

Feature Review

Reap the crop wild relatives for breeding future crops

Abhishek Bohra ¹, Benjamin Kilian ², Shoba Sivasankar,³ Mario Caccamo ⁴, Chikelu Mba,⁵ Susan R. McCouch ^{6,*} and Rajeev K. Varshney ^{7,8,*}

Crop wild relatives (CWRs) have provided breeders with several 'game-changing' traits or genes that have boosted crop resilience and global agricultural production. Advances in breeding and genomics have accelerated the identification of valuable CWRs for use in crop improvement. The enhanced genetic diversity of breeding pools carrying optimum combinations of favorable alleles for targeted crop-growing regions is crucial to sustain genetic gain. In parallel, growing sequence information on wild genomes in combination with precise gene-editing tools provide a fast-track route to transform CWRs into ideal future crops. Data-informed germplasm collection and management strategies together with adequate policy support will be equally important to improve access to CWRs and their sustainable use to meet food and nutrition security targets.

Using wild relatives to improve crops

Innovations in plant breeding and agronomy have been instrumental in achieving continuous growth in global food production for almost a century. However, feeding 34% more people and meeting their nutritional needs over the next 30 years will require an annual increase of 44 million tons of food, representing a 37% increase over the current production rate of 32 million metric tons per year [1]. Achieving this feat in a 2°C warmer world amid stiff competition for water, energy, and land presents a formidable challenge. Importantly, in the developing world, 80% of the yield increase must come from improvements in crop productivity and cropping intensity, given that only 20% can come from expanded use of arable land (www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf). In addition to producing more food, efforts will also be necessary to address the substantial food inequalities that are rife in current global food systems.

Domestication of wild species of food crops started nearly 10 000 years ago in the Middle East and Fertile Crescent. Subsequent genetic improvement led to the development of modern high-yielding crop varieties, many with resistance to pests and diseases as well as to abiotic forms of stress [2]. However, emphasis on delivering a steady stream of familiar, genetically uniform cultivars has focused breeding efforts on a small number of species and varieties at the expense of the wide range of locally adapted and genetically diverse traditional crops that were once cultivated. The FAO estimates that ~75% of the genetic diversity harbored in traditional agricultural crop varieties has been lost over the past century (www.cropwildrelatives.org/cwr/threats/). As opined by Abbo and coworkers [3], this important loss of diversity, caused by migration of crops from their centers of origin or from modern breeding, should be called post-domestication or breeding bottlenecks.

Wild ancestors and traditional **landraces** (see [Glossary](#)) are reservoirs of valuable traits, including diverse forms of resistance to both biotic and abiotic stresses ([Box 1](#)), which remain crucial for adaptation of modern cultivars to current and future climates. **CWRs** have been used for decades in crop improvement for enhancing plant performance. The annual contribution of CWRs

Highlights

Exotic genetic libraries in different crops are valuable genetic resources for genetic dissection of complex quantitative traits.

An informed choice of crop wild relatives (CWRs) for genetic studies and breeding can be made by taking account of the environmental variables of the collection sites.

New breeding tools such as genomic selection and optimum contribution selection help to achieve the optimal combinations of beneficial alleles in exotic × elite crosses.

Precise gene-editing tools open new avenues to broaden the array of current food crops by domesticating wild species *de novo*.

Regulating the known crossover suppressors through mutagenesis and ploidy-level change has great potential to disrupt linkage drag.

Systematic analysis of genebank collections would guide future germplasm collection strategies by prioritizing both target species and global sites.

¹ICAR–Indian Institute of Pulses Research (IIPR), 208024 Kanpur, India

²Global Crop Diversity Trust, 53113 Bonn, Germany

³International Atomic Energy Agency (IAEA), Vienna International Centre, 1400 Vienna, Austria

⁴NIAB, Cambridge CB3 0LE, UK

⁵Food and Agriculture Organization of the United Nations (FAO), Rome 00153, Italy

⁶Plant Breeding and Genetics, School of Integrative Plant Science, Cornell University, Ithaca, NY 14850, USA



Box 1.. Contributions of CWRs to enhance agricultural production

The importance of CWRs in bringing “game changing” characteristics to the breeding pool through wide hybridization is well recognized in plant breeding. Pioneering work with CWRs was performed in sugarcane in the early decades of the 20th century. Through interspecific hybridization, referred to as “**nobilization of canes**”, Indian sugarcane cultivars were imbued with high sugar content along with resistance to various biotic and abiotic stresses and wider adaptability. The hybridization between *Saccharum officinarum* (Noble cane, $2n=80$) and *S. spontaneum* ($2n=40-128$) resulted in a popular variety, Co 205, for the subtropical regions. Further expansion of this work spawned a series of high yielding cultivars that played a key role in revolutionizing the sugarcane industry in India and the world alike. Other notable cases that exemplify the contribution of CWRs to modern breeding have been reported in various crop species including wheat, rice, potato, tomato, sunflower, common bean etc.

A major proportion (c. 80%) of traits transferred from CWRs into cultivated crop varieties include resistance to diseases and pests [123]. Incorporating resistance to late blight [*Phytophthora infestans* (Mont.) de Bary] in potato represents a notable contribution to humankind as the Irish potato famine, caused by an extended epidemic of late blight reduced Ireland’s population by nearly 1.8 million over 7 years (1844-1851). Similarly, resistance to stem rust (caused by *Puccinia graminis*) introgressed from the wild wheat, *Aegilops tauschii*, played a significant role in enhancing wheat productivity as part of the “Green revolution” worldwide. A landmark rice variety, IR 36, incorporated resistance to the grassy stunt virus (GSV) derived from the wild rice, *Oryza nivara*, obtained after screening 7000 wild accessions for resistance to GSV [94]. *Oryza nivara* still serves as the sole source of GSV resistance in rice. Likewise, several rice varieties host a gene for bacterial blight resistance (*Xa21*) derived from the wild perennial rice, *Oryza longistaminata*.

Introgression breeding gained prominence in tomato thanks to the discovery of DNA marker technology. Several traits including pest resistance, productivity traits, higher nutritional value, and salt stress tolerance have been introduced from wild *Solanum pimpinellifolium* to the cultivated tomato. Like wild species, landraces also contain valuable traits that are absent from modern cultivated gene pools. For instance, submergence tolerance contributed by the *SUB1A* gene in modern rice varieties originated in an Indian landrace of *O. sativa*, FR13A. Similarly, the *mlo* gene from a barley landrace confers broad-spectrum resistance against powdery mildew in many modern barley varieties. Numerous genes conferring beneficial traits have been identified in CWR and traditional landraces for utilization in modern breeding programs of different crop species (Table S1). Recent advances in plant biology have facilitated the rapid discovery and utilization of valuable CWR genes for crop improvement which are helping to meet the global demand for nutritious food in a changing environment.

to the world economy is estimated to be ~\$186.3 billion [4]. For example, in a single crop species, genes from a wild tomato increased the content of soluble solids with a worth of ~\$250 million (www.cropwildrelatives.org/resources/in-situ-conservation-manual/).

Incorporation of genes from CWRs into the cultivated gene pool is not straightforward. Breeders are reluctant to use CWRs in commercial breeding programs owing to the challenges presented by crossability barriers, **linkage drag**, poor agronomic performance, and a wide range of phenotyping challenges. Advances in plant molecular biology have introduced new opportunities for overcoming several of these problems. For instance, genes can now be ‘edited’ *in situ* such that long-lost ancestral alleles known to confer desirable traits or phenotypes can be reintroduced into modern, elite cultivars without disrupting the constellation of genes in the genetic background that confer valuable characteristics essential to productivity in modern agricultural systems. In this review we briefly discuss the challenges of introgression breeding, and we present strategies to accelerate the discovery of valuable genes from CWRs and their efficient deployment in breeding programs. We also provide a perspective on the possible role of recombination-boosting methods for improving the efficiency of CWR introgression. Finally, we highlight the urgent need for systematic analysis of gaps in germplasm collections and for setting priorities for the future collection and systematic evaluation of genebank holdings.

Crucial challenges in expanding cultivated gene pools using CWRs and landraces

Key issues faced by breeders working to introgress new trait variation from wild or traditional germplasm into elite modern cultivars include biological barriers to compatibility and crossability, **F₁ generation** and **backcross** (BC1) sterility, infertility of offspring, and reduced recombination

⁷Centre of Excellence in Genomics and Systems Biology, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad 502324, India

⁸State Agricultural Biotechnology Centre, Centre for Crop and Food Innovation, Murdoch University, Murdoch, WA 6150, Australia

*Correspondence: srm4@cornell.edu (S.R. McCouch) and rajeev.varshney@murdoch.edu.au (R.K. Varshney).

between elite and CWR genomes [5]. Careful consideration of these challenges has opened the door to novel opportunities for managing male and female sterility in the development of hybrid crops. For example, male sterility caused by the disharmony of nuclear (cultivated) and cytoplasmic (wild) genomes in interspecific crosses of diverse crop species has proved to be a boon for the hybrid industry worldwide (Box 2).

In contrast to an elite breeding pool bred for performance across a range of environments in response to managed inputs, CWRs generally show poor adaptation beyond their natural distribution range [6]. The expression of valuable alleles may be masked when wild materials are grown outside their natural zone(s) of adaptation. Thus, the performance of wild accessions *per se*, and/or their derived interspecific progeny, might not appear to be phenotypically promising in standard breeding trials, and this can discourage further use of CWRs as a source of genetic variation in a breeding program [7]. Photoperiod sensitivity, asynchronous flowering, and phenological differences can all contribute to **maladaptation** to the artificial agricultural environments. For example, chickpea CWRs collected in temperate regions are ill-adapted to tropical or subtropical regions owing to large phenological differences [8]. However, the perception of agronomic potential may be deceptive because an agronomically inferior CWR (i.e., one with small seeds that shatter) may contain valuable alleles for a particular trait(s) (i.e., for disease resistance or abiotic stress tolerance). Using appropriate evaluation procedures, these beneficial alleles can be readily discovered in segregating populations derived from wild × elite crosses [9].

Strong linkage associations between favorable and unfavorable alleles (i.e., linkage drag) negatively influence the phenotypic assessment of progenies resulting from interspecific crosses and are a key deterrent to the use of CWRs in crop improvement. Linkage drag resulting from the use of CWRs has been well documented in the literature for numerous species and traits. Examples include yield penalties associated with the introgression of the resistance genes *Pch1* and *Pm16* from *Aegilops ventricosa* and *Aegilops speltoides*, respectively, which confer resistance to eyespot and powdery mildew in common wheat [10], or undesirable horticultural traits in tomato associated with *Phytophthora infestans* (late blight) resistance introduced from *Lycopersicon hirsutum* [11]. In these cases, marker technologies in combination with backcrossing have proven to be effective in eliminating linkage drag. More recently, Wang and coworkers [12] examined interspecific progenies of rice derived from crosses between *Oryza glaberrima* and *Oryza sativa*, and observed an association between the hybrid sterility locus,

Box 2. Male sterility trait to boost hybrid technology

Exploitation of heterosis or hybrid vigor is a key component of sustainable plant breeding. Manual emasculation presents the biggest challenge in hybrid seed production. To this end, male sterility systems in different crop species enable efficient hybrid seed production via the production of non-functional pollen grains. Compared to mutagenesis or genetic engineering, interspecific hybridization involving CWRs is the most widely used approach to induce **cytoplasmic male sterility (CMS)**, the most prominent among sterility systems in plants [124]. Hybrid breeding in sunflower offers an outstanding case of extensive utilization of CMS sources from CWRs, notable examples being *PET1* (*Helianthus petiolaris*), *GIG1* and *GIG2* (*Helianthus giganteus*), and *PEF1* (*H. petiolaris* ssp. *fallax*). Since its discovery in 1969, CMS-PET remains the chief source of male sterility in hybrid sunflower breeding programs. In rice, the predominant Wild Abortive (WA)-CMS was developed from weedy rice (*Oryza sativa* f. *spontanea*). Alternative CMS systems including HL, CW, FA, RT102A, and RT98 A were discovered in wild rice (*Oryza rufipogon*). The occurrence of CMS in interspecific crosses of wheat led to the development of T- and K-type cytoplasm from *Triticum timopheevii* and *Aegilops kotschyii*, respectively, which created novel possibilities for harnessing hybrid vigor. Similarly, hybrid breeding in sorghum and pearl millet gained momentum with the discovery of the Milo (A1) and Tift23A (A1) cytoplasm from landraces. Among legumes, CMS systems were discovered from interspecific crosses involving wild soybean (*Glycine soja*) and wild pigeonpea (*Cajanus cajanifolius* and *Cajanus scarabaeoides*). As evident from the susceptibility of maize T-CMS to northern corn leaf blight (caused by *Bipolaris maydis*) and pearl millet A1-CMS to downy mildew (caused by *Sclerospora graminicola*), hybrid breeding relying on a single CMS source poses a great risk for the occurrence of epidemics. This calls for diversifying the CMS sources by capitalizing on the vast array of distant CWRs that are available in genebank repositories.

Glossary

Advanced backcross nested association mapping (AB-NAM): a design that combines an advanced backcross quantitative trait locus (AB-QTL) method with a nested association mapping (NAM) population approach. AB-NAM involves crossing diverse wild accessions to a common popular cultivar to capture the maximum diversity of a CWR with adequate scope for minimizing the problem of CWR maladaptation.

Backcross: crossing an F₁ hybrid individual with one of the parents generates the first backcross (BC₁). The parent used repeatedly in the backcross scheme is known as the recurrent or recipient parent, whereas the other parent serves as the donor. Individuals from BC₁ are then crossed to the recurrent parent to create the second backcross (BC₂) population. Each backcross reduces the proportion of donor alleles by half.

Composite collection: a collection of germplasm selected to represent the entire genetic diversity of the species comprised of accessions from a mini-core, representatives from core collections, wild species, landraces, released cultivars, and donors of various biotic/abiotic stresses and agronomic traits.

Core collection: subset of a large germplasm collection that represents ~10% of the entire germplasm collection but captures >70% of the genetic variation.

Crop wild relatives (CWRs): related species with a different gene pool that display resilience due to natural selection and have many diverse beneficial traits that can contribute to producing a commercially viable cultivar.

Cytoplasmic male sterility: the inability of a plant to produce functional pollen as a result of impaired harmony between the cytoplasmic and nuclear genomes.

De novo domestication: the introduction of desirable genes directly into undomesticated wild ancestors using introgression breeding or genome editing.

De novo genome assembly: the construction of a new genome sequence by assembling short or long nucleotide sequence reads in the absence of a reference sequence.

Exotic genetic libraries: collections of backcross-derived introgression lines

S20, and genes conferring long sterile lemma (*G1-g*) and wide grains (*qGW7*), and proposed that linkage drag was accentuated owing to the gamete-eliminator mechanism of the S20 sterility locus, making it difficult for breeders to use recombination as a way to break linkage drag. In these cases, gamete elimination combined with reduced recombination in the vicinity of introgressed genomic segments greatly limited the possibility of identifying favorable recombinants.

Accelerating access to CWR diversity for breeding applications

Worldwide, a total of ~7.4 million accessions are archived in 1750 genebanks, and 130 gene banks each hold >10 000 accessions. Eleven genebanks of the Consultative Group for International Agricultural Research (CGIAR) network maintain 741 319 accessions of 3446 species of 612 different genera (www.fao.org/3/i1500e/i1500e03.pdf). Additional populations of wild species and traditional or landrace varieties are maintained *in situ* in wild or protected environments and/or in farmers' fields located in or near centers of diversity throughout the world. The benefit of *in situ* conservation is that they are genetically dynamic and continue to evolve in response to both natural and artificial selection, thereby enhancing their adaptation to the environments in which they are grown. However, *in situ* collections are vulnerable to habitat destruction and/or encroachment caused by civil strife, human settlement pressure, and natural disasters including wildfires, flooding, drought, and volcanic eruptions. *Ex situ* collections maintained in botanical gardens and genebanks complement *in situ* conservation efforts, but it can be challenging to maintain and propagate CWRs and traditional varieties outside their zones of adaptation, leading to some erosion of genetic variation over time. Regardless, identifying plant genetic resources (PGRs) that carry specific traits in large *in situ* or *ex situ* collections represents the proverbial problem of searching for a needle in a haystack. In this section we discuss strategies to efficiently narrow the search space when trying to identify accessions that carry traits or genes of interest.

Combining eco-geographic approaches and machine learning for gene discovery

To facilitate the investigation of large germplasm collections, including those containing CWRs, it is reasonable to begin by examining customized sets of germplasm that embody appropriate diversity and are of manageable size, such as **core collections**, **mini-core collections**, and **composite collections** [13,14]. However, the utility of using customized germplasm sets based on diversity to identify trait-specific genetic resources for breeding remains debatable. A slightly more targeted approach, referred to as the **focused identification of germplasm strategy (FIGS)** uses environmental data associated with germplasm collection sites to quantify trait–environment relationships in an effort to provide *a priori* information about 'best-bet germplasm' that has a high probability of carrying specific adaptive traits [15]. This eco-geographic approach is based on the premise that the adaptive evolution of plant traits is an outcome of natural selection operating on the diversity of genetic resources (i.e., on wild populations and early landraces) [16].

The utility of FIGS was best evidenced by a large-scale allele-mining experiment that identified the powdery mildew resistance gene *Pm3* in bread wheat based on examination of a subset of 1320 landraces. This set of landraces represented a FIGS-educated customized set selected from 16 089 accessions in the genebank [17]. Similarly, a strong relationship between the geographic distribution of resistance genes and environmental variables at collection sites enabled accurate prediction of novel sources of resistance against stem rust [15] and stripe rust [18] from wheat landraces stored in the genebank at the International Center for Agricultural Research in the Dry Areas (ICARDA), as well as for stripe rust from global spring wheat panel comprising landraces and improved accessions [19]. FIGS has facilitated the discovery of several genetic resources with valuable biotic stress resistance traits in wheat and barley [20,21], and abiotic

wherein each line contains a precisely defined segment of donor genome, and collectively these lines represent the entire donor genome.

Ex situ conservation: preservation of plant genetic resources in genebank facilities away from their natural habitats.

F₁ generation: the first filial generation, obtained by crossing two plants or animals with contrasting target traits.

F₂ generation: the second filial generation, created by random mating of the two F₁ individuals or self-pollinating of an F₁ individual.

Fast generation cycling system

(FGCS): the rapid development of homozygous lines by combining embryo culture with plant management in a breeding cycle while providing adequate scope for recombination.

Focused identification of

germplasm strategy (FIGS): the creation of manageable subsets of germplasm resources by using climate data for the collection sites for efficient mining of alleles controlling variation in target traits.

Genomic selection (GS): assessing the genetic worth of unobserved individuals from genotypic data alone by using prediction models trained on individuals having both genotypic and phenotypic records.

In situ conservation: the conservation of germplasms in their native habitats. *In situ* conservation is deemed more suitable for harnessing genetic diversity that is of potential use in improving crop adaptation to changing environmental conditions.

Landrace: a primitive heterogeneous genotype selected by farmers for its ability to adapt to local conditions and provide moderate yield in low-input systems.

Linkage drag: unintended coinheritance of undesirable CWR alleles with desirable alleles targeted for introduction into the elite pool.

Machine learning (ML): a branch of artificial intelligence that relies on learning algorithms, instead of being explicitly programmed, and that can find patterns and make decisions through training on large-scale data.

Maladaptation: failure to adapt or to show phenotypic plasticity in the wider context of environmental changes.

Mini-core collection: reducing the size of the core collection (10% of core) based on its evaluation for different traits suitable for enhanced germplasm use.

stress adaptation in faba bean [22] and soybean [23]. A more recent study in wheat applied FIGS to identify a subset of 52 accessions with potential to provide new genes for resistance to powdery mildew disease from a large collection of 19 460 accessions belonging to three taxa – *Triticum aestivum*, *Triticum durum*, and *Triticum dicoccum* [24].

Highly complex multi-dimensional and non-linear relationships between eco-geographic profiles and trait expression offer scope to apply **machine learning (ML)**-based approaches in these studies. For instance, ML algorithms such as random forest (RF), support vector machines (SVMs), and artificial neural networks (ANNs) were applied to classify germplasm sets with respect to the presence of adaptive traits [15,22]. Learning-based algorithms are apt for such studies because they circumvent the need for adherence to normality and/or linearity. Advances in ML can be used to improve the efficiency and predictive ability of the FIGS approach. To further refine the FIGS approach, Stenberg and Ortiz [16] proposed consideration of **non-adaptive evolutionary forces** (i.e., gene flow and genetic drift) in addition to historical natural selection. Application of targeted germplasm selection strategies improves the likelihood of finding sought-after adaptive traits in large germplasm collections (Figure 1). Moreover, such focused approaches become more important for accelerating the response time of breeders in the face of global climate change.

Construction of introgression libraries from CWR × elite crosses

A marked reduction in agronomic performance is typically observed in progenies derived from elite × CWR crosses, fueling the reluctance of breeders, particularly commercial breeders, to use exotic germplasm in crosses with elite lines for crop improvement. The large performance gaps that are characteristic of CWRs compared to elite lines inhibit short-term genetic gain [9], although the infusion of new variation may ultimately help breeders achieve long-term genetic gain. Using DNA marker technology in concert with quantitative trait locus (QTL) discovery methods, beneficial alleles from CWR and landraces can be efficiently introgressed into elite genetic backgrounds via pre-breeding [25]. This has been exploited by pre-commercial breeders to develop suites of **exotic genetic libraries** known as chromosome segment substitution lines (CSSLs). These CSSL libraries represent permanent genetic resources that are created by crossing and backcrossing an elite line (as the recurrent parent) with a wild or landrace accession (donor parent) followed by marker-assisted selection to maintain selected donor segments in the elite genetic background of backcrossed progenies. Selection of overlapping genomic segments from the donor allows construction of complete CSSL libraries that collectively contain the entire genome of the donor, where each individual introgression line carries only one or a few marker-defined segment(s) from the donor genome. CSSLs have been built in a variety of crops including rice, maize, barley, wheat, oat, rye, tomato, *Brassica*, cotton, groundnut, and soybean [26]. These genetic libraries offer significant advantages over other mapping populations for the genetic dissection of complex quantitative traits because their genetic structure effectively 'Mendelizes' individual QTLs and eliminates background noise. In addition, genomic loci associated with heterosis can be delineated by demonstrating the effects of heterozygosity on the phenotype in populations derived from homozygous CSSLs × inbred testers [5,27]. The availability of these community resources lays the foundation for identifying QTLs with both major and minor effects, validating QTL effects, fine mapping and cloning, and elucidating genetic interactions between multiple introgressed QTLs, between an introgressed QTL and the elite genetic background (epistasis), and/or between introgressed QTLs and the environment [gene × environment (G × E) effects]. The major limitation restricting the widespread use of these genetic libraries in crop research and breeding is the substantial time and resources required for their development. In this context, the **fast generation cycling system (FGCS)** using *in vitro* embryo culture may be used to accelerate progression of segregating generations derived from elite × CWR crosses [28]. Alternative rapid breeding solutions include **speed**

Multiparental advanced generation

inter-cross: a popular multiparent population design that involves multiple cycles of intermating of diverse founder parents followed by inbreeding to create large sets of homozygous lines representing mosaics of founder genomes.

Nobilization of canes: inspired by the work of Dutch breeders in Indonesia, interspecific hybridization between *Saccharum officinarum* (noble cane) and *S. spontaneum* (wild cane) resulted in varieties with a combination of high sugar content and resistance to a variety of biotic and abiotic stresses.

Non-adaptive evolutionary forces: evolutionary forces such as gene flow and genetic drift that oppose phenotypic plasticity and fitness of an individual in the population.

Optimum contribution selection

(OCS): a selection strategy that compromises on high-ranked individuals to maintain a balance between inbreeding and genetic diversity of the breeding program to sustain genetic gain in long term.

Pangenome: represents the entire gene repertoire of any given species. All individuals share the core genome, whereas the dispensable or variable component is only present in some individuals.

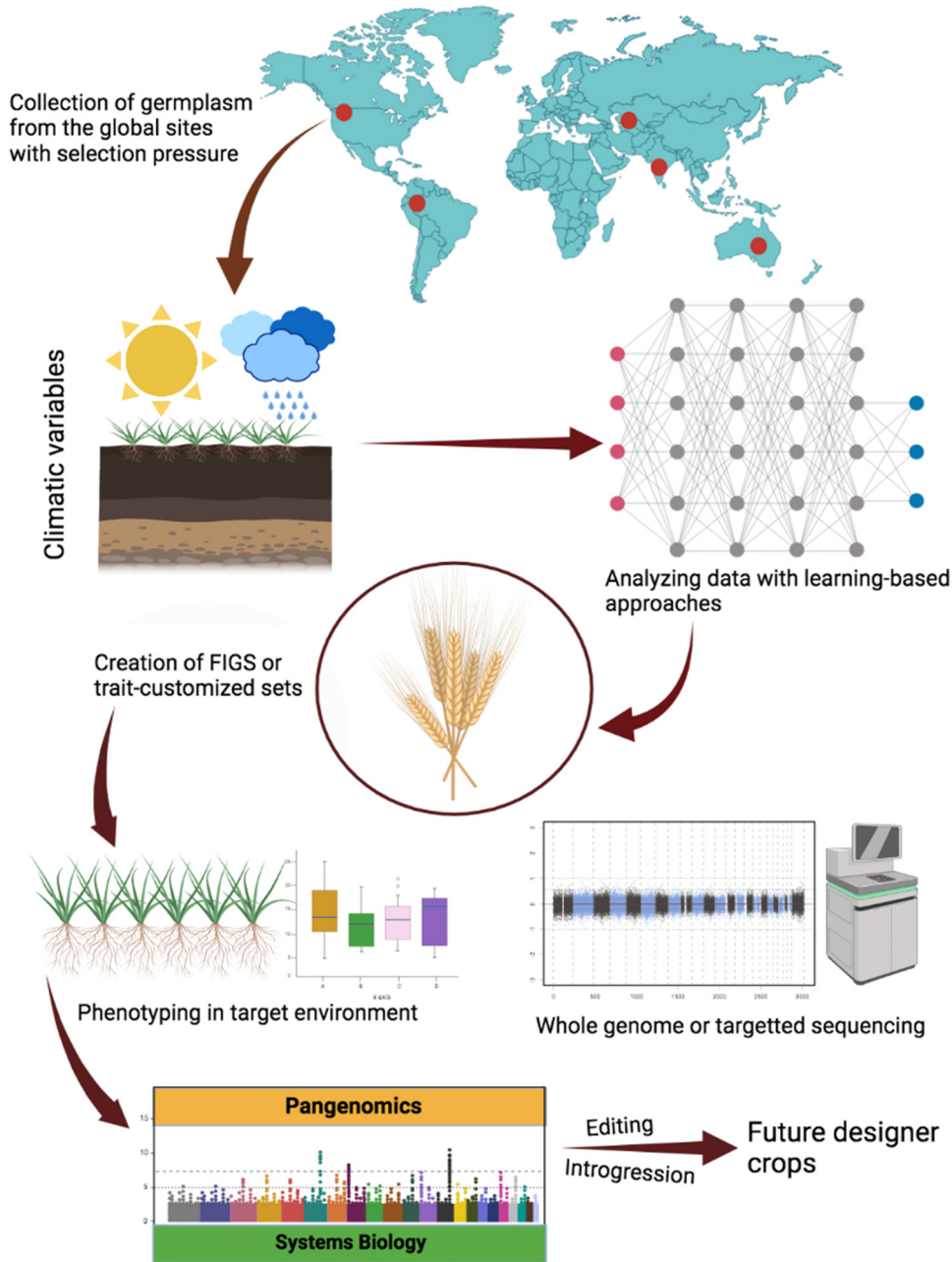
Reflective plant breeding paradigm

(RPBP): a conceptual framework that aims to establish close coordination between new breeding techniques, multidisciplinary science, and supply-value chain stakeholders by coupling plant breeding and commercialization, thus keeping the pace of germplasm development in sync with the progress of the commercial enterprise.

RenSeq: a rapid gene-cloning method that targets the NB-LRR family to discover resistance to a wide family of pathogens using a combination of NB-LRR sequence enrichment and high-throughput sequencing approaches.

Sequence-specific nucleases: enzymes that target specific genomic sites by introducing double-strand breaks (DSBs) followed by DSB repair by non-homologous end-joining (NHEJ) or homology-directed repair (HDR).

Speed breeding: a rapid breeding technique that reduces the generation time in genetic studies and breeding programs by accelerating plant growth and development. Manipulation of photoperiod length, harvesting of immature seed, and higher plant densities are



among the various means used in speed breeding to shorten generation intervals. **Structural variations (SVs):** modifications in the structure and organization of the genome by deletion, duplication, insertion, inversions, and translocations. **Transgressive improvement:** hybridization between distant parents leading to the recovery of extreme phenotypes resulting from novel combinations of superior alleles from both parents.

Figure 1. Targeted discovery of valuable genes from focused identification of germplasm strategy (FIGS)-educated crop wild relative (CWR) sets. The utilization of large collections of CWRs for genetic studies requires significant financial resources and man-hours if not carefully planned. Instead of random sampling, trait-specific CWRs may be selected from sites experiencing natural selection pressure for the trait(s) of interest. The hypothetical hot-spots or collection sites across the globe are shown as red circles in the figure. Analysis of climate datasets at these sites could help to pinpoint the accessions of value to future agriculture. Such data-informed subsets of CWRs can then be genotyped and phenotyped at scale and depth to reveal the corresponding genomic loci. Pangenomic and systems biology approaches may further increase the mapping resolution to pinpoint the causative loci for downstream applications in breeding. The image was created using BioRender (<https://biorender.com>).

Trends in Biotechnology

breeding based on manipulation of photoperiod length and temperature, which circumvents the need of cumbersome embryo culture practices. Because these genetic resources carry the genome of a single accession of a particular CWR species, this limits the amount of genetic variation. To enhance genetic diversity, CWRs may be included as founder parents in multiparent designs such as **multiparent advanced generation intercross** (MAGIC). The MAGIC lines created in rice (<https://gtr.ukri.org/projects?ref=BB%2FJ011754%2F1>) and tomato [29] using CWRs highlight the wide range of genetic variability that is available for analyzing complex plant traits of agronomic significance.

Sequencing wild genomes and pangenomics

Advances in sequencing technologies and analysis tools have facilitated the generation of high-quality reference genomes in many crops. These reference genome sequences have, in turn, catalyzed numerous downstream applications such as the development of genome-wide maps of genetic variants, the identification of gene–trait associations, and the cloning and characterization of genes of interest. However, the cost of sequencing has decreased and throughput and accuracy have increased, and the idea that a single reference genome was sufficient as a template to describe the full landscape of genetic diversity within a species was soon debunked. The growing amount of high-quality genome sequencing data for diverse accessions demonstrated that reference genome-based methods were insufficient to detect critical **structural variations (SVs)**, including presence–absence variation (PAV) and copy-number variation (CNV), that distinguish one accession from another. One implication of importance to breeders is that genomic segments introgressed from wild or exotic donors (i.e., in CSSLs) may actually carry genes that are not present in either the reference genome or the elite recurrent parent, or, conversely, exotic donors may fail to carry the genes that were targeted for introgression based on a reference genome.

De novo genome assembly and resequencing of various CWRs and landrace genomes have been reported in different crop species (Table 1) [30,31]. Examples of high-quality whole-genome assemblies of CWRs include *Triticum turgidum* [32], *Solanum pennellii* [33], *Glycine soja* [34], and 19 wild *Oryza* species [35–38]. In recent years, studies have been conducted with the aim of capturing the entire gene repertoire within a species (Table 2). Evidence from **pangenome** studies in different crops strongly supports a major role for SVs in crop domestication and breeding. For instance, pangenome analysis in rice based on *de novo* genome assemblies of 66 diverse accessions of the *O. sativa*–*Oryza rufipogon* species complex revealed extensive PAVs for genes controlling flowering time and hull color [39]. Thus, pangenome analysis offers novel opportunities to identify genetic diversity that has been lost, selected against, or only rarely brought into the domesticated gene pool during the process of crop domestication. A recent pangenome study involving 725 tomato accessions revealed a rare allele for the tomato lipoxygenase gene, *TomLoxC*, namely a ~4 kb substitution in the promoter region [40]. This study suggested that negative selection for the allele during domestication was accompanied by reintroduction into modern cultivars, possibly through wild introgression during recent breeding history. The study reported a total of 4873 genes that are absent from the reference genome 'Heinz 1706'. Similarly, pangenome analysis of 1961 cotton accessions uncovered 32 569 and 8851 genes that are not present in the reference genomes of *Gossypium hirsutum* (TM-1) and *Gossypium barbadense* (3–79), respectively [41]. These genes are described as belonging to the 'dispensable' portion of the genome of a species because they can be either present or absent without compromising the viability of the organism. As summarized by Zhao and coworkers [39] and Stein and coworkers [36] for rice, many genes in the 'dispensable rice genome' are associated with biotic and abiotic stress response, making them particularly interesting as plant breeders search for ways to enhance climate resilience in new crop varieties.

Table 1. Recent genome assemblies of wild species in different crops

Crop	Name of the wild species	Size of the assembly (Mb)	Scaffold/Contig N50 (Kb)	Number of genes	Refs
Barley	<i>Hordeum spontaneum</i>	4280	725	36 395	[100]
Rice	<i>Oryza brachyantha</i>	261	1612	32 038	[101]
	<i>O. eichingeri</i>	471	64	31 030	[36]
	<i>O. glaberrima</i>	316	217	33 164	[102]
	<i>O. granulata</i>	736.7	916.3	40 131	[103]
	<i>O. granulata</i>	777	262	40 116	[104]
	<i>O. longistaminata</i>	351	N/a ^a	34 389	[105]
	<i>O. meridionalis</i>	446.4	163	21 169	[106]
	<i>O. officinalis</i>	584	508	29 930	[36]
	<i>O. rhizomatis</i>	559	82	32 083	[36]
	<i>O. rufipogon</i>	380.5	30 200	34 830	[107]
	<i>O. rufipogon</i>	399.8	20 300	36 520	[108]
	<i>O. rufipogon</i>	384.8	219.4	22 035	[106]
Soybean	<i>Glycine soja</i>	1013.2	3300	89 477	[34]
Tomato	<i>Solanum pennellii</i>	942	1700	32 273	[33]
	<i>S. pennellii</i>	~1000	2500	N/a	[109]
	<i>S. chilense</i>	914	70.6	25 885	[110]
	<i>S. pimpinellifolium</i>	811	75.7	25 970	[111]
	<i>S. pimpinellifolium</i>	808.1	10 900	35 761	[112]
Wheat	<i>Aegilops tauschii</i>	4300	207.8	39 622	[113]
	<i>Triticum turgidum</i> ssp. <i>dicoccoides</i>	10 100	6955	62 813	[32]
	<i>Triticum urartu</i>	3900	64.54	34 879	[114]

^aAbbreviation: N/a, not available.

To better illustrate the wide-ranging diversity among CWR genomes, Khan and coworkers [42] suggested extending the current pangenome approach to a genus-level strategy based on constructing a 'super-pangenome' that technically reflects a pangenome of pangenomes. Relying on *de novo* assembly of selected representative accessions and resequencing of other accessions within the species, the development of a super-pangenome aims to assemble multiple species-specific pangenomes into a coherent graph to unleash the full range of genomic diversity that is present among CWRs. This universe of diversity otherwise remains obscure and inaccessible due to limitations inherent in current genome-wide approaches. A genus-level pangenome would potentially enable access to hitherto untapped genetic variation hidden within a specific CWR and permit exploration of the dispensable portion of the genome of a species, and this could greatly contribute to understanding crop genome evolution and adaptation. With growing refinements in informatics tools to manage pangenome data, the discovery of non-reference genomic elements would lay the foundation for identifying novel genotype–phenotype associations, with obvious potential for application in future crop breeding and improvement efforts.

Speed cloning of CWR genes through mutational genomics and association genetics

CWRs harbor a diversity of beneficial genes that confer resistance to pests, pathogens, and forms of abiotic stress that are of great value for sustainable crop improvement [43]. However, the problem of linkage drag between beneficial and deleterious alleles hampers their deployment in breeding programs through traditional approaches. The reduced number and diversity of

Table 2. Pangenome studies in crops involving CWRs

Crop	Number of accessions	Core genes	Dispensable genes	Refs
Barley	20 Accessions [cultivars, landraces, and one wild barley (<i>H. vulgare</i> subsp. <i>spontaneum</i>)]	25 228	14 948	[115]
<i>Brassica napus</i>	8 Cultivars	58 714	44 035	[116]
<i>Brassica oleracea</i>	8 Cultivars and 1 wild accession (<i>Brassica macrocarpa</i>)	49 895	11 484	[117]
Rice	66 Accessions of <i>O. sativa</i> and <i>O. rufipogon</i>	26 372	16 208	[39]
Sorghum	5 Wild accessions and 8 cultivars	15 867	28 212	[118]
Soybean	7 Wild accessions and 1 cultivar	28 716	30 364	[119]
Soybean	2898 (103 wild accessions, 1048 landraces, and 1747 cultivars)	28 786	28 706	[120]
Sunflower	493 Accessions (287 cultivars, 17 landraces, 189 wild accessions)	32 917	2464	[121]
Tomato	Accessions of <i>Solanum lycopersicum</i> (372), <i>S. lycopersicum</i> var <i>cerasiforme</i> (267), <i>S. pimpinellifolium</i> (78), <i>S. cheesmaniae</i> (3), and <i>S. galapagense</i> (8)	33 170	7199	[40]
Wheat	Chinese Spring and 18 diverse cultivars	89 795	26 711	[122]

genes conferring disease and pest resistance (R-genes) in modern cultivars renders them vulnerable to disease outbreaks [44] and evolving pathogen races. Access to a diversity of R-genes in CWRs and landrace accessions opens new avenues for pyramiding or stacking different genes in a single cultivar to impart broad-spectrum resistance to a wide array of pathogens and to enhance the durability of the resistance response. The nucleotide-binding site–leucine-rich repeat (NBS-LRR) gene family represents the most abundant class of R-genes in plants. These genes contain a central NBS domain and a C-terminal LRR domain [45]. The R-gene families typically have several members, often organized into tandem arrays along a chromosome, but genetic dissection of these complex loci based on recombination is challenging owing to their close proximity [43].

New strategies for identifying and cloning R-genes have been proposed in recent years. These include **RenSeq**, an approach that does not rely on positional cloning and provides a cost-effective alternative to whole-genome sequencing. RenSeq involves target gene enrichment using an NBS-LRR bait library accompanied by *de novo* assembly of enriched sequences generated by high-throughput sequencing [46]. The assembly is then used as a reference to detect variants between the susceptible and resistant (wild) types. Researchers have further modified RenSeq approaches, referred to as MutRenSeq and AgRenSeq, to accelerate the discovery and cloning of R-genes from CWRs. A combination of ethyl methane sulfonate (EMS)-mutagenesis and RenSeq was used for rapid cloning of the *Sr22* and *Sr45* genes in wheat and the study illustrates the potential of a mutational genomics approach for CWR gene discovery [43]. Circumventing the need for any synthetic or mutant population, AgRenSeq combines RenSeq with *k*-mer-based association mapping of a diversity panel. AgRenSeq was applied to 174 accessions of *Aegilops tauschii* ssp. *strangulata* and enabled rapid cloning of four *Sr* genes (*Sr33*, *Sr45*, *Sr46*, *SrTA1662*) that confer valuable levels of resistance against the wheat stem rust pathogen, *Puccinia graminis* f. sp. *tritici* [44]. Previous studies in wheat demonstrated mapping and introgression of stem rust resistance genes (*Sr33*, *Sr45*, *Sr46*, *SrTA1662*) from direct crosses between an elite wheat breeding line and resistant *Ae. tauschii* accessions (see Olson and coworkers [47] and references therein).

Further breakthroughs in long-read sequencing may help to alleviate the bottlenecks of RenSeq where CNV and high sequence similarity between paralogs and alleles make it difficult to accurately align short-read sequences [46]. By combining RenSeq with single-molecule real-time (SMRT) sequencing, Witek and coworkers [48] successfully cloned the *Rpi-amr3i* gene that confers resistance to late blight disease from diploid wild potato *Solanum americanum*. These examples suggest that sequencing of diversity panels at the whole-genome level is likely to provide a valuable tool for the identification of NBS-LRR genes that confer resistance to a wide array of pathogens.

Crop improvement strategies for deploying novel variation

Growing disease, insect, and environmental pressures under climate change threaten global crop production and amplify calls for expanding the range of genetic diversity available for use in breeding [25]. In the following section we describe different genetic designs that aim to facilitate targeted incorporation of the variation from CWRs into elite crop gene pools (Figure 2).

Bridging the gap between trait discovery and introgression

The advanced backcross QTL (AB-QTL) approach, pioneered in tomato and shortly thereafter demonstrated in rice, was proposed to allow simultaneous identification and introgression of superior alleles from unadapted germplasm into elite backgrounds, thus paving the way for **transgressive improvement** of the elite material. Delaying QTL identification to advanced stages (BC2/BC3) in AB-QTL creates avenues for the generation of lines that have greater similarity to the recurrent parent. This is based on early generation selection against pre-domestication traits and negative epistasis between the donor QTL and the elite genetic background [49]. The AB-QTL design permits more reliable measurements of quantitative trait variation owing to the higher proportion of adapted recurrent parent background, and to a higher likelihood of observing beneficial introgressions because of elimination of deleterious traits such as sterility, very late flowering, or extreme architectural variation, based on early-generation selection. Mapping of beneficial quantitative trait alleles from CWRs using conventional 'balanced' mapping populations is challenged by the high frequency of deleterious alleles and agronomically unadapted traits that hamper meaningful phenotyping as a result of an abundance of negative epistatic interactions and linkage drag. By bridging the gap between QTL discovery and transfer, AB-QTL accelerates the recovery of beneficial QTL near-isogenic lines (NILs) which serve as resource for isolating genes or can be used directly as parents in the variety release pipeline. For instance, a decade of research on AB-QTL in rice generated lines carrying *O. rufipogon* introgressions that were used in the development of hybrids and for positional cloning of the detected QTL, and provided functional markers for breeding applications [50]. In addition to tomato and rice, the AB-QTL approach has been used to source variation from CWRs for a variety of traits in other crops including wheat [51,52], maize [53], barley [54,55], cotton [56], common bean [57], groundnut [58], and pigeonpea [59].

The application of the AB-QTL approach to plant breeding is limited by a bias toward recurrent parent alleles, the creation of extensive linkage disequilibrium (LD), reduced opportunity for recombination, and, most importantly, deployment of exotic QTLs into breeding pipelines without extensive validation of their effects. A more recent design that combines backcrossing with nested association mapping (NAM) overcomes these bottlenecks by crossing diverse CWR accessions to a common elite parent(s) [60]. In barley, a genome-wide association study (GWAS) of 796 **advanced backcross nested association mapping (AB-NAM)** lines derived from crosses between 25 wild barleys and the six-rowed malting barley cultivar, Rasmusson, resulted in high-resolution identification of loci for glossy spike, glossy sheath, and black hull color [61]. In addition, loci were discovered for agronomically important traits such as days to heading, plant

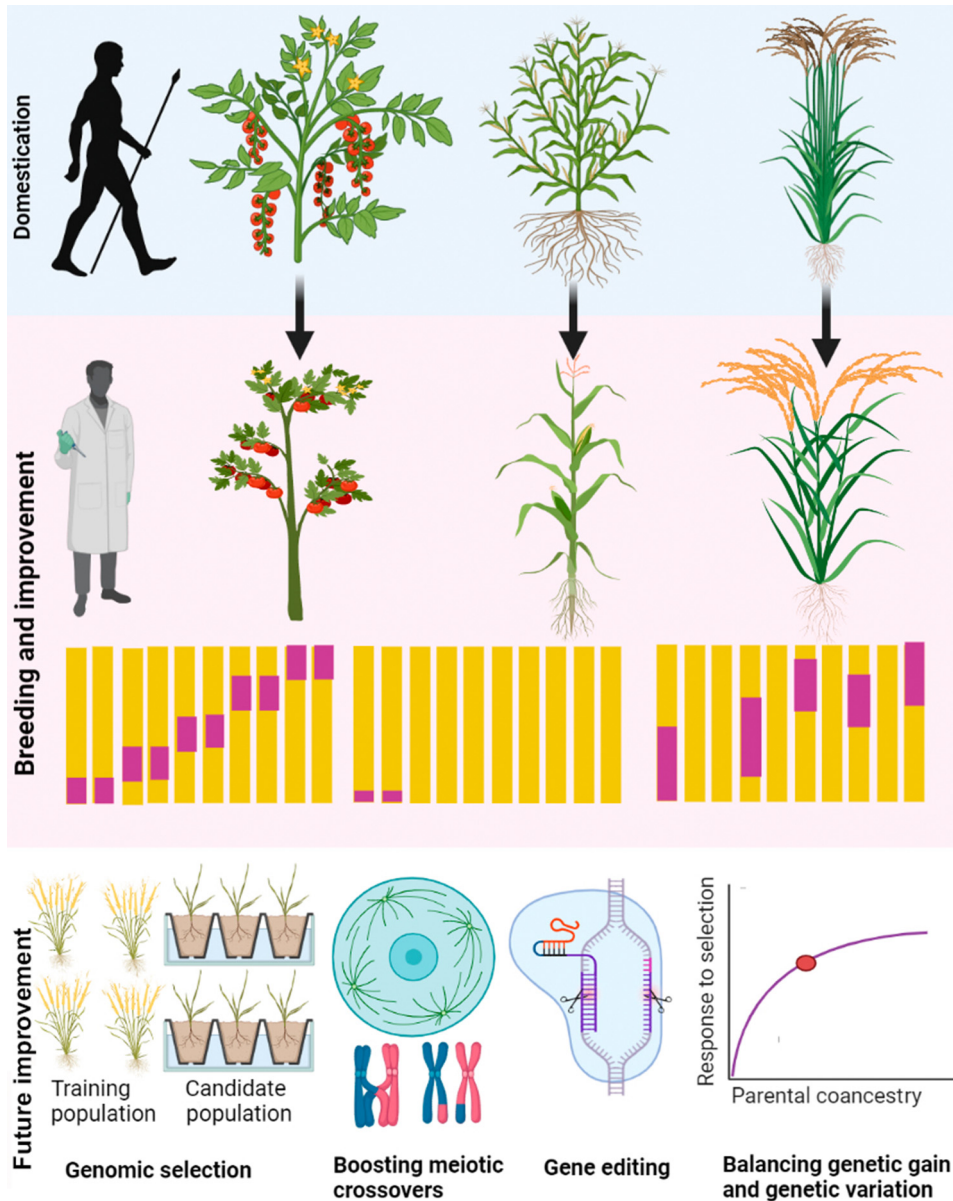


Figure 2. Domestication of crop wild relatives (CWRs) and their use for crop improvement. The domestication of CWRs of food crops started nearly 10 000 year ago. Subsequent breeding practices resulted in crop cultivars that are suitable for current agricultural settings. The arrows indicate the wild-to-cultivated transformation, where tomato, maize and rice are given as examples. Advances in genomic technologies have leveraged introgression breeding to a great extent by minimizing several of its bottlenecks, including linkage drag. A variety of community genetic resources harboring precisely defined chromosomal segments of CWRs are now available in different crops for use in research and breeding. The backcross populations, near-isogenic lines, and chromosome segment substitution lines are shown in the figure that contain varying levels of CWR genomes (magenta color) in a cultivated base (yellow color). New approaches including genomic selection, gene editing, optimum contribution selection, and improvement of recombination frequencies will unleash the true genetic potential of CWRs. The image was created using BioRender (<https://biorender.com/>).

height, number of productive tillers, seed weight, and yield [61]. Like the typical NAM design, AB-NAM offers the benefits of high genetic diversity together with balanced population structure and improved detection power for minor-frequency alleles.

Genomic selection in the context of pre-breeding

The availability of inexpensive, high-throughput genotyping has made it possible to implement genome-wide prediction to accelerate breeding progress in both plants and animals [62,63]. Implementation of this approach requires the development of a training set consisting of individuals that have been both genotyped and phenotyped to provide the basis for predicting the performance of similar individuals in a target population that has only genotypic information. The training set is used to obtain genomic estimated breeding values (GEBVs), which are used as the basis for recurrent selection in the process referred to as **genomic selection (GS)** [64–66]. The success of GS has been demonstrated in commercial breeding programs using a narrow pool of elite germplasm, such that the training set can be readily selected to optimize the prediction accuracy of GS. Several challenges must be addressed when applying GS to exotic × adapted crosses. One is to avoid narrowing the germplasm base too far, which can occur when only individuals with the largest GEBV are selected for mating in each generation. The second is to define the composition of the training population in a way that maximizes prediction accuracy in both elite and exotic materials. A solution proposed by Cowling and coworkers [67], further developed by Allier and colleagues [68], is to use **optimum contribution selection (OCS)** to improve the selection of exotic materials for use as parents while maintaining a desired level of diversity in the pre-breeding population as a whole. An increase in genetic variation with reduced inbreeding was reported in winter wheat for fructan content based on empirical evidence from two cycles of GS with OCS, although the gains were comparable between GS with simple truncated selection and GS with OCS [69]. Operationally, variations on OCS have been developed to consider information about pedigree relationships and/or genomic relationships to improve the prediction accuracies of both mean performance and genetic diversity maintenance in either the unselected or the selected fraction of the progeny [68].

In a related strategy, Yang and coworkers [70] proposed 'origin-specific genomic selection' (OSGS) in which favorable alleles are first partitioned according to the contributing parent (elite or exotic) and then marker effects are estimated, followed by prediction of breeding values of each candidate line using all markers. Unlike previous GS approaches aimed at selecting the best exotic line(s) for crossing, OSGS implements GS directly into the exotic × elite population to improve the 'introgression potential' of the exotic parent [70]. These approaches can help to extract favorable alleles from both parents and facilitate transgressive improvement.

In the post-next-generation sequencing era, the ability to generate genome-wide marker information on breeding populations in a cost- and time-efficient manner makes it possible to combine GS with OCS and/or OSGS in a pre-breeding program. The aim is to optimize the rate of genetic gain while recurrently introducing valuable exotic sources of diversity to broaden the genetic base of the elite breeding gene pool.

Extending genome-wide predictions to broaden the base of breeding populations

The ability of GS to harness the multitude of small-effect loci that contribute to polygenic variation in CWRs without requiring the development of targeted DNA markers makes this approach particularly suitable for broadening the genetic base of breeding populations. As discussed above, GS leverages information from training populations to derive genome-wide predictions that can hasten the breeding cycle and improve genetic gain. A simulation study by Bernardo [71] advocated for applying GS directly on elite × CWR **F₂ generation** populations rather than on

backcross (BC₁ and BC₂) populations (as previously outlined in the AB-QTL approach) to achieve a higher response to selection, even though the number of favorable alleles was significantly higher in the elite parent than in the exotic germplasm. The logic was that high trait heritability and large population size would ascertain a larger response from GS to accelerate population improvement. In the 'Seeds of Discovery' (SeeD; <http://seedsofdiscovery.org>) initiative a GS-based approach allowed shorter breeding cycles, which was an integral part of the effort to increase the frequency of favorable alleles in the CWR population before crossing with elite lines [72]. Large-scale genotyping of ~80 000 wheat accessions including 3903 CWRs performed under the SeeD initiative generated a valuable resource for the application of GS to breeding as well as for trait discovery in wheat [73]. The germplasm developed in the SeeD project includes semi-inbred lines of maize that are tolerant to heat, drought, and tar spot disease.

Novel breeding strategies for reclaiming lost genetic diversity

Since the earliest domestications of cereals (barley, rye, and several wheats) and pulses (lentil, pea, chickpea), full domestication has been recorded for nearly 300 plant species and 2500 species have experienced domestication at one time or another [74]. As discussed in previous sections, domestication and breeding efforts have delivered today's crops which produce high yields under current climatic conditions. However, several beneficial wild traits have been lost along the way including those that contribute to yield stability, including adaptation to harsh environments.

The **reflective plant breeding paradigm (RPBP)** is a theoretical framework to guide holistic germplasm development and the diversification of annual cropping systems to accelerate climate adaptation. The goal of the RPBP is to accelerate germplasm development through innovations in breeding, genomics, phenotyping, and agroecosystem modeling in close coordination with the design of new production systems and supply-value chains to ensure timely commercialization of the resultant germplasm [75]. A subsequent paper by de Haan and coworkers [76] addressed key aspects to consider when selecting candidates for domestication and the development of new crops to meet the agricultural targets defined in a given research pipeline.

In line with the RPBP concept of coordinated innovation and multi-stakeholder engagement, crop domestication is understood to be an iterative learning process that involves changes in human social relationships and the emergence of markets and trade along with agronomic innovation and broad genetic changes affecting multiple traits [77]. Genetic research on domestication in various crops such as tomato, maize, rice, and soybean has shown that a limited number of major-effect genes are responsible for the most obvious transitions from wild to domesticated forms [74]. The abundance of genome sequence information on wild and landrace accessions of many crops makes it possible to identify allele-frequency changes at key domestication loci that differentiate wild from cultivated material. Selective sweeps at these loci are often accompanied by shifts in allele frequencies at numerous other loci across the genome where the genomic targets and corresponding phenotypes cannot always be precisely identified. The ability to identify functional nucleotide polymorphisms in genes of known function opens the door to the possibility of editing genomic targets in virtually any variety or wild genetic background of interest (Table 3). Thus, instead of introducing wild alleles into elite material via backcrossing, it is possible to directly manipulate domestication-related genes in CWRs or other PGRs themselves (provided that they can be transformed and regenerated). This process has been termed **de novo domestication** [74]. Targeted interventions based on introgression breeding and cutting-edge genome-editing technology can facilitate *de novo* domestication [78].

The discovery of **sequence-specific nuclease** systems including zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and more recently clustered regularly

Table 3. Examples of *de novo* domestication in plants

Crop/CWR	Domestication traits	Genes targeted	Refs
<i>Oryza alta</i>	Seed shattering, awn length, plant height, ideal plant architecture, grain length and size, and heading date	<i>qSH1</i> , <i>An-1</i> , <i>SD1</i> , <i>IPA1</i> , <i>GS3</i> , <i>Ghd7</i> , <i>DTH7</i>	[84]
<i>Physalis peruinosa</i>	Plant architecture, flower production, and fruit size	<i>SELFPRUNING (SP)</i> , <i>SELFPRUNING 5G (SP5G)</i> , <i>CLAVATA (CLV)</i>	[83]
<i>Solanum pimpinellifolium</i>	Plant architecture, fruit size, and inflorescence branching	<i>Fasciated (fas)</i> , <i>locule number (lc)</i> , <i>CLV3</i>	[79]
<i>S. pimpinellifolium</i>	Plant architecture, fruit ripening, photoperiod insensitivity, fruit size, and vitamin C level	<i>SP</i> , <i>SP5G</i> , <i>SICLV3</i> , <i>SIWUS</i> , <i>SIGGP1</i>	[82]
<i>S. pimpinellifolium</i>	Plant growth habit, fruit shape, and size, fruit number, and nutritional quality	<i>SELFPRUNING (SP)</i> , <i>OVATE (O)</i> , <i>FRUIT WEIGHT 2.2 (FW 2.2)</i> , <i>LYCOPENE BETA CYCLASE (CycB)</i>	[81]

interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein-9 nuclease (Cas9) have revolutionized the field of genome editing. The rising popularity of CRISPR/Cas9 editing can be ascribed to its flexibility and expandability for introducing one or more double-strand breaks (DSBs) at any desired sequence in the genome [79,80]. Pioneering work in wild tomato (*Solanum pimpinellifolium*) illustrated the possibility of rapidly manipulating domestication phenotypes through editing of gene and regulatory regions by CRISPR/Cas9 [79,81] while consciously retaining wild traits such as salt tolerance and resistance to bacterial spot disease [82]. Similarly, major domestication traits were improved via CRISPR/Cas9-based editing of tomato orthologs in a related Solanaceae crop, groundcherry (*Physalis peruinosa*) [83]. A more recent study demonstrated the possibility of creating a novel cereal crop by *de novo* domestication of an allotetraploid wild rice *Oryza alta* (CCDD) [84]. These examples, combined with the availability of high-quality genome assemblies, gene-specific targets associated with phenotypes of interest, efficient tissue culture (and/or tissue culture-free), and transformation methodologies in a wide range of species, provide fertile ground for further CRISPR/Cas9 applications in stepwise strategies for domesticating wild crops *de novo*.

Boosting recombination rates for improving the efficiency of CWR introgression

The deployment of DNA marker technologies in a backcross program serves the dual purpose of accelerating recovery of the recurrent parent genome and reducing the amount of unwanted donor genome. Precise selection for a target locus helps to minimize linkage drag. Genetic recombination facilitated by meiotic crossovers (COs) is a key driver for disrupting linkage drag. However, in eukaryotes COs are limited, typically in the range of one to three per chromosome [85], and they are not evenly distributed along the chromosomes. For example, 80% of COs in *Arabidopsis thaliana* (arabidopsis) were detected in only one quarter of the genome, and recombination hot-spots containing CO clusters as well as recombination cold-spots, generally centromeric regions which are devoid of COs, have been well documented [86].

Programmed DSBs followed by repair during meiosis facilitate shuffling of genetic information between homologous non-sister chromatids. However, a majority of these DSBs remain unrepaired and result in the formation of non-crossovers (NCOs). Three pathways that suppress meiotic recombination have been identified in arabidopsis; these pathways rely on the activity of *Fanconi anemia of complementation group M (FANCM)* helicase, *Fidgetin-like protein 1 (FIGL1)*, and *RecQ helicase 4 (RECQ4)* that promote the formation of NCOs from unrepaired DSBs (see Pele and coworkers [87] and references therein). The evidence suggests that meiotic

recombination increases in response to manipulation of the above three pathways through the creation of mutation knockouts and/or a change in ploidy levels. Fernandes and coworkers [85] constructed single, double, and triple mutants for *FANCM*, *RECQ4*, and *FIGL1* in arabidopsis, and the hybrid double mutant (*recq4* and *figl1*) caused an increase of 2648 cM in the genetic map length, representing a 7.8-fold increase in genome-wide recombination. The approach of mutating *FANCM*, *RECQ4*, and *FIGL1* was extended to three other crops, rice, pea, and tomato, and a threefold greater CO frequency resulted from knocking out the *RECQ4* gene, showing that *RECQ4* suppresses CO in all three species [88]. Given the genome instability that could result from such a mutation, Mieulet and coworkers [88] advocated 'segregating out' the *RECQ4* mutation before the production of elite lines. The mutagenesis approach does not appear to influence recombination rates in centromeric regions, whereas a change in ploidy levels is reported to alter both recombination rates and the recombination landscape. For example, Pele and coworkers [87] demonstrated that the presence of nine C chromosomes in the allotriploid hybrid (AAC, $2n = 3x = 29$) resulted in an up to 3.4-fold increase in CO numbers in comparison to the diploid *B. rapa* (AA, $2n = 2x = 20$). The allotriploid hybrid is derived from a cross between the tetraploid *B. napus* (AACC, $2n = 4x = 38$) and the diploid *B. rapa*. The study demonstrated an increase in CO number for all 10 homologous A chromosomes in allotriploid hybrids, thus reinforcing the pattern previously reported for A07 chromosomes [89]. In addition, exploring genotypic differences in high or low recombining lines of maize [90] under differing environmental conditions, such as high temperature [91], shows promise with respect to controlling recombination rates for improving breeding efficiency.

Enhancing genome-wide recombination rates may help to address the problem of linkage drag, but the ability to ultimately target recombination to desired locations in the genome would be a major step forward [92]. A simulation study examining the potential of the above approaches to enhance backcross breeding showed promise only in cases where recombination rates were improved in cold-spots that contained target loci [93]. By contrast, increasing recombination rates in hot-spot regions may actually prove to be detrimental. Given that the *RECQ4* mutation caused only a limited increase in recombination targeted to genomic regions with high genetic divergence, the potential of these approaches remains to be seen in pre-breeding populations that incorporate highly diverse CWR genomes.

Conserving CWRs and landraces in genebanks: gap analysis and collection needs

The **ex situ conservation** of CWRs is becoming increasingly important in the light of threats to naturally occurring, *in situ* populations resulting from land-use change, competition from biofuel crops, urbanization, invasive species, pollution, mining, and climate change [94,95]. *Ex situ* conservation refers to a global system of safeguarding CWRs as seeds (or clonal propagules) in non-native settings such as genebanks, botanical gardens, and plant breeding programs (<https://colostate.pressbooks.pub/cropwildrelatives/>). A growing body of literature has provided evidence that CWRs are not sufficiently represented in global *ex situ* conservation systems. For example, nearly half of the material conserved *ex situ* has been categorized into cultivars, landraces, and CWRs, and landraces and CWRs comprise 44% and 17%, respectively, of the *ex situ* collection of known nature (www.fao.org/3/i1500e/i1500e03.pdf). Similarly, CWRs represent only 5.6% of total germplasm holdings in European genebanks based on information from the European Search Catalogue for Plant Genetic Resources (EURISCO; <http://eurisco.ecpgr.org>) that comprises 1.8 million samples from 43 European countries, representing more than half of the European and roughly 16% of total genebank holdings worldwide [96]. In another study, Maxted and Kell (www.fao.org/3/i1500e/i1500e18a.pdf) estimated that CWRs account for ~6% of global genebank *ex situ* collections, and that *ex situ* conservation was reported for only ~6% of the total

number of CWR species. Further, *ex situ* management and propagation of CWRs is more expensive and often less successful than for domesticated landrace materials, in part because of challenges related to flowering time, mating habits, and pollination requirements, as well as seed shattering, dormancy, and viability. All this points to an urgent need to improve *ex situ* conservation protocols as well as to identify gaps in current germplasm collections and to prioritize crucial germplasm and geographical hot-spot sites for future collection.

A systematic study performed under the CWR project (www.cwrdiversity.org/) analyzed the comprehensiveness of CWR conservation efforts involving 1076 taxa related to 81 crops [95]. The authors reported insufficient representation of >95% taxa based on their geographical and ecological representations. Importantly, for 29% of total taxa not a single accession was conserved *ex situ*. The study assigned high priority to 73% of the taxa and identified hot-spot regions for further collection of these high-priority materials. The priorities rely on the representativeness of the taxa/species in terms of sampling and geographic and ecological coverage. The study also revealed geographic areas that have crucial gaps. This approach, based on occurrence records, distribution modeling, and gap analysis, can be extended to prioritization at the species level. For instance, when applied to pigeonpea, it highlighted under-representation of pigeonpea CWRs in the *ex situ* collections and identified 80% of the species as high priority for collection. The recommendations for future collection included *Cajanus latisepalus*, *Cajanus cinereus*, and *Cajanus reticulatus* that emerged as high-priority species for collection based on gap analysis and expert assessments [97]. Further, areas with species richness (southern India and northern Australia) were identified as hot-spots for collection for prioritized pigeonpea species. The data pertaining to species occurrence and ecogeographic variables in these studies can further be harnessed to identify species with adaptive traits such as tolerance to temperature extremes, water logging, and drought. Equally importantly, comprehensive documentation of CWRs plays a crucial role while assessing their over- or under-representation in genebanks [98]. Given the ongoing large-scale projects on genebank genotyping, and the development of international information systems such as EURISCO and Genesys (www.genesys-pgr.org/), extending these web resources beyond passport and phenotyping data will be essential to realize the true potential of CWRs for research and breeding.

Access to PGRs requires sharing of benefits with resource providers

Access to PGRs and benefit-sharing derived from that access are essential to sustainable development, as outlined in the UN Sustainable Development Goals (<https://sdgs.un.org/goals>). Under public international law, the principles of 'access and benefit-sharing' (ABS) are to fairly distribute benefits arising from PGRs between the users of those resources (i.e., universities, agrobiotech companies, etc.) and provider countries. The ABS principles are codified in the Convention on Biological Diversity (CBD) 1992 (www.cbd.int/abs/theabsch.shtml), the Nagoya Protocol 2011 (www.cbd.int/abs/about/), and the International Treaty on Plant Genetic Resources (hereafter, the Treaty) (www.fao.org/plant-treaty/en/). The ABS principles guide the development of regulatory mechanisms that are designed and put in place by provider countries who are parties to the CBD and the Nagoya Protocol. Because each country interprets ABS in its own way, and designs its own regulatory system, benefit-sharing mechanisms are not uniform. For instance, there are monetary and non-monetary forms of benefit-sharing, and both are shaped by the national policies of the provider countries. For a thorough treatment of the subject, the reader may refer to Sirakaya [99]. This has led to ambiguity and confusion for both commercial and non-commercial users who seek to access PGRs for research, conservation and/or product development, as well as for providers who are required to give 'prior informed consent' for the use of PGR, but who are often vastly under-resourced and receive no direct benefit from the transaction. Despite the challenges, the Treaty provides a system through which PGR can be accessed

and benefit-sharing mechanisms agreed to, and ongoing efforts to streamline the process, lower the transaction costs, clarify expectations, and contribute to the ABS goals are slowly improving the system as a whole.

Concluding remarks

CWRs and traditional landrace varieties contain a vast array of beneficial traits that are essential to improve the resilience of crops in harsh climates and to sustain global food supplies. Diversified sources of male sterility from CWRs have helped the seed industry to avoid disease outbreaks and have supported the seed industry to enhance crop performance through hybrid vigor (Box 2). Advances in 'omic' and high-throughput phenotyping technologies have greatly increased the throughput of data generation and decreased the cost per data-point, leading to an abundance of digital information about genetic, biochemical, and physiological diversity at multiple growth stages in different crop species. Although these approaches have largely been applied to diversity panels and breeding populations, rigorous evaluation of structured introgression libraries is likely to yield returns, especially if automated phenotyping for a wide array of traits can be performed on plants grown in diverse environments. When combined with high-throughput genotyping and other omic analyses, these experiments would provide insight into the ways that different introgressions from diverse donors alter plant responses to stress. Sib-mated populations of CSSLs could be generated to further dissect loci of interest and to examine the effects of heterozygosity on particular regions of the genome. The high dimensionality of ecogeographic, climate, and phenotypic data provides immense scope for the adoption of ML tools (see Outstanding questions). Passport data from germplasm collections may be used to feed expression GWAS (eGWAS) models to accelerate the identification of adaptive variation for improving climate resilience in future crops. Germplasm enhancement strategies that create bridging germplasm can help to accelerate the use of PGRs in breeding programs with reduced challenges in adaptation. Achieving optimal allelic contributions from both elite line and unadapted CWR or landrace accessions will be key to long-term gain from pre-breeding programs. Pangenome studies empowered by increasing volumes of data on the diversity of plant genomes will further reveal hitherto untapped genetic variation for use in crop improvement. Pangenomes, in combination with previous research on domestication loci and genes governing traits of interest, will delineate breeding targets for manipulation through gene editing and/or introgression. Improving recombination frequencies could help to unlock the enormous genetic potential of CWRs; however, its utility for overcoming linkage drag and improving the resolution of desirable introgression has yet to be demonstrated empirically. Breakthroughs in genome-editing technologies applied to different plant species are already enabling scientists to target genomic changes, better understand the phenotypic and performance consequences of specific modifications, and overcome some of the inherent problems of local and genome-wide linkage drag. Technical improvements in genome editing will be necessary to be able to reliably and simultaneously target hundreds or thousands of loci for editing in plant genomes. This would make it possible to remove presumed deleterious variants from across the genome to minimize genetic load and determine to what extent they impose a drag on overall performance, especially in outcrossing species. A renewed focus to identify crucial gaps in genebank collections, improve seed conservation strategies particularly for CWRs, and inform future germplasm collection strategies is needed, together with support for long-term initiatives such as the CWR Project [25] to forestall genetic erosion. Substantial new investment should also be focused on the exploration of genetic and phenotypic variation in non-traditional species coupled with innovative uses of GS-assisted breeding strategies to adapt new crops for cultivation in mixtures or in rotation with traditional crops. These strategies can help to expand the range of species diversity in future agri-food systems, thereby providing new opportunities to buffer the effects of climate change and opening up new doors for introducing novel germplasm and incorporating it into sustainable cropping

Outstanding questions

Domestication and breeding have improved our current food crops. However, a cost is associated with these processes. To what extent will *de novo* domestication facilitated by gene editing or introgression be able to balance the trade-offs?

Will innovations in genomics, breeding, and phenotyping correct the long-standing perception of breeders who still judge the potential of CWRs by their agronomic performance in current agricultural settings?

Direct applications of genomic selection in pre-breeding are still rare. Given that adequate technical expertise is needed for genomic selection, how will this approach contribute equitably to germplasm enhancement in developing countries?

How can the research community be encouraged to deposit valuable genotypic and phenotypic data on community resources such as genetic libraries into easily accessible global repositories to facilitate future meta-analyses?

Novel methods that enhance meiotic recombinations seem fascinating, especially with respect to the long-standing problem of linkage drag. Because initial studies based on mutagenesis demonstrate a limited influence of these methods on highly diverse genomic regions and on cold-spots, to what extent will introgression breeding involving highly diverse accessions benefit from these innovations?

Large-scale ecogeographic approaches in combination with ML tools will be crucial to realize the true genetic potential of CWRs. Is the current plant breeding community in general adequately equipped to deal with the high dimensionality of these datasets?

systems. Together, these measures can significantly reduce the vulnerability of current crops to pests, diseases, and environmental stresses, and will greatly augment the potential to achieve food security targets around the world.

Acknowledgments

R.K.V. thanks the Science and Engineering Research Board (SERB) of the Department of Science and Technology (DST), Government of India, for providing a J C Bose National Fellowship (SB/S9/Z-13/2019). B.K. thanks the Government of Norway (QZA-14/0005) for funding the initiative “Adapting Agriculture to Climate Change: Collecting, Protecting and Preparing Crop Wild Relatives” (<https://www.cwrdiversity.org/project/pre-breeding/>). Thanks are also due to Rutwik Barmukh, from ICRIASAT, and Abha Tiwari from ICAR-IIPR for their help and support in the preparation of this article.

Declaration of interests

No interests are declared.

Supplemental information

Supplemental information associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.tibtech.2021.08.009>

References

- Tester, M. and Langridge, P. (2010) Breeding technologies to increase crop production in a changing world. *Science* 327, 818–822
- Tanksley, S.D. and McCouch, S.R. (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277, 1063–1066
- Abbo, S. *et al.* (2014) Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. *Trends Plant Sci.* 19, 351–360
- Tyack, N. *et al.* (2020) The potential of payment for ecosystem services for crop wild relative conservation. *Plants* 9, 1305
- Zamir, D. (2001) Improving plant breeding with exotic genetic libraries. *Nat. Rev. Genet.* 2, 983–989
- Cowling, W.A. (2013) Sustainable plant breeding. *Plant Breed.* 132, 1–9
- Wang, C. *et al.* (2017) Emerging avenues for utilization of exotic germplasm. *Trends Plant Sci.* 7, 624–637
- Warschafsky, E. *et al.* (2014) Back to the wilds: tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. *Am. J. Bot.* 101, 1791–1800
- Dempewolf, H. *et al.* (2017) Past and future use of wild relatives in crop breeding. *Crop Sci.* 57, 1070–1082
- Summers, R.W. and Brown, J.K.M. (2013) Constraints on breeding for disease resistance in commercially competitive wheat cultivars. *Plant Pathol.* 62, 115–121
- Brouwer, D.J. and St Clair, D.A. (2004) Fine mapping of three quantitative trait loci for late blight resistance in tomato using near isogenic lines (NILs) and sub-NILs. *Theor. Appl. Genet.* 108, 628–638
- Wang, M. *et al.* (2020) A hybrid sterile locus leads to the linkage drag of interspecific hybrid progenies. *Plant Divers.* 42, 370–375
- Upadhyaya, H. and Ortiz, R. (2001) A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. *Theor. Appl. Genet.* 102, 1292–1298
- Glaszmann, J.C. *et al.* (2010) Accessing genetic diversity for crop improvement. *Curr. Opin. Plant Biol.* 13, 167–173
- Bari, A. *et al.* (2012) Focused identification of germplasm strategy (FIGS) detects wheat stem rust resistance linked to environmental variables. *Genet. Resour. Crop. Evol.* 59, 1465–1481
- Stenberg, J.A. and Ortiz, R. (2021) Focused identification of germplasm strategy (FIGS): polishing a rough diamond. *Curr. Opin. Insect. Sci.* 45, 1–6
- Bhullar, N.K. *et al.* (2009) Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the Pm3 resistance locus. *Proc. Natl. Acad. Sci. U. S. A.* 106, 9519–9524
- Bari, A. *et al.* (2014) Predicting resistance to stripe (yellow) rust (*Puccinia striiformis*) in wheat genetic resources using focused identification of germplasm strategy. *J. Agric. Sci.* 152, 906–916
- Elbasyoni, I.S. *et al.* (2019) Evaluation of a global spring wheat panel for stripe rust: resistance loci validation and novel resources identification. *PLoS One* 14, e0222755
- Ei Bouhssini, M. *et al.* (2011) Sources of resistance in bread wheat to Russian wheat aphid (*Diuraphis noxia*) in Syria identified using the focused identification of germplasm strategy (FIGS). *Plant Breed.* 130, 96–97
- Endresen, D.T.F. *et al.* (2011) Predictive association between biotic stress traits and ecogeographic data for wheat and barley landraces. *Crop Sci.* 51, 2036–2055
- Khazaei, H. *et al.* (2013) The FIGS (focused identification of germplasm strategy) approach identifies traits related to drought adaptation in *Vicia faba* genetic resources. *PLoS One* 8, e63107
- Haupt, M. and Schmid, K. (2020) Combining focused identification of germplasm and core collection strategies to identify genebank accessions for central European soybean breeding. *Plant Cell Environ.* 43, 1421–1436
- Vikas, V.K. *et al.* (2020) Screening of 19,460 genotypes of wheat species for resistance to powdery mildew and identification of potential candidates using focused identification of germplasm strategy (FIGS). *Crop Sci.* 60, 2857–2866
- Kilian, B. *et al.* (2021) Crop Science special issue: adapting agriculture to climate change: a walk on the wild side. *Crop Sci.* 61, 32–36
- Balakrishnan, D. *et al.* (2019) Development and use of chromosome segment substitution lines as a genetic resource for crop improvement. *Theor. Appl. Genet.* 132, 1–25
- Semel, Y. *et al.* (2006) Overdominant quantitative trait loci for yield and fitness in tomato. *Proc. Natl. Acad. Sci. U. S. A.* 103, 2981–2986
- Mia, M.S. *et al.* (2019) Multiple near-isogenic lines targeting a QTL hotspot of drought tolerance showed contrasting performance under post-anthesis water stress. *Front. Plant Sci.* 10, 271
- Campanelli, G. *et al.* (2019) Multi-parental advances generation inter-cross population, to develop organic tomato genotypes by participatory plant breeding. *Agronomy* 9, 119
- 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, 1082–1088

31. 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
32. Avni, R. *et al.* (2017) Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* 97, 93–97
33. Bolger, A. *et al.* (2014) The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nat. Genet.* 46, 1034–1038
34. Xie, M. *et al.* (2019) A reference-grade wild soybean genome. *Nat. Commun.* 10, 1216
35. Kajiya-Kanegae, H. *et al.* (2021) OryzaGenome2.1: database of diverse genotypes in wild *Oryza* species. *Rice* 14, 24
36. Stein, J.C. *et al.* (2018) Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. *Nat. Genet.* 50, 285–296
37. Shenton, M. *et al.* (2020) Evolution and diversity of the wild rice *Oryza officinalis* complex, across continents genome types, and ploidy levels. *Genome Biol. Evol.* 12, 413–428
38. Brozynska, M. *et al.* (2017) Sequencing of Australian wild rice genomes reveals ancestral relationships with domesticated rice. *Plant Biotechnol. J.* 15, 765–774
39. Zhao, Q. *et al.* (2018) Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* 50, 278–284
40. Gao, L. *et al.* (2019) The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* 51, 1044–1051
41. Li, J. *et al.* (2021) Cotton pan-genome retrieves the lost sequences and genes during domestication and selection. *Genome Biol.* 22, 119
42. 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
43. Steuernagel, B. *et al.* (2016) Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat. Biotechnol.* 34, 652–655
44. Arora, S. *et al.* (2019) Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nat. Biotechnol.* 37, 139–143
45. McHale, L. *et al.* (2006) Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* 7, 212
46. Jupe, F. *et al.* (2013) Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* 76, 530–544
47. Olson, E.L. *et al.* (2013) Simultaneous transfer, introgression, and genomic localization of genes for resistance to stem rust race TTKSK (Ug99) from *Aegilops tauschii* to wheat. *Theor. Appl. Genet.* 126, 1179–1188
48. Witek, K. *et al.* (2016) Accelerated cloning of a potato late blight-resistance gene using RenSeq and SMRT sequencing. *Nat. Biotechnol.* 34, 656–660
49. Tanksley, S.D. and Nelson, J.C. (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theor. Appl. Genet.* 92, 191–203
50. McCouch, S.R. *et al.* (2007) Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa*. *Euphytica* 154, 317–339
51. Huang, X.Q. *et al.* (2004) Advanced backcross QTL analysis in progenies derived from a cross between a German elite winter wheat variety and a synthetic wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 109, 933–943
52. Buerstmayr, M. *et al.* (2011) Advanced backcross QTL mapping of resistance to Fusarium head blight and plant morphological traits in a *Triticum macha* × *T. aestivum* population. *Theor. Appl. Genet.* 123, 293–306
53. Ho, J.C. *et al.* (2002) Improvement of hybrid yield by advanced backcross QTL analysis in elite maize. *Theor. Appl. Genet.* 105, 440–448
54. Pillen, K. *et al.* (2003) Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 107, 340–352
55. von Korff, M. *et al.* (2006) AB-QTL analysis in spring barley. II. Detection of favourable exotic alleles for agronomic traits introgressed from wild barley (*H. vulgare* ssp. *spontaneum*). *Theor. Appl. Genet.* 112, 1221–1231
56. Wang, B. *et al.* (2017) QTL analysis of cotton fiber length in advanced backcross populations derived from a cross between *Gossypium hirsutum* and *G. mustelinum*. *Theor. Appl. Genet.* 6, 1297–1308
57. Blair, M.W. *et al.* (2006) QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean × wild common bean (*Phaseolus vulgaris* L.) cross. *Theor. Appl. Genet.* 112, 1149–1163
58. Khara, P. *et al.* (2019) Genetic imprints of domestication for disease resistance, oil quality, and yield component traits in groundnut (*Arachis hypogaea* L.). *Mol. Gen. Genomics.* 294, 36
59. Saxena, R. *et al.* (2020) Genotyping-by-sequencing and multi-location evaluation of two interspecific backcross populations identify QTLs for yield-related traits in pigeonpea. *Theor. Appl. Genet.* 133, 737–749
60. Mace, E.S. *et al.* (2021) A global resource for exploring and exploiting genetic variation in sorghum crop wild relatives. *Crop Sci.* 61, 150–162
61. 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
62. Whittaker, J.C. *et al.* (2000) Marker-assisted selection using ridge regression. *Genet. Res.* 75, 249–252
63. Meuwissen, T.H. *et al.* (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829
64. Heslot, N. *et al.* (2015) Perspectives for genomic selection applications and research in plants. *Crop Sci.* 55, 1–12
65. Crossa, J. *et al.* (2017) Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 22, 961–975
66. Voss-Fels, K.P. *et al.* (2019) Accelerating crop genetic gains with genomic selection. *Theor. Appl. Genet.* 132, 669–686
67. 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
68. Allier, A. *et al.* (2020) Optimized breeding strategies to harness genetic resources with different performance levels. *BMC Genomics* 21, 349
69. Veenstra, L.D. *et al.* (2020) Recurrent genomic selection for wheat grain fructans. *Crop Sci.* 60, 1499–1512
70. Yang, C.J. *et al.* (2020) Origin specific genomic selection: a simple process to optimize the favorable contribution of parents to progeny. *G3 (Bethesda)* 10, 2445–2455
71. Bernardo, R. (2009) Genome wide selection for rapid introgression of exotic germplasm in maize. *Crop Sci.* 49, 419–425
72. Gorjanc, G. *et al.* (2016) Initiating maize pre-breeding programs using genomic selection to harness polygenic variation from landrace populations. *BMC Genomics* 17, 30
73. Sansaloni, C. *et al.* (2020) Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. *Nat. Commun.* 11, 4572
74. Fernie, A.R. and Yan, J. (2019) De novo domestication: an alternative route toward new crops for the future. *Mol. Plant* 12, 615–631
75. Runck, B.C. *et al.* (2014) The reflective plant breeding paradigm: a robust system of germplasm development to support strategic diversification of agroecosystems. *Crop Sci.* 54, 1939–1948
76. de Haan, L.R. *et al.* (2016) A pipeline strategy for grain crop domestication. *Crop Sci.* 56, 917–930
77. van Tassel, D.L. *et al.* (2020) New food crop domestication in the age of gene editing: genetically, agronomic and cultural change remain co-evolutionarily entangled. *Front. Plant Sci.* 11, 789
78. Zsögön, A. *et al.* (2017) Genome editing as a tool to achieve the crop ideotype and de novo domestication of wild relatives: case study in tomato. *Plant Sci.* 256, 120–130
79. Rodríguez-Leal, D. *et al.* (2017) Engineering quantitative trait variation for crop improvement by genome editing. *Cell* 171, 470–480
80. Zhu, H. *et al.* (2020) Applications of CRISPR-Cas in agriculture and plant biotechnology. *Nat. Rev. Mol. Cell Biol.* 21, 661–677

81. Zsögön, A. *et al.* (2018) De novo domestication of wild tomato using genome editing. *Nat. Biotechnol.* 36, 1211–1216
82. Li, C. *et al.* (2019) A new rice breeding method: CRISPR/Cas9 system editing of the Xa13 promoter to cultivate transgene-free bacterial blight-resistant rice. *Plant Biotechnol. J.* 18, 313–315
83. Lemmon, Z.H. *et al.* (2018) Rapid improvement of domestication traits in an orphan crop by genome editing. *Nat. Plants* 4, 766–770
84. Yu, H. *et al.* (2021) A route to de novo domestication of wild allotetraploid rice. *Cell* 184, 1156–1170
85. Fernandes, J.B. *et al.* (2018) Unleashing meiotic crossovers in hybrid plants. *Proc. Natl. Acad. Sci. U. S. A.* 115, 2431–2436
86. Choi, K. *et al.* (2013) *Arabidopsis* meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. *Nat. Genet.* 45, 1327–1336
87. Pele, A. *et al.* (2017) Amplifying recombination genome-wide and reshaping crossover landscapes in *Brassicacae*. *PLoS Genet.* 13, e1006794
88. Mieulet, D. *et al.* (2018) Unleashing meiotic crossovers in crops. *Nat. Plants* 4, 1010–1016
89. Leflon, M. *et al.* (2010) Crossovers get a boost in *Brassica* allotriploid and allotetraploid hybrids. *Plant Cell* 22, 2253–2264
90. Bauer, E. *et al.* (2013) Intraspecific variation of recombination rate in maize. *Genome Biol.* 14, R103
91. Phillips, D. *et al.* (2015) The effect of temperature on the male and female recombination landscape of barley. *New Phytol.* 208, 421–429
92. Reynolds, M. *et al.* (2021) Addressing research bottlenecks to crop productivity. *Trends Plant Sci.* 26, 607–630
93. Tourrette, E. *et al.* (2021) Enhancing backcross programs through increased recombination. *Genet. Sel. Evol.* 53, 25
94. Ford-Lloyd, B.V. *et al.* (2011) Crop wild relatives – undervalued, underutilized and under threat? *Bioscience* 61, 559–565
95. Castaneda-Alvarez, N.P. *et al.* (2016) Global conservation priorities for crop wild relatives. *Nat. Plants* 2, 16022
96. Weise, S. *et al.* (2017) EURISCO: the European search catalogue for plant genetic resources. *Nucleic Acids Res.* 45, D1003–D1008
97. Khoury, C.K. *et al.* (2015) Crop wild relatives of pigeonpea [*Cajanus cajan* (L.) Mill sp.]: distributions, ex situ conservation status, and potential genetic resources for abiotic stress tolerance. *Biol. Conserv.* 184, 259–270
98. Weise, S. *et al.* (2020) Document or lose it – on the importance of information management for genetic resources conservation in genebanks. *Plants* 9, 1050
99. Sirakaya, A. (2019) Balanced options for access and benefit-sharing: stakeholder insights on provider country legislation. *Front. Plant Sci.* 10, 1175
100. Liu, M. *et al.* (2020) The draft genome of a wild barley genotype reveals its enrichment in genes related to biotic and abiotic stresses compared to cultivated barley. *Plant Biotechnol. J.* 18, 443–456
101. Chen, J. *et al.* (2013) Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying *Oryza* genome evolution. *Nat. Commun.* 4, 1595
102. Wang, M. *et al.* (2014) The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nat. Genet.* 46, 982–988
103. Shi, C. *et al.* (2020) The draft genome sequence of an upland wild rice species, *Oryza granulata*. *Sci. Data* 7, 131
104. Wu, Z. *et al.* (2018) De novo genome assembly of *Oryza granulata* reveals rapid genome expansion and adaptive evolution. *Commun. Biol.* 1, 84
105. Reuscher, S. *et al.* (2018) Assembling the genome of the African wild rice *Oryza longistaminata* by exploiting synteny in closely related *Oryza* species. *Commun. Biol.* 1, 162
106. Brozynska, M. *et al.* (2016) Genomics of crop wild relatives: expanding the gene pool for crop improvement. *Plant Biotechnol. J.* 14, 1070–1085
107. Li, W. *et al.* (2020) SMRT sequencing of the *Oryza rufipogon* genome reveals the genomic basis of rice adaptation. *Commun. Biol.* 3, 167
108. Xie, X. *et al.* (2021) A chromosome-level genome assembly of the wild rice *Oryza rufipogon* facilitates tracing the origins of Asian cultivated rice. *Sci. China Life Sci.* 64, 282–293
109. Schmidt, H.W. *et al.* (2017) De novo assembly of a new *Solanum pennellii* accession using nanopore sequencing. *Plant Cell* 29, 2336–2348
110. Stam, R. *et al.* (2019) The de novo reference genome and transcriptome assemblies of the wild tomato species *Solanum chilense* highlights birth and death of NLR genes between tomato species. *G3 (Bethesda)* 9, 3933–3941
111. Razali, R. *et al.* (2018) The genome sequence of the wild tomato *Solanum pimpinellifolium* provides insights into salinity tolerance. *Front. Plant Sci.* 9, 1402
112. Wang, X. *et al.* (2020) Genome of *Solanum pimpinellifolium* provides insights into structural variants during tomato breeding. *Nat. Commun.* 11, 5817
113. Luo, M.C. *et al.* (2017) Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature* 551, 498–502
114. Ling, H.Q. *et al.* (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* 496, 87–90
115. Jayakodi, M. *et al.* (2020) The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature* 588, 284–289
116. 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
117. Golicz, A. *et al.* (2016) The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat. Commun.* 7, 13390
118. Tao, Y. *et al.* (2021) Extensive variation within the pan-genome of cultivated and wild sorghum. *Nat. Plants* 7, 766–773
119. Li, Y. *et al.* (2014) De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat. Biotechnol.* 32, 1045–1052
120. Liu, Y. *et al.* (2020) Pan-genome of wild and cultivated soybeans. *Cell* 182, 162–176
121. Hübner, S. *et al.* (2019) Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nat. Plants* 5, 54–62
122. Montenegro, J.D. *et al.* (2017) The pangenome of hexaploid bread wheat. *Plant J.* 90, 1007–1013
123. Hajjar, R. and Hodgkin, T. (2007) The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica* 156, 1–13
124. Bohra, A. *et al.* (2016) Cytoplasmic male sterility (CMS) in hybrid breeding in field crops. *Plant Cell Rep.* 35, 967–993