Trends in
Plant Science

CellPress
OPEN ACCESS

Special Issue: Feeding the World: The Future of Plant Breeding

Feature Review

Designing Future Crops: Genomics-Assisted Breeding Comes of Age

Rajeev K. Varshney [1,2,*] Abhishek Bohra [3] Jianming Yu [4] Andreas Graner [5] Qifa Zhang [6] and Mark E. Sorrells [7]

Over the past decade, genomics-assisted breeding (GAB) has been instrumental in harnessing the potential of modern genome resources and characterizing and exploiting allelic variation for germplasm enhancement and cultivar development. Sustaining GAB in the future (GAB 2.0) will rely upon a suite of new approaches that fast-track targeted manipulation of allelic variation for creating novel diversity and facilitate their rapid and efficient incorporation in crop improvement programs. Genomic breeding strategies that optimize crop genomes with accumulation of beneficial alleles and purging of deleterious alleles will be indispensable for designing future crops. In coming decades, GAB 2.0 is expected to play a crucial role in breeding more climate-smart crop cultivars with higher nutritional value in a cost-effective and timely manner.

Fifteen Years of Genomics-Assisted Breeding (GAB): Concept to Product Delivery

Ensuring a sustainable increase in global food production with finite resources for an increasing human population is a great challenge. In the wake of the enormous genomic advances, 15 years back we proposed the concept of GAB for accelerating crop improvement [1]. Interestingly, the proposition coincided with the release of a high-quality genome sequence assembly of rice (*Oryza sativa*), representing the first genome sequence of any crop plant [2]. Subsequently, a vast array of genomic tools and technologies have now become available for applications in crop breeding (Table 1). Parallel to the advancements in genomic technologies, innovative genetic designs based on multi-parent synthetic populations were implemented for trait discovery that impart benefits of both association mapping and linkage analysis, such as higher genetic diversity, controlled structure, greater power for quantitative trait locus (QTL) detection and improved mapping accuracy [3,4].

GAB approaches have contributed to comprehensive characterization of allelic variation underlying important agronomic traits and their efficient incorporation in the germplasm enhancement and cultivar development processes. In this article, we discuss the improved crop products delivered through GAB and the potential opportunities that the latest genomics and breeding innovations offer to sustain recent gains in the coming decades [i.e., GAB 2.0 or genomic breeding (GB)]. We highlight approaches that fast-track targeted manipulation of broad allelic variation to create novel diversity for breeding and selection.

Breakthrough Success Stories

Over the past 15 years, GAB has expedited timelines of breeding progress across a broad range of crop species, with the development of more than 130 publicly bred cultivars of different crops [5]. The majority of the noteworthy crop products delivered by GAB applied in a variety of breeding programs include improved cultivars having elevated resistance levels against important diseases such as bacterial blight and blast in rice and rust in wheat (*Triticum aestivum*). Among

Highlights

Availability of reference genomes and genome-wide surveys on comprehensive diversity panels pave the way to associate the allelic variation with phenotypes.

Methods are now available to evaluate the genetic worth of the vast genetic resources archived in gene banks and streamline application of these resources in crop improvement programs.

Precise genome editing technologies in concert with enhanced trait architectures enable innovative solutions to engineer complex trait variation.

High-throughput phenotyping methods are beginning to alleviate the challenge of accurate, precise, and large-scale measurements of plant performance.

Optimized speed breeding protocols remain crucial to accelerating breeding advance when applied with genomic breeding approaches.

Sustaining gains from genomic breeding seeks fast-tracking exploitation of the minor effect alleles, accumulation of favorable alleles, and purging of deleterious alleles.

[1]Center of Excellence in Genomics and Systems Biology (CEGSB), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India
[2]State Agricultural Biotechnology Centre, Centre for Crop and Food Innovation, Food Futures Institute, Murdoch University, Murdoch, Western Australia, Australia
[3]Crop Improvement Division, ICAR- Indian Institute of Pulses Research (ICAR- IIPR), Kanpur, India
[4]Department of Agronomy, Iowa State University, Ames, IA, USA

biotic stresses, tolerance to submergence, salinity, and drought remained the key target traits for improvement using GAB. A similar impact of GAB approaches has been witnessed in improvement of quality traits in several crop species (Box 1).

Major Technological Advances Leveraging Genomics-Assisted Breeding and Key Lessons Learnt

The past decade has seen a remarkable rise in throughput and accuracy of genome sequencing technologies. **Third generation sequencing technologies** (see Glossary) facilitated development of contiguous, chromosome-scale genome assemblies in many crops. The increased genome sequence information in crops has improved gene mapping strategies used to discover and map genome-wide allelic variation. This, in combination with adoption of more efficient family-based linkage designs/large diversity panels, multi-omics assays, and high-throughput phenotyping (HTP) platforms, has contributed to bridging gaps in genome–phenome maps. Resultant acceleration of gene and trait discovery has in turn imparted precision and efficiency to crop breeding programs. As mentioned in the preceding section, a variety of improved crop products are now available for cultivation in farmers' fields. At the same time, limitations and challenges began to surface with the acquisition of high-throughput and high-dimensional datasets. The caveats associated with fragmented genome assemblies came to the fore and a pressing need was to construct more genome sequences representative of species (**pangenome**) or even the entire genus (super-pangenome) in order to capture a comprehensive view of genetic diversity that spans the entire crop gene pool. Genetic improvement of complex traits demanded efficient breeding methods to facilitate identification and subsequent exploitation of hitherto unexplained trait variation attributable to a vast number of small-effect QTLs. Consequently, breeding methods like **genomic selection (GS)**, that exploit genome-wide marker information, became more relevant to continuous population improvement and improving the rate of **genetic gain** [6]. Likewise, optimization and adoption of techniques fast-tracking the generation turnover by manipulating the plant growth environment is noteworthy. Advances in genome editing have greatly enhanced our capacity to perform accurate and rapid alterations in plant genomes.

Linking Genome Diversity and Complex Trait Phenotypes

Acquisition of End-to-End Crop Genomes

As mentioned in the preceding section, the increased throughput and declining cost of next-generation sequencing (NGS) platforms have caused a surge in whole-genome sequence information in various crops (Figure 1). Following the establishment of the first high-quality reference genome sequence of rice (*O. sativa* v. g. *japonica* cv. Nipponbare) in 2005 (IRGSP RefSeq), Wing *et al.* [7] estimated that nearly 10 000 rice accessions have been sequenced so far. Besides decoding new genome sequences, crop communities strove hard to improve the contiguity and completeness of the fragmented genome assemblies based on short NGS reads. Even the gold standard genome assemblies could resolve **haplotypes** only at critical regions (https://phasegenomics.com/the-era-of-platinum-genomes-has-arrived/). In this context, recently invented long-read sequencers have helped to acquire end-to-end information for all chromosomes in combination with improved haplotype resolution for complete genome assemblies, referred to as **platinum standard** reference genome sequences (psRfs). More recently, Zhou *et al.* [8] constructed psRfs for 12 rice genomes by sequencing them at 100X depth followed by validation with Bionano optical maps. With an average number of 30 contigs and 12 gaps, 12 genome assemblies had an average N50 value of 23.10 Mb and an average 97.9% match with the Benchmarking Universal Single-Copy Orthologs (BUSCO) reference gene set. Using BUSCO sets to assess the completeness of a genome sequence was deemed robust in comparison with conventional parameters, including *k*-mer

[5]Leibniz Institute of Plant Genetics and Crops Plant Research (IPK), Gatersleben, Germany
[6]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China
[7]Department of Plant Breeding and Genetics, Cornell University, Ithaca, NY, USA

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

Table 1. Genome Resources in Ten Topmost Food Crops[a]

Crop	Area (mha)[b]	Production (mmt)[b]	Assembled genomes (Mb)	SNP array	Genomic variation databases	Gene expression atlas	Pan-genome
Wheat (*Triticum aestivum*)	215.9	765.7	14 500 [61]	Wheat 9K iSelect [62] Wheat 90K iSelect [63] Wheat 660K Axiom Wheat HD genotyping array [64] Wheat breeder's genotyping array (Affymetrix Axiom 35K) [65]	Wheat Genome Variation Database (WGVD) [66]	WheatExp [67]	18 Cultivars [68]
Maize (*Zea mays*)	197.2	1148.4	2048 [69]	MaizeSNP50 BeadChip (Illumina Infinium 50K) [70] Subset of MaizeSNP50 (Illumina Infinium 3K) [71] Axiom 600K [72] Axiom 55K [73]	MaizeSNPDB [74]	36 207 Genes [75]	503 Inbred lines [76]
Rice (*Oryza sativa*)	162	755.4	371 [2]	Affymetrix (1M) [77] RiceSNP50 (Illumina Infinium 50K) [78] RICE6K (Illumina Infinium 6K) [79] OsSNPnks (Affymetrix Axiom 50K) [80] Affymetrix GeneChip (44K) [81]	SNP-Seek [82]	[83,84]	66 Accessions [85]
Soybean (*Glycine max*)	120.5	333.7	973 [86]	SoySNP50K [87] SoyaSNP180K Axiom [88]	SoyKB [89]	55 616 Genes [90]	26 Accessions [91]
Barley (*Hordeum vulgare*)	51.1	158.9	4980 [92]; 4790 [93]	9K Illumina Infinium iSelect Custom Genotyping BeadChip [94] 50K Illumina Infinium iSelect [95]	BarleyVarDB [96]	21 439 Genes [97]	20 Accessions [98]
Sorghum (*Sorghum bicolor*)	40	57.9	739 [99]	3K SNP Infinium array [100]	SorGSD [101]	27 577 Genes [102]	n.d.
Rapeseed (*Brassica napus*)	34	70.5	849.7 [103]	International Brassica SNP Consortium (60K) [104]	BnaGVD [105]	101 040 Genes [106]	8 Accessions [107]
Dry beans (*Phaseolus vulgaris*)	33	28.9	473 [108]	BARCBean6K_1, BARCBean6K_2, BARCBean6K_3 [109]	PhaseolusGenes (http://phaseolusgenes.bioinformatics.ucdavis.edu/)	[110]	n.d.
Groundnut (*Arachis hypogaea*)	29.6	48.8	2540 [111]; 2540 [112]	'Axiom_Arachis' SNP array with 58K SNP [113]	n.d.	57 344 Transcripts [114]	n.d.
Sugarcane (*Saccharum officinarum*)	26.7	1949.3	800–900 (Monoploid)	76K SNPs [115] 84K SNPs [116] Axiom Sugarcane100K SNP array [117]	n.d.	n.d.	n.d.

[a]Abbreviation: n.d., no data.
[b]Source: http://www.fao.org/faostat/en/#data/QC.

distributions and contig N50 [9]. The psRfs data acquisition is an essential step toward developing a comprehensive view of genetic diversity available in any crop at the genus level, referred to as genus-level pangenome or the super-pangenome [10]. Availability of psRfs facilitate cataloguing of structural variations (SVs), including insertion/deletion (InDEL), copy number variation (CNV), and presence-absence variation. A growing body of literature on sequencing multiple genomes suggests a greater role for these SVs in crop evolution, domestication, and breeding.

Box 1. Key Products Delivered by Genomics-Assisted Breeding in Some Crops

GAB for Biotic Stress Resistance

Simply inherited traits under the influence of strong-effect QTL, such as disease resistance, remained most preferred for introgression through GAB approaches. The GAB products in rice include 'Improved Samba Mahsuri' (ISM) carrying three bacterial blight (BB) disease (*Xanthomonas oryzae* pv. *oryzae*) genes (*Xa21*, *xa13*, and *xa5*) [132]. Two major blast disease (*Magnaporthe oryzae*) resistance genes (*Pi-2* and *Pi-54*) and a BB gene (*Xa38*) were further stacked into 'ISM' [133,134]. 'Pusa Basmati 1' pyramided with two (*Pi2+Pi5*) and three (*Pi54+Pi1+Pita*) blast genes [135] and improved version of 'Pusa Basmati 1121' and 'Pusa Basmati 6' carrying blast (*Pi2* and *Pi54*) and BB resistance genes (*xa13* and *Xa21*) are among others [136].

A variety of DNA markers were applied in wheat breeding for improving stress response and other agronomic and quality-related traits (http://maswheat.ucdavis.edu/protocols/index.htm). Examples include improved versions of hard red winter wheat (HRWW) cultivars 'Jagger' and 'Overley' carrying genes *Yr40/Lr57* and *Lr58*, respectively [137] and spring wheat cultivar 'HUW510' carrying *Lr34* [138]. In pearl millet, 'HHB 67-improved' represented a downy mildew resistant version of 'HHB 67', which was released for commercial cultivation in India in 2005 (see Rai *et al.* [139]). Other success stories demonstrating potential of GAB in cereal breeding included transfer of eyespot (*Rhizoctonia cerealis*) resistance gene *Pch1*, the recessive resistance genes *rym4/ rym5* against barley yellow mosaic viruses, and *mlo* for barley powdery mildew (*Blumeria graminis* f. sp. *hordei*).

Unlike cereals, GAB in grain legume crops has lagged behind in terms of product delivery; however, genotyping-based selections are now increasingly embraced in breeding programs. For instance, pyramiding resistance against multiple soybean cyst nematode (*Heterodera glycines*) races (2, 3, 5, and 14) in soybean at USDA-ARS has led to the development and registration of high-yielding and multiple disease resistant genotypes 'JTN 5503', 'JTN 5303', 'DS 880', and 'JTN 5109' [140–143]. Similarly, Varshney *et al.* [144] obtained a set of 20 introgression lines in groundnut (*Arachis hypogaea*) showing higher yield and increased rust (*Puccinia arachidis*) resistance through transferring a major effect QTL for rust resistance into the background of three susceptible cultivars ('ICGV 91114', 'JL 24', and 'TAG 24'). In chickpea, simultaneous improvement of resistance to both wilt (*Fusarium oxysporum* f. sp. *ciceris*) and blight (*Ascochyta rabiei*) was shown for a popular chickpea cultivar C 214 [145].

GAB for Abiotic Stress Tolerance

The immense utility of GAB for improving abiotic stress response of crop genotypes is exemplified by the recent release of improved rice cultivars with QTL controlling submergence tolerance (*sub1*), salt tolerance (*Saltol*), and drought tolerance introgressed into them. *Sub1* QTL was introgressed into 'Swarna', a popular high-yielding variety from India, within a short span of 2 years [146]. In Vietnam, nearly ten improved lines were obtained from the cross OM1490/IR64-Sub1 showing 90–99% revival under field conditions [147]. Higher survival rates of improved versions of several mega-varieties, including 'Samba Mahsuri' (BPT 5204), 'CR 1009' from India, 'Thadokkham 1' (TDK1) from Laos, and 'BR 11' from Bangladesh were also evident following the QTL-introgression (see Hasan *et al.* [148]).

The *Saltol* QTL was introgressed into various rice genotypes in different countries, and the candidate varieties targeted for QTL-introgression were 'Pusa Basmati 1121', 'Pusa Basmati 6', 'AS 996', 'BT 7', 'Bacthom 7', 'Q5DB', and 'BRRI-Dhan 49' (see Waziri *et al.* [149]). Successful pyramiding of *Sub1*, *Saltol*, blast (*Pi2*, *Pi9*), and gall midge (*Orseolia oryzae*) genes (*Gm1*, *Gm4*) into 'Improved Tapaswini', which is a gene pyramid (*Xa 4*, *xa5*, *xa13*, *Xa21*) of highly popular indica cultivar 'Tapaswini', was demonstrated [150].

Similar to the above-mentioned examples of breeding for submergence and salinity tolerance, pyramiding of two major-effect QTLs controlling drought tolerance into 'Sabitri' (a popular high-yielding yet drought-susceptible indica variety of Nepal) yielded improved variants with good grain type that are suitable for cultivation in rain-fed areas in Nepal and other countries of South Asia [151]. The availability of stable QTLs having large effects on drought tolerance traits has facilitated development and release of GAB products in other crops as well. For instance, introgression of the '*QTL hotspot*' region controlling drought tolerance traits into the 'Pusa 372' led to the development of the first pulse molecular breeding product in India, 'Pusa 10216' (https://icar.org.in/content/development-two-superior-chickpea-varieties-genomics-assisted-breeding).

GAB for Quality Traits

One of the major breakthroughs in quality improvement in crop plants using GAB involves introduction of the *Gpc-B1* (grain protein content) gene into tetra and hexaploid wheat that has caused creation of high GPC cultivars in different countries viz. USA ('Farnum', 'Lassik', 'Westmore', and 'Desert King-High Protein'), Canada ('Lillian', 'Somerset', 'Burnside'), and Australia (improved introgression lines of 'Wyalkatchem', 'Gladius', 'VR 1128') (see Mitrofanova and Khakimova [152] and references therein). The transfer of variant alleles of *badh2* and *Wx* from basmati into 'Manawthukha' (an elite rice cultivar of Myanmar) resulted in improved fragrance and intermediate amylose content [153]. By reducing the breeding cycles up to 3 years, Chu *et al.* [154] developed 'Tifguard High O/L' groundnut with high oleic acid content and nematode resistance. More recently, GAB varieties of groundnut having improved oil quality combined with resistance to rust and late leaf spot (*Phaeoisariopsis personata* Berk. & Curtis) have been released for commercial cultivation [155,156].

Glossary

AB-QTL: an approach that identifies and transfers QTL in a backcross population derived from a cross between an elite line and an unadapted germplasm. During population development, selection is practiced against unadapted traits in early generations and QTLs are mapped in the advanced stage.

AgRenSeq: R-gene enrichment sequencing; a technique for the identification of polymorphisms co-segregating with functional resistance (R)-gene alleles. It represents a robust and reproducible technique for the hybridization-based specific capture of fragments up to 7 kb in any genomic context and it could be used for gap filling, other types of genome finishing, or structural variation verification.

Chromosome segment substitution lines: a series of near isogenic lines having homozygous and stable segment (s) of the chromosome from the donor parent in the background of recipient parent and which can be used for fine mapping of the QTL under study.

CRISPR-Cas9 technology: a genome editing technology that introduces mutations in the form of insertions and/ or deletions (InDels) or substitutions in the target sequences in the genome.

***De novo* domestication:** a novel strategy for crop breeding that refers to the introduction of domestication alleles into non-domesticated plants.

Deleterious allele: a version of a gene that, on average, decreases the fitness of the organism carrying it.

ExpressEdit: a system that incorporates gene editing directly in the speed breeding system. It holds the potential to bypass the bottlenecks of *in vitro* manipulation of plant materials.

Genetic gain: the improvement in average genetic value in a population or the improvement in average phenotypic value due to selection within a population over cycles of breeding.

Genetic hitchhiking: a process whereby a gene possessing a neutral value achieves a high value, or even fixation, within a population because it is closely linked to a gene that is being selected for.

Genetic drift: the random change in the genetic composition of a population due to chance events causing unequal participation of individuals in producing succeeding generations. Along with

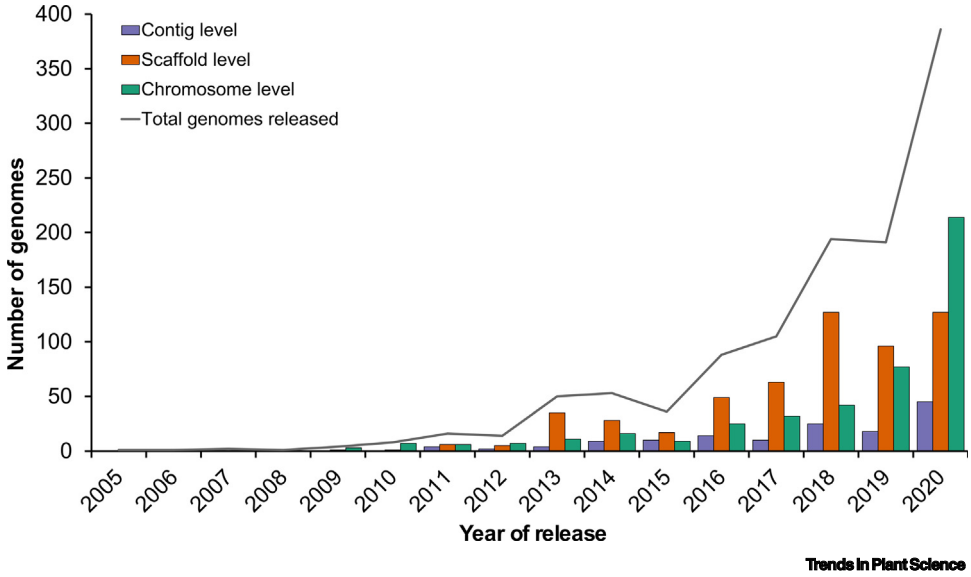

Figure 1. Recent Trend in Whole-Genome Assemblies of Land Plants in Public Domain. Construction of plant genome assemblies has witnessed a surge since the publication of the rice genome in 2005. Whole-genome assemblies of chromosome- or pseudomolecule-level are becoming increasingly available in the public domain owing to the advances in the sequencing technologies and computational tools. The data depicted in the figure was downloaded from the NCBI GENOME REPORTS (https://ftp.ncbi.nlm.nih.gov/genomes/GENOME_REPORTS) and the entries in the 'Land Plants' subgroup with release year between 2005 and 2020 were selected. The entries for mitochondria, chloroplasts, and plastids were removed. We also removed entries with size <100 Mb, which primarily consisted of partial or targeted regions of genome sequenced.

Whole-Genome Surveys and Phenotyping to Access Natural Allelic Variation

High-density genotyping systems enable the survey of comprehensive diversity panels available in many crops to associate the genome information with variations in important phenotypes. Though dependent on the length of the **linkage disequilibrium (LD)** blocks/decay in the genome, the historical recombination events accessible through the diverse collections greatly enhance the resolution of gene mapping. In parallel, multi-parental populations with controlled genetic structures have also emerged as a valuable tool for conducting genome-wide association studies (**GWAS**) [11]. Some recent examples of comprehensive GWAS in crops include GWAS of 10 000 soybean (*Glycine max*), accessions for protein and oil content [12], 3010 rice accessions for grain width and grain length [13], 7887 wheat lines for 50 traits [14], and 4471 maize (*Zea mays*) **landraces** for flowering time adaptation [15].

A considerable outcome of all these genetic mapping experiments is acquisition of genotypic and phenotypic information on large germplasm collections or populations [16]. As in human genetics (https://atlas.ctglab.nl/), the increasing volume of data, resources, and knowledge on gene-trait associations in plants have fostered community web resources such as AraGWAS (https://aragwas.1001genomes.org/#/) and GWASAtlas (https://bigd.big.ac.cn/gwas/) that make GWAS results accessible for thousands of gene-trait associations underlying hundreds of traits across different plant species. The sequence-based GWAS in many crops and the high level of mapping resolution that these studies achieve has been reviewed elsewhere [17]. Recently, proposed quantitative genetic frameworks integrating GWAS revealed the contribution of high-order interactions, including epistasis and **pleiotropy** [18]. In parallel, targeted approaches that integrate association genetics with enrichment sequencing to better investigate broad germplasm collections were proposed, such as **AgRenSeq** for rapid cloning of the disease resistance genes [19].

natural selection, genetic drift is a principal force in evolution.

Genotype–environment interaction (G × E): refers to the positive or negative influence of the environment on the performance of genotypes with respect to quantitative traits.

Genomic selection (GS): a GAB approach in which genome-wide markers are used to estimate the effects of all loci and thereby compute a genomic estimated breeding value (GEBV), so as to achieve more comprehensive and reliable selection.

GWAS: evaluates genetic variants across the genomes of diverse individuals to identify genotype–phenotype associations.

Haplotype: refers to a set of alleles in an organism that are usually inherited together from a single parent.

Haplotype-based breeding (HBB): a promising breeding approach to develop custom-made crop varieties, which deals with identification of superior haplotypes and their deployment in breeding programs.

Identical-by-descent: refers to similar nucleotide sequences present in two or more than two individuals through replication of same ancestral copy of respective sequences.

Intermediate phenotypes: refer to traits positioned somewhere between genetic variation and the terminal phenotype such as yield, disease. They represent a target for attempts to find causal genetic variants and elucidation of mechanisms.

Landrace: a cultivated, and genetically heterogenous line of a plant that has evolved in a particular eco-geographical region and is therefore adapted to the edaphic and environmental conditions and to its traditional management practices and uses.

Linkage disequilibrium (LD): the non-random association of alleles at different loci in a given population. Genetic loci are said to be in linkage disequilibrium when the frequency of association of their different alleles is higher or lower than what would be expected if the genetic loci were independent and associated randomly.

Linkage drag: the random association of undesired region(s) with target genes during introgression. It is more frequently observed during the transfer of target genes from wild species to modern cultivars.

Large structural variations: represent a significant amount of the genetic

Accounting for environmental variation has been a long-standing challenge in complex trait dissection and prediction. Similar to the proposition of 'landscape genomics' in ecology, environmental GWAS in crops uses the climatic conditions as phenotype to identify single nucleotide polymorphisms (SNPs) associated with the accession's environment of origin. To this end, a more recent study by Li *et al*. [20] provided an integrated framework to conduct GWAS and GS in crops with an environmental dimension. Finding patterns in environmental index and associating these patterns with changes in underlying genomic determinants have great implications for understanding complex traits in plants and their prediction for future climates.

Renewed Focus to Identify and Exploit Genomic Loci with Smaller Effects

Many plant traits of agronomic significance are controlled by a large number of QTLs, with variable effects on the corresponding phenotypes. Domestication and early breeding of crops inadvertently targeting alleles with large effect have caused their fixation in elite germplasm. Beneficial alleles with strong effects, such as *teosinte glume architecture1* (*tga1*) and *teosinte branched1* (*tb1*), were brought to fixation during domestication or early breeding in maize, a cross-pollinated crop, whereas in self-pollinating species, the large-effect loci are still evolving [21]. Genetic mapping experiments in different crops also highlight notable contributions of minor-effect QTL to genetic architecture of a variety of traits, such as flowering time, leaf architecture, inflorescence architecture in maize, grain size in rice, and fruit weight in tomato (*Lycopersicon esculentum*). To this end, high-density genotyping systems, in combination with large experimental populations, provide enhanced opportunity to detect QTL that explain relatively small proportions of the phenotypic variation (PV). For instance, genotyping-by-sequencing (GBS) analysis of 1021 maize recombinant inbreds capturing more than 50 000 recombination events revealed 51 QTL explaining an average of 2% of the PV for six plant architecture-related traits [21].

DNA marker-based solutions like marker-assisted backcrossing (MABC) have improved plant traits by facilitating transfer of QTL with strong effects. However, the potential of MABC for improving genetic gain is limited by the number of loci that can be addressed. Wallace *et al.* [16] consider that the extensive presence of minor-effect QTL is the possible reason explaining 'diminishing returns' of current crop breeding practices. Therefore, success of future crop improvement would rely on harnessing the variation attributable to minor-effect loci given that experimental populations with these loci segregating could be created when the majority of major-effect loci have been fixed [21]. In this context, simultaneous improvement of hundreds of minor-effect loci could be achieved by genome-wide prediction and GS that exploit genome-wide marker information instead of only QTL markers.

Molecular Cataloguing of Germplasm Repositories

Crop germplasm repositories are crucial to maintaining genetic diversity that has otherwise been lost as a result of past domestication and current breeding activities intended to develop crops with homogeneity or uniformity [22]. The importance of germplasm is highlighted by the fact that the wild rice (*Oryza nivara*) served as the single source for developing rice varieties resistant to grassy stunt virus (RSGV 1) for more than 30 years, until the emergence of RGSV 2 [23]. Similarly, other important traits sourced from wild relatives and landraces include salinity and submergence tolerance in rice, powdery mildew resistance in barley (*Hordeum vulgare*), and resistance to late blight in potato (*Solanum tuberosum*). Consequently, gene banks conserving useful genetic variation as landraces and wild relatives remain vital to secure future food supply. Worldwide, a total of 7.4 million accessions are stored across 1750 gene banks [24]. However, management of huge collections of crop genetic resources archived in global germplasm repositories is becoming challenging. Despite investing tremendous resources, these gene banks are still plagued by redundancy within and between collections. Identification of these

diversity within a population. They often involve rearrangements of a large region of a chromosome and are known to cause a number of genetic conditions.
Long-day plants: flower only when the amount of daylight lasts longer than the critical threshold (about 12 hours).
Nested association mapping (NAM): a popular multi-parent design comprising various interconnected bi-parental families obtained by crossing diverse parents with a common founder.
Optimal contribution selection (OCS): a selection method that is effective at achieving a balance between rate of inbreeding and genetic gain. This selection process maximizes genetic gain in the next generation while constraining the rate of inbreeding via restriction of relatedness among offspring.
Pangenome: in molecular biology and genetics, a pangenome refers to the entire set of genes and genetic variations present within a species.
Platinum standard: a reference-quality genome that is distinguished from remaining draft assemblies by completeness, low error rates, and a high percentage of the sequences assembled into chromosome-length scaffolds.
Pleiotropy: refers to a single genetic variant that affects two or more seemingly unrelated phenotypic traits.
PAGE: promotion of allele through genome editing; a strategy to enable rapid increases in the frequency of favorable alleles, for improving quantitative traits that are controlled by multiple quantitative trait loci (QTLs).
Quantitative trait nucleotide (QTN): polymorphism that is responsible for the QTL effect (the proportion of the genetic variance, as observed in a segregating population, which is explained by the QTL) and provides useful information about gene function and QTL architecture. Predicting the effects and estimating the accuracy of QTN enhances the rate of genetic gain.
RAGE: removal of allele through genome editing; a strategy for the removal of deleterious variants in the genome that offspring accumulate through inheritance or *de novo* mutagenesis.
See-through phenotyping technology: refers to imaging or phenotyping plant components within tissues or soils.
Short-day plants: bloom when the length of daylight (the photoperiod)

duplicate accessions has always been resource intensive. The marked redundancy within and between germplasm collections has been highlighted by high-density genotyping (GBS and DarTSeq) of 1143 *Aegilops tauschii* accessions from Wheat Genetics Resource Center (WGRC, USA), CIMMYT (Mexico), and Punjab Agricultural University (PAU, India) [25]. In this context, the growing numbers of reference sequences in concert with resequencing facilitate acquisition of molecular data on large germplasm collections (Table 2). Besides supplementing the traditional passport records, the bio-digitalization of gene banks would help to devise better germplasm management strategies and realize the 'evolutionary potential' of gene banks [26]. According to Mascher *et al.* [27], developing molecular profiles of germplasms will help preserve genetic diversity that is lost to spontaneous mutations, **genetic drift** resulting from small population size, and new selection pressure occurring during storage/evaluation in these gene banks.

In parallel, the high-density genetic profiles and readily available passport data and/or historical phenotypic records of these genetic resources facilitate linking of genotype to phenotype, thus revealing the novel haplotypes for downstream applications in breeding programs. For instance,

drops below a certain critical threshold (about 12 hours).

Systems biology: a holistic approach for deciphering the complexity of biological systems that starts from the understanding that the networks that form the whole of living organisms are more than the sum of their parts. It aims to understand the original behavioral properties of the system as a whole, usually through the extensive characterization of the components of the system coupled to mathematical modeling.

Third generation sequencing technologies: a class of DNA sequencing technologies that possess the capability to produce substantially longer reads than second generation sequencing and to produce genome assemblies of unprecedented quality.

Table 2. Examples of Resequencing-Based Investigations on Diverse Germplasm Collections in Plants[a]

Crop	Number of accessions	Average sequencing depth	Genetic variants detected	Refs
Rice (*Oryza sativa*)	3010	n.d.	29 million SNPs, 2.4 million InDels and 90 000 SVs	[13]
Rice	529	n.d.	6.4 million SNPs	[78]
Rice	616	3X to 30X	16.4 million SNPs and 4.8 million InDels	[118]
Rice	176	5.8X	426 337 SNPs and 67 544 InDels	[119]
Soybean (*Glycine max*)	1007	14X	12 197 920 SNPs and 1 873 299 InDels	[120]
Soybean	809	8.3X	10 415 168 SNPs and 1 033 071 small InDels	[121]
Soybean	302	>11X	9 790 744 SNPs and 876 799 InDels	[122]
Brassica napus	991	6.6X	5.56 million SNPs and 1.86 million InDels	[123]
Cotton (*Gossypium* L.)	419	6.55X	3.66 million SNPs	[124]
Cotton	318	5X	8 621 073 SNPs	[125]
Sunflower (*Helianthus annuus*)	287	5X to 25X	5 830 734 SNPs	[126]
Tomato (*Solanum lycopersicum*)	360	5.7X	11 620 517 SNPs and 1 303 213 small InDels	[127]
Grape (*Vitis vinifera*)	472	15.5X	77 726 929 SNPs, 10 278 017 InDels, and 25 000 CNVs	[128]
Citrus (*Citrus* sp.)	100	30X	n.d.	[129]
Cassava (*Manihot esculenta*)	241	30X	25.9 million SNPs and 1.9 million InDels	[51]
Chickpea (*Cicer arietinum* L.)	429	6.84X	4.97 million SNPs	[130]
Pigeonpea (*Cajanus cajan*)	292	5X to 12X	15.1 million SNPs and 2.1 million InDels	[131]

[a]Abbreviation: CNVs, Copy number variations; InDels, Insertion-deletions; SNPs, Single nucleotide polymorphisms; SVs, Structural variations; n.d., no data.

Milner *et al.* [26] performed GBS analysis of 22 626 barley accessions and combined the molecular profiles with legacy passport data of informed subsets to identify GWAS signals for important traits such as spikelet fertility, flowering time, and resistance to bymoviruses. The causative genes thus delineated may be further subjected to allele mining strategies to elucidate other valuable alleles existing at the causative loci. With the help of accurate phenotyping, the phenotypic effects of the allelic variation can be identified and, subsequently, targeted breeding can efficiently recover the superior genotypes having desirable chromosome haplotypes or allelic variation.

Enormous resources required for phenotyping and genotyping of large germplasm collections present an obstacle to identification and utilization of the useful genetic variation locked in gene banks. The challenge is exacerbated by the unpredicted outcomes that are often encountered in the crosses involving exotic germplasms. Methods like GS facilitate not only assessment of the genetic worth of the vast genetic resources archived in the gene banks but also ease genetic exchange in crop germplasm enhancement programs (Box 2). In this context, introgression populations such as **chromosome segment substitution lines** allow examination of the interaction of wild type or exotic QTL with the genome of the cultivated (recipient) parent, thus enabling validation of the phenotypic effects of QTL from wild relatives [28]. Innovative designs combining advanced backcross (AB) with **nested association mapping (NAM)** offer unique opportunities to discover novel variation from exotic germplasm while addressing the problem of population structure and rare allele in association studies. AB-NAM approach in barley based on one cultivar crossed with 25 wild parents facilitated high-resolution mapping of three important quantitative traits [29]. These immortal genetic resources are extremely useful for prebreeding and breeding purposes as these can be shared among research communities.

Genomic Breeding (GB) for Designing Future Crops

For designing future crops, we believe that one or more of the following GB approaches, namely marker-assisted selection (MAS), MABC, marker-assisted recurrent selection (MARS),

Box 2. Genome-wide Prediction and Genomic Selection for Prebreeding

Both genotyping and phenotyping can be performed on the representative sets or the training sets and prediction accuracy may be assessed with validation sets. The resulting genome-wide prediction models in combination with genotypic data of untested germplasm will generate predicted phenotypes of these accessions [27].

Recent simulation and empirical evidence on the application of genomic prediction to gene bank accessions is encouraging. For instance, Yu *et al.* [24] created a 299-accession training set covering 75% diversity of the 962 accessions of reference collection of sorghum. Besides recording high prediction accuracies (0.35–0.78) of the predicted genetic effects (PGE) using validation sets, the potential of PGEs was demonstrated in a broader germplasm context by using 580 exotic germplasm accessions. Expanding genomic predictions to microscopic phenotypes, a follow-up study in maize advocated for implementing double selection based on 'prediction' and 'reliability' to inform decisions while choosing candidates for phenotyping from large gene bank collections [35]. Similarly, Crossa *et al.* [157] assayed 40 000 SNPs on 8416 Mexican landrace accessions and 2403 Iranian landrace accessions of CIMMYT's wheat gene bank to compute prediction accuracies for different traits. The genetic resources with high PGEs can be crossed with each other (conversion) or with an elite line (introgression) for germplasm enhancement.

In GAB 1.0 [1], we anticipated success for **AB-QTL** approach given its ability for simultaneous discovery and introgression of exotic QTL; however, the scheme only registered modest success with few examples in tomato and beans because of limited population sizes and cost. Generating these PGEs for germplasm accessions serves as an excellent tool in the case of landraces since their maladaption to current climates might plague their phenotypic evaluations. Genomic predictions could also underpin the **optimal contribution selection (OCS)** in 'evolving gene banks' of self-pollinating crops by addressing the problems of linkage drag and rapid reconstruction of elite genomes that hampers migration of favorable alleles in exotic × elite populations [158]. Also, GS imparts greater accuracy and rapidity to selection procedures in 'evolving gene banks' even in the absence of pedigree records. When applied with GS, OCS and truncation selection (TS) have shown comparable gains for grain fructan content in wheat, however, OCS retained higher genetic variance and lower inbreeding levels in comparison with TS [159].

haplotype-based breeding, promotion/removal of allele through genome editing (**PAGE**/**RAGE**), and GS can be used in combination with speed breeding. We present GAB 2.0 or GB as an extension of GAB that includes previously discussed approaches as well new approaches that have emerged. Some approaches, like MAS, MABC, and MARS approaches, have been discussed earlier [30]; here we discuss HBB, genomic designing and optimization, PAGE/RAGE, GS, and speed breeding approaches.

Haplotype-based Breeding (HBB)

Advances in plant genome biology have inspired innovative approaches to expedite the progress of assembling desirable phenotypes in crop breeding programs (Figure 2). The haplotype

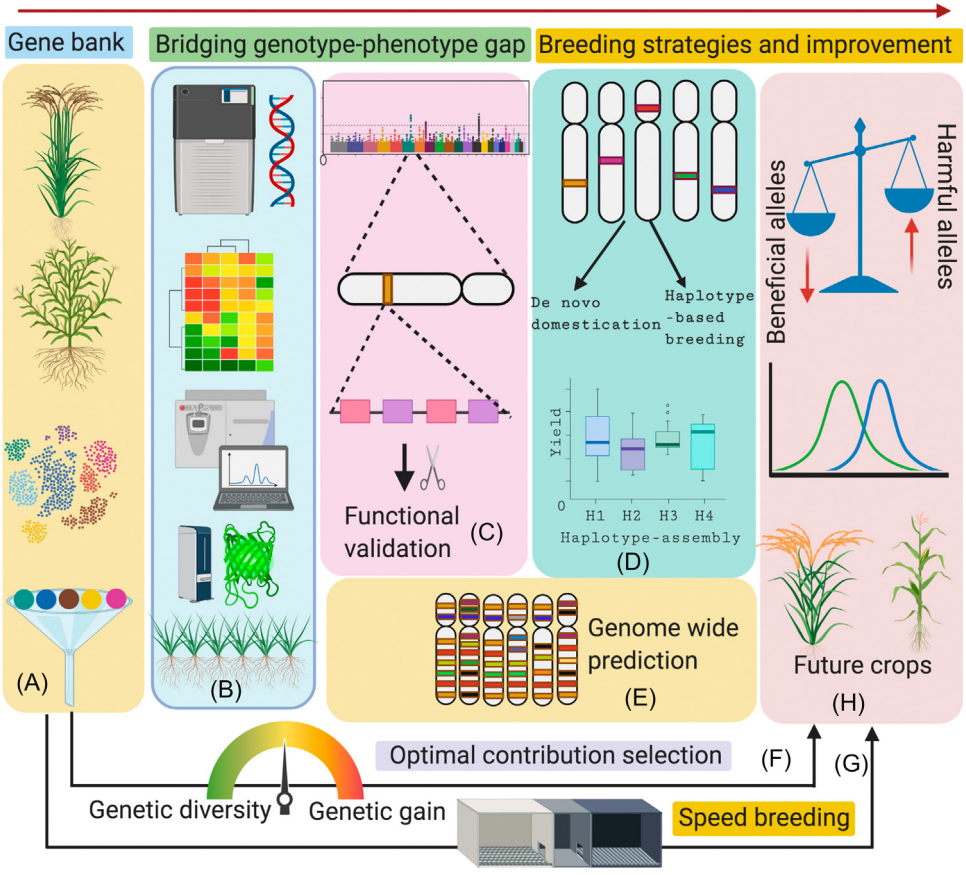

Figure 2. An Overview of Genomics-Assisted Breeding v 2.0 (GAB 2.0) to Deliver Future Crops. The figure illustrates a holistic approach that aims to accumulate favorable alleles or purge deleterious alleles in the plant genomes for designing future crops. (A) Germplasm collections archived in gene banks provide both superior (beneficial) and deleterious effect (harmful) alleles. (B) High-throughput sequencing in combination with multi-omics assays and field phenotyping provides a powerful means to connect genomic variations with the important phenotypes. (C) Once a gene-trait association is identified, functional validation leads to a causative gene. (D) Information about the causative genes affecting key plant traits paves the way for haplotype-based breeding/genomic breeding or *de novo* domestication. (E) In parallel, genomic prediction approach based on genome-wide genotyping information can also be used to make informed decisions in breeding programs. (F) Methods like optimal contribution selection (OCS) that maintain a balance between the rate of genetic gain and genetic diversity/inbreeding will be crucial for prebreeding and breeding purposes. (G) Speed breeding will help expedite crop breeding progress. (H) Implementation of these new breeding tools and approaches will help in accumulating beneficial alleles or purging harmful alleles in breeding population and improving genetic gains of breeding program. This breeding strategy will pave the way for designing future crops. The image was created using BioRender (https://biorender.com/).

assembly approach of Bevan *et al.* [31] from germplasm sequencing data commences with identification and validation of phenotypic effects of the 'component' haplotypes. The development of haplotype-informed DNA markers enables selection of new haplotype combinations. In other words, a set of such haplotype-defining markers provides crop breeders with increased opportunity to attain optimized genetic combinations for improved performance and disrupt **linkage drag**. Large-scale whole-genome resequencing (WGRS) datasets, in combination with *haplo-pheno* analysis, have uncovered useful haplotypes for future breeding in rice [32] and pigeonpea (*Cajanus cajan*) [33]. There is a need to track the inheritance patterns of haplotypes in breeding pedigrees. This is important for assembling new genomic combinations as it helps identify optimal parents for crosses that contain desired combinations of features.

A HBB approach becomes particularly relevant in case of crops having genomes with extended LD blocks. A recent analysis of haplotype blocks in wheat highlighted the inability of SNP markers to distinguish seven haplotypes (H1–H7) predicted for the highly conserved genome region on chromosome 6A that contains the gene *TaGW2-A* controlling various yield-related traits [34]. The study provides evidence for the need to develop haplotype-informed genetic markers for crop improvement instead of relying on SNP markers that are often not causative. We expect HBB approach to be used extensively in coming years with the availability of whole-genome sequence data of germplasm collections in several crops.

Genomic Designing and Optimization

Based on haplotypes, the 'genomic design' concept was given by Yu *et al.* [35] in the case of 'green super rice'. The genome design concept involves enlisting target genes controlling the phenotypes of interest and the germplasm resources where these genes can be sourced. Technically, the 'selection system' comprises a target gene-specific functional marker (R) and the DNA markers (M1 and M2) flanking the target gene, and selection for recurrent parent background (A). Functional selection for R and selection for M1 or M2 and A-background in initial backcross generations, followed by the selections against M1 and M2 (while retaining R gene) in later generations, facilitate speedy recovery of homozygous lines with the precisely incorporated R gene [7]. The selection system facilitated by gene-specific genotyping systems and low-density SNP chips (such as 5–6K) could yield a series of near isogenic lines carrying different target genes within a span of 2–3 years. A series of near-isogenic lines (NILs) in the background of a popular cultivar Kongyu131 carrying four blast resistance genes (*Pi1*, *Pi2*, *Pi9*, and *Pigm*) have been developed in rice.

Concerning hybrid breeding, Jiang *et al.* [36] suggested *in silico* construction of an optimized genome representing a virtual assembly of all superior haplotypes. Analysis of the sequence data of parental lines offers recombination maps having **identical-by-descent** bins. These bins are then used for GWAS and bins associated with the desirable phenotype are identified. The superior haplotypes within these bins are accumulated to synthesize a virtual optimized genome. The parental lines are then ranked according to their contribution of bins to the optimized genome. The high-ranking parental lines thus identified are mated to develop hybrids that are likely to show higher yield than the check variety. In this way, the optimized genome presents a virtual target for heterosis exploitation in any breeding program.

More Predictions and Selective Measurement to Inform Breeding Decisions

Advances in sequencing and phenotyping represent two remarkable breakthroughs of the past decade that are being leveraged in plant science [37]. Both genotyping and phenotyping platforms continue to evolve. Though cost of genotyping has plummeted drastically in recent times, high-density genotyping of entire collections will take some time to become a reality. At

the same time, phenotyping still remains a major stumbling block in breeding progress, although HTP methods are beginning to alleviate this bottleneck (Box 3). **Intermediate phenotypes** are more amenable to HTP analysis than terminal phenotypes (Box 4). In this context, genomic prediction based on the genome-wide marker data and phenotyping will inform breeding decisions (Figure 2). We cite the example of hybrid breeding owing to its reliance on extensive field trials of large-scale testcrosses, a process that is resource-intensive and constrained by the technical problems faced in generating large quantities of hybrid seeds, such as pollen availability from cleistogamous flowers in crops like wheat, pigeonpea, etc. Also, genomic profiles of the hybrids can be easily deduced from that of corresponding parents and a subset of the hybrids may be genotyped to validate the deduced fingerprinting. In wheat, genomic profiles of 1604 single cross hybrids deduced from 135 parental lines, in combination with the extensive phenotyping data (hybrids + parents), allowed prediction of performance of a total of 7441 nonphenotyped hybrids with a high prediction accuracy of 0.89 [38]. In maize, predictions for seven heterotic traits were made with high accuracy (up to 0.80) using genome-wide SNPs and metabolite data of 570 testcrosses that were phenotyped over seven environments [39]. A more recent analysis in pigeonpea predicted performance of a huge set of 78 210 possible single-cross hybrids based on the genomic prediction models trained using WGRS dataset and field phenotyping data of 396 parental lines and 435 single-cross hybrids (R.K. Saxena *et al.* unpublished). Genomic predictions for hybrid performance have been applied in several other crops, including rice [40], pearl millet (*Pennisetum glaucum*) [41], sunflower (*Helianthus annuus*) [42], sorghum (*Sorghum bicolor*) [43], and sugar beet (*Beta vulgaris*) [44].

These genome-wide predictions thus obtained laid the foundation for developing heterotic groups and identification of heterotic patterns to sustain gains in crops from hybrid breeding in

Box 3. Integrating High-Throughput Phenotyping with Genomics-Assisted Breeding to Accelerate Crop Breeding Progress

In the post-NGS era, the phenotyping challenges present the key bottleneck hindering the breeding progress. Manual measurements/visual scoring of plant traits, despite being labor-intensive and error-prone, still play a dominant role to influence key decisions in breeding programs, especially in public sector programs of less-developed countries. The inaccuracies of manual measurements became evident from recent studies on maize streak virus and stay green trait following adoption of HTP technologies (see Araus *et al.* [160]). Growing reliance of modern breeding strategies, including GS models on phenotyping, calls for improvements in accuracy, precision, and throughput of the phenotyping platforms. In other words, enhanced phenotyping capabilities will be crucial to realize the full potential of GAB approaches summarized here. Implications of HTP on various components of a breeder's equation [selection intensity (i), selection accuracy (r), genetic variation (σa), and breeding cycle time (L)] reflect its relevance in context to improving genetic gains in a breeding program [160].

Advances in HTP technologies, particularly imaging and sensor technologies, facilitate high-resolution measurements of plant traits at temporal and spatial scales. A variety of active and passive sensors (photosynthesis, fluorescence, stereo, LiDAR, RGB, multispectral, hyperspectral, and thermal) have been deployed at different levels (leaf, canopy, and air-borne) to monitor the expression of a variety of plant traits in both controlled-environment and field conditions. The technical details, applications, and limitations of these sensor technologies have been thoroughly reviewed by Jin *et al.* [161]. The availability of noninvasive methods for below-ground phenotyping (X-ray microcomputed tomography) and '**see-through phenotyping technologies**' (Terahertz and WiFi holography) [162] creates new avenues for gene discovery. Emphasis is given to field-based phenotyping of a large population to obtain phenotypic information closer to real world settings.

Several national and international initiatives and facilities, including EMPHASIS (https://emphasis.plant-phenotyping.eu/), EPPN (https://eppn2020.plant-phenotyping.eu/), NPPN (https://nordicphenotyping.org/), PhenomUK (https://www.phenomuk.net/), North American HTP facilities (http://nappn.plant-phenotyping.org/high-throughput-phenotyping-facilities/), and Australian plant phenomics facility (https://www.plantphenomics.org.au/about-us/#about-the-appf), have been lunched recently to facilitate acquisition, integration, management, and sharing of the phenomic information. Addressing the current phenotyping challenges will require strong collaborations among diverse fields of science, including engineers, breeders, physiologists, phenotypers, and manufacturers. Also, advancements in machine learning such as neural network have shown promising results in extracting meaningful information from the image data.

Box 4. Multi-omics/Systems Biology Platforms and Mapping of Intermediate Phenotypes

Understanding the genetic make-up of 'terminal phenotypes' such as yield has dominated the genetic mapping experiments in crop plants [163]. Though identification of a few key QTL controlling a trait may be useful to fast-track selection and breeding for yield, such complex traits are seldom resolved to a single gene level. Intermediate phenotypes or endophenotypes serve as bridges to link the genomic variation and terminal phenotypes. For example, transcript-wide association studies in combination with GWAS have shown great potential for prioritization of candidate genes for their association with plant traits [164]. A variety of omics platforms are available that enable accurate and high-throughput measurements of such intermediate phenotypes, including transcripts (RNA-Seq), metabolites (GC-MS, LC-MS, FTICR-MS, NMR) and proteins (MALDI-TOF-MS, iTRAQ). Integration of the multi-omics data with the genetic mapping strategies could greatly accelerate discovery of functional allelic variation [165,166]. In human genetics, intermediate phenotypes have played a pivotal role for genetic mapping of psychiatric disease. Lesser vulnerability of intermediate phenotypes to measurement errors, **genotype–environment interaction (G × E)**, LD, and multiple pathway effects make them more amenable for trait mapping, even with the small sample sizes [164,167]. One should focus on 'intermediate phenotypes' to reach a high-resolution level of genetic mapping [16]. Likewise, 'trait decomposition' allows complex traits to be divided into several subtraits or component traits (such as number of panicles per plant, number of grains per panicle, and grain weight in case of grain yield per plant in rice) for scoring in a high-throughput and accurate manner [163].

the long term. Approaches relying on data mining and design thinking would play a significant role in streamlining GS for hybrid breeding through facilitating an optimal training set design for predicting performance of untested hybrids from a 'large search space' [45]. Design thinking is the iteration process so that we keep improving the solutions based on earlier attempts. Recent improvements in our capacity in genomics, biotechnologies, and phenomics make the process of rethinking and redesigning of the selection and breeding more relevant.

Similarly, deployment of genomic prediction models in variety development pipelines to facilitate evaluation and selection of superior lines for the next cycle in early generations instead of the inbred lines will make breeding programs more efficient and more responsive to current needs. A holistic approach applicable to a broad array of crops was proposed recently by Varshney *et al.* [46] that aims to improve the rate of genetic gains in crop breeding programs, relying more on predictions and less phenotyping.

Male sterility defined by production of nonfunctional pollen represents a promising system to exploit hybrid vigor for improving plant productivity. Among different male sterility systems, cytoplasmic male sterility (CMS) is extensively used for heterosis breeding across different crops. Advances in plant molecular biology have contributed to refine our understanding of CMS and fertility restoration (*Rf*) in crop plants. Map-based approaches have facilitated cloning of a variety of *Rf* genes in different plant species, including rice, sorghum, radish (*Raphanus sativus*), and sugar beet [47]. A more recent study in wheat has demonstrated restoration of *Triticum timopheevii*-type CMS through cleavage of the mitochondrial *orf279* transcript by two nuclear genes, *Rf1* and *Rf3*, that code for pentatricopeptide repeat proteins [48]. Deeper insights into plant CMS-*Rf* system will be crucial to make hybrid breeding more accurate and productive. For instance, diagnostic DNA markers based on the causative Rf locus could help avoid extensive field testing of the large germplasm collections for testing their potential for pollen fertility restoration.

Managing Quantitative Trait Alleles for Greater Response and More Fitness

Enhanced resolution of trait discovery in concert with growing accuracy of GS and precise genome editing technologies [49] paves the way for innovative methods to accumulate favorable alleles in genotypes. Agronomic traits are controlled by a number of **quantitative trait nucleotides (QTNs)** having variable effects on the phenotype. Accumulating these QTNs with conventional breeding may not be efficient due to the loss of favorable alleles driven by **genetic hitchhiking** or drift and reduced selection pressures on the rest of the QTNs [50]. Of the vast

number of QTNs, those having large effects on the phenotypes can be better addressed using genome editing technologies, referred to as promotion of alleles by genome editing (PAGE). A simulation study in livestock comparing different scenarios, including genome selection alone and genome selection supplemented by PAGE (GS+PAGE) elucidated the potential of PAGE in shifting frequency of favorable alleles in both short and long terms [50]. Schemes implementing GS+PAGE had 4.2 times higher response to selection as compared with the GS alone scheme. The authors proposed drift-driven loss of favorable alleles before their fixation in the population as the possible reason underlying this difference. The most dramatic change in allele frequencies occurred in a case where a subset of 20 major-effect QTN were considered for gene editing. No change in allele frequencies was observed when all QTNs having edits were considered for the simulation. The application of PAGE for improving response to selection will require QTNs, especially those having strong effect to be identified with extremely high precision. Toward this end, increased sequence and phenotypic information on comprehensive collections will facilitate GWAS. Besides effect and size, the other features that could be factored while prioritizing alleles for PAGE include their frequencies in the current generation and their vulnerability to recombination [50].

The inefficiency of conventional breeding in purging **deleterious alleles** was evident based on the genomic evolutionary rate profiling and amino acid conservation modeling on the sequence data of diverse accessions. For instance, in cassava (*Manihot esculenta*), domestication and subsequent breeding relying on clonal propagation has increased deleterious alleles by 26% and maintained the deleterious load in the heterozygous state instead of purging this load [51]. In long term, this inadvertent 'shielding' might lead to 'mutation meltdown' or extinction due to gradual accumulation of deleterious alleles in small populations of asexual plants. This makes purging of deleterious alleles a key target in future breeding of crops that harbor substantial genetic load. A recent simulation study based on RAGE suggests that in the long term multiplex genome editing may be a potent and necessary tool to purge deleterious alleles instead of practicing selection against carriers [52]. The potential of PAGE and RAGE relying on simultaneous editing of multiple variants has been assessed in livestock breeding and the tools could potentially extend to plant breeding where breeding efforts may have also reduced fitness and genetic variation of our current crops. Targeted manipulation of quantitative variation through genome editing has shown great potential in ***de novo* domestication** of crop plants (Box 5).

Speed Breeding (SB)

Long generation time of crops to reach fixed homozygous state following hybridization considerably slows down the progress of basic and applied research. Efforts to reduce the generational interval time in different crops relied primarily on off-season nursery/shuttle breeding, double haploid technology, and *in vitro*/embryo culture [53]. Embryo rescue technique, when applied in combination with management of water regimes, light, and temperature, led to eight generations of wheat and nine generations of barley in a year [54]. Similarly, a simplified biotron breeding system based on regulating CO_2 level, photoperiod, and root volume allowed four crossing cycles in rice [55]. To this end, a recently proposed SB technique circumvents the need for cumbersome tissue culture procedures to accelerate time to harvest and instead relies on optimizing plant growth environment (photoperiod, temperature) combined with the application of growth regulators, high-density planting, and early harvesting [56]. Particularly suited for **long-day plants** or day-neutral plants, SB protocols have allowed researchers to grow four to six generations in 1 year of crops, including chickpea (*Cicer arietinum*), pea (*Pisum sativum*), wheat, barley, canola (*Brassica napus*), etc. SB methods markedly reduce breeding cycle time and are considered to be an efficient means to improve selection gains in crop breeding programs. In a collaboration between the University of Queensland and

Box 5. Genome/Gene Editing to Accelerate Crop Domestication

Since its conception in 1968, ideotypes are predicated to have optimal phenotypic manifestations under defined environmental conditions [168]; it has always served as a fascinating crop model to inspire plant breeding communities worldwide. In fact, the ideal phenotype is the one that adapts to every environment in a unique way and therefore cannot be defined. Unlike domestication, the concept relied on accumulating specific traits in the crop instead of performing selection against the exotic features such as shattering, etc. [169]. Based on the morphology and physiology, crop ideotypes were suggested in many crops, including rice, wheat, barley, and maize and witnessed success to a considerable extent. In GAB 1.0, 'Breeding by design' intended to design superior genotypes by capturing allelic variation at all the loci controlling agronomically important traits through: (i) locus identification, (ii) allele mining and exploiting the allelic variation present at all these loci, followed by (iii) crossing the genotypes in a designed way. With recent advances in genomic technologies and characterization of domestication traits in different crops, the 'hypothetical biological model' proposed once as a crop ideotype now seems a realistic proposition (i.e., GAB 2.0).

Resigning plant types for improving agricultural productivity requires changes in plant systems that range from 'straightforward' to 'seemingly fanciful' and enabling technologies in synthetic biology is likely to assume a central role in the latter case [170]. Genome sequence information, in combination with previous genetic mapping experiments on domestication traits, have elucidated gene(s)/loci underlying crop domestication. Notably, the monogenic nature of these genes renders these most suitable for targeted manipulation. In this context, *de novo* domestication, a process of introgressing the domestication loci back into exotic germplasms, is being viewed as a viable alternative to developing future crops [169]. Domestication is a prolonged process that has resulted in inadvertent accumulation of deleterious alleles along with some useful traits into our current crops. As has been described as a 'cost of domestication', the process of domestication is often associated with a marked loss of both fitness and genetic diversity of the domesticated species [171]. *De novo* domestication facilitated by targeted genome editing or introgression breeding could fast-track the recovery of domesticated forms while retaining beneficial exotic traits. Recent studies in different crops have demonstrated the potential of genome editing technology for *de novo* domestication of wild types. For example, by reverse-engineering six domestication genes (*SELFPRUNING*, *OVATE*, *FASCIATED*, *FRUIT WEIGHT*, *MULTIFLORA*, *LYCOPENE BETA CYCLASE*) with **CRISPR-Cas9 technology**, Zsögön *et al.* [172] improved wild tomato ancestor *Solanum pimpinellifolium* with respect to fruit number (10X), size (3X), nutrient content (500% more lycopene accumulation), and plant architecture. In another independent study, *de novo* domestication of *S. pimpinellifolium* was achieved by CRISPR-Cas9-based editing of five target genes (*SP*, *SP5G*, *SlCLV3*, *SlWUS*, *SlGGP1*), resulting in enhanced fruit size and nutrient content while retaining stress tolerance (resistance to bacterial spot disease and salt tolerance) of the wild ancestor [173]. Earlier, Rodríguez-Leal *et al.* [174] obtained novel quantitative variation for key domestication traits (fruit size, inflorescence branching, and plant architecture) in tomato through creating cis-regulatory mutations in three genes using CRISPR/Cas9 technology.

In view of the fact that 90% of the world calories are supplied by only 20 plant species [175], we need to broaden the array of cultivated crops. In order to diversify cropping systems, Massawe *et al.* [175] suggest use of orphan or underutilized crops that are better adapted to local or marginalized environments. Lemmon *et al.* [176] performed *de novo* domestication for a Solanaceae orphan crop groundcherry (*Physalis pruinosa*) by using CRISPR-Cas9-based targeted modification of homologs of tomato domestication and improvement genes, *SELFPRUNING (SP)*, *SELF-PRUNING 5G* (*SP5G*), and *CLAVATA* (*CLV*)–*WUSCHEL*. However, optimized tissue culture and transformation protocols should be in place to make genome editing a routine tool for plant improvement [176].

DOW Agrosciences (https://www.theland.com.au/story/4623477/tackling-pre-harvest-sprouting-in-wheat/), SB facilitated the development of an Australian high-protein milling wheat variety, DS Faraday, that has tolerance to preharvest sprouting and leaf, stripe, and stem rust. A recent study extended the possibility to apply SB to **short-day plants** like soybean using a 10-h photoperiod and a light spectrum enriched with blue light [57]. However, SB experiments on other short-day crops (rice and amaranth *Amaranthus* spp.) highlighted the need for crop-specific optimization of light quality/intensity to improve generation turnover. The importance of both quality and quality of light on plant photo-morphogenic response was evident from an earlier study on grain legumes [58]. The widespread applications of standardized SB systems include assisting hybridization, phenotyping of adult plant traits, MAS for target traits, generation advancement through single seed descent (SSD), and gene editing in several crops.

A new approach, **ExpressEdit**, provides ample scope to combine gene editing and SB systems in a tissue-culture free manner [59]. In ExpressEdit, 'preassembled Cas9-single guide RNA (sgRNA) ribonucleoproteins' are delivered into plant shoot apical meristems using particle

bombardment or biolistic DNA delivery. The plants lacking Cas9 and having the desired trait are selected using MABC. The 'CRISPR-ready' plants thus obtained can be targeted for sgRNA spray.

Concluding Remarks: Way Forward and Challenges Ahead

There are many challenges facing plant breeders in the coming years. Increasing weather volatility, apparent yield plateaus in some crops, connecting genotype to phenotype, predicting genotype by environment interactions, and improved phenotyping methods are just a few examples. Innovative tools and technologies have been instrumental in improving our understanding of genome structure and function, providing the genetic underpinnings of important trait architectures. We anticipate continued improvement in the rate of genetic gains in crop breeding programs across the globe, in spite of climate change, as our capacity to measure and exploit quantitative trait variation in elite varieties, our germplasm repositories and novel variation created using targeted genetic recombination [60] and genome editing. However, achieving desired phenotypic manifestations and improved plant performance by targeted manipulation of a considerable number of major- and minor-effect QTLs also demands a **systems biology** approach in which breeders will need to carefully prioritize traits for a specified target population of environments. Since precise manipulation of the causal loci will require in-depth understanding of the genetic make-up of various traits, equal emphasis should be placed on elucidation of trait architectures in combination with black box approaches like GS that do not necessarily rely on the underlying biology of traits. The potential of GAB 2.0 will be realized in the optimization of the methods of using crop genome information for crop improvement. Such optimization for obtaining desirable phenotypes requires continued innovation in plant breeding methods resulting from new knowledge and improved technologies (see Outstanding Questions).

Acknowledgments

R.K.V. thanks Bill and Melinda Gates Foundation, USA, for supporting projects related to genomics-assisted breeding at ICRISAT and acknowledges Science & Engineering Research Board (SERB), Department of Science & Technology, Government of India for awarding JC Bose National Fellowship. Thanks are also due to Rutwik Barmukh, Prasad Bajaj, Pushpesh Joshi from ICRISAT, and Abha Tiwari from ICAR-IIPR for their help and support while preparing this article.

Declaration of Interests

No interests are declared.

References

1. Varshney, R.K. *et al.* (2005) Genomics-assisted breeding for crop improvement. *Trends Plant Sci.* 10, 621–630
2. International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436, 793–800
3. Kover, P.X. *et al.* (2009) A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet.* 5, e1000551
4. Yu, J. *et al.* (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics* 178, 539–551
5. Vogel, B. (2014) *Marker-Assisted Selection: A Biotechnology for Plant Breeding Without Genetic Engineering*, Greenpeace International
6. Crossa, J. *et al.* (2017) Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 22, 961–975
7. Wing, R.A. *et al.* (2018) The rice genome revolution: from an ancient grain to Green Super Rice. *Nat. Rev. Genet.* 19, 505–517
8. Zhou, Y. *et al.* (2020) A platinum standard pan-genome resource that represents the population structure of Asian rice. *Sci. Data* 7, 113
9. Simão, F.A. *et al.* (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212
10. Khan, A.W. *et al.* (2020) Super-pangenome by integrating the wild-side of a species for accelerated crop improvement. *Trends Plant Sci.* 25, 148–158
11. Scott, M.F. *et al.* (2020) Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. *Heredity* 125, 396–416
12. Bandillo, N. *et al.* (2015) A population structure and genome-wide association analysis on the USDA soybean germplasm collection. *Plant Genome* 8, 1–13
13. Wang, W. *et al.* (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557, 43–49
14. Juliana, P. *et al.* (2019) Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. *Nat. Genet.* 51, 1530–1539
15. Navarro, J.A.R. *et al.* (2017) A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat. Genet.* 49, 476–480
16. Wallace, J.G. *et al.* (2018) On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics. *Annu. Rev. Genet.* 52, 421–444
17. Bohra, A. *et al.* (2020) Genomic interventions for sustainable agriculture. *Plant Biotechnol. J.* 18, 2388–2405

Outstanding Questions

Haplotype-based breeding aims to deliver a designer cultivar by assembling superior haplotypes into one genotype. How will the process take into account novel genetic interactions that the transferred haplotypes establish with the recipient genetic background?

The past decade has witnessed genomic selection (GS) studies focusing on close proximity/genetic relatedness between individuals from training and breeding populations for better accuracy. Given this focus, how can the potential of wild relatives, less represented in current breeding populations, be harnessed for genome-wide predictions?

Does the term '*de novo* domestication' not trivialize a process that has taken hundreds of years of selection and breeding? In this respect, to what extent will systems biology and pangenomic approaches help achieve the goal of obtaining future crops?

How can international service capacity be built to support the breeding programs of lesser developed countries, with limited resources and poor expertise to deploy the high-throughput technologies like GS and HTP that lie at the core of genomics-assisted breeding?

Will speed breeding short-day plants witness similar success to breeding long-day or day-neutral plants? How can we resolve/alleviate the genotype-dependence of current protocols to facilitate their widespread applications?

18. Jiang, Y. *et al.* (2017) A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. *Nat. Genet.* 49, 1741–1746
19. Arora, S. *et al.* (2019) Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nat. Biotechnol.* 37, 139–143
20. Li, X. *et al.* (2021) An integrated framework reinstating the environmental dimension for GWAS and genomic selection in crops. *Mol. Plant* Published online March 10, 2021. https://doi.org/10.1016/j.molp.2021.03.010
21. Wang, B. *et al.* (2018) Identification of minor effect QTLs for plant architecture related traits using super high density genotyping and large recombinant inbred population in maize *(Zea mays)*. *BMC Plant Biol.* 18, 17
22. McCouch, S. *et al.* (2020) Mobilizing crop biodiversity. *Mol. Plant* 13, 1341–1344
23. McCouch, S. *et al.* (2013) Agriculture: feeding the future. *Nature* 499, 23–24
24. Yu, X. *et al.* (2016) Genomic prediction contributing to a promising global strategy to turbocharge genebanks. *Nat. Plants* 2, 16150
25. Singh, N. *et al.* (2019) Efficient curation of genebanks using next generation sequencing reveals substantial duplication of germplasm accessions. *Sci. Rep.* 9, 650
26. Milner, S.G. *et al.* (2019) Genebank genomics highlights the diversity of a global barley collection. *Nat. Genet.* 51, 2
27. Mascher, M. *et al.* (2019) Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. *Nat. Genet.* 51, 1076–1081
28. Dempewolf, H. *et al.* (2017) Past and future use of wild relatives in crop breeding. *Crop Sci.* 57, 1070–1082
29. Nice, L.M. *et al.* (2016) Development and genetic characterization of an advanced backcross - nested association mapping (AB-NAM) population of wild × cultivated barley. *Genetics* 203, 1453–1467
30. Varshney, R.K. *et al.* (2012) Can genomics boost productivity of orphan crops? *Nat. Biotechnol.* 30, 1172–1176
31. Bevan, M.W. *et al.* (2017) Genomic innovations for crop improvement. *Nature* 543, 346–354
32. Abbai, R. *et al.* (2019) Haplotype analysis of key genes governing grain yield and quality traits across 3K RG panel reveals scope for the development of tailor-made rice with enhanced genetic gains. *Plant Biotechnol. J.* 17, 1612–1622
33. Sinha, P. *et al.* (2020) Superior haplotypes for haplotype-based breeding for drought tolerance in pigeonpea (*Cajanus cajan* L.). *Plant Biotechnol. J.* 18, 2482–2490
34. Brinton, J. *et al.* (2020) A haplotype-led approach to increase the precision of wheat breeding. *Commun. Biol.* 3, 712
35. Yu, S. *et al.* (2020) Genomic breeding of Green Super Rice varieties and their deployment in Asia and Africa. *Theor. Appl. Genet.* 133, 1427–1442
36. Jiang, H. *et al.* (2019) Development and evaluation of improved lines with broad-spectrum resistance to rice blast using nine resistance genes. *Rice* 12, 29
37. Fernie, A.R. and Gutierrez-Marcos, J. (2019) From genome to phenome: genome-wide association studies and other approaches that bridge the genotype to phenotype gap. *Plant J.* 97, 5–7
38. Zhao, Y. *et al.* (2015) Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. *Proc. Natl. Acad. Sci. U. S. A.* 112, 15624–15629
39. Riedelsheimer, C. *et al.* (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat. Genet.* 44, 217–220
40. Cui, Y. *et al.* (2020) Hybrid breeding of rice via genomic selection. *Plant Biotechnol. J.* 18, 57–67
41. Jarquin, D. *et al.* (2020) Enhancing hybrid prediction in pearl millet using genomic and/or multi-environment phenotypic information of inbreds. *Front. Genet.* 10, 1294
42. Reif, J.C. *et al.* (2013) Genomic prediction of sunflower hybrid performance. *Plant Breed.* 132, 107–114
43. Ishimori, M. *et al.* (2020) Impacts of dominance effects on genomic prediction of sorghum hybrid performance. *Breed. Sci.* 70, 605–616
44. Würschum, T. *et al.* (2013) Genomic selection in sugar beet breeding populations. *BMC Genet.* 14, 85
45. Guo, T. *et al.* (2019) Optimal designs for genomic selection in hybrid crops. *Mol. Plant* 12, 390–401
46. Varshney, R.K. *et al.* (2019) Towards sequence-based breeding in legumes in postgenome sequencing era. *Theor. Appl. Genet.* 132, 797–816
47. Bohra, A. *et al.* (2016) Cytoplasmic male sterility (CMS) in hybrid breeding in field crops. *Plant Cell Rep.* 35, 967–993
48. Melonek, J. *et al.* (2021) The genetic basis of cytoplasmic male sterility and fertility restoration in wheat. *Nat. Commun.* 12, 1036
49. Zhu, H. *et al.* (2020) Applications of CRISPR–Cas in agriculture and plant biotechnology. *Nat. Rev. Mol. Cell Biol.* 21, 661–677
50. Jenko, J. *et al.* (2015) Potential of promotion of alleles by genome editing to improve quantitative traits in livestock breeding programs. *Genet. Sel. Evol.* 47, 55
51. Ramu, P. *et al.* (2017) Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. *Nat. Genet.* 49, 959–963
52. Johnsson, M. *et al.* (2019) Removal of alleles by genome editing (RAGE) against deleterious load. *Genet. Sel. Evol.* 51, 14
53. Bhatta, M. *et al.* (2021) Need for speed: manipulating plant growth to accelerate breeding cycles. *Curr. Opin. Plant Biol.* 60, 101986
54. Zheng, Z. *et al.* (2013) A procedure allowing up to eight generations of wheat and nine generations of barley per annum. *Euphytica* 191, 311–316
55. Tanaka, J. *et al.* (2016) A practical, rapid generation-advancement system for rice breeding using simplified biotron breeding system. *Breed. Sci.* 66, 542–551
56. Ghosh, S. *et al.* (2018) Speed breeding in growth chambers and glasshouses for crop breeding and model plant research. *Nat. Protoc.* 13, 2944–2963
57. Jähne, F. *et al.* (2020) Speed breeding short-day crops by LED-controlled light schemes. *Theor. Appl. Genet.* 133, 2335–2342
58. Croser, J.S. *et al.* (2016) Time to flowering of temperate pulses in vivo and generation turnover in vivo-in vitro of narrow leaf lupin accelerated by low red to far red ratio and high intensity in the far red region. *Plant Cell Tissue Organ Cult.* 127, 591–599
59. Hickey, L.T. *et al.* (2019) Breeding crops to feed 10 billion. *Nat. Biotechnol.* 37, 744–754
60. Taagen, E. *et al.* (2020) Counting on crossovers: controlled recombination for plant breeding. *Trends Plant Sci.* 25, 455–465
61. International Wheat Genome Sequencing Consortium (IWGSC) (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361, eaar7191
62. Cavanagh, C.R. *et al.* (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc. Natl. Acad. Sci. U. S. A.* 110, 20
63. Wang, Y. *et al.* (2014) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* 32, 947–951
64. Winfield, M.O. *et al.* (2016) High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol. J.* 14, 1195–1206
65. Allen, A.M. *et al.* (2017) Characterization of a wheat breeders' array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*). *Plant Biotechnol. J.* 15, 390–401
66. Wang, J. *et al.* (2020) WGVD: an integrated web-database for wheat genome variation and selective signatures. *Database* 2020, baaa090
67. Pearce, S. *et al.* (2015) WheatExp: an RNA-seq expression database for polyploid wheat. *BMC Plant Biol.* 15, 299
68. Montenegro, J.D. *et al.* (2017) The pangenome of hexaploid bread wheat. *Plant J.* 90, 1007–1013
69. Schnable, P.S. *et al.* (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115
70. Ganal, M.W. *et al.* (2011) A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping,

and genetic mapping to compare with the B73 reference genome. *PLoS One* 6, e28334
71. Rousselle, Y. *et al.* (2015) Study on essential derivation in maize: III. Selection and evaluation of a panel of single nucleotide polymorphism loci for use in European and North American germplasm. *Crop Sci.* 55, 1170–1180
72. Unterseer, S. *et al.* (2014) A powerful tool for genome analysis in maize: development and evaluation of the high density 600 K SNP genotyping array. *BMC Genomics* 15, 823
73. Xu, C. *et al.* (2017) Development of a maize 55 K SNP array with improved genome coverage for molecular breeding. *Mol. Breed.* 37, 20
74. Zhou, W. *et al.* (2019) MaizeSNPDB: a comprehensive database for efficient retrieve and analysis of SNPs among 1210 maize lines. *Comput. Struct. Biotechnol. J.* 17, 1377–1383
75. Hoopes, G.M. *et al.* (2019) An updated gene atlas for maize reveals organ-specific and stress-induced genes. *Plant J.* 97, 1154–1167
76. Hirsch, C.N. *et al.* (2014) Insights into the maize pan-genome and pan-transcriptome. *Plant Cell* 26, 121–135
77. McCouch, S.R. *et al.* (2010) Development of genome-wide SNP assays for rice. *Breed. Sci.* 60, 524–535
78. Chen, W. *et al.* (2014) Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.* 46, 714–721
79. Yu, H. *et al.* (2014) A whole-genome SNP array (RICE6K) for genomic breeding in rice. *Plant Biotechnol. J.* 12, 28–37
80. Singh, N. *et al.* (2015) Single-copy gene based 50 K SNP chip for genetic studies and molecular breeding in rice. *Sci. Rep.* 5, 11600
81. Tung, C. *et al.* (2010) Development of a research platform for dissecting phenotype genotype associations in rice (*Oryza* spp.). *Rice* 3, 205–217
82. Alexandrov, N. *et al.* (2015) SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Res.* 43, D1023–D1027
83. Wang, L. *et al.* (2010) A dynamic gene expression atlas covering the entire life cycle of rice. *Plant J.* 61, 752–766
84. Cao, P. *et al.* (2012) The Rice Oligonucleotide Array Database: an atlas of rice gene expression. *Rice* 5, 17
85. Zhao, Q. *et al.* (2018) Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* 50, 278–284
86. Schmutz, J. *et al.* (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463, 7278
87. Song, Q. *et al.* (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS One* 8, e54985
88. Lee, Y.G. *et al.* (2015) Development, validation and genetic analysis of a large soybean SNP genotyping array. *Plant J.* 81, 625–636
89. Joshi, T. *et al.* (2014) Soybean knowledge base (SoyKB): a web resource for integration of soybean translational genomics and molecular breeding. *Nucleic Acids Res.* 42, D1245–D1252
90. Libault, M. *et al.* (2010) An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *Plant J.* 63, 86–99
91. Liu, Y. *et al.* (2020) Pan-genome of wild and cultivated soybeans. *Cell* 182, 162–176
92. International Barley Genome Sequencing Consortium (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491, 711–716
93. Mascher, M. *et al.* (2017) A chromosome conformation capture ordered sequence of the barley genome. *Nature* 544, 427–433
94. Comadr`an, J. *et al.* (2012) Natural variation in a homolog of *Antirrhinum CENTRORADIALIS* contributed to spring growth habit and environmental adaptation in cultivated barley. *Nat. Genet.* 44, 1388–1392
95. Bayer, M.M. *et al.* (2017) Development and evaluation of a barley 50k iSelect SNP array. *Front. Plant Sci.* 8, 1792
96. Tan, C. *et al.* (2020) BarleyVarDB: a database of barley genomic variation. *Database* 2020, baaa091
97. Druka, A. *et al.* (2006) An atlas of gene expression from seed to seed through barley development. *Funct. Integr. Genomics* 6, 202–211
98. Jayakodi, M. *et al.* (2020) The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature* 588, 284–289
99. Paterson, A.H. *et al.* (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457, 551–556
100. Bekele, W.A. *et al.* (2013) High-throughput genomics in sorghum: from whole-genome resequencing to a SNP screening array. *Plant Biotechnol. J.* 11, 1112–1125
101. Luo, H. *et al.* (2016) SorGSD: a sorghum genome SNP database. *Biotechnol. Biofuels* 9, 6
102. Shakoor, N. *et al.* (2014) A *Sorghum bicolor* expression atlas reveals dynamic genotype-specific expression profiles for vegetative tissues of grain, sweet and bioenergy sorghums. *BMC Plant Biol.* 14, 35
103. Chalhoub, B. *et al.* (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345, 950–953
104. Clarke, W.E. *et al.* (2016) A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single locus markers in the allotetraploid genome. *Theor. Appl. Genet.* 129, 1887–1899
105. Yan, T. *et al.* (2021) BnaGVD: a genomic variation database of rapeseed (*Brassica napus*). *Plant Cell Physiol.* Published online January 5, 2021. https://doi.org/10.1093/pcp/pcaa169
106. Chao, H. *et al.* (2020) BrassicaEDB: a gene expression database for *Brassica* crops. *Int. J. Mol. Sci.* 21, 5831
107. Song, J.M. *et al.* (2020) Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat. Plants* 6, 34–45
108. Schmutz, J. *et al.* (2014) A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46, 707–713
109. Song, Q. *et al.* (2015) SNP assay development for linkage map construction, anchoring whole-genome sequence, and other genetic and genomic applications in common bean. *G3 (Bethesda)* 5, 2285–2290
110. O'Rourke, J.A. *et al.* (2014) An RNA-Seq based gene expression atlas of the common bean. *BMC Genomics* 15, 866
111. Bertioli, D.J. *et al.* (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat. Genet.* 48, 438–446
112. Zhuang, W. *et al.* (2019) The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat. Genet.* 51, 865–876
113. Pandey, M.K. *et al.* (2017) Development and evaluation of a high density genotyping *Axiom_Arachis* array with 58 K SNPs for accelerating genetics and breeding in groundnut. *Sci. Rep.* 7, 40577
114. Sinha, P. *et al.* (2020) *Arachis hypogaea* gene expression atlas for *fastigiata* subspecies of cultivated groundnut to accelerate functional and translational genomics applications. *Plant Biotechnol. J.* 18, 2187–2200
115. Yang, X. *et al.* (2017) Mining sequence variations in representative polyploid sugarcane germplasm accessions. *BMC Genomics* 18, 594
116. Balsalobre, T.W.A. *et al.* (2017) GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane. *BMC Genomics* 18, 72
117. You, Q. *et al.* (2019) Development of an Axiom Sugarcane100K SNP array for genetic map construction and QTL identification. *Theor. Appl. Genet.* 13, 2829–2845
118. Higgins, J. *et al.* (2020) Resequencing of 672 native rice accessions to explore genetic diversity and trait associations along Vietnam. *Res. Square* Published online September 3, 2020. http://doi.org/10.21203/rs.3.rs-70697/v1
119. Yano, K. *et al.* (2016) Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* 48, 927
120. Torkamaneh, D. *et al.* (2020) Soybean (*Glycine max*) haplotype map (GmHapMap): a universal resource for soybean translational and functional genomics. *Plant Biotechnol. J.* 19, 324–334
121. Fang, C. *et al.* (2017) Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* 18, 1–14

122. Zhou, Z. *et al.* (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* 33, 408–414
123. Wu, D. *et al.* (2019) Whole-genome resequencing of a worldwide collection of rapeseed accessions reveals the genetic basis of ecotype divergence. *Mol. Plant* 12, 30–43
124. Ma, Z. *et al.* (2018) Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat. Genet.* 50, 803–813
125. Fang, L. *et al.* (2017) Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat. Genet.* 49, 1089
126. Hübner, S. *et al.* (2019) Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nat. Plants* 5, 54–62
127. Lin, T. *et al.* (2014) Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* 46, 1220–1226
128. Liang, Z. *et al.* (2019) Whole-genome resequencing of 472 *Vitis* accessions for grapevine diversity and demographic history analyses. *Nat. Commun.* 10, 1–12
129. Wang, X. *et al.* (2017) Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nat. Genet.* 49, 765–772
130. Varshney, R.K. *et al.* (2019) Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nat. Genet.* 51, 857–864
131. Varshney, R.K. *et al.* (2017) Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nat. Genet.* 49, 1082
132. Sundaram, R.M. *et al.* (2008) Marker assisted introgression of bacterial blight resistance in Samba Mahsuri, an elite indica rice variety. *Euphytica* 160, 411–422
133. Madhavi, K.R. *et al.* (2016) Marker assisted introgression of blast (Pi-2 and Pi-54) genes in to the genetic background of elite, bacterial blight resistant indica rice variety, improved Samba Mahsuri. *Euphytica* 212, 331–342
134. Yugander, A. *et al.* (2018) Incorporation of the novel bacterial blight resistance gene *Xa38* into the genetic background of elite rice variety Improved Samba Mahsuri. *PLoS One* 13, e0198260
135. Khanna, A. *et al.* (2015) Development and evaluation of near-isogenic lines for major blast resistance gene(s) in Basmati rice. *Theor. Appl. Genet.* 128, 1243–1259
136. Ellur, R.K. *et al.* (2016) Improvement of Basmati rice varieties for resistance to blast and bacterial blight diseases using marker assisted backcross breeding. *Plant Sci.* 242, 330–341
137. Kuraparthy, V. *et al.* (2009) Development of a PCR assay and marker-assisted transfer of leaf rust and stripe rust resistance genes *Lr57* and *Yr40* into hard red winter wheats. *Crop Sci.* 49, 120–126
138. Vasistha, N. *et al.* (2017) Molecular introgression of leaf rust resistance gene *Lr34* validates enhanced effect on resistance to spot blotch in spring wheat. *Euphytica* 213, 262
139. Rai, K.N. *et al.* (2008) Adaptation and quality traits of a germplasm-derived commercial seed parent of pearl millet. *Plant Genet. Resour. Newsl.* 154, 20–24
140. Arelli, P. *et al.* (2006) Registration of high yielding and multiple disease resistant germplasm JTN5503. *Crop Sci.* 46, 2723–2724
141. Arelli, P. *et al.* (2017) Registration of conventional soybean germplasm JTN-4307 with resistance to nematodes and fungal diseases. *J. Plant Regist.* 11, 192–199
142. Arelli, P.R. and Young, L.D. (2009) Inheritance of resistance in soybean PI 567516C to LY1 nematode population infecting cv. Hartwig. *Euphytica* 165, 1–4
143. Smith, J. (2010) *USDA, ARS, National Genetic Resources Program*, Germplasm Resources Information Network
144. Varshney, R.K. *et al.* (2014) Marker-assisted introgression of a QTL region to improve rust resistance in three elite and popular varieties of peanut (*Arachis hypogaea* L.). *Theor. Appl. Genet.* 127, 1771–1781
145. Varshney, R.K. *et al.* (2014) Marker-assisted backcrossing to introgress resistance to *Fusarium* wilt race 1 and *Ascochyta* blight in C 214, an elite cultivar of chickpea. *Plant Genome* 7, 1–11
146. Neeraja, C. *et al.* (2007) A marker-assisted backcross approach for developing submergence-tolerant rice cultivars. *Theor. Appl. Genet.* 115, 767–776
147. Lang, N.T. *et al.* (2011) Marker-assisted backcrossing (MAB) for rice submergence tolerance in Mekong delta. *Omonrice* 18, 11–21
148. Hasan, M.M. *et al.* (2015) Marker-assisted backcrossing: a useful method for rice improvement. *Biotechnol. Biotechnol. Equip.* 29, 237–254
149. Waziri, A. *et al.* (2016) Saltol QTL and their role in salinity tolerance in rice. *Austin. J. Biotechnol. Bioeng.* 3, 1–5
150. Das, G. *et al.* (2018) Improved Tapaswini having four BB resistance gene pyramided with six genes/QTLs, resistance/tolerance to biotic and abiotic stresses in rice. *Sci. Rep.* 8, 2413
151. Dixit, S. *et al.* (2017) Marker-assisted breeding to develop the drought-tolerant version of Sabitri, a popular variety from Nepal. *Euphytica* 213, 184
152. Mitrofanova, O.P. and Khakimova, A.G. (2017) New genetic resources in wheat breeding for increased grain protein content. *Russ. J. Genet. Appl. Res.* 7, 477–487
153. Yi, M. *et al.* (2009) Marker assisted backcross breeding to improve cooking quality traits in Myanmar rice cultivar Manawthukha. *Field Crops Res.* 113, 178–186
154. Chu, Y. *et al.* (2011) Marker-assisted selection to pyramid nematode resistance and the high oleic trait in peanut. *Plant Genome* 4, 110–117
155. Janila, P. *et al.* (2016) Molecular breeding for introgression of fatty acid desaturase mutant alleles (*ahFAD2A* and *ahFAD2B*) enhances oil quality in high and low oil containing peanut genotypes. *Plant Sci.* 242, 203–213
156. Yaduru, S. *et al.* (2020) Improvement of three popular Indian groundnut varieties for foliar disease resistance and high oleic acid using SSR markers and SNP array in marker-assisted backcrossing. *Crop J.* 8, 1–15
157. Crossa, J. *et al.* (2016) Genomic prediction of gene bank wheat landraces. *G3: Genes, Genomes, Genetics* 6, 1819–1834
158. Cowling, W.A. *et al.* (2017) Evolving gene banks: improving diverse populations of crop and exotic germplasm with optimal contribution selection. *J. Exp. Bot.* 68, 1927–1939
159. Veenstra, L.D. *et al.* (2020) Recurrent genomic selection for wheat grain fructans. *Crop Sci.* 60, 1499–1512
160. Araus, J.L. *et al.* (2018) Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci.* 23, 451–466
161. Jin, X. *et al.* (2020) High-throughput estimation of crop traits: a review of ground and aerial phenotyping platforms. *IEEE Trans. Geosci. Remote Sens.* 9, 200–231
162. Roitsch, T. *et al.* (2019) Review: new sensors and data-driven approaches—a path to next generation phenomics. *Plant Sci.* 282, 2–10
163. Ramstein, G.P. *et al.* (2018) Breaking the curse of dimensionality to identify causal variants in breeding. *Theor. Appl. Genet.* 132, 559–567
164. Kremling, K.A.G. *et al.* (2019) Transcriptome-wide association supplements genome-wide association in *Zea mays*. *G3 (Bethesda)* 9, 3023–3033
165. Weckwerth, W. *et al.* (2020) PANOMICS meets germplasm. *Plant Biotechnol. J.* 18, 1507–1525
166. Pazhamala, L. *et al.* (2021) Systems biology for crop improvement. *Plant Genome* http://doi.org/10.1002/tpg2.20098
167. Flint, J. *et al.* (2014) Assessing the utility of intermediate phenotypes for genetic mapping of psychiatric disease. *Trends Neurosci.* 37, 733–741
168. Donald, C.M. (1968) The breeding of crop ideotypes. *Euphytica* 17, 385–403
169. Fernie, A.R. and Yan, J.B. (2019) De novo domestication: an alternative route toward new crops for the future. *Mol. Plant* 12, 615–631
170. Ort, D.R. *et al.* (2015) Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc. Natl. Acad. Sci. U. S. A.* 112, 8529–8536
171. Lu, J. *et al.* (2006) The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends Genet.* 22, 126–131

172. Zsögön, A. *et al.* (2018) De novo domestication of wild tomato using genome editing. *Nat. Biotechnol.* 36, 1211–1216
173. Li, T. *et al.* (2018) Domestication of wild tomato is accelerated by genome editing. *Nat. Biotechnol.* 36, 1160–1163
174. Rodríguez-Leal, D. *et al.* (2017) Engineering quantitative trait variation for crop improvement by genome editing. *Cell* 171, 470–480
175. Massawe, F. *et al.* (2016) Crop diversity: an unexploited treasure trove for food security. *Trends Plant Sci.* 21, 365–368
176. Lemmon, Z.H. *et al.* (2018) Rapid improvement of domestication traits in an orphan crop by genome editing. *Nat. Plants* 4, 766–770