Trends in Genetics

Review

Fast-forward breeding for a food-secure world

Rajeev K. Varshney [ID], [1,2,3,*] Abhishek Bohra [ID], [4] Manish Roorkiwal [ID], [1,3] Rutwik Barmukh [ID], [1] Wallace A. Cowling [ID], [3] Annapurna Chitikineni, [1] Hon-Ming Lam, [5] Lee T. Hickey [ID], [6] Janine S. Croser, [3] Philipp E. Bayer, [3,7] David Edwards, [3,7] José Crossa, [8] Wolfram Weckwerth, [9] Harvey Millar, [3,10] Arvind Kumar, [11] Michael W. Bevan, [12] and Kadambot H.M. Siddique [3]

Crop production systems need to expand their outputs sustainably to feed a burgeoning human population. Advances in genome sequencing technologies combined with efficient trait mapping procedures accelerate the availability of beneficial alleles for breeding and research. Enhanced interoperability between different omics and phenotyping platforms, leveraged by evolving machine learning tools, will help provide mechanistic explanations for complex plant traits. Targeted and rapid assembly of beneficial alleles using optimized breeding strategies and precise genome editing techniques could deliver ideal crops for the future. Realizing desired productivity gains in the field is imperative for securing an adequate future food supply for 10 billion people.

Need for food security

Safeguarding a person's right to adequate and nutritious food requires intensive research efforts and innovative solutions to breed nutritious crops with improved productivity and resilience [1]. However, a major challenge is the uneven distribution of resources, resulting in a huge gap in supply and demand for food. Crop productivity and harvest are improved by access to modern infrastructure and technologies, including breeding for improved varieties, agronomic practices, and machinery for farm preparation, harvest, processing, and marketing.

Regions with high populations and low crop production should be studied to address these uneven distribution challenges and provide equitable opportunities. Lessons learned from the pandemic highlight the need for self-sustainability, with less dependence on imports, especially for agriculture. For instance, a vast portion of the entire global population resides in low-income food deficit countries (32.23%), least developed countries (12%), and net food-importing developing countries (20.15%)[i,ii]. Therefore, enhancing crop productivity and addressing the worldwide zero hunger and nutrition food security challenges through modern breeding technologies, infrastructure, agronomic practices, and soil improvement remains essential.

Sequencing and phenotyping technologies for understanding genomic variation

A high-quality **reference genome** (see Glossary) is a prerequisite for genomics studies in a given crop to attain accurate and precise results on crop performance [2]. High-confidence variant calling facilitated by the availability of a high-quality reference genome, is crucial for genetic studies such as gene discovery and manipulation. 'Democratization' of sequencing technologies in concert with advanced informatics tools has improved the contiguity and completeness of existing and genome assemblies. Since a single reference genome cannot capture all genomic variations of a species, an increased number of gold- or platinum-standard reference genomes have become available for several crops. Long-read or linked-read sequencing platforms, such as PacBio, 10X Chromium, and Oxford Nanopore, supplemented by short reads from next-generation sequencing (NGS), allow the assembly of long contigs with high base-to-base precision

Highlights

The rapid advances in plant genome sequencing and phenotyping have enhanced trait mapping and gene discovery in crops.

Increasing adoption of machine learning algorithms is crucial to derive meaningful inferences from complex multidimensional phenotyping data.

Emerging breeding approaches like optimal contribution selection, alone or in combination with genomic selection, will enhance the genetic base of breeding programs while accelerating genetic gain.

Integrating speed breeding with new-age genomic breeding technologies holds promise to relieve the long-standing bottleneck of lengthy crop breeding cycles.

Haplotype-based breeding, genomic prediction, and genome editing will hasten targeted assembly of superior alleles in future cultivars for sustainable agricultural development and long-term food security.

[1]Centre of Excellence in Genomics and Systems Biology, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad 502324, India
[2]State Agricultural Biotechnology Centre, Centre for Crop and Food Innovation, Food Futures Institute, Murdoch University, Murdoch WA 6150, Western Australia, Australia
[3]The UWA Institute of Agriculture, The University of Western Australia, Perth, WA 6009, Australia
[4]ICAR-Indian Institute of Pulses Research (IIPR), Kanpur, India
[5]School of Life Sciences and Center for Soybean Research of the State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Shatin, Hong Kong, China
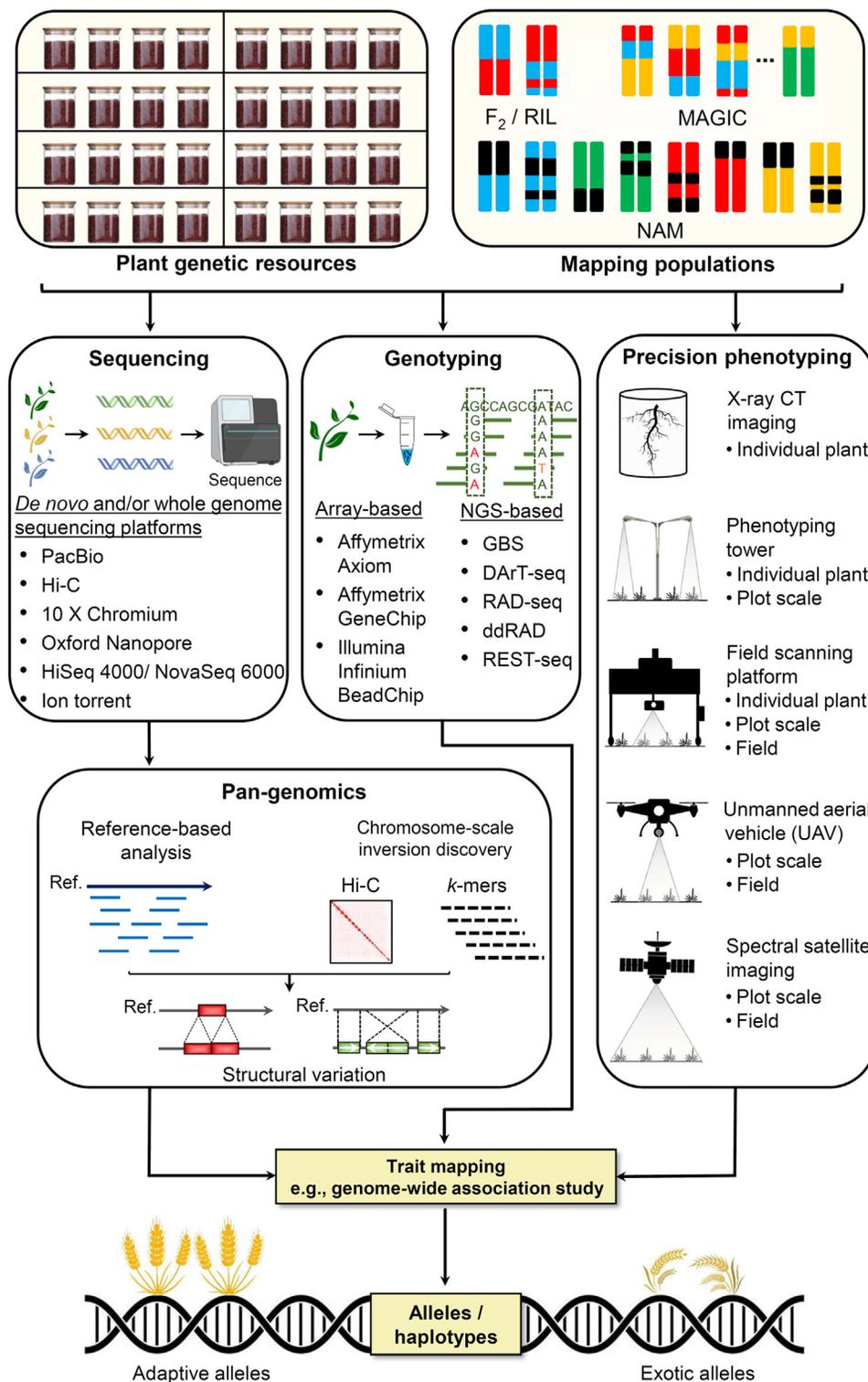
[6]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, QLD, Australia

[7]School of Biological Sciences, The University of Western Australia, Crawley, WA, Australia

[8]International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico

[9]Department of Ecogenomics and Systems Biology, Vienna Metabolomics Center (VIME), University of Vienna, Vienna, Austria

[10]ARC Centre of Excellence in Plant Energy Biology, The University of Western Australia, Crawley, WA, Australia

[11]Deputy Director General's Office, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad 502324, India

[12]John Innes Centre, Norwich Research Park, Norwich, UK

*Correspondence:
r.k.varshney@cgiar.org
rajeev.varshney@murdoch.edu.au
(R.K. Varshney).

(See figure legend at the bottom of the next page.)

(Figure 1). Hi-C sequencing [3] and Bionano Genomics Optical Mapping [4] have facilitated genome assemblies with greater contiguity by dramatically improving **haplotype phasing** and **haplotype scaffolding**, especially in polyploid genomes [5].

Due to the reduction in sequencing costs, high-density genotyping is now affordable for assaying large samples with high precision [6]. A large suite of genotyping platforms (e.g., Affymetrix Axiom, Affymetrix GeneChip, and Illumina Infinium BeadChip) with varying **single nucleotide polymorphisms (SNPs)** are now available for most crop species [7]. Several genome-wide genotyping approaches integrating deep sequencing with reduced representation methods, such as genotyping-by-sequencing, restriction site associated DNA sequencing, double-digest RAD, and restriction fragment sequencing, have led to major innovations in marker discovery in various plant species (Figure 1). Compared with array-based or NGS methods, whole genome resequencing can simultaneously detect known and uncatalogued SNPs and large **structural variations (SVs)**, such as **presence/absence variations (PAVs)** and **copy number variations (CNVs)**.

Concurrent advances in plant phenotyping driven mainly by innovations in image and sensor technologies, with cost- and time-efficient acquisition of massive spatial and temporal data, have leveraged the field of predictive phenomics [8]. Automated platforms equipped with **plant-to-sensor** and **sensor-to-plant** modes can monitor the dynamic response of plant at the organ, whole plant, and field scales (Figure 1). For instance, 3D structural imaging applications, such as X-ray computed tomography, allow *in situ* phenotyping of root system architecture, alleviating underground phenotyping bottlenecks. The growing need to nondestructively monitor plant performance in the real world has led to fully automated field-based phenotyping facilities. In contrast, aerial platforms include unmanned aerial vehicles, manned aerial vehicles, and satellites with varying levels of payload capacity and image resolution (Figure 1) [9]. Fewer than 20% of mechanized phenotyping platforms established worldwide are field-based infrastructures [10]. In African countries, instead of such specialized facilities, setting up field stations and surveying local pathogens with a regional, as opposed to an international foundation, needs to be considered.

Mining superior haplotypes for breeding traits

Fast-tracking mining of beneficial alleles in **plant genetic resources (PGRs)**, such as wild relatives and landraces conserved in genebanks, is crucial for ensuring future food supplies. Current crops have accumulated a suite of traits suitable for modern agriculture and human consumption through human-mediated domestication and improvement of wild species. For trait dissection, a range of genetic populations, including biparental and multiple parental populations such as **multi-parent advanced generation inter-cross (MAGIC) populations** and **nested association mapping populations**, have been developed for many crops [6,11]. With the availability of new sequencing/genotyping technologies, PGRs and genetic populations can be assayed with sequencing and genotyping technologies and evaluated for agronomic and nutrition traits in

Figure 1. Accelerating the discovery of beneficial alleles from plant genetic resources. Plant genetic resources, including accessions archived in genebanks and experimental populations, serve as valuable sources of new genetic variation. Long-read sequencing platforms generate high-quality reference genomes and facilitate pangenomic analyses. High-density genotyping approaches generate genome-wide marker information on these panels. Parallel developments in image and sensor technologies allow acquisition of precise phenotyping data. New genes/haplotypes discovered from analyzing sequence information and phenotyping data will pave the way for enhanced crop improvement. Abbreviations: CT, computed tomography; GBS, genotyping-by-sequencing; MAGIC, multi-parent advanced generation inter-cross; NAM, nested association mapping; RAD-seq, restriction site associated DNA sequencing; REST-seq, restriction fragment sequencing.

Glossary

Artificial neural network: a machine learning algorithm inspired by the human nervous system that includes multiple processing elements, which receive inputs and deliver outputs on the basis of their predefined activation functions.

Breeder's equation: $\Delta G = \frac{h^2 \sigma_p i}{L}$; a simplified form of the quantitative genetics concept to describe change of rate of genetic gain (ΔG) in a given plant breeding population relying on the parameters like heritability (h^2), phenotypic variation within the population (σ_p), selection intensity (i), and length of the breeding cycle (L).

Convolutional neural network: a class of artificial neural networks for processing data containing a grid pattern (e.g., images), which is designed to automatically and adaptively learn spatial hierarchies of features ranging from low- to high-level patterns.

Copy number variation (CNV): sequences that are present in different number of copies between individuals belonging to the same species and encompass duplications, insertions, and deletions.

CRISPR activation: a genome editing method exploiting modified versions of CRISPR effectors devoid of endonuclease activity, with additional transcriptional activators on catalytically inactive Cas9 (dCas9) or single guide RNA (sgRNA).

CRISPR interference: a genome editing method utilizing catalytically inactive Cas9 (dCas9) and single guide RNA (sgRNA) to repress sequence-specific genes without the need of gene knockout.

Deep learning: a subfield of machine learning that deals with algorithms inspired by the structure and function of brain known as artificial neural networks.

Doubled haploid technology: a breeding technique that allows production of completely homozygous lines by chromosome doubling within a short time and thus dramatically reduces the time to establish new cultivars.

Findable, accessible, interoperable, and reusable (FAIR): a concise and measurable set of data principles for helping data producers and publishers to maximize the added-value gained by contemporary, formal scholarly digital publishing.

Genomic selection (GS): a genomics-assisted breeding approach that utilizes genome-wide marker data to predict

Box 1. Pangenomics to bring genes back from the past

Sequencing of multiple plant genomes has popularized the 'pangenome' or set of genes present within a species, including the core genes present in all individuals and dispensable genes absent from at least one individual. The growing literature on multiple genome sequencing and pangenomics has provided strong evidence for the contribution of large SVs, to evolutionary processes that have shaped adaptive diversity in plants. In plants, PAVs within a species range from 7.8% in rice [80] to up to 40% in wheat [81], with many diploid plants having ~30–50% variable genes [82]. *De novo* genome sequencing of wild species accessions has enabled the development of pangenome of different species, called '**super-pangenome**' [83].

Gene PAVs are associated with environmental adaptation, domestication, and breeding [84], and annotations for variable genes in many pangenomes are often enriched for agronomic traits, such as biotic and abiotic stress. Many variable genes have been lost during domestication and breeding bottlenecks; identifying and characterizing these genes can support their targeted reintroduction into breeding programs. Calling PAVs across wild and domesticated lines helps to retrace the impact of domestication and breeding on the pangenome; genes with negative effects can decrease in frequency during selection, while those with agronomic benefits increase (Figure 1). Due to low effective population size in many plant breeding programs and ineffective recombination, many PAVs between wild and cultivated lines may have been lost through genetic drift during breeding bottlenecks.

Knowledge of pangenomes enables us to fine-tune breeding outcomes by constructing the optimal gene content for a crop variety. This requires building species-wide or even genus-wide super-pangenomes representing all genes and allelic variants for breeding the next generation of crops. Pangenomes teach us which haplotypes to combine to produce superior haplotype and gene combinations, enabling plant breeders to shift from useful individuals to useful haplotypes and useful genes when planning breeding programs. Several important agronomic traits have been associated with PAVs and the selective reintroduction of these genes into elite germplasm has led to improved varieties. This knowledge enables breeders to quickly breed cultivars with novel phenotypic attributes by simply backcrossing major genes into elite varieties. Alternatively, GE systems now provide precise molecular tools to modify key genes that have been drivers of crop domestication; for instance, *de novo* domestication of wild rice *Oryza alta* (CCDD) was achieved through gene editing of six agronomically important genes [85].

multiple environments and seasons [2]. Superior genes/alleles for traits of interest can be identified through **pangenomics** (Box 1) or **genome-wide association studies (GWAS)** (Table S1 in the supplemental information online, Figure 1). Concerning modernizing agriculture in developing countries, local needs should be addressed to identify and conserve the germplasm of local crops and wild relatives and undertake genomic breeding programs for accelerating crop improvement.

Systems biology for identifying genes and pathways

Resolving complex quantitative trait loci (QTLs) at the gene level using multi-omics approaches

Transcriptomics, proteomics, metabolomics, and epigenomics provide windows into molecular variation in breeding lines beyond the actual or interpretable genetic variation they contain [12,13]. These windows are closer to phenotype, narrow the genome to phenome divide, and provide independent sets of markers to complement genetic markers as breeding tools (Figure 2).

Associative transcriptomics examines correlations of phenotypic variation with variations in both DNA sequence and transcript abundance [14]. In maize, *cis* expression QTLs (eQTLs) contribute to phenotypic diversity for several traits, including domestication and adaptation [15]. Expression read depth GWAS and transcriptome-wide association studies test associations of mRNA expression with phenotypic diversity [16,17]. Unlike genetic variants, the transcript levels are independent of linkage disequilibrium across the genome; these methods provide deep insights into the regulatory mechanisms of complex traits and enable better prioritization of causal candidate genes [17]. Proteomics approaches can also be used in multiple ways for refining the QTLs underlying complex traits (Box 2).

Alterations in gene expression can be attributed to heritable epigenetic changes that do not involve DNA sequences, including DNA methylation, histone modification, and noncoding RNAs

the genetic worth of an unobserved candidate in a breeding population via estimating the effects of all genetic markers.

Genome-wide association study (GWAS): an approach used in genetics research to detect associations between genetic variants and traits of interest in natural population.

Genomics-assisted breeding: a strategy that integrates genomic tools with high-throughput phenotyping to support breeding practices via molecular markers and to enable prediction of phenotype from genotype.

Haplotypes: a group of alleles within an organism that are inherited together from a single parent.

Haplotype-based breeding: a genomic breeding approach for developing tailored crop varieties, which includes identification of superior haplotypes and their deployment in breeding programs.

Haplotype phasing: the process of reconstruction of haplotype sequences from the genome data.

Haplotype scaffolding: a technique to link together a noncontiguous series of haplotype sequences into a scaffold, which consists of sequences, separated by gaps of known length.

Machine learning (ML): the method of data analysis that provides computers the capability to learn without being explicitly programmed.

Multi-parent advanced generation inter-cross (MAGIC) population: a multi-parent population design in plants includes intercrossing of several founder lines over multiple generations before selfing to develop inbred lines. Unlike biparental populations, MAGIC populations incorporate multiple alleles and provide enhanced recombination and mapping resolution.

Nested association mapping population: an integrated multi-parent population approach that creates a series of interconnected families by crossing multiple founders with a common parent. Like MAGIC, this strategy combines the benefits of linkage mapping and association mapping for high-resolution mapping of complex traits.

Optimal contributions selection (OCS): a selection method that is effective at increasing genetic gain, controlling the rate of inbreeding and enabling maintenance of genetic diversity.

Pangenome: a comprehensive representation of the genetic variation present in the entire species or population as opposed to a single individual. It consists of a core genome, comprising of

[18]. Recent advances in NGS-based protocols, including methylated DNA immunoprecipitation and bisulfite sequencing, have facilitated large-scale analysis of methylation levels, the most common form of epigenetic polymorphisms. In plants, there is evidence for the role of epigenetic modifications on gene expression regulation. High-throughput analysis can screen many lines during breeding cycles, and proteomics and metabolomics have reached a technical standard for application in these studies. Specific metabolomic technologies, such as gas chromatography or liquid chromatography, coupled with mass spectrometry, allow a more cost-effective analysis than NGS techniques [12]. Metabolites are not only important traits in GWAS but are causally involved in stress resistance and thus could serve as important biomarkers that can be rapidly measured and implemented in breeding cycles [19].

Deep learning and artificial intelligence

Machine learning (ML) uses statistical techniques so that computer systems can 'learn' from current data and make predictions on new data [20]. ML tools allow systematic integration of information flowing across the multiple omics layers to provide a comprehensive view of biological mechanisms involving causative genes and regulatory networks (Figure 2). In this context, data-driven network analysis such as multiplex network and the interconnected network would help to elucidate the genes and their complex functional relationships at the systems level [21]. To relieve the 'big-data' challenge while dealing with high-throughput phenotyping data, advanced 'computational ecosystems' backed by deep ML tools allowing storage, processing, and analysis of the data will remain pivotal to derive meaningful inferences from the multidimensional datasets [22]. This is exemplified by the recent use of **artificial neural network/convolutional neural network** and **support vector machine** for the identification and prediction of pests and diseases in plants from high-throughput image/hyperspectral data (Table 1). High-throughput image recognition has fueled the recent advancements in ML [23]. Besides its role in genotype–phenotype associations, the performance of ML tools has been assessed for genome-wide predictions [24], and 'deepGS' has outperformed the traditional **genomic selection (GS)** models in cases where nonadditive variances had significant influence [25]. An array of ML applications in plant science and breeding include its use in defining genomic regions and genome function, regulatory network inference, and understanding the complexity in plant response to stresses [23]. However, the current lack of high-quality labeled data on large populations presents the major challenge in adopting ML for accelerating crop breeding [26]. Auto ML approaches and synthetic data generation may help to alleviate this bottleneck.

Making the most of these multiscale experiments calls for the development of cross-scale meta-analyses [27]. Equally important is enhanced accessibility to highly valued molecular- to field-scale datasets that strictly follow agreed standards, such as **findable, accessible, interoperable, and reusable (FAIR)** norms [28]. Realizing the enormous potential of fast-forward breeding needs effective communication and collaboration among the diversity of disciplines involving end-users, biologists, engineers, data scientists, and manufacturers. Improved understanding of the genomic basis of plant traits of agronomic relevance is crucial for accelerating breeding, along with implementing cutting edge 'black box' approaches, such as GS and ML. For instance, novel insights into global translational reprogramming during growth-to-defense switch in plants [29] paved the way for researchers to use 'TBF1 cassette' to obtain immune-boosted rice with no associated penalty in agronomic performance [30]. Previous research on engineering rice with the *NPR1* gene could enhance disease resistance; the engineered rice had significant fitness costs associated with the resistance.

Accelerated development of crop varieties

Once superior **haplotypes**/alleles and causal genes for agronomic, climate resilience, and nutrition traits are identified by pangenomics, GWAS, and systems biology approaches, the following breeding approaches can be used to accelerate the development of superior varieties (Figure 3).

sequences present in all individuals of the species, and the dispensable genome that is shared by only some individuals.

Pangenomics: the study of all genes and genetic variation within a species.

Plant-to-sensor: a phenotyping technique practiced at high-throughput phenotyping platforms where the imaging station occupies a fixed position during the measurement routine and the plants are transported to the imaging setup.

Plant genetic resources (PGRs): the most valuable and crucial basic plant genetic materials required to meet the current and future needs of crop improvement programs.

Presence/absence variation (PAV): a class of genome structure variation that is used to describe sequences that are present in one genome but entirely missing in the other genome.

Reference genome: a high-quality genome sequence that is characterized by its completeness (less number of gaps), low error rate, and relatively high proportion of sequence assembled into chromosomes.

Sensor-to-plant: a phenotyping technique practiced at high-throughput phenotyping platforms where plants occupy a permanent position during a measurement routine and an imaging setup moves to each of these locations.

Shuttle breeding: a breeding strategy that uses diverse ecological environments to develop improved crop varieties with higher adaptability.

Single nucleotide polymorphism (SNP): genetic variation of a single base pair at a specific position in the genome.

Single seed descent (SSD): a breeding method used with segregating populations of self-pollinated species in which plants are advanced by single seeds from one generation to the next generation.

Speed breeding: a breeding technology that involves growing plant populations under controlled conditions to accelerate generation advancement and shorten the crop breeding cycle.

Structural variations: large-scale structural differences in the genomic DNA, which are inherited and polymorphic in a particular species.

Super-pangenome: an approach of developing a pangenome of the pangenomes of diverse species for a given genus. It offers a comprehensive genomic variation repertoire of a genus and provides exceptional opportunities for crop improvement.

Haplotype-based breeding

While **genomics-assisted breeding** has successfully developed superior cultivars through marker-assisted selection (MAS) and backcrossing [31], most of these studies have identified trait-associated DNA markers using biparental populations. In recent years, whole genome sequence combined with extensive phenotypic records can identify the diversity and structures of key haplotypes associated with breeding decisions and validate their phenotypic effects [32]. The corresponding haplotypes with well-known phenotypic effects are then assembled precisely using the genetic variants that define them. **Haplotype-based breeding** has shown potential for trait improvement in several crops (e.g., rice [33], wheat [34], and pigeonpea [35]). Targeted assembly of haplotypes can reduce the trade-offs of conventional introgression methods to incorporate 'compound' loci into different genetic backgrounds.

Genomic prediction

Advances in sequencing technologies have augmented the speed, throughput, and cost-effectiveness of genotyping. In many cases, it is now cheaper to genotype a breeding line at high density than to evaluate its performance in the field. Access to improved sequencing and genotyping technologies at lower cost has developed ways to leverage genotypic information in breeding programs. The use of new cost-effective genome-wide sequencing combined with precise phenotype data allows calculating genomic estimated breeding values (GEBVs) that help the breeder to identify offspring that can serve as parents for the next generation (improvement cycle). The

> **Support vector machine:** a supervised machine learning algorithm that provides analysis of data for classification and regression analysis.
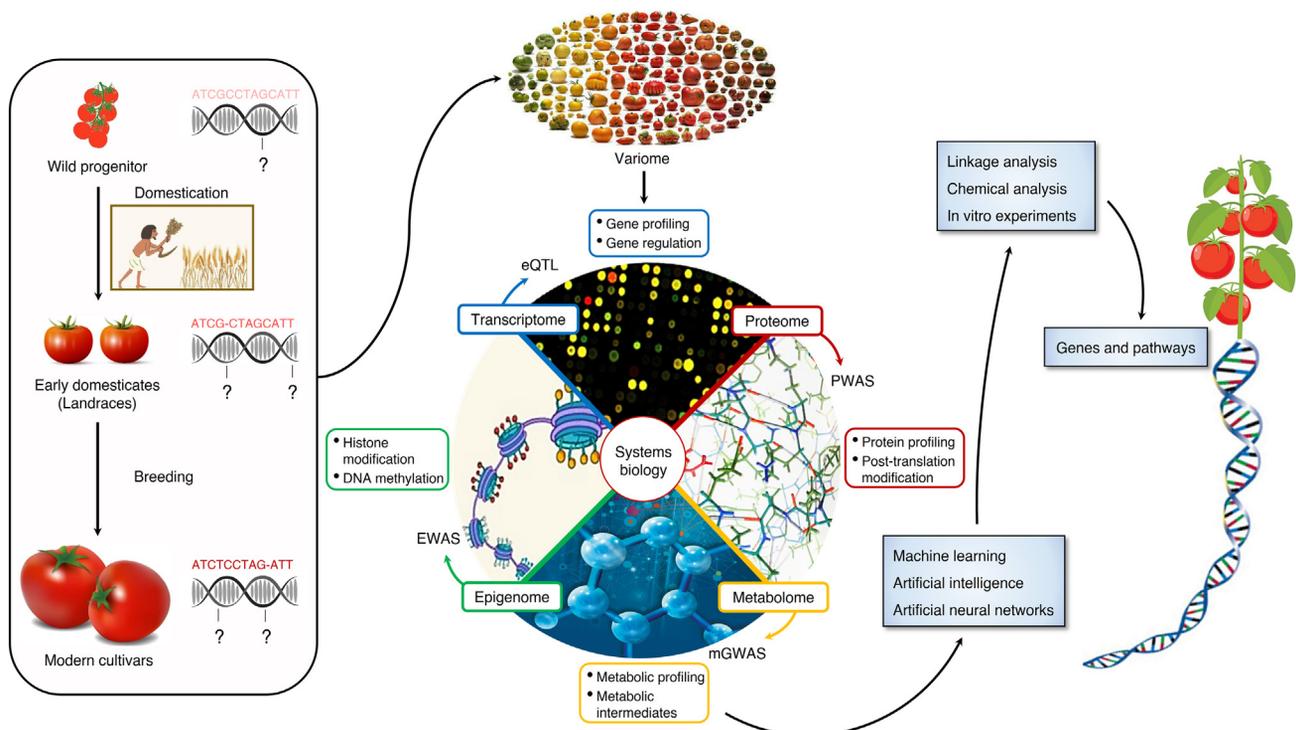
Trends in Genetics

Figure 2. Multi-omics platforms and machine learning tools to develop a systems-level understanding of complex plant phenotypes. Recent advances in high-throughput technologies covering various layers of 'central dogma', including genomics, epigenomics, transcriptomics, proteomics, and metabolomics, facilitate analysis of global genetic diversity (variomes). Modern machine learning tools and artificial intelligence approaches systematically integrate the deluge of information flowing through these multiomics layers. Thus, system-level understanding will help elucidate functional variations and regulatory networks underlying complex phenotypes of agricultural importance. Abbreviations: eQTL, expression qualitative trait locus; EWAS, epigenome-wide association study; mGWAS, metabolite genome-wide association study; PWAS, proteome-wide analysis of SNPs.

Box 2. Proteomics to refine QTLs for complex traits

Proteomics, based on mass spectrometry identification of peptides and matching them to translated gene sequences, can be used for resolving QTLs for complex traits in various ways:

1. Reference crop proteomes can be compared with actual or predicted proteomes of crops to evaluate species or genotype-specific protein differences via their orthology [86,87] and identify INDELS (missing or additional proteins) that are translated in genotypes.
2. Quantitative proteomics approaches can be used in the same way as gene expression studies in breeding for single or multiple gene discoveries associated with specific traits [88]. Proteomics can identify genes of interest that have not been identified through other means because variation in protein abundance can be difficult to detect from gene structure or gene expression studies due to post-transcriptional modification processes [89].
3. Major proteomics-based resources such as protein–protein interaction maps and protein coexpression maps can link gene products as functional units or responses [90]. This can help reduce the solution space to resolve single mechanisms that underlie multiple QTL associations with a trait that may fail to reach statistical significance on its own [91].
4. Specific traits that are traditionally poorly accessible to QTL analysis (e.g., postharvest physiology and seed germination traits) can be sought using proteomics because the accumulated proteome of plant tissues is critical for the timing of traits [92].
5. Traits involving post-translational processes, such as cascades of activation/deactivation by kinases/phosphatases or protein degradation studies, are usually resolved by gene loci studies alone, needing direct analysis of phosphopeptides [93] or protein turnover [94]. Such analyses can provide insight into potential solutions for maintaining target levels (e.g., stabilizing target protein, overexpression of target protein) or altering signaling (phosphomimic alterations), which are important for gene editing in plant breeding).
6. Proteomics now offers data-independent acquisition modes that allow targeted analysis of protein abundance in breeding populations of hundreds to thousands of lines [e.g., sequential window acquisition of all theoretical fragment ion spectra (SWATH) and multiple-reaction monitoring (MRM)] to directly assess protein markers in line selection during breeding cycles (e.g., Jacoby *et al.* [95]).

use of GEBVs in the context of genome-wide prediction promises to help accelerate the rate of genetic gain in breeding [36]. Heffner *et al.* [37] showed that genomic prediction accuracy of 0.5 in maize and wheat increases genetic gains per year in two- to three-fold compared with MAS and traditional selection methods (phenotypic and pedigree based).

The development of sequencing technologies delivered a large amount of marker data, posing challenges when incorporating these into prediction models. GS calculates the GEBV of lines using genome-wide marker profiling and allows the selection of lines prior to field-phenotyping, thereby shortening the breeding cycle. GS techniques are already used in commercial crop breeding programs [38] and are currently being established in many public programs [39]. One of the main advantages of GS is the time saved by selecting parents earlier in the variety development pipeline by predicting the genetic merit of untested individuals or lines. One challenge that GS has already been shown to be well-suited for is the prediction of GEBVs across multiple environments [40]. To accurately make such predictions, GS models are typically augmented with additional terms to account for variability attributable to environments and their interaction with the genotype.

The **breeder's equation** indicates that response to selection is dependent on interactions between selection intensity, the accuracy of prediction, genetic variation, and duration of cycles [36]. Response to selection may be accelerated in the short-term by increasing selection intensity, but linear increases in selection intensity are accompanied by exponential increases in population inbreeding and loss in genetic diversity [41], which compromises long-term genetic gain. GS is very efficient at detecting high-performing crossing candidates, but these tend to be closely related, especially when selection is based on the truncation of GEBVs [42]. In summary, fast-forward breeding for grain yield and abiotic stress tolerance will require some form of assisted ML, based on the evolutionary algorithm (EA) or **deep learning**, to ensure that breeding goals are achieved in the long term.

Table 1. Machine learning for plant phenotyping

Machine learning algorithm	Sensor type	Crop	Trait(s)	Refs
Abiotic stress				
Object-based image analysis	Trichromatic (RGB) images and multispectral unmanned aerial vehicle images	Wild tomato (*Solanum pimpinellifolium*)	Salinity and yield	[67]
NN and RF	RGB images and red, green, near infrared images	Soybean	Iron deficiency chlorosis	[68]
SVM variant	Scanning images	Rice	Nitrogen, phosphorus, and potassium stress	[69]
Biotic stress				
CNN	RGB images	Cucumber	Powdery mildew	[70]
SiVM and SVM	Hyperspectral images	Barley	Powdery mildew	[71]
DCNN	RGB images	Soybean	Bacterial blight (*Pseudomonas savastanoi* pv. *glycinea*), bacterial pustule (*Xanthomonas axonopodis* pv. *glycines*), sudden death syndrome (*Fusarium virguliforme*), *Septoria* brown spot (*Septoria glycines*), frogeye leaf spot (*Cercospora sojina*), iron deficiency chlorosis, potassium deficiency, and herbicide injury	[72]
Genetic algorithm and SVM	Hyperspectral images	Soybean	Charcoal rot	[73]
CNN	RGB images	Maize	Northern leaf blight	[74]
SVM	Hyperspectral images	Rice	Bakanae disease, *Fusarium fujikuroi*	[75]
Crop quality and agronomy				
SVM	Hyperspectral images	Cotton	Crop quality (common types of botanical and nonbotanical foreign matter that are embedded inside the cotton lint)	[76]
CNN	Normalized difference vegetation index images	Lettuce	Crop counting and yield-related traits	[77]
SfM	Multiview images	Sugar beet	Plant height, maximum canopy area, convex hull volume, total leaf area, and individual leaf length.	[78]
FCN and R-CNN	RGB images	Tomato	Plant-part segmentation	[79]

Abbreviations: CNN, convolutional neural network; DCNN, deep convolutional neural network; FCN, fully convolutional network; NN, neural network; R-CNN, region-based convolutional neural network; RF, random forest; SfM, structure from motion; SiVM, simplex volume maximization; SVM, support vector machine.

Optimal contributions selection

Crop improvement programs continue to remain interested in enriching the genetic base with exotic alleles through prebreeding. However, the genetic exchange between exotic and elite pools is hampered by various factors, including linkage drag associated with positive alleles, unanticipated outcomes resulting from exotic loci interaction with elite background, and loss of target locus due to drift in small prebreeding populations [43]. The migration of exotic alleles to the elite pool can be accelerated using efficient approaches, such as **optimal contributions selection (OCS)**, that strike a balance between genetic gain and genetic relatedness in exotic × elite populations. Based on different types of assisted ML, several OCS methods can optimize
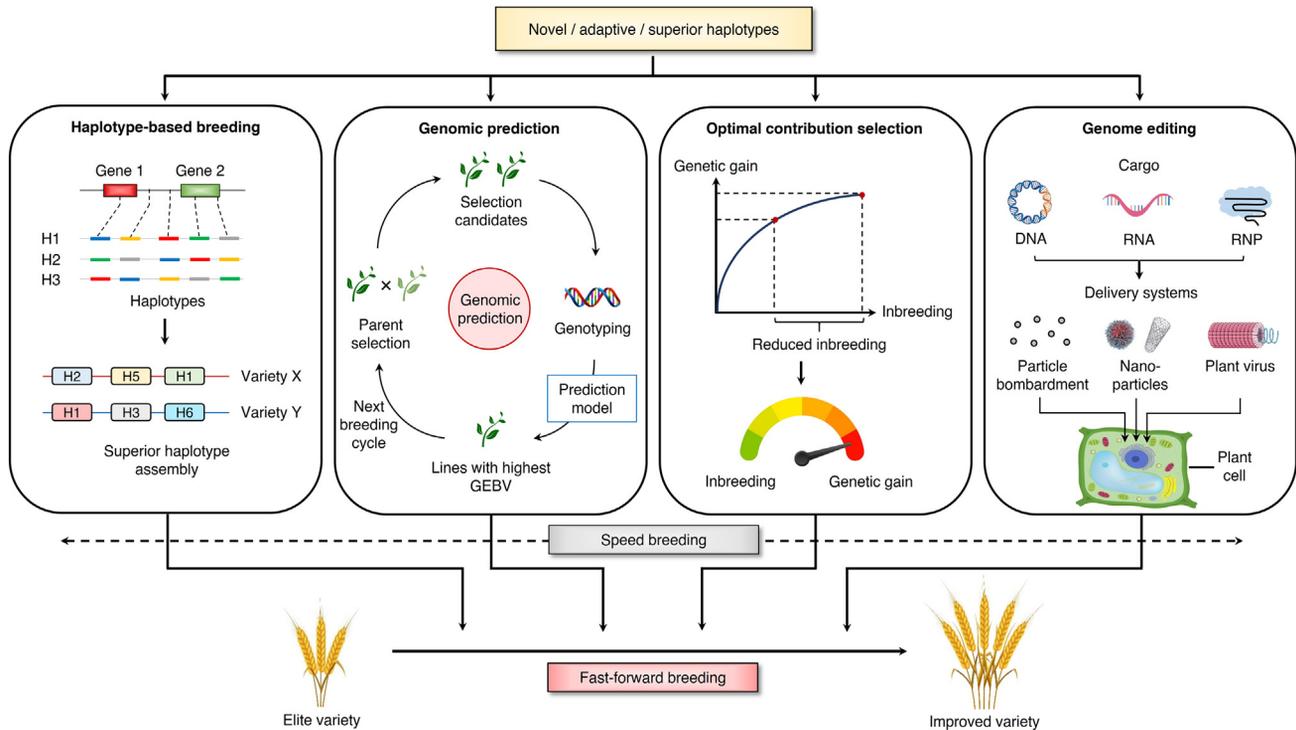
Figure 3. Fast-forward breeding to develop cultivars for future food supply. Rapid, precise, and targeted manipulation of important plant traits is crucial for delivering new cultivars. Haplotype-based breeding and genomic editing approaches rely on a set of significant genes, whereas breeding methods such as genomic selection use genome-wide marker information rather than *a priori* gene-trait associations. Genomic selection is likely to reduce the genetic diversity of a breeding program in the long-term. Hence, maintaining genetic diversity in breeding programs will be crucial for sustaining genetic gains from breeding innovations. Integrating optimal selection contributions with these approaches may help retain genetic diversity while improving genetic gains. Abbreviations: GEBV, genomic estimated breeding value.

contributions from individuals over multiple breeding cycles [44]. An important area of application for assisted ML is the genetic algorithm from which the EA was derived. EA-based search strategies can be integrated with several ML and DL techniques to assist optimization. EA has been adapted for the complex problem of mate selection in animal and plant breeding; one such development uses differential evolution (DE) in the optimization engine [45]. DE-based OCS was used in the simulated genetic improvement of complex traits in self-pollinating grain crops with pedigree and genomic information [42,43]. The same approach was used to model genetic improvement in heat-stress tolerance of wheat, together with grain yield, disease resistance, and other complex traits, to protect wheat yields for the next 60 years of global warming [46]. Simulated stacking of beneficial wheat haplotypes showed the value of genetic algorithms in crossing designs for long-term genetic gain [47].

ML can be used to optimize mating designs with non-inbred lines, reduce cycle time, and accelerate breeding [48]. Typically, non-inbred lines have lower prediction accuracy than pure lines but may improve with pedigree or genomic information as in animal breeding [49]. Cowling *et al*. [43] used pedigree information and Gorjanc *et al*. [42] used genomic information, combined with OCS, to model long-term genetic progress in selfing crops based on S_0 (F_1) recurrent selection. The time taken to double the economic index (including yield, disease resistance, and other economic traits) was similar in S_0- and S_3-derived recurrent selection combined with OCS when accounting for an additional year per cycle selfing the S_3 [43]. These studies confirm that non-inbred

selection is valuable for 'fast-forward breeding' for grain yield and abiotic stress tolerance combined with OCS.

Genome editing

Genome editing (GE), based on CRISPR-Cas9, is a revolutionary tool for editing the plant genome directly. Continuous technological breakthroughs in the GE toolkit have expanded the scope of this technique in crops. For instance, cytosine and adenine base editors are being used to generate point mutations in important field-grown crops [50,51]. Conventional CRISPR-Cas9-mediated delivery technologies, mainly based on DNA, are associated with obnoxious genetic changes [52]. Hence, DNA-free GE holds great potential for developing edited crops with a reduced risk of obnoxious off-target effects. DNA-mediated or DNA-free GE is being applied in crops using delivery systems such as particle bombardment [53], nanoparticles [54], and plant viruses [55] (Figure 3). While the type II CRISPR-Cas9 system is straightforward and efficient, it is limited to target sites upstream of 5'-NGG-3' protospacer adjacent motifs (PAMs). To this end, type V CRISPR-Cas12a (formerly Cpf1) [56] and Cas12b (formerly C2c1) [57] systems can target T-rich PAMs and create cohesive ends for plant GE.

The efficient application of GE in crops requires detailed knowledge of genetic content to minimize off-target effects and optimize phenotypic outcomes. An exciting update to the CRISPR-Cas9 protocol called 'prime editing' has enabled search-and-replace-editing instead of single base substitutions [58], allowing for small insertions, deletions, point mutations, and combination edits. This mechanism has so far been used in human cells but is expected to work in plants. This has brought us one step closer to the complete fine-grained control of gene content in crops. **Pangenomes** can teach us about the required prime editing changes. It is feasible to use prime editing to transform one resistance gene into a different allelic version, as identified from the pangenome, than cross-breed that resistance gene into elite lines, since cross-breeding takes and can lead to random changes in the resulting cultivars. Furthermore, GE plays a key role in regulating gene expression in crops. For example, **CRISPR activation** and **CRISPR interference** systems have been used to activate or repress transcription of plant genes with Cas9 [59], Cas12a [56], and Cas12b [57]. With the advances being made at a breathtaking pace, GE holds enormous potential in expediting design-based crop improvement and in meeting future food security.

Speeding up the breeding cycle

While any of the above-mentioned genomic breeding approaches can be used for trait improvement, it is important to reduce generation cycle time to enhance the rate of genetic gain [60]. In the last half of the 20th century, **doubled haploid technology** revolutionized the time required to achieve genome fixation for many crop species. Unfortunately, most key crops in developing countries remain largely recalcitrant to doubled haploid techniques [61]. Further, in species responsive to doubled haploidy, the linkage of genes, need for further recombination, or lack of specialized tissue culture facilities demands an alternative approach to gene fixation. In such situations, modified pedigree breeding methods, such as **single seed descent (SSD)**, facilitated faster generation cycling and, combined with **shuttle breeding** between complementary environments, were a key component of genetic improvements that led to crop productivity gains during the Green Revolution. Accelerated SSD can be achieved in some plant species by rapidly completing the full plant cycle *in vitro*. Difficulties with *in vitro* genotype dependence, the requirement for specialized facilities, and the need to reduce costs per plant for large-scale breeding have driven the development of alternative platforms, such as fast generation cycling system [62] and **speed breeding** [60]. Species important within global agricultural systems from the Poaceae, Fabaceae, and Brassicaceae families are amenable to rapid generation turnover in

soil under controlled environment conditions. These 'speed breeding' conditions can be rudimentary and within reach of all genetic improvement programs (e.g., day length extension improving generation turnover by one or two generations per year).

More complex SSD platforms have recently arisen that use high-quality controlled environments [62] or a combination of temperature and photoperiod control with light quality manipulation to exploit ancient plant light responses, such as shade avoidance, to achieve rapid, synchronized flowering across varied phenotypes [63,64]. The optimal conditions for rapid elicitation of flowering time vary with species; however, there are opportunities within related species to create broadly applicable growth conditions. The accelerated SSD strategy is particularly beneficial in situations where simultaneous selection is required for several characteristics with different heritability and is more efficient than conventional in-season generation turnover.

'Speed breeding' or other 'SSD platforms' can be combined with proposed breeding approaches such as OCS, HHB, GS, and GE. For instance, simulation studies suggest that combining GS with rapid gene fixation techniques can reduce the length of the breeding cycle and maximize genetic gain per unit time [47,65]. This represents a powerful new integrative approach to plant improvement, particularly in neglected species such as food legumes, and represents the next step-change in genetic improvement.

Concluding remarks

Extracting novel genetic variation from wide germplasm for plant breeding applications has now become a reality, owing to the recent breakthroughs in sequencing and phenotyping. In this context, learning-based approaches allow understanding and integration of large-scale datasets to find meaningful patterns for guiding future breeding strategies. Also, GE in combination with pangenomics and systems biology provides an alternative route to engineer designer crops. As the paradigm shifts from individual DNA marker to haplotypes, future cultivar development will be driven by the breeding strategies that expeditiously create and incorporate superior haplotypes in breeding populations. The efficient crop breeding programs with an enhanced genetic base will accelerate the progress of forward breeding. By rapid production of climate-resilient crop varieties, the above-mentioned approaches offer huge opportunities to improve breeder's response to growing challenges that crop improvement faces. While private sector and several public sector breeding programs are already deploying these approaches in crop improvement programs, public breeding programs, especially in developing countries, need to accelerate adoption of these approaches. However, there are still some outstanding questions on the utilization of these approaches for molecular dissection of complex traits and development of climate resilient varieties in the face of a rapidly changing world (see Outstanding questions). Furthermore, it is important to note that sustainable food security requires more than development of superior varieties. For example, a robust seed system for delivering improved varieties to replace old cultivars, appropriate agronomy practices, and mechanization is required for harvesting higher produce [66]. Similarly, farmers' access to better markets, value addition, and food processing will generate more income to farmers and deliver better products to consumers. Notwithstanding this, the approaches mentioned here and their integration will fast-forward breeding for accelerated crop improvement to contribute to a food-secure world.

Outstanding questions

Transcriptomic, proteomic, and metabolic variations during crop domestication and improvement processes are different in diverse species. Can these variations be dissected at a greater resolution by including additional landraces/wild relatives to identify if the domestication process triggered similar changes at the transcriptome, proteome, or metabolome level? Can such variations contribute towards designing climate-resilient crops for the future?

What are the effective methods to integrate CRISPR/Cas systems into large-scale breeding programs for accelerating crop genetic improvement?

How can the rate of deployment of machine learning algorithms be enhanced across different crops and program goals to help diverse research objectives by developing beneficial evaluation technologies?

To what extent can present and future genetic gain in crop productivity be fully realized by optimizing the length of breeding cycles via speed breeding and its integration with modern genomic breeding technologies?

How can we ensure that modern genomic breeding technologies will actually reach the resource poor via favorable technology transfer processes?

Declaration of interests

No interests are declared.

Supplemental information

Supplemental information associated with this article can be found online https://doi.org/10.1016/j.tig.2021.08.002.

Resources

[i]www.worldometers.info/world-population/population-by-country/

[ii]www.fao.org/News/FACTFILE/IMG/FF9712-e.pdf

References

1. Siddique, K.H.M. *et al.* (2021) Re-discovering Asia's forgotten crops to fight chronic and hidden hunger. *Nat. Plants* 7, 116–122
2. Varshney, R.K. *et al.* (2020) 5Gs for crop genetic improvement. *Curr. Opin. Plant Biol.* 56, 190–196
3. Belton, J.M. *et al.* (2012) Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* 58, 268–276
4. Pendleton, M. *et al.* (2015) Assembly and diploid architecture of an individual human genome via single-molecule technologies. *Nat. Methods* 12, 780–786
5. Zhuang, W. *et al.* (2019) The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat. Genet.* 51, 865–876
6. Huang, B.E. *et al.* (2015) MAGIC populations in crops: current status and future prospects. *Theor. Appl. Genet.* 128, 999–1017
7. Rasheed, A. *et al.* (2017) Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. *Mol. Plant* 10, 1047–1064
8. Mir, R.R. *et al.* (2012) Integrated genomics, physiology and breeding approaches for improving drought tolerance in crops. *Theor. Appl. Genet.* 125, 625–645
9. Jin, X. *et al.* (2020) High-throughput estimation of crop traits: a review of ground and aerial phenotyping platforms. *IEEE Trans. Geosci. Remote Sens.* 9, 200–231
10. Yang, W. *et al.* (2020) Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Mol. Plant* 13, 187–214
11. Scott, M.F. *et al.* (2020) Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. *Heredity* 125, 396–416
12. Weckwerth, W. *et al.* (2020) PANOMICS meets germplasm. *Plant Biotechnol. J.* 18, 1507–1525
13. Pazhamala, L.T. *et al.* (2021) Systems biology for crop improvement: prospects and challenges. *Plant Genome* 14, e20098
14. Harper, A.L. *et al.* (2012) Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. *Nat. Biotechnol.* 30, 798–802
15. Lemmon, Z.H. *et al.* (2014) The role of cis regulatory evolution in maize domestication. *PLoS Genet.* 10, e1004745
16. Lin, H. *et al.* (2017) Substantial contribution of genetic variation in the expression of transcription factors to phenotypic variation revealed by eRD-GWAS. *Genome Biol.* 18, 192
17. Kremling, K.A.G. *et al.* (2019) Transcriptome-wide association supplements genome-wide association in *Zea mays*. *G3 (Bethesda)* 9, 3023–3033
18. Hu, Y. *et al.* (2015) Prediction of plant height in *Arabidopsis thaliana* using DNA methylation data. *Genetics* 201, 779–793
19. Ghatak, A. *et al.* (2018) Metabolomics in plant stress physiology. In *Plant Genetics and Molecular Biology* (Varshney, R.K. *et al.*, eds), pp. 187–236, Springer
20. Samuel, A.L. (1959) Some studies in machine learning using the game of checkers. *IBM J. Res. Dev.* 3, 210–229
21. Hass, H. *et al.* (2017) Mathematical model of early Reelin-induced Src family kinase-mediated signaling. *PLoS One* 12, 1–16
22. Singh, A. *et al.* (2016) Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci.* 21, 110–124
23. van Dijk, A.D.J. *et al.* (2021) Machine learning in plant science and plant breeding. *iScience* 24, 101890
24. Liu, F. *et al.* (2016) PEDLA: predicting enhancers with a deep learning-based algorithmic framework. *Sci. Rep.* 6, 28517
25. Abdollahi-Arpanahi, R. *et al.* (2020) Deep learning versus parametric and ensemble methods for genomic prediction of complex phenotypes. *Genet. Sel. Evol.* 52, 12
26. Bayer, P.E. and Edwards, D. (2020) Machine learning in agriculture: from silos to marketplaces. *Plant Biotechnol. J.* 19, 648–650
27. Tardieu, F. *et al.* (2017) Plant phenomics, from sensors to knowledge. *Curr. Biol.* 27, 770–783
28. Wilkinson, M.D. *et al.* (2016) The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3, 160018
29. Xu, G. *et al.* (2017) Global translational reprogramming is a fundamental layer of immune regulation in plants. *Nature* 545, 487–490
30. Xu, G. *et al.* (2017) uORF-mediated translation allows engineered plant disease resistance without fitness costs. *Nature* 545, 491–494
31. Varshney, R.K. *et al.* (2021) Designing future crops: genomics-assisted breeding comes of age. *Trends Plant Sci.* 26, 631–649
32. Bevan, M.W. *et al.* (2017) Genomic innovation for crop improvement. *Nature* 543, 346–354
33. Abbai, R. *et al.* (2019) Haplotype analysis of key genes governing grain yield and quality traits across 3K RG panel reveals scope for the development of tailor-made rice with enhanced genetic gains. *Plant Biotechnol. J.* 17, 1612–1622
34. Brinton, J. *et al.* (2020) A haplotype-led approach to increase the precision of wheat breeding. *Commun. Biol.* 3, 712
35. Sinha, P. *et al.* (2020) Superior haplotypes for haplotype-based breeding for drought tolerance in pigeonpea (*Cajanus cajan* L.). *Plant Biotechnol. J.* 18, 2482–2490
36. Sinha, P. *et al.* (2021) Genomics and breeding innovations for enhancing genetic gain for climate resilience and nutrition traits. *Theor. Appl. Genet.* 134, 1829–1843
37. Heffner, E.L. *et al.* (2011) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome* 4, 65–75
38. Gaffney, J. *et al.* (2015) Industry-scale evaluation of maize hybrids selected for increased yield in drought-stress conditions of the US Corn Belt. *Crop Sci.* 55, 1608–1618
39. Belamkar, V. *et al.* (2018) Genomic selection in preliminary yield trials in a winter wheat breeding program. *G3 (Bethesda)* 8, 2735–2747
40. Burgueno, J. *et al.* (2012) Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. *Crop Sci.* 52, 707–719
41. Woolliams, J.A. and Meuwissen, T.H.E. (1993) Decision rules and variance of response in breeding schemes. *Anim. Prod.* 56, 179–186
42. Gorjanc, G. *et al.* (2018) Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. *Theor. Appl. Genet.* 131, 1953–1966

43. Cowling, W.A. *et al.* (2017) Evolving gene banks: improving diverse populations of crop and exotic germplasm with optimal contribution selection. *J. Exp. Bot.* 68, 1927–1939

44. Woolliams, J.A. *et al.* (2015) Genetic contributions and their optimization. *J. Anim. Breed. Genet.* 132, 89–99

45. Kinghorn, B.P. (2011) An algorithm for efficient constrained mate selection. *Genet. Sel. Evol.* 43, 4

46. Cowling, W.A. *et al.* (2019) Modeling crop breeding for global food security during climate change. *Food Energy Secur.* 8, e00157

47. Voss-Fels, K.P. *et al.* (2019) Breeding improves wheat productivity under contrasting agrochemical input levels. *Nat. Plants* 5, 706–714

48. Cobb, J.N. *et al.* (2019) Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. *Theor. Appl. Genet.* 132, 627–645

49. Goddard, M.E. and Hayes, B.J. (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10, 381–391

50. Yan, F. *et al.* (2018) High-efficient A-T to G-C base editing by Cas9n-guided tRNA adenosine deaminase in rice. *Mol. Plant* 11, 631–634

51. Li, C. *et al.* (2018) Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol.* 19, 59

52. Zhang, Y. *et al.* (2018) Applications and potential of genome editing in crop improvement. *Genome Biol.* 19, 210

53. Liang, Z. *et al.* (2017) Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nat. Commun.* 8, 14261

54. Lv, Z. *et al.* (2020) Nanoparticle-mediated gene transformation strategies for plant genetic engineering. *Plant J.* 104, 880–891

55. Ma, X. *et al.* (2020) Highly efficient DNA-free plant genome editing using virally delivered CRISPR–Cas9. *Nat. Plants* 6, 773–779

56. Tang, X. *et al.* (2017) A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat. Plants* 3, 17018

57. Ming, M. *et al.* (2020) CRISPR–Cas12b enables efficient plant genome engineering. *Nat. Plants* 6, 202–208

58. Anzalone, A.V. *et al.* (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157

59. Papikian, A. *et al.* (2019) Site-specific manipulation of *Arabidopsis* loci using CRISPR-Cas9 SunTag systems. *Nat. Commun.* 10, 729

60. Watson, A. *et al.* (2018) Speed breeding is a powerful tool to accelerate crop research and breeding. *Nat. Plants* 4, 23–29

61. Croser, J.S. *et al.* (2007) Toward doubled haploid production in the Fabaceae: progress, constraints, and opportunities. *Crit. Rev. Plant Sci.* 25, 139–157

62. Yan, G. *et al.* (2017) Accelerated generation of selfed pure line plants for gene identification and crop breeding. *Front. Plant Sci.* 8, 1786

63. Croser, J.S. *et al.* (2016) Time to flowering of temperate pulses in vivo and generation turnover in vivo-in vitro of narrow leaf lupin accelerated by low red to far red ratio and high intensity in the far red region. *Plant Cell Tissue Organ Cult.* 127, 591–599

64. Jähne, F. *et al.* (2020) Speed breeding short-day crops by LED-controlled light schemes. *Theor. Appl. Genet.* 133, 2335–2342

65. Jighly, A. *et al.* (2019) Boosting genetic gain in allogamous crops via speed breeding and genomic selection. *Front. Plant Sci.* 10, 1364

66. Varshney, R.K. *et al.* (2018) Accelerating genetic gains in legumes for the development of prosperous smallholder agriculture: integrating genomics, phenotyping, systems modelling and agronomy. *J. Exp. Bot.* 69, 3293–3312

67. Johansen, K. *et al.* (2019) Unmanned aerial vehicle-based phenotyping using morphometric and spectral analysis can quantify responses of wild tomato plants to salinity stress. *Front. Plant Sci.* 10, 370

68. Dobbels, A.A. and Lorenz, A.J. (2019) Soybean iron deficiency chlorosis high throughput phenotyping using an unmanned aircraft system. *Plant Methods* 15, 97

69. Chen, L. *et al.* (2014) Identification of nitrogen, phosphorus, and potassium deficiencies in rice based on static scanning technology and hierarchical identification method. *PLoS One* 9, e113200

70. Lin, K. *et al.* (2019) Deep learning-based segmentation and quantification of cucumber powdery mildew using convolutional neural network. *Front. Plant Sci.* 10, 155

71. Thomas, S. *et al.* (2018) Quantitative assessment of disease severity and rating of barley cultivars based on hyperspectral imaging in a non-invasive, automated phenotyping platform. *Plant Methods* 14, 45

72. Ghosal, S. *et al.* (2018) An explainable deep machine vision framework for plant stress phenotyping. *Proc. Natl. Acad. Sci. U. S. A.* 115, 4613–4618

73. Nagasubramanian, K. *et al.* (2018) Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems. *Plant Methods* 14, 86

74. DeChant, C. *et al.* (2017) Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. *Phytopathology* 107, 1426

75. Chung, C.L. *et al.* (2016) Detecting Bakanae disease in rice seedlings by machine vision. *Comput. Electron. Agric.* 121, 404–411

76. Zhang, M. *et al.* (2017) Classification of foreign matter embedded inside cotton lint using short wave infrared (SWIR) hyperspectral transmittance imaging. *Comput. Electron. Agric.* 139, 75–90

77. Bauer, A. *et al.* (2019) Combining computer vision and deep learning to enable ultra-scale aerial phenotyping and precision agriculture: a case study of lettuce production. *Hortic. Res.* 6, 70

78. Xiao, S. *et al.* (2020) Image-based dynamic quantification of aboveground structure of sugar beet in field. *Remote Sens.* 12, 269

79. Shi, W. *et al.* (2019) Plant-part segmentation using deep learning and multi-view vision. *Biosyst. Eng.* 187, 81–95

80. Schatz, M.C. *et al.* (2014) Whole genome *de novo* assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of aus and indica. *Genome Biol.* 15, 506

81. Montenegro, J.D. *et al.* (2017) The pangenome of hexaploid bread wheat. *Plant J.* 90, 1007–1013

82. Gao, L. *et al.* (2019) The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* 51, 1044–1051

83. Khan, A.W. *et al.* (2020) Super-pangenome by integrating the wild side of a species for accelerated crop improvement. *Trends Plant Sci.* 25, 148–158

84. Tan, S. *et al.* (2012) Variation of presence/absence genes among *Arabidopsis* populations. *BMC Evol. Biol.* 12, 86

85. Yu, H. *et al.* (2021) A route to *de novo* domestication of wild allotetraploid rice. *Cell* 184, 1156–1170

86. Ghatak, A. *et al.* (2017) Proteomics survey of Solanaceae family: current status and challenges ahead. *J. Proteome* 169, 41–57

87. Hooper, C.M. *et al.* (2020) CropPAL for discovering protein subcellular location divergence in crops to support strategies for molecular crop breeding. *Plant J.* 104, 812–827

88. Hoehenwarter, W. *et al.* (2011) MAPA distinguishes genotype-specific variability of highly similar regulatory protein isoforms in potato tuber. *J. Proteome Res.* 7, 2979–2991

89. Millar, A.H. *et al.* (2019) The scope, functions, and dynamics of posttranslational protein modifications. *Annu. Rev. Plant Biol.* 70, 119–151

90. Duncan, O. *et al.* (2017) Resource: mapping the *Triticum aestivum* proteome. *Plant J.* 89, 601–616

91. Weckwerth, W. *et al.* (2004) Differential metabolic networks unravel the effects of silent plant phenotypes. *Proc. Natl Acad. Sci. USA* 101, 7809–7814

92. Vanderschuren, H. *et al.* (2013) Proteomics of model and crop plant species: status, current limitations and strategic advances for crop improvement. *J. Proteome* 93, 5–19

93. Chen, Y. and Weckwerth, W. (2020) Mass spectrometry untangles plant membrane protein signaling networks. *Trends Plant Sci.* 25, 930–944

94. Nelson, C.J. and Millar, H. (2015) Protein turnover in plant biology. *Nat. Plants* 1, 15017

95. Jacoby, R.P. *et al.* (2013) Application of selected reaction monitoring mass spectrometry to field-grown crop plants to allow dissection of the molecular mechanisms of abiotic stress tolerance. *Front. Plant Sci.* 4, 20