Journal of Applied Genetics
https://doi.org/10.1007/s13353-025-01034-7

PLANT GENETICS • REVIEW

Accelerating genetic gain through integrated genomic selection in crop plants

Bandela Edukondalu[1] · Nunavath Aswini[2] · Amaresh[3] · Gopalareddy Krishnappa[4] · Buruka Soundharya[1] · Gottimukkala Nikhitha[5] · T. Lakshmi Pathy[3] · Kasanaboina Krishna[6] · Yadla Hari[1] · Vinayaka[3]

Received: 24 August 2025 / Revised: 8 November 2025 / Accepted: 20 November 2025

Abstract
Meeting the projected 70% rise in agricultural output by 2050 to sustain a global population of 9.6 billion poses a formidable challenge amid intensifying biotic and abiotic stresses. Traditional breeding methods, although foundational, are limited in their ability to improve complex polygenic traits such as yield, stress tolerance, and disease resistance. Genomic selection (GS) has emerged as a transformative approach that leverages genome-wide markers to predict breeding values with higher accuracy and efficiency. Unlike marker-assisted selection (MAS) and genome-wide association studies (GWAS), which emphasize major-effect loci, GS captures the cumulative contribution of numerous small-effect loci, enabling faster genetic gains for complex traits. This review outlines the conceptual framework, evolution, and integration of GS with cutting-edge technologies such as high-throughput genotyping, phenomics, multi-omics, and machine learning. It also discusses key achievements, implementation strategies, and the potential of GS to enhance selection accuracy, shorten breeding cycles, and develop climate-resilient, high-yielding cultivars. The integration of GS within modern breeding pipelines represents a paradigm shift toward sustainable crop improvement and global food security in an era of climatic uncertainty.

Keywords Genomic selection · Marker-assisted breeding · Genomic estimated breeding value · Climate resilience · Sustainable agriculture

Introduction

The global population is projected to reach 9.6 billion by 2050, requiring a 70% increase in agricultural productivity to meet food demand (Godfray and Garnett 2014). Achieving this goal is increasingly constrained by biotic and abiotic stresses that severely impact global crop yields (Hossain et al. 2020; Nunavath et al. 2025a). Pests and diseases alone account for nearly 30–40% of total yield losses, with annual global losses exceeding USD 470 billion (Savary et al. 2019; Skendžić et al. 2021). Fungal infections cause 10–15% yield reductions, while abiotic factors like heat, drought, nutrient depletion, and salinity further threaten productivity (Munns and Tester 2008). For instance, each 1 °C rise in temperature reduces wheat yield by ~ 6%, and drought can lower yields by up to 40% (Lesk et al. 2016). Moreover, poor soil fertility contributes to 30–50% of yield gaps in developing regions (Mueller et al. 2012). Addressing these constraints requires developing stress-resilient, high-yielding, and resource-efficient cultivars through advanced breeding strategies (Kopeć, 2024).

Communicated by: Izabela Pawłowicz

✉ Amaresh
amaresh541@gmail.com

1 Professor Jayashankar Telangana Agricultural University (PJTAU), Hyderabad, Telangana 500030, India

2 ICAR – National Institute of Biotic Stress Management, Raipur 493225, India

3 ICAR – Sugarcane Breeding Institute, Coimbatore, Tamil Nadu 641007, India

4 ICAR – National Institute of Seed Science and Technology, Regional Station, Bengaluru 560065, India

5 Department of Crop Sciences, University of Hohenheim, 70599 Stuttgart, Germany

6 International Crops Research Institute for the Semi-Arid Tropics, Patancheru, Hyderabad, Telangana 502324, India

Published online: 02 January 2026

Traditional breeding, primarily based on phenotypic and pedigree selection, has been instrumental in crop improvement but remains slow, imprecise, and environment-dependent. The introduction of molecular markers like Simple Sequence Repeats (SSRs) and Single Nucleotide Polymorphisms (SNPs) ushered in marker-assisted selection (MAS), enhancing selection efficiency (Collard et al. 2005; Collard and Mackill 2008). However, MAS remains effective mainly for traits governed by major-effect loci and less so for polygenic, environment-sensitive traits (Bernardo 2008). The subsequent development of genome-wide association studies (GWAS) improved resolution but still focused on large-effect variants and faced issues of population structure and missing heritability (Zhu et al. 2008; Tibbs Cortes et al. 2021). The advent of Genomic Selection (GS) marked a major shift in breeding paradigms. By utilizing genome-wide marker data to predict breeding values, GS captures the collective effect of numerous small-effect loci, enabling accurate early selection and accelerated breeding cycles (Meuwissen et al. 2001; Crossa et al. 2017a, b). When integrated with high-throughput genotyping, advanced phenomics, and computational tools such as machine learning, GS facilitates continuous, data-driven improvement of complex traits. Consequently, GS has become central to modern crop improvement programs aimed at enhancing genetic gain and resilience under changing climatic and resource-limited conditions.

In summary, the transition from conventional, phenotype-based breeding toward genomic approaches reflects a critical paradigm shift driven by the need to accelerate genetic gains and improve the predictability of complex traits. Traditional methods, while foundational, remain slow, resource-intensive, and ineffective for traits governed by multiple small-effect loci. This historical context underscores the emergence of molecular and statistical tools that leverage genome-wide data to enhance selection precision. The subsequent section examines in detail the inherent limitations of pre-genomic selection methodologies, establishing the scientific rationale for adopting Genomic Selection as a more integrative and predictive breeding framework.

Limitations of pre-genomic selection methodologies

Crop improvement has long relied on phenotype-based selection, which, while foundational, is inherently limited by its dependence on environmental variability, long breeding cycles, and the inability to precisely dissect genetic contributions to complex traits. The advent of molecular markers and quantitative genetics led to the development of marker-assisted selection (MAS), quantitative trait loci (QTL) mapping, and later genome-wide association studies (GWAS). These tools represented major advances in linking genotype to phenotype; however, they also revealed critical limitations that hindered their utility for complex, polygenic traits central to modern crop breeding. Marker-assisted selection (MAS) is effective for traits governed by one or a few major QTLs, such as disease resistance or simple quality attributes, but it performs poorly for polygenic traits where many loci each contribute a small effect. The need for prior identification of major QTLs limits its scalability, and the context-dependency of QTL effects often reduces transferability across populations and environments.

Similarly, QTL mapping relies on biparental populations with restricted genetic diversity, leading to low mapping resolution and poor applicability to broader germplasm. While GWAS expanded the scope to diverse populations, it too faces challenges: stringent multiple-testing corrections reduce detection power, and identified loci explain only a small fraction of total phenotypic variance, the so-called "missing heritability" problem (Huang and Han 2024). Fine mapping and candidate gene validation are resource-intensive, requiring extensive recombination and large populations, thereby slowing the breeding process. Collectively, these limitations constrain the ability of pre-genomic methods to capture the additive and non-additive genetic variance that drives complex trait expression. These challenges underscore the necessity for a paradigm shift toward genomic selection (GS), which utilizes genome-wide marker information to predict breeding values without prior locus identification. The emergence of GS thus represents a logical and scientifically grounded evolution from traditional and marker-based methods toward faster, data-driven, and more accurate crop improvement. Collectively, the limitations of conventional selection, marker-assisted selection (MAS), GWAS, and fine-mapping approaches emphasize their partial capacity to resolve the genetic basis of complex traits. The reliance on few major-effect QTLs, low transferability across environments, and inefficiency in polygenic trait prediction have constrained their long-term impact on breeding efficiency. These gaps have driven the evolution toward whole-genome prediction models that treat all available markers as contributors to genetic variance. In this context, Genomic Selection (GS) emerged as a comprehensive solution capable of integrating genome-wide marker effects to predict breeding values more accurately and efficiently. The following section discusses the conceptual framework and implementation strategies that define the operational principles of GS in plant breeding.

Evolution of genomic selection

Genomic selection (GS), proposed by Meuwissen et al. (2001), revolutionized breeding by predicting breeding values using genome-wide markers, minimizing dependence on extensive phenotyping. Initially applied in animal breeding, GS soon transformed plant breeding by addressing limitations of marker-assisted selection (MAS), such as its inefficacy for polygenic traits and long breeding cycles. Traits like yield, disease resistance, and drought tolerance governed by numerous small-effect loci benefit substantially from GS through early selection based on genomic potential rather than observable phenotypes (Lorenz et al. 2011; Banks 2022). GS relies on predictive models trained on populations with both genotypic and phenotypic data. Once marker effects are estimated, genomic estimated breeding values (GEBVs) of untested individuals can be accurately predicted, expediting the selection process (Meuwissen et al. 2001). Its efficiency has been enhanced by high-throughput genotyping technologies such as SNP arrays and next-generation sequencing (Thomson 2014), and by machine learning algorithms that improve predictive accuracy (Kumar et al. 2024).

The Genomic Selection framework represents a fundamental advancement in quantitative genetics, enabling early and accurate selection based on genomic estimated breeding values (GEBVs). By combining genome-wide markers, robust statistical models, and optimized training populations, GS enhances prediction accuracy and accelerates breeding cycles. However, the success of GS depends on appropriate model calibration, population structure, and data quality factors that directly affect prediction reliability. The next section evaluates the practical applications and real-world successes of GS across diverse crops, providing empirical evidence of its effectiveness in improving yield, stress tolerance, and overall breeding efficiency.

Framework of genomic selection: training and testing populations

Concept and workflow

Genomic selection integrates molecular and phenotypic data from a training population (TRN), which is both genotyped and phenotyped to train prediction models that estimate GEBVs for a testing population (TST), which is genotyped but not phenotyped (Torres et al. 2019). This framework drastically reduces the time, cost, and environmental dependency of traditional selection cycles, particularly for low-heritability traits (Fig. 1). Prediction accuracy depends on the relatedness between TRN and TST populations, marker density, and model robustness. Accurate phenotyping of the TRN is essential for reliable model training. High-throughput phenotyping platforms such as drones, imaging systems, and remote sensing capture multi-environment trait variation efficiently (Dhanya et al. 2024; Kumar et al. 2025). TRNs designed with broad phenotypic diversity and structured

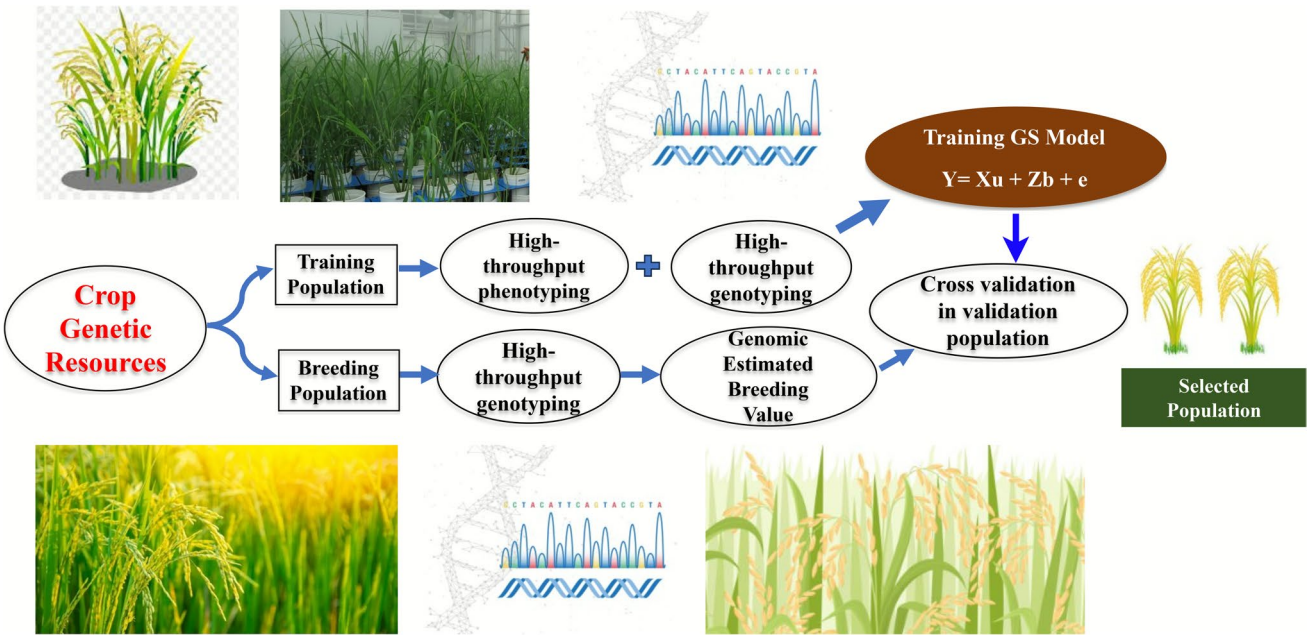

Fig. 1 Basic flow diagram of genomic selection in crop plants

sampling improve model generalizability. Genotyping is performed using dense SNP arrays or sequencing platforms to capture genome-wide variation. Statistical models (GBLUP, Bayesian, ML) estimate marker effects, which are then used to predict GEBVs in the TST. Multi-trait and multi-environment models further enhance prediction accuracy (Burgueño et al. 2012; Gaynor et al. 2017).

Phenotyping and genotyping

Phenotyping provides the baseline data for GS model calibration. Precision phenotyping through remote sensing and multi-environment trials captures genotype × environment (G × E) interactions and reduces error variance (Araus and Cairns 2014; Jarquín et al. 2014). Stratified sampling of phenotypic extremes further strengthens model accuracy (Isidro et al. 2015). Genotyping supplies the molecular data necessary for estimating marker effects. High-density SNP arrays and next-generation sequencing (NGS) deliver genome-wide coverage, enabling the detection of both major and minor QTLs (Rafalski 2002; Norman et al. 2018). Cost-effective approaches like low-density genotyping combined with imputation preserve accuracy while reducing expenses (Pryce et al. 2014). High-density SNP datasets improve prediction of polygenic traits and rare alleles, with applications across major crops such as wheat, maize, and rice (Goddard et al. 2011; Bassi et al. 2016).

Statistical models in genomic selection

The success of GS depends on the predictive accuracy of statistical models trained on the TRN. A major challenge is the "$p > n$" condition, where the number of markers far exceeds the number of individuals, risking overfitting. Models such as Genomic BLUP (GBLUP), ridge regression, and Bayesian regressions address this through shrinkage and regularization (Jannink et al. 2010). GBLUP extends the traditional BLUP framework using a genomic relationship matrix to model additive genetic effects efficiently (Meuwissen et al. 2001), while Bayesian models (BayesA, BayesB, BayesCπ) flexibly model heterogeneous marker effects and sparsity (Zhu et al. 2021). Emerging machine learning (ML) approaches, including random forests, support vector machines, and deep neural networks capture nonlinear relationships and epistatic interactions beyond the capacity of linear models (Montesinos-López et al. 2021). Mixed linear models (MLMs) incorporating G × E interactions and population structure corrections further improve prediction robustness (Crossa et al. 2017a, b; Jarquín et al. 2014). Model choice should align with trait architecture, dataset size, and genetic complexity. Linear models remain effective for highly polygenic traits, while nonlinear and hybrid (deep + GBLUP) frameworks are advantageous when omics or environmental data are integrated (Wang et al. 2025; Montesinos-López et al. 2025).

Statistical models for genomic selection (GS): equations, worked setup, and pipeline flowcharts

BLUP and genomic BLUP (GBLUP) The model of the BLUP and GBLUP represented below:

$$\begin{aligned} \mathrm{y} &= \mathrm{Xb} + \mathrm{Zu} + \mathrm{e} \\ \mathrm{u} &\sim \mathcal{N}\left(0, \mathrm{K}\sigma_g^2\right) \\ \mathrm{e} &\sim \mathcal{N}\left(0, \mathrm{I}\sigma_e^2\right) \end{aligned}$$

y: phenotype vector; b: fixed effects (trial, block, covariates); u: additive genomic breeding values; K: genomic relationship matrix (GRM; e.g., $\mathrm{K} = \frac{1}{m}\mathrm{WW}^\top$ with centered/standardized genotype matrix W); σ_g^2, σ_e^2: variance components estimated by REML. BLUP uses pedigree-based A; GBLUP replaces A with K to exploit genome-wide SNPs. Predictive use: Posterior mean of u is given by: $\hat{\mathrm{u}} = \sigma_g^2 \mathrm{KZ}^\top \left(\mathrm{Z}\sigma_g^2 \mathrm{KZ}^\top + \mathrm{R}\right)^{-1} (\mathrm{y} - \mathrm{X}\hat{\mathrm{b}})$.

Ridge/Bayesian ridge regression (BRR) a.k.a. RR-BLUP The Marker regression form is also given as:

$$\begin{aligned} \mathrm{y} &= 1\mu + \mathrm{M}\beta + \varepsilon \\ \beta &\sim \mathcal{N}\left(0, \tau^2 \mathrm{I}\right) \\ \varepsilon &\sim \mathcal{N}(0, \sigma^2 \mathrm{I}) \end{aligned}$$

Equivalent to RR-BLUP; M is standardized marker matrix. Closed-form ridge solution or Bayesian Gibbs sampling for $(\mu, \beta, \tau^2, \sigma^2)$.

Bayesian whole-genome regressions

a) BayesA: All SNPs have effects; marker-specific variances $\beta_j \sim \mathcal{N}(0, \sigma_j^2)$, $\sigma_j^2 \sim$ Scaled-Inv-χ^2.
b) BayesB: Spike-and-slab; a proportion $1 - \pi$ of SNPs have zero effect; others as in BayesA.
c) BayesC π/BayesR: Common variance per included SNP (BayesC) or finite mixture of normal components (BayesR) for flexible architectures.

Kernel/RKHS and non-additive modelling The expression of the RKHS regression model:

$$\begin{aligned} \mathrm{y} &= 1\mu + \mathrm{K}_\kappa \alpha + \varepsilon, \\ \alpha &\sim \mathcal{N}\left(0, \sigma_\alpha^2 \mathrm{I}\right) \end{aligned}$$

with kernel $\kappa(w_i, w_j)$, e.g., Gaussian or arc-cosine, to approximate dominance/epistasis via nonlinear similarity.

Machine learning (ML) and deep Learning (DL) for GS Support vector regression (SVR) criterion:

$$\min_{w,b} \frac{1}{2} \| w\|^2 + C \sum_i \xi_i \text{s.t.}$$
$$| y_i - \left(w^\top \phi(m_i) + b\right) | \leq \epsilon + \xi_i, \xi_i \geq 0$$

Kernel trick $\kappa(m_i, m_j) = \langle \phi(m_i), \phi(m_j) \rangle$.
Gradient Boosting/XGBoost: Additive trees f_t minimizing

$$\mathcal{L}^{(t)} = \sum_i \ell \left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i) \right) + \Omega(f_t)$$

with shrinkage and column/row subsampling-useful for heterogeneous, nonlinear signals.

Neural networks: Multi-layer perceptrons for SNPs; CNNs for image phenotypes; multimodal nets concatenating genomic (M), phenomic (UAV indices, hyperspectral), and meta-environment features.

End-to-end pipelines A flowchart illustrating the complete end-to-end pipeline for various genomic selection (GS) models is presented in Fig. 2, depicting the workflows for GBLUP/RR-BLUP, Bayesian/RKHS, and ML/DL integrated with multi-omics and phenomics data.

(A) Classical GS (GBLUP/RR-BLUP) pipeline
Raw genotypes → QC (MAF, missingness, Hardy–Weinberg) → Imputation
→ Standardize markers (0/1/2 → centered/scaled)
→ Build GRM K (VanRaden)
→ Model fit (GBLUP/RR-BLUP; REML)
→ Cross-validation (CV-kfold; CV1/CV2; LOEO for G × E)
→ Predict GEBV for TST → Selection & mating design
(B) Bayesian/RKHS pipeline
Same QC/imputation → Choose prior/kernel (BayesB, BayesR, RKHS)
→ Hyperparameter tuning (π in BayesB; bandwidth σ in Gaussian kernel)
→ MCMC or EM → Posterior means of SNP effects or genetic values
→ CV & independent validation → GEBV
(C) ML/DL + multi-omics/phenomics pipeline
Genotypes (SNP/haps)
→ Feature scaling / PCs / LD-pruning (optional)
Omics (RNA/protein/metabolites) → Normalization / batch correction
Phenomics (UAV RGB/HSI) → CNN/feature extractor → latent features
Envirotyping (weather/soil) → engineered covariates
↓
Multimodal fusion (concat/attention)
↓
Model (XGBoost / MLP / CNN / deepGBLUP hybrid)
CV (nested; LOEO for new environment) → Predictions

Model selection and rationale

Models deployed in genomic selection have taken a huge leap in the past decade, with numerous models being proposed to enhance accuracy at computational cost. However, the fundamental model used was simple linear regression, where the phenotype was regressed on marker alleles. The simplest and commonly used model is GBLUP, based on which several advanced models have evolved (Wang et al. 2025). Regularisation was incorporated in the modelling to counter overfitting, resulting in increased accuracy in the breeding population. Thereafter, advanced models, such as Bayesian alphabets and machine learning algorithms, were developed to capture higher-order gene effects (Montesinos-López et al. 2021). To summarise, model accuracy depends on the kind of data in the training and breeding population. More specifically, kind of populations used, the relationship between the populations, the number of genes affecting the trait the nature of gene action and the number of markers and population size, etc., among other minor factors. For instance, when the training and breeding populations are not related, then Bayesian or machine learning models perform better as they do not depend on the relationship among genotypes as in the case of GBLUP (Li et al. 2023). If the number of genes governing the trait is large, with minor effects, GBLUP is reported to outperform others, but if the trait is oligogenic, like flowering time, resistance to diseases and pests, where a few major effect genes are responsible, then machine learning and Bayesian methods are considered better (Meher et al. 2022). Moreover, advanced models tend to capture non-additive effects, yielding better accuracy than GBLUP, which captures additive effects (Pathy 2024). Therefore, if the gene action is unknown, using multiple models is an effective option to estimate marker effects accurately.

Population size, marker density, and model validation

Population size and marker density jointly determine GS performance. Larger training populations improve marker effect estimation and prediction accuracy, though gains

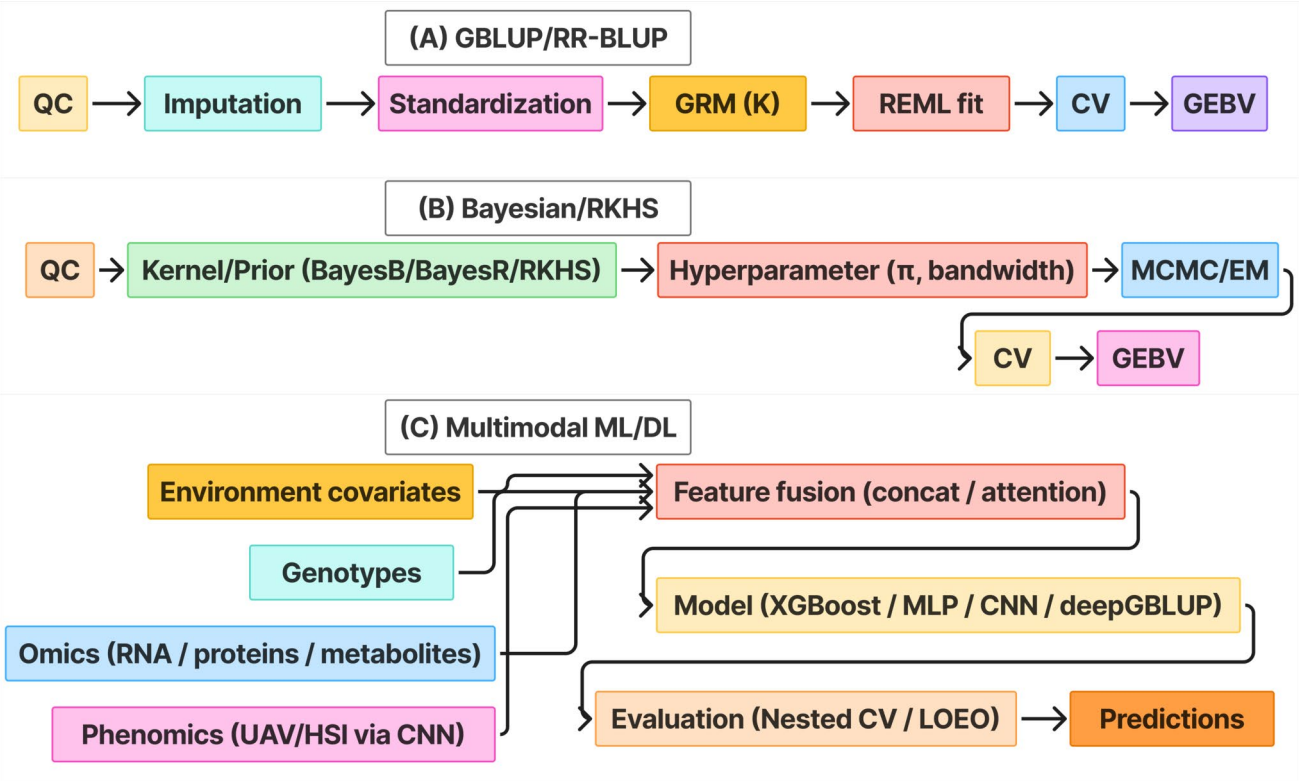

Fig. 2 Data processing pipelines for GS models **A** GBLUP/RR-BLUP **B** Bayesian/RKHS pipeline **C** ML/DL + multi-omics/phenomics pipeline

plateau beyond moderate sizes (Daetwyler et al. 2013). High marker density enhances linkage disequilibrium coverage and predictive resolution but must be balanced against cost. Imputation and optimized sampling strategies effectively maintain accuracy in resource-limited programs (Habier et al. 2007; Li et al. 2023). Validation ensures the reliability of GS models in independent populations or environments. Cross-validation (k-fold, leave-one-out, or stratified CV) is widely used to assess performance and prevent overfitting (Meuwissen et al. 2001; Heslot et al. 2015). Independent validation datasets provide unbiased accuracy estimates, typically measured by correlation (r) or coefficient of determination (R^2) between predicted and observed values (Crossa et al. 2010). Continuous model updating incorporating new genotypic, phenotypic, and environmental data improves prediction stability across breeding cycles (Jarquín et al. 2014; Hickey et al. 2017). Modern ML-based updating methods allow dynamic integration of new data streams, enhancing long-term model adaptability (Montesinos-López et al. 2018).

The body of evidence summarized in this section demonstrates the versatility and efficacy of GS across cereals, pulses, oilseeds, fibre, sugar, fruit, and vegetable crops. Through improved prediction accuracies, GS has substantially enhanced selection efficiency for both simple and complex traits under various environmental conditions. The observed variability in prediction accuracy among studies underscores the importance of optimizing population structure, marker density, and heritability estimation. Collectively, these findings affirm that GS is not only a statistically powerful tool but also a practical component of modern breeding pipelines. Building on these successes, the next section explores how GS can be integrated with emerging technologies such as multi-omics, high-throughput phenotyping, and AI-driven modelling—to further accelerate genetic gain.

Genomic selection as a strategy to accelerate genetic gain

GS accelerates crop improvement by predicting genetic potential early, thereby shortening breeding cycles and enhancing genetic gain (Meuwissen et al. 2001). Unlike phenotypic selection or MAS, GS captures genome-wide additive effects, allowing accurate prediction of polygenic traits such as yield, drought tolerance, and disease resistance (Crossa et al. 2017a, b; Heffner et al. 2009a, b). The

magnitude of genetic gain depends on selection accuracy, selection intensity, genetic variance, and cycle time. By reducing phenotyping requirements, GS enables recurrent selection and rapid-cycle breeding, thereby increasing annual genetic gain (Das et al. 2020; Allier et al. 2019). However, maintaining genetic diversity is critical to prevent inbreeding depression and preserve long-term adaptability (Wray and Thompson 1990; Goddard et al. 2010). Additive GBLUP models remain the standard in GS, but incorporating non-additive genetic variance i.e., dominance and epistasis, can enhance predictions, particularly in hybrid or cross-pollinated crops. Extended models such as GBLUP+D and kernel-based methods (e.g., RKHS) efficiently capture these effects (de los Campos et al. 2013). Hybrid architectures like deepGBLUP combine linear additive modelling with nonlinear feature extraction to improve accuracy across diverse datasets (Li et al. 2023). Ultimately, the choice of model should reflect trait architecture and breeding objectives: additive models for self-pollinated crops and extended dominance/epistatic models for hybrid systems. Regardless of framework, rigorous validation across environments ensures robustness and generalizability. Collectively, GS reduces breeding cycles by up to 50% and enhances genetic gain by 20–45% compared with conventional approaches (Lorenz et al. 2011). Its integration into rapid-cycle and multi-environment genomic prediction pipelines marks a paradigm shift toward data-driven, predictive crop improvement.

Overall, integrated genomic selection represents a transformative breeding paradigm that unites genomic prediction with high-throughput genotyping, phenomics, machine learning, and genome editing. The convergence of these technologies facilitates data-driven selection and climate-resilient crop improvement at unprecedented speed and precision. However, realizing this potential requires overcoming challenges related to data integration, computational scalability, and cross-environmental validation. The next section introduces the workflow and computational strategies required to incorporate multi-omics data into GS pipelines, outlining best practices and critical bottlenecks in real-world implementation.

Overview of successful genomic selection strategies across crop species

Table 1 summarizes key advances in the application of genomic selection (GS) across major crop groups, including cereals, pulses, oilseeds, sugar, fibre, fruit, and vegetable crops. These studies collectively demonstrate the wide applicability and growing precision of GS for improving complex traits such as yield, quality, and resistance to biotic and abiotic stresses. Reported prediction accuracies range from moderate to high, depending on population type, trait complexity, and model used, spanning conventional RR-BLUP and GBLUP to advanced Bayesian, kernel-based, and machine learning algorithms. This compilation highlights GS as a robust framework to accelerate genetic gain and enhance selection efficiency in modern breeding programs.

Prediction accuracy in GS is primarily influenced by population structure, marker density, and trait heritability. Generally, higher heritability yields greater accuracy since phenotypic values more accurately reflect genetic effects (Bernardo 2016). For low-heritability traits, increasing the training population size enhances prediction power. Multi-trait genomic selection (MTGS) models further improve accuracy by exploiting information from correlated, highly heritable traits (Pathy 2024). A fundamental assumption of GS is that at least one marker is in linkage disequilibrium (LD) with the QTL controlling the trait (Krishnappa et al. 2021). Accordingly, as marker density increases, prediction accuracy initially improves but eventually plateaus due to marker redundancy. Norman et al. (2018) reported a sharp increase in prediction accuracy up to ~ 5,000 markers, followed by diminishing returns. Self-pollinated crops generally require fewer markers than cross-pollinated species, where LD decays faster. Optimizing marker number not only enhances prediction efficiency but also minimizes genotyping costs (Alemu et al. 2024). Prediction accuracy also depends on the genetic relationship between training and test populations. Greater relatedness improves prediction due to shared alleles and similar LD patterns (Senthilkumar et al. 2022). Conversely, high population structure manifested as variable allelic frequencies can bias prediction accuracy (Sidhu et al. 2025). Expanding training population size mitigates these effects, particularly when training and test sets are less related (Norman et al. 2018). Traits under strong selection pressure similarly benefit from incorporating multiple subpopulations to increase allelic diversity.

In summary, the successful integration of multi-omics data into GS frameworks demands a balanced design encompassing biological relevance, computational efficiency, and data standardization. Harmonizing genomics with transcriptomics, proteomics, and metabolomics enables a systems-level understanding of genotype–phenotype relationships, enhancing predictive power for complex traits. Despite these advancements, technical challenges, such as dimensionality reduction, batch effects, and imputation must be carefully managed to maintain model robustness. The next section critically discusses the current bottlenecks and practical constraints that hinder large-scale GS adoption, including genomic resource gaps, computational infrastructure, and cost barriers.

Table 1 Notable achievements of genomic selection in cereal, pulse, oilseed, sugar, fibre, fruit, and vegetable crops

Crop	Trait studied	Population type	GS model	Prediction accuracy (r)	Reference
Cereals					
Rice	Panicle weight and Yield	Inter-related synthetic	Bayesian LASSO	0.30 and 0.32	Grenier et al. (2015)
	Field grain and variance of field grain	128 Japanese cultivars	GBLUP	0.30 and 0.53	Yabe et al. (2018)
	Grain yield and thousand grain weight	NCD II	GBLUP, SVM, LASSO, and PLS	0.5 and 0.28	Xu et al. (2018)
	Blast resistance	Rice lines	RRBLUP and GBLUP	0.17 to 0.73	Huang et al. (2019)
	Drought tolerance	Germplasm	GBLUP and RKHS	0.22 to 0.80	Bhandari et al. (2019)
	Weight of panicle and Nitrogen balance index	Breeding lines	GBLUP, RKHS, and Bayes B	0.30 and 0.21	Hassen et al. (2018)
Wheat	Grain yield	Germplasm and Advanced breeding lines	RRBLUP, RKHS, and Bayesian LASSO	0.49 to 0.61	Crossa et al. (2010)
	Grain yield	Breeding lines	RRBLUP, Bayes A, Bayes B, Bayes C, LASSO, NN, and RKHS	0.6 to 0.7	Pérez-Rodríguez et al (2012)
	Leaf, stem and yellow rust resistance	Germplasm	GBLUP and BRR	0.35, 0.27, and 0.44	Daetwyler et al. (2014)
	Fusarium head blight and Deoxynivalenol resistance	Breeding lines	RR, RKHS and RF	0.463 and 0.575	Rutkoski et al. (2012)
	Fusarium head blight and Septoria leaf blotch	Winter wheat breeding lines	Bayesian LASSO and multiple linear regression	0.6 and 0.5	Mirdita et al. (2015)
	Grain yield	F_1 population	RRBLUP, BayesA, BayesB, and BayesC	0.3 to 0.6	Zhao et al. (2017)
Maize	Grain yield	F_1 population from half diallel and test crosses	GBLUP	0.58	Zhao et al. (2012)
	Grain yield	test crosses	Bayes A, Bayes B, Bayes C, LASSO, and RKHS	0.78	Rio et al. (2019
	Striga resistance and drought tolerance	Inbred and test cross lines	BLUP	0.58 and 0.65	Badu-Apraku et al. (2019)
	Anthesis–silking interval (ASI)	Asian and African inbreds	BLUP	0.35–0.43	Vivek et al. (2017)
	Water-logging tolerance	Inbreds and half diallel lines	RRBLUP and GBLUP	0.53 to 0.84	Das et al. (2020)
Barley	Grain yield, fusarium head blight and Deoxynivalenol (DON) resistance	Barley Breeding lines	RRBLUP	0.57, 0.74 and 0.72	Sallam et al. (2015)
	Thousand kernel weight	Barley Breeding lines	GBLUP and RKHS	0.67	Abed et al. (2018
Oat	Grain yield, winter hardiness, Ear emergence, mildew	F_2 mapping population of 194 cross winter oat varieties 'Buffalo' and 'Tardis'	RR-BLUP, DiPR	0.63–0.69, 0.43–0.48, 0.24–0.29, 0.49–0.52	Mellers et al. (2020)
Pulses					
Chickpea	Drought tolerance	Breeding lines	RRBLUP, Bayesian LASSO and BRR	0.56 to 0.61	Li et al. (2018)
	Seed yield, hundred seed weight, days to 50% flowering and days to maturity	320 elite breeding lines	RR-BLUP, Kinship GAUSS, Bayes Cπ, Bayes B, Bayesian LASSO) and RF	0.52, 0.89, 0.71 and 0.81	Roorkiwal et al. (2016)
	Ascochyta blight resistance	2790 chickpea lines	GBLUP	0.40–0.90	Lin et al. (2025)
Lentil	Crop duration	RILs and diversity panel	RRBLUP GBLUP Bayes A Bayes B Bayes Cπ Bayesian LASSO BRR and RKHS	0.58–0.84	Haile et al. (2020)
	Grain yield, ascochyta blight resistance, botrytis grey mould resistance, salinity and boron stress tolerance, 100-grain weight, seed size index and protein content	2,081 breeding lines	BayesR	0.34, 0.49, 0.63, 0.47, 0.82, 0.82	Gebremedhin et al. (2024)

Table 1 (continued)

Crop	Trait studied	Population type	GS model	Prediction accuracy (r)	Reference
Common bean	Cooking time	RIL, multi-parent advanced generation inter-cross (MAGIC), germplasm	GBLUP, Bayes A, Bayes B, Bayes C, Bayesian LASSO, and BRR	0.22 to 0.55	Diaz et al. (2021)
	Root rot resistance	Breeding lines	RKHS	0.52	Diaz et al. (2021)
Oilseeds					
Groundnut	Yield Protein Rust resistance Late leaf spot resistance	Breeding lines	Bayesian generalized linear regression	0.49 to 0.60, 0.41 to 0.46, 0.74–0.75, and 0.57–0.65	Pandey et al. (2020
	Aflatoxin	Recombinant inbred line (RIL) populations derived from 3 different crosses	RRBLUP	0.25	Gimode et al. (2024)
Brassica napus	Seed yield Oil content Lodging resistance	F_1 population	RRBLUP	0.45, 0.81, and 0.39	Jan et al. (2016)
	Days to 50% flowering, Days to maturity, Plant height, 1000 kernel weight	Nested Association Mapping population (2,572 RILs)	rrBLUP and RKHS	0.71, 0.46, 0.63, 0.40	Perumal et al. (2024)
Sunflower	Oil content	F_1 population factorial mating	LASSO and multi-kernel BLUP	0.78	Mangin et al. (2017
Soybean	Yield and Oil, Protein, content	RILs	RBLUP	0.68, 0.76 and 0.76	Beche et al. (2021)
	Oil and Protein content	Breeding lines	Bayes B and Bayesian LASSO	0.30 and 0.50	Stewart-Brown et al. (2019)
	Seed yield	Panel of 250 soybean genotypes	RRBLUP fused with RBF, SVR, and RF	7% increase over normal	Yoosefzadeh-Najafabadi et al. (2022)
	Seed yield	Advanced breeding lines	BayesB, BayesRR, RRBLUP, et al. EGBLUP	0.40	Miller et al. (2023)
Sugar crops					
Sugarcane	Tonnes of cane per hectare (TCH), Commercial cane sugar (CCS), Fiber content	1,318 clones from Final Assessment Trials (FAT)	GBLUP, RKHS (Gaussian kernel), Arc-cosine kernel (AK4), BayesR, among others	0.5 to 0.7	Yadav et al. (2020);
Sugar beet	White sugar yield, sugar content, root yield, sodium content, potassium content, α-amino nitrogen content	924 inbred lines comprising 676 diverse lines and 248 lines from four biparental families	RR-BLUP	0.48 to 0.72	Würschum et al. (2013)
Fiber crops					
Cotton	Fiber quality traits (length, strength, elongation, micronaire, uniformity, short fiber index)	550 recombinant inbred lines from a MAGIC population	BayesB, Bayesian LASSO, RKHS	0.50 to 0.70	Islam et al. (2020)
	Fiber quality and yield traits	1,385 breeding lines	Bayesian regression methods	0.40 to 0.70	Li et al. (2022)
Vegetables					
Tomato	Fruit weight, Firmness, Soluble solids, Sugar content, Acidity and Biochemical profile	Germplasm	RRBLUP	0.814, 0.614, 0.714, 0.649, 0.619 and 0.126–0.705	Duangjit et al. (2016)
	Soluble solids content and total fruit weight	96 big-fruited F_1 lines	GBLUP, Bayesian Lasso, wBSR, Bayes C, RKHS, RF	0.43 to 0.59	Yamamoto et al. (2017)
	Yield	Breeding lines	rrBLUP and SVM	0.4 to 0.6	Bhandari et al. (2023)
Capsicum	Fruit length Fruit width Fruit shape and Fruit weight	Core collection and RIL population	GBLUP, Bayesian LASSO, Bayes B, Bayes C, and RKHS	0.32, 0.50, 0.34 and 0.48	Hong et al. (2020)
Onion	Simulated quantitative trait (additive QTLs)	F_2 population derived from a single F_1 plant	RRBLUP	Varied across cycles	Sekine and Yabe (2020)
Carrot	Root length, maximum width, aspect ratio, root fill	Diverse germplasm panel of 662 accessions	GBLUP	0.67	Brainard et al. (2022)
Spinach	White rust resistance	346 USDA-GRIN germplasm	rrBLUP, gBLUP, cBLUP, Bayesian methods, RF, SVM	Up to 0.84	Shi et al. (2022)

Table 1 (continued)

Crop	Trait studied	Population type	GS model	Prediction accuracy (r)	Reference
Cauliflower	Curd width, cluster width, number of branches, apical shoot length, nearest branch length, days to budding	174 gene bank accessions	GBLUP, RRBLUP, BayesB	0.09 to 0.66	Thorwarth et al. (2018)
Pumpkin	Fruit weight (FWT), Growth vigor (GV), Leaf size (LS)	320 parental lines; 119–120 hybrids in C. maxima, 89–111 hybrids in C. moschata	Bayesian methods (Bayes A, B, C, TA, TB, TC), Linear Mixed Models (LMM)	0.71 to 0.91	Wu et al. (2019)
Fruit crops					
Apple	Firmness and Soluble solids	F_1 population factorial mating	RRBLUP and Bayesian LASSO	0.83 and 0.89	Kumar et al. (2012)
	10 key traits of quality	Full-sib (FS) families	Bayesian	0.19 to 0.5	Muranty et al. (2015)
	Firmness	Germplasm and biparental families	RRBLUP	0.81	Roth et al. (2020)
Citrus	10 key traits of quality	180 individual hybrid trees	BLUP	0.53 to 0.64	Gois et al. (2016)
	Fruit weight, Sugar content and Acid content	F_1s	GBLUP	0.650, 0.519, and 0.666	Imai et al. (2019)
	Fruit weight distribution	Varieties and their full sib families	GBLUP	0.89	Minamikawa et al. (2017)
Grapevine	15 traits related to yield, berry composition, phenology, and vigor	Diversity panel of 277 genotypes and a half-diallel of 622 genotypes	GBLUP	0.6	Brault et al. (2022)
Apricot	Glucose and Ethylene content	F_1 pseudo-testcross population	GBLUP, Bayes A, Bayes B, Bayes C, Bayesian LASSO, and BRR	0.31 and 0.78	Nsibi et al. (2020)
Pear	Crispness and Sweetness	Full sib families	GBLUP	0.32 and 0.62	Kumar et al. (2019)
Guava	Soluble solids content (SSC), fruit mass (FM), pulp mass (PM), number of fruits per plant (NF), production per plant (PROD)	Guava population	Bayesian Ridge Regression	SSC: 0.65; FM: 0.88; PM: 0.97; NF: 0.69; PROD: 0.51	da Silva et al. (2021)
Peach	Fruit Weight (FW), Soluble Solids Content (SSC), Titratable Acidity (TA)	1,147 individuals from 11 biparental European peach progenies	GBLUP	FW: up to 0.84; SSC: up to 0.78; TA: up to 0.83	Biscarini et al. (2017)
Banana	15 traits related fruit quality	307 genotypes	BayesB, BayesC, BRR, RKHS, RKHS_PM	0.47–0.75 for fruit filling and fruit bunch traits	Nyine et al. (2018)
Blueberry	8 traits related to fruit quality	1,847 autotetraploid individuals	GBLUP with various relationship matrices	0.20 to 0.49 across traits and models	de Bem Oliveira et al. (2019)
Strawberry	Resistance to gray mold	Breeding populations	GBLUP	LD: 0.28–0.59; EM: 0.37–0.47	Petrasch et al. (2022)
Kiwifruit	Red intensity, fruit weight, yield, fruit firmness	88 diploid individuals	RR-BLUP, GBLUP, LASSO	0.48–0.91	Cheng et al. (2019)
Sour passion fruit	Economically important traits	Population genotyped with 183 polymorphic markers	RR-BLUP, Bayes A, Bayes B, Bayes B2, Bayes Cπ, and Bayesian Lasso	0.55–0.75	Viana et al. (2017)

Integrated genomic selection: a framework for accelerated genetic gain

Figure 3 illustrates a comprehensive framework integrating genomic selection with emerging genomic, phenomic, and computational technologies. At its core, GS leverages genome-wide markers to predict breeding values, enabling early, data-driven selection decisions and reducing breeding cycle duration. A broad and diverse germplasm base forms the foundation of this approach, ensuring wide allelic representation and enhancing GS model accuracy (Salgotra and Chauhan 2023). Recent innovations such as pangenomes and super-pangenomes capture full species-level genetic diversity, including rare alleles and structural variants (Li et al. 2023; Jayakodi et al. 2024; He et al. 2025). High-throughput genotyping (e.g., SNP arrays, GBS, KASP assays) and phenotyping systems (e.g., UAV- and sensor-based imaging) provide rapid, precise, and large-scale data acquisition (D'Agostino and Tripodi 2017; You et al. 2018).

An overview of software tools and computational platforms for multi-omics data integration with applications in genomic prediction is depicted in Table 2, and the common data formats used for integrating multi-omics datasets into genomic selection pipelines are listed in Table 3. Integration of phenomics with machine learning has expanded GS potential. Image- and sensor-derived phenotypes (e.g., vegetation indices, hyperspectral features) can serve as secondary traits in GS models, improving prediction of complex, low-heritability traits (Montesinos-López et al. 2021). Advanced modelling of genotype × environment interactions (GEI) using approaches such as GGE biplots, factor analytic models, environmental covariates, and reaction norm frameworks further refines prediction across environments (Jarquín et al. 2014; Nguyen et al. 2025).

Accelerated breeding strategies including doubled haploid (DH) production and speed breeding complement GS by reducing generation time and rapidly fixing desirable alleles (Germana 2011; Watson et al. 2018; Pandey et al. 2022). Multi-omics integration combining genomic, transcriptomic, proteomic, and metabolomic data adds biological depth to GS models, capturing regulatory and metabolic components of trait expression (Vazquez et al. 2016; Yang et al. 2021; Nunavath et al. 2025b,c). Machine learning and Bayesian frameworks facilitate this integration, improving accuracy for complex traits (Jia and Jannink 2012). Artificial intelligence (AI) and deep learning (DL) are emerging as powerful tools for modelling non-linear relationships and multi-dimensional datasets (Vinayaka and Prasad 2024; Vinayaka et al. 2025). While linear models such as GBLUP remain competitive for highly polygenic traits and

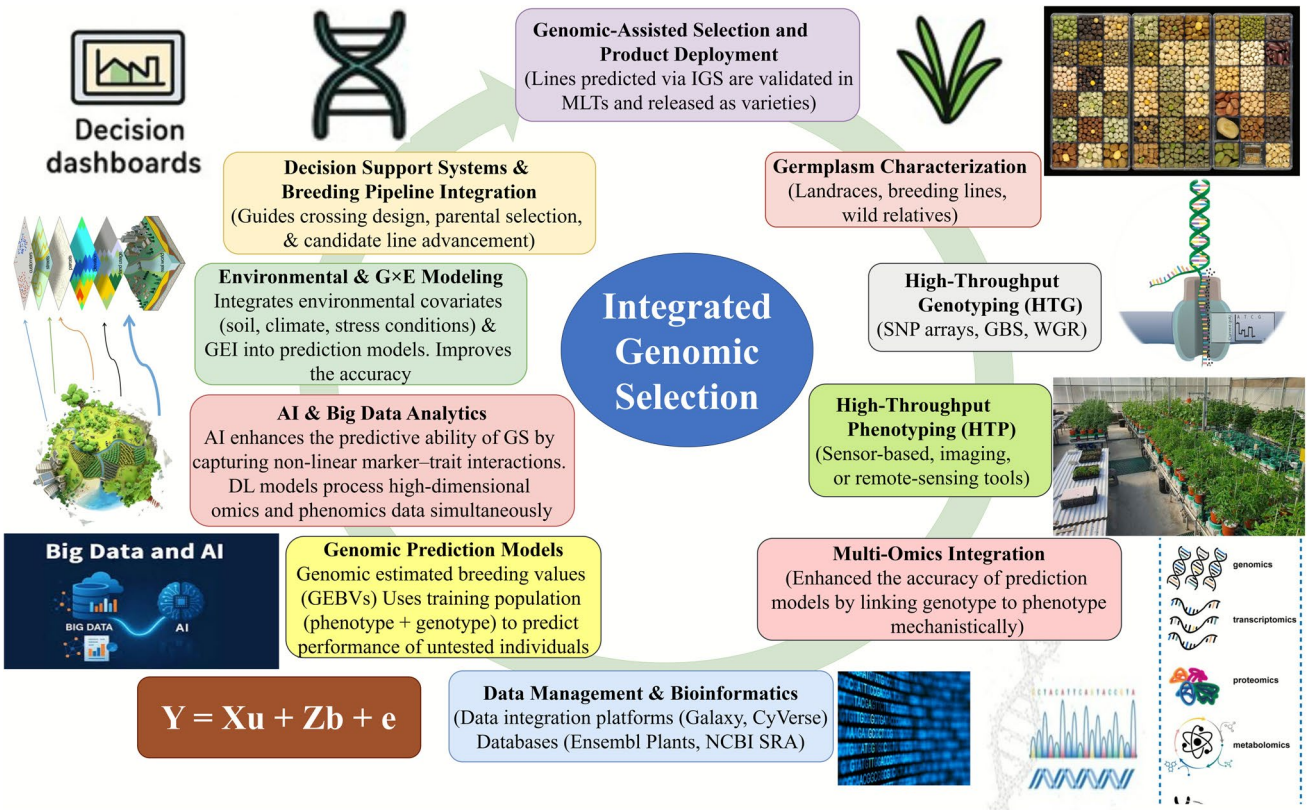

Fig. 3 Conceptual framework of integrated genomic selection for accelerating genetic gain in crops

Table 2 Overview of software tools and computational platforms for multi-omics data integration with applications in genomic prediction

Platform/Tool	Key features for multi-omics integration	Strengths & limitations in GS context	Reference
MiBiOmics	Interactive web app for exploring up to three omics layers; network inference (e.g. WGCNA-based), feature extraction, visualization	Good for exploratory phases; less designed for breeding prediction or handling very large marker datasets	Zoppi et al. 2021
BiomiX	Focus on "middle integration" using MOFA (Multi-Omics Factor Analysis); supports standard omics matrices; helps in interpretability and handles missing data; FAIR compliance	Useful for combining data and reducing dimensionality; in GS, may need coupling with prediction models (GBLUP, Bayesian, etc.)	Iperi et al. 2025
NetGP	Deep learning model combining transcriptomic + SNP (genomic) data; includes feature extraction via Pearson-Collinearity Selection; shows improved predictive performance in several plants	Good example of how integrating transcriptomics with genomics can increase accuracy; but requires matched omics data and substantial computational resources	Zhao et al. 2025
MLLASSO	Iterative integration: first predict transcript/metabolite layers from genetic markers, then combine genetically predictable transcripts (or metabolites) into prediction of phenotype	Demonstrates a hierarchical/layered strategy; may reduce noise by filtering out omic features poorly predictable by genotype	Hu et al. 2019
Genomic prediction powered by multi-omics data	Evaluation of many integration strategies (early fusion, model-based, concatenation) across datasets; shows that model-based or sophisticated fusion often outperform simple concatenation	Good benchmarking; suggests which integration modes are promising; but requires well-matched datasets and cross-validation	Montesinos-López et al. 2025
Joint modelling	Joint modelling of genomic, transcriptomic, and methylomic (epigenomic) data to predict complex traits in *Arabidopsis*	Very useful for breeding programs with genotypic + phenotypic data; highlights gap that many GS pipelines do not yet include transcript, metabolite, etc	Wang et al. 2025

small datasets, DL and ensemble models often outperform them under complex, cross-environment, or multimodal conditions (Montesinos-López et al. 2018). Hybrid models such as deepGBLUP combine the interpretability of linear models with the adaptability of neural networks (Li et al. 2023). Complementary technologies like genome editing (CRISPR/Cas9, TALENs, ZFNs) and transgenic methods directly modify or introduce desirable alleles, providing functional validation and rapid trait improvement (Shan et al. 2013; Amaresh et al. 2025a,b); Wang et al. 2025). Collaborative, open-access GS networks that share genotyping, phenotyping, and predictive resources can significantly lower implementation costs and enhance global adoption (Krishnappa et al. 2021). Collectively, this integrated GS paradigm represents a data-driven, accelerated pathway for developing climate-resilient, high-yielding, and nutritionally enhanced crops—redefining the future of precision breeding.

Practical implementation pathways for developing regions

While the integrated genomic selection (GS) framework proposed in this study presents a technologically advanced vision for accelerating crop improvement, its immediate implementation in developing regions remains constrained by economic, infrastructural, and human-capacity limitations. High genotyping costs, limited computational facilities, and the shortage of skilled bioinformatic personnel often hinder large-scale deployment of advanced prediction models. Recognizing these realities, this framework is re-envisioned as a modular and scalable system, adaptable to varying levels of resource availability.

At the foundational level, low-cost genotyping strategies, such as targeted SNP arrays, genotyping-by-sequencing (GBS), or low-density panels combined with statistical imputation, can effectively capture essential genomic variation at reduced expense. Similarly, low-throughput but informative phenotyping approaches, including digital imaging and field-based physiological assessments, can support initial GS implementation in regional breeding stations. Open-source analytical tools like rrBLUP, BGLR, and GAPIT offer cost-effective computational alternatives to proprietary platforms, enabling accurate model development even on standard computing systems.

To strengthen technical capacity, collaborative networks and shared data infrastructures, for example, regional genotyping hubs or cloud-based phenomic repositories can

Table 3 Common data formats used for integrating multi-omics datasets into genomic selection pipelines

Omics type	Typical raw/Processed data formats	Key transformations to make them usable in GS	Reference
Genomics	Raw reads (FASTQ); variants (VCF, BCF, binary formats like PLINK.bed/.bim/.fam, pgen, etc.); imputed genotype matrices; SNP-chips outputs	QC (filtering by MAF, missingness, HWE), imputation, conversion to relationship matrices (G matrix), standardization (e.g. centering/scaling). Needed to align sample IDs, ensure correspondence with phenotype/transcriptome/metabolome samples	Subramanian et al. 2020
Transcriptomics	Raw reads (FASTQ), aligned reads (BAM/SAM), counts (e.g. raw counts, FPKM/TPM/RPKM), expression matrices (CSV/TSV/tabular formats), microarray data (.CEL,.CEL/.txt) etc	Normalization (library size/batch effect/technical confounders), filtering low-expression genes; possibly transforming (log/variance stabilizing); mapping gene IDs to reference; matching to genotype samples; in layered strategies, screening for 'genetically predictable' transcripts	Dávila-Fernández et al. 2024
Proteomics	MS raw files (vendor formats, mzML/mzXML/RAW), identification/quantification outputs (e.g. peptide/protein level, CSV/txt tables), spectral count or intensity/differential abundance tables	Mapping peptides to proteins/gene IDs; normalization; handling missing values; scaling; alignment of sample IDs; possibly defining protein co-expression networks; selecting features	Sanches et al. 2024
Metabolomics	Raw MS/NMR data (e.g.. mzML,.mzXML, etc.), peak lists, annotation files (e.g. with metabolite names/identifiers like CHEBI), processed abundance matrices (CSV/TSV/tabular), metadata	Peak processing; annotation; normalization; batch correction; mapping metabolites to pathways; filtering by quality; matching to same samples; merging with transcript/proteomic/genomic features; scaling	Kriegsmann and Deininger 2023

facilitate access to advanced tools and data without the need for heavy institutional investment. Moreover, training programs focused on quantitative genetics, data curation, and bioinformatics will be crucial to building a sustainable human resource base. In this context, the hyper-integrated genomic selection framework illustrated in Fig. 3 represents a progressive roadmap toward next-generation crop breeding rather than an immediately implementable system. Through phased adoption, beginning with low-cost genomic tools and moving toward full integration, public-sector breeding programs in developing regions can gradually transition to efficient, data-driven selection pipelines aimed at delivering climate-resilient and high-yielding crop varieties. Collectively, the factors outlined above, ranging from incomplete genomic resources and phenotyping limitations to computational and economic constraints highlight that the widespread adoption of GS remains uneven across regions and crops. Addressing these barriers requires coordinated investment in digital infrastructure, open-access data platforms, and human capacity building. Moreover, ethical considerations related to data ownership, sharing, and benefit distribution must be incorporated into future GS frameworks. The following section synthesizes these insights, outlining both the challenges and forward-looking prospects of GS as a cornerstone of next-generation plant breeding.

Challenges and limitations of integrated genomic selection

While Genomic Selection (GS) and its integration with multi-omics, machine learning (ML), and environmental modelling have revolutionized predictive breeding, several challenges constrain their scalability, interpretability, and equitable application, particularly in public breeding systems. These limitations, spanning technological, computational, and ethical domains, highlight the need for continued methodological refinement and institutional support.

High cost and data intensity of multi-omics integration

Integrated GS frameworks increasingly combine genomics with transcriptomics, metabolomics, phenomics, and environmental datasets to capture the multi-layered architecture of complex traits (Cobb et al. 2013; Varshney et al. 2021). However, such multi-omics integration remains cost-prohibitive for most breeding programs, especially in developing countries. High-throughput sequencing, RNA-Seq, and metabolomic profiling demand substantial financial investment and sophisticated infrastructure. Moreover, multi-omics datasets differ in scale, noise level,

and temporal resolution, complicating normalization and feature alignment. Integrating heterogeneous data often requires specialized bioinformatic pipelines and statistical frameworks, increasing computational burden and the risk of overfitting when sample sizes are limited. These factors restrict routine multi-omics-assisted GS to well-funded research centres rather than operational breeding pipelines.

Model complexity and interpretability in deep learning-based GS

Machine and deep learning algorithms have significantly enhanced GS accuracy by capturing nonlinear interactions, epistasis, and G × E effects (Montesinos-López et al. 2018; Sandhu et al. 2022). However, these models, especially convolutional and recurrent neural networks, often function as "black boxes," providing limited interpretability regarding biological mechanisms. The inability to trace predictions back to causal genomic regions hinders biological insight and validation, limiting confidence among breeders and geneticists seeking mechanistic understanding. Moreover, model training requires extensive datasets and hyperparameter tuning, demanding computational expertise rarely available in conventional breeding programs. Without transparent frameworks for model explainability, such as SHAP values (SHapley Additive exPlanations) or attention-based visualization, adoption of deep learning models risks prioritizing predictive accuracy over biological interpretability.

Computational bottlenecks and reproducibility constraints

Large-scale GS implementation, particularly when combined with dense genomic and phenomic datasets, generates massive computational loads. Storing, managing, and analyzing terabytes of data requires high-performance computing (HPC) infrastructure, cloud storage, and advanced data pipelines (Crossa et al. 2021). Many public-sector breeding programs lack sustained access to such resources, creating disparities in technological capability between private and public research institutions. Furthermore, computational pipelines for GS often lack standardization. Variations in preprocessing, imputation, model training, and cross-validation protocols lead to inconsistent results across studies (Browning and Browning 2016). Reproducibility remains a major challenge, compounded by insufficiently documented code and metadata. Without open-source frameworks and reproducible workflows, scaling GS from experimental datasets to breeding-scale populations remains inefficient and error-prone.

Ethical, legal, and data-sharing concerns

The integration of GS with large genomic and phenotypic datasets raises ethical and governance issues, particularly regarding data ownership, privacy, and equitable benefit sharing. Public breeding institutions often depend on international collaborations and donor-funded projects, where data access policies vary widely (Zhao et al. 2017). Ambiguities surrounding intellectual property and data sovereignty can hinder collaboration, delay data sharing, or restrict access to predictive models. Additionally, the increasing use of digital phenotyping, UAV imaging, and environmental sensors introduces new data privacy and consent concerns, especially when field sites involve smallholder communities. Clear data governance frameworks are therefore essential to ensure transparency, protect farmer data, and promote open-access repositories that enable equitable use of GS technologies across regions.

Institutional and capacity barriers

Beyond technical constraints, institutional capacity remains a key limiting factor for widespread GS deployment. The transition from conventional breeding to data-driven pipelines requires multidisciplinary expertise spanning quantitative genetics, bioinformatics, machine learning, and agronomy (Crossa et al. 2017a, b). Many breeding programs lack the trained personnel to design, maintain, and interpret GS systems effectively. Furthermore, long-term funding mechanisms are often absent, resulting in fragmented implementation and underutilized datasets once initial projects end. Building sustainable GS capacity will require targeted investment in breeder training, open-access computational tools, and collaborative data platforms to ensure equitable participation in the genomic era.

Collectively, the factors outlined above, ranging from incomplete genomic resources and phenotyping limitations to computational and economic constraints, highlight that the widespread adoption of GS remains uneven across regions and crops. Addressing these barriers requires coordinated investment in digital infrastructure, open-access data platforms, and human capacity building. Moreover, ethical considerations related to data ownership, sharing, and benefit distribution must be incorporated into future GS frameworks. The following section synthesizes these insights, outlining both the challenges and forward-looking prospects of GS as a cornerstone of next-generation plant breeding.

Future prospects and research priorities

Genomic selection (GS) represents a paradigm shift in plant breeding, transforming the process from empirical selection to predictive, data-driven decision-making. Over the past decade, GS has demonstrated significant potential in accelerating genetic gain by 30–50% and shortening breeding cycles by 2–4 years across several staple crops, including wheat, maize, rice, and chickpea (Crossa et al. 2017a, b; Varshney et al. 2021). Yet, despite remarkable progress, its full impact remains unrealized, particularly in under-resourced breeding systems and minor crops. Future research must therefore focus on integrating GS with artificial intelligence (AI), pan-genomics, multi-omics, and sustainability frameworks to maximize predictive accuracy, biological interpretability, and global equity in breeding outcomes.

AI-driven predictive breeding

The next generation of genomic selection will be defined by the convergence of AI, deep learning, and systems biology. Advanced machine learning architectures such as convolutional neural networks (CNNs), graph neural networks (GNNs), and transformer-based models can capture nonlinear and hierarchical interactions across genotypic, phenotypic, and environmental data (Montesinos-López et al. 2021; Azodi et al. 2020). Empirical studies already show that AI-augmented GS improves prediction accuracy for complex traits by 10–20% over conventional GBLUP and Bayesian methods, especially under stress environments. The next decade should prioritize developing interpretable AI models capable of identifying biologically meaningful features and epistatic networks. Moreover, real-time adaptive models integrating satellite, soil, and weather data can enable "climate-smart" breeding allowing breeders to simulate responses to future climatic scenarios. These systems are projected to reduce time-to-market for improved cultivars by up to 40%, representing one of the most cost-effective strategies for enhancing food security under climate change.

Pan-genomic and pangenotypic modelling

As the diversity gap in reference genomes narrows, pan-genomics will redefine the genetic foundation of selection. Traditional GS models often rely on a single reference genome, missing structural variants (SVs) and presence–absence variations (PAVs) that can explain up to 30% of unexplained heritability (Golicz et al. 2020). Integrating pan-genomic data across multiple populations allows more comprehensive allele representation, particularly of rare, adaptive, or locally selected variants. In maize and rice, inclusion of pan-genomic markers has increased predictive ability by up to 18% for yield and drought tolerance (Gao et al. 2019). Future research should aim to construct global pan-genome databases for at least 50 major and orphan crops by 2035, coupled with pangenotypic prediction models that account for intra-species diversity and region-specific allele-environment interactions. This would enable "localized genomic selection" breeding strategies fine-tuned to specific agro-climatic niches.

Multi-omics and environmental integration

Complex traits such as yield stability, nutrient use efficiency, and abiotic stress tolerance are governed by multi-layered regulatory networks. Integrating genomic, transcriptomic, proteomic, metabolomic, and epigenomic data within GS models can enhance prediction accuracies by 15–25% (Tardieu et al. 2017; Roorkiwal et al. 2020). Emerging frameworks such as integrated multi-omics GS (iGS) can elucidate the causal pathways linking genotype to phenotype, improving model interpretability. The coupling of omics datasets with high-throughput phenotyping (drones, hyperspectral imaging, IoT-based monitoring) and environmental metadata enables robust modelling of G × E interactions across spatiotemporal scales. Developing global open-access multi-omics repositories—supported by initiatives such as CGIAR Excellence in Breeding, AG2PI, and the European Plant Phenotyping Network should be a key strategic goal to democratize data access and accelerate predictive model training.

Sustainable and equitable implementation

Despite the proven potential of GS, global adoption remains uneven. While genotyping costs have declined from $120 per sample (2015) to < $15 (2025), less than 10% of public breeding programs in developing countries have fully integrated GS pipelines (Bassi et al. 2016). To ensure global impact, the next phase of research must focus on low-cost, scalable genotyping, open-source prediction platforms, and regional data-sharing consortia. The deployment of portable sequencing and imputation-based low-density marker panels could cut operational costs by another 70–80%, enabling resource-limited programs to benefit from GS innovation. Embedding sustainability indices—such as carbon footprint, nutrient-use efficiency, and resilience score into GS pipelines can align breeding priorities with UN SDGs 2 (Zero Hunger) and 13 (Climate Action). A fivefold expansion in GS adoption across the Global South by 2035 could directly contribute to a 20% rise in yield stability and 25% reduction in input dependency, reinforcing food and nutritional security.

Integrative decision support and digital breeding systems

The integration of GS outputs into digital decision support systems (DSS) represents a critical step toward predictive and automated breeding. AI-assisted DSS can synthesize genomic predictions with crop growth, soil fertility, and climate simulation models to prioritize crosses, allocate resources, and optimize field trials. Cloud-based breeding dashboards and mobile breeder interfaces are emerging as practical tools for real-time data management. Quantitatively, such systems can reduce breeding cycle times by 25–35% and enhance cross-location coordination in multi-environment trials. Future programs should invest in interoperable digital infrastructures linking genomic databases with phenomic and agronomic data streams, establishing a "breeding-as-a-service" ecosystem that supports collaborative innovation across institutions.

Ethical, data, and governance frameworks

As genomic data proliferates, ethical data management and fair benefit sharing become crucial. Adopting FAIR data principles (Findable, Accessible, Interoperable, Reusable) across breeding networks will ensure long-term sustainability of data-driven innovation (Varshney et al. 2021). By 2035, coordinated global frameworks—possibly under the joint stewardship of FAO, CGIAR, and regional genebanks should establish standardized policies for genomic data ownership, privacy, and equitable access. Moreover, equitable training programs in bioinformatics and AI for young breeders, particularly in Africa and South Asia, could triple the efficiency of data use while diversifying the global talent base in genomic breeding.

Concluding perspective

In summary, the coming decade will mark the maturation of genomic selection from a predictive framework into a holistic breeding ecosystem. The integration of AI, pangenomics, and multi-omics data will not only enhance the biological precision of GS but also bridge the gap between genotypic prediction and real-world performance. To fully harness its potential, global efforts must converge on data democratization, interpretability, sustainability, and inclusivity. If pursued strategically, GS could enable a doubling of genetic gain by 2035, foster resilient agricultural systems, and position genomic-assisted breeding as a cornerstone of climate-smart agriculture and global food security.

Author contributions **B.E., N.A.,** and **A.,** conceived the idea; **B.E., N.A., A., B.S., G.N.,** and **K.K.** contributed to writing the original draft; **N.A.** and **A.** prepared the figures and tables; **G.K., T.L.P., Y.H.,** and **V.,** contributed to reviewing and editing the manuscript. All authors read and approved the final version of the manuscript.

Funding Not applicable.

Data availability All data generated or analysed in this study are included in the manuscript.

Declarations

Conflict of interest The authors declare no competing interests.

References

Abed A, Pérez-Rodríguez P, Crossa J, Belzile F (2018) When less can be better: how can we make genomic selection more cost-effective and accurate in barley? Theor Appl Genet 131:1873–1890. https://doi.org/10.1007/s00122-018-3120-8

Alemu A, Åstrand J, Montesinos-Lopez OA, Sanchez JI et al (2024) Genomic selection in plant breeding: key factors shaping two decades of progress. Mol Plant 17(4):552–578

Allier A, Lehermeier C, Charcosset A, Moreau L, Teyssèdre S (2019) Improving short- and long-term genetic gain by accounting for within-family variance in optimal cross-selection. Front Genet 10:1006. https://doi.org/10.3389/fgene.2019.01006

Amaresh AN, Krishnappa G, Durai AA, Manimekalai R et al (2025a) Next-generation molecular breeding tools to harness higher genetic gains in sugarcane. Planta 262(5):122. https://doi.org/10.1007/s00425-025-04842-7

Amaresh NA, Appunu C et al (2025b) Advanced genome editing technologies: potentials and prospects in improvement of sugar crops. Sugar Tech 27:14–28. https://doi.org/10.1007/s12355-024-01447-4

Araus JL, Cairns JE (2014) Field high-throughput phenotyping: the new crop breeding frontier. Trends Plant Sci 19(1):52–61. https://doi.org/10.1016/j.tplants.2013.09.008

Azodi CB, Tang J, Shiu SH (2020) Machine learning approaches for genomic prediction. Plant Genome 13(3):e20001

Badu-Apraku B, Talabi AO, Fakorede MAB, Fasanmade Y, Gedil M, Magorokosho C et al (2019) Yield gains and associated changes in an early yellow bi-parental maize population following genomic selection for *Striga* resistance and drought tolerance. BMC Plant Biol 19(1):129. https://doi.org/10.1186/s12870-019-1740-z

Banks R (2022) Evolution of genetics organisations' strategies through the implementation of genomic selection: learnings and prospects. Agriculture 12:1524. https://doi.org/10.3390/agriculture12101524

Bassi FM, Bentley AR, Charmet G, Ortiz R, Crossa J (2016) Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). Plant Sci 242:23–36. https://doi.org/10.1016/j.plantsci.2015.08.021

Beche E, Gillman JD, Song Q, Nelson R, Beissinger T, Decker J, Shannon G, Scaboo AM (2021) Genomic prediction using training population design in interspecific soybean populations. Mol Breed 41(2):15

Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci 48(5):1649–1664. https://doi.org/10.2135/cropsci2008.03.0131

Bernardo R (2016) Bandwagons i, too, have known. Theor Appl Genet 129(12):2323–2332

Bhandari A, Bartholomé J, Cao-Hamadoun T-V, Kumari N, Frouin J, Kumar A et al (2019) Selection of trait-specific markers and multi-environment models improve genomic predictive ability in rice. PLoS One 14(1):e0208871. https://doi.org/10.1371/journal.pone.0208871

Bhandari P, Kim J, Lee TG (2023) Genetic architecture of fresh-market tomato yield. BMC Plant Biol 23(1):18

Biscarini F, Nazzicari N, Bink M, Arús P, Aranzana MJ, Verde I et al (2017) Genome-enabled predictions for fruit weight and quality from repeated records in European peach progenies. BMC Genomics 18(1):1–15. https://doi.org/10.1186/s12864-017-4392-7

Brainard SH, Ellison SL, Simon PW, Dawson JC, Goldman IL (2022) Genetic characterization of carrot root shape and size using genome-wide association analysis and genomic-estimated breeding values. Theor Appl Genet 135:605–622. https://doi.org/10.1007/s00122-021-03988-8

Brault C, Lazerges J, Doligez A et al (2022) Interest of phenomic prediction as an alternative to genomic prediction in grapevine. Plant Methods 18(1):108. https://doi.org/10.1186/s13007-022-00940-9

Browning BL, Browning SR (2016) Genotype imputation with millions of reference samples. Am J Hum Genet 98(1):116–126

Burgueño J, de Los Campos G, Weigel K, Crossa J (2012) Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. Crop Sci 52:707–719

Cheng CH, Datson PM, Hilario E, Deng CH, Manako KI, McNeilage M, Hoeata K (2019) Genomic predictions in diploid *Actinidia chinensis* (kiwifruit). Eur J Hortic Sci 84(4):213–217

Cobb JN, Declerck G, Greenberg A, Clark R, McCouch S (2013) Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. Theor Appl Genet 126:867–887. https://doi.org/10.1007/s00122-013-2066-0

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142(1–2):169–196. https://doi.org/10.1007/s10681-005-1681-5

Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc Lond B Biol Sci 363(1491):557–572. https://doi.org/10.1098/rstb.2007.2170

Crossa J, de los Campos G, Pérez P, Gianola D, Burgueño J, Araus JL (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186(2):713–724. https://doi.org/10.1534/genetics.110.118521

Crossa J, Fritsche-Neto R, Montesinos-Lopez OA, Costa-Neto G, Dreisigacker S, Montesinos-Lopez A, Bentley AR (2021) The modern plant breeding triangle: optimizing the use of genomics, phenomics, and enviromics data. Front Plant Sci 12:651480

Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de Los Campos G, Burgueño J, González-Camacho JM, Pérez-Elizalde S, Beyene Y, Dreisigacker S (2017a) Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci 22(11):961–975. https://doi.org/10.1016/j.tplants.2017.08.011

Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López OA, Jarquín D, de Los Campos G, Burgueno J, González-Camacho JM, Pérez-Elizalde S, Beyene Y, Dreisigacker S, Singh R, Zhang X, Gowda M, Roorkiwal M, Rutkoski J, Varshney RK (2017b) Genomic selection in plant breeding: methods, models, and perspectives. Crop Sci 57(1):1–20

D'Agostino N, Tripodi P (2017) NGS-based genotyping, high-throughput phenotyping and genome-wide association studies laid the foundations for next-generation breeding in horticultural crops. Diversity 9(3):38. https://doi.org/10.3390/d9030038

da Silva FA, Viana AP, Correa CCG, Santos EA, de Oliveira JAVS, Andrade JDG, Ribeiro RM, Glória LS (2021) Bayesian ridge regression shows the best fit for SSR markers in *Psidiumguajava* among Bayesian models. Sci Rep 11:13639. https://doi.org/10.1038/s41598-021-93120-z

Daetwyler HD, Bansal UK, Bariana HS, Hayden MJ, Hayes BJ (2014) Genomic prediction for rust resistance in diverse wheat landraces. Theor Appl Genet 127:1795–1803. https://doi.org/10.1007/s00122-014-2341-8

Daetwyler HD, Calus MP, Pong-Wong R, de los Campos G, Hickey JM (2013) Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics 193(2):347–365. https://doi.org/10.1534/genetics.112.147983

Das RR, Vinayan MT, Patel MB, Phagna RK, Singh SB, Shahi JP (2020) Genetic gains with rapid-cycle genomic selection for combined drought and waterlogging tolerance in tropical maize (*Zea mays* L.). Plant Genome 13:1–15. https://doi.org/10.1002/tpg2.20035

Dávila-Fernández M, Creixell P, Villar R (2024) Model-based multi-omics data integration to characterize multi-level systems: the INTEGRATE computational framework. PLoS Comput Biol 20:e1009337. https://doi.org/10.1371/journal.pcbi.1009337

de Bem Oliveira I, Resende MFR, Ferrão LFV, Amadeu RR, Endelman JB, Kirst M, Muñoz PR (2019) Genomic prediction of autotetraploids: influence of relationship matrices, allele dosage, and continuous genotyping calls in phenotype prediction. G3 Genes|genomes|genetics 9:1189–1198. https://doi.org/10.1534/g3.119.400059

de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MP (2013) Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics 193:327–345. https://doi.org/10.1534/genetics.112.143313

Dhanya VG, Subeesh A, Susmita C, Amaresh SSJ, Dilsha C, Kumar S et al (2024) High throughput phenotyping using hyperspectral imaging for seed quality assurance coupled with machine learning methods: principles and way forward. Plant Physiol Rep 29:749–768. https://doi.org/10.1007/s40502-024-00839-8

Díaz S, Ariza-Suarez D, Ramdeen R, Aparicio J, Arunachalam N, Hernández C, Raatz B (2021) Genetic architecture and genomic prediction of cooking time in common bean (*Phaseolus vulgaris* L.). Front Plant Sci 11:2257. https://doi.org/10.3389/fpls.2020.622213

Duangjit J, Causse M, Sauvage C (2016) Efficiency of genomic selection for tomato fruit quality. Mol Breed 36(3):29

Gao L, Gonda I, Sun H, Ma Q, Bao K, Tieman DM, Fei Z (2019) The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. Nat Genet 51(6):1044–1051

Gaynor RC, Gorjanc G, Bentley AR, Ober ES, Howell P, Jackson R, Hickey JM (2017) A two-part strategy for using genomic selection to develop inbred lines. Crop Sci 57:2372–2386

Gebremedhin A, Li Y, Shunmugam AS, Sudheesh S, Valipour-Kahrood H, Hayden MJ, Kaur S (2024) Genomic selection for target traits in the Australian lentil breeding program. Front Plant Sci 14:1284781

Germana MA (2011) Anther culture for haploid and doubled haploid production. Plant Cell Tissue Organ Cult 104:283–300

Gimode D, Wallace J, Holbrook C, Isleib T, Chu Y, Virk S, Porter W, Ozias-Akins P (2024) Genomic selection as an approach to select for reduced aflatoxin contamination in peanut under terminal drought stress. Peanut Sci 51:18–31. https://doi.org/10.3146/0095-3679-51-PS23-4

Goddard ME, Hayes BJ, Meuwissen TH (2010) Genomic selection in livestock populations. Genet Res. https://doi.org/10.1017/s0016672310000613

Goddard ME, Kemper KE, MacLeod IM, Chamberlain AJ, Hayes BJ (2011) Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms, and genetic architecture. Proc R Soc B Biol Sci 281(1796):20122352. https://doi.org/10.1098/rspb.2012.2352

Godfray HCJ, Garnett T (2014) Food security and sustainable intensification. Philos Trans R Soc Lond B Biol Sci 369(1639):20120273. https://doi.org/10.1098/rstb.2012.0273

Gois IB, Borém A, Cristofani-Yaly M, De Resende MDV, Azevedo CF, Bastianel M, Machado MA (2016) Genome-wide selection in *Citrus* breeding. Genet Mol Res 15:gmr15048863. https://doi.org/10.4238/gmr15048863

Golicz AA, Bayer PE, Bhalla PL, Batley J, Edwards D (2020) Pangenomics comes of age: from bacteria to plant and animal applications. Trends Genet 36(2):132–145

Grenier C, Cao T-V, Ospina Y, Quintero C, Châtel MH, Tohme J, Ahmadi N (2015) Accuracy of genomic selection in a rice synthetic population developed for recurrent selection breeding. PLoS One 10:e0136594. https://doi.org/10.1371/journal.pone.0136594

Habier D, Fernando RL, Dekkers JC (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–2397. https://doi.org/10.1534/genetics.107.081190

Haile TA, Heidecker T, Wright D, Neupane S, Ramsay L, Vandenberg A, Bett KE (2020) Genomic selection for lentil breeding: empirical evidence. Plant Genome 13:1–30. https://doi.org/10.1002/tpg2.20002

Hassen BM, Bartholomé J, Valè G, Cao TV, Ahmadi N (2018) Genomic prediction accounting for genotype by environment interaction offers an effective framework for breeding simultaneously for adaptation to an abiotic stress and performance under normal cropping conditions in rice. G3 (Bethesda) 8(7):2319–2332. https://doi.org/10.1534/g3.118.200098

He W, Li X, Qian Q, Shang L (2025) The developments and prospects of plant super-pangenomes: demands, approaches, and applications. Plant Commun 6:2

Heffner EL, Sorrells ME, Jannink JL (2009a) Genomic selection for crop improvement. Crop Sci 49:1–12

Heffner EL, Sorrells ME, Jannink JL (2009b) Plant breeding with genomic selection: gain per unit time and cost. Crop Sci 49:1–12

Heslot N, Jannink JL, Sorrells ME (2015) Perspectives for genomic selection applications and research in plants. Crop Sci 55:1–12. https://doi.org/10.2135/cropsci2014.03.0249

Hickey JM, Chiurugwi T, Mackay I, Powell W (2017) Implementing genomic selection in crop breeding: strategies, structure, and practical considerations. Heredity 118:100–109. https://doi.org/10.1038/hdy.2016.37

Hong JP, Ro N, Lee HY, Kim GW, Kwon JK, Yamamoto E, Kang BC (2020) Genomic selection for prediction of fruit-related traits in pepper (*Capsicum* spp.). Front Plant Sci 11:570871

Hossain A, Krupnik TJ, Timsina J, Mahboob MG, Chaki AK, Farooq M, Hasanuzzaman M (2020) Agricultural land degradation: processes and problems undermining future food security. Environment, climate, plant and vegetation growth. Springer, Cham, pp 17–61

Hu X, Xie W, Wu C, Xu S (2019) A directed learning strategy integrating multiple omic data improves genomic prediction. Plant Biotechnol J 17(10):2011–2020

Huang M, Balimponya EG, Mgonja EM, McHale LK, Luzi-Kihupi A, Wang G-L, Sneller CH (2019) Use of genomic selection in breeding rice (*Oryza sativa* L.) for resistance to rice blast (*Magnaporthe oryzae*). Mol Breed 39:1–16. https://doi.org/10.1007/s11032-019-1023-2

Huang X, Han B (2024) Natural variations and genome-wide association studies in crop plants. Annu Rev Plant Biol 65:531–551. https://doi.org/10.1146/annurev-arplant-050213-035715

Imai A, Kuniga T, Yoshioka T, Nonaka K, Mitani N, Fukamachi H, Hiehata N, Yamamoto M, Hayashi T (2019) Single-step genomic prediction of fruit-quality traits using phenotypic records of non-genotyped relatives in citrus. PLoS One 14(8):e0221880

Iperi C, Fernández-Ochoa Á, Barturen G, Pers JO, Foulquier N, Bettacchioli E et al (2025) BiomiX, a user-friendly bioinformatic tool for democratized analysis and integration of multiomics data. BMC Bioinformatics 26(1):8

Isidro J, Jannink JL, Akdemir D, Poland J, Heslot N, Sorrells ME (2015) Training set optimization under population structure in genomic selection. Theor Appl Genet 128:145–158. https://doi.org/10.1007/s00122-014-2418-4

Islam MS, Fang DD, Jenkins JN, Guo J, McCarty JC, Jones DC (2020) Evaluation of genomic selection methods for predicting fiber quality traits in upland cotton. Mol Genet Genomics 295:67–79

Jan HU, Abbadi A, Lücke S, Nichols RA, Snowdon RJ (2016) Genomic prediction of testcross performance in canola (*Brassicanapus*). PLoS One 11(1):e0147769. https://doi.org/10.1371/journal.pone.0147769

Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics 9(2):166–177. https://doi.org/10.1093/bfgp/elq001

Jarquín D, Crossa J, Lacaze X, Cheyron PD, Daucourt J, Lorgeou J (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet 127(3):595–607. https://doi.org/10.1007/s00122-013-2243-1

Jayakodi M, Shim H, Mascher M (2024) What are we learning from plant pangenomes? Annu Rev Plant Biol 76:663–686

Jia Y, Jannink JL (2012) Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192:1513–1522. https://doi.org/10.1534/genetics.112.144246

Kopeć P (2024) Climate change—the rise of climate-resilient crops. Plants 13(4):490

Kriegsmann J, Deininger S-O (2023) Multi-omics integration in biomedical research – a metabolomics-centred review. Metabolomics 19:14. https://doi.org/10.1007/s11306-023-02001-8

Krishnappa G, Savadi S, Tyagi BS, Singh SK, Mamrutha HM, Kumar S, Singh GP (2021) Integrated genomic selection for rapid improvement of crops. Genomics 113:1070–1086. https://doi.org/10.1016/j.ygeno.2021.02.007

Kumar R, Das SP, Choudhury BU, Kumar A, Prakash NR, Verma R, Chakraborti M, Devi AG, Bhattacharjee B, Das R, Das B (2024) Advances in genomic tools for plant breeding: harnessing DNA molecular markers, genomic selection, and genome editing. Biol Res 57:80. https://doi.org/10.1186/s40659-024-00562-6

Kumar R, Meena MR, Dhansu P, Chhabra ML, Appunu C, Sreenivasa V, Amaresh G (2025) Disclosure of productive and stable sugarcane genotypes for cane yield and quality traits using various stability statistics. Trop Plant Biol 18(1):11

Kumar S, Banks TW, Cloutier S (2012) SNP discovery through next-generation sequencing and its applications. Front Genet 3:256. https://doi.org/10.3389/fgene.2012.00256

Kumar S, Kirk C, Deng CH, Shirtliff A, Wiedow C, Qin M, Wu J, Brewer L (2019) Marker-trait associations and genomic predictions of interspecific pear (*Pyrus*) fruit characteristics. Sci Rep 9(1):9072

Lesk C, Rowhani P, Ramankutty N (2016) Influence of extreme weather disasters on global crop production. Nature 529(7584):84–87. https://doi.org/10.1038/nature16467

Li N, He Q, Wang J, Wang B, Zhao J, Huang S, Yu Q (2023) Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. Nat Genet 55:852–860

Li Y, Ruperao P, Batley J, Edwards D, Khan T, Colmer TD, Sutton T (2018) Investigating drought tolerance in chickpea using genome-wide association mapping and genomic selection based on whole-genome resequencing data. Front Plant Sci 9:190. https://doi.org/10.3389/fpls.2018.00190

Li Z, Liu S, Conaty W, Zhu QH, Moncuquet P, Stiller W, Wilson I (2022) Genomic prediction of cotton fibre quality and yield traits using Bayesian regression methods. Heredity 129:103–112

Lin Z, Li Y, Riaz A, Sudheesh S, Yazdifar S, Atieno J, Kaur S (2025) Assessing the utility of genomic selection to breed for

durable Ascochyta blight resistance in chickpea. Plant Genome 18(2):e70023
Lorenz AJ, Chao S, Asoro FG, Heffner EL, Hayashi T, Iwata H, Jannink JL (2011) Genomic selection in plant breeding: knowledge and prospects. Adv Agron 110:77–123. https://doi.org/10.1016/B978-0-12-385531-2.00002-5
Mangin B, Bonnafous F, Blanchet N, Boniface MC, Bret-Mestries E, Carrère S, Cottret L, Legrand L, Marage G, Pegot-Espagnet P, Munos S (2017) Genomic prediction of sunflower hybrids oil content. Front Plant Sci 8:1633
Meher PK, Rustgi S, Kumar A (2022) Performance of Bayesian and BLUP alphabets for genomic prediction: analysis, comparison and results. Heredity (Edinb) 128(6):519–530. https://doi.org/10.1038/s41437-022-00539-9
Mellers G, Mackay I, Cowan S, Griffiths I, Martinez-Martin P, Poland JA, Howarth CJ (2020) Implementing within-cross genomic prediction to reduce oat breeding costs. Plant Genome 13:e20004
Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4):1819–1829. https://doi.org/10.1093/genetics/157.4.1819
Miller MJ, Song Q, Fallen B, Li Z (2023) Genomic prediction of optimal cross combinations to accelerate genetic improvement of soybean (*Glycine max*). Front Plant Sci 14:1171135
Minamikawa MF, Nonaka K, Kaminuma E, Kajiya-Kanegae H, Onogi A, Goto S, Yoshioka T, Imai A, Hamada H, Hayashi T, Matsumoto S (2017) Genome-wide association study and genomic prediction in citrus: potential of genomics-assisted breeding for fruit quality traits. Sci Rep 7(1):4721
Mirdita V, He S, Zhao Y, Korzun V, Bothe R, Ebmeyer E, Jiang Y (2015) Potential and limits of whole genome prediction of resistance to Fusarium head blight and Septoria tritici blotch in a vast Central European elite winter wheat population. Theor Appl Genet 128:2471–2481. https://doi.org/10.1007/S00122-015-2602-1
Montesinos-López OA, Montesinos-López A, Crossa J, Gianola D (2018) Multi-trait, multi-environment deep learning modeling for genomic-enabled prediction. Genes. https://doi.org/10.1534/g3.118.200728
Montesinos-López OA, Montesinos-López A, Gianola D, Crossa J, Hernández-Suárez CM, Nogales A (2021) Multi-trait genomic-enabled prediction enhanced with deep learning, gradient boosting, and support vector machine methods. G3 (Bethesda) 11(2). https://doi.org/10.1093/g3journal/jkab006
Montesinos-López OA, Montesinos-López A, Mosqueda-González BA, Delgado-Enciso I, Chavira-Flores M, Crossa J, Dreisigacker S, Sun J, Ortiz R (2025) Genomic prediction powered by multi-omics data. Front Genet 16:1636438. https://doi.org/10.3389/fgene.2025.1636438
Mueller ND, Gerber JS, Johnston M, Ray DK, Ramankutty N, Foley JA (2012) Closing yield gaps through nutrient and water management. Nature 490(7419):254–257. https://doi.org/10.1038/nature11420
Munns R, Tester M (2008) Mechanisms of salinity tolerance. Annu Rev Plant Biol 59:651–681. https://doi.org/10.1146/annurev.arplant.59.032607.092911
Muranty H, Troggio M, Sadok IB, Rifaï MA, Auwerkerken A, Banchi E, Bink MC (2015) Accuracy and responses of genomic selection on key traits in apple breeding. Hortic Res. https://doi.org/10.1038/hortres.2015.60
Nguyen HT, Khan MAR, Nguyen TT, Pham NT, Nguyen TTB, Anik TR, Ha CV (2025) Advancing crop resilience through high-throughput phenotyping for crop improvement in the face of climate change. Plants 14(6):907
Norman A, Taylor J, Edwards J, Kuchel H (2018) Optimizing genomic selection in wheat: effect of marker density, population size, and population structure on prediction accuracy. G3 Genes|genomes|genetics 8:2889–2899. https://doi.org/10.1534/g3.118.200311
Nsibi M, Gouble B, Bureau S, Flutre T, Sauvage C, Audergon JM, Regnard JL (2020) Adoption and optimization of genomic selection to sustain breeding for apricot fruit quality. G3 Genes|genomes|genetics 10(12):4513–4529
Nunavath A, Moniusha J, Keerthana S, Amaresh NM, Prathima PT, Manimekalai R (2025a) Genomic exploration for the sucrose content in sugarcane. Trop Plant Biol 18:44. https://doi.org/10.1007/s12042-025-09413-4
Nunavath A, Murugan N, Keerthana S, Kumari S, Singaravelu B, Sundar AR, Manimekalai R (2025b) Transcription factors in plant biotic and abiotic stress responses: potentials and prospects in sugarcane. Trop Plant Biol 18:1–13
Nunavath A, Nandhini M, Keerthana S, Kumar R, Gujjar RS, Gomathi R, Manimekalai R (2025c) Genome-wide expression profiling of cytochrome P450 genes in response to oxidative stress in *Saccharum* spp. Sugar Tech 27:378–392
Nyine M, Uwimana B, Blavet N, Hřibová E, Vanrespaille H, Batte M, Doležel J (2018) Genomic prediction in a multiploid crop: genotype by environment interaction and allele dosage effects on predictive ability in banana. Plant Genome 11:170090
Pandey MK, Chaudhari S, Jarquin D, Janila P, Crossa J, Patil SC, Varshney RK (2020) Genome-based trait prediction in multi-environment breeding trials in groundnut. Theor Appl Genet 133:3101–3117. https://doi.org/10.1007/s00122-020-03658-1
Pandey S, Singh A, Parida SK, Prasad M (2022) Combining speed breeding with traditional and genomics-assisted breeding for crop improvement. Plant Breed 141(3):301–313
Pathy TL (2024) Establishing marker trait associations, and predicting and cross validating genomic estimated breeding values for grain yield and its attributing traits in cowpea (Vigna Unguiculata L. Walp). Doctoral dissertation, University of Agricultural Sciences, GKVK, Bangalore
Pérez-Rodríguez P, Gianola D, González-Camacho JM, Crossa J, Manès Y, Dreisigacker S (2012) Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. G3 (Bethesda) 2:1595–1605. https://doi.org/10.1534/g3.112.003665/-/DC1
Perumal S, Higgins E, Sra S, Khedikar Y, Moore J, Chaudary R, Parkin IA (2024) Harnessing genome prediction in Brassica napus through a nested association mapping population. bioRxiv 2024–07
Petrasch S, Mesquida-Pesci SD, Pincot DDA, Feldmann MJ, López CM, Famula R, Hardigan MA, Cole GS, Knapp SJ, Blanco-Ulate B (2022) Genomic prediction of strawberry resistance to postharvest fruit decay caused by the fungal pathogen *Botrytiscinerea*. G3 Genes|genomes|genetics. https://doi.org/10.1093/g3journal/jkab378
Pryce JE, Johnston J, Hayes BJ, Sahana G, Weigel KA, McParland S, Spurlock D, Krattenmacher N, Spelman RJ, Wall E, Calus MPL (2014) Imputation of genotypes from low density (50,000 markers) to high density (700,000 markers) of cows from research herds in Europe, North America, and Australasia using 2 reference populations. J Dairy Sci 97:1799–1811. https://doi.org/10.3168/jds.2013-7368
Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100. https://doi.org/10.1016/S1369-5266(02)00240-6
Rio S, Mary-Huard T, Moreau L, Charcosset A (2019) Genomic selection efficiency and a priori estimation of accuracy in a structured dent maize panel. Theor Appl Genet 132:81–96. https://doi.org/10.1007/s00122-018-3196-1
Roorkiwal M, Rathore A, Das RR, Singh MK, Jain A, Srinivasan S, Varshney RK (2016) Genome-enabled prediction models for yield-related traits in chickpea. Front Plant Sci 7:1666
Roorkiwal M, Rathore A, Das RR, Pathania D, Singh NP, Varshney RK (2020) Integrating genomics for chickpea improvement: achievements and opportunities. Theor Appl Genet 133(6):1703–1720. https://doi.org/10.1007/s00122-020-03584-2

Roth M, Muranty H, Di Guardo M, Guerra W, Patocchi A, Costa F (2020) Genomic prediction of fruit texture and training population optimization towards the application of genomic selection in apple. Hortic Res. https://doi.org/10.1038/s41438-020-00370-5

Rutkoski J, Benson J, Jia Y, Brown-Guedira G, Jannink J-L, Sorrells M (2012) Evaluation of genomic prediction methods for Fusarium head blight resistance in wheat. Plant Genome 5:51–61. https://doi.org/10.3835/PLANTGENOME2012.02.0001

Salgotra RK, Chauhan BS (2023) Genetic diversity, conservation, and utilization of plant genetic resources. Genes 14(1):174

Sallam AH, Endelman JB, Jannink JL, Smith KP (2015) Assessing genomic selection prediction accuracy in a dynamic barley breeding population. Plant Genome 8:20. https://doi.org/10.3835/PLANTGENOME2014.05.0020

Sanches PHG, de Melo NC, Porcari AM, de Carvalho LM (2024) Integrating molecular perspectives: strategies for comprehensive multi-omics integrative data analysis and machine learning applications in transcriptomics, proteomics, and metabolomics. Biology 13(11):848

Sandhu KS, Shiv A, Kaur G, Meena MR, Raja AK, Vengavasi K, Kumar S (2022) Integrated approach in genomic selection to accelerate genetic gain in sugarcane. Plants 11:2139

Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A (2019) The global burden of pathogens and pests on major food crops. Nat Ecol Evol 3(3):430–439. https://doi.org/10.1038/s41559-018-0793-y

Sekine D, Yabe S (2020) Simulation-based optimization of genomic selection scheme for accelerating genetic gain while preventing inbreeding depression in onion breeding. Breed Sci 70:594–604. https://doi.org/10.1270/jsbbs.20047

Senthilkumar S, Vinod KK, Parthiban S, Thirugnanasambandam P, Lakshmi Pathy T, Banerjee N, Sarath Padmanabhan TS, Govindaraj P (2022) Identification of potential MTAs and candidate genes for juice quality-and yield-related traits in *Saccharum* clones: a genome-wide association and comparative genomic study. Mol Genet Genomics 297(3):635–654

Shan Q, Wang Y, Li J, Gao C (2013) Genome editing in rice and wheat using the CRISPR/Cas system. Nat Protoc 8:2368–2385

Shi A, Bhattarai G, Xiong H, Avila CA, Feng C, Liu B, Correll JC (2022) Genome-wide association study and genomic prediction of white rust resistance in USDA GRIN spinach germplasm. Hortic Res 9:uhac069. https://doi.org/10.1038/s41438-022-00768-5

Sidhu NS, Prasad G, MS S, ME K, Krishnappa G, Pathy TL (2025) Genome-wide association studies for next-generation crop breeding in the postgenomic era. Plant Breed 8:1-31. https://doi.org/10.1111/pbr.70017

Skendžić S, Zovko M, Živković IP, Lešić V, Lemić D (2021) The impact of climate change on agricultural insect pests. Insects 12:440. https://doi.org/10.3390/insects12050440

Stewart-Brown BB, Song Q, Vaughn JN, Li Z (2019) Genomic selection for yield and seed composition traits within an applied soybean breeding program. G3 Genes|genomes|genetics 9(7):2253–2265

Subramanian I, Verma S, Kumar S, Jere A, Anamika K (2020) Multi-omics data integration, interpretation, and its application. Bioinform Biol Insights 14:1177932219899051

Thomson MJ (2014) High-throughput SNP genotyping to accelerate crop improvement. Plant Breed Biotechnol 2(3):195–212

Thorwarth P, Yousef EA, Schmid KJ (2018) Genomic prediction and association mapping of curd-related traits in gene bank accessions of cauliflower. G3 Genes Genomes Genet 8:707–718. https://doi.org/10.1534/g3.117.300771

Tibbs Cortes L, Zhang Z, Yu J (2021) Status and prospects of genome-wide association studies in plants. Plant Genome 14(1):e20077. https://doi.org/10.1002/tpg2.20077

Torres LG, de Vilela Resen MD, Azevedo CF, Fonseca e Silva F, de Oliveira EJ (2019) Genomic selection for productive traits in biparental cassava breeding populations. PLoS One 14:e0220245. https://doi.org/10.1371/journal.pone.0220245

Varshney RK, Roorkiwal M, Sinha P, Punna R, Saxena RK, Bohra A, Pandey MK, Chitikineni A, Kim JS, Rathore A, Das RR, Boyd LA, Kilian B, Singh NK, Cook DR (2021) Accelerating genetic gains in crop breeding. Nat Genet 53:789–798

Vazquez AI, de los Campos G, Klimentidis YC, Rosa GJ, Gianola D (2016) Multi-trait genomic selection methods for plant and animal breeding. Genetics 202(1):9–30. https://doi.org/10.1534/genetics.115.183442

Viana AP, de Lima e Silva FH, Gloria LS, Ribeiro RM, Krause W, Boechat MSB (2017) Implementing genomic selection in sour passion fruit population. Euphytica 213:1–13

Vinayaka M, Prasad PRC (2024) AI-enhanced remote sensing applications in Indian sugarcane research: a comprehensive review. Sugar Tech 26(3):609–628. https://doi.org/10.1007/s12355-024-01409-w

Vinayaka PPRC, Avinash G, Amaresh KRA, Murali P, Palaniswami C, Govindaraj P (2025) Harnessing AI and remote sensing for precision sugarcane farming: tackling water stress, salinity, and nitrogen challenges. Front Agron 7:1681294. https://doi.org/10.3389/fagro.2025.1681294

Vivek BS, Krishna GK, Vengadessan V, Babu R, Zaidi PH, Kha LQ, Crossa J (2017) Use of genomic estimated breeding values results in rapid genetic gains for drought tolerance in maize. Plant Genome 10. https://doi.org/10.3835/PLANTGENOME2016.07.0070/FORMAT/PDF

Wang P, Si H, Li C, Xu Z, Guo H, Jin S, Cheng H (2025) Plant genetic transformation: achievements, current status and future prospects. Plant Biotechnol J. https://doi.org/10.1111/pbi.70028

Watson A, Ghosh S, Williams MJ, Cuddy WS, Simmonds J, Rey MD, Hickey LT (2018) Speed breeding is a powerful tool to accelerate crop research and breeding. Nat Plants 4:23–29

Wray NR, Thompson R (1990) Prediction of rates of inbreeding in selected populations. Genet Res 55(1):41–54. https://doi.org/10.1017/s0016672300025180

Wu PY, Tung CW, Lee CY, Liao CT (2019) Genomic prediction of pumpkin hybrid performance. Plant Genome 12:180082. https://doi.org/10.3835/plantgenome2018.10.0082

Würschum T, Reif JC, Kraft T, Janssen G, Zhao Y (2013) Genomic selection in sugar beet breeding populations. BMC Genet 14:1–8

Xu Y, Wang X, Ding X, Zheng X, Yang Z, Xu C, Hu Z (2018) Genomic selection of agronomic traits in hybrid rice using an NCII population. Rice 11:32–10. https://doi.org/10.1186/S12284-018-0223-4/FIGURES/5

Yabe S, Yoshida H, Kajiya-Kanegae H, Yamasaki M, Iwata H, Ebana K, Nakagawa H (2018) Description of grain weight distribution leading to genomic selection for grain-filling characteristics in rice. PLoS One 13(11):e0207627. https://doi.org/10.1371/JOURNAL.PONE.0207627

Yadav S, Jackson P, Wei X, Ross EM, Aitken K, Deomano E, Atkin F, Hayes BJ, Voss-Fels KP (2020) Accelerating genetic gain in sugarcane breeding using genomic selection. Agronomy 10(4):585

Yamamoto E, Matsunaga H, Onogi A, Ohyama A, Miyatake K, Yamaguchi H, Fukuoka H (2017) Efficiency of genomic selection for breeding population design and phenotype prediction in tomato. Heredity 118(2):202–209

Yang Y, Saand MA, Huang L, Abdelaal WB, Zhang J, Wu Y, Wang F (2021) Applications of multi-omics technologies for crop improvement. Front Plant Sci 12:563953

Yoosefzadeh-Najafabadi M, Rajcan I, Eskandari M (2022) Optimizing genomic selection in soybean: an important improvement in agricultural genomics. Heliyon. https://doi.org/10.1016/j.heliyon.2022.e11873

You Q, Yang X, Peng Z, Xu L, Wang J (2018) Development and applications of a high throughput genotyping tool for polyploid crops: single nucleotide polymorphism (SNP) array. Front Plant Sci 9:104

Tardieu F, Cabrera-Bosquet L, Pridmore T, Bennett M (2017) Plant phenomics, from sensors to knowledge. Curr Biol 27(15):R770–R783

Zhao Y, Gowda M, Liu W, Würschum T, Maurer HP, Longin FH, Ranc N, Reif JC (2012) Accuracy of genomic selection in European maize elite breeding populations. Theor Appl Genet 124:769–776. https://doi.org/10.1007/S00122-011-1745-Y

Zhao C, Liu B, Piao S, Wang X, Lobell DB, Huang Y, Asseng S (2017) Temperature increase reduces global yields of major crops in four independent estimates. Proc Natl Acad Sci USA 114(35):9326–9331. https://doi.org/10.1073/pnas.1701762114

Zhao L, Tang P, Luo J, Liu J, Peng X, Shen M, Wang C, Zhao J, Zhou D, Fan Z, Chen Y (2025) Genomic prediction with NetGP based on gene network and multi-omics data in plants. Plant Biotechnol J 23(4):1190–1201

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. Plant Genome 1(1):5–20. https://doi.org/10.3835/plantgenome2008.02.0089

Zhu S, Guo T, Yuan C, Liu J, Li J, Han M, Zhao H, Wu Y, Sun W, Wang X, Wang T (2021) Evaluation of Bayesian alphabet and GBLUP based on different marker density for genomic prediction in Alpine Merino sheep. G3 11(11):jkab206. https://doi.org/10.1093/g3journal/jkab206

Zoppi J, Guillaume JF, Neunlist M, Chaffron S (2021) Mibiomics: an interactive web application for multi-omics data exploration and integration. BMC Bioinformatics 22(1):6