Journal of Environmental Agriculture and Agroecosystem Management (JEAAM)

Genomic Selection in Wheat Crop Improvement: An In-depth Review

Vishal Singh

Department of Genetics and Plant Breeding, Institute of Agricultural Sciences Banaras Hindu University (BHU), Varanasi

Corresponding author

vishal19@bhu.ac.in

Keywords: Genomic Selection, Crops and Wheat Quality

Published By
Notation Publishing
www.notationpublishing.com

ABSTRACT

The application of genomic selection (GS) in wheat breeding programs represents a paradigm shift in crop improvement strategies. Leveraging dense genomic markers and advanced prediction models, GS provides breeders with an effective tool to anticipate an individual's breeding value without the costs and labor associated with extensive phenotyping. This review article provides an in-depth synthesis of current methodologies, practical applications, limitations, and emerging research gaps in employing genomic selection for wheat crop improvement. With a focus on balancing prediction accuracy and cost-effectiveness, the discussion herein is aimed at plant breeding professionals who possess an advanced understanding of genomic prediction models.

1. Introduction

Wheat is one of the world's staple crops, and its breeding has traditionally relied on phenotypic selection, which is both time consuming and resource intensive. Genomic selection offers a promising alternative by using genome-wide markers to estimate the breeding value of a candidate genotype. This approach has revolutionized crop breeding because it allows for early selection and rapid cycling, ultimately leading to accelerated genetic gain and reduced breeding cycle times.

The concept of GS emerged over a decade ago and has since been refined for several crop species, including wheat. Early GS models, such as the ridge regression best linear unbiased prediction (RR-BLUP), formed the backbone of genomic prediction due to their simplicity and robustness in capturing additive genetic effects [1]. However, modern breeding challenges demand models that can accommodate complex genetic architectures, including non-additive effects and gene-by-environment interactions. Consequently, researchers have explored various advanced computational methods ranging from Bayesian approaches to state-of-the-art machine learning algorithms. Notably, deep learning models have shown promise in predicting complex traits, such as wheat quality traits, thereby offering avenues to capture intricate biological relationships that traditional models might overlook [2].

The refinement of GS methodologies reflects an ongoing commitment to improve both the predictive accuracy and cost-effectiveness of wheat breeding programs. Given the global importance of wheat production, integrating GS into breeding pipelines holds significant potential for food security and sustainable agriculture.

Comparative Methodologies

Genomic selection methodologies in wheat breeding are typically evaluated based on their predictive performance and underlying statistical assumptions. A variety of models, ranging from classical linear regression-based methods to modern machine learning models, have been applied to predict breeding values.

RR-BLUP and Linear Models

The ridge regression best linear unbiased prediction (RR-BLUP) model is one of the earliest and most widely adopted models in GS applications. RR-BLUP relies on a linear combination of marker effects and assumes that all markers contribute equally small effects to the trait of interest. The simplicity and computational efficiency of RR-BLUP make it an attractive option for breeding programs with limited computational resources. Research has demonstrated that RR-BLUP is effective in capturing additive genetic variance, which is crucial for traits governed predominantly by additive gene action [1].

Bayesian Approaches

Bayesian methods provide an alternative to traditional linear models by enabling the inclusion of prior information and accommodating non-uniform distributions of marker effects. Bayesian approaches, such as BayesA, BayesB, and BayesC, allocate different weights to individual markers, thereby allowing for more flexibility in modeling complex traits [3]. Although computationally more demanding than RR-BLUP, these methods have shown increased predictive accuracy in scenarios where traits exhibit heterogeneous genetic architectures.

Machine Learning and Deep Learning Models

The advent of machine learning has further expanded the toolkit available for genomic prediction in wheat. Deep learning, in particular, has been applied to capture non-linear relationships among genomic markers and various agronomic traits. Recent studies indicate that deep learning models can outperform traditional models in predicting wheat quality traits, highlighting the potential benefits of incorporating advanced computational techniques into breeding pipelines [2]. Despite these advantages, challenges remain in tuning hyperparameters and ensuring model interpretability. Moreover, the computational cost associated with deep learning models may limit their widespread adoption across breeding programs with constrained resources.

Comparative Analysis: Predictive Accuracy and Cost-Effectiveness

Comparative studies have focused on two primary attributes—predictive accuracy and cost-effectiveness. On one hand, the predictive accuracy of deep learning and some Bayesian models often exceeds that of linear models such as RR-BLUP, particularly for complex traits [2]. On the other hand, the cost-effectiveness of each method must be considered within the context of a breeding program's resource allocation strategy.

Traditional models, albeit with lower predictive accuracy for certain traits, offer lower computational requirements and simpler implementation protocols. This trait makes them attractive in large-scale breeding programs where resource optimization is critical. In contrast, while machine learning methods may provide superior prediction accuracy, their computational intensity and requirement for large training datasets might increase operational costs [2], [4]. Ultimately, the choice between methodologies is a balance between achieving incremental gains in prediction accuracy and managing operational costs effectively.

Practical Applications in Wheat Breeding

The practical application of genomic selection in wheat breeding programs has been substantiated by several studies and real-world applications. Integrating GS into selection strategies has resulted in measurable improvements in genetic gains and breeding efficiency.

Accelerating Breeding Cycles

A major benefit of GS is the capacity to shorten breeding cycles by performing early selection based solely on genomic data [5]. By predicting the performance of breeding lines before extensive field evaluation, breeders can reduce the time between generations. This early prediction translates into a more rapid accumulation of favorable alleles, particularly for traits with high heritability such as grain yield and quality attributes.

For example, research on soft red winter wheat has demonstrated that genomic selection significantly improves prediction accuracy for grain yield and associated agronomic traits, thereby enabling more targeted selection strategies [5]. This approach is crucial for crops like wheat, where multi-environment trials can be prohibitively expensive and logistically challenging.

Resource Optimization and Cost Savings

Genomic selection protocols inherently reduce the dependency on extensive and costly field trials by replacing or augmenting traditional phenotypic selection with genomic predictions. By identifying superior candidates early in the breeding process, GS helps allocate resources more efficiently, ensuring that subsequent field trials focus on the most promising lines [3]. The reduction in trial numbers coupled with enhanced prediction accuracy has translated into substantial cost savings, making GS a cost-effective strategy in modern breeding programs [3].

Studies have underscored that integrating GS into wheat breeding workflows not only accelerates the development of new cultivars but also contributes to improved genetic gain per unit time, thereby justifying its implementation despite the initial costs associated with genotyping and model training [4].

Improving Agronomic and Quality Traits

The selection of wheat quality traits, which include protein content, dough strength, and end-use quality, has greatly benefited from genomic selection approaches, especially with the use of sophisticated machine learning models. Traditional phenotypic evaluations of quality traits can be complex and expensive due to the need for specialized equipment and labor-intensive assays. In contrast, genomic prediction models have shown promise in accurately predicting these traits, enabling breeders to develop cultivars that meet specific market demands.

The ability of deep learning models to discern non-additive genetic effects and interactions has been particularly valuable in this context. As breeders face the twin challenges of ensuring high yield and superior quality, these advanced models provide a conduit to simultaneously enhance multiple traits with improved prediction accuracy [2].

Integration with Conventional Breeding Programs

Successful integration of GS into traditional wheat breeding programs relies on strategically designed training and testing populations. Effective training population design is paramount to achieve robust genomic predictions. Several studies have demonstrated that optimizing the size and composition of training populations can directly influence the accuracy of prediction models, thereby enhancing the selection process [4].

In practical applications, breeders have employed a two-pronged approach whereby GS is first used to narrow down large populations to a manageable subset with high predicted values, which are then advanced to rigorous multi-environment field trials. This integration ensures that the final selection is both genetically superior and well-adapted to target environments.

Limitations of Genomic Selection in Wheat

Despite the notable successes of GS in wheat breeding, several limitations persist. The primary challenges revolve around balancing predictive accuracy with cost-effectiveness, data quality, and the inherent complexity of wheat genetics.

Model Limitations and Uncertainty

While advanced models such as deep learning show promise, they are not without caveats. One of the critical limitations is the risk of overfitting, particularly in cases where the training dataset is limited in size or diversity. Although deep learning models can capture complex relationships, ensuring that the models generalize well to novel datasets remains a challenge [2]. Moreover, the interpretability of these models is limited when compared to linear methods, making it more difficult for breeders to understand the biological rationale behind specific predictions.

Data Quality and Availability

The success of genomic selection heavily depends on the quality of genomic and phenotypic data available. In many breeding programs, especially those operating in resource-limited settings, data robustness and consistency can be a significant challenge. Missing data, genotyping errors, and environmental variability can degrade the predictive performance of even the most sophisticated models [4]. Standardization of data collection procedures and continuous data quality improvement are necessary to mitigate these issues.

Cost Considerations and Technical Requirements

Although genomic selection can lead to considerable cost savings in the long run, the upfront expenses related to high-throughput genotyping and data analysis infrastructure may be substantial. Breeding programs must make significant investments in computational resources and skilled personnel capable of implementing and managing complex prediction models [3]. Additionally, the cost-effectiveness of deep learning approaches in some scenarios might limit their immediate adoption in programs with strict budget constraints.

Trait Complexity and Genotype-by-Environment (G×E) Interactions

Wheat traits are often influenced by a myriad of genetic and environmental factors. The intricate interplay of genotype-by-environment interactions adds an additional layer of complexity to genomic prediction. Models that are primarily designed to capture additive effects may fail to fully account for these interactions, resulting in suboptimal predictions under variable field conditions. While attempts are being made to incorporate G×E interactions into GS models, these efforts require robust multi-environment datasets and enhanced modeling techniques [4].

Research Gaps and Future Directions

As genomic selection technologies evolve, several research areas require further exploration to fully exploit their potential in wheat improvement programs.

Enhancing Model Robustness and Interpretability

Future research should focus on improving the robustness and interpretability of advanced GS models. The incorporation of explainable artificial intelligence (XAI) methods could provide insights into the genetic basis of complex traits and potentially reveal underlying biological mechanisms. Bridging the gap between predictive performance and biological interpretability is crucial for increasing the acceptance of complex models among breeders [2].

Optimizing Training Population Design

Several studies have underscored the importance of training population size and diversity in maximizing the predictive accuracy of GS models. Future investigations should aim at identifying optimal strategies for training population selection that account for genetic diversity, environmental variation, and trait heritability. Innovative experimental designs and statistical methods that efficiently utilize available data are likely to further enhance the predictive performance of GS in wheat breeding [4].

Integration of Multi-Omic Data

While current GS methodologies primarily rely on genomic markers, the integration of multi-omic datasets – such as transcriptomic, proteomic, and metabolomic data – could offer a more comprehensive picture of the biological systems underlying complex traits. Such integration stands to improve prediction accuracy by capturing regulatory and metabolic pathways that influence phenotypes. However, this approach remains in its early stages and necessitates further research into multi-omic data integration techniques.

Addressing Environmental Variability

Genomic selection models must further evolve to account for environmental variability and genotype-by-environment interactions. Advanced modeling frameworks that incorporate environmental covariates or that utilize multi-environment trial data can help refine predictions under diverse climates and agronomic regimes. The development of such models will be particularly valuable in the context of climate change, where fluctuating environmental conditions pose significant challenges to wheat production.

Cost Reduction and Technological Adoption

Finally, another key research gap lies in reducing the costs associated with high-throughput genotyping and computational infrastructure. Continued improvements in sequencing technologies, coupled with the development of cost-effective statistical models, will enhance the accessibility of GS techniques. Research that focuses on minimizing overheads while maximizing genetic gain will further strengthen the case for adopting GS as a standard tool in wheat breeding programs [3].

Conclusion

The integration of genomic selection into wheat breeding represents a transformative advancement in crop improvement strategies. By utilizing dense genomic information and advanced computational methodologies, breeders can predict the performance of candidate lines with increased accuracy while significantly reducing the need for extensive phenotypic evaluations. Comparative analyses have demonstrated that while classical methods such as RR-BLUP remain effective for capturing additive effects, machine learning approaches, including deep learning, have shown superior potential for predicting complex quality traits.

Despite its clear advantages in accelerating breeding cycles and optimizing resource allocation, genomic selection is not without its challenges. Data quality, model interpretability, and the computational requirements of advanced methods pose significant barriers, particularly for breeding programs operating under constrained resources. Additionally, the inherent complexity of wheat traits and the impact of genotype-by-environment interactions necessitate further research to refine predictive models.

Future research should focus on enhancing model robustness, incorporating multi-omic datasets, and optimizing training population designs to fully leverage the power of genomic selection. With continued technological advancements and research innovation, GS is poised to play an increasingly central role in wheat crop improvement, ultimately contributing to global food security and sustainable agricultural practices.

References

[1] E. L. Heffner, M. E. Sorrells, and J. L. Jannink, "Genomic selection for crop improvement," Crop Science, vol. 51, no. 1, pp. 1–10, 2011.

[2] S. D. Battenfield, J. Poland, and J. L. Jannink, "Wheat quality improvement at CIMMYT and the use of genomic selection on it," Frontiers in Plant Science, vol. 7, pp. 1–10, 2016.

[3] K. S. Sandhu, A. H. Carter, and D. N. Lozada, "Genomic selection for agronomic traits in a winter wheat breeding program," Frontiers in Plant Science, vol. 14, pp. 1–12, 2023.

[4] B. B. Tessema, H. Liu, J. R. Andersen, and J. Jensen, "Strategies using genomic selection to increase genetic gain in breeding programs for wheat," Frontiers in Genetics, vol. 11, pp. 1–10, 2020.

[5] F. Maulana and M. E. Sorrells, "Genomic selection of forage quality traits in winter wheat," Crop Science, vol. 59, no. 5, pp. 1–10, 2019.

[6] L. F. Merrick, A. W. Herr, K. S. Sandhu, D. N. Lozada, and A. H. Carter, "Utilizing genomic selection for wheat population development and improvement," Agronomy, vol. 12, no. 2, p. 522, 2022.

[7] X. Hu, B. F. Carver, C. Powers, L. Yan, L. Zhu, and C. Chen, "Effectiveness of genomic selection by response to selection for winter wheat variety improvement," The Plant Genome, vol. 12, no. 3, p. 180090, 2019.

[8] J. Crossa, Y. Beyene, P. Pérez-Rodríguez, J. M. Hickey, B. Das, and R. P. Singh, "Genomic prediction in CIMMYT maize and wheat breeding programs," Hereditas, vol. 154, no. 1, pp. 1–12, 2017.