Dordrecht: Springer, 143–154. doi: 10.1007/978-94-007-5687-8_7
- The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Available online at: www.nature.com.
- Bagge, M., Xia, X., and Lübberstedt, T. (2007). Functional markers in wheat. *Curr. Opin. Plant Biol.* 10, 211–216. doi: 10.1016/j.pbi.2007.01.009
- Barkley, N. A., and Wang, M. L. (2008). *Application of TILLING and EcoTILLING as Reverse Genetic Approaches to Elucidate the Function of Genes in Plants and Animals*. Available online at: www.ncbi.nlm.nih.gov.
- Beales, J., Turner, A., Griffiths, S., Snape, J. W., and Laurie, D. A. (2007). A Pseudo-Response Regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 115, 721–733. doi: 10.1007/s00122-007-0603-4
- Bhat, J. A., Yu, D., Bohra, A., Ganie, S. A., and Varshney, R. K. (2021). Features and applications of haplotypes in crop breeding. *Commun. Biol.* 4, 1266. doi: 10.1038/s42003-021-02782-y
- Bonnett, D., Rebetzke, G., and Spielmeier, W. (2005). Strategies for efficient implementation of molecular markers in wheat breeding. *Mol. Breeding* 15, 75–85. doi: 10.1007/s11032-004-2734-5
- Browning, S. R., and Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81, 1084–1097. doi: 10.1086/521987
- Cain, A. K., Barquist, L., Goodman, A. L., Paulsen, I. T., Parkhill, J., and van Opijnen, T. (2020). A decade of advances in transposon-insertion sequencing. *Nat. Rev. Genet.* 21, 526–540. doi: 10.1038/s41576-020-0244-x
- Cerrudo, D., Cao, S., Yuan, Y., Martinez, C., Suarez, E. A., Babu, R., et al. (2018). Genomic selection outperforms marker assisted selection for grain yield and physiological traits in a maize doubled haploid population across water treatments. *Front. Plant Sci.* 9, doi: 10.3389/fpls.2018.00366
- Chand, G., Muthusamy, V., Zunjare, R. U., Hossain, F., Baveja, A., Chauhan, H. S., et al. (2024). Molecular analysis of opaque2 gene governing accumulation of lysine and tryptophan in maize endosperm. *Euphytica* 220, 155. doi: 10.1007/s10681-024-03414-2
- Chang, H. H. Y., Pannunzio, N. R., Adachi, N., and Lieber, M. R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* 18, 495–506. doi: 10.1038/nrm.2017.48
- Chen, L., Duan, L., Sun, M., Yang, Z., Li, H., Hu, K., et al. (2023a). Current trends and insights on EMS mutagenesis application to studies on plant abiotic stress tolerance and development. *Front. Plant Sci.* 13, doi: 10.3389/fpls.2022.1052569
- Chen, L., Huang, L., Min, D., Phillips, A., Wang, S., Madgwick, P. J., et al. (2012). Development and characterization of a new TILLING population of common bread wheat (*Triticum aestivum* L.). *PLoS One* 7, e41570. doi: 10.1371/journal.pone.0041570
- Chen, Z. Q., Klingberg, A., Hallingbäck, H. R., and Wu, H. X. (2023c). Preselection of QTL markers enhances accuracy of genomic selection in Norway spruce. *BMC Genomics* 24, 147. doi: 10.1186/s12864-023-09250-3
- Chen, N. C., Kolesnikov, A., Goel, S., Yun, T., Chang, P. C., and Carroll, A. (2023b). Improving variant calling using population data and deep learning. *BMC Bioinf.* 24, 197. doi: 10.1186/s12859-023-05294-0
- Chen, Y.-R., Frei, U. K., and Lübberstedt, T. (2025). Multi-trait ridge regression BLUP with de novo GWAS improves genomic prediction for haploid induction ability of haploid inducers in maize. *Front. Plant Sci.* 16, 1614457. doi: 10.3389/fpls.2025.1614457
- Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2019). CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu. Rev. Plant Biol.* 70, 667–697. doi: 10.1146/annurev-arplant-050718
- Chien, P. S., Chen, P. H., Lee, C. R., and Chiou, T. J. (2023). Transcriptome-wide association study coupled with eQTL analysis reveals the genetic connection between gene expression and flowering time in *Arabidopsis*. *J. Exp. Bot.* 74, 5653–5666. doi: 10.1093/jxb/erad262
- Claros, M. G., Bautista, R., Guerrero-Fernández, D., Benzerki, H., Seoane, P., and Fernández-Pozo, N. (2012). Why assembling plant genome sequences is so challenging. *Biol. (Basel)* 1, 439–459. doi: 10.3390/biology1020439
- Cobb, J. N., Biswas, P. S., and Platten, J. D. (2019). Back to the future: revisiting MAS as a tool for modern plant breeding. *Theor. Appl. Genet.* 132, 647–667. doi: 10.1007/s00122-018-3266-4
- Collard, B. C. Y., and Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. B: Biol. Sci.* 363, 557–572. doi: 10.1098/rstb.2007.2170
- Comai, L., and Henikoff, S. (2006). TILLING: Practical single-nucleotide mutation discovery. *Plant J.* 45, 684–694. doi: 10.1111/j.1365-3113X.2006.02670.x
- Csörgő, B., León, L. M., Chau-Ly, I. J., Vasquez-Rifo, A., Berry, J. D., Mahendra, C., et al. (2020). A compact Cascade-Cas3 system for targeted genome engineering. *Nat. Methods* 17, 1183–1190. doi: 10.1038/s41592-020-00980-w
- Cuperus, J. T. (2022). Single-cell genomics in plants: current state, future directions, and hurdles to overcome. *Plant Physiol.* 188, 749–755. doi: 10.1093/plphys/kiab478
- DeBruin, J., Aref, T., Tirado Tolosa, S., Hensley, R., Underwood, H., McGuire, M., et al. (2025). Breaking the field phenotyping bottleneck in maize with autonomous robots. *Commun. Biol.* 8, 467. doi: 10.1038/s42003-025-07890-7
- Delaneau, O., Zagury, J. F., Robinson, M. R., Marchini, J. L., and Dermitzakis, E. T. (2019). Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* 10, 5436. doi: 10.1038/s41467-019-13225-y
- Delzer, B., Liang, D., Szwerdszarf, D., Rodriguez, I., Mardones, G., Elumalai, S., et al. (2024). Elite, transformable haploid inducers in maize. *Crop J.* 12, 314–319. doi: 10.1016/j.cj.2023.10.016
- Denyer, T., Ma, X., Klesen, S., Scacchi, E., Nieselt, K., and Timmermans, M. C. P. (2019). Spatiotemporal developmental trajectories in the *Arabidopsis* root revealed using high-throughput single-cell RNA sequencing. *Dev. Cell* 48, 840–852.e5. doi: 10.1016/j.devcel.2019.02.022
- Dixit, S., Singh, A., Sta Cruz, M. T., Maturan, P. T., Amante, M., and Kumar, A. (2014). Multiple major QTL lead to stable yield performance of rice cultivars across varying drought intensities. *BMC Genet.* 15, 16. doi: 10.1186/1471-2156-15-16
- Dong, C., Dalton-Morgan, J., Vincent, K., and Sharp, P. (2009). A modified TILLING method for wheat breeding. *Plant Genome* 2, 39–47. doi: 10.3835/plantgenome2008.10.0012
- Dwivedi-Yu, J. A., Oppler, Z. J., Mitchell, M. W., Song, Y. S., and Brisson, D. (2023). A fast machine-learning-guided primer design pipeline for selective whole genome amplification. *PLoS Comput. Biol.* 19(4), e1010137. doi: 10.1371/journal.pcbi.1010137
- Farooq, M. A., Gao, S., Hassan, M. A., Huang, Z., Rasheed, A., Hearne, S., et al. (2024). Artificial intelligence in plant breeding. *Trends Genet.* 891–908. doi: 10.1016/j.tig.2024.07.001
- Farooq, M. S., Uzair, M., Raza, A., Habib, M., Xu, Y., Yousuf, M., et al. (2022). Uncovering the research gaps to alleviate the negative impacts of climate change on food security: A review. *Front. Plant Sci.* 13, doi: 10.3389/fpls.2022.927535
- Ferreira, P., and Choupina, A. B. (2022). CRISPR/Cas9 a simple, inexpensive and effective technique for gene editing. *Mol. Biol. Rep.* 49, 7079–7086. doi: 10.1007/s11033-022-07442-w
- Flint-Garcia, S. A., Thornsberry, J. M., and Edwards, S. B. IV (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54, 357–374. doi: 10.1146/annurev.arplant.54.031902.134907
- Flint-Garcia, S. A., Thuillet, A. C., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., et al. (2005). Maize association population: A high-resolution platform for quantitative trait locus dissection. *Plant J.* 44, 1054–1064. doi: 10.1111/j.1365-3113X.2005.02591.x
- Frisch, M., and Melchinger, A. E. (2001). Marker-assisted backcrossing for introgression of a recessive gene. *Crop Science* 41, 1485–1494. doi: 10.2135/cropsci2001.4151485x
- Fu, D., Uauy, C., Distelfeld, A., Blechl, A., Epstein, L., Chen, X., et al. (2009). A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. *Sci. (New York N.Y.)* 323, 1357–1360. doi: 10.1126/science.1166289
- Fukushima, K. (1980). Biological cybernetics neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. doi: 10.1007/BF00344251
- Gaj, T., Gersbach, C. A., and Barbas, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31, 397–405. doi: 10.1016/j.tibtech.2013.04.004
- Ganal, M. W., Altmann, T., and Röder, M. S. (2009). SNP identification in crop plants. *Curr. Opin. Plant Biol.* 12, 211–217. doi: 10.1016/j.pbi.2008.12.009
- Garvin, M. R., and Gharrett, A. J. (2007). DEco-TILLING: an inexpensive method for single nucleotide polymorphism discovery that reduces ascertainment bias. *Mol. Ecol. Notes* 7, 735–746. doi: 10.1111/j.1471-8286.2007.01767.x
- Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D. M., et al. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* 51, 1044–1051. doi: 10.1038/s41588-019-0410-2
- Ghorbani, A., Rostami, M., Ashrafi-Dehkordi, E., and Guzzi, P. H. (2025). AutoPVPrimer: A comprehensive AI-Enhanced pipeline for efficient plant virus primer design and assessment. *PLoS One* 20(1), e0317918. doi: 10.1371/journal.pone.0317918
- Gilchrist, E. J., and Haughn, G. W. (2005). TILLING without a plough: A new method with applications for reverse genetics. *Curr. Opin. Plant Biol.* 8, 211–215. doi: 10.1016/j.pbi.2005.01.004
- Gilchrist, E., and Haughn, G. (2010). Reverse genetics techniques: Engineering loss and gain of gene function in plants. *Brief Funct. Genomic Proteomic* 9, 103–110. doi: 10.1093/bfpg/elp059
- Gill, T., Gill, S. K., Saini, D. K., Chopra, Y., de Koff, J. P., and Sandhu, K. S. (2022). A comprehensive review of high throughput phenotyping and machine learning for plant stress phenotyping. *Phenomics* 2, 156–183. doi: 10.1007/s43657-022-00048-z
- Guo, B., Sleper, D. A., and Beavis, W. D. (2010). Nested association mapping for identification of functional markers. *Genetics* 186, 373–383. doi: 10.1534/genetics.110.115782
- Hasan, N., Choudhary, S., Naaz, N., Sharma, N., and Laskar, R. A. (2021). Recent advancements in molecular marker-assisted selection and applications in plant breeding programmes. *J. Genet. Eng. Biotechnol.* 19, 128. doi: 10.1186/s43141-021-00231-1
- Hickey, J. M., Bruce, C., Whitelaw, A., and Gorjanc, G. (2016). Promotion of alleles by genome editing in livestock breeding programmes. *J. Anim. Breed. Genet.* 133, 83–84. doi: 10.1111/jbg.12206

- Hickey, L. T., Hafeez, N., Robinson, H., Jackson, S. A., Leal-Bertioli, S. C. M., Tester, M., et al. (2019). Breeding crops to feed 10 billion. *Nat. Biotechnol.* 37, 744–754. doi: 10.1038/s41587-019-0152-9
- Hirsch, C. N., Foerster, J. M., Johnson, J. M., Sekhon, R. S., Muttoni, G., Vaillancourt, B., et al. (2014). Insights into the maize pan-genome and pan-transcriptome. *Plant Cell* 26, 121–135. doi: 10.1105/tpc.113.119982
- Huang, L., Gao, G., Jiang, C., Guo, G., He, Q., Zong, Y., et al. (2023). Generating homozygous mutant populations of barley microspores by ethyl methanesulfonate treatment. *abioTECH* 4, 202–212. doi: 10.1007/s42994-023-00108-6
- Hurni, S., Scheuermann, D., Krattinger, S. G., Kessel, B., Wicker, T., Herren, G., et al. (2015). The maize disease resistance gene Htn1 against northern corn leaf blight encodes a wall-associated receptor-like kinase. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8780–8785. doi: 10.1073/pnas.1502522112
- Jeong, S., Kim, J. Y., and Kim, N. (2020). GMStool: GWAS-based marker selection tool for genomic prediction from genomic data. *Sci. Rep.* 10, 19653. doi: 10.1038/s41598-020-76759-y
- Jiang, Y., and Li, C. (2020). Convolutional neural networks for image-based high-throughput plant phenotyping: A review. *Plant Phenomics* 2020, 4152816. doi: 10.34133/2020/4152816
- Johnson, A., Mcassey, E., Diaz, S., Reagin, J., Redd, P. S., Parrilla, D. R., et al. (2021). Development of mPing-based activation tags for crop insertional mutagenesis. *Plant Direct* 5(1), e00300. doi: 10.1002/pld3.300
- Kage, U., Kumar, A., Dhokane, D., Karre, S., and Kushalappa, A. C. (2016). Functional molecular markers for crop improvement. *Crit. Rev. Biotechnol.* 36, 917–930. doi: 10.3109/07388551.2015.1062743
- Kantor, A., McClements, M. E., and Maclaren, R. E. (2020). Crispr-cas9 dna base-editing and prime-editing. *Int. J. Mol. Sci.* 21, 1–22. doi: 10.3390/ijms21176240
- Knudsen, S., Wendt, T., Dockter, C., Thomsen, H. C., Rasmussen, M., Jørgensen, M. E., et al. (2022). FIND-IT: Accelerated trait development for a green evolution. *Sci. Adv.* 8, eabq2266. doi: 10.1126/sciadv.abq2266
- Kumar, S., Kirk, C., Deng, C. H., Shirliff, A., Wiedow, C., Qin, M., et al. (2019). Marker-trait associations and genomic predictions of interspecific pear (*Pyrus*) fruit characteristics. *Sci. Rep.* 9, 9072. doi: 10.1038/s41598-019-45618-w
- Lagudah, E. S., Krattinger, S. G., Herrera-Foessel, S., Singh, R. P., Huerta-Espino, J., Spielmeier, W., et al. (2009). Gene-specific markers for the wheat gene Lr34/Yr18/Pm38 which confers resistance to multiple fungal pathogens. *Theor. Appl. Genet.* 119, 889–898. doi: 10.1007/s00122-009-1097-z
- Lande, R., and Thompson, R. (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124, 743–756. doi: 10.1093/genetics/124.3.743
- Li, P., Liu, Q., Wei, Y., Xing, C., Xu, Z., Ding, F., et al. (2024). Transcriptional landscape of cotton roots in response to salt stress at single-cell resolution. *Plant Commun.* 5, 100740. doi: 10.1016/j.xplc.2023.100740
- Lin, S., Staahl, B. T., Alla, R. K., and Doudna, J. A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* 3, e04766. doi: 10.7554/eLife.04766
- Liu, Y., Gao, X., Liu, H., Yang, X., Liu, X., Xu, F., et al. (2025). Constraint of accessible chromatin maps regulatory loci involved in maize speciation and domestication. *Nat. Commun.* 16, 2477. doi: 10.1038/s41467-025-57932-1
- Lu, Y., Zhang, S., Shah, T., Xie, C., Hao, Z., Li, X., et al. (2010). Joint linkage-linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19585–19590. doi: 10.1073/pnas.1006105107
- Mahmood, U., Li, X., Fan, Y., Chang, W., Niu, Y., Li, J., et al. (2022). Multi-omics revolution to promote plant breeding efficiency. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1062952
- Malagol, N., Rao, T., Werner, A., Töpfer, R., and Hausmann, L. (2025). A high-throughput ResNet CNN approach for automated grapevine leaf hair quantification. *Sci. Rep.* 15, 1590. doi: 10.1038/s41598-025-85336-0
- Marsh, J. I., Hu, H., Gill, M., Batley, J., and Edwards, D. (2021). Crop breeding for a changing climate: integrating phenomics and genomics with bioinformatics. *Theor. Appl. Genet.* 134, 1677–1690. doi: 10.1007/s00122-021-03820-3
- McCallum, C. M., Comai, L., Greene, E. A., and Henikoff, S. (2000). Targeted screening for induced mutations A B C. Available online at: <http://biotech.nature.com>.
- Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., et al. (2013). Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res.* 23, 1233–1236. doi: 10.1038/cr.2013.123
- Negus, K. L., Li, X., Welch, S. M., and Yu, J. (2024). “The role of artificial intelligence in crop improvement,” in *Advances in Agronomy* (Cambridge, MA: Academic Press Inc), 1–66. doi: 10.1016/bs.agron.2023.11.001
- Obeng-Bio, E., Badu-Apraku, B., Elorhor Ifie, B., Danquah, A., Blay, E. T., and Dadzie, M. A. (2020). Phenotypic characterization and validation of provitamin A functional genes in early maturing provitamin A-quality protein maize (*Zea mays*) inbred lines. *Plant Breed.* 139, 575–588. doi: 10.1111/pbr.12798
- Oladosu, Y., Rafii, M. Y., Abdullah, N., Hussin, G., Ramli, A., Rahim, H. A., et al. (2016). Principle and application of plant mutagenesis in crop improvement: A review. *Biotechnol. Biotechnol. Equip.* 30, 1–16. doi: 10.1080/13102818.2015.1087333
- Peng, J. (2019). Gene redundancy and gene compensation: An updated view. *J. Genet. Genomics* 46, 329–333. doi: 10.1016/j.jgg.2019.07.001
- Pinto, F., Zaman-Allah, M., Reynolds, M., and Schulthess, U. (2023). Satellite imagery for high-throughput phenotyping in breeding plots. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1114670
- Poplin, R., Chang, P. C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., et al. (2018). A universal snp and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* 36, 983. doi: 10.1038/nbt.4235
- Portwood, J. L., Woodhouse, M. R., Cannon, E. K., Gardiner, J. M., Harper, L. C., Schaeffer, M. L., et al. (2019). Maizegdb 2018: The maize multi-genome genetics and genomics database. *Nucleic Acids Res.* 47, D1146–D1154. doi: 10.1093/nar/gky1046
- Radecka-Janusik, M., Piechota, U., Piaskowska, D., Górka, T., and Czembor, P. (2022). Evaluation of fusarium head blight resistance effects by haplotype-based genome-wide association study in winter wheat lines derived by marker backcrossing approach. *Int. J. Mol. Sci.* 23, 14233. doi: 10.3390/ijms232214233
- Robertson, C. D., Hjortshøj, R. L., and Janss, L. L. (2019). Genomic selection in cereal breeding. *Agronomy* 9, 95. doi: 10.3390/agronomy9020095
- Rogers, A. R., and Holland, J. B. (2022). Environment-specific genomic prediction ability in maize using environmental covariates depends on environmental similarity to training data. *G3: Genes Genomes Genet.* 12, jkab440. doi: 10.1093/g3journal/jkab440
- Saika, H., Oikawa, A., Matsuda, F., Onodera, H., Saito, K., and Toki, S. (2011). Application of gene targeting to designed mutation breeding of high-tryptophan rice. *Plant Physiol.* 156, 1269–1277. doi: 10.1104/pp.111.175778
- Sahu, P. K., Sao, R., Mondal, S., Vishwakarma, G., Gupta, S. K., Kumar, V., et al. (2020). Next generation sequencing based forward genetic approaches for identification and mapping of causal mutations in crop plants: a comprehensive review. *Plants* 9, 1355. doi: 10.3390/plants9101355
- Salgotra, R. K., Gupta, B. B., and Stewart, C. N. (2014). From genomics to functional markers in the era of next-generation sequencing. *Biotechnol. Lett.* 36, 417–426. doi: 10.1007/s10529-013-1377-1
- Salgotra, R. K., and Neal Stewart, C. (2020). Functional markers for precision plant breeding. *Int. J. Mol. Sci.* 21, 1–33. doi: 10.3390/ijms21134792
- Sargent, D. J., Rys, A., Nier, S., Simpson, D. W., and Tobutt, K. R. (2007). The development and mapping of functional markers in *Fragaria* and their transferability and potential for mapping in other genera. *Theor. Appl. Genet.* 114, 373–384. doi: 10.1007/s00122-006-0441-9
- Schubert, M. S., Thommandru, B., Woodley, J., Turk, R., Yan, S., Kurgan, G., et al. (2021). Optimized design parameters for CRISPR Cas9 and Cas12a homology-directed repair. *Sci. Rep.* 11, 19482. doi: 10.1038/s41598-021-98965-y
- Schwartz, C., Lenderts, B., Feigenbutz, L., Barone, P., Llaca, V., Fengler, K., et al. (2020). CRISPR–Cas9-mediated 75.5-Mb inversion in maize. *Nat. Plants* 6, 1427–1431. doi: 10.1038/s41477-020-00817-6
- Shen, Y. J., Jiang, H., Jin, J. P., Zhang, Z. B., Xi, B., He, Y. Y., et al. (2004). Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. *Plant Physiol.* 135, 1198–1205. doi: 10.1104/pp.103.038463
- Shi, J., Gao, H., Wang, H., Lafitte, H. R., Archibald, R. L., Yang, M., et al. (2017). ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. *Plant Biotechnol. J.* 15, 207–216. doi: 10.1111/pbi.12603
- Singh, G., Gopala Krishnan, S., Kumar, A., Vinod, K. K., Bollinedi, H., Ellur, R. K., et al. (2022). Molecular profiling of BADH2 locus reveals distinct functional allelic polymorphism associated with fragrance variation in Indian aromatic rice germplasm. *Physiol. Mol. Biol. Plants* 28, 1013–1027. doi: 10.1007/s12298-022-01181-9
- Sivabharathi, R. C., Rajagopalan, V. R., Suresh, R., Sudha, M., Karthikeyan, G., Jayakanthan, M., et al. (2024). Haplotype-based breeding: A new insight in crop improvement. *Plant Sci.* 346, 112129. doi: 10.1016/j.plantsci.2024.112129
- Slade, A. J., Fuerstenberg, S. I., Loeffler, D., Steine, M. N., and Facciotti, D. (2005). A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nat. Biotechnol.* 23, 75–81. doi: 10.1038/nbt1043
- Spindel, J., Begum, H., Akdemir, D., Collard, B., Redoña, E., and Jannink, J. L. (2016). Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. *Heredity*, 116, 395–408. doi: 10.1038/hdy.2015.113
- Studer, A., Zhao, Q., Ross-Ibarra, J., and Doebley, J. (2011). Identification of a functional transposon insertion in the maize domestication gene tb1. *Nat. Genet.* 43, 1160–1163. doi: 10.1038/ng.942
- Su, Z., Jin, S., Zhang, D., and Bai, G. (2018). Development and validation of diagnostic markers for Fhb1 region, a major QTL for Fusarium head blight resistance in wheat. *Theor. Appl. Genet.* 131, 2371–2380. doi: 10.1007/s00122-018-3159-6
- Sun, Y., Shang, L., Zhu, Q. H., Fan, L., and Guo, L. (2022). Twenty years of plant genome sequencing: achievements and challenges. *Trends Plant Sci.* 27, 391–401. doi: 10.1016/j.tplants.2021.10.006
- Szurman-Zubrzycka, M., Kurowska, M., Till, B. J., and Szarejko, I. (2023). Is it the end of TILLING era in plant science? *Front. Plant Sci.* 14, 1160695. doi: 10.3389/fpls.2023.1160695
- Tausen, M., Clausen, M., Moeskjær, S., Shihavuddin, A. S. M., Dahl, A. B., Janss, L., et al. (2020). Greenotyper: image-based plant phenotyping using distributed computing and deep learning. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.01181
- Technow, F., Bürger, A., and Melchinger, A. E. (2013). Genomic prediction of northern corn leaf blight resistance in maize with combined or separated training sets

- for heterotic groups. *G3: Genes Genomes Genet.* 3, 197–203. doi: 10.1534/g3.112.004630
- Thomson, M. J., Biswas, S., Tsakirpaloglou, N., and Septiningsih, E. M. (2022). Functional allele validation by gene editing to leverage the wealth of genetic resources for crop improvement. *Int. J. Mol. Sci.* 23, 6565. doi: 10.3390/ijms23126565
- Trentin, H. U., Krause, M. D., Zunjare, R. U., Almeida, V. C., Peterlini, E., Rotarenco, V., et al. (2023). Genetic basis of maize maternal haploid induction beyond MATRILINEAL and ZmDMP. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1218042
- Tsai, H., Howell, T., Nitcher, R., Missirian, V., Watson, B., Ngo, K. J., et al. (2011). Discovery of rare mutations in populations: Tilling by sequencing. *Plant Physiol.* 156, 1257–1268. doi: 10.1104/pp.110.169748
- Udage, A. C. (2021). Introduction to plant mutation breeding: Different approaches and mutagenic agents. *J. Agric. Sci. - Sri Lanka* 16, 466–483. doi: 10.4038/jas.v16i03.9472
- Varshney, R. K., Bohra, A., Yu, J., Graner, A., Zhang, Q., and Sorrells, M. E. (2021). Designing future crops: genomics-assisted breeding comes of age. *Trends Plant Sci.* 26, 631–649. doi: 10.1016/j.tplants.2021.03.010
- Wallace, J. G., Bradbury, P. J., Zhang, N., Gibon, Y., Stitt, M., and Buckler, E. S. (2014). Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS Genet.* 10, e1004845. doi: 10.1371/journal.pgen.1004845
- Wang, T., Wang, F., Deng, S., Wang, K., Feng, D., Xu, F., et al. (2025b). Single-cell transcriptomes reveal spatiotemporal heat stress response in maize roots. *Nat. Commun.* 16. doi: 10.1038/s41467-024-55485-3
- Wang, Y., Wang, S., Lu, D., Chen, M., Li, B., Li, Z., et al. (2025c). Genome-wide association study and candidate gene mining of husk number trait in maize. *Int. J. Mol. Sci.* 26, 3437. doi: 10.3390/ijms26073437
- Wang, C., Wu, S., Yao, Z., Luo, E., Deng, J., and Liu, J. (2025a). QTG-LGBM: A method of prioritizing causal genes in quantitative trait loci in maize. *Crop J.* 16, 177. doi: 10.1016/j.cj.2025.03.004
- Washburn, J. D., Mejia-Guerra, M. K., Ramstein, G., Kremling, K. A., Valluru, R., Buckler, E. S., et al. (2019). Evolutionarily informed deep learning methods for predicting relative transcript abundance from DNA sequence. *Proc. Natl. Acad. Sci. U.S.A.* 116, 5542–5549. doi: 10.1073/pnas.1814551116
- Watanabe, S., Hideshima, R., Zhengjun, X., Tsubokura, Y., Sato, S., Nakamoto, Y., et al. (2009). Map-based cloning of the gene associated with the soybean maturity locus E3. *Genetics* 182, 1251–1262. doi: 10.1534/genetics.108.098772
- Weil, C., and Monde, R. (2009). TILLING and point mutation detection. In: Benetzen, J. L., and Hake, S. (eds) *Handbook of Maize: Genetics and Genomics*. New York, NY: Springer, 639–647. doi: 10.1007/978-0-387-77863-1_29
- Xia, D., Xu, P., Zheng, S., Xue, L., Yin, Y., Yao, Z., et al. (2025). Haplotype analysis of the GL7/GW7/SLG7 gene and its application in improving the grain quality of three-line hybrid rice. *Rice* 18, 38. doi: 10.1186/s12284-025-00786-4
- Xu, Y., Lin, Q., Li, X., Wang, F., Chen, Z., Wang, J., et al. (2021). Fine-tuning the amylose content of rice by precise base editing of the Wx gene. *Plant Biotechnol. J.* 19, 11–13. doi: 10.1111/pbi.13433
- Xu, K., Xu, X., Fukao, T., Canlas, P., Maghirang-Rodriguez, R., Heuer, S., et al. (2006). Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* 442, 705–708. doi: 10.1038/nature04920
- Xu, Y., Lu, Y., Xie, C., Gao, S., Wan, J., and Prasanna, B. M. (2012). Whole-genome strategies for marker-assisted plant breeding. *Mol. Breed.* 29, 833–854. doi: 10.1007/s11032-012-9699-6
- Yan, J., Shah, T., Warburton, M. L., Buckler, E. S., McMullen, M. D., and Crouch, J. (2009). Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS One* 4, e8451. doi: 10.1371/journal.pone.0008451
- Yan, J., Warburton, M., and Crouch, J. (2011). Association mapping for enhancing maize (*Zea mays* L.) genetic improvement. *Crop Sci.* 51, 433–449. doi: 10.2135/cropsci2010.04.0233
- Yang, H., Ren, S., Yu, S., Pan, H., Li, T., Ge, S., et al. (2020). Methods favoring homology-directed repair choice in response to crispr/cas9 induced-double strand breaks. *Int. J. Mol. Sci.* 21, 1–20. doi: 10.3390/ijms21186461
- Yoosefzadeh-Najafabadi, M., Eskandari, M., Torabi, S., Torkamaneh, D., Tulpan, D., and Rajcan, I. (2022). Machine-learning-based genome-wide association studies for uncovering QTL underlying soybean yield and its components. *Int. J. Mol. Sci.* 23, 5538. doi: 10.3390/ijms23105538
- Yu, J., Hu, S., Wang, J., Ka-Shu Wong, G., Li, S., Liu, B., et al. A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. indica). *Science* 296, 79–92. doi: 10.1126/science.1068037
- Yu, L., Nie, Y., Jiao, J., Jian, L., and Zhao, J. (2021). The sequencing-based mapping method for effectively cloning plant mutated genes. *Int. J. Mol. Sci.* 22, 6224. doi: 10.3390/ijms22126224
- Zanke, C. D., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., Korzun, V., et al. (2014). Whole genome association mapping of plant height in winter wheat (*Triticum aestivum* L.). *PLoS One* 9, e113287. doi: 10.1371/journal.pone.0113287
- Zhang, W., Boyle, K., Brule-Babel, A., Fedak, G., Gao, P., Djama, Z. R., et al. (2021). Evaluation of genomic prediction for fusarium head blight resistance with a multi-parental population. *Biol. (Basel)* 8, 756. doi: 10.3390/biology10080756
- Zhang, H., Zhang, J., Lang, Z., Botella, J. R., and Zhu, J. K. (2017). Genome editing —Principles and applications for functional genomics research and crop improvement. *CRC Crit. Rev. Plant Sci.* 36, 291–309. doi: 10.1080/07352689.2017.1402989
- Zhang, Y., Wan, J., He, L., Lan, H., and Li, L. (2019) Genome-wide association analysis of plant height using the maize F1 population. *Plants (Basel)* 8:432. doi: 10.3390/plants8100432
- Zhang, Y., Zhang, M., Ye, J., Xu, Q., Feng, Y., Xu, S., et al. (2023). Integrating genome-wide association study into genomic selection for the prediction of agronomic traits in rice (*Oryza sativa* L.). *Mol. Breed.* 43, 81. doi: 10.1007/s11032-023-01423-y
- Zhou, L., Chen, S., Yang, G., Zha, W., Cai, H., Li, S., et al. (2018). A perfect functional marker for the gene of intermediate amylose content Wx-in in rice (*Oryza sativa* L.) ARTICLE. *Crop Breed. Appl. Biotechnol.* 18, 103–109. doi: 10.1590/1984
- Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., et al. (2017). Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* 35, 438–440. doi: 10.1038/nbt.3811