Genomic interventions for sustainable agriculture

Abhishek Bohra¹,* 1, Uday Chand Jha¹, Ian D. Godwin² and Rajeev Kumar Varshney³,4,* 2

¹ICAR-Indian Institute of Pulses Research (IIPR), Kanpur, India
²Centre for Crop Science, Queensland Alliance for Agriculture and Food Innovation (QAAFI), The University of Queensland, Brisbane, Qld, Australia
³International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India
⁴The UWA Institute of Agriculture, The University of Western Australia, Perth, Australia

Summary
Agricultural production faces a Herculean challenge to feed the increasing global population. Food production systems need to deliver more with finite land and water resources while exerting the least negative influence on the ecosystem. The unpredictability of climate change and consequent changes in pests/pathogens dynamics aggravate the enormity of the challenge. Crop improvement has made significant contributions towards food security, and breeding climate-smart cultivars are considered the most sustainable way to accelerate food production. However, a fundamental change is needed in the conventional breeding framework in order to respond adequately to the growing food demands. Progress in genomics has provided new concepts and tools that hold promise to make plant breeding procedures more precise and efficient. For instance, reference genome assemblies in combination with germplasm sequencing delineate breeding targets that could contribute to securing future food supply. In this review, we highlight key breakthroughs in plant genome sequencing and explain how the presence of these genome resources in combination with gene editing techniques has revolutionized the procedures of trait discovery and manipulation. Adoption of new approaches such as speed breeding, genomic selection and haplotype-based breeding could overcome several limitations of conventional breeding. We advocate that strengthening varietal release and seed distribution systems will play a more determining role in delivering genetic gains at farmer’s field. A holistic approach outlined here would be crucial to deliver steady stream of climate-smart crop cultivars for sustainable agriculture.

Keywords: genetic gains, genome sequencing, genomic selection, gene editing, speed breeding, varietal turnover, seed replacement.

Introduction
The current food production systems are under immense pressure to double their productivity in order to feed the ever-increasing global population. The current annual yield gains (~1%) reported for major crops, that is wheat, rice, maize and soybean remain less than what is projected (~2.4%) to reach the goal of doubling global production (Ray et al., 2013). Climate change further aggravates the challenge that the global food production system is facing, and the global yields of aforementioned commodities are likely to reduce in response to every degree Celsius rise in global mean temperature (Varshney et al., 2020). Importantly, this remarkable increase in food production has to be achieved with finite or even depleting land resources and water systems, while meeting the demand for ecosystem preservation (Ronald, 2014). Prevalence of extreme weather conditions is projected to influence pests/pathogens dynamics and compromise the plant defence response (Atlin et al., 2017).

Traditional plant breeding systems have been in place for decades and delivered a series of widely adopted high-yielding crop cultivars worldwide. However, longer time invested in variety development and breeding cycles presents a stumbling block to an accelerated response of plant breeders to growing demands for food production (Lenaerts et al., 2019). Improving the rates of crop productivity through breeding seeks transformational changes in our current plant breeding operations and decisions (Santantonio et al., 2020). Recent progress in genomics technologies has imparted greater strength to the breeders’ toolbox (Bohra et al., 2014a,b, 2020; Bohra and Singh, 2015; Varshney et al., 2019a). In this review, we highlight the key milestones in plant genome sequencing and discuss how sequencing data have helped illuminate trait architectures and trait alteration. Genomics technologies, when accommodated within new methods like gene editing, rapid generation turnover, including genomic selection and haplotype-based breeding are likely to increase the rate of genetic gains in breeding programmes. We also underline the significance of varietal release and seed distribution systems in pursuing our goal of sustainable food production.

Key breakthroughs in plant genome sequencing
A contiguous and well-annotated genome sequence is the foundation for downstream analyses such as gene/trait discovery, genome dynamics, phylogenetic and evolutionary studies, and better cataloguing of repeat elements (van de Peer, 2018). Advances in DNA sequencing technologies have paved way for decoding of whole genomes for a variety of plant species. Currently, over 400 genomes of different species of land plants are deposited in GenBank (https://www.ncbi.nlm.nih.gov/).
In 2000, Arabidopsis became the first multicellular organism sequenced by a multinational consortium using a bacterial artificial chromosome (BAC)-by-BAC approach that relies on construction of a minimum tiling path (MTP) based on overlapping BAC clones (AGI, 2000). As reviewed by Kersey (2019), the Arabidopsis genome assembly is of the highest accuracy ‘gold standard’, with the latest version having only 161 gaps. A similar BAC-based approach was used to sequence the rice crop in 2005 (IRGSP, 2005). A technological breakthrough in genome sequencing was achieved with the whole genome shotgun (WGS) strategy in which the genomic DNA is sheared followed by sequencing and assembly of these fragments. For instance, Tuscan et al. (2006) assembled 434.29 Mb genome of poplar (Populus trichocarpa) using WGS strategy. However, this strategy yielded a fragmented assembly and proved costly at that time due to its reliance on Sanger chemistry (Bolger et al., 2014).

Post-Sanger sequencing approaches based on next-generation sequencing (NGS) leveraged the WGS strategy by dramatically improving sequencing throughput at a much reduced time and cost for genome sequencing projects (Varshney et al., 2009). The first plant genomes that were created using a combination of Sanger and NGS approaches were grape (Velasco et al., 2007) and cucumber (Huang et al., 2009a), with short reads generated, respectively, by 454 and Illumina platforms. The first de novo whole genome assembly created solely with short-read technologies was strawberry (Fragaria vesca), and the authors used 454, Illumina and SOLiD platforms to decode the whole genome (Shulaev et al., 2011).

The NGS platforms have been employed to build reference genome sequences not only for model plants but also for a range of orphan crops such as chickpea (Varshney et al., 2013), pigeonpea (Varshney et al., 2012) and Vigna crops (Kang et al., 2014, 2015; Yang et al., 2015). Sequencing of reference genomes in different plant species has enabled access to massive genome-wide genetic markers that are indispensable tool for genomics-assisted breeding. For example, the reference genome sequences have facilitated development of high-density genotyping arrays tiled with 1K to 820K single nucleotide polymorphisms (SNPs) spread over the entire genome in important crop plants including rice, wheat, maize, barley, soybean, sorghum, groundnut, chickpea and pigeonpea (Rasheed et al., 2017). Also, mapping-by-sequencing approaches guided by the reference genome sequence greatly augment the gene discovery in plants. These mapping-by-sequencing approaches have been thoroughly reviewed elsewhere (Davey et al., 2011; Schneeberger, 2014).

The Illumina platforms based on sequencing-by-synthesis still remain the most preferred NGS system for sequencing. However, the short reads generated by the NGS platforms pose challenges in de novo genome assembly, particularly in case of complex genomes with polyploidy, heterozygosity and abundant repeat sequences (Bolger et al., 2014; Hu et al., 2018). This is evident from the fact that several of the draft genome assemblies built on NGS reads still remain incomplete and fragmented (Paajanen et al., 2019). In this context, Belser et al. (2018) discussed the varying levels of contiguity in current genome assemblies and they observed that only six plant species have genome assemblies with contig N50 greater than 5 Mb. A constant quest towards overcoming these issues has led to the development of third-generation sequencing (TGS) technologies (van Dijk et al., 2018). The most widely used TGS technologies the PacBio single molecule real-time (SMRT) sequencing and Oxford Nanopore (MinION/ PromethION) generate read length up to 100 kb and 1 Mb, respectively, with an average of 10–15 kb as against the usual average Illumina read length of 125–300 bp (Hu et al., 2018).

The long-read sequencers in combination with optical maps (Schwartz et al., 1993) are being used to generate high-quality chromosome level genome assemblies (Jiao et al., 2017; Paajanen et al., 2019; Tang et al., 2015). Recently, PacBio RS II system was applied for construction of 2.5 Gb genome assembly of peanut (Arachis hypogaea, an allotetraploid) with a contig N50 of 1.5 Mb (Zhuang et al., 2019). The long-range scaffolding techniques such as high-throughput chromosome conformation capture (Hi-C) facilitate chromosome-scale assembly of the contigs. In this respect, recently built genome assemblies of Brassica rapa (529 Mb), B. oleracea (630 Mb) and Musa schizocarpa (587 Mb) showed up to 450-fold improvement in contiguity over the existing assemblies (Belser et al., 2018). Similarly, relative to a new maize genome assembly (PH 207) based on illumina short read, improved genome sequence of the maize inbred line B73 generated using PacBio RS II system with contig N50 of 1.2 Mb offers a 240-fold improvement in contig length (Jiao et al., 2017). The remarkable improvement in contiguity was achieved in a more recent 2.16-Gb genome assembly of small-kernel (SK) maize line based on the long-read PacBio system, which has a contig N50 of 15.78 Mb (Yang et al., 2019). The assembly has 238 gaps as compared to 2,522 of improved B73 assembly. Belser et al. (2018) discuss that a combination of Oxford Nanopore, Bionano Genomics, and Illumina could generate a sequence of 500–600 Mb for around US$ 6,000. The cost involved here is remarkably smaller than the 120 Mb genome assembly of Arabidopsis, which was generated at an approximate cost of $100 million over a period of 10 years (Goff et al., 2014).

Stimulated by the technological innovations, researchers are undertaking ambitious projects that intend to offer deeper insights into the genomic architectures and evolution (Liu et al., 2019). For example, the 3,000 Rice Genomes Project (Wang et al., 2018), 1000 plants project (1 KP, Matasci et al., 2014; https://sites.google.com/a/ualberta.ca/onekp/) , 3000 chickpea genome sequencing initiative (unpublished) etc. Notable in this context is recently proposed 10 000 plant genomes sequencing project (10 KP) with the aim to deliver more than 10 000 genome sequences across plants and eukaryotic microbes (https://db.cngb.org/10kp/). 10KP is a key component of EarthBioGenome project (https://www.earthbiogenome.org/) with the aim to generate sequence data for 1.5 million known eukaryotic species over a 10-year period.

Sequencing multiple genomes to leverage pangenomics

Genetic diversity acts as raw material for crop improvement programmes. According to Mascher et al. (2019), exploitation of genetic variation from landraces in crop breeding programmes has met with modest success, with dwarfing genes in rice and wheat and mlo alleles in barley being the notable cases. The narrow genetic variation of current crop breeding programmes is because of domestication and modern breeding. In recent years, genome-scale investigations of wide germplasm panels have served as a great resource to study genomic variation dynamics during domestication and selective breeding (Zhou et al., 2015).

For instance, recent sequencing of multiple accessions in various crop species in concert with genome-wide association study (GWAS) has facilitated identification of key genomic regions
associated with crop domestication and selection/improvement (Varshney et al., 2017).

Availability of the reference genome sequence has stimulated sequencing of multiple accessions of a plant species to enable genome-scale investigations. For instance, Morrelli et al. (2012) highlight the importance of comparative genome analyses with the proposition that ‘the future of crop improvement will be centred on comparisons of individual plant genomes’. Sequencing of multiple genomes opens new avenues for pan-genomic studies that aim to identify core and indispensable genes in crop species. Also, panogenomics has great potential in identifying larger structural variations (SVs) particularly copy number variation (CNV) and presence/absence variation (PAV) that significantly contribute towards phenotypic diversity. Identification of such SVs otherwise remains difficult through analysis of a single reference genome or reference-based resequencing studies (Tao et al., 2019). Sequencing of 292 pigeonpea accessions highlighted the role of evolutionary transitions in shaping structural variation and the association of SVs with the genome regions affected by domestication and modern breeding (Varshney et al., 2017). Concerning the identification of the large SVs at chromosomal scale, modern systems based on optical mapping technology such as the Bionano Genomics Saphyr system have remarkable sensitivity towards detection of genome-wide SVs (https://bionanogenomics.com/support-page/saphyr-system/).

More recently, we have proposed a concept of super-pangenome to capture a complete view of genetic diversity present in a genus. In this approach, first different species-level pangenomes are constructed and then these pangenomes are combined to obtain a ‘pangenome of pangenomes’ or a genus-level pangenome. For developing a species-level pangenome, the most diverse accessions of a species are identified and selected. Then, the genome of one of these accessions is sequenced and assembled de novo, which serves as a reference for the mapping of resequencing data from the remaining accessions. The super-pangenome thus constructed offers better insights into the indispensable genome set and hence has a greater utility for crop improvement (Khán et al., 2020).

Genomic technologies facilitate efficient characterization and utilization of germplasm stored in global repositories. Creation of subsets of germplasm collections such as core and mini core has been proposed to bring the number of germplasm accessions to manageable level (10% and 1% of the total accessions in core and mini core, respectively) while encompassing high diversity of a species (Upadhyaya and Ortiz, 2001). In the context, DNA marker data were also used for the development of mini core collections in different crops including rice, maize, soybean, peanut, chickpea and pigeonpea (see Guo et al., 2014). Cost-effectiveness of recent high-throughput genotyping technologies has inspired researchers to perform genome-wide characterization of global germplasm collections instead of relying on limited subset of collections such as core or mini core. Large-scale characterization of germplasm collections was carried out in a variety of crops including soybean (14 430 accessions typed with 52 041 SNPs; Bandillo et al., 2015) and maize (2815 accessions typed with 681 257 SNPs; Romay et al., 2013). A more recent study based on genotyping-by-sequencing (GBS) of 22 626 barley accessions from ex situ genebank presents opportunities not only for the discovery of novel beneficial genes but also to take informed decisions for germplasm management (Milner et al., 2019). In this context, Mascher et al. (2019) recommend to transform genebanks into ‘biodigital resource centres’ which would be instrumental in linking genomic information with the plant performance of each stored accession. Creation of biodigital resource centres will greatly help researchers to make informed choices for pre-breeding programmes that lead to product delivery. Furthermore, for crop improvement applications, we propose to develop crop diversity panels (CDPs) based on germplasm sequencing data. These CDPs can be evaluated and used for mining the haplotypes for the genes for different target traits. Germplasm lines carrying superior haplotypes can be used in breeding programmes for transferring these unexplored haplotypes and broadening genetic base of elite gene pool.

**Trait discovery in the post-NGS era**

High-throughput methods for rapid gene/QTL discovery

Conventional quantitative trait loci (QTL) mapping methods suffer from limited genetic resolution besides having low throughput, being labour-intensive and time-consuming in nature. Presence of whole genome sequence in concert with advances in DNA sequencing technologies and computational biology has greatly empowered trait analysis and gene discovery in plants (Jagannathan et al., 2020). Last decade has seen emergence of a series of such trait mapping approaches such as SHOREmap, SNP ratio mapping (SRM), next-generation mapping (NGM), MutMap and QTL-seq that harness the immense potential of reference genome sequences (Bohra, 2013; Varshney et al., 2014; Zhang et al., 2019). As a result, candidate QTL regions can be resolved now to a level of few kbs through either sequencing genomes of all individuals of the mapping populations or integrating bulked segregants analysis (BSA) with whole genome resequencing (WGRS). For example, Huang et al. (2016) sequenced genomes of more than 10 000 F2 individuals from 17 representative hybrid rice crosses and mapped QTLs mostly within 300 kb. The study provided important insights into genomic architecture of heterosis as well as occurrence of partial dominance and overdominance at the loci contributing to heterotic advantage. Similar examples include mapping of plant height QTL and GWS gene to a 100-kb (Huang et al., 2009b) and 200-kb region (Xie et al., 2010), respectively, in rice and QTL controlling resistance against southern root-knot nematode within a bin size of 29.7 kb region in soybean (Xu et al., 2013) following resequencing of 150, 238 and 246 RILs, respectively. In a biparental population, the mapping resolution of the QTL region achieved by the WGRS was 16.7–144.5 times higher as compared to the conventional QTL mapping using SNP and SSR markers (Xu et al., 2013). These studies have resolved candidate genomic regions to a level that is comparable to sequence-based GWAS of diverse genotypes. For example, GWAS of 302 sequenced genotypes in soybean could narrow down a known QTL region (12-Mb) for pod dehiscence to a 190-kb region harbouring 14 genes (Zhou et al., 2015).

Alternative approaches based on sequencing of selected/bulked individuals such as QTL-seq have been widely applied for trait mapping across different crop species owing to its inherent ability to address both qualitative and quantitative traits (Table 1). To this end, Zhang et al. (2019) have proposed a new strategy called as quantitative trait gene sequencing (QTG-seq) to improve genetic resolution achieved by the QTL-seq. In the QTG-seq, target QTL selection in the first generation of backcross (BC1F2) is accompanied by sequencing of selected BC1F2 pools at relatively high coverage. This allows a quantitative trait to be analysed in a ‘near qualitative’ fashion. Using this strategy, these researchers located a plant height QTL of maize (qPH7) to a 150-kb genomic...
interval harbouring a causal gene that codes for an NF-YC transcription factor.

Harnessing high-power mapping resources

With high-throughput genotyping systems coming within grasp of even small-scale laboratories, the type of the genetic material being employed for trait mapping studies assumes greater significance (Stadlermeier et al., 2018). Biparental QTL mapping has seen tremendous success in understanding the genetic architecture of various important traits in different crop species (Bohra et al., 2014a, b). Subsequently, association genetics of diverse panels was proposed to overcome the inherent caveats of biparental analysis such as low mapping resolution, limited allelic diversity and need of artificially created populations. As illustrated in Fig 1, a greater need to resolve the complex genetic architecture of traits has caused a methodological shift towards broad-based mapping resources that accommodate diverse founders and abundant recombination events while retaining benefits of linkage-based designs (Chen et al., 2019a). These designs involving multi-parents impart rich allelic content, higher genetic resolution, large phenotypic diversity and better estimation of allelic effects (Scott et al., 2020). Two such designs, that is nested association mapping (NAM) and multi-parent advanced generation intercross ( MAGIC), have been adopted in various crops for trait mapping (Table 2). Even a simplified MAGIC panel with modest population size (394 RILs) is shown to capture nearly 70% of the diversity of German wheat breeding gene pool (Stadlermeier et al., 2018). Similarly, sorghum NAM design with 2214 RILs had captured ~70% of global diversity and shown three times more power than the association panel of the same 2214 RILs had captured nearly 70% of the diversity of German wheat breeding gene pool (Stadlermeier et al., 2018). Similarly, sorghum NAM design with 2214 RILs had captured ~70% of global diversity and shown three times more power than the association panel of the same size to detect QTL for adaptive traits (Bouchet et al., 2017).

Nested association mapping comprises a series of connected half-sib families derived from crossing diverse parents with a common reference parent. MAGIC encompasses cycles of structured intermaturing among founders and advancement, yielding mosaics of genome blocks from all founders (Huynh et al., 2018). The highly recombined nature of these populations has been strongly supported from linkage disequilibrium (LD) patterns inferred from high-density genotyping (Mackay et al., 2014; Ongom and Ejeta, 2018; Scott et al., 2020). Inherent to the nature of the mating scheme, recovery of novel QTL combinations is limited in NAM because of the biparental derivation of the constituent RILs. Huang et al. (2015) proposed to combine MAGIC with recurrent selection where marker–trait associations (MTAs) are identified and then deployed in the same MAGIC panel to select lines with greater number of positive lines only to be recombined for 2–3 cycles, leading to the development of lines carrying maximum number of positive alleles.

High-resolution genome-wide association studies

Genome-wide markers such as SNPs/CNVs unleashed from WGRS efforts have greatly empowered GWAS for delineating the smallest possible genome region associated with phenotypic variation in large germplasm sets. Recent instances of WGRS-based GWAS are worth mentioning in rice (Huang et al., 2010; Yano et al., 2016), foxtail millet (Jia et al., 2013), soybean (Zhou et al., 2015), sesame (Wei et al., 2015), chickpea (Varshney et al., 2019b), pigeonpea (Varshney et al., 2017) and cotton (Ma et al., 2018) for discovering highly resolved MTAs related to traits of economic importance including plant domestication traits (Table 3).

A GWA study of more than 500 sequenced landraces in rice elucidated a total 80 MTAs for 14 different traits related to grain yield and quality, physiology and drought stress (Huang et al., 2010). Similarly, GWAS based on WGRS of 176 rice accessions uncovered four genes viz., LOC_Os01g15370 and LOC_Os06g15420 (MTAs) are identified and then deployed in the same MAGIC panel to select lines with greater number of positive lines only to be recombined for 2–3 cycles, leading to the development of lines carrying maximum number of positive alleles.

### Table 1 A list of some key NGS-based trait discovery studies in some crops

| Crop       | Population | Trait analysed                                    | QTL/Gene mapped                                                                                     | References                                      |
|------------|------------|--------------------------------------------------|------------------------------------------------------------------------------------------------------|------------------------------------------------|
| Rice       | NIL-13B4 × GH998 (F2) | Nitrogen use efficiency | 266.5-kb qNUE6 (LOC_Os06g15370 and LOC_Os06g15420)                                                | Yang et al. (2017)                              |
| Soybean    | Zhonghuang × Jiyu 102 (F2) | Seed cotyledon colour | qCC1 (30.7-kb) and qCC2 (67.7-kb)                                                                | Song et al. (2017)                              |
|            | Jikedou 2 × Huachun 18 (F2) | Phytophthora resistance | 146-kb RpsHC18 (RpsHC18-NBL1 and RpsHC18-NBL2)                                                   | Zhong et al. (2018)                             |
| Brassica napus | Huyou19 × Purler(F2) | Branch angle | branch angle 1 (BnaA0639380D, a homolog of AYUCCA6)                                               | Wang et al. (2016)                              |
| Peanut     | ZH8 × ZH9 (RIL) | Testa colour | AhtC1, encoding a MYB transcription factor                                                        | Zhao et al. (2019)                              |
|            | TAG 24 × GBPD 4 (RIL) | Rust and late leaf spot resistance | qRust80D_06, qRust90D_06, qRust 80D_07, qRust 90D_07, qRust 90D_08, qRust 80D_09, qRust 90D_09, qLLS50D_08, qLLS 90D_08, qLLS 90D_09 | Pandey et al. (2017)                             |
|            | ICGV 00350 × ICGV 97045 (RIL) | Fresh seed dormancy | RING-H2 finger protein and zeaxanthin epoxidase                                                     | Kumar et al. (2019)                             |
| Chickpea   | Yuanza 9102 × Xuzhou 68-4 (RIL) | Shelling percentage | 10 SNPs in nine candidate genes                                                                  | Luo et al. (2019)                               |
|            | ICC 4958 × ICC 1882 (RIL) | 100-seed weight | Ca_0436 and Ca_04607                                                                             | Singh et al. (2016a)                            |
|            | ICCV 96029 × CDC Frontier (RIL) | Ascochyta blight | Six candidate genes                                                                             | Deokar et al. (2019)                            |
|            | ICCV 96029 × Amit (RIL) | Fusarium wilt and sterility mosaic disease resistance | C. cajan_03203 and C. cajan_01839                                                                 | Singh et al. (2016b)                            |
| Pigeonpea  | ICPL 20096 × ICPL 332 (RIL) | Fusarium wilt and sterility mosaic disease resistance | C. cajan_03203 and C. cajan_01839                                                                 | Singh et al. (2016b)                            |
lines allowed fine-dissection of fibre-related traits and the flowering time trait (Ma et al., 2018).

In legumes, WGRS-based GWAS has been successfully applied for delineating new QTLs/candidate gene(s)/MTAs along with validating the loci identified previously through QTL mapping or association studies. In soybean, GWAS on 302 sequenced genotypes identified several new MTAs that remain congruent with the previously identified QTLs controlling a range of domestication-related traits (Zhou et al., 2015). Another GWA study in soybean with WGRS-SNPs on 106 lines revealed 401 and 328 SNPs significantly associated leaf scorch score (LSS) and leaf chlorophyll content, respectively, contributing to salinity tolerance (Patil et al., 2016). Interestingly, the most significant SNP related to LSS was pinpointed in *GmCHX1* gene, which explained 63% variation of the phenotype (Patil et al., 2016). Likewise, GWAS of 234 lines elucidated genomic architecture of salinity tolerance in soybean with significant MTAs for leaf scorch score, chlorophyll content ratio, leaf sodium content and leaf chloride content (Do et al., 2019).

A recent GWA study of a 429-line global reference set of chickpea elucidated important candidate genes underlying 262 MTAs controlling various traits that confer heat and drought stress tolerance (Varshney et al., 2019b). In legumes, other high-resolution trait mapping studies combining GWAS and WGRS were performed for drought stress in 132 lines of chickpea (Li et al., 2018), yield/seed traits and anthracnose resistance in 683 lines of common bean (Wu et al., 2020) and adaptive traits in 292 pigeonpea accessions (Varshney et al., 2017). The GWAS has been greatly benefited by the enhanced marker density of WGRS, and however, the mapping resolution of GWAS depends on the extent of LD and recombination rate, which vary in different plant species (self-pollinated or cross-pollinated), and among different populations (wild, landraces and improved cultivars) and within the genome (euchromatin and heterochromatin regions) of a given species (Chang et al., 2018; Zhou et al., 2015).

**Genomics-informed gene editing**

Gene editing technologies include a number of powerful tools to directly change genetic sequences in coding and/or regulatory regions to create new alleles, most effectively without introducing new transgenes (Zhang et al., 2018). The most frequently applied techniques include CRISPR (Clustered Regularly Interspaced Palindromic Repeats) or TALEN (Transcription Activator Like Effector Nucleases), with the CRISPR/Cas system being the simplest gene editing system to apply. The basic techniques and applications for crop gene editing have been well-described elsewhere (Chen et al., 2019b; Schindele et al., 2020; Zhang et al., 2018; Zhang et al., 2019). As a complementary tool to genomics, gene editing can resolve questions as to gene identity and function, as well as provide novel allelic variants not available within the crop species or interfertile relatives in the domesticated primary or wild secondary gene pools.

Gene editing techniques can be used to knockout genes, usually by inducing small insertions or deletions, which lead to frameshift mutations causing premature stop codons. The most frequent approach relies on non-homologous end joining (NHEJ) edits. This type of editing targets a region of the coding regions or sometimes a regulatory sequence. CRISPR/Cas9 induces a double-strand break in the DNA and relies on the cell’s endogenous DNA repair mechanisms to religate the broken strands. While majority of DNA repair mechanisms are intrinsically accurate, it is error prone, and it is these errors that make new alleles. As already stated, these frequently result in new non-functional alleles. However, at a low frequency they may simply cause a single non-synonymous amino acid change. These repair mechanisms can also delete nucleotide in multiple of three, which will lead to

---

*Figure 1* Adoption of new-generation genetic resources for enhanced trait discovery. The power and precision of trait discovery have improved several folds with increasing adoption of multi-parent populations and association panel. Importantly, mapping populations derived from multiple founders retain benefits of both linkage analysis and association mapping. CC: Collaborative cross; GWAS: genome-wide association study; LD: linkage disequilibrium; MAGIC: multi-parent advanced generation intercross; NAM: nested association mapping; RIL: recombinant inbred line. [Colour figure can be viewed at wileyonlinelibrary.com]
deletions of amino acids or even peptides and these may alter the
gene expression without actually creating a null allele. In a
number of jurisdictions, including the United States, Australia,
Japan and Argentina, these types of changes give rise to plants
that are not regarded as transgenic and hence can be rapidly used
in breeding programmes.

It is also possible to apply gene editing such that a specific
repair template is used to change the coding or regulatory

| Crop | Founders involved | Size | Markers assayed | Trait mapped | Approach used | References |
|------|-------------------|------|-----------------|--------------|--------------|------------|
| Magic Cowpea | 8 | 305 | 51, 128 SNPs | Flowering time, growth habit, seed size, maturity, photoperiod | Interval QTL mapping | Huynh et al. (2018) |

**Table 2 Details of multi-parent populations and their applications in trait mapping in some crops**

| Crop | Founders involved | Size | Markers assayed | Trait mapped | Approach used | References |
|------|-------------------|------|-----------------|--------------|--------------|------------|
| Faba bean (Go‘ttingen Winter Bean Population, GWBP) | 11 | 400 | 156 SNPs | Morphological traits, fatty acid composition, shoot water content | Association mapping | Sallam and Martsc (2015) |

Rice

| Crop | Founders involved | Size | Markers assayed | Trait mapped | Approach used | References |
|------|-------------------|------|-----------------|--------------|--------------|------------|
| Japan-MAGIC (JAM) | 8 | 981 | 16 345 SNPs | Days to heading, culm length | GWAS | Ogawa et al. (2018) |

| Crop | Founders involved | Size | Markers assayed | Trait mapped | Approach used | References |
|------|-------------------|------|-----------------|--------------|--------------|------------|
| MAGIC, MAGIC plus, japonica MAGIC, Global MAGIC | 8, 16 | 500–1328 | 17 387 SNPs | Submergence tolerance, bacterial blight, grain quality | GWAS | Ongom and Ejeta (2018) |

| Crop | Founders involved | Size | Markers assayed | Trait mapped | Approach used | References |
|------|-------------------|------|-----------------|--------------|--------------|------------|
| Sorghum | 19 | 1000 | 79 728 SNPs | Plant height | GWAS | Khazaei et al. (2018) |

| Crop | Founders involved | Size | Markers assayed | Trait mapped | Approach used | References |
|------|-------------------|------|-----------------|--------------|--------------|------------|
| Wheat | 8 | 8 | 90 000 SNPs | Powdery mildew | Interval QTL mapping | Stadler et al. (2018) |

| Crop | Founders involved | Size | Markers assayed | Trait mapped | Approach used | References |
|------|-------------------|------|-----------------|--------------|--------------|------------|
| Maize | 8 | 1636 | 54 234 SNPs | Days to pollen shed, plant height, ear height and grain yield | Linkage mapping and association mapping | Ongom and Ejeta (2018) |

**NAM**

| Crop | Founders involved | Size | Markers assayed | Trait mapped | Approach used | References |
|------|-------------------|------|-----------------|--------------|--------------|------------|
| Maize (B73) | 26 | 5000 | 3641 SNPs | Flowering time | Joint linkage analysis and GWAS | Buckler et al. (2009) |

| Crop | Founders involved | Size | Markers assayed | Trait mapped | Approach used | References |
|------|-------------------|------|-----------------|--------------|--------------|------------|
| NAM (W22) | 6 | 1257 | 51 544 | Domestication and agronomic traits | Joint linkage analysis and GWAS | Cook et al. (2012) |

| Crop | Founders involved | Size | Markers assayed | Trait mapped | Approach used | References |
|------|-------------------|------|-----------------|--------------|--------------|------------|
| Soybean (IA3023) | 41 | 5600 | 5303 SNPs | Grain yield stability | GWAS | Xavier et al. (2018) |

| Crop | Founders involved | Size | Markers assayed | Trait mapped | Approach used | References |
|------|-------------------|------|-----------------|--------------|--------------|------------|
| Wheat (Berkut) | 29 | 2100 | 800 000 | – | – | Jordan et al. (2018) |

| Crop | Founders involved | Size | Markers assayed | Trait mapped | Approach used | References |
|------|-------------------|------|-----------------|--------------|--------------|------------|
| Wheat (Asassa) | 51 | 6280 | 13 000 SNPs | Phenology traits and plant height | GWAS | Kidane et al. (2019) |

| Crop | Founders involved | Size | Markers assayed | Trait mapped | Approach used | References |
|------|-------------------|------|-----------------|--------------|--------------|------------|
| Rice (IR64) | 11 | 1879 | 7152 SNPs | Northern leaf blight | – | Fragoso et al. (2017) |

| Crop | Founders involved | Size | Markers assayed | Trait mapped | Approach used | References |
|------|-------------------|------|-----------------|--------------|--------------|------------|
| Sorghum (RTx430) | 11 | 2214 | 90 000 SNPs | Flowering time and plant height | – | Bouchet et al. (2017) |

| Crop | Founders involved | Size | Markers assayed | Trait mapped | Approach used | References |
|------|-------------------|------|-----------------|--------------|--------------|------------|
| Barley (Barke) | 26 | 1420 | 27 000 SNPs | Glossy spike, glossy sheath and black hull colour | – | Nice et al. (2016) |

| Crop | Founders involved | Size | Markers assayed | Trait mapped | Approach used | References |
|------|-------------------|------|-----------------|--------------|--------------|------------|
| Pea (Cameor) | 8 | 927 | 13 204 SNPs | Yield-related traits | – | Sharma et al. (2018) |

| Crop | Founders involved | Size | Markers assayed | Trait mapped | Approach used | References |
|------|-------------------|------|-----------------|--------------|--------------|------------|
| Groundnut (NAM-Tfrunner & NAM-Florida-07) | 5, 5 | 581 496 | 58 000 SNPs | 100-pod weight and 100-seed weight | Linkage mapping and association mapping | Gangurde et al. (2019) |

*Common parent of the NAM population is shown in parentheses.*
| Crop         | Trait                                                                 | MTA/QTL/candidate genes identified | SNP/indels | Number of genotypes | LGs/Chromosomes | References          |
|--------------|----------------------------------------------------------------------|-----------------------------------|------------|---------------------|-----------------|---------------------|
| Rice         | Agronomic traits                                                    | 80 significant MTAs               | 3.6 million SNPs | 517                | 1–12            | Huang et al. (2010) |
| Rice         | Amylose content and seed length, pericarp colour                    | Os03g0407400, Os06g0133000, Os10g0536400 | 2.3 million SNPs | 203                | 6, 7, 10        | Wang et al. (2016)  |
| Rice         | Days to heading date, awn length, panicles per plant, plant height, panicle length, spikelet number per panicle, leaf blade width | LOC_Os01g62780, LOC_Os11g08410, LOC_Os04g52479, LOC_Os08g37890 | 426 337 SNPs, 67 544 indels | 176                | 1, 4, 6, 8, 11     | Yano et al. (2016)  |
| Rice         | Grain shape, length and width                                       | GWi7.1, GWi7.2, GL2.1, GWi5.1, GW11.1 | 2.9 million SNPs, 3.9 million indels | 591                | 3, 5, 7, 10, 11   | Misra et al. (2017) |
| Rice         | Alkalinity                                                          | Eight QTLs                        | 788 396 SNPs | 295                | 3               | Li et al. (2019)    |
| Rice         | 17 mineral elements                                                 | 72 loci                           | 6.4 million SNPs | 529                | 1-12            | Yang et al. (2018)  |
| Rice         | Heading date, grain mass, straw biomass, harvest index              | 115 QTLs                          | ~2 million SNPs | 266                | 1-12            | Norton et al. (2018) |
| Rice         | Pericarp colour, amylose content, protein content, panicle number   | 2 046 529 SNPs                    | 137        | 2, 3, 5, 6, 7, 9   |                | Kim et al. (2016)   |
| Rice         | Grain width, grain length                                           | MTAs coincided with GS3, GW5, and qGL7, OsFD | 223 743 SNPs | 3,010               | 1, 3, 5, 6, 7, 9, 11 | Wang et al. (2018) |
| Rice         | Bacterial blight                                                    | Xa26                              | 148 999 SNPs | 365                | 7, 11           | Fuentes et al. (2019) |
| Rice         | Seed coat colour, grain length                                      | Rc locus, LONG KERNEL 3 gene       | 889 903 SNPs | 302                |                | Zhou et al. (2015)  |
| Soybean      | Oil content, plant height, domestication traits, pubescence form, flower colour, cyst nematode tolerance, seed weight | –                                 | 9 790 744 SNPs | 876 799 indels     |                | Patil et al. (2016) |
| Soybean      | Salinity tolerance                                                  | 401 and 328 MTAs for leaf scorch score and leaf chlorophyll content, respectively | 5 million SNPs | 106                | –               | Patil et al. (2016) |
| Soybean      | 84 traits                                                           | 245 loci                          | 4 million SNPs | 809                | 1–20           | Fang et al. (2017)  |
| Soybean      | Salinity tolerance                                                  | 51 significant MTAs               | 3.7 million SNPs | 234                | 1, 2, 3, 5, 6, 8, 14,15,16, 18, 19, 20 | Do et al. (2019)   |
| Chickpea     | Ascochyta blight                                                    | AB4.1 and 12 candidate genes and 20 significant SNPs | 144 000 SNPs | 132                | 4               | Li et al. (2017)    |
| Chickpea     | Yield-related traits under drought stress                           | 38 significant SNPs               | 144 777 SNPs | 132                | 3, 4, 5, 6      | Li et al. (2018)    |
| Chickpea     | Traits related to drought and heat stress                            | 262 MTAs and several candidate genes including TIC, REF6, aspartic protease, cc-NBS-LRR, RGA3, Ca_13671, Ca_13939 | 3.65 million SNPs | 429                | –               | Varshney et al. (2019) |
| Pigeonpea    | Agronomic traits                                                    | 241 MTAs, homologs of LIGULELESS1, SHATTERING1 and EARLY FLOWERING3 (ELF3) | 15.1 million SNPs, 2.1 million indels | 292                | 1–11            | Varshney et al. (2017) |
sequence in a specific manner. These homology directed repair (HDR) edits can be extremely powerful tools to edit genes and create novel variants. New or different amino acids may be introduced, from a single amino acid up to hundreds of amino acids, depending on the templates used. However, there are considerable restrictions to their usage, and in many jurisdictions, the introduction of any new DNA to the host is sufficient for these to be classified as transgenic. In others, such as in the United States and Japan, they may be considered as non-transgenic on a case-by-case basis.

Biallelic editing using CRISPR/Cas9 and Mendelian inheritance of these edits was first reported in Arabidopsis and crop plants including rice (Zhang et al., 2014) and tomato (Brooks et al., 2014). This paved the way for gene editing to be broadly applied across species. It was soon demonstrated that multiple genes and gene combinations could be edited simultaneously. Wang et al. (2014) demonstrated that simultaneous editing of 3 homeoealleles in hexaploid wheat could be performed to develop powdery mildew resistance in wheat. Indeed, the wild relatives of crops can also be edited to increase their utility as either new crops or sources of novel genetic variation. Solanum pimpinellifolium, a wild relative of cultivated tomato, was edited at six independent loci to produce plants more closely resembling the domesticated S. lycopersici for key fruit traits. These gene-edited plants produced more flowers and fruits, with larger fruits, fewer seeds and higher lycopersicin content in the fruits than the wild species (Zsögön et al., 2018). It quickly became evident that producing gene-edited plants became more straightforward than detecting edited plants, particularly when the altered phenotype was not evident visually. Various groups have developed rapid phenotyping tests to more efficiently screen plants for the most desirable edit(s) (Peng et al., 2018).

Used in concert with genomics techniques, gene editing is a particularly elegant tool for gene discovery. Indeed, many gene-edited crop plants have been produced based on either gene identification in other species, quite often in model species. Where genomic approaches have also been used for gene discovery purposes, it can be a laborious process to increase the recombination events around the desired haplotype. It is not infrequent that a region associated with a trait or QTL may be in the order of 100-500 genes, dependent on the LD in a species/population. Hence, the ability to identify the true causative gene among many potential candidates can be time-consuming. The use of classical transgenics has been useful and informative, yet imprecise because of variables such as position effect and gene dosage where the transgene inserts into the host genome.

The availability of gene editing techniques offers considerable advantages in identifying candidate genes and genetic interactions to elucidate gene action in the understanding of QTL regions. The edit(s) can be made in the actual gene, and hence, there are no position or dosage effects. Gene expression can be totally knocked out, which has previously been difficult using RNAi approaches, which usually lead to a diminution of gene expression but rarely to zero (Eamens et al., 2008). This also means that editing of candidate genes enables clear identification of single gene action. As another advantage, multiple candidate genes can be targeted in a single experiment. For example, three genes, A, B and C, can be edited and the independent progenies will include lines with the individual genes edited and all possible combinations (A + B, B + C, A + C, A + B + C) provided sufficient lines are produced. This can be extremely effective to identify candidate genes in a linkage block, to elucidate specific interactions in a multigene pathway, to uncover evidence of epistasis and to determine instances of pleiotropy and close linkage.

A current limitation of the power of gene editing is the reliance on tissue culture techniques for editing to be performed in most crop species. As a result, gene editing can be extremely genotype limited. The development of tools and broadly applicable means to edit genes without the need of in vitro plant regeneration will enable the application of gene editing significantly more efficiently and rapidly. A number of techniques are currently being used to overcome the genotypic bottleneck of gene editing, as reviewed in Hickey et al. (2019).

### Breeding strategies to deliver higher genetic gains

Genetic gains from a selection programme can be expressed in the form of breeder’s equation, that is \( \Delta G = R \times h^2 \times S = a_i \times i \times r/L \). Following the equation, the gain \( \Delta G \) or response to selection \( R \) can be improved by tweaking additive genetic variation \( a_i \) or narrow sense heritability \( h^2 \), selection intensity \( i \) and selection accuracy \( r \) and length of the breeding cycle \( L \). In the following section, we discuss the new breeding methodologies that address different components of the breeder’s equation and improve the rate of genetic gain in a breeding programme.

### Genomic selection

The paradigm ‘genotype once phenotype many times’ has dominated genetic studies for the past two decades owing to
the high cost of genotyping. The increased availability of homozygous immortal genetic populations allowing replications across time and locations further fuelled this paradigm (Srivastava et al., 2017). With the development of NGS, genome-wide marker assays are now affordable, accurate and high throughput. However, acquisition of accurate and precise phenotyping data on sizeable individuals presents a major bottleneck in plant breeding programmes. This has stimulated adoption of new breeding techniques that optimize phenotyping requirements for improving complex traits controlled by a number of small-effect QTL (Akdemir and Isidro-Sanchez, 2019).

Genomic selection (GS) improves genetic gain by enhancing selection intensity (i) and selection accuracy (r) and reducing the breeding cycle length (L). GS predicts genetic merit of unobserved phenotypes from target population based on the breeding values (GEBVs) computed from genome-wide information of a training set that has been scored phenotypically.

Since the concept was originally proposed by Meuwissen et al. (2001), GS implementation has seen tremendous success in animal breeding, and some of the GS studies show 50%–100% increase in genetic gain per year for yield traits in dairy cattle and 35% increase in pig breeding programme (Edwards et al., 2019). The key factors underlying success of GS in animal breeding are greater economic returns from early selections and reduced generation intervals, weaker genotype–environment interactions (G × E) and easily controllable environments, higher individual value, large training populations with stronger genetic relatedness between training and breeding individuals, access to both cost-efficient genotyping systems and historical phenotypic records, greater significance of additive genetic effects and the straightforward incorporation of existing best linear unbiased predictor (BLUP)-based approaches into the prediction models (Jonas and de Koning, 2013; Santantonio et al., 2020; Xu et al., 2020). In plants, recent simulation and empirical evidence has established superiority of GS over traditional selection methods including phenotypic, pedigree and marker-assisted selections (Crossa et al., 2017). For long-term selection gains in hybrid breeding, genome-wide predictions have been used for identification of heterotic groups and establishment of heterotic patterns in various crops including wheat (Zhao et al., 2015), rice (Beukert et al., 2017) and pigeonpea (R. K. Saxena, et al., Unpublished data).

A variety of factors are known to influence GS prediction accuracy, that is the degree to which GEBVs relate to estimated genetic values (Akdemir and Isidro-Sanchez, 2019), which include training population size, relatedness between training and test individuals, DNA marker type and density, trait architectures and heritability, statistical models and population structure (Roorklaw et al., 2018; Thorwarth et al., 2017; Xu et al., 2020; Zhang et al., 2017). Optimization of these factors has shown improvements in GS prediction accuracies.

Studies suggest that using multi-environmental settings and incorporating GxE interactions into GS models improve prediction accuracies (Jarquin et al., 2014; Roorklaw et al., 2018; Sukumar et al., 2017). Though GS unlike MAS does not need a set of DNA markers associated with the trait, incorporating information about the significant markers is shown to improve prediction accuracies (Spindel et al., 2016). In a recent GS study in chickpea, Li et al. (2018) obtained twofold improvement in prediction accuracy with a subset of SNPs informed by GWA as compared to using all WGRS-SNPs. Of the various models used to predict the genetic worth of unobserved individuals, GBLUP remains the most extensively used (Table 4). Further improvement in prediction accuracy is expected with advances in high-throughput phenotyping such as hyperspectral imaging (Crossa et al., 2017). However, application of deep machine-learning methods for genome-wide prediction awaits further research.

Since the public breeding programmes in developing countries are severely constrained by the lack of resources and appropriate technical skills, Santantonio et al. (2020) recommend a phased GS implementation in order to adopt GS as a routine strategy for crop breeding. The initial phase involves informatics development and genotyping of lines that are the most relevant to breeding programmes such as the lines entering in the variety release system. In the second phase, GS is applied to enhance selection intensity in varietal development programmes, while the final phase focuses on rapid cycle recurrent selection. Such optimized approaches that allow the efficient use of resources and technical expertise will be crucial for large-scale implementation of GS in breeding programmes of public sectors.

**Rapid generation turnover**

Traditional plant breeding methods have delivered a series of high-yielding crop cultivars suited to diverse agro-climatic conditions worldwide. However, reliance of these traditional methods on repeated cycles of crossing and inbreeding requires 10–15 years for developing and releasing a new crop cultivar. The lengthy crop breeding cycles have been described as a ‘high entry barrier’ in accelerating crop research with modern tools and technologies (Watson et al., 2018).

As mentioned in the previous section, manipulating parameters of breeder’s equation could improve rate of genetic gain. However, approaches that could shorten the length of breeding cycle are considered to substantially influence ΔG in comparison with manipulating other parameters of the equation (Cobb et al., 2019; Li et al., 2018). The protocols collectively grouped under ‘speed breeding’ (SB) aim to accelerate plant development and shorten breeding cycle time via optimizing *in vivo* growth conditions such as light, photoperiod, temperature, humidity in combination with enhanced plant density and early seed harvesting (Ghosh et al., 2018). To reduce time to anthesis, application of *in vitro* protocols is recommended for germination of immature seeds (Crosser et al., 2016). Optimized SB recipes have proven to be effective in different crops including wheat (Watson et al., 2018), barley (Hickey et al., 2017), chickpea (Samineni et al., 2020) and pea (Mobini and Warkentin, 2016) for obtaining multiple generations in a single year. The technology has great potential to accelerate breeding programmes for rapid delivery of crop cultivars. However, the SB protocols do not represent a ‘one size fits all’ system and need to be tailored according to both crop behaviour and resources at hand. Also, further experimentation is needed to extend these protocols to short-day plants such as rice, maize (Watson et al., 2018). In the context, preliminary results in pigeonpea, a short-day plant, show the possibility to achieve four generations per year using immature seed harvest, single pod descent and controlled light/humidity conditions (Saxena et al., 2017, 2019). Exhaustive survey of the photoperiod response of different genotypes sets an essential prerequisite for adoption of SB protocols in crop research and breeding. Also, genotype independence of these protocols still remains to be established, which will in turn confirm the broader applicability of this technique across diverse crops and crop genotypes.
| Crop       | Training population                                   | Markers used | Traits analysed                                                                 | Predictive ability/Prediction accuracy | Model                          | References                       |
|------------|-------------------------------------------------------|--------------|--------------------------------------------------------------------------------|---------------------------------------|---------------------------------|----------------------------------|
| Wheat      | 1100 PYT lines (F₃ₑ)                                  | 27 000 SNPs  | Grain yield                                                                     | 0.17–0.28                             | GBLUP                           | Belamkar et al. (2018)           |
|            | 10 375 lines                                          | 18 101 SNPs  | Grain yield, relative maturity, glaucousness and thousand-kernel weight          | 0.59–0.98                             | Maximal model (GBLUP)           | Norman et al. (2018)            |
|            | 330 lines from HarvestPlus Association Mapping (HPAM) panel | 24 497 SNPs  | Grain zinc and iron concentrations, thousand-kernel weight and days to maturity  | 0.324–0.76                            | GBLUP using the reaction norm model | Velu et al. (2016)              |
|            | 208 lines                                             | 6211 DA/seq-SNPs | Grain yield, thousand-grain weight, grain number, days to anthesis, days to maturity, plant height and normalized difference vegetation index at vegetative and grain filling | 0.34–0.68                             | GBLUP                           | Sukumaran et al. (2018)         |
|            | 287 advanced elite lines (WAMI panel)                | 15 000 SNPs  | Grain yield, thousand-grain weight, grain number, thermal time for flowering     | 0.27–0.63                             | GBLUP                           | Sukumaran et al. (2017)         |
|            | 1378 breeding lines                                   | –            | Grain yield and yield stability                                                  | up to 0.54                            | Reaction norm models            | Jarquín et al. (2017)           |
|            | 2992 F2:4 lines                                       | 25 000 SNPs  | Grain yield                                                                     | 0.125–0.127                           | GBLUP                           | Edwards et al. (2019)          |
|            | 2325 inbred lines                                     | 12 642 SNPs  | Fusarium head blight, Septoria tritici blotch                                   | up to 0.6                             | RR-BLUP, Bayes Cr, RKHS, EG-BLUP | Mirdita et al. (2015)          |
| Soybean    | 301 elite breeding lines                              | 52 349 SNPs  | Grain yield                                                                     | 0.43–0.68                             | G-BLUP, G°G, Kaa, G°G, G_Kaa   | Jarquín et al. (2014)           |
| Maize      | 169 doubled haploid lines and 190 testcrosses         | 20 473 SNPs  | Grain yield, plant height, anthesis-silking interval, normalized difference vegetative index (NDVI), the green leaf area duration (GLAD) | 0.16–0.48                             | rBLUP                           | Cerrudo et al. (2018)          |
|            | 4120 lines from 22 biparental populations             | 200 SNPs     | Grain yield, anthesis date, plant height                                         | 0.18–0.38                             | Bayesian LASSO (BL) function neural network (RBFNN), reproducing kernel Hilbert space (RKHS) | Zhang et al. (2017)            |
|            | 284 inbred lines                                      | 55 000 SNPs  | Female flowering, male flowering, grain yield, anthesis-silking interval         | 0.28–0.84                             | nBLUP                           | Crossa et al. (2014)           |
| Barley     | 750 lines                                             | 11 203 SNPs  | Earing, hectolitre weight, spikes per square metre, thousand-kernel weight and yield | 0.31–0.71                             | GBLUP                           | Thorwart et al. (2017)          |
| Pea        | 315 RILs                                              | 400–500 SNPs | Grain yield                                                                     | 0.4–0.5                               | BL, nBLUP, support vector regression (SVM) | Annicihiacono et al. (2017) |
|            | 339 accessions                                         | 13 200 SNPs  | Thousand seed weight, the number of seeds per plant and the date of flowering    | up to 0.83                             | Kernel partial least squares regression (kPLSR), least absolute shrinkage and selection operator (LASSO), genomics | Tayeh et al. (2015)            |
The unique abilities of the GS and SB to shorten breeding cycle time could be harnessed synergistically to further enhance the rate of genetic gain per unit time, a strategy termed as ‘SpeedGS’ (Voss-Fels et al., 2019). Simulation study by Voss-Fels et al. (2019) compared different scenarios [phenotypic selection (PS) and GS alone and SpeedGS] and the authors observed that schemes integrating GS with SB witness 30% more genetic gain after 30 years as compared to the PS alone. However, authors suggested introgression of new diversity into the SpeedGS scheme in order to sustain the gain in long term. A simulation study in fescue also reported higher genetic gains in speedGS than that of PS (Jighly et al., 2019). Importantly, the improvement in genetic gain was higher in the case of low-heritability traits and with higher number of SB cycles. Recent empirical evidence in wheat demonstrates the potential of SpeedGS for rapid population improvement where phenotyping of SB traits in combination with multivariate GS could guide the selection of lines for field trials or next breeding cycle (Watson et al., 2019). These recent studies highlight the immense scope for ‘customizing the breeding pipelines’ (Voss-Fels et al., 2019) in order to accommodate SB and GS to achieve higher rate of genetic gains in crop breeding programmes.

**Haplotype-based breeding**

Agricultural traits are controlled by genomic loci that are ‘compound’ in nature. In other words, these loci contain several candidate genes that exert influence of varying degree and nature on the associated phenotype. Because of this, unexpected outcomes are often witnessed while transferring genomic regions through routine MAS/MABC technique. In the context, Bevan et al. (2017) have proposed a haplotype-based approach that capitalizes on the deluge of whole genome sequencing data and extensive phenotypic records in order to allow such ‘compound’ loci incorporated efficiently in breeding programmes. Here, different haplotypes for the given locus may be defined as combinations of genes and genetic polymorphisms that are inherited together.

Presence of multi-year and multi-location phenotypic data enables a genome-scale analysis of haplotypes for their phenotypic validation. As has been demonstrated in rice, a panel of sequenced lines capturing the maximum diversity is deemed suitable for phenotypic validation of haplotypes defining key traits (Abbai et al., 2019). A similar haplo-phenotype analysis in pigeonpea validated superior haplotypes of three genes for drought tolerance that were identified by mining of the WGRS data set and candidate gene-based association analysis (Sinha et al., 2020). The study also identified a set of promising lines carrying these superior haplotypes. Introgression of superior haplotypes in breeding has been referred as haplotype-based breeding (Sinha et al., 2020; Varshney et al., 2020).

Tracking sequence variation that marks the validated haplotype, in breeding programme will facilitate synthesis of an ideal line harbouring novel combinations of such established haplotypes. Retrospectively, targeted analysis of superior haplotypes across mega-varieties may help revealing combinations of superior haplotypes that explain the genetic basis of the high-performance of these lines. In pigeonpea, Sinha et al. (2020) found complete absence of superior haplotypes for drought tolerance in popular varieties Maruti (ICP 8863) and Jagriti (ICPL 151), thus offering possibilities for further improvement of such high-yielding varieties. In parallel, increasing sequencing data on
wild relatives will aid in the discovery of new haplotypes that the cultivated pool currently lacks.

Accelerating rates of varietal and seed replacements

Since high-yielding semi-dwarf varieties of wheat and rice heralded the ‘Green Revolution’ in the late 1960s, mega-varieties of major staple crops have received widespread adoption (Pingali, 2012; Singh, 2017; Singh et al., 2020). Farmers cultivate these old varieties and landraces for decades, particularly in the underdeveloped and developing countries in South Asia and sub-Saharan Africa. The average age of rice varieties in South Asia (14–25 years; Pandey et al., 2015) and sub-Saharan Africa (15.8 years; Walker et al., 2015) confirms this trend. A recent study reported cultivation of even 25-year-old wheat varieties in major wheat-growing states in India (Pavithra et al., 2020; Singh, 2017; Singh et al., 2017). Breeding techniques have yielded more than 500 maize varieties in major staple crops have received widespread adoption (Pingali, 2012; Singh, 2017; Singh et al., 2017). In case of maize, the average age of cultivars is 14–24 years in Kenya (Walker et al., 2015) and 18 years in sub-Saharan Africa (Witcombe et al., 2016).

According to Singh et al. (2020), farmer’s preference for older varieties in India is evident from the quantity of breeder seed (BS) indented. It is observed that yield gains of these obsolete cultivars are severely deteriorating due to growing prevalence of extreme weather conditions and resurgence of new diseases and pests (Atlin et al., 2017). Such a scenario, varietal replace race (VRR) could be a key driver for accelerating the genetic gain (Spielman and Melinda, 2017). The VRR reflects the pace with which new varieties with enhanced yield and resilience are deployed at farmers’ field to replace the existing cultivars.

Farmers in the USA, China and Europe have now higher accessibility to newly released varieties that are better adapted to the current situations (Atlin et al., 2017). The varietal turnover period of hybrid maize in the USA has been reduced to 3–4 years from that of eight years in the early 1990s (Abate et al., 2017). Likewise, variety turnover time in tropical countries viz., Mexico, Brazil and Argentina is reported to be 3–4 years in comparison with 5–7 years in the subtropics and in Asia (Abate et al., 2017). The high average age of the predominant hybrids (13 years) in sub-Saharan Africa has greatly hampered achieving potential yield gain in maize (Abate et al., 2017). A comparative assessment of cultivar adoption among three African countries suggested Ethiopia as having the lowest percentage of farmers (25%) adopting improved maize cultivars, while Tanzania (58%) and Malawi (61%) had the higher proportions (Westengen et al., 2019). Replacing older maize varieties with improved drought-tolerant varieties is reported to enhance maize yields and reduce poverty by 13.3% and 12.9%, respectively, in rural Nigeria (Wossen et al., 2017). Higher genetic gains and resistance levels from higher VRR have been evident from the data of semi-dwarf high-yielding wheat varieties adopted during 1960 and 1970 in India (Byerlee and Heisey, 1990). Farmers’ awareness about improved varieties showed positive association with the adoption of improved pulses’ varieties in Tanzania and Ethiopia (Abate et al., 2012; Amare et al., 2012). For replacing the existing popular variety, modern plant breeders have to develop market-oriented ‘product profiles’ with clearly defined ‘trait package’ that may help encouraging farmers to accept new variety (Cobb et al., 2019; Ragot et al., 2018; https://excellenceinbreeding.org/blog/product-profiles-are-blueprint-breeding-impact/). Engaging farmers in selection in crop breeding trials and nursery through participatory plant breeding and participatory varietal selection could also contribute to enhancing VRR (Atlin et al., 2017).

Like VRR, availability of quality seed and seed replacement ratio (SRR) could contribute to improving genetic gain. Low SRR in India despite increased availability of quality seed is due to farmers being accustomed to use >70% farm-saved seed for raising the succeeding crop (Pattanaik, 2013). Recently, the SRR of various crops including cereals, pulses and oilseeds has seen a notable rise in India following implementation of national seed policy (2002) that encouraged farmer’s access to seeds of newly developed varieties and replacement of old varieties (Singh et al., 2017). In this context, recent initiatives by Department of Agriculture Cooperation and Farmers Welfare (DACFW), India and Indian Council of Agricultural Research (ICAR), India on enhancing availability of quality seeds to farmers are noteworthy, such as creation of seed hubs for major pulse, millet and oilseed crops.

Seed certification being an essential step for seed quality control (QC) merits attention of both public and private agencies. Flexible systems for seed certification are warranted such that of quality declared system (QDS) adopted in countries such as Kenya and Zambia, where seed certification is licensed to private institutions (Varshney et al., 2018). With the increasing number of cultivars being released in different crops, the morphological descriptors used for discriminating these become increasingly limited and the procedure of testing the genetic purity of cultivars (grow-out test) is time-consuming, costly and prone to environmental fluctuations. In this context, modern genomic technologies owing to their high throughput and environmental independent nature facilitate cost-effective and reliable examination of genetic purity and identity, complementing the quality assurance (QA) and quality control (QC) system of various seed companies and seed certification agencies (Bohar et al., 2020). For instance, low-density SNP assays optimized for several crops facilitate data generation of 10-100 SNPs in US $ 1–5 per sample including DNA extraction (http://cegsb.icrisat.org/high-throughput-genotyping-project/htgp/). More recently, specific-locus amplified fragment sequencing (SLAF-seq) technology and customized SNP array (maizeSNP3072) were optimized to support varietal identification in soybean (Zhang et al., 2020) and maize (Tian et al., 2015), respectively. Similarly, Pemberton et al. (2016) demonstrated the utility of the GBS technology in testing seed purity of ryegrass cultivars by detecting the mismarked seed lots. Recognizing the immense potential of genomic technologies to address seed quality-related issues, the International Union for the Protection of New Varieties of Plants (UPOV) has also set guidelines for using marker technologies in distinctness, uniformity and stability (DUS) testing (https://www.upov.int/edocs/tgpdocs/en/tgp_15.pdf).

Collectively, increased genetic gain for meeting the rising demand of food grain could be achieved through a holistic approach covering re-orientation of public–private programme related to seed business, implementation of sound seed policies and farm innovation to farmers’ awareness (Alwang et al., 2017; Siddique et al., 2012). As has been adopted recently in India, seed production of obsolete cultivars should be discouraged through denotifying/decertification the obsolete varieties or varieties older than 10 years (Shiferaw et al., 2013). Farmers’ access to newly developed varieties also depends upon the streamlining and accelerating the varietal release and notification processes. Extension activities also need attention for disseminating the information on the latest released varieties with the package of
practices clearly highlighting their unique advantages over the obsolete varieties (Atlin et al., 2017; Singh et al., 2020).

Conclusion and prospects

Recent progress in genomics research has provided geneticists, biologists and breeders with a number of modern tools and technologies that impart precision and efficiency to breeding programmes. Reference genome assemblies are increasingly becoming available, and consequently, methods of gene discovery and trait manipulation have been transformed. Genomics research is also advancing gene editing methods in plants for elucidating candidate genes and genetic interactions.

Breeding techniques such as marker-assisted back crossing (MABC) are suited more for defect elimination of mega-varieties; however, enhancing genetic gains per unit time warrants rapid population improvement informed by genome-wide predictions and associations (Varshney et al., 2019a). Increasing access to the deluge of multi-omics information and high-dimensional phenotypic data are also revealing the potential challenges associated with handling and interpretation of the data. Plant breeders need to be trained adequately, and this would play a significant role in embracing more sophisticated approaches such as systems biology-driven breeding for crop improvement (Lavarenne et al., 2018). Adopting these new approaches would fast track the development of climate-smart cultivars. Notwithstanding this, enhanced variety release and seed distribution systems remain instrumental to deploy these new climate-smart cultivars at the farmers’ field, concurrent with the replacement of old obsolete cultivars. Such coordinated efforts involving multiple disciplines would be central to provide solutions for sustainable agriculture.

Acknowledgements

AB and UCI acknowledge support from Indian Council of Agricultural Research (ICAR), New Delhi. RKV is grateful to Science & Engineering Research Board (SERB), Department of Science & Technology, Government of India for awarding JC Bose National Fellowship. RKV thanks Bill and Melinda Gates Foundation for partial support.

Conflict of interest

The authors declared that they have no conflict of interest.

Author contributions

AB and RKV jointly developed the conceptual structure. AB prepared the original draft. UCI and IDG contributed specific sections. RKV edited manuscript. All authors read and approved the final manuscript.

References

Abate, T., Alene, A.D., Bergvinson, D., Shiferaw, B., Sim, S., Orr, A. and Asfaw, S. (2012) Tropical grain legumes in Africa and south Asia: knowledge and opportunities. Nairobi, Kenya: International Crops Research Institute for the Semi-Arid Tropics.

Abate, T., Fisher, M., Abdoulaye, T., Kassie, G.T., Lunduka, R., Marenya, P. and Asnake, W. (2017) Characteristics of maize cultivars in Africa: how modern are they and how many do smallholder farmers grow? Agric. Food Secur. 6, 1–17.

Abbai, R., Singh, V.K., Nachimuthu, V.V., Sekaraj, R., Vippalari, A.K., Singh, A.K., Singh, U.M. 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.

Akdemir, D. and Isidro-Sanchez, J. (2019) Design of training populations for selective genotyping in genomic prediction. Sci. Rep. 9, 1446.

Alwang, J., Gotor, E., Theile, G., Hareau, G., Jaleta, M. and Chamberlin, J. (2017) Pathways from research on improved staple crop germplasm to poverty reduction for small- holder farmers. Agric. Syst. 172, 16–27.

Amare, M., Asfaw, S. and Shiferaw, B. (2012) Welfare impacts of maize-pigeon pea intensification in Tanzania. Agric. Econ. 43, 27–43.

Annichiarico, P., Nazzini, N., Pace, L., Romani, M., Ferrari, B., Wei, Y. and Brummer, E.C. (2017) GBS-based genomic selection for pea grain yield under severe terminal drought. Plant Genome, 10, 2.

Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature, 408, 796–815.

Atlin, G.N., Cairns, J.E. and Das, B. (2017) Rapid breeding and varietal replacement are critical to adaptation of cropping systems in the developing world to climate change. Global Food Secur. 12, 31–37.

Baghill, N., Jarquin, D., Song, O., Nelson, R., Cregan, P., Specht, J. and Lorenz, A. (2015) A population structure and genome-wide association analysis on the USDA soybean germplasm collection. Plant Genome, 8, 1–13.

Belamkara, V., Guttieri, M.J., Hussain, W., Janqui, D., El-bayyoni, I., Poland, J., Lorenz, A.J. et al. (2018) Genomic selection in preliminary yield trials in a winter wheat breeding program. G3 Genes Genome Genet. 8, 2735–2747.

Belser, C., Istance, B., Denis, E., Dubarry, M., Baurum, F.C., Falentin, C., Genete, A. et al. (2018) Chromosome-scale assemblies of plant genomes using nanopore long. Nat. Plants, 4, 879–887.

Beukert, U., Li, Z., Liu, G., Zhao, Y., Ramachandra, N., Mirdita, V., Pita, F. et al. (2017) Genome-based identification of heterotic patterns in rice. Rice 10, 22.

Bevan, M.W., Uauy, C., Wulff, B.B.H., Zhou, J., Krasileva, K. and Clark, M.D. (2017) Genomic innovation for crop improvement. Nature 543, 346–354.

Bohra, A., Chitimkini, A. and Varshney, R.K. (2020) Genetic molecular markers to accelerate genetic gains in crops. Biotechniques 2020, https://doi.org/10.2144/tnn-2020-0066.

Bohra, A. (2013) Emerging paradigms in genomics-based crop improvement. Sci. World J. 2013, 585-587.

Bohra, A., Jha, U.C., Kavi Kishor, P.B., Pandey, S. and Singh, N.P. (2014b) Genomics and molecular breeding in lesser explored pulse crops: Current trends and future opportunities. Biotechnol. Adv. 32, 1410–1428.

Bohra, A., Pandey, M.K., Jha, U.C., Singh, B., Singh, I.P., Datta, D., Chaturvedi, S.K. et al. (2014a) Genomics-assisted breeding in the four major pulse crops of developing countries: Present status and prospects. Theor. Appl. Genet. 127, 1263–1291.

Bohra, A., Saxena, K.B., Varshney, R.K. & Saxena, R.K. (2020) Genomics assisted breeding for pigeonpea improvement. Theor. Appl. Genet. 133, 1721–1737.

Bohra, A. and Singh, N.P. (2015) Whole genome sequences in pulse crops: A global community resource to expedite translational genomics and knowledge-based crop improvement. Biotechnol. Lett. 37, 1529–1539.

Bolger, M.E., Weisshaar, B., Scholtz, U., Stein, N., Usadel, B. and Mayer, K.F. (2014) Plant genome sequencing - applications for crop improvement. Curr. Opin. Biotechnol. 26, 31–37.

Bouchet, S., Olatoye, M.O., Maria, S.R., Perual, R., Tesio, T., Yu, J., Tuijnstra, M. et al. (2017) Increased power to dissect adaptive traits in global sorghum diversity using a nested association mapping population. Genetics, 206, 573–586.

Brocks, C., Nekrasov, V., Lippman, Z.B. and Van Eck, J. (2014) Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. Plant Physiol. 166, 1292–1297.

Buckler, E.S., Holland, J.B., Bradbury, P.J., Acharya, C.B., Brown, P.J., Browne, C. et al. (2009) The genetic architecture of maize flowering time. Science, 325, 714–718.
Byerlee, D. and Heisey, P.W. (1990) Wheat varietal diversification over time and space as factors in yield gains and rust resistance in the Punjab. In Accelerating the transfer of wheat breeding gains to farmers: A study of the dynamics of varietal replacement in Pakistan (P.W. Heisey ed), pp. S-24. CIMMYT Research Report No. 1 Mexico, D.F. Mexico.

Cerrudo, D., Cao, S., Yuan, Y., Martinez, C., Suarez, E.A., Babu, R., Zhang, X. et al. (2018) Genomic selection outperforms marker assisted selection for grain yield and physiological traits in a maize doubled haploid population across water treatments. Front. Plant Sci. 9, 366.

Chang, F., Guo, C., Sun, F., Zhang, J., Wang, Z., Kong, J., He, Q. et al. (2018) Genome-wide association studies for dynamic plant height and number of nodes on the main stem in summer-sown soybeans. Front. Plant Sci. 9, 1184.

Chen, K., Wang, Y., Zhang, R., Zhang, H. and Gao, C. (2019b) CRISP/Cas genome editing and precision plant breeding in agriculture. Annu. Rev. Plant Biol. 70, 667–697.

Chen, Q., Yang, C.J., York, A.M., Xue, W., Daskalaki, L.L., DevKalk, C.A., Krueger, K.W. et al. (2019a) TeonAM: a nested association mapping population for domestication and agronomic trait analysis in Maize. Genetics, 213, 1065–1078.

Cobb, J.N., Juma, R.U., Biswas, P.S., Arbelaez, J.D., Rutkoski, J., Atlin, G., Hagen, T. 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.

Cook, J.P., McMullen, M.D., Holland, J.B., Tian, F., Bradbury, P., Ross-Ibarra, J., Buckler, E.A. et al. (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol. 158, 824–834.

Crosier, J.S., Pazos-Navarro, M., Bennett, R.G., Tschirek, S., Edwards, K., Erskine, W., Creasy, R. and Ribalta, F.M. (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.

Crossa, J., Perez, P., Hickey, J., Burgueño, J., Ornella, L., Cerón-Rojas, J., Zhang, X. et al. (2014) Genomic prediction in CIMMYT maize and wheat breeding programs. Hered. (Edinb) 112, 48-60.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquin, D., de los Campos, G., Burgueño, J. et al. (2017) Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci. 22, 961-975.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. and Blaxter, M.L. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12, 499–510.

Dell’Acqua, M., Gatti, D.M., Pea, G., Cattonaro, F., Coppins, F., Magris, G., Hilaire, A.L. et al. (2015) Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. Genome Biol. 16, 167.

Deokar, A., Sagi, M., Daba, K. and Taran, B. (2019) QTL sequencing strategy to Fang, C., Ma, Y., Wu, S., Liu, Z., Wang, Z., Yang, R., Hu, G. et al. (2014) Genomic association studies dissect the genetic networks underlying agronomical traits in soybean. Genome Biol. 18, 161.

Fuentes, R.R., Chebotarav, D., Dastam, J., Smith, S., De la Hoz, J.F., Mohiyuddin, M., Wing, R.A. et al. (2019) Structural variants in 3000 rice genomes. Genome Res. 29, 870–880.

Gangurde, S.S., Wang, Y., Yaduru, S., Pandey, M.K., Fountain, J.K., Chu, Y., Isleib, T. et al. (2019) Nested-association mapping (NAM)-based genetic dissection uncovers candidate genes for seed and pod weights in peanut (Arachis hypogaea). Plant Biotechnol. J. 18, 1457–1471.

Ghosh, S., Watson, A., Gonzalez-Navarro, O.E., Ramirez-Gonzalez, R.H., Yanes, L., Mendoza-Suárez, M., Simmonds, J. et al. (2018) Speed breeding in growth chambers and glasshouses for crop breeding and model plant research. Nat. Protoc. 13, 2944–2963.

Goff, S.A., Schnable, J.C. and Feldmann, K.A. (2014) The evolution of plant gene and genome sequencing. Adv. Bot. Res. 69, 47–90.

Guo, D., Jiang, H., Yan, W., Yang, L., Ye, J., Wang, Y., Yan, Q. et al. (2020) Resequencing 200 Flx cultivated accessions identifies candidate genes related to seed size and weight and reveals signatures of artificial selection. Front. Plant Sci. 10, 1682.

Guo, Y., Li, Y., Hong, H. and Qiu, L.J. (2014) Establishment of the integrated applied core collection and its comparison with mini core collection in soybean (Glycine max). Crop J. 2, 38–45.

Hickey, L.T., Germán, S.E., Pereyra, S.A., Díaz, J.E., Zimms, L.A., Fowler, R.A., Platz, G.J. et al. (2017) Speed breeding for multiple disease resistance in barley. Euphytica 203, 71–83.

Huang, B., Hafeez, A.N., Robinson, H., Jackson, S.A., Leal-Bertioli, S.C.M., Tester, M., Gao, C. et al. (2019) Breeding crops to feed 10 billion. Nat. Biotechnol. 37, 744–754.

Hu, H., Scheben, A. and Edwards, D. (2018) Advances in integrating genomics and bioinformatics in the plant breeding pipeline. Agriculture 8, 75.

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J. et al. (2009b) High-throughput genotyping by whole-genome resequencing. Genome Res. 19, 1068–1076.

Huang, B.E., George, A.W., Forrest, K.L., Killian, A., Hayden, M.J., Morell, M.K., Cavanagh, C.R. (2012) A multiparent advanced generation inter-cross population for genetic analysis in wheat. Plant Biotechnol. J. 10, 826–839.

Huang, S.W., Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., Lucas, W.J. et al. (2009a) The genome of the cucumber, Cucumis sativus L. Nat. Genet. 41, 1275–1281.

Huang, B.E., Verbyla, K.L., Verbyla, A.P., Raghavan, C., Singh, V.K., Gaur, P., Leung, H. et al. (2015) MAGIC populations in crops: current status and future prospects. Theor. Appl. Genet. 129, 999–1017.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C. et al. (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. 42, 961–967.

Huang, X., Yang, S., Gong, J., Zhao, Q., Feng, Q., Zhan, Q., Zhao, Y. et al. (2016) Genomic architecture of heterosis for yield traits in rice. Nature 537, 329–333.

Huynh, B., Ehlers, J.D., Huang, B.E., Muruzam-Amaatiani, M., Lonardi, S., Santos, J.R.P., Ndue, A. et al. (2018) A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (Vigna unguiculata L. Walp.). Plant J. 93, 1129–1142.

IRGSP. (2005) The map-based sequence of the rice genome. Nature 436, 793–800.

Jagannathan, D., Boha, A., Thudi, M. and Varshney, R.K. (2020) Fine mapping and gene cloning in the post-NGS era: Advances and prospects. Theor. Appl. Genet. 133, 1791–1810.

Jarquin, D., Kocak, K., Posadas, L., Hyma, K., Jedlicka, J., Graet, G. and Lorez, A. (2014) Genotyping by sequencing for genomic prediction in a soybean breeding population. BMC Genom. 15, 740.

Jarquin, D., Lemes da Silva, C., Gaynor, R.C., Poland, J., Fritz, A., Howard, R., Battenfield, S. and Crossa, J. (2017) Increasing genomic-enabled prediction accuracy by modeling genotype × environment interactions in Kansas wheat. Plant Genome, 10, 2.

Jia, G., Huang, X., Zhi, H., Zhao, Y., Zhao, Q., Li, W. et al. (2013) A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (Setaria italica): Nat. Genet. 45, 957–961.

Jiao, Y., Peluso, P., Shi, J., Liang, T., Sitzter, M.C., Wang, B. et al. (2017) Improved maize reference genome with single-molecule technologies. Nature 546, 524–527.

Jighy, A., Lin, Z., Pemberton, L.W., Cogan, N.O.I., Spangenberg, G.C., Hayes, B.J. et al. (2019) Boosting genetic gain in allogamous crops via speed breeding and genomic selection. Front. Plant Sci. 10, 1364.
Abhishek Bohra et al.

Jonas, E. and de Koning, D.J. (2013) Does genomic selection have a future in plant breeding? Trends Biotechnol. 31, 497–504.

Jordan, K.W., Wang, S., He, F., Chao, S., Lun, Y., Faux, E. et al. (2018) The genetic architecture of genome-wide recombination rate variation in allopolyploid wheat revealed by nested association mapping. Plant J. 95, 1039–1054.

Kang, Y.J., Kim, S.K., Kim, M.Y., Lestari, P., Kim, K.H., Ha, B.K. et al. (2014) Genome sequence of mungbean and insights into evolution within Vigna species. Nat. Commun. 5, 5443.

Kang, Y.J., Sataywa, D., Shim, S., Lee, T., Lee, J., Hvang, W.I. et al. (2015) Draft genome sequence of adzuki bean, Vigna angularis. Sci. Rep. 5, 8069.

Kesey, P.J. (2019) Plant genome sequences: past, present, future. Curr. Opin. Plant Biol. 48, 1–8.

Khan, A.W., Garg, V., Rookkival, M., Golicz, A.A., Edwards, D. and Varsnhey, R.K. (2020) Super-pangenome by integrating the wild-side of a species for accelerated crop improvement. Trends Plant Sci. 25, 148–158.

Khazaei, H., Stoddard, F.L., Purves, R.W. and Vendenberg, A. (2018) A multi-parent faba bean (Vicia faba L.) population for future genomic studies. Plant Genet Resour: Character Utilizat. 16, 419–423.

Kidane, Y.G., Gesesse, C.A., Hailiemariam, B.N., Desta, E.A., Mengistu, D.K., Fadda, C., Pi, M.E. et al. (2019) A large nested association mapping population for breeding and quantitative trait locus mapping in Ethiopian durum wheat. Plant Biotechnol. J. 17, 1380–1393.

Kim, T.S., He, Q., Kim, K.W., Yoon, M.Y., Ra, W.H., Li, F.P. et al. (2016) Genome-wide resequencing of KRICE_CORE reveals their potential for future breeding, as well as functional and evolutionary studies in the post-genomic era. BMC Genom. 17, 1.

Kumar, R., Janila, P., Vishwakarma, M.K., Khan, A.W., Manohar, S.S., Gangurde, S.S. et al. (2019) Whole genome re-sequencing-based QTL-seq identified candidate genes and molecular markers for fresh seed dormancy in groundnut. Plant Biotechnol. J. 18, 992–1003.

Lavarenne, J., Guyomar, S., Sallaud, C., Gantet, P. and Lucas, M. (2018) The spring of systems biology-driven breeding. Trends Plant Sci. 23, 706–720.

Lenaerts, B., Collard, B.C.Y. and Demont, M. (2019) Improving global food security through accelerated plant breeding, as well as functional and evolutionary studies in the post-genomic era. BMC Genom. 17, 1.

Mascher, M., Schreiber, M., Scholz, U., Graner, A., Reif, J.C. and Stein, N. (2013) Does genomic selection have a future in plant breeding, as well as functional and evolutionary studies in the post-genomic era. BMC Genom. 17, 1.

Mackay, I.J., Bansept-Basler, P., Barber, T., Bentley, A.R., Cockram, J., Gosman, T. et al. (2017) Genome analysis identified novel candidate genes for ascochyta blight and Septoria tritici blotch in a vast Central European elite winter wheat population. Theor. Appl. Genet. 128, 2471–2481.

Mira, G., Badoni, S., Anacleto, R., Graner, A., Alexandrov, N. and Sreenivasa, N. (2017) Whole genome sequencing-based association study to unravel genetic architecture of cooked grain width and length traits in rice. Sci. Rep. 7, 12478.

Mobini, S.H. and Warkentin, T.D. (2016) A simple and efficient method of in vivo rapid generation technology in pea (Pisum sativum L.). In Vitro Cell Dev. Biol. 52, 530–536.

Morrell, P.L., Buckley, E.S. and Ross-Ibarra, J. (2012) Crop genomics: advances and applications. Nat. Rev. Genet. 13, 85–96.

Nishimura, T.H.E., Hayes, B.J. and Goddard, M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157, 1819–1829.

Müller, S.G., Jost, M., Taketa, S., Rey Mazon, E., Himmelbach, A., Oppermann, M. et al. (2019) Genebank genomics reveals the diversity of a global barley collection. Nat. Genet. 51, 321–326.

Mirdita, V., He, S., Zhao, Y., Korzun, V., Bör, R., Ebmeyer, E. et al. (2015) Potential and limits of whole genome prediction of resistance to Fusarium head blight and Septoria tritici blotch in a vast Central European elite winter wheat population. Theor. Appl. Genet. 128, 2471–2481.

Mugulo, I.W., Low, M.C., Rain, M., Inch, C., Spangenberg, G.C., Wang, J. et al. (2016) Targeted genotyping-by-sequencing permits cost-effective identification and discrimination of pasture grass species and cultivars. Theor. Appl. Genet. 129, 991–1005.

Peng, C., Wang, H., Xu, X., Wang, X., Chen, X., Wei, W., Lai, Y. et al. (2018) High-throughput detection and screening of plants modified by gene editing using quantitative real-time polymerase chain reaction. Plant J. 95, 557–567.
Singh, R.P. (2017) Improving seed systems resiliency at local level through participatory approach for adaptation to climate change. *Adv. Plant Agric.* 6, 00200.

Saxena, K.B., Saxena, R.K. and Varshney, R.K. (2017) Use of immature seed for breeding decisions: a framework. *Theor. Appl. Genet.* 136, 954–957.

Schindele, A., Donn, A. and Puchtla, H. (2020) CRISPR/Cas brings plant biology and breeding into the fast lane. *Curr. Opin. Biotech.* 61, 7–14.

Schneeburger, K. (2014) Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nat. Rev. Genet.* 15, 662–676.

Schwarz, D.C., Li, X., Hernandez, L.I., Ramainain, S.P., Hulff, E.J. and Wang, Y.K. (1993) Ordered restriction maps of Saccaromyces cerevisiae chromosomes constructed by optical mapping. *Science*. 262, 110–114.

Scott, M.F., Ladejobi, O., Amer, S., Bentley, A.R., Biernaskie, J., Scott, B.A., Clark, M. et al. (2020) Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. *Heredity*. https://doi.org/10.1038/s41437-020-0336-6.

Sharma, R., Dracchio, F., Bull, H., Herzig, P., Maurer, A., Pillen, K. Thomas, W.T.B. and Flavell, A.J. (2018) Genome-wide association of yield traits in a nested association mapping population of barley reveals new gene diversity for future breeding. *J. Exp. Bot.* 69, 3811–3822.

Shiferaw, B., Smale, M., Braun, H.J., Duveller, E., Reynolds, M. and Murchio, G. (2013) Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Sec.* 5, 291–317.

Shulaev, V., Sargent, D.J., Crowhurst, R.N., Mockler, T.C., Polkerts, O., Delcher, A.J. et al. (2013) The genome of woodland strawberry (*Fragaria vesca*). *Nat. Genet.* 45, 109–116.

Siddique, K.H., Johansen, C., Turner, N.C., Jeffroy, M.-H., Hashem, A., Sakar, D., Gan, Y. et al. (2012) Innovations in agronomy for food legumes. *A review. Agron. Sustain. Dev.* 32, 45–64.

Singh, R.P. (2017) Improving seed systems resiliency at local level through participatory approach for adaptation to climate change. *Adv. Plant Agric.* 6, 00200.
Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., Yu, Y. et al. (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat. Biotechnol. 33, 408–414.

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M.K., Zhang, C., Chang, W.C. et al. (2019) The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nat. Genet. 51, 865–876.

Zsögön, A., Cermák, T., Naves, E.R., Notini, M.M., Edel, K.H., Weinl, S., Freschi, L., et al. (2018) De novo domestication of wild tomato using genome editing. Nat. Biotechnol. 36, 1211–1216.