GMR
The Original

OMICS- AND AI-ASSISTED BREEDING IN PULSE CROPS

Sudharsan, S[1], Chandirakala, R[1*], Anandhi, K[2], Karthiba, L[2], Shanmugam, P.S[2], Manivannan, N[2], Ramya S[3].

[1]Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India
[2]Department of Pulses, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India
[3]Department of Fruit Science, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India
*Corresponding Author Mail ID: chandirakala2009@gmail.com

Abstract
Pulses, including chickpea, pigeonpea, lentil, mungbean, blackgram, pea, cowpea and common bean, are an affordable source of dietary protein and enhance soil fertility through biological nitrogen fixation. The productivity of pulses remains low owing to drought, heat stress, diseases, insect pests, a narrow genetic base and long breeding cycles. Recent advances in the application of omics tools and artificial intelligence (AI) to pulse breeding are summarised herein. Reference genomes are now available for most major pulses, ranging in size from less than 0.6 Gb in mungbean and blackgram to approximately 13 Gb in faba bean. In addition, 3,366 chickpea accessions have been resequenced. Pan-genomes, single-nucleotide polymorphism (SNP) arrays, quantitative trait locus sequencing (QTL-seq) and genome-wide association studies (GWAS) have accelerated gene and marker discovery. Marker-assisted backcrossing has delivered Fusarium wilt-resistant chickpea varieties and drought-tolerant introgression lines in India. Transcriptomics, proteomics, metabolomics and phenomics provide complementary insights into gene function and plant performance under stress. Genomic selection (GS) studies in chickpea, pea, common bean and lentil have reported prediction accuracies of 0.14–0.63 for grain yield and up to 0.91 for seed weight. Machine learning (ML) and deep learning (DL) are increasingly applied to genomic prediction, image-based phenotyping, disease diagnosis and the integration of genomic, phenomic and environmental data. These models do not consistently outperform conventional linear models. Major challenges include small training populations, poor-quality phenotypic data, limited computational infrastructure and skilled personnel, and the need for explainable models. Integrating these tools with speed breeding, genome editing and envirotyping can shorten breeding cycles and accelerate the development of climate-resilient, nutritionally enhanced pulse varieties.

Keywords: pulse crops; genomic selection; pan-genome analysis; marker-assisted selection; machine learning; high-throughput phenotyping; climate resilience.

1. Introduction

Pulses, the dry seeds of grain legumes, are eaten almost every day in India and in many parts of Asia and Africa (Considine et al., 2017; Semba et al., 2021). They are a cheap source of protein, fibre, minerals and vitamins, and in vegetarian and low-income households they often supply most of the protein in the diet (Foyer et al., 2016; Semba et al., 2021). Being legumes, they also fix nitrogen from the air with the help of rhizobia. This cuts down the fertiliser bill and leaves the soil in better shape for the crop that follows (Stagnari et al., 2017). Chickpea alone is cultivated in over fifty countries, and India is both its biggest producer and its biggest consumer (Merga and Haji, 2019).

Even so, pulse yields have not grown as fast as those of rice and wheat (Varshney et al., 2018). Most pulses are raised on poor soils under rainfed conditions. Here, drought at the end of the season, heat during flowering, salinity and diseases like Fusarium wilt, Ascochyta blight and yellow mosaic take a heavy toll (Jha et al., 2014; Nair et al., 2019). A narrow genetic base makes the problem worse. Chickpea, for instance, lost much of its diversity through a series of bottlenecks during domestication and its later spread (Abbo et al., 2003). Conventional breeding has given farmers many good varieties, but the process is slow. More than ten years may pass between the first cross and the release of a variety, and selecting for complex traits such as yield under drought takes a lot of time and money (Bohra et al., 2014; Samineni et al., 2020).

Varshney et al. (2005) used the phrase genomics-assisted breeding for the use of genome information and DNA markers in breeding programmes. Until the end of that decade, pulses had so few genomic resources that they were called orphan crops (Varshney et al., 2009). Things changed quickly once next-generation sequencing became affordable, and today most major pulses have a reference genome, resequencing data and single-nucleotide polymorphism (SNP) genotyping platforms (Pandey et al., 2016; Thudi et al., 2021). Other omics layers, namely transcriptomics, proteomics, metabolomics, ionomics and epigenomics, tell us what happens between the gene and the final trait (Langridge and Fleury, 2011; Yang et al., 2021). Phenomics, in which sensors and cameras are used to measure plant traits on a large scale, has grown alongside them (Furbank and Tester, 2011; Yang et al., 2020).

All these tools generate huge and complicated datasets. Machine learning (ML), a branch of artificial intelligence (AI), lets a computer pick up patterns from examples rather than from rules written by hand, and deep learning (DL) takes this further by using neural networks with many layers (LeCun et al., 2015; Libbrecht and Noble, 2015). Plant breeders now use ML and DL to predict breeding values, to score traits from images, to diagnose diseases and to bring omics and weather data together (Harfouche et al., 2019; Wang et al., 2020; Xu et al., 2022). In legumes, most of the early AI studies were done in soybean. Reports from chickpea, lentil, pea and common bean are, however, becoming more common (Singh et al., 2018; Marzougui et al., 2019; Keller et al., 2020).
In this review, we bring together what is known about omics resources and AI methods for pulse crops, and we discuss how they can be used together to breed better varieties in less time. We begin with genomic resources and trait mapping, move on to functional omics and phenomics, and then take up genomic selection and AI. The later sections cover speed breeding and genome editing, databases, the challenges that remain and the way forward. Figure 1 shows the overall framework that we have followed (Varshney et al., 2021b; Xu et al., 2022).

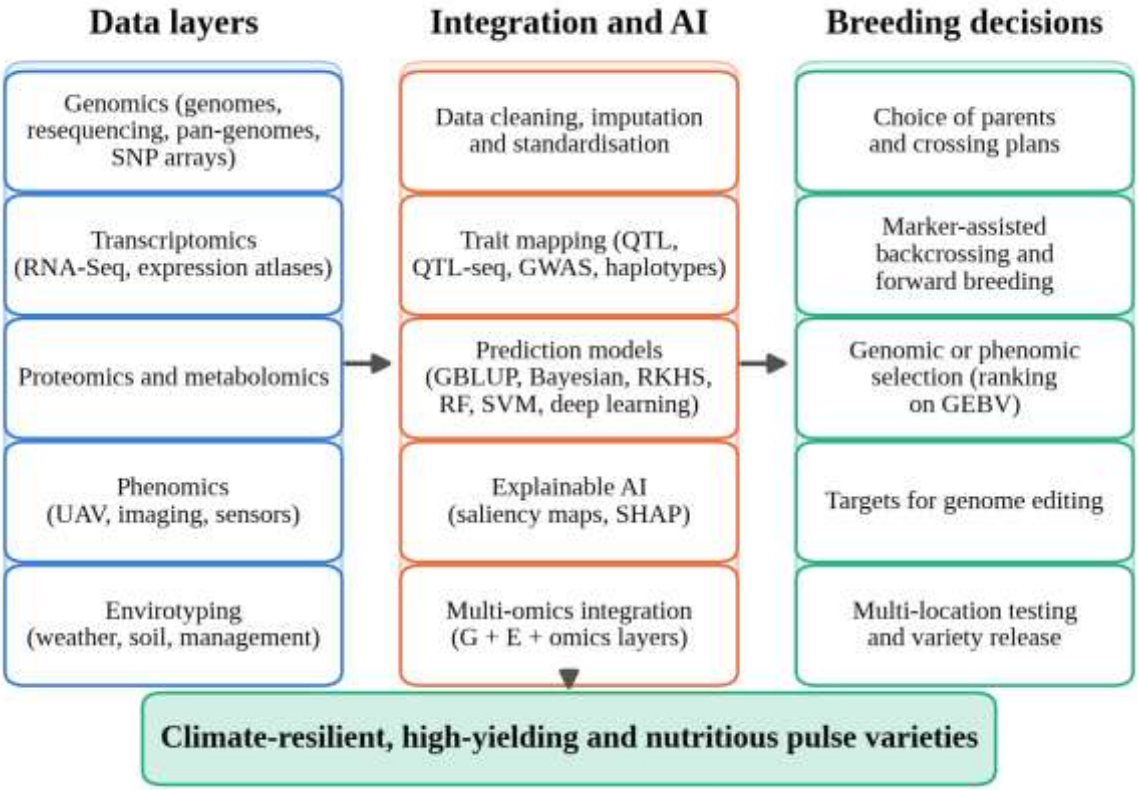


Figure 1. Overall framework for omics- and AI-assisted breeding in pulse crops. Data from several layers are cleaned and merged, analysed with statistical and machine learning models, and then used to make breeding decisions.

2. Pulse crops and their breeding constraints

The main pulses grown in India are chickpea (Cicer arietinum L.), pigeonpea (Cajanus cajan (L.) Millsp.), lentil (Lens culinaris Medik.), mungbean (Vigna radiata (L.) R. Wilczek), blackgram (Vigna mungo (L.) Hepper), field pea (Pisum sativum L.), cowpea (Vigna unguiculata (L.) Walp.) and common bean (Phaseolus vulgaris L.) (Bohra et al., 2014; Nair et al., 2019). Outside India, faba bean (Vicia faba L.) is a major protein crop in West Asia, North Africa, Europe and China (Jayakodi et al., 2023). Soybean (Glycine max (L.) Merr.) is normally counted among the oilseeds. Still, as a close legume relative with a large genomic toolkit, it is the crop in which many methods now used in pulses were first tried (Schmutz et al., 2010; Singh et al., 2018).
Abiotic stress is the first hurdle. Chickpea loses most of its yield to drought at the end of the season and to high temperature at the reproductive stage, while cold and salinity matter in certain areas (Jha et al., 2014). Mungbean is hit by heat, waterlogging and yellow mosaic disease in the field and by bruchids in storage (Nair et al., 2019). In pigeonpea, the two most serious diseases are Fusarium wilt and sterility mosaic (Singh et al., 2016b), and in several lentil and pea-growing areas, Aphanomyces root rot limits production (Marzougui et al., 2019). Common bean breeders have to deal with a long list of diseases along with drought and phosphorus-poor soils (Assefa et al., 2019; Keller et al., 2020). For pea, Smýkal et al. (2012) have reviewed the available genetic resources and genomic tools in detail.

The second hurdle is the narrow genetic base. As noted in the Introduction, chickpea lost much of its diversity through several bottlenecks (Abbo et al., 2003). Wild relatives do carry valuable genes for resistance and tolerance. However, crossing barriers and linkage drag make them hard to use, and so pre-breeding and genomic tools are needed (Sharma et al., 2013; Bohra et al., 2022). The third hurdle is time. Most pulses are self-pollinated, and a line has to be tested over several seasons and locations before it can be released (Hickey et al., 2019; Samineni et al., 2020). It is mainly for these reasons that breeders have started looking at omics and AI tools to make selection faster as well as more accurate (Varshney et al., 2018; Varshney et al., 2021a).

3. Genomic resources for pulse crops

3.1 Reference genome sequences

Soybean cultivar Williams 82 was the first legume crop to have its genome assembled up to the chromosome level. The assembly had 950 Mb of sequence and 46,430 predicted protein-coding genes (Schmutz et al., 2010). Pigeonpea was the first pulse to be sequenced, and the genome of ICPL 87119 (Asha) covered 605.78 Mb out of an estimated 833.07 Mb, with 48,680 predicted genes (Varshney et al., 2012). A year later, two chickpea genomes came out: one of the kabuli variety CDC Frontier with about 28,269 genes (Varshney et al., 2013b), and one of the desi line ICC 4958 with 27,571 genes (Jain et al., 2013).

The common bean genome of the Andean landrace G19833, with 473 Mb assembled out of 587 Mb, made it possible to compare the two independent domestications of this crop, one in Mesoamerica and the other in the Andes (Schmutz et al., 2014). In mungbean, the genome of VC1973A (431 Mb and 22,427 high-confidence genes) offered the first detailed look at how genomes have evolved within the genus Vigna (Kang et al., 2014). The cowpea genome of IT97K-499-35 followed (Lonardi et al., 2019). For blackgram, the Pant U-31 assembly spans about 475 Mb and contains 42,115 predicted protein-coding genes, and its authors paid particular attention to the resistance (R) genes (Jegadeesan et al., 2021).

Cool-season pulses have far bigger genomes, most of which are made up of repeated sequences. Assembling the roughly 4.45 Gb genome of pea cultivar Caméor needed Illumina and PacBio reads together with bacterial artificial chromosome (BAC) profiling, sorted chromosomes, optical maps and a dense genetic map (Kreplak et al., 2019). The faba bean genome is larger still, at about 13 Gb, and ranks among the largest diploid crop genomes; Jayakodi et al. (2023) reported an 11.9 Gb assembly of the line Hedin/2. Table 1 and Figure 2 compare the estimated genome size with the assembled length for the main pulse and related legume crops.

Table 1. Reference genome assemblies of major pulse and related grain legume crops.

Crop (species)	Genotype	Estimated genome size (Mb)	Assembled size (Mb)	Predicted genes	Reference
Soybean (Glycine max)	Williams 82	1,115	950	46,430	Schmutz et al. (2010)
Pigeonpea (Cajanus cajan)	ICPL 87119 (Asha)	833.07	605.78	48,680	Varshney et al. (2012)
Chickpea, kabuli (Cicer arietinum)	CDC Frontier	~738	544.73	~28,269	Varshney et al. (2013b)
Chickpea, desi (Cicer arietinum)	ICC 4958	~740	~520	27,571	Jain et al. (2013)
Common bean (Phaseolus vulgaris)	G19833	587	473	27,197	Schmutz et al. (2014)
Mungbean (Vigna radiata)	VC1973A	579	431	22,427	Kang et al. (2014)

Cowpea (Vigna unguiculata)	IT97K-499-35	640.6	519	–	Lonardi et al. (2019)
Pea (Pisum sativum)	Caméor	~4,450	3,920	44,756	Kreplak et al. (2019)
Blackgram (Vigna mungo)	Pant U-31	~574	~475	42,115	Jegadeesan et al. (2021)
Faba bean (Vicia faba)	Hedin/2	~13,000	11,900	34,221	Jayakodi et al. (2023)

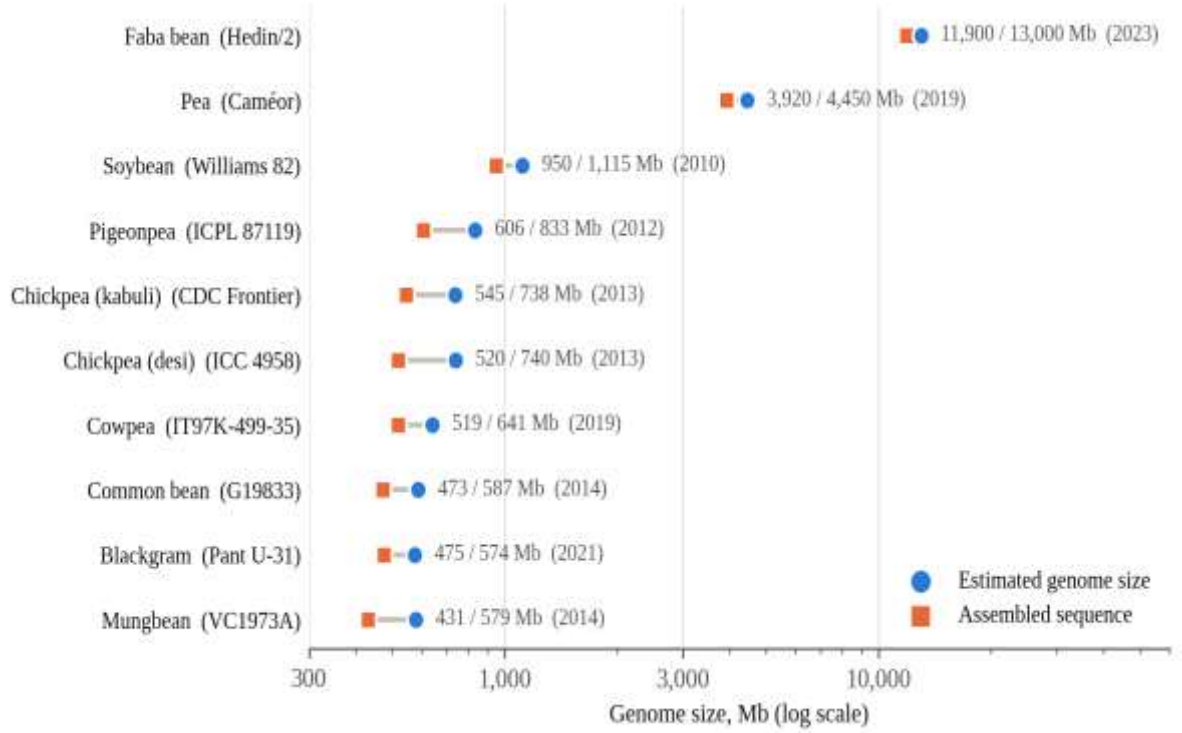


Figure 2. Estimated genome size and assembled length of the reference genomes of pulse and related legume crops, plotted on a log scale; the year of publication is shown in brackets.

3.2 Resequencing, pan-genomes and haplotypes

One reference genome cannot hold all the variation found in a species, so the next step is to resequence a wide range of germplasm (Varshney et al., 2021b). In pigeonpea, resequencing of 292 accessions helped to locate parts of the genome that changed during domestication, as well as regions linked with yield-related traits (Varshney et al., 2017). A similar study on 429 chickpea accessions collected from 45 countries shed light on the diversity of the crop and on how it was domesticated (Varshney et al., 2019). The effort was later scaled up to 3,366 chickpea genomes, cultivated as well as wild, which gave a detailed map of genetic variation and a large pool of haplotypes for breeders (Varshney et al., 2021c).

A pan-genome brings together the core genes shared by all members of a species and the variable genes that only some of them carry. In pigeonpea, many of the variable genes turned out to be associated with agronomically useful traits (Zhao et al., 2020). The soybean pan-genome, built from both wild and cultivated lines, uncovered large structural variants that would be missed if only one reference were used (Liu et al., 2020). For pea, an improved reference together with a pan-genome has thrown light on the structure and evolution of the genome (Yang et al., 2022). With such resources, breeders can move towards haplotype-based breeding, in which they choose the best combination of alleles within a genomic block instead of tracking one marker at a time (Bhat et al., 2021). The main milestones are summarised in Figure 3.

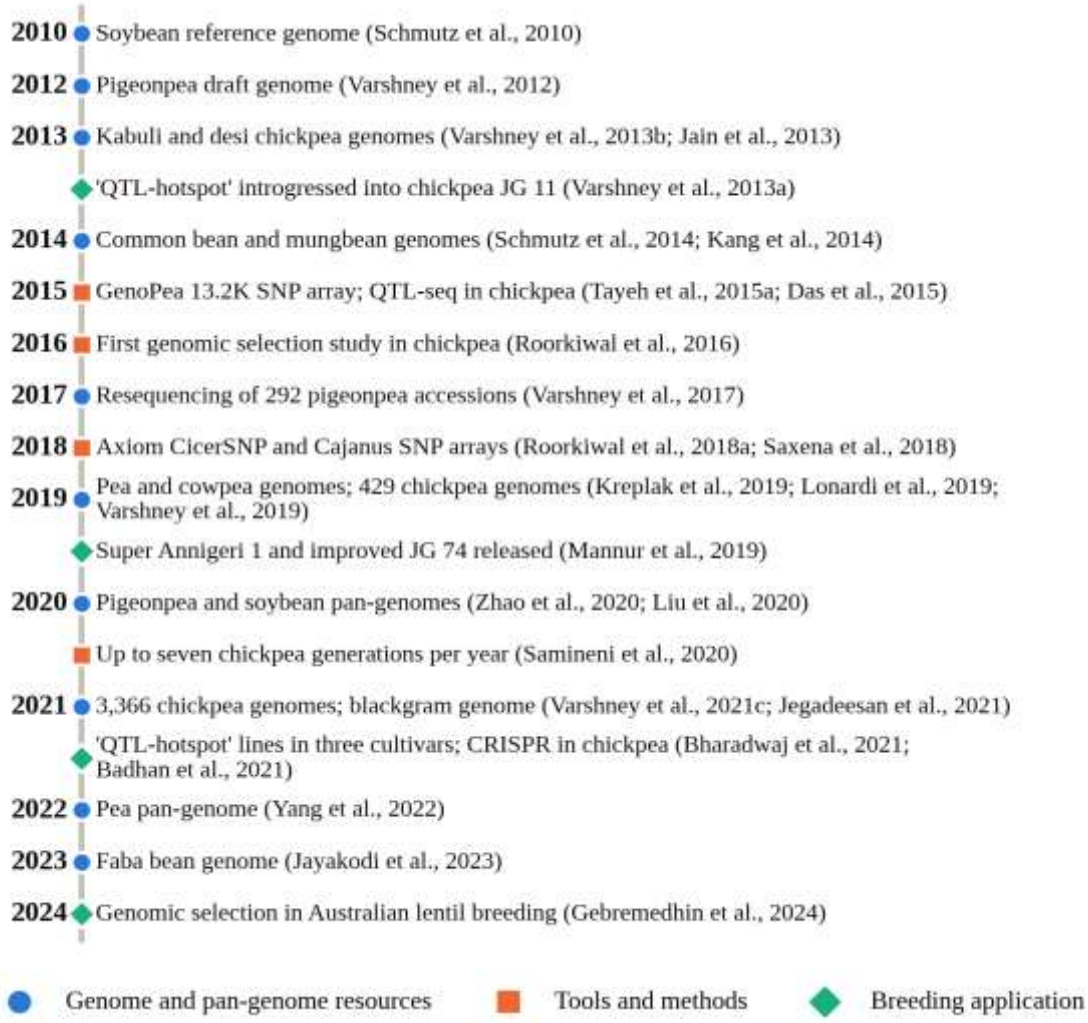


Figure 3. Important milestones in genomic resources and tools, and in their use for breeding pulse and related legume crops, from 2010 to 2024.

3.3 Molecular markers and genotyping platforms

The first markers used in pulses, such as restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) and simple sequence repeat (SSR) markers, were limited in number and covered the genome poorly, which reduced their value in breeding (Varshney et al., 2009; Pandey et al., 2016). SNPs, being the most frequent kind of DNA variation, are far better suited to automation. In genotyping-by-sequencing (GBS), restriction enzymes cut the genome into a smaller and repeatable fraction before sequencing, so that thousands of SNPs can be scored cheaply in large populations (Elshire et al., 2011).

Most pulses now have fixed SNP arrays. In chickpea, the Axiom CicerSNP array was designed both for fine genetic mapping and for routine breeding work (Roorkiwal et al., 2018a). A 56K Axiom array for pigeonpea was used to study the genetic make-up of released cultivars and of the founder lines behind them (Saxena et al., 2018). Pea workers have the GenoPea 13.2K array along with a dense consensus map (Tayeh et al., 2015a), and in common bean, SNP assays were developed for building linkage maps and for anchoring the genome sequence (Song et al., 2015). Because these platforms give the same result in any laboratory and in any year, their data can be pooled, and this matters a great deal for genomic selection (Thudi et al., 2021).

4. Trait mapping and marker-assisted breeding

4.1 QTL mapping and sequencing-based mapping

For many years, complex traits were mapped mainly through quantitative trait locus (QTL) analysis in populations derived from two parents. In chickpea, one such study found a region on linkage group CaLG04 that controlled several root and drought-related traits at once, and this region came to be known as the QTL-hotspot (Varshney et al., 2014). Using sequence and expression data, Kale et al. (2015) later narrowed down the most likely candidate genes within it.

Sequencing has made mapping much quicker. QTL-seq, first tested in rice, pairs bulked segregant analysis with the resequencing of DNA pools drawn from plants at the two extremes of a trait (Takagi et al., 2013). Das et al. (2015) applied this method in chickpea and were able to pin down a likely candidate gene for a major QTL within a short time. In pigeonpea, sequencing-based approaches pointed to candidate genes for resistance to Fusarium wilt and to sterility mosaic (Singh et al., 2016b).

4.2 Genome-wide association studies

Genome-wide association studies (GWAS) make use of the recombination that has built up over many generations in a diverse set of lines, and so they can map traits more finely than a two-parent cross (Yu et al., 2006). Mixed linear models, which account for population structure and for relatedness among the lines, keep false positives in check (Yu et al., 2006). More recent methods like FarmCPU and BLINK add statistical power and run fast even on very large datasets (Liu et al., 2016; Huang et al., 2019).

In chickpea, association mapping carried out across the whole genome and on selected candidate genes found markers linked with drought and heat tolerance traits (Thudi et al., 2014). Li et al. (2018) combined GWAS with resequencing data from 132 chickpea varieties and breeding lines, and found SNPs associated with yield-related traits under drought. In mungbean, a diversity panel was studied for the extent of linkage disequilibrium and for its population structure, basic information that is needed before any association study (Noble et al., 2018). The large resequencing panels of pigeonpea and chickpea have also been used for association analysis of agronomic traits (Varshney et al., 2017; Varshney et al., 2019).

4.3 Marker-assisted breeding

The true test of any breeding tool is whether it puts a better variety in the farmer's field. In chickpea, the QTL-hotspot was moved into the popular variety JG 11 through marker-assisted backcrossing (MABC). The resulting lines had better root traits, and some of them also yielded more under rainfed conditions (Varshney et al., 2013a). The same region was later transferred into three elite Indian cultivars, and under drought the new lines did better than their recurrent parents for both stress tolerance and grain yield (Bharadwaj et al., 2021). Against Fusarium wilt, MABC gave two resistant lines, Super Annigeri 1 and an improved version of JG 74, which were released for cultivation in India (Mannur et al., 2019). These cases show that MABC can deliver a better variety in a fairly short time, provided the QTL is well defined and good diagnostic markers are at hand (Varshney et al., 2021b).

Most traits of economic value, yield and adaptation in particular, are controlled by a large number of genes with small effects, and MABC on its own cannot handle them (Heffner et al., 2009). This is the reason genomic selection has been taken up, as discussed in Section 7 (Crossa et al., 2017).

5. Functional omics in pulse crops

Knowing the DNA sequence tells us which genes a plant has, but not when or where those genes are switched on. Functional omics fill this gap by measuring transcripts, proteins, metabolites and epigenetic marks (Langridge and Fleury, 2011; Scossa et al., 2021).

5.1 Transcriptomics

RNA sequencing (RNA-Seq) profiles the expression of every gene in a single experiment and has largely replaced microarrays (Wang et al., 2009). For chickpea, an expression atlas built with RNA-Seq across many tissues and growth stages showed how gene activity shifts in space and time as the plant develops (Kudapa et al., 2018). A comparable atlas in pigeonpea helped identify genes that influence pollen fertility, which is important for hybrid breeding (Pazhamala et al., 2017). When chickpea lines differing in tolerance were compared, their transcriptomes responded to drought and salt stress in ways that depended on both the genotype and the growth stage (Garg et al., 2016). Data of this kind help to shortlist candidate genes within QTL regions, as was done for the chickpea QTL-hotspot (Kale et al., 2015). Much of the chickpea expression data can be freely accessed through the Chickpea Transcriptome Database, CTDB (Verma et al., 2015).

5.2 Proteomics

Proteins perform most of the work in a cell, and the abundance of a protein does not always match that of its transcript (Ramalingam et al., 2015). In legumes, mass spectrometry-based proteomics has been used to study stress responses, seed development and nodule formation (Ramalingam et al., 2015). In chickpea seedlings, leaf proteins were monitored as dehydration intensified, and the levels of several proteins rose or fell in parallel with the stress (Vessal et al., 2020). Proteins that differ in tolerant lines can serve as markers or be pursued for further functional studies (Ramalingam et al., 2015).

5.3 Metabolomics and ionomics

Metabolites are the end products of gene activity and therefore lie close to the phenotype (Fernie and Schauer, 2009). Metabolomics can support breeding in two ways, either by identifying metabolic markers for a trait or by mapping QTL that control metabolite levels (Fernie and Schauer, 2009; Kumar et al., 2017). In legumes, metabolite profiling has helped explain stress tolerance and seed quality traits, such as protein content and anti-nutritional factors (Ramalingam et al., 2015). A study in hybrid maize showed that the metabolite profiles of parental lines could predict complex heterotic traits in their hybrids, highlighting the promise of metabolite-based prediction (Riedelsheimer et al., 2012). Ionomics concerns the full set of mineral nutrients and trace elements in a plant, and it is directly useful when pulses are to be biofortified with iron and zinc (Salt et al., 2008).

5.4 Epigenomics

DNA methylation, histone modification and other epigenetic marks can alter gene expression without changing the DNA sequence, and some of these changes are inherited by the next generation (Gallusci et al., 2017). These mechanisms contribute to stress memory and to the ability of plants to adapt to their surroundings, both of which are relevant for crops grown under variable conditions (Gallusci et al., 2017). Very few epigenomic studies have been conducted on pulses so far, and this area clearly needs more attention (Yang et al., 2021).

5.5 Putting the omics layers together

No single omics layer provides the full picture, and combining layers explains complex traits better (Yang et al., 2021; Scossa et al., 2021). In maize, for example, transcript levels were used to predict complex traits and complemented genetic markers (Azodi et al., 2020). Deep learning tools have also been built specifically to use several omics layers for genomic prediction in plants (Wang et al., 2023), as discussed in Section 8.2. Table 2 lists some of the important omics resources and studies in pulses.

Table 2. Selected omics resources and studies in pulse crops.

Omics layer	Crop	Resource or approach	Main outcome	Reference
Genomics (resequencing)	Pigeonpea	292 accessions resequenced	Genomic regions tied to domestication and yield-related traits	Varshney et al. (2017)
Genomics (resequencing)	Chickpea	429 accessions from 45 countries	Diversity, domestication history and trait associations	Varshney et al. (2019)
Genomics (variation map)	Chickpea	3,366 genomes sequenced	Genome-wide variation map and haplotypes	Varshney et al. (2021c)
Pan-genome	Pigeonpea	Pan-genome assembly	Variable genes linked with useful traits	Zhao et al. (2020)
Pan-genome	Pea	Improved reference and pan-genome	Structure and evolution of the pea genome	Yang et al. (2022)
Genotyping array	Chickpea	Axiom CicerSNP array	Fine mapping and routine breeding use	Roorkiwal et al. (2018a)
Genotyping array	Pigeonpea	56K Axiom array	Genetic make-up of cultivars and founder lines	Saxena et al. (2018)
Genotyping array	Pea	GenoPea 13.2K array	Dense consensus genetic map	Tayeh et al. (2015a)
Trait mapping	Chickpea	QTL-seq	Likely candidate gene for a major QTL	Das et al. (2015)
Transcriptomics	Chickpea	RNA-Seq expression atlas	Tissue- and stage-specific expression	Kudapa et al. (2018)
Transcriptomics	Pigeonpea	Expression atlas	Genes influencing pollen fertility	Pazhamala et al. (2017)
Transcriptomics	Chickpea	RNA-Seq under stress	Genotype- and stage-dependent responses to drought and salt	Garg et al. (2016)
Proteomics	Chickpea	Leaf proteome under dehydration	Proteins changing with stress severity	Vessal et al. (2020)
Proteomics and metabolomics	Legumes	Review of platforms	Uses in legume improvement	Ramalingam et al. (2015)

6. Phenomics and high-throughput phenotyping

Genotyping is now cheap, but measuring traits accurately on thousands of plots is still slow and costly. Furbank and Tester (2011) called this the phenotyping bottleneck. High-throughput phenotyping tries to remove it with

sensors, cameras and automated platforms that measure plants without harming them and can repeat the measurement many times in a season (Araus and Cairns, 2014; Tardieu et al., 2017). Each sensor captures something different. Red–green–blue (RGB) cameras record plant size, colour and the number of organs; multispectral and hyperspectral sensors pick up reflectance linked to pigments, water and nitrogen; thermal cameras give canopy temperature; and light detection and ranging (LiDAR) and 3D imaging capture plant architecture (Zhao et al., 2019; Yang et al., 2020).

At the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), the LeasyScan platform pairs 3D laser scanning of leaf area with lysimeters, so that the traits governing how much water a plant uses can be studied (Vadez et al., 2015). Unmanned aerial vehicles (UAVs) carrying high-resolution cameras can fly low over large field trials and image them quickly and repeatedly (Sankaran et al., 2015). In pea, remote sensing data were used to estimate biomass and yield-related traits (Quirós Vargas et al., 2019), and image-based methods were worked out to measure how heavily pea and chickpea plants flower (Zhang et al., 2020). In lentil, resistance to Aphanomyces root rot is usually scored by eye, but RGB and hyperspectral imaging made it possible to score it in numbers (Marzougui et al., 2019).

The volume of phenomic data is so large that it has to be analysed automatically, and this is exactly where machine learning comes in (Singh et al., 2016a; Tardieu et al., 2017). Phenomic data can also be used for prediction on their own. In phenomic selection, spectra such as near-infrared spectra take the place of markers as predictors, an idea that worked well in wheat and poplar (Rincent et al., 2018). In soybean, canopy traits recorded by sensors served as phenomic predictors of yield and helped in making selection decisions (Parmley et al., 2019).

7. Genomic selection in pulse crops

7.1 Principle and statistical models

Genomic selection (GS) was put forward by Meuwissen et al. (2001). The idea is to use all the markers spread across the genome, at the same time, to estimate the breeding value of an individual. A training population that has been both genotyped and phenotyped is used to estimate the marker effects, and the model then predicts the genomic estimated breeding values (GEBVs) of selection candidates, which need only be genotyped (Heffner et al., 2009; Crossa et al., 2017). As lines can be selected before they are tested in the field, GS shortens the breeding cycle and raises the genetic gain achieved per year (Crossa et al., 2017).

The most widely used GS models are ridge regression best linear unbiased prediction (rrBLUP) and genomic best linear unbiased prediction (GBLUP), both of which assume that every marker has a small effect of about the same size (VanRaden, 2008; Endelman, 2011). Bayesian models such as BayesA, BayesB, BayesC and Bayesian LASSO let the marker effects vary in size (Habier et al., 2011), and reproducing kernel Hilbert space (RKHS) regression can pick up non-additive effects (Gianola and van Kaam, 2008). Freely available packages like BGLR have made these models easy to run (Pérez and de los Campos, 2014). When lines are tested across many environments, models that include genotype-by-environment (G×E) interaction and environmental covariates, such as the reaction norm model, are especially useful (Jarquín et al., 2014).

7.2 Genomic selection studies in pulses

In the first GS study on chickpea, 320 elite breeding lines were genotyped with about 3,000 Diversity Arrays Technology sequencing (DArTseq) markers and six models were compared. Accuracy was 0.138 to 0.606 for seed yield and 0.797 to 0.912 for 100-seed weight, and the models gave similar results for any given trait (Roorkiwal et al., 2016). Working with multi-environment trials, the same team found that adding G×E interaction to the model changed the accuracy, which varied from 0.087 to 0.670 depending on the trait and the validation scheme (Roorkiwal et al., 2018b). Li et al. (2018), using resequencing data on 132 chickpea lines, got accuracies of just 0.148 to 0.186 for yield under rainfed conditions when all SNPs were used. When they used only the SNPs picked up by GWAS, the accuracy for three yield-related traits went up to 0.56 to 0.61.

Work on pea gave accuracies of 0.83 to 0.86 for thousand seed weight and 0.65 to 0.79 for beginning of flowering with the GenoPea array, and showed that the number of markers and the size and make-up of the training set influence accuracy (Tayeh et al., 2015b). Annicchiarico et al. (2019) obtained accuracies of 0.30 for grain yield and 0.65 for flowering onset within pea populations, but lower values when one population was used to predict another. In common bean, prediction ability for seed yield across 481 Andean breeding lines was 0.25 to 0.60, and modelling G×E improved the predictions under drought and under low phosphorus (Keller et al., 2020). For lentil, a breeding programme in Australia reported accuracies of 0.25 to 0.63 for grain yield and 0.45 to 0.64 for ascochyta blight resistance, and GS promised larger genetic gains than selection based on phenotype alone (Gebremedhin et al., 2024). Table 3 and Figure 4 bring these results together.

A few broad lessons come out of these studies. Traits of high heritability, such as seed weight and flowering time, are predicted well, whereas grain yield is predicted less accurately (Roorkiwal et al., 2016; Tayeh et al., 2015b). Accuracy also drops when the training and test populations are not closely related, so the training set must represent the material that is actually being selected (Tayeh et al., 2015b; Annicchiarico et al., 2019). Lastly, accounting for G×E and adding prior information from GWAS can make the predictions better (Li et al., 2018; Keller et al., 2020).

Table 3. Genomic prediction accuracies reported in pulse crops.

Crop	Population	Markers	Models	Trait: accuracy range	Reference
Chickpea	320 elite breeding lines	~3,000 DArTseq markers	rrBLUP, kinship GAUSS, BayesCπ, BayesB, Bayesian LASSO, random forest	Seed yield 0.138–0.606; 100-seed weight 0.797–0.912; days to flowering 0.556–0.851	Roorkiwal et al. (2016)
Chickpea	320 breeding lines, multi-environment trials	1,568 DArTseq and 88,845 GBS SNPs	13 models with and without G×E	Eight traits 0.087–0.670	Roorkiwal et al. (2018b)
Chickpea	132 varieties and breeding lines	144,777 SNPs from resequencing	rrBLUP, Bayesian LASSO, Bayesian ridge regression	Rainfed yield (all SNPs) 0.148–0.186; yield traits (GWAS SNPs) 0.56–0.61	Li et al. (2018)
Pea	339 accessions (training) and 746 recombinant inbred lines (RILs)	9,824 SNPs (GenoPea array)	kernel partial least squares regression, LASSO, GBLUP, BayesA, BayesB	Thousand seed weight 0.83–0.86; seeds per plant 0.68–0.73; flowering 0.65–0.79	Tayeh et al. (2015b)
Pea	306 RILs from three crosses	6,058 GBS SNPs	Bayesian LASSO, rrBLUP, GBLUP	Grain yield 0.19–0.30; flowering onset 0.40–0.65; seed weight 0.28–0.64	Annicchiarico et al. (2019)
Common bean	481 Andean elite lines	4,962 SNPs	Genomic prediction with and without G×E	Seed yield 0.25–0.60; days to flowering and maturity up to about 0.8	Keller et al. (2020)
Lentil	Australian breeding programme	Genome-wide SNPs	Bayesian models	Grain yield 0.25–0.63; boron tolerance 0.47–0.72; ascochyta blight 0.45–0.64	Gebremedhin et al. (2024)

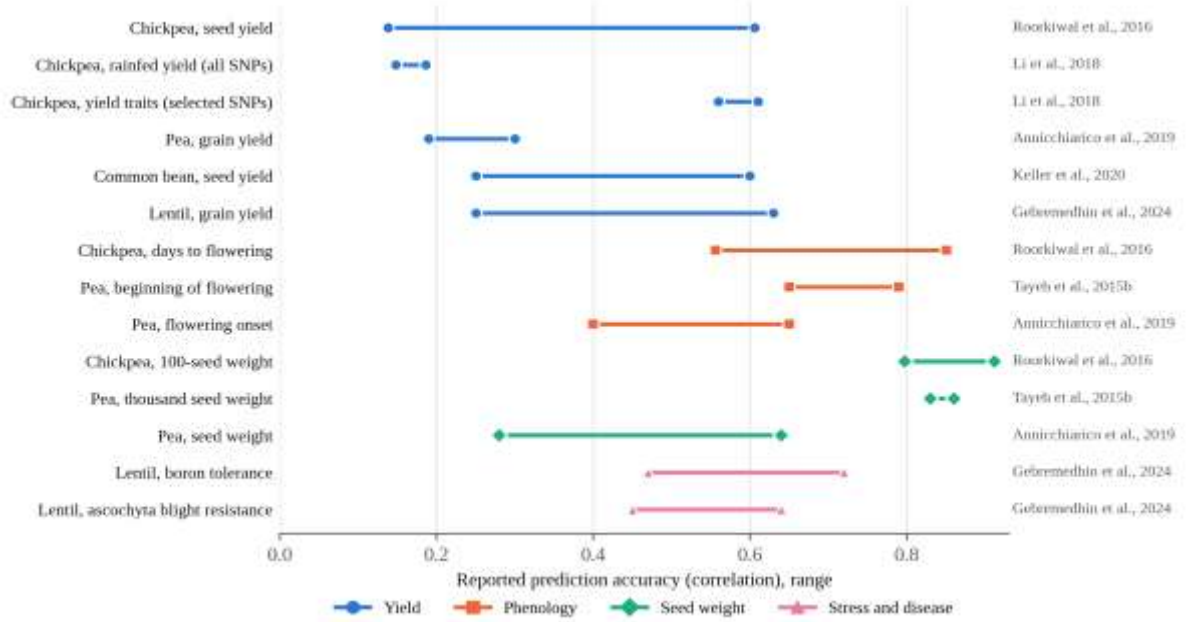


Figure 4. Reported range of genomic prediction accuracy for different traits in chickpea, pea, common bean and lentil. Each line runs from the lowest to the highest value given in the cited study across models, environments or validation schemes.

8. Artificial intelligence and machine learning in pulse breeding

8.1 Basic concepts

In machine learning, an algorithm is shown a set of examples in which both the inputs (say, SNP genotypes or image pixels) and the output (say, yield) are known. It learns the relationship between them and then applies it to new cases (Libbrecht and Noble, 2015). Among the supervised methods popular with breeders are random forests, which take the average of many decision trees (Breiman, 2001), support vector machines, and gradient boosting methods like XGBoost (Chen and Guestrin, 2016). Deep learning relies on neural networks with many layers. Convolutional neural networks (CNNs) work particularly well on images, since they learn edges, colours and shapes straight from the raw pixels (LeCun et al., 2015). In genomics, researchers are using deep learning to predict gene expression, regulatory regions, and phenotypes from DNA sequence (Zou et al., 2019; Wang et al., 2020).

8.2 Machine learning for genomic prediction

Several studies have put ML methods side by side with the standard GS models. Ogutu et al. (2011) tested random forest, boosting and support vector machines on a simulated dataset and found that they could hold their own against the usual approaches. When Azodi et al. (2019) benchmarked many algorithms in six plant species, no single method came out on top for every trait, and plain linear models often did as well as the more elaborate ML models. A review on deep learning in GS arrived at much the same view. Deep networks can capture non-additive effects and accept many kinds of input, but they do not always beat conventional models, and their performance depends on how much data is available and how well the model is tuned (Montesinos-López et al., 2021).

In chickpea, random forest was among the six models tried by Roorkiwal et al. (2016), and it performed about as well as the parametric models. A deep CNN developed for soybean predicted phenotypes from genotypes and, at the same time, highlighted the genomic regions tied to each trait, so that prediction and association were handled in one framework (Liu et al., 2019). The DNNGP method, which is based on a deep neural network, can take genomic, transcriptomic and other omics data together as input for prediction in plants (Wang et al., 2023). Since training populations in pulses usually contain only a few hundred lines, a well-tuned linear model remains a strong starting point, and ML methods are likely to help most where non-linear effects or several types of data are involved (Azodi et al., 2019; Montesinos-López et al., 2021).

8.3 Image-based phenotyping and disease diagnosis

Image-based phenotyping is where deep learning has made its largest impact so far (Singh et al., 2018). Mohanty et al. (2016) trained CNNs on 54,306 photographs of healthy and diseased leaves from the open PlantVillage collection (Hughes and Salathé, 2015), covering 14 crop species and 26 diseases, and the models classified 99.35% of the held-out images correctly (Figure 5). The authors also pointed out that accuracy fell sharply on pictures taken from other sources under different conditions, a reminder that training data must be varied (Mohanty et al., 2016). Pound et al. (2017) showed that deep learning did better than earlier approaches in tasks such as locating root tips and shoot features in images, and platforms like Deep Plant Phenomics have made these methods easier for plant scientists to adopt (Ubbens and Stavness, 2017).



Figure 5. Photographs of healthy and diseased leaves covering the 38 crop and disease classes (14 crop species, soybean among them) that were used to train convolutional neural networks to identify diseases automatically.

Among grain legumes, soybean has served as the test crop. A CNN trained on soybean leaf images could identify and classify foliar stresses, and the explainable part of the framework showed which areas of each image drove the decision (Ghosal et al., 2018). A three-dimensional CNN run on hyperspectral images picked out charcoal rot in soybean stems and indicated which wavelengths carried the most information (Nagasubramanian et al., 2019). In breeding trials, pictures of soybean plots taken from several angles were merged by deep learning to estimate yield (Riera et al., 2021), and ML models predicted soybean yield from hyperspectral reflectance, with a stacked ensemble doing better than any single model (Yoosefzadeh-Najafabadi et al., 2021).
In pulses, the image-based studies described in Section 6 follow the same approach (Marzougui et al., 2019; Quirós Vargas et al., 2019; Zhang et al., 2020). Automated scoring of this kind cuts down human error and lets breeders screen many more lines in the same time (Singh et al., 2016a; Harfouche et al., 2023). Figure 6 outlines the usual steps in an AI-based phenotyping pipeline, and Table 4 lists examples of ML and DL applications.

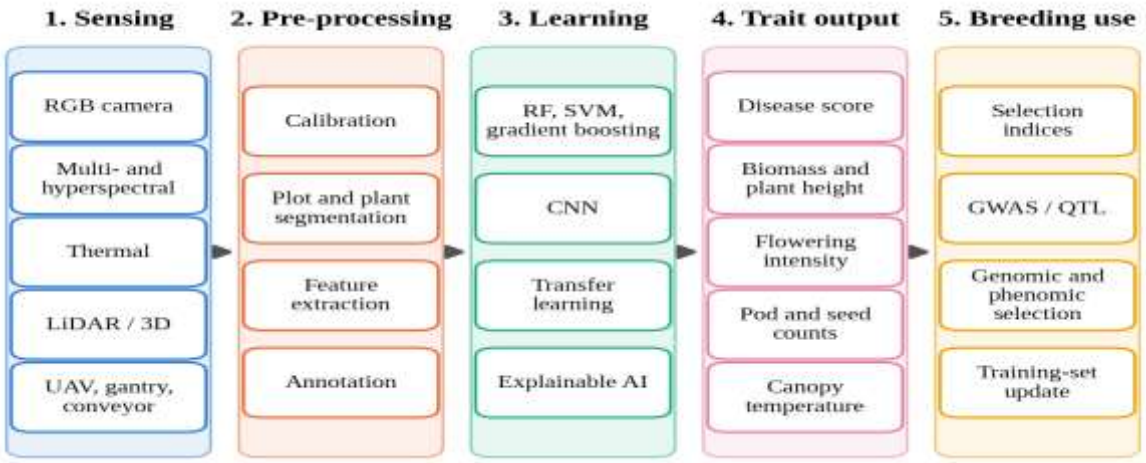


Figure 6. Usual steps in an AI-assisted high-throughput phenotyping pipeline, from capturing the images to using the trait data in breeding.

Table 4. Examples of machine learning and deep learning applications in pulse and related legume crops.

Crop	Task	Data type	Method	Main outcome	Reference
Chickpea	Genomic prediction of yield traits	DArTseq markers	rrBLUP, Bayesian models, random forest	Random forest about as good as parametric models	Roorkiwal et al. (2016)
Soybean	Genomic prediction with association	SNP genotypes	Deep CNN	Predictions plus trait-linked genomic regions	Liu et al. (2019)
Soybean	Leaf stress identification	RGB leaf images	Explainable deep CNN	Stress classes with visual explanation	Ghosal et al. (2018)
Soybean	Charcoal rot detection	Hyperspectral stem images	3D CNN with saliency maps	Disease detected; key wavelengths found	Nagasubramanian et al. (2019)
Soybean	Plot-level yield estimates	Plot images from several angles	Deep learning image fusion	Yield estimates used to rank genotypes	Riera et al. (2021)
Soybean	Yield prediction	Hyperspectral canopy reflectance	Multilayer perceptron, support vector machine, random forest, stacked ensemble	Ensemble better than single models	Yoosefzadeh-Najafabadi et al. (2021)
Soybean	Phenomic-assisted selection	Canopy traits from sensors	Machine learning	Optimised phenomic predictors of yield	Parmley et al. (2019)
Lentil	Root rot scoring	RGB and hyperspectral images	Image analysis and data-driven models	Disease scored in numbers instead of by eye	Marzougui et al. (2019)
Pea	Biomass and yield traits	UAV remote sensing data	Remote sensing models	Biomass and yield traits estimated	Quirós Vargas et al. (2019)
Pea and chickpea	Flowering intensity	Aerial images	Image processing	Flowering measured from the air	Zhang et al. (2020)
14 crops including soybean	Leaf disease diagnosis	54,306 leaf images	CNN	99.35% accuracy on held-out images	Mohanty et al. (2016)

8.4 Explainable AI

ML models, and deep networks in particular, are often criticised for working like a black box (Harfouche et al., 2023). A breeder would like to know why a line is predicted to do well, and a biologist would like to connect the prediction with genes and pathways. Explainable AI methods are meant to answer such questions. SHapley Additive exPlanations (SHAP) values break an individual prediction down into the share contributed by each input feature (Lundberg and Lee, 2017), while saliency and activation maps show which parts of an image, or which wavelengths, a CNN relied upon (Ghosal et al., 2018; Nagasubramanian et al., 2019). In genomic prediction, feature importance scores from random forest, or saliency from CNNs, can point to genomic regions worth a closer look (Breiman, 2001; Liu et al., 2019).

9. Integration of multi-omics, environment and AI

How a line performs depends not only on its genotype but also on the environment and on the interaction between the two. Envirotyping refers to recording detailed information on weather, soil and crop management, so that the response of genotypes can be understood and predicted (Xu, 2016). Such enviromic information can be fed into prediction models along with genomic data, as was done in the reaction norm model (Jarquín et al., 2014) and in common bean (Keller et al., 2020).

Xu et al. (2022) used the term smart breeding for an approach in which large genomic, phenomic and enviromic datasets are combined with AI to make joint genomic-enviromic predictions. Harfouche et al. (2019) put forward a similar idea when they called for next-generation AI to breed climate-resilient crops, and Varshney et al. (2021a) described it as fast-forward breeding. For pulses, most of the pieces are already in place: reference genomes and pan-genomes (Varshney et al., 2021c; Yang et al., 2022), gene expression atlases (Pazhamala et al., 2017; Kudapa et al., 2018), phenotyping platforms (Vadez et al., 2015) and a growing body of GS work (Gebremedhin et al.,

2024). What is needed now is a single data system that links all of them and that breeders can use in their routine work (Varshney et al., 2021b).

There are a few ways of combining data from different omics layers. All the layers may be joined into one large input, separate models may be built for each layer, and their predictions merged, or knowledge of pathways and networks may be used to connect the layers (Yang et al., 2021; Scossa et al., 2021). Deep networks can accept several layers as input (Wang et al., 2023), and the use of transcript data in maize is a good example of adding an intermediate layer to genomic data (Azodi et al., 2020). Where money for genotyping is short, phenomic selection based on spectra offers a cheaper route (Rincent et al., 2018; Parmley et al., 2019).

10. Shortening the breeding cycle: speed breeding and genome editing

GS and AI give their full benefit only when the breeding cycle is short (Hickey et al., 2019). Speed breeding relies on long photoperiods and controlled temperatures to fit several generations into one year, and it has been shown to work in a number of crops, chickpea and pea included (Watson et al., 2018). A rapid generation advance protocol for chickpea managed as many as seven generations in a year, which can considerably cut down the time needed to develop a variety (Samineni et al., 2020). If speed breeding is combined with low-cost genotyping and selection on GEBVs, parents for the next cycle can be picked without waiting for field trials, as illustrated in Figure 7 (Varshney et al., 2021a).

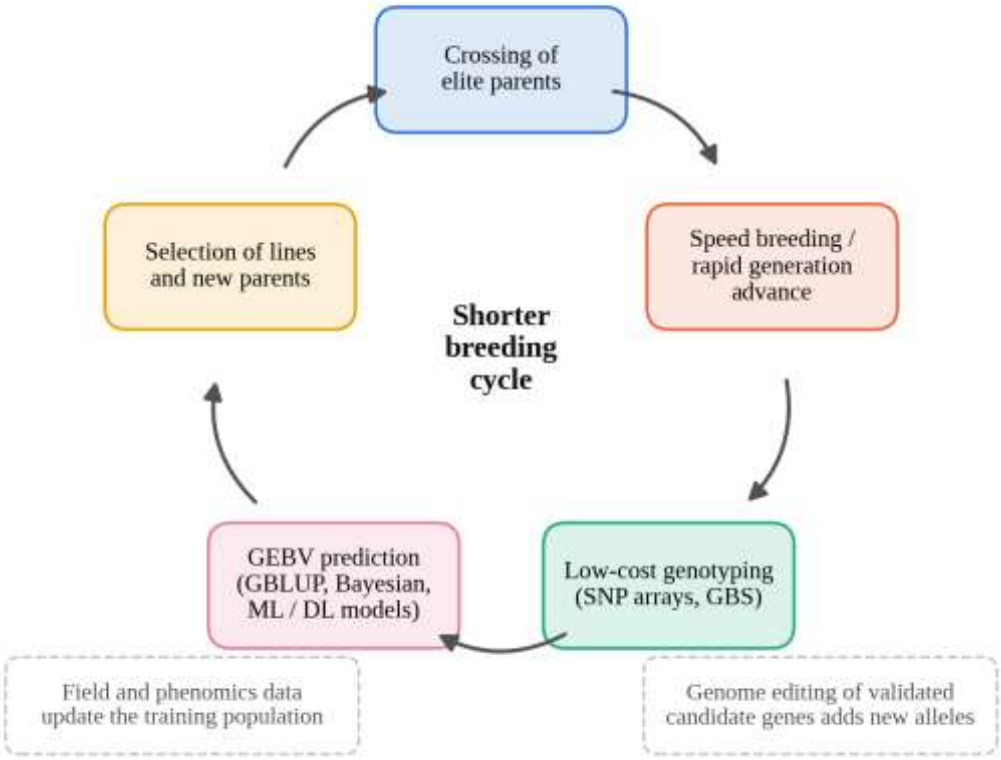


Figure 7. Shortening the breeding cycle in pulses by combining speed breeding, low-cost genotyping and genomic prediction.

Genome editing offers a new way of creating variation in genes whose function is already known. Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) systems can make precise changes at chosen sites and are in use in many crops (Chen et al., 2019; Gao, 2021). In chickpea, the 4CL and RVE7 genes were edited with CRISPR/Cas9 in protoplasts without introducing foreign DNA, which was the first report of its kind for this crop (Badhan et al., 2021). Candidate genes found through GWAS, QTL-seq and expression studies become obvious targets for editing once their role is confirmed (Kale et al., 2015; Varshney et al., 2021c). Even so, efficient transformation and regeneration remain a bottleneck in many crops, most pulses among them (Chen et al., 2019; Gao, 2021).

11. Databases and computational resources

Public databases are what make omics data usable by breeders. The Legume Information System (LegumeInfo.org) is one of a group of linked resources that together cover a large number of legume species (Dash et al., 2016). SoyBase caters to soybean researchers (Grant et al., 2010), while Phytozome offers comparative genomics across green plants, several legumes included (Goodstein et al., 2012). Chickpea workers can turn to CicArVarDB for SNP and insertion–deletion (InDel) variants (Doddamani et al., 2015) and to CTDB

for transcriptome data (Verma et al., 2015). Software for GWAS and GS, for example FarmCPU, BLINK, rrBLUP and BGLR, is free to use (Endelman, 2011; Pérez and de los Campos, 2014; Liu et al., 2016; Huang et al., 2019), and deep learning tools such as Deep Plant Phenomics and DNNGP have been released for plant data (Ubbens and Stavness, 2017; Wang et al., 2023). Table 5 lists these resources.

Table 5. Databases and software commonly used in omics- and AI-assisted breeding of pulse crops.

Name	Type	Main use	Reference
LegumeInfo.org (Legume Information System)	Database	Genomes, genetic maps and markers for many legumes	Dash et al. (2016)
SoyBase	Database	Soybean genetics and genomics	Grant et al. (2010)
Phytozome	Database	Comparative genomics of green plants	Goodstein et al. (2012)
CicArVarDB	Database	Chickpea SNP and InDel variants	Doddamani et al. (2015)
CTDB	Database	Chickpea transcriptome data	Verma et al. (2015)
FarmCPU	Software	Multi-locus GWAS	Liu et al. (2016)
BLINK	Software	Fast GWAS for very large datasets	Huang et al. (2019)
rrBLUP	R package	Ridge regression and kernel-based genomic selection	Endelman (2011)
BGLR	R package	Bayesian genomic regression and prediction	Pérez and de los Campos (2014)
Deep Plant Phenomics	Software	Deep learning for plant phenotyping tasks	Ubbens and Stavness (2017)
DNNGP	Software	Deep learning genomic prediction using several omics layers	Wang et al. (2023)

12. Challenges

Small training sets. Most GS studies in pulses have worked with a few hundred lines at most (Roorkiwal et al., 2016; Annicchiarico et al., 2019), which is too few for deep learning models that need a lot of data (Montesinos-López et al., 2021). Image collections for pulse diseases are also small, and a model trained on pictures taken under controlled conditions may fail in a farmer's field (Mohanty et al., 2016; Singh et al., 2018).

Phenotyping quality and standards. Poor or uneven phenotyping lowers prediction accuracy. Data from different programmes often cannot be merged either, because each programme follows its own protocols and defines traits in its own way (Tardieu et al., 2017; Yang et al., 2020).

Transfer of models. A model that works well in one population may predict poorly in an unrelated population or in a new environment (Tayeh et al., 2015b; Annicchiarico et al., 2019). The training set therefore needs regular updating with new lines and fresh field data (Crossa et al., 2017).

Infrastructure and skills. Public breeding programmes in developing countries face real problems here: the cost of genotyping, the need for computing facilities and a shortage of people trained in both breeding and data science (Varshney et al., 2018; Thudi et al., 2021).

Interpretation and validation. Complex models are hard to interpret, and breeders will trust their predictions only after these have been confirmed in field trials (Lundberg and Lee, 2017; Harfouche et al., 2023). Candidate genes coming out of omics studies also have to be validated for their function, and whether genome editing can be used depends on the rules of each country (Chen et al., 2019; Gao, 2021).

13. Future prospects

Several developments can be expected in the next few years. More chromosome-level assemblies and pan-genomes, including those of wild relatives, will help breeders to find and use new alleles for stress tolerance and quality (Bohra et al., 2022; Yang et al., 2022). The chickpea variation map, and similar resources for other pulses, will support haplotype-based breeding, in which superior haplotypes are identified and brought together (Bhat et al., 2021; Varshney et al., 2021c). Deep learning models that take in several omics layers at once may improve the prediction of complex traits once enough data have been gathered (Montesinos-López et al., 2021; Wang et al., 2023).

When genomic prediction is linked with envirotyping, breeders will be able to predict how a line will perform in a particular place and season, which is valuable for a country like India with its many agro-climatic zones (Xu, 2016; Xu et al., 2022). Cheap sensors, disease diagnosis on smartphones and phenomic selection could bring these tools within the reach of national programmes (Mohanty et al., 2016; Rincent et al., 2018; Parmley et al., 2019). Bringing speed breeding, GS and genome editing into a single pipeline, as proposed under fast-forward breeding, could shorten the time from the first cross to the release of a variety (Watson et al., 2018; Varshney et al., 2021a). Last but not least, open databases and shared data standards will be needed, so that data from many programmes can be pooled to train better models (Dash et al., 2016; Tardieu et al., 2017).

14. Conclusion

In a little over a decade, pulses have moved from being genomic orphans to crops with chromosome-level reference genomes, pan-genomes, resequencing data on thousands of accessions and high-density SNP arrays (Varshney et al., 2009; Roorkiwal et al., 2018a; Varshney et al., 2021c). These resources are already reaching farmers. Marker-assisted backcrossing has produced Fusarium wilt-resistant chickpea varieties released in India (Mannur et al., 2019), and introgression of the "QTL-hotspot" region has improved drought tolerance and grain yield in elite chickpea cultivars (Bharadwaj et al., 2021). Genomic selection studies in chickpea, pea, common bean and lentil report useful prediction accuracies, particularly for traits of moderate to high heritability (Tayeh et al., 2015b; Roorkiwal et al., 2016; Keller et al., 2020; Gebremedhin et al., 2024).

The next gains are likely to come less from any single technology than from linking them together. Functional omics shows how genes act, high-throughput phenotyping measures how plants perform, and machine learning offers ways to combine these layers with genomic and environmental data (Yang et al., 2021; Xu et al., 2022; Harfouche et al., 2023). For now, however, deep learning does not consistently outperform well-tuned linear models for genomic prediction, largely because training populations in pulses remain small (Azodi et al., 2019; Montesinos-López et al., 2021). Combined with speed breeding and, where regulations allow, genome editing, these tools could markedly shorten the breeding cycle (Watson et al., 2018; Varshney et al., 2021a). Realising this will require larger and regularly updated training populations, standardised phenotyping, open and connected databases, and breeders trained in data science, especially in public programmes of developing countries (Tardieu et al., 2017; Varshney et al., 2018). If these needs are met, omics- and AI-assisted breeding can help deliver climate-resilient and nutritious pulse varieties to farmers sooner than conventional breeding alone.

Acknowledgments Nil.

Author Contributions Conceptualisation, S.S. and R.C.; investigation and literature review, S.S., K.A. and L.K.; writing-original draft preparation, S.S.; writing-review and editing, R.C., P.S.S., N.M. and S.R.; supervision, R.C. All authors have read and agreed to the published version of the manuscript.

Funding This study was not supported by any funding.

Data Availability Statement No data were generated or analysed. All data discussed are available in the cited references.

Conflict of interest: The authors declare that they have no conflict of interest.

References

1. Abbo, S., Berger, J., and Turner, N.C. (2003). Evolution of cultivated chickpea: four bottlenecks limit diversity and constrain adaptation. Functional Plant Biology, 30(10), 1081–1087. https://doi.org/10.1071/FP03084
2. Annicchiarico, P., Nazzicari, N., Pecetti, L., et al. (2019). Pea genomic selection for Italian environments. BMC Genomics, 20, 603. https://doi.org/10.1186/s12864-019-5920-x
3. Araus, J.L., and Cairns, J.E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. Trends in Plant Science, 19(1), 52–61. https://doi.org/10.1016/j.tplants.2013.09.008
4. Assefa, T., Assibi Mahama, A., Brown, A.V., et al. (2019). A review of breeding objectives, genomic resources, and marker-assisted methods in common bean (Phaseolus vulgaris L.). Molecular Breeding, 39, 20. https://doi.org/10.1007/s11032-018-0920-0
5. Azodi, C.B., Bolger, E., McCarren, A., et al. (2019). Benchmarking parametric and machine learning models for genomic prediction of complex traits. G3: Genes, Genomes, Genetics, 9(11), 3691–3702. https://doi.org/10.1534/g3.119.400498
6. Azodi, C.B., Pardo, J., VanBuren, R., et al. (2020). Transcriptome-based prediction of complex traits in maize. The Plant Cell, 32(1), 139–151. https://doi.org/10.1105/tpc.19.00332
7. Badhan, S., Ball, A.S., and Mantri, N. (2021). First report of CRISPR/Cas9 mediated DNA-free editing of 4CL and RVE7 genes in chickpea protoplasts. International Journal of Molecular Sciences, 22(1), 396. https://doi.org/10.3390/ijms22010396
8. Bharadwaj, C., Tripathi, S., Soren, K.R., et al. (2021). Introgression of "QTL-hotspot" region enhances drought tolerance and grain yield in three elite chickpea cultivars. The Plant Genome, 14(1), e20076. https://doi.org/10.1002/tpg2.20076
9. Bhat, J.A., Yu, D., Bohra, A., et al. (2021). Features and applications of haplotypes in crop breeding. Communications Biology, 4, 1266. https://doi.org/10.1038/s42003-021-02782-w
10. Bohra, A., Pandey, M.K., Jha, U.C., et al. (2014). Genomics-assisted breeding in four major pulse crops of developing countries: present status and prospects. Theoretical and Applied Genetics, 127(6), 1263–1291. https://doi.org/10.1007/s00122-014-2301-3

11. Bohra, A., Kilian, B., Sivasankar, S., et al. (2022). Reap the crop wild relatives for breeding future crops. Trends in Biotechnology, 40(4), 412–431. https://doi.org/10.1016/j.tibtech.2021.08.009
12. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
13. Chen, K., Wang, Y., Zhang, R., et al. (2019). CRISPR/Cas genome editing and precision plant breeding in agriculture. Annual Review of Plant Biology, 70, 667–697. https://doi.org/10.1146/annurev-arplant-050718-100049
14. Chen, T., and Guestrin, C. (2016). XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785
15. Considine, M.J., Siddique, K.H.M., and Foyer, C.H. (2017). Nature's pulse power: legumes, food security and climate change. Journal of Experimental Botany, 68(8), 1815–1818. https://doi.org/10.1093/jxb/erx099
16. Crossa, J., Pérez-Rodríguez, P., Cuevas, J., et al. (2017). Genomic selection in plant breeding: methods, models, and perspectives. Trends in Plant Science, 22(11), 961–975. https://doi.org/10.1016/j.tplants.2017.08.011
17. Das, S., Upadhyaya, H.D., Bajaj, D., et al. (2015). Deploying QTL-seq for rapid delineation of a potential candidate gene underlying major trait-associated QTL in chickpea. DNA Research, 22(3), 193–203. https://doi.org/10.1093/dnares/dsv004
18. Dash, S., Campbell, J.D., Cannon, E.K.S., et al. (2016). Legume information system (LegumeInfo.org): a key component of a set of federated data resources for the legume family. Nucleic Acids Research, 44(D1), D1181–D1188. https://doi.org/10.1093/nar/gkv1159
19. Doddamani, D., Khan, A.W., Katta, M.A.V.S.K., et al. (2015). CicArVarDB: SNP and InDel database for advancing genetics research and breeding applications in chickpea. Database, 2015, bav078. https://doi.org/10.1093/database/bav078
20. Elshire, R.J., Glaubitz, J.C., Sun, Q., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE, 6(5), e19379. https://doi.org/10.1371/journal.pone.0019379
21. Endelman, J.B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome, 4(3), 250–255. https://doi.org/10.3835/plantgenome2011.08.0024
22. Fernie, A.R., and Schauer, N. (2009). Metabolomics-assisted breeding: a viable option for crop improvement? Trends in Genetics, 25(1), 39–48. https://doi.org/10.1016/j.tig.2008.10.010
23. Foyer, C.H., Lam, H.-M., Nguyen, H.T., et al. (2016). Neglecting legumes has compromised human health and sustainable food production. Nature Plants, 2, 16112. https://doi.org/10.1038/nplants.2016.112
24. Furbank, R.T., and Tester, M. (2011). Phenomics – technologies to relieve the phenotyping bottleneck. Trends in Plant Science, 16(12), 635–644. https://doi.org/10.1016/j.tplants.2011.09.005
25. Gallusci, P., Dai, Z., Génard, M., et al. (2017). Epigenetics for plant improvement: current knowledge and modeling avenues. Trends in Plant Science, 22(7), 610–623. https://doi.org/10.1016/j.tplants.2017.04.009
26. Gao, C. (2021). Genome engineering for crop improvement and future agriculture. Cell, 184(6), 1621–1635. https://doi.org/10.1016/j.cell.2021.01.005
27. Garg, R., Shankar, R., Thakkar, B., et al. (2016). Transcriptome analyses reveal genotype- and developmental stage-specific molecular responses to drought and salinity stresses in chickpea. Scientific Reports, 6, 19228. https://doi.org/10.1038/srep19228
28. Gebremedhin, A., Li, Y., Shunmugam, A.S.K., et al. (2024). Genomic selection for target traits in the Australian lentil breeding program. Frontiers in Plant Science, 14, 1284781. https://doi.org/10.3389/fpls.2023.1284781
29. Ghosal, S., Blystone, D., Singh, A.K., et al. (2018). An explainable deep machine vision framework for plant stress phenotyping. Proceedings of the National Academy of Sciences USA, 115(18), 4613–4618. https://doi.org/10.1073/pnas.1716999115
30. Gianola, D., and van Kaam, J.B.C.H.M. (2008). Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. Genetics, 178(4), 2289–2303. https://doi.org/10.1534/genetics.107.084285
31. Goodstein, D.M., Shu, S., Howson, R., et al. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research, 40(D1), D1178–D1186. https://doi.org/10.1093/nar/gkr944
32. Grant, D., Nelson, R.T., Cannon, S.B., et al. (2010). SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Research, 38(suppl_1), D843–D846. https://doi.org/10.1093/nar/gkp798
33. Habier, D., Fernando, R.L., Kizilkaya, K., et al. (2011). Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics, 12, 186. https://doi.org/10.1186/1471-2105-12-186
34. Harfouche, A.L., Jacobson, D.A., Kainer, D., et al. (2019). Accelerating climate resilient plant breeding by applying next-generation artificial intelligence. Trends in Biotechnology, 37(11), 1217–1235. https://doi.org/10.1016/j.tibtech.2019.05.007
35. Harfouche, A.L., Nakhle, F., Harfouche, A.H., et al. (2023). A primer on artificial intelligence in plant digital phenomics: embarking on the data to insights journey. Trends in Plant Science, 28(2), 154–184. https://doi.org/10.1016/j.tplants.2022.08.021

36. Heffner, E.L., Sorrells, M.E., and Jannink, J.-L. (2009). Genomic selection for crop improvement. Crop Science, 49(1), 1–12. https://doi.org/10.2135/cropsci2008.08.0512
37. Hickey, L.T., Hafeez, A.N., Robinson, H., et al. (2019). Breeding crops to feed 10 billion. Nature Biotechnology, 37(7), 744–754. https://doi.org/10.1038/s41587-019-0152-9
38. Huang, M., Liu, X., Zhou, Y., et al. (2019). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. GigaScience, 8(2), giy154. https://doi.org/10.1093/gigascience/giy154
39. Hughes, D.P., and Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint, arXiv:1511.08060. https://doi.org/10.48550/arXiv.1511.08060
40. Jain, M., Misra, G., Patel, R.K., et al. (2013). A draft genome sequence of the pulse crop chickpea (Cicer arietinum L.). The Plant Journal, 74(5), 715–729. https://doi.org/10.1111/tpj.12173
41. Jarquín, D., Crossa, J., Lacaze, X., et al. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theoretical and Applied Genetics, 127(3), 595–607. https://doi.org/10.1007/s00122-013-2243-1
42. Jayakodi, M., Golicz, A.A., Kreplak, J., et al. (2023). The giant diploid faba genome unlocks variation in a global protein crop. Nature, 615(7953), 652–659. https://doi.org/10.1038/s41586-023-05791-5
43. Jegadeesan, S., Raizada, A., Dhanasekar, P., et al. (2021). Draft genome sequence of the pulse crop blackgram [Vigna mungo (L.) Hepper] reveals potential R-genes. Scientific Reports, 11, 11247. https://doi.org/10.1038/s41598-021-90683-9
44. Jha, U.C., Chaturvedi, S.K., Bohra, A., et al. (2014). Abiotic stresses, constraints and improvement strategies in chickpea. Plant Breeding, 133(2), 163–178. https://doi.org/10.1111/pbr.12150
45. Kale, S.M., Jaganathan, D., Ruperao, P., et al. (2015). Prioritization of candidate genes in "QTL-hotspot" region for drought tolerance in chickpea (Cicer arietinum L.). Scientific Reports, 5, 15296. https://doi.org/10.1038/srep15296
46. Kang, Y.J., Kim, S.K., Kim, M.Y., et al. (2014). Genome sequence of mungbean and insights into evolution within Vigna species. Nature Communications, 5, 5443. https://doi.org/10.1038/ncomms6443
47. Keller, B., Ariza-Suarez, D., de la Hoz, J., et al. (2020). Genomic prediction of agronomic traits in common bean (Phaseolus vulgaris L.) under environmental stress. Frontiers in Plant Science, 11, 1001. https://doi.org/10.3389/fpls.2020.01001
48. Kreplak, J., Madoui, M.-A., Cápal, P., et al. (2019). A reference genome for pea provides insight into legume genome evolution. Nature Genetics, 51(9), 1411–1422. https://doi.org/10.1038/s41588-019-0480-1
49. Kudapa, H., Garg, V., Chitikineni, A., et al. (2018). The RNA-Seq-based high resolution gene expression atlas of chickpea (Cicer arietinum L.) reveals dynamic spatio-temporal changes associated with growth and development. Plant, Cell and Environment, 41(9), 2209–2225. https://doi.org/10.1111/pce.13210
50. Kumar, R., Bohra, A., Pandey, A.K., et al. (2017). Metabolomics for plant improvement: status and prospects. Frontiers in Plant Science, 8, 1302. https://doi.org/10.3389/fpls.2017.01302
51. Langridge, P., and Fleury, D. (2011). Making the most of 'omics' for crop breeding. Trends in Biotechnology, 29(1), 33–40. https://doi.org/10.1016/j.tibtech.2010.09.006
52. LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539
53. Li, Y., Ruperao, P., Batley, J., et al. (2018). Investigating drought tolerance in chickpea using genome-wide association mapping and genomic selection based on whole-genome resequencing data. Frontiers in Plant Science, 9, 190. https://doi.org/10.3389/fpls.2018.00190
54. Libbrecht, M.W., and Noble, W.S. (2015). Machine learning applications in genetics and genomics. Nature Reviews Genetics, 16(6), 321–332. https://doi.org/10.1038/nrg3920
55. Liu, X., Huang, M., Fan, B., et al. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genetics, 12(2), e1005767. https://doi.org/10.1371/journal.pgen.1005767
56. Liu, Y., Wang, D., He, F., et al. (2019). Phenotype prediction and genome-wide association study using deep convolutional neural network of soybean. Frontiers in Genetics, 10, 1091. https://doi.org/10.3389/fgene.2019.01091
57. Liu, Y., Du, H., Li, P., et al. (2020). Pan-genome of wild and cultivated soybeans. Cell, 182(1), 162–176. https://doi.org/10.1016/j.cell.2020.05.023
58. Lonardi, S., Muñoz-Amatriaín, M., Liang, Q., et al. (2019). The genome of cowpea (Vigna unguiculata [L.] Walp.). The Plant Journal, 98(5), 767–782. https://doi.org/10.1111/tpj.14349
59. Lundberg, S.M., and Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774. https://doi.org/10.48550/arXiv.1705.07874
60. Mannur, D.M., Babbar, A., Thudi, M., et al. (2019). Super Annigeri 1 and improved JG 74: two Fusarium wilt-resistant introgression lines developed using marker-assisted backcrossing approach in chickpea (Cicer arietinum L.). Molecular Breeding, 39, 2. https://doi.org/10.1007/s11032-018-0908-9

61. Marzougui, A., Ma, Y., Zhang, C., et al. (2019). Advanced imaging for quantitative evaluation of Aphanomyces root rot resistance in lentil. Frontiers in Plant Science, 10, 383. https://doi.org/10.3389/fpls.2019.00383
62. Merga, B., and Haji, J. (2019). Economic importance of chickpea: production, value, and world trade. Cogent Food and Agriculture, 5(1), 1615718. https://doi.org/10.1080/23311932.2019.1615718
63. Meuwissen, T.H.E., Hayes, B.J., and Goddard, M.E. (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
64. Mohanty, S.P., Hughes, D.P., and Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, 1419. https://doi.org/10.3389/fpls.2016.01419
65. Montesinos-López, O.A., Montesinos-López, A., Pérez-Rodríguez, P., et al. (2021). A review of deep learning applications for genomic selection. BMC Genomics, 22, 19. https://doi.org/10.1186/s12864-020-07319-x
66. Nagasubramanian, K., Jones, S., Singh, A.K., et al. (2019). Plant disease identification using explainable 3D deep learning on hyperspectral images. Plant Methods, 15, 98. https://doi.org/10.1186/s13007-019-0479-8
67. Nair, R.M., Pandey, A.K., War, A.R., et al. (2019). Biotic and abiotic constraints in mungbean production – progress in genetic improvement. Frontiers in Plant Science, 10, 1340. https://doi.org/10.3389/fpls.2019.01340
68. Noble, T.J., Tao, Y., Mace, E.S., et al. (2018). Characterization of linkage disequilibrium and population structure in a mungbean diversity panel. Frontiers in Plant Science, 8, 2102. https://doi.org/10.3389/fpls.2017.02102
69. Ogutu, J.O., Piepho, H.-P., and Schulz-Streeck, T. (2011). A comparison of random forests, boosting and support vector machines for genomic selection. BMC Proceedings, 5(Suppl 3), S11. https://doi.org/10.1186/1753-6561-5-S3-S11
70. Pandey, M.K., Roorkiwal, M., Singh, V.K., et al. (2016). Emerging genomic tools for legume breeding: current status and future prospects. Frontiers in Plant Science, 7, 455. https://doi.org/10.3389/fpls.2016.00455
71. Parmley, K., Nagasubramanian, K., Sarkar, S., et al. (2019). Development of optimized phenomic predictors for efficient plant breeding decisions using phenomic-assisted selection in soybean. Plant Phenomics, 2019, 5809404. https://doi.org/10.34133/2019/5809404
72. Pazhamala, L.T., Purohit, S., Saxena, R.K., et al. (2017). Gene expression atlas of pigeonpea and its application to gain insights into genes associated with pollen fertility implicated in seed formation. Journal of Experimental Botany, 68(8), 2037–2054. https://doi.org/10.1093/jxb/erx010
73. Pérez, P., and de los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. Genetics, 198(2), 483–495. https://doi.org/10.1534/genetics.114.164442
74. Pound, M.P., Atkinson, J.A., Townsend, A.J., et al. (2017). Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. GigaScience, 6(10), gix083. https://doi.org/10.1093/gigascience/gix083
75. Quirós Vargas, J.J., Zhang, C., Smitchger, J.A., et al. (2019). Phenotyping of plant biomass and performance traits using remote sensing techniques in pea (Pisum sativum, L.). Sensors, 19(9), 2031. https://doi.org/10.3390/s19092031
76. Ramalingam, A., Kudapa, H., Pazhamala, L.T., et al. (2015). Proteomics and metabolomics: two emerging areas for legume improvement. Frontiers in Plant Science, 6, 1116. https://doi.org/10.3389/fpls.2015.01116
77. Riedelsheimer, C., Czedik-Eysenberg, A., Grieder, C., et al. (2012). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nature Genetics, 44(2), 217–220. https://doi.org/10.1038/ng.1033
78. Riera, L.G., Carroll, M.E., Zhang, Z., et al. (2021). Deep multiview image fusion for soybean yield estimation in breeding applications. Plant Phenomics, 2021, 9846470. https://doi.org/10.34133/2021/9846470
79. Rincent, R., Charpentier, J.-P., Faivre-Rampant, P., et al. (2018). Phenomic selection is a low-cost and high-throughput method based on indirect predictions: proof of concept on wheat and poplar. G3: Genes, Genomes, Genetics, 8(12), 3961–3972. https://doi.org/10.1534/g3.118.200760
80. Roorkiwal, M., Rathore, A., Das, R.R., et al. (2016). Genome-enabled prediction models for yield related traits in chickpea. Frontiers in Plant Science, 7, 1666. https://doi.org/10.3389/fpls.2016.01666
81. Roorkiwal, M., Jain, A., Kale, S.M., et al. (2018a). Development and evaluation of high-density Axiom® CicerSNP Array for high-resolution genetic mapping and breeding applications in chickpea. Plant Biotechnology Journal, 16(4), 890–901. https://doi.org/10.1111/pbi.12836
82. Roorkiwal, M., Jarquin, D., Singh, M.K., et al. (2018b). Genomic-enabled prediction models using multi-environment trials to estimate the effect of genotype × environment interaction on prediction accuracy in chickpea. Scientific Reports, 8, 11701. https://doi.org/10.1038/s41598-018-30027-2
83. Salt, D.E., Baxter, I., and Lahner, B. (2008). Ionomics and the study of the plant ionome. Annual Review of Plant Biology, 59, 709–733. https://doi.org/10.1146/annurev.arplant.59.032607.092942
84. Samineni, S., Sen, M., Sajja, S.B., et al. (2020). Rapid generation advance (RGA) in chickpea to produce up to seven generations per year and enable speed breeding. The Crop Journal, 8(1), 164–169. https://doi.org/10.1016/j.cj.2019.08.003

85. Sankaran, S., Khot, L.R., Espinoza, C.Z., et al. (2015). Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: a review. European Journal of Agronomy, 70, 112–123. https://doi.org/10.1016/j.eja.2015.07.004
86. Saxena, R.K., Rathore, A., Bohra, A., et al. (2018). Development and application of high-density Axiom Cajanus SNP array with 56K SNPs to understand the genome architecture of released cultivars and founder genotypes. The Plant Genome, 11(3), 180005. https://doi.org/10.3835/plantgenome2018.01.0005
87. Schmutz, J., Cannon, S.B., Schlueter, J., et al. (2010). Genome sequence of the palaeopolyploid soybean. Nature, 463(7278), 178–183. https://doi.org/10.1038/nature08670
88. Schmutz, J., McClean, P.E., Mamidi, S., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. Nature Genetics, 46(7), 707–713. https://doi.org/10.1038/ng.3008
89. Scossa, F., Alseekh, S., and Fernie, A.R. (2021). Integrating multi-omics data for crop improvement. Journal of Plant Physiology, 257, 153352. https://doi.org/10.1016/j.jplph.2020.153352
90. Semba, R.D., Ramsing, R., Rahman, N., et al. (2021). Legumes as a sustainable source of protein in human diets. Global Food Security, 28, 100520. https://doi.org/10.1016/j.gfs.2021.100520
91. Sharma, S., Upadhyaya, H.D., Varshney, R.K., et al. (2013). Pre-breeding for diversification of primary gene pool and genetic enhancement of grain legumes. Frontiers in Plant Science, 4, 309. https://doi.org/10.3389/fpls.2013.00309
92. Singh, A., Ganapathysubramanian, B., Singh, A.K., et al. (2016a). Machine learning for high-throughput stress phenotyping in plants. Trends in Plant Science, 21(2), 110–124. https://doi.org/10.1016/j.tplants.2015.10.015
93. Singh, A.K., Ganapathysubramanian, B., Sarkar, S., et al. (2018). Deep learning for plant stress phenotyping: trends and future perspectives. Trends in Plant Science, 23(10), 883–898. https://doi.org/10.1016/j.tplants.2018.07.004
94. Singh, V.K., Khan, A.W., Saxena, R.K., et al. (2016b). Next-generation sequencing for identification of candidate genes for Fusarium wilt and sterility mosaic disease in pigeonpea (Cajanus cajan). Plant Biotechnology Journal, 14(5), 1183–1194. https://doi.org/10.1111/pbi.12470
95. Smýkal, P., Aubert, G., Burstin, J., et al. (2012). Pea (Pisum sativum L.) in the genomic era. Agronomy, 2(2), 74–115. https://doi.org/10.3390/agronomy2020074
96. Song, Q., Jia, G., Hyten, D.L., et al. (2015). SNP assay development for linkage map construction, anchoring whole-genome sequence, and other genetic and genomic applications in common bean. G3: Genes, Genomes, Genetics, 5(11), 2285–2290. https://doi.org/10.1534/g3.115.020594
97. Stagnari, F., Maggio, A., Galieni, A., et al. (2017). Multiple benefits of legumes for agriculture sustainability: an overview. Chemical and Biological Technologies in Agriculture, 4, 2. https://doi.org/10.1186/s40538-016-0085-1
98. Takagi, H., Abe, A., Yoshida, K., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. The Plant Journal, 74(1), 174–183. https://doi.org/10.1111/tpj.12105
99. Tardieu, F., Cabrera-Bosquet, L., Pridmore, T., et al. (2017). Plant phenomics, from sensors to knowledge. Current Biology, 27(15), R770–R783. https://doi.org/10.1016/j.cub.2017.05.055
100. Tayeh, N., Aluome, C., Falque, M., et al. (2015a). Development of two major resources for pea genomics: the GenoPea 13.2K SNP Array and a high-density, high-resolution consensus genetic map. The Plant Journal, 84(6), 1257–1273. https://doi.org/10.1111/tpj.13070
101. Tayeh, N., Klein, A., Le Paslier, M.-C., et al. (2015b). Genomic prediction in pea: effect of marker density and training population size and composition on prediction accuracy. Frontiers in Plant Science, 6, 941. https://doi.org/10.3389/fpls.2015.00941
102. Thudi, M., Upadhyaya, H.D., Rathore, A., et al. (2014). Genetic dissection of drought and heat tolerance in chickpea through genome-wide and candidate gene-based association mapping approaches. PLoS ONE, 9(5), e96758. https://doi.org/10.1371/journal.pone.0096758
103. Thudi, M., Palakurthi, R., Schnable, J.C., et al. (2021). Genomic resources in plant breeding for sustainable agriculture. Journal of Plant Physiology, 257, 153351. https://doi.org/10.1016/j.jplph.2020.153351
104. Ubbens, J.R., and Stavness, I. (2017). Deep Plant Phenomics: a deep learning platform for complex plant phenotyping tasks. Frontiers in Plant Science, 8, 1190. https://doi.org/10.3389/fpls.2017.01190
105. Vadez, V., Kholová, J., Hummel, G., et al. (2015). LeasyScan: a novel concept combining 3D imaging and lysimetry for high-throughput phenotyping of traits controlling plant water budget. Journal of Experimental Botany, 66(18), 5581–5593. https://doi.org/10.1093/jxb/erv251
106. VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. Journal of Dairy Science, 91(11), 4414–4423. https://doi.org/10.3168/jds.2007-0980
107. Varshney, R.K., Graner, A., and Sorrells, M.E. (2005). Genomics-assisted breeding for crop improvement. Trends in Plant Science, 10(12), 621–630. https://doi.org/10.1016/j.tplants.2005.10.004
108. Varshney, R.K., Close, T.J., Singh, N.K., et al. (2009). Orphan legume crops enter the genomics era! Current Opinion in Plant Biology, 12(2), 202–210. https://doi.org/10.1016/j.pbi.2008.12.004

109. Varshney, R.K., Chen, W., Li, Y., et al. (2012). Draft genome sequence of pigeonpea (Cajanus cajan), an orphan legume crop of resource-poor farmers. Nature Biotechnology, 30(1), 83–89. https://doi.org/10.1038/nbt.2022
110. Varshney, R.K., Gaur, P.M., Chamarthi, S.K., et al. (2013a). Fast-track introgression of "QTL-hotspot" for root traits and other drought tolerance traits in JG 11, an elite and leading variety of chickpea. The Plant Genome, 6(3). https://doi.org/10.3835/plantgenome2013.07.0022
111. Varshney, R.K., Song, C., Saxena, R.K., et al. (2013b). Draft genome sequence of chickpea (Cicer arietinum) provides a resource for trait improvement. Nature Biotechnology, 31(3), 240–246. https://doi.org/10.1038/nbt.2491
112. Varshney, R.K., Thudi, M., Nayak, S.N., et al. (2014). Genetic dissection of drought tolerance in chickpea (Cicer arietinum L.). Theoretical and Applied Genetics, 127(2), 445–462. https://doi.org/10.1007/s00122-013-2230-6
113. Varshney, R.K., Saxena, R.K., Upadhyaya, H.D., et al. (2017). Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. Nature Genetics, 49(7), 1082–1088. https://doi.org/10.1038/ng.3872
114. Varshney, R.K., Thudi, M., Pandey, M.K., et al. (2018). Accelerating genetic gains in legumes for the development of prosperous smallholder agriculture: integrating genomics, phenotyping, systems modelling and agronomy. Journal of Experimental Botany, 69(13), 3293–3312. https://doi.org/10.1093/jxb/ery088
115. Varshney, R.K., Thudi, M., Roorkiwal, M., et al. (2019). Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. Nature Genetics, 51(5), 857–864. https://doi.org/10.1038/s41588-019-0401-3
116. Varshney, R.K., Bohra, A., Roorkiwal, M., et al. (2021a). Fast-forward breeding for a food-secure world. Trends in Genetics, 37(12), 1124–1136. https://doi.org/10.1016/j.tig.2021.08.002
117. Varshney, R.K., Bohra, A., Yu, J., et al. (2021b). Designing future crops: genomics-assisted breeding comes of age. Trends in Plant Science, 26(6), 631–649. https://doi.org/10.1016/j.tplants.2021.03.010
118. Varshney, R.K., Roorkiwal, M., Sun, S., et al. (2021c). A chickpea genetic variation map based on the sequencing of 3,366 genomes. Nature, 599(7886), 622–627. https://doi.org/10.1038/s41586-021-04066-1
119. Verma, M., Kumar, V., Patel, R.K., et al. (2015). CTDB: an integrated chickpea transcriptome database for functional and applied genomics. PLoS ONE, 10(8), e0136880. https://doi.org/10.1371/journal.pone.0136880
120. Vessal, S., Arefian, M., and Siddique, K.H.M. (2020). Proteomic responses to progressive dehydration stress in leaves of chickpea seedlings. BMC Genomics, 21, 523. https://doi.org/10.1186/s12864-020-06930-2
121. Wang, H., Cimen, E., Singh, N., et al. (2020). Deep learning for plant genomics and crop improvement. Current Opinion in Plant Biology, 54, 34–41. https://doi.org/10.1016/j.pbi.2019.12.010
122. Wang, K., Abid, M.A., Rasheed, A., et al. (2023). DNNGP, a deep neural network-based method for genomic prediction using multi-omics data in plants. Molecular Plant, 16(1), 279–293. https://doi.org/10.1016/j.molp.2022.11.004
123. Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics, 10(1), 57–63. https://doi.org/10.1038/nrg2484
124. Watson, A., Ghosh, S., Williams, M.J., et al. (2018). Speed breeding is a powerful tool to accelerate crop research and breeding. Nature Plants, 4(1), 23–29. https://doi.org/10.1038/s41477-017-0083-8
125. Xu, Y. (2016). Envirotyping for deciphering environmental impacts on crop plants. Theoretical and Applied Genetics, 129(4), 653–673. https://doi.org/10.1007/s00122-016-2691-5
126. Xu, Y., Zhang, X., Li, H., et al. (2022). Smart breeding driven by big data, artificial intelligence, and integrated genomic-enviromic prediction. Molecular Plant, 15(11), 1664–1695. https://doi.org/10.1016/j.molp.2022.09.001
127. Yang, T., Liu, R., Luo, Y., et al. (2022). Improved pea reference genome and pan-genome highlight genomic features and evolutionary characteristics. Nature Genetics, 54(10), 1553–1563. https://doi.org/10.1038/s41588-022-01172-2
128. Yang, W., Feng, H., Zhang, X., et al. (2020). Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. Molecular Plant, 13(2), 187–214. https://doi.org/10.1016/j.molp.2020.01.008
129. Yang, Y., Saand, M.A., Huang, L., et al. (2021). Applications of multi-omics technologies for crop improvement. Frontiers in Plant Science, 12, 563953. https://doi.org/10.3389/fpls.2021.563953
130. Yoosefzadeh-Najafabadi, M., Earl, H.J., Tulpan, D., et al. (2021). Application of machine learning algorithms in plant breeding: predicting yield from hyperspectral reflectance in soybean. Frontiers in Plant Science, 11, 624273. https://doi.org/10.3389/fpls.2020.624273
131. Yu, J., Pressoir, G., Briggs, W.H., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics, 38(2), 203–208. https://doi.org/10.1038/ng1702
132. Zhang, C., Craine, W.A., McGee, R.J., et al. (2020). Image-based phenotyping of flowering intensity in cool-season crops. Sensors, 20(5), 1450. https://doi.org/10.3390/s20051450
133. Zhao, C., Zhang, Y., Du, J., et al. (2019). Crop phenomics: current status and perspectives. Frontiers in Plant Science, 10, 714. https://doi.org/10.3389/fpls.2019.00714

134. Zhao, J., Bayer, P.E., Ruperao, P., et al. (2020). Trait associations in the pangenome of pigeon pea (Cajanus cajan). Plant Biotechnology Journal, 18(9), 1946–1959. https://doi.org/10.1111/pbi.13354
135. Zou, J., Huss, M., Abid, A., et al. (2019). A primer on deep learning in genomics. Nature Genetics, 51(1), 12–18. https://doi.org/10.1038/s41588-018-0295-5