Popular Agriculture 2(3) 53-57(May, 2026)
Published at popularagriculture.com

Global E-Magazine

POPULAR ARTICLE

Genes, genomes and numbers: how statistics drives modern plant breeding

T. Vignesuwar, K. R. Aadithyan

School of Agricultural sciences, Karunya Institute of Technology and Sciences, Coimbatore – 641114 Tamil Nadu, India

Corresponding authors email: ttvignesuwar@gmail.com

Manuscript received: April 18, 2026; Decision on manuscript, April 28, 2025; Manuscript accepted: May 15, 2026

Intricacies of the topic

Throughout the history of plant breeding, the objective has been to choose the superior plants for superior crops. Over the years, breeders have developed crop varieties that are higher yielding, of higher quality and resistant to pests and diseases by using careful observation, experience and multiple field trials. In today's day and age, however, agriculture is on a new path where the decision making is no longer made just on what can be seen in the field. With the advances of genomics, the genetic information of plants has become practically limitless, with breeders now able to peer directly into a plant's genes and understand what genes are responsible for significant traits. Decoding a genome is just the first step. With modern breeding there are millions of genetic markers and huge amounts of data that must be statistically analysed to interpret the data. From the raw genomic data to the actual breeding decisions, statistics has emerged as the link between the two stages, enabling scientists to select the best performing genotypes and to better predict the result of breeding and to speed up the process of developing climate-resilient, high-yielding and nutritionally better crops. In the era of AI, big data, and precision breeding, the combination of genes, genomes, and numbers is revolutionizing plant breeding by shifting it from a field-driven science to a data-driven practice, all with the promise of a more sustainable and food-secure future.

The evolution of plant breeding: From selection to genomics

The breeding of plants has come a long way in the last few hundred years. Early farmers were able to select the healthiest and best yielding plants and use their seeds to improve their crops. When Mendelian genetics were discovered, this practice became a science and breeders were able to create better varieties using controlled hybridization. In the future, quantitative genetics and statistical analyses aided breeders in more accurately analysing complex traits like yield, stress tolerance, and disease resistance. Crop improvement has progressed a further step in the genomic era. Recent progress in the areas of DNA sequencing, molecular markers and genomic selection mean that breeders can now choose desirable traits directly from a plant's genome, which shaves time off the breeding process and makes the process more efficient. Now that genomic information is playing an important role in the plant breeding, the problem is not the production of data, it is the interpretation of data. Here, statistics can be invaluable, enabling millions of data points on DNA to be used in making meaningful breeding decisions.

Decoding the genome: The language of DNA

Each plant has a distinct genome or full complement of DNA that controls its growth and development, yield, resistance to disease, and responses to the environment. Scientists have been using a new generation of technologies involving genome sequencing and molecular markers to determine the genes involved in these desirable traits with unprecedented accuracy. Breeders can now look at a crop's genetic make-up at an early stage, rather than waiting for mature plants to show desirable traits, to speed up and make crop improvement more efficient. But a genome's decoding generates a lot of data, however, typically thousands of plants and millions of DNA markers. These gene sequences, taken individually, tell little or nothing. The real power of their use lies in the application of statistical techniques to detect meaningful patterns, to foretell breeding efficiency, and to tease out genetic factors from environment factors. Today, the genome carries the information in plant breeding, and statistics turns information into knowledge to allow plant breeders to make informed decisions based on data.

From data to decisions: Why statistics matters in plant breeding

Modern Plant Breeding involves not only genetics, but statistics. The breeding programmes yield enormous amounts of data from field experiments, molecular markers and genome sequencing. The raw data, however, does not allow identification of the best-performing genotype nor does it predict its performance under a variety of environmental conditions. These observations are then analysed statistically to transform them into reliable scientific evidence that allows breeders to differentiate between the genetic variation and environmental variation, and to make informed selection decisions. In crop improvement, statistics is used at every step of the process, ranging from planning field experiments and estimating heritability, to analysing G×E interactions, the selection of superior parental lines and the prediction of breeding values by genomic selection. Analysis of Variance (ANOVA), correlation, regression and stability analysis are among the techniques that assist breeders in assessing complex traits such as yield, drought resistance, disease resistance and nutritional quality with greater precision. Today, plant breeding is faster, more precise and more efficient, using statistical modelling to incorporate genomic data and produce higher-yielding and more climate-resistant cultivars.

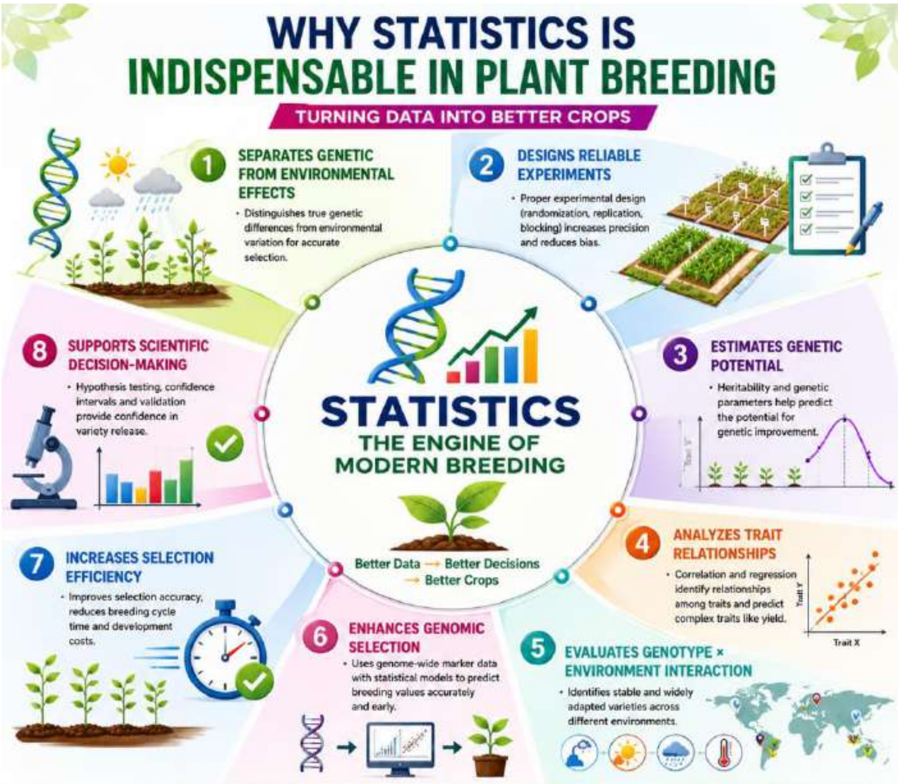
Genomic statistics: Turning DNA into decisions

The pace of genome sequencing has increased so quickly that plant breeders are now able to capture much more genetic data, including millions of DNA markers linked to desirable traits, for example, in crops. But these markers are of little use if they cannot be connected to meaningful aspects of biology. Genomic statistics fills this void by combining statistical models and genomic data to identify useful genes, to estimate the impact of these genes, and to make predictions of what a breeding line will be able to do based on the genomic data with an extremely high level of accuracy. Breeders are able to estimate the genome wide marker estimated breeding value (GEBV) of plants without the need for long field trials by combining GWMs with prediction models. This data-driven approach enhances the accuracy of selection, reduces breeding cycles and speeds up the production of better crop varieties that yield more, are more resitant to disease and climate stresses and have better nutritional quality. In the era of precision agriculture, genomic statistics is playing an increasingly vital role in next-generation crop improvement, enabling farmers to make accurate breeding choices from complex genomic data.

Table 1: Role of statistics in plant breeding and genomics techniques

Genomic data	Statistical role	Breeding benefit
DNA markers (SNPs)	Detect marker–trait associations	Identify desirable genes
Whole-genome data	Predict breeding values (GEBV)	Early selection of superior plants
Multi-environment data	Model genotype × environment interactions	Develop stable, widely adapted varieties
Phenotypic + genomic data	Build genomic prediction models	Higher selection accuracy
Large breeding populations	Analyze complex traits	Faster genetic gain

Fig: Figure showing the importance of statics



Accelerating crop improvement with genomic selection

In traditional plant breeding, it takes several crop seasons to assess performance of traditional varieties and choose superior plants for the next round of varieties. However, with the advent of Genomic Selection (GS), breeders can now estimate the genetic potential of a plant, before it is even mature, by using genome-wide DNA markers. Genomic Estimated Breeding Value (GEBV) is calculated using a combination of genomic data and powerful statistical models, enabling the selection of the best individuals at an early stage. Genomic Selection works differently than Marker-Assisted selection (MAS) that works on a few genes affecting large traits, because it uses thousands of genes that affect complex traits, such as yield, drought tolerance, disease resistance, and nutritional quality. This will bring about a substantial reduction in breeding cycles, higher selection accuracy and a faster genetic improvement. GS is now an established method in several major crops like maize, wheat, rice, barley, and soybean and is one of the most powerful technologies for producing climate-smart, high-yielding, and sustainable plant varieties. Plant breeders are still struggling to develop new cultivars at the rate they would like to. The rate at which plant breeders can develop new cultivars is still lagging. Genomic Selection is transforming Plant Breeding by early selection, selection based on DNA without full plant growth, increased accuracy of prediction for polygenetic, complex traits, shorter breeding cycles, that is, a shorter time required for releasing new varieties, more genetic gain – through the selection of superior lines more efficiently, cost-effective as it does not require multiple season's field evaluations, contributes to drought, heat, and salinity resistant climate-smart breeding and disease resistance and it seamlessly integrates with AI, genomic statistics, for next-generation precision breeding.

AI, big data and the future of plant breeding

The combination of artificial intelligence (AI), big data and genomics is revolutionizing plant breeding into a very data- and prediction-driven science. Milling tons of data from genome sequencing, molecular markers, field trials, drones, sensors, satellite imaging, and high-throughput phenotyping platforms, today's breeding programmes produce vast amounts of data. These various sets of data can be used by AI and machine learning algorithms to swiftly analyze the data, discovering relationships between genes and traits, forecasting breeding results and selecting the best parental breeding with greater speed and accuracy than traditional analytical methods. In addition to analysing the genomic information, AI can combine phenotypic, environmental, and management data to assist in predictive breeding, allowing breeders to make predictions on how crop varieties will perform under various climatic and growing conditions, prior to comprehensive field testing. The integration of genomics, transcriptomics, proteomics, and metabolomics into comprehensive decision-support systems has further contributed to breeding efficiency with the help of emerging technologies like digital twins, multi-omics integration, and advanced data analytics. These innovations are helping to shorten the time and expense of breeding, and to speed up the development of high-yielding, climate-resilient, disease resistant and nutritious crop varieties. With the continued evolution of these computational technologies, the combination of AI, Big Data, genomics, and statistics will shape the future of plant breeding, leading to more sustainable and resilient agriculture.

Challenges in the age of big data

While genomics, statistics and AI have revolutionised plant breeding, the process also has some tricky challenges. To manage large volumes of these data, to create advanced statistical tools, and to apply computational resources to genetics, statistics, and bioinformatics, it is necessary to have a high level of expertise, a strong team of scientists and engineers, and advanced computational infrastructure, as well as standardized data management. Maintaining data quality and correct interpretation of complex data is still essential to make appropriate breeding decisions. However, data-driven breeding remains limited due to the high cost of these advanced technologies, lack of computational power, and issues with sharing data ownership rights. It will be crucial to overcome these challenges by investing in infrastructure, collaboration among disciplines and capacity building if the potential of genomics and big data is to be realised in the creation of resilient crop varieties with high yield.

Hence in conclusion, future of plant breeding is now not anymore dependent on the field of genetics but more on the strength of genes, genomes and statistics. DNA forms the map of crop improvement and statistical techniques help derive information from large amounts of genomics data, helping plant breeders make faster and more accurate decisions. Artificial intelligence and digital tools, in addition to this method of deriving valuable knowledge from data, are making it possible for crop breeders to produce climate-resilient crops which yield more and have higher nutritional values. As we live in an era of climate change where there is increasing pressure on food supply, data has become as much of an asset as the seed itself. Through biological innovation and statistical intelligence, plant breeding is becoming a predictive science.