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. Sorrells7

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.

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

Table 1. Genome Resources in Ten Topmost Food Crops

<table>
<thead>
<tr>
<th>Crop</th>
<th>Area (mha)</th>
<th>Production (mmt)</th>
<th>Assembled genomes (Mb)</th>
<th>SNP array</th>
<th>Genomic variation databases</th>
<th>Gene expression atlas</th>
<th>Pan-genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maize (Zea mays)</td>
<td>197.2</td>
<td>1148.4</td>
<td>2048 [69]</td>
<td>MaizeSNP50 BeadChip (Illumina Infinium 50K) [70]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Subset of MaizeSNP50 (Illumina Infinium 3K) [71]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rice (Oryza sativa)</td>
<td>162</td>
<td>755.4</td>
<td>371 [7]</td>
<td>Affymetrix (1M) [77]</td>
<td>SNP-Seek [82]</td>
<td>[83,84]</td>
<td>66 Accessions [85]</td>
</tr>
<tr>
<td>Soybean (Glycine max)</td>
<td>120.5</td>
<td>333.7</td>
<td>973 [86]</td>
<td>SoySNP50K [87] SoySNP180K Axiom [88]</td>
<td></td>
<td>[89]</td>
<td>26 Accessions [91]</td>
</tr>
<tr>
<td>Barley (Hordeum vulgare)</td>
<td>51.1</td>
<td>158.9</td>
<td>4980 [92]; 4790 [93]</td>
<td>9K Illumina Infinium Select Custom Genotyping BeadChip [94]</td>
<td>BarleyVarDB [96]</td>
<td>21 439 Genes [97]</td>
<td>20 Accessions [98]</td>
</tr>
<tr>
<td>Sorghum (Sorghum bicolor)</td>
<td>40</td>
<td>57.9</td>
<td>739 [99]</td>
<td>3K SNP Infinium array [100]</td>
<td>SorGSD [101]</td>
<td>27 577 Genes [102]</td>
<td>n.d.</td>
</tr>
<tr>
<td>Rapeseed (Brassica napus)</td>
<td>34</td>
<td>70.5</td>
<td>849.7 [103]</td>
<td>International Brassica SNP Consortium (60K) [104]</td>
<td>BnaGVD [105]</td>
<td>101 040 Genes [106]</td>
<td>8 Accessions [107]</td>
</tr>
<tr>
<td>Sugarcane (Saccharum officinarum)</td>
<td>26.7</td>
<td>1949.3</td>
<td>800–900 (Monoploid)</td>
<td>76K SNPs [115] 84K SNPs [116]</td>
<td></td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
</tbody>
</table>

*Abbreviation: n.d., no data.

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.html). Examples include improved versions of hard red winter wheat (HRWW) cultivars ‘Jagger’ and ‘Overly’ carrying genes Yr40/Lr57 and Lr58, respectively [137] and spring wheat cultivar ‘HUL510’ 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 Ph1, the recessive resistance genes rym4/rym5 against barley yellow mosaic viruses, and mlb for barley powdery mildew (Blumeria graminis f. sp. hordei).

Unlike cereals, GAB in grain legume collections 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 (‘JTNT 5503’, ‘JTNT 5303’, ‘DS 880’, and ‘JTNT 5109’ [140–143]). Similarly, Vansheney et al. [144] obtained a set of 20 introgression lines in groundnut (Arachis hypogaea) showing higher yield and increased rust (Fusarium arachidicola) 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 with (Fusarium oxysporum f. sp. ciceri) 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 tolerance submergence (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 OM1-490/Rf64-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, ‘Thaddockham 1’ (TDK1) from Laos, and ‘IBR 11’ from Bangladesh were also evident following the QTL-introgression (see Hasan et al. [148]).

The Saltol QTL was introgressed into various rice genomes in different countries, and the candidate varieties targeted for QTL-introgression were ‘Pusa Basmati 1121’, ‘Pusa Basmati 6’, ‘AS 996’, ‘BT 7’, ‘Bachthon 7’, ‘QSDS’, and ‘BRRI-Dhan 49’ (see Waziri et al. [149]). Successful pyramiding of Sub1, Saltol, blast (P2, P8), and gall midge (Orseola 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 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’, ‘Gladias’, ‘VR 1128’) (see Mitrofanova and Khakimova [152] and references therein). The transfer of variant alleles of baxh2 and Wx from basmati into ‘Manawthukham’ (an elite rice cultivar of Myanmar) resulted in improved fragrant and intermediate amylose content [153]. By reducing the breeding cycles up to 3 years, Chui et al. [154] developed ‘Tiguard High O/L’ groundnut with high oleic acid content and nematode resistance. More recently, GAB varieties of groundnut having improved quality combined with resistance to rust and late leaf spot (Phaeosphaeria personata Berk. & Curtis) have been released for commercial cultivation [155,156].
Trends in Plant Science

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.cglab.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].
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 (RGSV 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,

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

*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 [33]. 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 maladaptation 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...
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-phen 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 (a), genetic variation (v), and breeding cycle time (L)] reflect its relevance in context to improving genetic gains in a breeding program [162].

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 launched 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.
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 CO₂ 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 ‘hypothesis 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 germplasm, 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 CYCLASE) with CRISPR-Cas9 technology, Zsógón et al. [172] improved wild tomato ancestor Solanum pimpinellifolium with respect to fruit number (10X), size (5X), 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, SCLV3, SWUS, SIGGPI), 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, Rodriguez-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
8. Zhou, Y. et al. (2020) A platinum standard pan-genome resource that represents the population structure of Asian rice. Sci. Data 7, 113

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/mediate the genotype-dependence of current protocols to facilitate their widespread applications?
trends of crop domestication. Nature 543, 711–716


121. Liu, Y. et al. (2005) Pan-genome of wild and cultivated groundnut to accelerate genetics and breeding in groundnut. Crop Sci. 95, 289–299


127. Yang, X. et al. (2017) Mining sequence variations in representative polyploid sugarcane germplasm accessions. BMC Genom. 18, 594

128. 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 Genom. 18, 72


133. Fang, C. et al. (2017) Genome-wide association studies dissect the genetic networks underlying agronomic traits in soybean. Genome Biol. 18, 1–14
153. Madhavi, K.R. et al. (2016) Marker assisted introgression of blast (P2-P and P1-P4) genes in to the genetic background of elite, bacterial blight resistant indica rice variety, improved Samba Mahsuri. Euphytica 212, 331-343
160. Areli, P. et al. (2006) Registration of high yielding and multiple disease resistant germplasm JTN5003. Crop Sci. 46, 2723-2724
167. Lang, N.T. et al. (2011) Marker-assisted backcrossing (MAB) for rice submergence tolerance in Mekong delta. Oomocia 18, 11-21
171. Dixit, S. et al. (2017) Marker-assisted breeding to develop the drought-tolerant version of Sabitri, a popular variety from Nepal. Euphytica 213, 184
188. Donald, C.M. (1968) The breeding of crop ideotypes. Euphytica 17, 365-403