Beyond DNA Markers: Making Genomic Selection More Accessible to Small-Scale Crop Breeding Programmes

Kashish Preet Kaur[1]*, Lavanya Prabha Sahu[2] and Hemlata Ohdar[3]

[1,2,3]*Indira Gandhi Krishi Vishwavidyalya, Krishak Nagar, Raipur, (C.G.), 492012*

Corresponding Author
Kashish Preet Kaur
Email: kashishpreetkaur51@gmail.com

OPEN ACCESS

Keywords

Genomic selection, pulse crop improvement, plant breeding, genomic prediction, training population, cost-effective genotyping, climate resilience

How to cite this article:

Kaur, K. P., Sahu, L. P. and Ohdar, H. 2026. Beyond DNA Markers: Making Genomic Selection More Accessible to Small-Scale Crop Breeding Programmes. *Vigyan Varta* 7 (11): 185-193.

ABSTRACT

Genomic selection (GS) offers a promising approach to modern plant breeding by using genome-wide molecular marker information to predict the genetic potential of candidate plants. It is particularly relevant to complex traits such as grain yield, drought tolerance, disease resistance and adaptation to variable environments, which are often influenced by many genes and environmental factors. However, the adoption of GS in small-scale crop breeding programmes remains challenging because of genotyping costs, limited training populations, inadequate infrastructure and shortages of specialised expertise. These constraints are particularly important for pulse crops, including chickpea, pigeon pea, lentil and mung bean, which are essential sources of dietary protein and contribute to sustainable agricultural systems. This article examines practical ways to make GS more accessible through shared genotyping facilities, carefully designed training populations, cost-effective marker technologies, reliable field phenotyping, collaborative research and accessible computational tools. It also considers the limitations of genomic prediction, including reduced accuracy when candidate plants differ substantially from the training population or when environmental conditions change. Rather than replacing conventional breeding, GS can complement field evaluation and established selection methods. A collaborative, crop-specific approach could help smaller breeding programmes use genomic information more

efficiently and make progress towards improved productivity, resilience and nutritional quality.

INTRODUCTION: A New Direction for Crop Improvement

Crop improvement has always depended on identifying plants with desirable characteristics and using them to develop better varieties. For generations, plant breeders have selected plants with higher yields, improved grain quality, resistance to diseases and better adaptation to local conditions. Although conventional selection remains essential, developing superior varieties can be time-consuming, especially when the traits of interest are complex and strongly influenced by the environment.

Modern genetics has expanded the tools available to breeders. Molecular markers allow researchers to study differences in DNA among plants and use this information in selection decisions. One important development is genomic selection (GS), which uses information from markers distributed across the genome to predict the genetic potential of breeding candidates.

The significance of this approach lies in its potential to support selection before every candidate has undergone extensive field evaluation. However, genomic selection is not simply a matter of collecting DNA data and identifying the best plant. Its success depends on good-quality phenotypic data, a suitable training population, appropriate statistical models and validation under relevant growing conditions (Jannink *et al*., 2010; Escamilla *et al*., 2025).

A central question therefore remains: How can the benefits of genomic selection be extended beyond well-resourced breeding programmes to smaller public institutions, regional research centres and programmes working with crops that receive comparatively limited research investment?

Answering this question is particularly important for pulse crops, where breeding programmes must improve productivity and resilience while working within practical financial and infrastructure constraints.

1. Understanding Genomic Selection: Looking Beyond Individual DNA Markers

To understand genomic selection, it is useful to distinguish it from marker-assisted selection (MAS).

Marker-assisted selection generally uses markers associated with particular genes or genomic regions of interest. It is especially useful when a target gene or region has a sufficiently reliable association with a trait. For example, a breeder may use a validated marker linked to a disease-resistance gene to help identify plants carrying that resistance.

Genomic selection follows a broader approach. Instead of relying only on a small number of selected markers, it uses genome-wide marker information to estimate the genetic merit of individuals. This is particularly useful for complex traits, such as yield, that are commonly influenced by many genetic factors, each potentially contributing a relatively small effect.

How does genomic selection work?

The process can be understood through four main steps.

1. Step 1: Develop a training population. A representative group of plants is evaluated

for the traits of interest and genotyped using molecular markers. The quality of the field observations is critical because the model learns from these data.

2. Step 2: Build a prediction model. Statistical methods use the relationship between marker information and observed performance to estimate patterns that can help predict genetic merit.

3. Step 3: Predict candidate performance. Marker data from new candidate plants are analysed using the trained model. The model produces genomic predictions, commonly expressed as genomic estimated breeding values (GEBVs), depending on the statistical framework and breeding objective.

4. Step 4: Select and validate candidates. Breeders use the predictions alongside field observations, agronomic knowledge and other selection criteria to identify promising material. Selected candidates must still be evaluated appropriately before decisions about advancement or variety release.

The key advantage is that the breeder may be able to screen a larger set of candidates or make some selection decisions earlier in the breeding cycle. This can improve the use of resources when the model predicts the target trait reliably (Jannink *et al*., 2010; Sandhu *et al*., 2022).

However, a genomic prediction is not a direct measurement of yield, drought tolerance or disease resistance. It is an estimate based on available data. Its accuracy must be assessed rather than assumed.

2. Why Small-Scale Breeding Programmes Matter

Crop improvement is not carried out only by large international research centres or commercial seed companies. Public agricultural universities, regional breeding stations and national crop research programmes play important roles in developing varieties suited to specific agroclimatic conditions and farming systems.

Smaller programmes may work with limited budgets, modest field facilities and restricted access to high-throughput genotyping. They may also focus on crops or local varieties that do not attract the same level of commercial investment as major cereals.

Yet these programmes can have a direct connection with farmers' needs. They may understand local production constraints, preferred grain characteristics, cropping patterns and the stresses that frequently affect regional agriculture.

Pulse crops illustrate why this work matters. Chickpea, pigeon pea, mung bean, urd bean and lentil contribute dietary protein and can form important components of crop rotations. As legumes, many pulse crops also participate in symbiotic nitrogen fixation, although the extent of this contribution depends on the crop, soil conditions, compatible rhizobia and crop management.

Pulse breeding objectives may include higher and more stable yield, early maturity, resistance to major diseases and insect pests, tolerance to heat or drought, improved seed quality and suitability for local farming systems. Progress in these areas can support both farm productivity and food and nutritional security.

Genomics-assisted breeding has already received considerable attention in major pulse crops, but genomic resources and research capacity are not equally available for every crop or breeding programme (Bohra *et al*., 2014). Making genomic selection more accessible should therefore be viewed as a

practical breeding challenge, not merely a technological one.

3. Making Genomic Selection More Accessible

Wider adoption of genomic selection does not necessarily require every breeding programme to establish a complete in-house genomics laboratory. A more realistic approach is to identify which parts of the process need local capacity and which can be supported through collaboration.

3.1 Shared genotyping facilities and research partnerships: Genotyping infrastructure can be expensive to establish and maintain. Equipment, laboratory consumables, quality-control procedures, data storage and trained personnel all require investment.

One possible solution is to develop shared facilities through agricultural universities, public research institutions, national crop research programmes and collaborative research networks. A smaller breeding programme could collect plant samples and arrange genotyping through a partner facility rather than independently maintaining every component of the laboratory workflow.

Such arrangements may reduce duplicated costs and allow several programmes to benefit from the same technical resources. However, they require clear agreements on sample handling, data ownership, timelines, quality standards and access to results.

The aim should not be to outsource all scientific decision-making. Local breeders must remain involved in defining breeding objectives, interpreting predictions and deciding which plants merit further evaluation.

3.2 Choosing a cost-effective genotyping strategy*:* Not every breeding programme needs the highest possible marker density or the most expensive sequencing technology. The appropriate strategy depends on the crop, available genetic resources, the diversity of the breeding population, the target traits and the intended use of the predictions.

Depending on these factors, programmes may investigate suitable SNP arrays, reduced-representation sequencing, genotyping-by-sequencing or other validated marker platforms. Low-density genotyping may also be considered where an appropriate reference population and a reliable imputation strategy are available.

However, a lower-cost technology is not automatically an effective one. Marker coverage, data quality, missing genotypes, population structure and the relationship between training and candidate plants can all affect prediction performance.

Before adopting a genotyping platform, a programme should compare the total cost with the quality and usefulness of the information it produces. The goal is to obtain sufficient reliable information for the breeding objective, not simply to generate the largest possible dataset.

3.3 Building useful training populations: The training population is one of the most important components of genomic selection. A prediction model can only learn useful relationships from the data available to it.

A training population should represent the genetic material to which the model will be applied. If the candidate plants are genetically distant from the training population, or if the genetic relationships between them are weak, prediction accuracy may decline.

For smaller programmes, a well-designed training population may be more valuable than a much larger but poorly characterised one. Breeders can consider combining historical

field data, existing breeding material and newly evaluated lines, provided the records are comparable and the data quality is adequate.

Collaborative training populations may be useful when several institutions work on the same crop or related breeding objectives. Such collaboration can broaden the genetic diversity represented in the data and share the cost of phenotyping and genotyping.

Nevertheless, combining data from different programmes requires care. Differences in field protocols, scoring systems, environments and experimental designs can introduce inconsistencies. Data harmonisation and appropriate statistical analysis are therefore essential.

3.4 Treating phenotyping as a priority: A common misconception is that genomic selection reduces the need for field experiments. In reality, reliable phenotyping remains essential because the prediction model needs dependable observations from which to learn.

For complex traits such as yield, drought response and disease resistance, field performance is affected by the interaction between genotype and environment. A plant that performs well in one location or season may not perform equally well under different conditions.

Small-scale programmes can improve phenotyping by using consistent protocols, suitable experimental designs, appropriate checks, careful record keeping and accurate identification of each line. Data from multiple seasons or representative environments may be particularly valuable when the objective is broad adaptation or yield stability.

Digital tools, including mobile data collection and image-based phenotyping, may help improve the efficiency of data collection where they are affordable and appropriate. More advanced sensing tools can also be useful, but they should be adopted only when they provide dependable information relevant to the breeding objective.

Genomic selection cannot compensate for poor field data. Reliable phenotyping is the foundation on which useful genomic predictions are built (Sandhu *et al*., 2022).

3.5 Using accessible computational tools and developing expertise: Genomic prediction involves data processing, statistical analysis, model fitting and validation. These tasks require a combination of plant-breeding knowledge, quantitative genetics and computational skills.

Open-source statistical software and established genomic prediction packages can lower software costs. However, access to a tool does not automatically provide the expertise needed to use it correctly. Breeders and researchers must understand data quality, model assumptions, validation methods and the interpretation of prediction accuracy.

Training workshops, university collaborations, shared analytical support and interdisciplinary teams can help build these skills. A practical strategy is to begin with a clearly defined breeding objective and a manageable dataset, then expand as technical capacity develops.

Model selection should be based on performance under suitable validation rather than on the popularity or complexity of a particular algorithm. A sophisticated machine-learning model is not necessarily better than a simpler statistical model for every crop or dataset.

4. Applications in Pulse Crop Improvement

Pulse crops provide important opportunities for exploring genomic selection because many traits of interest are complex and influenced by multiple genes and environmental conditions.

Nevertheless, the feasibility of genomic selection must be assessed separately for each crop, population and breeding objective.

4.1 Chickpea

Chickpea improvement may target grain yield, maturity, seed quality, resistance to diseases such as Fusarium wilt, and adaptation to moisture-limited environments. Genomic prediction could help prioritise candidate lines for further testing when a suitable training population and reliable phenotypic data are available.

For drought-related traits, the breeding programme must define what it intends to improve. Selection for yield under water-limited conditions, for example, is not identical to selection for a physiological measure of drought response. Predictions should therefore be evaluated against the specific breeding objective.

4.2 Pigeon pea

Pigeon pea is important in several tropical and subtropical farming systems. Breeding objectives may include higher yield, improved maturity duration, resistance to major diseases and insect pests, and adaptation to different production environments.

Genomic selection could be explored to rank candidate lines for complex traits or to support selection at particular stages of a breeding programme. However, it would require suitable genotyping data, well-characterised populations and reliable field evaluation. It should not be assumed that a model developed for one pigeon pea population will perform equally well in another.

4.3 Mung bean and lentil

In mung bean and lentil, breeding programmes may seek improvements in yield, maturity, seed characteristics, disease resistance and adaptation to local growing conditions.

Genomic prediction may be particularly worth investigating where breeders routinely evaluate large numbers of lines or where field evaluation is expensive and time-consuming. Its practical value would depend on the quality of existing records, the genetic diversity of the material and whether the predictions improve selection decisions compared with current methods.

4.4 A shared opportunity, not a universal solution

Research on genomics-assisted breeding in pulse crops demonstrates the growing importance of molecular tools for crop improvement (Bohra *et al.*, 2014). However, the existence of genomic resources does not mean that every pulse crop already has a validated, affordable genomic-selection system.

For each crop, researchers need to establish whether the available population, marker data and phenotypic records are adequate for the intended application. Where these conditions are not met, building reliable foundational data may be more useful than immediately introducing a complex prediction model.

5. Challenges That Cannot Be Ignored

Genomic selection offers potential benefits, but several limitations must be addressed before it can be implemented effectively in smaller breeding programmes.

5.1 Prediction accuracy is not guaranteed

Prediction accuracy depends on factors such as the size and composition of the training population, genetic relationships between training and candidate plants, marker coverage, trait heritability, data quality and the prediction model used.

A model that performs well in one breeding cycle may lose accuracy when applied to a different population or generation. Regular validation and, where necessary, updating of the training population are important.

5.2 Environmental variation complicates selection

Traits such as yield and stress response can vary across locations, seasons and management systems. Genotype-by-environment interaction may affect both observed performance and the accuracy of predictions.

Where possible, training data should represent the target environments. Breeders should also examine whether predictions remain useful across relevant locations or seasons instead of relying on a single overall accuracy estimate.

5.3 Initial investment and operational costs remain important

Although some technologies and software options may reduce costs, genomic selection still requires funding for genotyping, phenotyping, data management, analysis and staff training.

A programme should assess the full cost of implementation and compare it with the expected benefit. Savings should not be assumed merely because selection decisions can be made earlier.

5.4 Data governance and continuity matter

Collaborative programmes need reliable systems for storing, documenting and sharing data. Missing records, inconsistent line identification, unclear ownership and interrupted funding can undermine long-term progress.

Agreements on data access, authorship, sample use and responsibilities should be established before large collaborative datasets are assembled.

5.5 Genomic selection must serve the breeding objective

Prediction accuracy is not the only measure of success. A breeding programme ultimately needs to develop varieties that meet farmers' and users' needs.

Selection decisions must consider the balance among yield, stability, disease resistance, maturity, seed quality, nutritional characteristics and other relevant traits. A high genomic prediction for one trait does not automatically identify the best overall variety.

6. A Practical Roadmap for Smaller Breeding Programmes

A phased approach can help smaller institutions evaluate genomic selection without making unnecessarily large initial investments.

1. First, define a specific breeding problem. Identify the crop, target population, trait and production environment. A clearly defined objective makes it easier to assess whether genomic selection is suitable.

2. Second, assess existing resources. Review historical field data, available breeding lines, marker information, field facilities and staff expertise. Good existing records can provide a valuable starting point.

3. Third, strengthen phenotyping. Standardise field protocols, maintain accurate line identification and ensure that observations are collected using appropriate experimental designs.

4. Fourth, develop a representative training population. Select material that reflects the candidate population and the intended breeding objective. Where appropriate, collaborate with other institutions to broaden the available data.

5. Fifth, choose a proportionate genotyping strategy. Compare available technologies, data quality, total cost and the intended use of the predictions. Seek external support where establishing in-house facilities would not be practical.

6. Sixth, validate predictions before relying on them. Use a suitable validation strategy that reflects the intended application. Where possible, evaluate performance in material or environments that were not used to train the model. This helps provide a more realistic assessment of its usefulness.

7. Seventh, compare genomic selection with current practice. Assess whether it improves selection decisions, reduces unnecessary evaluation costs, shortens relevant stages of the breeding cycle or helps identify promising material more effectively. The comparison should consider both the benefits and the additional costs.

Finally, expand only when the evidence supports it. If the approach provides useful results, the programme can gradually extend it to additional traits, breeding stages or populations. If prediction performance is weak, improving the training data or phenotyping system may be the appropriate next step.

This roadmap places breeding objectives and evidence ahead of technology adoption. It also allows institutions to learn through manageable projects rather than treating genomic selection as an all-or-nothing investment.

CONCLUSION: Making Advanced Breeding More Inclusive

Genomic selection represents an important development in plant breeding because it uses genome-wide marker information to predict the genetic potential of candidate plants. For complex traits, it can support more informed selection and may help breeding programmes evaluate candidates more efficiently.

However, its benefits depend on more than the availability of molecular markers. Suitable training populations, reliable phenotyping, appropriate prediction models, relevant validation and sustained technical capacity are all essential. These requirements are especially important for small-scale breeding programmes, where resources must be used carefully. Pulse crops provide a valuable context for exploring these possibilities. Improvements in chickpea, pigeon pea, mung bean, lentil and other legumes can contribute to more productive and resilient farming systems. Yet genomic selection should be adopted crop by crop and objective by objective, based on evidence that it improves breeding decisions. The way forward is not to replace conventional breeding with technology, but to combine established field expertise with useful genomic information. Shared facilities, collaborative datasets, accessible computational tools and investment in local scientific capacity can help make this combination more feasible. The future of genomic selection should not be measured only by how advanced the technology becomes, but also by how effectively it helps breeding programmes develop better crops under real-world conditions.

REFERENCES

Bohra, A., Pandey, M. K., Jha, U. C., Singh, B., Singh, I. P., Datta, D., Chaturvedi, S. K., Nadarajan, N., & Varshney, R. K. (2014). Genomics-assisted breeding in four major pulse crops of developing countries: Present status and prospects. Theoretical and Applied Genetics, 127, 1263–1291. https://doi.org/10.1007/s00122-014-2301-3

Escamilla, D. M., Li, D., Negus, K. L., Kappelmann, K. L., Kusmec, A., Vanous, A. E., Schnable, P. S., Li, X., & Yu, J. (2025). Genomic selection: Essence, applications, and prospects. The Plant Genome, 18(2), e70053. https://doi.org/10.1002/tpg2.70053

Jannink, J.-L., Lorenz, A. J., & Iwata, H. (2010). Genomic selection in plant breeding: From theory to practice. Briefings in Functional Genomics, 9(2), 166–177. https://doi.org/10.1093/bfgp/elq001

Sandhu, K. S., Merrick, L. F., Sankaran, S., Zhang, Z., & Carter, A. H. (2022). Prospectus of genomic selection and phenomics in cereal, legume and oilseed breeding programs. Frontiers in Genetics, 12, 829131. https://doi.org/10.3389/fgene.2021.829131