



POPULAR ARTICLE

Artificial intelligence and machine learning in genomic selection for crop improvement

T. Vijay¹, M.R. Madhusudhan², H. Akshata³, S.N. Navya⁴, R. Jayasankaramoorthy⁵

1 Department of Genetics and Plant Breeding, Junagadh Agricultural University, Junagadh 362001, Gujarat, India

2 Department of Genetics and Plant Breeding, University of Agricultural Sciences, GKVK, Bengaluru 560065, Karnataka, India

3 Departments of Genetics and Plant Breeding, Keladi Shivappa Nayaka University of Agricultural and Horticultural Sciences, Iruvakki-577412, Shivamogga, Karnataka, India

4 Departments of Genetics and Plant Breeding, Kerala Agriculture University, Vellayani 695522, Kerala, India

5 Department of Plant Pathology, Coffee Board, Kolli hills 637411, Namakkal, Tamil Nadu, India

Corresponding authors email: vijayraajputthakur51297@gmail.com

Manuscript received: August 13, 2026; Decision on manuscript, August 30, 2026; Manuscript accepted: September 9, 2026

Intricacies of the subject

Trehalose is a non-reducing disaccharide that functions as a multifunctional component of plant metabolic regulation and environmental adaptation. Although present at relatively low concentrations in most vascular plants, trehalose metabolism is associated with carbohydrate homeostasis, cellular protection, photosynthetic performance, nitrogen utilization and abiotic stress responses. Its relevance to carbon–nitrogen (C-N) coordination arises from the dependence of nitrogen assimilation on carbon-derived skeletons, cellular energy and reducing power. Exogenous trehalose has been reported to enhance nitrate and ammonium assimilation, activities of key nitrogen-assimilatory enzymes, chlorophyll content, photosynthesis and growth under nitrogen limitation.

Trehalose-associated responses to drought, salinity, temperature extremes and heavy-metal stress further involve osmotic adjustment, reactive oxygen species (ROS) homeostasis, antioxidant defence and preservation of cellular integrity. Trehalose-6-phosphate (T6P), an intermediate of trehalose biosynthesis, has an established role in carbon-status signalling, particularly through its relationship with sucrose and SnRK1; however, T6P and trehalose have distinct functions. This article focuses primarily on trehalose and summarizes its roles in carbon metabolism, C-N coordination, nitrogen assimilation, photosynthetic fitness, plant growth and abiotic stress adaptation. Collectively, available evidence supports trehalose as a metabolic hub that contributes to the integration of carbon availability, nitrogen utilization and stress responses.



Introduction

Genomic selection begins with a training population in which both the genetic profile and the actual performance of each individual are known. DNA markers are collected through genotyping, while traits such as grain yield, disease resistance or quality are recorded through field or controlled-environment evaluation. These data are used to train a statistical model that learns the relationship between marker patterns and observed performance. After training, the model can generate genomic estimated breeding values (GEBVs) for new individuals using their marker information, even before those individuals are evaluated extensively in the field. GBLUP (genomic best linear unbiased prediction) remains one of the most widely used approaches for genomic prediction. It is based largely on the assumption that many genes, each with relatively small effects, contribute to variation in a trait. Because of its strong performance and established use in breeding programs, GBLUP continues to serve as an important benchmark when newer AI-based prediction methods are evaluated.

Where machine learning improves on classical genomic prediction

Traditional prediction methods such as GBLUP and ridge-regression BLUP are powerful, but they generally rely on relatively simple relationships between markers and traits. In reality, important agricultural traits can be influenced by gene interactions, genotype-by-environment effects and other non-additive relationships. Machine learning offers a way to model some of these more complicated patterns without requiring breeders to define the genetic architecture in advance.

Statistical and ensemble machine learning

Methods such as random forests, support vector machines and gradient boosting can model non-linear relationships between genomic markers and traits. Their flexibility makes them useful when the relationship between genotype and phenotype is more complicated than a simple linear model can represent. The increasing availability of large genomic and phenomic datasets, together with more accessible software, is also making these approaches easier to explore in breeding programs.

Deep learning

Deep learning takes genomic prediction a step further by using architectures such as convolutional neural networks and multilayer perceptrons to learn complex patterns from marker data. Recent comparative work has examined deep-learning models alongside GBLUP across multiple plant breeding datasets, with results showing that deep learning can capture non-linear genetic patterns and may provide an advantage in some datasets. Models such as DeepGS and DNNGP have also been developed specifically for genomic prediction, although their performance depends on the data and model design.

Multi-omics and data fusion

A plant's performance is determined by much more than its DNA sequence. Gene expression, metabolites and the environment can all influence the final phenotype. For this reason, breeding research is increasingly moving toward the combined analysis of genomic, transcriptomic, metabolomic and high-throughput phenotypic information. AI and ML are well suited to this type of data integration because they can handle large and diverse datasets within a common predictive



framework. Multi-omics selection is therefore emerging as an important direction in data-driven crop improvement. data integration because they can handle large and diverse datasets within a common predictive framework. Multi-omics selection is therefore emerging as an important direction in data-driven crop improvement.

Practical applications in breeding programs

Trait prediction and parent selection: AI models can use genomic information to help identify individuals likely to carry useful traits such as drought tolerance, disease resistance or improved grain quality. This allows breeders to narrow down potential parents before investing resources in extensive field trials. Disease resistance screening: Image-based AI, including convolutional neural networks, can identify and classify disease symptoms from plant images. Digital phenotyping can identify and classify disease symptoms from plant images. Digital phenotyping can reduce the dependence on repeated manual scoring and can provide faster and more consistent measurements. Hybrid performance forecasting: Machine learning can learn from historical multi-environment trial data to estimate the likely performance of untested hybrid combinations. Such predictions can help breeders decide which crosses deserve further evaluation. Marker-assisted and genomic selection pipelines: AI-supported software is also making genomic prediction workflows more accessible, helping breeding programs manage and interpret large datasets more efficiently.

From the field: why this matters for breeders: The greatest practical benefit of these approaches is the possibility of reducing the number of breeding cycles required to identify and fix useful combinations of genes. Earlier marker-assisted selection work, for

example, helped reduce the number of generations needed for developing some maize hybrids. AI-enhanced genomic selection extends this idea by allowing breeders to evaluate large populations computationally and prioritize the most promising lines for field testing. Instead of treating every breeding line as equally deserving of extensive evaluation, breeders can use prediction as an early filter. Thousands of candidates can be screened using genomic information, while the time, land, labour and cost required for detailed field testing can be concentrated on the most promising material.

Challenges and open questions: AI-driven genomic selection is not a simple plug-and-play technology. Its success depends heavily on the quality and size of the training data. Deep learning models, in particular, generally benefit from large and well-characterized datasets, while many breeding programs working with minor or regionally important crops do not yet have sufficient genomic and phenotypic information. Another important issue is interpretability. Classical statistical models are often easier to understand, whereas complex deep-learning models can behave like a black box, making it difficult to explain why a particular genotype receives a high prediction. In addition, better performance in one dataset does not necessarily guarantee better results across all crops and breeding environments. AI and ML therefore need to be evaluated alongside well-established prediction methods under realistic breeding conditions.

The road ahead

The next stage of crop improvement is likely to involve closer integration of AI with genomic selection, high-throughput phenotyping and environmental information. Such integration could move breeding programme towards more predictive decision



making, where genomic and environmental data are considered together rather than relying only on repeated field evaluation. Researchers are also beginning to explore large-scale foundation models and other advanced architectures for biological sequence analysis. Although these approaches are still developing, they point toward a future in which computational tools can help breeders interpret increasingly complex biological datasets. For a discipline traditionally measured in growing seasons, the ability to make reliable predictions earlier in the breeding cycle could substantially shorten the path from genetic variation to improved varieties.

Conclusion

Artificial intelligence and machine learning are not replacing the basic principles of plant breeding. Instead, they are providing breeders

with new ways to apply those principles more effectively. Ensemble learning can improve genomic prediction, deep learning can explore complex genetic relationships, image-based systems can support disease screening, and multi-omics approaches can bring efficiency. Ensemble learning can improve genomic prediction, deep learning can explore complex genetic relationships, image-based systems can support disease screening, and multi-omics approaches can bring different layers of biological information together. The real strength of these technologies will come from combining computational predictions with biological knowledge and careful field validation. As datasets improve and prediction models become more practical, AI-assisted genomic selection has the potential to help breeders identify promising material earlier and develop high-performing, resilient crop varieties more efficiently.

References

1. Ali, Q. and Ashraf, M. 2011. Induction of drought tolerance in maize (*Zea mays* L.) due to exogenous application of trehalose: Growth, photosynthesis, water relations and oxidative defence mechanism. *J. Agron. Crop Sci.*, 197(4): 258–271.
2. Elkelish, A., Alhudhaibi, A.M., Hossain, A.S., Haouala, F., Alharbi, B.M., El-Banna, M.F., Rizk, A., Badji, A., AlJwaizea, N.I. and Sayed, A.A.S. 2024. Alleviating chromium-induced oxidative stress in *Vigna radiata* through exogenous trehalose application: Insights into growth, uptake, and reactive oxygen species scavenging enzyme activity enhancement. *BMC Plant Biol.*, 24: 460.
3. Figueroa, C.M. and Lunn, J.E. 2016. A tale of two sugars: Trehalose 6-phosphate and sucrose. *Plant Physiol.*, 172(1): 7–27.
4. Hassan, M.U., Nawaz, M., Shah, A.N., Raza, A., Barbanti, L., Skalicky, M., Hashem, M., Brestic, M., Pandey, S., Alamri, S., Mostafa, Y.S., El Sabagh, A.E.L. and Qari, S.H. 2023. Trehalose: A key player in plant growth regulation and tolerance to abiotic stresses. *J. Plant Growth Regul.*, 42(8): 4935–4957.