



# Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation

Joshua N. Cobb<sup>1</sup> · Roselyne U. Juma<sup>1,2</sup> · Partha S. Biswas<sup>1,3</sup> · Juan D. Arbelaez<sup>1</sup> · Jessica Rutkoski<sup>1</sup> · Gary Atlin<sup>4</sup> · Tom Hagen<sup>5,6</sup> · Michael Quinn<sup>5,6</sup> · Eng Hwa Ng<sup>5,6</sup>

Received: 6 October 2018 / Accepted: 21 February 2019 / Published online: 1 March 2019  
© The Author(s) 2019

## Abstract

**Key message** The integration of new technologies into public plant breeding programs can make a powerful step change in agricultural productivity when aligned with principles of quantitative and Mendelian genetics.

**Abstract** The breeder's equation is the foundational application of quantitative genetics to crop improvement. Guided by the variables that describe response to selection, emerging breeding technologies can make a powerful step change in the effectiveness of public breeding programs. The most promising innovations for increasing the rate of genetic gain without greatly increasing program size appear to be related to reducing breeding cycle time, which is likely to require the implementation of parent selection on non-inbred progeny, rapid generation advance, and genomic selection. These are complex processes and will require breeding organizations to adopt a culture of continuous optimization and improvement. To enable this, research managers will need to consider and proactively manage the, accountability, strategy, and resource allocations of breeding teams. This must be combined with thoughtful management of elite genetic variation and a clear separation between the parental selection process and product development and advancement process. With an abundance of new technologies available, breeding teams need to evaluate carefully the impact of any new technology on selection intensity, selection accuracy, and breeding cycle length relative to its cost of deployment. Finally breeding data management systems need to be well designed to support selection decisions and novel approaches to accelerate breeding cycles need to be routinely evaluated and deployed.

---

Communicated by Lee Hickey.

---

Joshua N. Cobb and Roselyne U. Juma are co-first authors.

---

✉ Joshua N. Cobb  
j.cobb@irri.org

- <sup>1</sup> International Rice Research Institute, Los Banos, Laguna, Philippines
- <sup>2</sup> Kenya Agricultural and Livestock Research Organization, Nairobi, Kenya
- <sup>3</sup> Bangladesh Rice Research Institute, Gazipur, Bangladesh
- <sup>4</sup> Bill and Melinda Gates Foundation, Seattle, WA, USA
- <sup>5</sup> CGIAR Excellence in Breeding Platform (EiB), El Batan, Mexico
- <sup>6</sup> International Maize and Wheat Improvement Center (CIMMYT), El Batan, Mexico

## Introduction

In the face of climate change, annual reductions in arable land, and localized malnutrition (Ritchie et al. 2018), plant breeding will play an essential role in feeding 9 billion people sustainably by 2050 (Godfray et al. 2010). However, many public-sector plant breeding programs in both developed and developing countries have struggled to keep pace with technological change since the Green Revolution (Jain 2010; Pingali 2012; Baranski 2015). Relatively few reliable estimates of the rate of genetic gain delivered by public plant breeding programs that serve most farmers in the developing world have been published, but these report rates of well under 1% annually under conditions faced by most small-holder farmers (Lopes et al. 2012; Masuka et al. 2017), and do not appear to have increased in the last 30 years.

Advances in statistics (Drovandi et al. 2017), quantitative and population genetics (Hill 2010; Walsh 2014), molecular biology (Moose and Mumm 2008), genomics (Dekkers 2012; Bevan and Uauy 2013; Barabaschi et al. 2016),

phenomics (Finkel 2009; Cobb et al. 2013; Lobos et al. 2017), other -omics (Carnielli et al. 2015; VanEmon 2016), and most recently machine learning and artificial intelligence (Singh et al. 2016; Chlingaryan et al. 2018) offer the potential of transforming plant breeding programs toward a data-rich, evidence-based, and team-oriented process and away from the romantic tradition of an individual breeder as an artist. However, applying these new technologies to the problem of increasing the rate of genetic gain delivered in farmers' fields will require careful attention to their impact on the parameters of the breeder's equation.

The reduced cost and increased throughput of data acquisition means that the primary challenge facing plant breeding teams in the 2020s will not be access to modern technology [though in some developing countries that may still be the case (Badiane 2017)], but rather developing a framework for assembling relevant technological options into an optimized product development pipeline. This manuscript is intended to provide a framework to help breeding teams make wise investments by using the genetic gain equation (Lush 1937) as a guide. Commonly known as “the breeder's equation,” the expression is a useful starting point because it articulates the parameters breeding teams manipulate as part of the crop improvement process. The equation itself is fully described by multiplying the additive genetic variation within the population ( $\sigma_a$ ), selection intensity ( $i$ ), and selection accuracy ( $r$ ). Eberhart (1970) later introduced the number of years per cycle ( $L$ ) into the denominator as a way to evaluate efficiency by expressing the response to selection as change over time.

$$\Delta G = (\sigma_a)(i)(r)/L$$

The equation serves as a useful mental framework for considering investment priorities because it elegantly distills theory down to parameters that a breeding program aims to manipulate. With the increased complexity of plant breeding programs, in terms of the genetic and phenotypic information acquired to make selection decisions, there are important implications for the organization, management, and incentivization of breeding teams that also need to be considered. Some aspects of breeding team management will also therefore be discussed.

### Product profiles as a variety replacement strategy

Before a breeding pipeline can be designed or optimized, it must have very clear objectives with respect to the type of product it will produce. Many breeding programs refer to this as the *product profile*. The product profile describes the trait package needed to replace a specific reference or target variety. A product profile, defined by Ragot et al. (2018), is “a set of targeted attributes that a new plant variety is expected to meet in order to be successfully released onto

a market segment.” Rather than an imaginative venture to design the perfect variety, a product profile attempts to focus breeding efforts on the key traits that drive incremental value creation. While a product profile could take several forms, one simplified approach is to create a document that (1) identifies an existing reference variety already grown by a majority of farmers in a region/market, (2) evaluates what farmers, consumers, and other value chain actors like about that variety, and (3) articulates the complaints key stakeholders have about the reference variety. Opinions and preferences of stakeholders can be queried via surveys and focus group discussions, and may be distilled into the product profile, a convenient list of “must-have” traits and “value-added” traits that allow a breeder to meet market demands, add value, and deliver an incrementally improved variety quickly, rather than delaying time to market by many years in an attempt to create the ideal variety. For example, a new rain-fed rice variety for South Asia might aim to replace the dominant variety *Swarna*. The product profile might specify that the new product should be 20 days shorter in duration than *Swarna*, have resistance to predominant rice blast races, out-yield *Swarna* by 10% under favorable conditions and 20% under drought conditions, and have the same grain size, shape, amylose content, and texture. This description provides breeding teams with most of the guidance they need to design the breeding pipeline and acquire needed traits. It should be noted that for many traits (e.g., yield) the product profile does not specify an absolute target but rather an advantage, in a particular target population of environments (TPE), over the variety to be replaced. This means that that variety must be used as a check in agronomic testing, and provides some guidance on the extent of testing required to detect the targeted advantage at a particular level of acceptance probability.

Based on the product profile, breeding teams can make transparent advancement decisions when a potential new variety meets all the “must-have” criteria and possesses at least one “value-added” trait that differentiates it in the marketplace. The product profile provides breeding teams a starting point for setting investment priorities that are aligned with the perspectives and opinions of multiple value chain actors. Updating of product profiles should be done regularly and involve as many diverse stakeholders as possible, but should always focus on identifying the reference variety, its attributes that must be retained in a new variety, and the opportunities for value addition. For a more thorough review of using product profiles to enable breeding decisions, see Ragot et al. (2018).

In most private-sector breeding programs, the product profile is designed by a marketing team, with critical feedback from business units, sales agronomists, and marketing departments to help translate feedback on product performance into tangible trait targets that drive adoption

and align new varieties with emerging needs expressed by the farmer, processor, or consumer. Public-sector breeding institutes have not traditionally had business development units or marketing departments and therefore have relied informal relationships with customers to achieve a good understanding of the constraints faced by farmers in the target region. Without consistent and accurate feedback indicating whether breeding objectives are drifting off track, a public-sector breeder typically takes at best an academic approach or at worst a speculative approach to determining trait targets and breeding strategy (Laske et al. 2012). In an ideally, funded organization articulating well-informed breeding objectives involves market research, close and regular interaction with key stakeholders, and an understanding of the challenges associated with breeding for various characteristics. In some cases, social science teams at institutions such as CGIAR centers have developed protocols for assigning relative values to traits in interactions with particular subsets of farmers, but this has rarely carried through to the design of formal product profiles. The participatory plant breeding (PPB) and participatory varietal selection (PVS) movements, which brought farmers into breeding trials and nurseries to make selections, were an effort to provide some farmer preference and requirement information into the selection process, but relying heavily on farmer selection is neither efficient nor particularly effective, since farmer feedback was obtained only for those traits visible at the time and place of the exercise, and do not incorporate the requirements of millers, urban consumers, and other stakeholders (Atlin et al. 2001). Well-constructed and frequently updated product profiles, designed in consultation with men and women farmers, marketers, processors, and end users, distill the requirements of all stakeholders into a blueprint for varietal development.

### Resource allocation and breeding pipeline optimization

Once breeding targets have been set and the product profile developed with the input of key stakeholders, the design and optimization of a pipeline to deliver the profile are the responsibility of the plant breeder. Because most important agronomic and quality traits are polygenically inherited, only modest and incremental improvements in these traits can be made in each breeding step. Breeding for quantitative traits is best conceived as an iterative process, in which each generation incrementally increases the frequency of favorable alleles in the gene pool under selection, thereby increasing the probability of extracting a superior cultivar. Below, we consider the steps in optimizing a breeding pipeline.

## Additive genetic variation

### Managing and maintaining genetic variance

The first step in establishing a breeding pipeline is the selection of elite parents as founders of the program. Elite germplasm can be defined as a reproductively compatible set of genotypes disproportionately enriched for favorable alleles that improve breeding value (i.e., the mean performance of the progeny of a given parent) in a particular environment or market. Breeding values are used regularly in the context of animal breeding since the breeding product is not a sire itself, but rather its progeny. A breeding value uses pedigree or genome-wide marker data to borrow information from related lines in a phenotypic data set to estimate the additive value of an individual. While a BLUP value for phenotypic performance accounts for both the additive and nonadditive genetic values of a line, a breeding value uses the relationship matrix to determine the additive value of a line, which is the primary source of genetic variance passed on to its offspring (Henderson 1976). This is critical information for parental selection decisions and determining the relative “eliteness” of a line, and is under-utilized in public plant breeding programs. For many plant breeders, the term elite is used indiscriminately without specific terms of reference to the purpose of characterizing and managing genetic diversity in a plant breeding program. “Elite” is a status best inherited from elite parents or earned through rigorous testing of the individual per se and related lines.

Selection theory suggests that the process of generating genetic gain in source populations should be treated separately from the process of product extraction. Doing so allows a breeder to manage genetic diversity, set the desired rate of genetic gain, rapidly increase the frequency of favorable alleles in source populations, and extract products from those populations as often as needed to serve the needs of farmers or stay competitive in the marketplace. In reviewing how genomic selection unifies plant and animal breeding, Hickey et al. (2017) point out that plant and animal breeding have taken very different theoretical approaches to managing elite genetic variation. They highlight that while many plant breeders focus on the intentional introgression of valuable alleles in a single line, animal breeders have practiced a quantitative population improvement approach, wherein parents of each generation are selected on the basis of high additive breeding value, determined either by pedigree or genomic methods. As a result, plant breeders often have a tendency to conflate the genetic improvement of source populations with the process of extracting commercial products from those populations in a single step that focuses on obtaining a specific complex haplotype through the manipulation of Mendelian genetics. Achieving this complex haplotype (say, a combination of several disease resistance alleles

and QTLs for abiotic stress tolerance) in high frequency in a segregating population in a single step can require very large population sizes. Also, many plant breeders use complex crossing schemes with exotic or diverse materials in an attempt to bring together disparate traits of interest. This comes at a genetic cost as the constant introduction of novel alleles, the crossing with older or exotic material with low breeding value, and the extended breeding cycles from complex backcrossing reduce the response to selection per year relative to what could be achieved through selective mating of parents with high breeding value in a closed population. As such, genetic variance for an economically important trait within clearly elite breeding material is much more valuable to a breeder than is genetic variation per se. Recurrent selection schemes focused on recycling the best genetics in a closed breeding system have been demonstrated to produce high rates of genetic gain (Bressegello et al. 2009; Shelton et al. 2015), and some of the most successful plant breeding programs in the private-sector currently manage their genetic diversity in this way (Smith et al. 2015). These programs either impose strict rules about parent choice or use a closed recurrent selection approach without overlapping generations to ensure that the parents of each new breeding cycle have higher additive breeding value than the last. This is critical to ensuring that genetic gain will be achieved in each breeding cycle.

One common concern among plant breeders when confronted with this approach is that the genetic base may be too narrow, thus permitting short-term gain from selection, but at the expense of long-term progress, arguing that while the gene pool becomes increasingly elite, it also becomes increasingly inbred. Fortunately, for quantitative traits governed by complex additive genetic architectures (Boyle et al. 2017), it has been shown that even starting from a relatively small effective population size, many decades of breeding progress can still be achieved (Guzman 1998; Bernardo et al. 2006). Inbreeding is, of course, inevitable as favorable alleles become fixed in the gene pool. However, new mutations replenish genetic variance lost to allele fixation and are an important source of genetic variance for selection over long timescales (Lind et al. 2018).

High rates of genetic gain have been reported in many experiments in a range of species using closed recurrent selection schemes. In maize, Coors (1999) reviewed over 90 recurrent selection experiments and found that gains per year from reciprocal recurrent selection exceeded those produced by commercial hybrid breeders. Recurrent selection has also been very effectively used in rice, notably by the EMBRAPA upland and lowland rice breeding programs in Brazil. Bressegello et al. (2009) reported gains of 3.6% per year over three cycles of recurrent selection in upland rice. In irrigated lowland rice, Júnior et al. (2017) reported, over three cycles, an average rate of yield gain per year of

1.98%. A recurrent selection strategy for managing elite genetic diversity not only enables evolutionary processes that increase the frequency of favorable alleles over time (Falk 2010), but also permits several clear applications of modern technologies to the breeding process. Notably, the application of whole genome sequence information allows for very clear inferences to be made about the elite breeding germplasm and permits the systematic tracking and evaluation of genetic variation over successive cycles of breeding (Beissinger et al. 2014). A key step toward modernization is to gather all available data and rigorously define a core set of high-performing elite lines. Once this is achieved, a systematic characterization of the resulting material, including the generation of high-density genotype information, will create a framework for understanding, tracking, and managing elite genetic diversity.

### Application of high-density marker data in a recurrent selection program

While the topic is worthy of its own review, the full application of high-density genotype data to a plant breeding program would include characterization of genetic variation (Wright 1931, 1933; Wang et al. 2016; Prieur et al. 2017), development of a framework for tracking identity by descent (Vela-Avitúa et al. 2015), validating the accuracy of trait markers for MAS (Li et al. 2013; Javid et al. 2015), imputation of marker information from low-density data sets (Pryce et al. 2014; Moghaddar et al. 2017), haplotype phase inference (Hess et al. 2017), and the application of optimum contribution selection to crossing decisions (Meuwissen 1997; Wang et al. 2017). For brevity, only monitoring and managing genetic variance and validating trait markers will be touched upon here, but breeding programs that manage genetic diversity carefully and rapidly recycle parents with high breeding values are well positioned to leverage insight from these data sets to appropriately manage genetic diversity such that genetic gain is maximized while the long-term erosion of genetic variance is minimized.

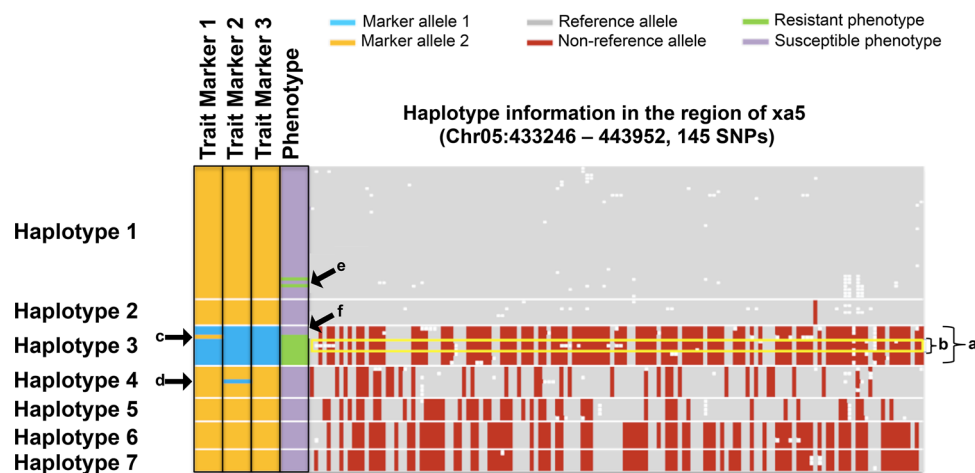
Effective population size ( $N_e$ ), a concept introduced by Wright (1931) and further refined by Crow and Kimura (1970), is one of the most commonly used indicators in population genetics and evolutionary biology for describing the rate of inbreeding in a population and assessing genetic diversity. Plant breeders should know the effective population size of their program and actively monitor it through successive cycles of breeding to ensure the long-term viability of the breeding effort. The smaller the effective population size, the faster the population will become inbred and thus no longer respond to selection. While  $N_e$  can be estimated using both pedigree and marker data (Fernández et al. 2005), the latter is preferred (Wang 2016). Knowledge of  $N_e$  helps design efficient selection



schemes in both animal and plant breeding (Caballero et al. 1991) and based on the effective population size breeders can redesign (if necessary) the parental combinations to maintain sufficient genetic variance among future selection candidates. One strategy for using marker data to preserve the capacity for long-term genetic gain used extensively in animal breeding but with still limited use by plant breeders is optimum contribution selection (OCS) (Meuwissen 1997; Grundy et al. 1998; Sonesson and Meuwissen 2000; Hallander and Waldmann 2009). As opposed to a static parental selection and crossing strategy which does not consider the value of differential parental contribution to the next generation, OCS seeks to balance gain from selection with preventing erosion of genetic variance by employing a dynamic parental selection strategy that seeks to optimize the contribution of each selection candidate in the next breeding cycle consistent with its predicted contribution to the rate of inbreeding in the next cycle. Sonesson and Meuwissen (2000) suggested that non-random mating among the selected parents might improve the family genetic variances in the progenies of next generations and thus affect the results of subsequent selection. OCS schemes in non-random mating populations can reduce rates of inbreeding (Caballero et al. 1996), control the levels of kinship among the progeny (Caballero et al. 1996; Fernández and Toro 1999; Toro and Pérez-Enciso 1990), increase the rate of genetic response (Caballero et al. 1996; Wang et al. 2017) and ultimately help maintain genetic diversity in the selection candidates (Wang

et al. 2017) through restricting the average co-ancestry of the population. As plant breeding programs shift toward more quantitative approaches, the proactive exchange of elite material and monitoring and enhancement of genetic variation will become increasingly important.

In addition to monitoring and managing the effective population size of the program, another key advantage to applying high-density marker information in a recurrent selection breeding scheme is that it provides clear information about the haplotypic diversity available within an elite gene pool, and the association of that diversity with key trait markers targeted for marker-assisted selection. Reviewed by Cobb et al. (2018) in a companion paper, the effective application of MAS to a breeding program depends on having accurate marker systems capable of properly identifying QTL[+] and QTL[-] lines. The use of diagnostic markers based on the functional polymorphism is obviously ideal (Barr 2009), but many times such knowledge is unavailable. Using linked markers then becomes necessary, but the efficacy of a linked marker depends on linkage phase, allele frequency, genomic region, and the nature of the QTL interval. With sequence data available on lines representing the key pedigrees within a breeding program, haplotypes can easily be determined and breeding program specific error rates for markers of potential value can be determined (see Fig. 1). Quality metrics for assessing the value of a marker in the context of an elite breeding program had recently suggested by Platten et al. (2019).



**Fig. 1** Accuracy determination of three linked SNP markers for a bacterial leaf blight resistance gene (xa5) among resequenced lines in the IRRI Irrigated breeding program. **a** The haplotype of the validated donor line indicates the known resistant haplotype. **b** IRBB 60, IRBB 61 and IRBB 64 as validated trait donors for the resistant xa5 allele. **c** A breeding line with a negative marker score for trait marker 1, but harboring the QTL[+] haplotype (i.e., a false negative

for marker 1). **d** A breeding line with a positive marker score for trait marker 2, but possessing a QTL[-] haplotype (i.e., a false positive for trait marker 2). **e** Line exhibiting a favorable phenotypic response from another unlinked locus or phenotyping error. **f** Seed source variation (error) between the phenotyped source and the sequenced source showing susceptible phenotypes among the QTL[+] haplotype group

## Germplasm enhancement and pre-breeding

Careful management of genetic variation, as important as it is, will not compensate for when heritable variation is simply not present among elite lines—for example when a major disease resistance allele is not present in the elite gene pool. In these cases, pre-breeding or germplasm enhancement is necessary and careful strategies need to be deployed in order to introgress the needed variation without unnecessarily contaminating the elite breeding gene pool with parents of average or low value. The need to focus this effort and to resource it independently is discussed above, but some tactical considerations for pre-breeding activities are worth noting.

The most common approach to pre-breeding is to use genetic mapping to identify QTL for key phenotypic variation and to introgress those QTL into the elite gene pool. While genetic mapping is a powerful strategy to resolve complex traits into discrete Mendelian units (quantitative trait loci or QTL; Barr 2009), it is important to recognize that the effect size of the QTL affects the deployment strategy substantially. QTLs with small effects rarely warrant marker-assisted introgression. The expense of fine mapping and marker development for small-effect QTL is rarely worthwhile when they constitute only a small proportion of the genetic variation for the trait in existing breeding germplasm (Hill et al. 2008; Boyle et al. 2017). For many species, the primary value of exotic genetic variation is the identification and deployment of rare alleles with large effects that can be introduced into elite breeding programs via a thoughtful implementation of marker-assisted selection (see Cobb et al. 2018 for a more in-depth discussion on QTL deployment).

## Selection intensity

### New technologies for increased selection intensity

In terms of the breeder's equation, generating and testing more selection candidates while holding the number of selected candidates constant lead to higher selection intensity ( $i$ ) which in turn increases the rate of genetic gain. Selection intensity can also be increased by selecting fewer parents; however, it is usually more advisable to determine the number of parents to select based on whether the objective of the breeding program is long- or short-term genetic gain (Bernardo et al. 2006). Thus, increasing  $i$  by way of increasing population sizes requires that either budgets be increased, or a reduction in the cost of testing each selection candidate.

As budgets are usually fixed, several new technologies have been proposed to help reduce the cost of testing as a way to increase  $i$ , thereby increasing gain from selection. For example, genomic selection proposed by Meuwissen

et al. (2001) could be used to increase the total number of selection candidates with a fixed budget if genotyping is less costly than phenotyping. Further, sparse testing designs, wherein individual lines are unreplicated or partially replicated across locations, but relatives are randomized among locations to allow estimates of haplotype  $\times$  environment effects, can reduce the replication of selection candidates within and across environments. This reduces field costs and would allow a larger number of selection candidates to be tested (Endelman et al. 2014; Roorkiwal et al. 2018). Simulation studies by Lorenz (2013) and Riedelsheimer et al. (2013) found that the application of genomic prediction generally led to greater response to selection because phenotyping all selection candidates, even at reduced levels of replication, increased both the accuracy and intensity of selection.

Aside from genomic selection, low-cost screening of large population sizes prior to more advanced yield testing based on conventional MAS and/or phenotyping remains useful and can lead to high  $i$  for specific traits such as disease resistance. However, it is possible that selection imposed in early generations or prior to more extensive phenotyping could actually reduce gain from selection for traits evaluated in more advanced stages of testing such as grain yield and quality due to unfavorable genetic correlations between traits evaluated in early generations and traits evaluated subsequently. For example, intense selection for early flowering in early generations would actually lead to less response to selection for grain yield in subsequent stages due the effect that direct and indirect selection has on reducing the genetic variance (Bulmer 1971) prior to inter-mating. Additionally, intense selection based on traits or markers prior to advanced testing could reduce the selection candidate population from thousands to hundreds, substantially reducing  $i$  and gain from selection for traits tested at advanced stages. Thus, large-scale screening and selection prior to advanced testing should be done with careful consideration of the impacts it will have on the overall gain from selection for all traits of interest. It should be noted that in some breeding programs where high priority “must-have” traits are at low frequency among elite lines, it may be worth sacrificing the highest yielding individuals, in order to increase the frequency of such a trait in the short term before refocusing on agronomic performance.

In addition to marker-based solutions to increasing selection intensity, one emergent technology that could enable greater gain from selection by increasing  $i$  is high-throughput phenotyping (HTP). Rapid and low-cost data capture using sensors can be converted into variables that may be useful as secondary traits for indirect selection either on their own or as part of pedigree or genomic prediction schemes. Studies by Rutkoski et al. (2016) and Sun et al. (2017) found that pedigree and genomic prediction accuracies for grain

yield in wheat approached those of direct phenotypic selection when data on secondary traits measured using an aerial HTP platform were utilized. This suggests that it may be possible to increase gain from selection with a fixed budget by increasing the number of selection candidates planted in yield trials, phenotyping secondary traits using HTP, and only harvesting a portion of the yield trial. Such an approach may be especially advantageous where the cost of harvesting is high, such as when trials are conducted in remote locations and/or when the harvesting must be done manually.

It should be noted, however, that genetic gain is not linearly related to the number of selection candidates. Increasing the number of selection candidates tenfold, from 10 out of 100 to 10 out of 1000, increases standardized selection differential, and therefore genetic gain, by only 52% (Table 1). The difficulty and expense of trying to increase genetic gain through increased selection intensity relative to reduced cycle time are discussed in more detail later.

### Selection accuracy

Phenotyping is currently (and is likely to remain) the most expensive component of a plant breeding operation. HTP may have some applications for helping to increase selection intensity, but may have more important effects on enhancing selection accuracy. The value of improvements in phenotyping is usually expressed by citing increases in broad sense heritability ( $H^2$ ). While this metric is useful for comparing two phenotyping strategies, it is important to recall that genetic gain is proportional to the genetic accuracy, which is the square root of the narrow-sense heritability ( $h^2$ ). This has big implications for deciding how to invest a breeding program's limited resources. For most breeding programs, the best way to increase heritability is to better sample the targeted population of environments by increasing the number of yield trial locations. This turns out to be a very expensive option and is limited by physical capacity and partnerships as much as it is by budgets. Thus, most innovations in phenotyping have focused on extracting more information from existing yield trials. This can lead to making large expenditures in capital equipment and/or digital devices, and as

such, careful thought needs to be put into a phenotyping strategy for a breeding program (Cobb et al. 2013).

### HTP and the digitization of data collection

Breeding programs in recent years have witnessed an influx of engineering solutions such as robotics, imaging systems, and unmanned vehicles to increase the throughput of phenotyping (Tattaris and Reynolds 2016; Awada et al. 2018). These solutions have provided significant cost savings and efficiency gains to breeding programs via automation of routine processes, addition of new data types for more unbiased selection, and the scaling up of data collection processes at a reduced cost. While these technologies are the focus of several reviews, one unifying principle across them that is seldom discussed is the value of digitized data collection. Digitization of breeding data, both phenotypic and genotypic, is crucial for breeding programs to scale up in the twenty-first century. With cost of computing power decreasing exponentially, it has enabled genotyping data (already in digital form) to be used with ease and minimal manual curation. However, this is not the same for phenotypic data as many public plant breeding programs to date are still relying on paper and pen to record and subsequently transcribe data into usable formats (Rife and Poland 2015). Manual data collection strategies like this render phenotypic data collection expensive, labor-intensive and error-prone. Mackay and Caligari (1999) went so far as to simulate the effect of common typographical errors on the rate of genetic gain and found significant reductions in response to selection for error rates as low as 1%. Paradoxically, they showed that in many cases, response to selection would actually increase if selection intensity is reduced simply because misrepresented data points are disproportionately found among selected candidates. The idea to digitize phenotypic data collection in plant breeding was discussed as early as the 1990s by Berke and Baenziger (1992), but due to the high cost of field-ready electronic devices at the time and lack of technical skills to deploy them, adoption rates of digital field data collection remained low until the smartphone revolution of last decade. The advent of high-performance computing that fits in the palm of our hand has led many teams to develop software

**Table 1** Relationship between proportion selected, standardized selection intensity ( $i$ ), and genetic gain

Effective population size ( $N_e$ )	Proportion selected	Total population	Standardized selection differential	Genetic gain relative to $N = 100$
10	0.1	100	1.75	1
10	0.05	200	2.063	1.18
10	0.01	1000	2.665	1.52
10	0.005	2000	2.892	1.65
10	0.001	10,000	3.367	1.92

that permits the digital collection of phenotype data in the field. One such example of this is the PhenoApps (<http://phenoapps.org/apps/>). The PhenoApps tools are completely open-source Android-based applications that aid breeding programs in digitizing data collection and maintaining the chain of custody (Rife and Poland 2015). Additional software tools include Phenobook (Crescente et al. 2017) and the PhenoTyper (Köhl and Jürgen 2015).

### Breeding data management

It should be intuitively recognized that digitizing the collection of data is actually a fairly easy problem to solve relative to the grand challenge of bringing all sources of breeding data together to inform selection decisions. Platforms need to be built that bring phenotype data, genotype data, pedigree data, and ultimately climate and weather data together in interoperable ways that permit the analysis of multi-year, multi-location, and multivariate information on generations of related lines. Sophisticated breeding informatics systems must be created and incorporated into breeding workflows to enable better predictions on the performance of different genotypes across various environments. These systems need to be robust in their architecture, computationally powerful, accessible even in remote locations, and user-friendly. There are currently very few systems available to public-sector breeding programs at the time of this publication fully capable of accomplishing this task, and no single system is in widespread use across public plant breeding programs; however, Rathore et al. (2018) provide a thorough overview of some of the strongest contenders. A fully functional and integrated system should seamlessly: (1) define, display, and permit the revision of distinct breeding zones, their target markets, and corresponding product profiles for more accurately selecting and advancing germplasm, classifying trial locations, and advancing selection candidates against defined criteria; (2) permit simplified workflows for trial design, trial management, and phenotyping procedures that allow breeding teams to quickly and easily implement robust statistical designs; (3) employ yield trial analytics with phenotypic spatial correction, location data quality evaluations, multi-location analytical capabilities, and the generation and display of pedigree estimated breeding values (pEBV) and genomic estimated breeding values (gEBV); (4) intuitively integrate pedigree, genotype, phenotype, and climate information across years, geographies, and generations to more accurately predict, summarize, and interpret variance components such as genotype, location, year, agronomic management regimes, and their interaction terms; (5) simply, intuitively, and powerfully permit creating and tracking the progression of germplasm through the breeding program;

and (6) effectively track breeding operations, budgets, and activities at different timescales.

Such systems would be central and formative to breeding teams endeavoring to efficiently conduct their operations, make selection decisions, optimize resources, summarize outputs, and explore the genetic consequences of breeding decisions. In order to maintain flexibility, avoid obsolescence, and leverage extensibility, these systems should ideally be modular in design with advanced data models, generic, extendable, and scalable software logic, and built using an architected approach with clear specifications. Figure 2 illustrates some suggested data and analysis modules that, together, could fully enable breeding operations.

### Cycle time

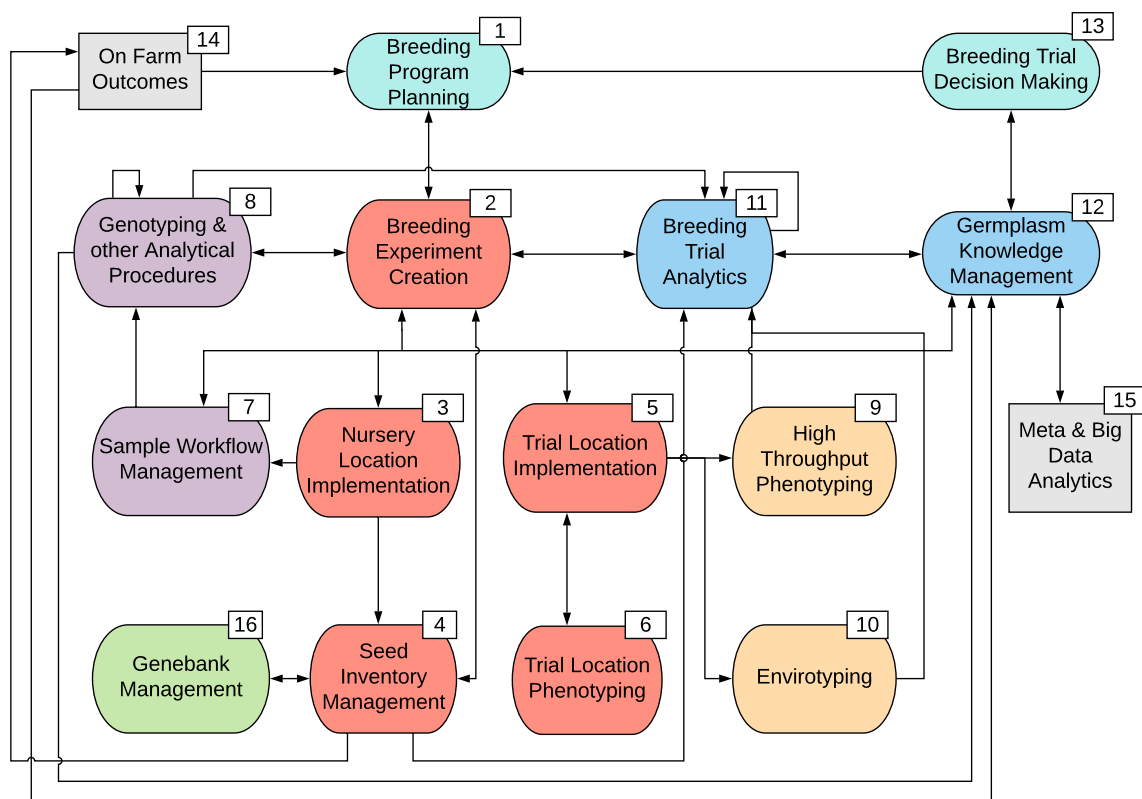
Given the complexity of the other parameters in the breeder's equation, cycle time is the easiest to understand, cheapest to manipulate, and the most powerful parameter for increasing genetic gain. Cycle time (or generation interval) involves recycling breeding material back into the crossing block as quickly as a breeder can determine that a genotype is above average in breeding value for a desired quantitative trait. Despite its simplicity, manipulating cycle time requires careful planning and consideration of breeding strategy because a breeding team can move as quickly in the wrong direction as they can in the right one.

### Rapid generation advance for inbred and hybrid crops

For many inbred and hybrid crops, public-sector plant breeding has maintained a strong legacy of pedigree breeding strategies focused on heavy visual selection during segregating generations. While effective for the identification of semidwarf progenies during the Green Revolution, visual selection during inbreeding is no longer effective when plant type is fixed (Atlin et al. 2017). In self-pollinated crops, approaches for rapidly developing fixed lines, such as doubled haploid (Maluszynski et al. 2003; Asif 2013) or single seed descent (Collard et al. 2017) methodologies, effectively partition genetic variation exclusively between lines, permitting the rapid evaluation and promotion of high-performing material. In general, modern breeding methods for self-pollinated crops should aim to develop fixed lines at the lowest possible cost and in the minimum possible amount of time.

Rapid generation advance (RGA) technique was first proposed by Goulde (1939) and later modified by Grafius (1965). Its most recent iteration has been in the form of "speed breeding." (Watson et al. 2018). An RGA system shortens generation cycle through modified environments and early seed harvest in  $F_2$ – $F_6$  generations (Collard et al. 2017). These methods are regularly reviewed, and new modifications are often published (Choo et al. 1985; Snape 1989;





**Fig. 2** Modular design of breeding system functional capabilities. Sky blue-colored modules address breeding strategy and objectives; orange boxes correspond to the management and creation of breeding experiments; purple boxes illustrate genotyping workflows; yellow boxes for modules enabling phenotypic and environmental data col-

lection; green boxes represent pre-breeding and gene bank management; blue boxes highlight modules for breeding analytics. Detailed explanations of numbered modules are provided in the body of the manuscript (color figure online)

Maluszynski et al. 2003; Forster et al. 2007; Touraev et al. 2009; Tadesse et al. 2012; Dwivedi et al. 2015; Humphreys and Knox 2015; van Ginkel and Ortiz 2018; Watson et al. 2018). Depending on the crop, the strategy, the growing season, and the budget, specialized facilities are sometimes needed to be able to accelerate generations and achieve the desired cycle time. The particular biology of the crop dictates the most successful approaches, but many techniques have been applied, including harvesting of immature seed in soybeans (Carandang et al. 2006) and pigeon pea (Saxena et al. 2017) and light manipulation in chickpea (Gaur et al. 2007), sorghum (Rizal et al. 2014) and rice (Tanaka et al. 2016). More recently, seven generations of oats and triticale had been developed in 1 year through RGA, (Liu et al. 2016) using stress to induce flowering followed by embryo rescue. Likewise Watson et al. (2018) had recently achieved six generations per year in wheat, barley, and chickpea.

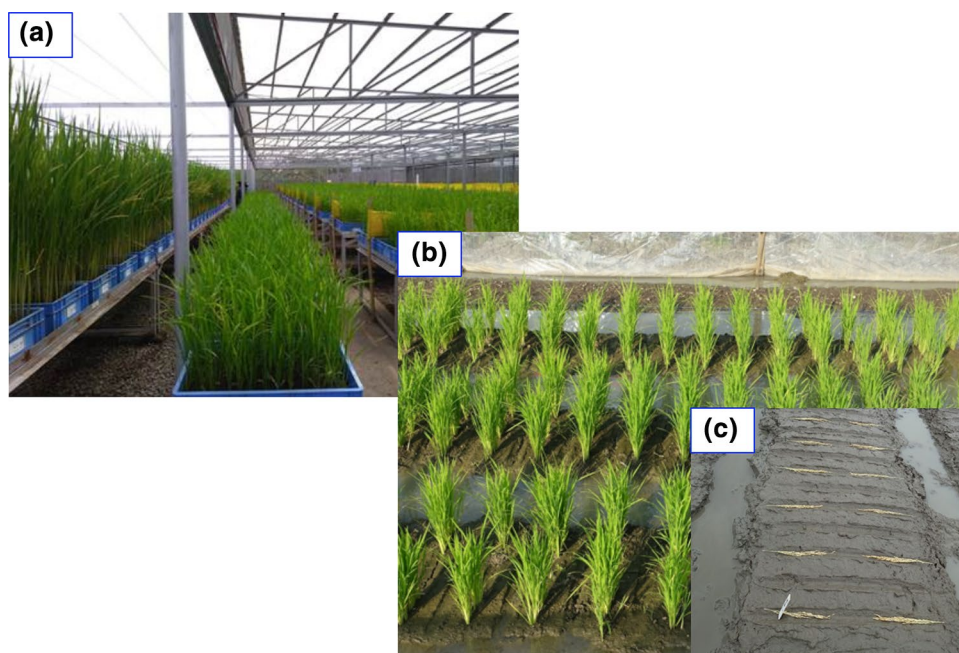
In addition to the speed, one of the primary advantages of RGA compared to other methods can be the reduction in the cost per new recombinant relative to pedigree breeding, which has a very high land and labor requirement for each fixed line generated. Embryo rescue, growth chambers, and

other technologies with a high cost of entry or skilled labor requirement will be most successful where high volumes and economies of scale can spread the cost among many programs. For individual breeding programs deploying RGA as a single seed descent method, usually a greenhouse facility or even a field RGA will suffice. Such a setup was developed at the International Rice Research Institute (IRRI), reducing the cost of line fixation from hundreds of dollars (under pedigree selection) to USD \$0.74 using a greenhouse-based RGA and USD \$0.29 using field RGA (Fig. 3; Collard et al. 2017).

### Emphasizing cycle time over other selection parameters

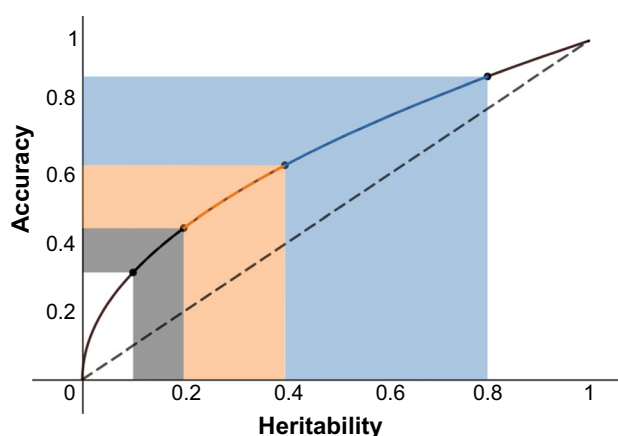
Accelerating breeding cycles can be the most efficient way to increase the rate of genetic gain, but it has been underexploited in most breeding programs, which have focused more on the other three parameters of the breeder's equation, namely heritability, selection intensity, and additive genetic variance. Unfortunately, manipulating these variables is effective initially, but they are subject to rapidly diminishing returns on investment and thus offer an expensive pathway

**Fig. 3** Rapid generation advance strategies at the International Rice Research Institute. **a** RGA greenhouse facility. **b** Field RGA nursery in vegetative stage. **c** Field RGA nursery at panicle seeding



to increased rates of genetic gain. In the case of heritability, once modest levels are achieved (in the range of 0.5–0.7) it is rarely worth the often substantial investment required to secure additional testing locations needed to increase it further. Heritability does not scale linearly with replication (i.e., doubling the number of yield trial locations will not double the heritability). Compounding this effect is the fact that neither does genetic gain scale linearly with heritability (i.e., doubling the heritability does not translate to double the response to selection). Genetic gains are proportional to the accuracy, which is the square root of heritability, and thus scales according to a square root function with heritability (Fig. 4). Thus, increasing heritability from 0.6 to 0.8 comes at significant cost but only increases the rate of genetic gain by  $\sqrt{(0.8/0.6)}$ , or 15%. An even more extreme form of diminishing returns affects efforts to increase the rate of genetic gain by increasing population size and selection pressure. Increasing selection intensity tenfold (i.e., requiring a tenfold increase in the size of the population generated and screened if effective population size ( $N_e$ ) remains constant) will increase the standardized selection differential (and therefore genetic gain), by only 52%, assuming the same heritability (see Table 1). A 52% increase in genetic gain for a tenfold increase in program size and cost is an untenable solution for most breeding programs.

The impact of reducing breeding cycle length for most programs is potentially much greater than increasing heritability above about 0.6, or reducing the selected proportion below 5%. While the RGA methods described above allow breeding cycles to be reduced to three or 4 years compared to pedigree selection, there is an opportunity for breeding cycles to be shortened much further, potentially to just 1 year



**Fig. 4** Diminishing returns on selection accuracy relative to increases in heritability. Since selection accuracy is expressed as a square root function (solid line) and not a linear function (dotted line) relative to heritability, linear increases in heritability ( $x$ -axis), which require significant financial investment, have diminishing impacts on selection accuracy ( $y$ -axis). Even modest heritabilities can command sufficient selection accuracy to drive genetic gain (orange shaded portion) (color figure online)

in the many annual cereal, oilseed, and legume crops that are not strongly photoperiod-sensitive. Most breeding programs in self-pollinated crops unnecessarily lengthen the breeding cycle by initiating selection for breeding value (i.e., selecting parents) only among highly homozygous lines. This adds significant time even relative to schemes that assess breeding value and select on non-inbred candidates. Most programs advance lines only two seasons per year and therefore would take 3 years to generate a fixed line with enough seed

to enter in a yield trial before choosing to recycle lines as parents.

The breeding cycle could be greatly accelerated by selecting  $F_{2,3}$  (or  $S_{0,1}$ ) lines for use as parents in a rapid-cycle recurrent selection program to improve the breeding value of a population, rather than selecting highly inbred lines that can be used directly as cultivars. This amounts to a partial decoupling of parent selection for, and commercial line extraction from, elite breeding populations. This is not a novel idea; most traditional phenotypic recurrent selection programs were based on recombination of parents from the first or second generation of self-pollination, rather than highly inbred lines that can be directly commercialized. In recent years, the development and simulation of genomic selection protocols have demonstrated that breeding cycles can be completely decoupled from commercial line development, by selecting parents purely on the basis of GEBV (Heffner et al. 2009; Gaynor et al. 2016), but it has not yet been widely demonstrated experimentally that this extreme decoupling of population advancement from phenotyping will permit effective estimation of breeding value in crop breeding programs. Recurrent selection involving crossing, field testing of non-inbred progeny, and recombination in a single year is amenable to genomic selection for breeding value, especially after several cycles have been completed and linkage disequilibrium increases among parental selection candidates due to the increased relatedness of the elite source population.

An aggressive but effective single-year recurrent selection program applicable to most self-pollinated species, including rice, was developed by K.J. Frey and his students in the 1970s (Frey et al. 1988), consisting of the following steps:

- (1) A closed breeding population is created by inter-mating a set of selected lines chosen for high breeding value and complementary defensive and quality traits; the initial population must undergo two rounds of inter-mating to generate segregating  $S_0$  progeny.
- (2)  $S_{0,1}$  lines are generated by allowing  $S_0$  plants to self-pollinate.
- (3)  $S_{0,1}$  lines are phenotyped for yield and other traits in agronomic trials. While in some species seed may be limited at this juncture, in many of the cereal species with high tillering capacity, enough seeds for several replicates of field testing could potentially be produced from a single plant.
- (4) The best-performing  $S_{0,1}$  lines are inter-mated.
- (5) Steps 2 through 4 are then repeated for several cycles.

A genetic simulation can be used to show that radically shortening the breeding cycle from 4 years to one will greatly increase selection response even if selection intensity and heritability are reduced. Using a deterministic genetic

gain simulation tool developed by Rutkoski (2018), different scenarios were simulated using the range of phenotypic values for yield from the IRRRI 2016 Observational yield trial of the Drought Breeding Network. Several scenarios demonstrate the viability of ultrashort breeding cycles to drive increased genetic gain:

- (1) A conventional breeding program involving selection among 1000  $F_{5,6}$  lines evaluated at  $h^2$  of 0.3, with  $N=10$  parents selected per cycle, and a breeding cycle of 4 years, over a single cycle.

This scenario represents a typical, if fairly aggressive, self-pollinated crop breeding program based on the selection of parents from among highly homozygous selection candidates.

- (2) A rapid-cycle (1 year per cycle) recurrent selection program involving selection among 100  $S_{0,1}$  lines per cycle, with  $N=10$ , evaluated at  $h^2=0.2$ , over four cycles.

This represents the fastest recurrent selection program possible in rice that still permits phenotypic evaluation each cycle. Narrow-sense heritability is assumed to be lower than the larger, conventional program posited in (1), because the quantity of seed per  $S_{0,1}$  line available for field phenotyping would likely be smaller than the quantity available from  $F_{5,6}$  lines, limiting replication over environments, and because the additive genetic variance among  $F_{5,6}$  lines is somewhat greater than that among  $S_{0,1}$  ( $F_{2,3}$ ) lines.

- (3) A rapid-cycle (1 year per cycle) recurrent selection program involving selection among 100  $S_{0,1}$  lines per cycle, with  $N=10$ , evaluated at  $h^2=0.15$ , over four cycles.

This scenario is identical to (2), but with an even more severe reduction in  $h^2$ .

The results (Table 2) indicate that rapid-cycle recurrent selection in a population is only 10% as large in each cycle as in the conventional program resulted in substantially greater genetic gains, even assuming considerably reduced heritability. Even with heritability only 50% as high as in the conventional 4-year program of a single cycle of selection in fully inbred progeny, the rapid-cycle RS program, over 4 years, resulted in 50% higher rates of gain per year than the conventional program.

A key and under-recognized element of rapid-cycle recurrent selection is that it delivers higher gains than long cycle, large population breeding at a much lower cost, because much smaller populations are evaluated each cycle and in total. In the current example, the conventional program

**Table 2** A comparison of simulated genetic gains based on three breeding scenarios

Metrics	One cycle, 4 years/ cycle, $h^2=0.3$	Four cycles, 1 year/ cycle, $h^2=0.2$	Four cycles, 1 year/cycle, $h^2=0.15$
Number of lines phenotyped per cycle	1000	100	100
Total genetic gain (kg ha <sup>-1</sup> )	455	954	686
Annual genetic gain (kg ha <sup>-1</sup> yr <sup>-1</sup> )	113.75	238.5	171.5
Genetic standard deviation units per year	0.32	0.82	0.68
% genetic gain per year	2.22	4.66	3.35

requires the generation and phenotyping of 1000 selection candidates during the single 4-year cycle. In the rapid-cycle recurrent selection programs, 100 selection candidates are generated and phenotyped per year, for a total of 400. Perhaps 10 sublines would be extracted from each of 10 selected parents at the second and fourth cycles, to develop 100 fixed lines as candidate varieties for potential commercialization every second year. In total, the 4 years of recurrent selection would involve the creation and testing of 600 lines compared to the 1000 lines for the conventional program, while delivering much higher rates of genetic gain. In oats, Frey et al. (1988) achieved yield gains of 5.4% per cycle and therefore per year, using this protocol. Less aggressive schemes requiring 2 or 3 years per cycle have been widely used in self-pollinated crops, usually generating gains per year that are at least as high as those achieved from pedigree breeding programs. For example, Payne et al. (1986) achieved gains of 3.8% per cycle or 1.28% per year over three cycles, using a protocol requiring 3 years per cycle in oats. In soybean, Kenworthy and Brim (1979) evaluated a plan identical to that used in oats by Frey et al. (1988), but with cycles completed over 2 years. Over three cycles, an average gain of 5% per cycle or 2.5% per year was achieved for grain yield.

Shifting away from the long cycles inherent in pedigree breeding strategies toward a more accelerated 3–4-year cycle based on rapid generation advance (for inbred and hybrid crops) has the potential to make a step change in the rate at which varietal improvement is occurring. Further accelerating that by abandoning inbreeding to near fixation before selection for breeding value and redeploying the ultra-rapid single-year recurrent selection cycles will enhance the productivity of plant breeding programs by yet another step.

## Breeding program management

Successful modern breeding programs are complex and moderately expensive scientific enterprises, wherein products are developed and delivered by teams charged with delivering a steady stream of incrementally improved varieties that meet the market and farmer requirements captured in the product profile. It is a critical function of the

management of breeding organizations to properly incentivize, support, and monitor the effectiveness of these teams, and to instill a culture of continuous improvement wherein the breeding pipeline is constantly assessed and new methods integrated as they become available.

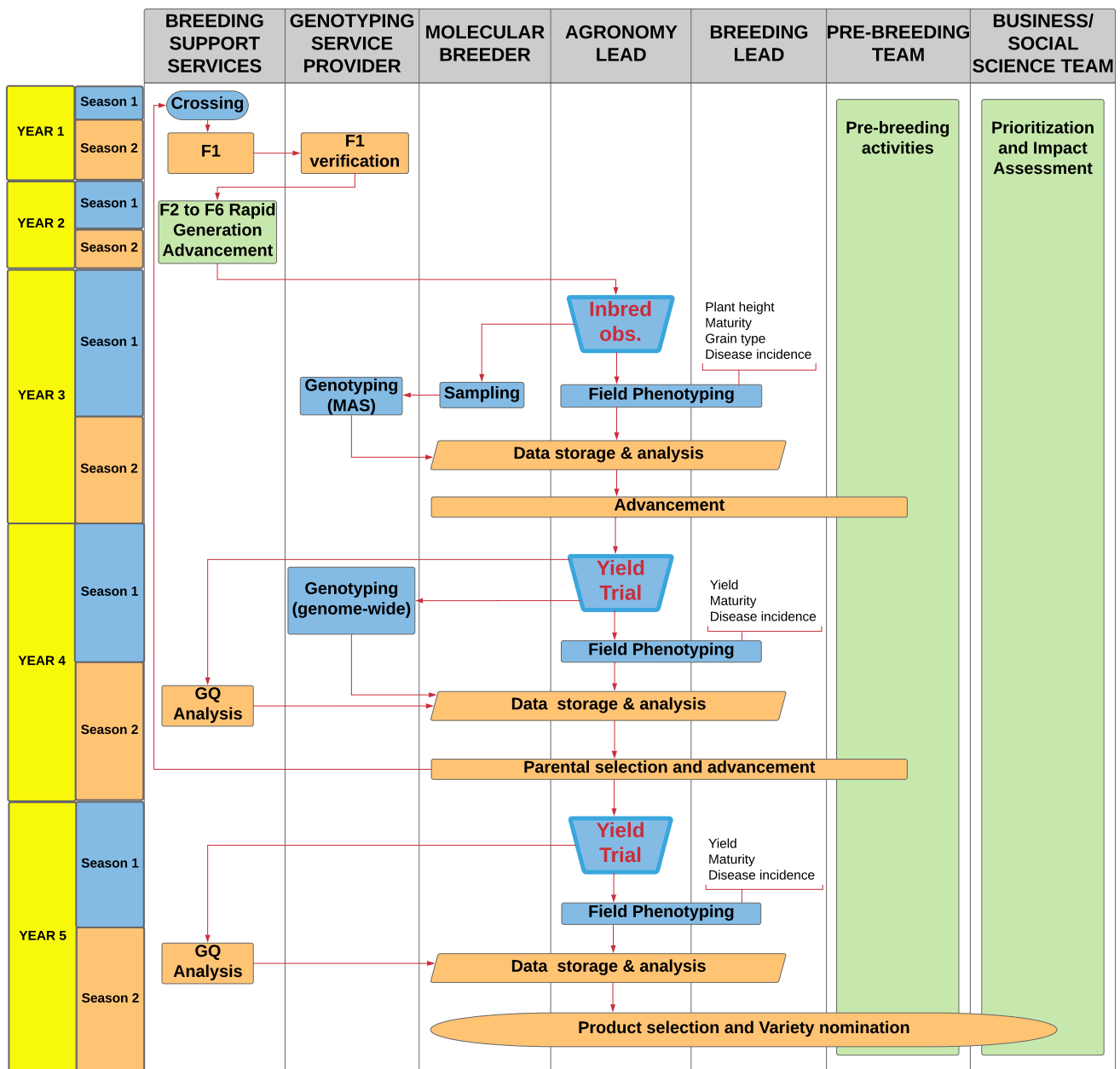
## Diagramming the breeding process

One of the most valuable exercises a breeding team undertakes in initiating its continuous improvement process is to articulate and communicate the breeding strategy clearly and formally. However, this is rarely done, especially in public programs, wherein the details of the breeding strategy are often unarticulated and subject to change on the breeder's whim. Figure 5 illustrates one example of how a multi-year RGA-based breeding strategy might be sketched out in terms of activities, accountability, and the flow of information and genetic material. While the illustrated schema is generalized and may not be applicable to any one program, it serves as a model for visualizing whom in a breeding team is responsible for which activities, and when they occur. When such a diagram is created and communicated with others, it allows breeders, pathologists, physiologists, agronomists, geneticists, and other members of the breeding team to understand where they interact with the strategy, how breeding activities fit together, identify and alleviate bottlenecks, and most importantly where innovations might be applied.

## Accountability and metrics of success

A process diagram of the breeding strategy also allows for transparency and accountability, and helps breeding management teams develop key performance indicators that assess the value of activities in delivering genetic gains. As part of a culture of continuous improvement, all breeding programs should be aiming to increase their rate of genetic gain by optimizing breeding pipelines, and demonstrating improvement by regularly measuring their outputs and progress against robust and relevant metrics. Often plant breeding programs are measured against the number of varieties nominated or released per year to the local government's variety release pipelines. This metric often over-estimates





**Fig. 5** Process flow diagram for a generalized inbred breeding program based on single seed descent. Columns correspond to people, teams, or service providers. Colors indicate seasonal activities. Arrows indicate the flow of information and/or breeding material through the pipeline

success rates because many released varieties are never adopted by farmers (Maredia and Raitzer 2010). In private-sector breeding programs, numbers of units sold or market share are common metrics of success, but such information is difficult to obtain for public-sector programs, particularly in developing countries, and comes with the added socio-economic complication of calculating who benefits and how equitably the benefit is distributed. Ceccarelli (2015) provides a good review of this topic and suggests that cost/benefit ratios should be used to measure plant breeding efficiency. In public-sector programs, direct estimation of genetic trend

(Piepho et al. 2014), delivered under farmer management, is the likely the best method for assessing long-term breeding program effectiveness. This metric can be captured relatively easily if the breeding organization continuously evaluates newly released materials in on-farm agronomy trials.

The role of senior managers and funders in driving the efficiency of public plant breeding programs cannot be overstated. Optimization of a breeding program will not likely occur without explicit leadership and direction that incentivizes innovation, efficiency, and the delivery of genetic gains under farmer management. Moreover, it is imperative for

senior management and financial stakeholders to be aware of performance against key indicators such as the average age of parental material, selection intensities (i.e., number of parents in the crossing block), selection accuracies (i.e., heritability in multi-location trials), number of lines advanced to each stage, length of the breeding cycle, and percentage of external germplasm used as parental material. Furthermore, it is worth mentioning the need for long-term, stable funding to support public breeding programs to deliver on the intended targets. The fragmented, restricted, and short-term funding models that are prevalent in most public breeding grants (especially in the developing world) incur the risk of short-circuiting breeding efficiency through frequent realignments of breeding targets or driving variable selection intensities and accuracies due to unpredictable and fluctuating budgets.

## Considerations for the future

Public-sector plant breeders and managers have a tremendous responsibility to the world as the sector often serves the communities most in need of increased rates of genetic gains. In many cases, public breeding programs are often the sole source of improved germplasm for the farmers they serve (Lopez-Pereira and Filippello 1995). Despite being immensely important, increasing rates of genetic gain in public plant breeding programs can be extremely challenging. Making changes to the breeding process in the name of increasing rates of genetic gain is often seen as very risky. Breeders often view their current pipelines as close to optimal and worry that if changes fail and the breeding program in turn fails, there will be dire consequences for the world's most vulnerable people. This leads to a level of conservatism that can make change extraordinarily difficult. But of course change must occur in order to improve. In fact, rates of genetic gain and varietal replacement in much of the developing world have been very low since the end of the Green Revolution period, and the potential positive impact of optimizing breeding pipelines to modestly increase selection differential and accuracy, and substantially reduce breeding cycle time, greatly outweighs the risk of continuing to rely on visual selection in outmoded pedigree breeding schemes.

## Strengthening public breeding through collaboration and purposeful innovation

As plant breeding teams navigate the early twenty-first century, a few key changes need to be made to the political and social landscapes to ensure continued success. First, funding the modernization of public-sector breeding programs as a dedicated activity must take place. Given the challenges of implementing modern approaches in public

breeding programs, an argument can be made for donors to fund the modernization process itself as separate and focused activity. This is the approach being taken by the Bill and Melinda Gates Foundation (BMGF) and the CGIAR through the establishment of the Excellence in Breeding Platform (EiB; [https://www.cgiar.org/wp/wp-content/uploads/2018/05/SC6-04\\_Multi-Funder-Breeding-Initiative-update.pdf](https://www.cgiar.org/wp/wp-content/uploads/2018/05/SC6-04_Multi-Funder-Breeding-Initiative-update.pdf)), a clearing house and consultancy for supporting the implementation of best practices in breeding programs serving the developing world. Second, unlike private-sector organizations, public plant breeding programs struggle to leverage and benefit from the collective investment, skills and experience across crops that large transnational breeding companies can command. As a result, public plant breeding programs must form interactive communities of practice that allow them to aggregate demand and stimulate the development of low-cost genotyping, phenotyping, and open-access IT systems for storage, management, analysis, and exchange of data (Spindel and McCouch 2016). Here again, the Bill and Melinda Gates Foundation is leading an effort within the CGIAR through the High Throughput Genotyping Project (<http://cegsb.icrisat.org/high-throughput-genotyping-project-http/>) and the Genomics Open Breeding Informatics Initiative (GOBii; <http://gobiiproject.org/>). In an effort to deal with diminishing research funds and increased costs, many public breeding programs targeting the same agroecologies will need to work together in coordinated breeding networks. Through this process, and other collaborative exercises, public plant breeding communities need to come together to lift one another up to a higher standard of operating.

By coming together in the next few decades, public plant breeding communities will have an opportunity to develop a common way of understanding challenges, opportunities, and progress against standardized sets of metrics. They will need to establish industry-wide standards, common vocabulary, and shared protocols for integrating new technology into the breeding process. Breeding teams and senior management of breeding institutes can both be optimistic about opportunities to include new technologies and improved ways of operating and increasing rates of genetic gain per dollar invested. It is reasonable to expect through the thoughtful and consistent application of new technologies to breeding strategies, funders of public plant breeding efforts will gain increased confidence in the approaches being adopted and in the capacity of public plant breeding programs to meet the demands of an increasingly complex world.

## Conclusions

Public-sector plant breeding programs serving farmers in the developing world can deliver much higher rates of genetic gain if breeding programs are optimized to select for quantitative traits. Traditional pedigree breeding methods based on visual selection do not work well after plant type is fixed, and therefore, breeding pipelines serving smallholder farmers in the developing world must be modernized to optimize the key components of the breeder's equation. Accuracy of selection for yield and other quantitative traits must be increased by testing more selection candidates in multi-location trials earlier in the breeding process, using experimental designs that effectively control field noise at low levels of within trial replication (e.g., p-rep designs). Selection intensity can be increased by replacing slow and ineffective pedigree selection, which requires visual selection of widely spaced plants in each inbreeding generation and is therefore extremely expensive in terms of time and labor, with single seed descent and bulk generation advancement techniques that rapidly move lines to fixation without selection, relying on MAS and a single visual selection step to ensure that only lines with appropriate plant type, phenology, and high-value haplotypes for disease resistance, stress tolerance, and quality are advanced to expensive multi-location trials. It should be noted, however, that returns on investment in both accuracy and selection intensity offer expensive pathways to increased genetic gain; investments in accuracy through increased replication rapidly run into diminishing returns once heritability exceeds 0.5 or so, and the relationship between genetic gain and selection intensity (i.e., population size) is roughly logarithmic rather than linear, meaning that a tenfold increase in program size is needed to double genetic gain. Modestly scaled breeding programs can achieve good genetic gains with stage one yield trials consisting of four or five well-managed p-rep trials conducted at locations representative of the TPE, with roughly 200 entries already fixed for phenology, plant type, and must-have qualitative traits, applying a selection intensity of 5% to maintain an effective population size of at least 10 per cycle.

The most underutilized pathway to increased genetic gains is likely reduced cycle time. Breeding cycles can be accelerated by immediately advancing parents selected from stage one testing for use as parents of the next cycle; additional years of testing are unlikely to increase accuracy of breeding value estimation enough to compensate for slower breeding cycles (of course, additional testing is needed before commercialization). Much greater reductions in cycle time can be achieved by selecting parents on the basis of breeding value before they are inbred to near

fixation, an approach whose efficacy was confirmed in a large number of recurrent selection experiments in maize, small grains, and legume crops in the 1960s through the 1990s. Classic, closed recurrent selection breeding plans can be made much more effective by integrating genomic selection. Accelerating the breeding cycle in such schemes can deliver genetic gains equivalent to those delivered in much larger and more expensive breeding programs that cycle more slowly and apply higher selection intensity each cycle. The greatest reductions in cycle time are achievable with pure genomic selection breeding plans (e.g., Gaynor et al. 2016), in which no inbreeding or phenotyping is conducted between cycles of recombination; however, such plans require extremely large and expensively produced training populations to deliver selection accuracy, and are unlikely to be feasible in many crop species for some years.

The changes outlined in this discussion will require some initial investment on the part of most breeding programs, but will result in greater genetic gains per dollar of operating budget. More importantly than additional investment, the modernization of public breeding programs will require strong support and guidance from research managers. Research managers in the CGIAR and national breeding programs must clearly convey to breeding teams that they will be supported in and held accountable for the delivery of genetic gains in product profiles valued by farmers, processors, and consumers. External evaluation and consultancy will often be needed to help programs design and implement these changes, as programs rarely have all the necessary skills in-house.

A critical change that must be supported by research managers is shifting away from performance evaluation systems that emphasize the number of journal articles published or varieties released and toward the contribution of team members to the overall objective of generating genetic gain in the product profile. This will require understanding how individual team members contribute to the overall process and designing performance metrics accordingly.

Taken together, the improvements in product focus, selection accuracy, selection intensity, and cycle length, driven by the effective application of new genotyping, phenotyping, and decision support technologies, have the potential to raise the current rate of genetic gain in the staple food crops produced by farmers in the developing world from a current rate that is likely well under 1% annually (and in many instances not significantly different than zero) to at least 2%. In the process, farmers will be better protected against a rapidly changing climate and better able to adapt to rapidly commercializing production systems.

**Authors contribution statement** JNC and RUJ conceived and drafted the outline and ideas for the review and provided

extensive editing and revision. PSB provided additional formatting and editing support. JNC, RUJ, PSB, JDA, JR, GA, TH, MQ, and EH each contributed the primary content for one or more sections. GA provided additional support for editing.

**Acknowledgements** The authors wish to thank Princess Dela Cruz for assistance in various aspects of this work and Damien Platten for valuable discussion and refining of ideas. This work was supported in part by the Transforming Rice Breeding project from the Bill and Melinda Gates Foundation (Grant No. OPP1076488).

## Compliance with ethical standards

**Conflict of interest** All authors declare that they have no conflict of interest.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Asif M (2013) Progress and opportunities of doubled haploid production, vol 6. Springer briefs in plant science. Springer, Heidelberg. <https://doi.org/10.1007/978-3-319-00732-8>
- Atlin GN, Cooper M, Bjornstad A (2001) A comparison of formal and participatory breeding approaches using selection theory. *Euphytica* 122:463–475. <https://doi.org/10.1023/A:1017557307800>
- Atlin GN, Cairns JE, Das B (2017) Rapid breeding and varietal replacement are critical to adaptation of cropping systems in the developing world to climate change. *Global Food Secur* 12C(2017):31–37
- Awada L, Phillips P, Smyth SJ (2018) The adoption of automated phenotyping by plant breeders. *Euphytica*. <https://doi.org/10.1007/s10681-018-2226-z>
- Badiane O (2017) Africa must invest now to harness future agricultural technologies, IFPRI. <http://www.ifpri.org/blog/africa-must-invest-now-harness-future-agricultural-technologies>. Accessed 20 Sept 2018)
- Barabaschi D, Tondelli A, Desiderio F, Volante A, Vaccino P, Valè G, Cattivelli L (2016) Next generation breeding. *Plant Sci* 242:3–13. <https://doi.org/10.1016/j.plantsci.2015.07.010>
- Baranski MR (2015) The wide adaptation of green revolution wheat: international roots and Indian context of new plant breeding ideal, 1960-1970. *Stud Hist Philos Biol Biomed Sci* 50:41–50. <https://doi.org/10.1016/j.shpsc.2015.01.004>
- Barr AR (2009) Marker-assisted selection in theory and practice. In: Ceccarelli S, Guimaraes EP, Weltzien E (eds) *Plant breeding and farmer participation*. FAO, Rome, pp 479–517
- Beissinger TM, Hirsch CN, Vaillancourt B, Deshpande S, Barry K, Buell CR, Kaeppler SM, Gianola D, de Leon N (2014) A genome-wide scan for evidence of selection in a maize population under long-term artificial selection for ear number. *Genetics* 196:829–840. <https://doi.org/10.1534/genetics.113.160655>
- Berke TG, Baenziger PS (1992) Portable and desktop computer integrated field book and data collection system for agronomists. *Agron J* 84:119–121. <https://doi.org/10.2134/agronj1992.00021962008400010023x>
- Bernardo R, Moreau L, Charcosset A (2006) Number and fitness of selected individuals in marker-assisted and phenotypic recurrent selection. *Crop Sci* 46:1972–1980. <https://doi.org/10.2135/cropsci2006.01-0057>
- Bevan MW, Uauy C (2013) Genomics reveals new landscapes for crop improvement. *Genome Biol* 14:206. <https://doi.org/10.1186/gb-2013-14-6-206>
- Boyle EA, Li YI, Pritchard JK (2017) An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169:1177–1186. <https://doi.org/10.1016/j.cell.2017.05.038>
- Breseghele F, Morais OP, Castro EM, Prabhu AS, Bassinello PZ, Pereira JA et al (2009) Recurrent selection resulted in rapid genetic gain for upland rice in Brazil. *Int Rice Res Newsl* 34:4
- Bulmer MG (1971) The effect of selection on genetic variability. *Am Nat* 105:201–211
- Caballero A, Keightley PD, Hill WG (1991) Strategies for increasing fixation probabilities of recessive mutations. *Genet Res* 58:129–138. <https://doi.org/10.1017/S0016672300029785>
- Caballero A, Santiago E, Toro MA (1996) Systems of mating to reduce inbreeding in selected populations. *Anim Sci* 62(3):431–442. <https://doi.org/10.1017/S1357729800014971>
- Carandang FM, Shanmugasundaram S, Carpena AL (2006) Rapid generation advancement in soybeans using immature seeds. *Phil J Crop Sci* 31:53–59
- Carnielli CM, Winck FV, Leme AFP (2015) Functional annotation and biological interpretation of proteomics data. *Biochem Biophys Acta Prot Proteom* 1854:46–54. <https://doi.org/10.1016/j.bbap.2014.10.019>
- Ceccarelli S (2015) Efficiency of plant breeding. *Crop Sci* 55(1):87
- Chlingaryan A, Sukkarieh S, Whelan B (2018) Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. *Comput Electron Agric* 151:61–69. <https://doi.org/10.1016/j.compag.2018.05.012>
- Choo TM, Reinbergs E, Kasha KJ (1985) Use of haploids in breeding barley. *Plant Breed Rev* 3:219–252
- Cobb JN, DeClerck G, Greenberg A, Clark R, McCouch SR (2013) Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theor Appl Genet* 126:867–887. <https://doi.org/10.1007/s00122-013-2066-0>
- Cobb JN, Biswas PS, Platten JD (2018) Back to the future: Revisiting MAS as a tool for modern plant breeding. *Theor Appl Genet*. <https://doi.org/10.1007/s00122-018-3266-4>
- Collard BCY, Berozo JC, Lenaerts B, Mendoza R, Santelices R, Lopena V, Verdeprado H, Raghavan C, Gregorio G, Vial L, Demont M, Biswas P, Iftekharruddaula K, Rahman MA, Cobb J, Islam MR (2017) Revisiting rice breeding methods: evaluating the use of rapid generation advance (rga) for routine rice breeding. *Plant Prod Sci* 20:337–352. <https://doi.org/10.1080/1343943X.2017.1391705>
- Coors JG (1999) Selection methodology and heterosis. In: Coors JG, Pandey S (eds) *The genetics and exploitation of heterosis in crops*. American Society of Agronomy Inc, Madison
- Crescente JM, Guidobaldi F, Demichelis M, Formica MB, Helguera M, Vanzetti LS (2017) Phenobook: an open source software for phenotypic data collection. *GigaScience* 6:1–5. <https://doi.org/10.1093/gigascience/giw019>
- Crow JF, Kimura M (1970) *An introduction to population genetics theory*. Blackburn Press, Caldwell
- Dekkers JCM (2012) Application of genomics tools to animal breeding. *Curr Genom* 13:207–212. <https://doi.org/10.2174/13892021200543057>



- Drovandi CC, Holmes C, McGree JM, Mengersen K, Richardson S, Ryan EG (2017) Principles of experimental design for big data analysis. *Stat Sci* 3:385–404. <https://doi.org/10.1214/16-STS604>
- Dwivedi SL, Britt AB, Tripathi L (2015) Haploids: constraints and opportunities in plant breeding. *Biotechnol* 33:812–829. <https://doi.org/10.1016/j.biotechadv.2015.07.001>
- Eberhart SA (1970) Factors affecting efficiencies of breeding methods. *Afr Soils* 15:655–680
- Endelman JB, Atlin GN, Beyene Y, Semagn K, Zhang X, Sorrells M, Jannink JL (2014) Optimal design of preliminary yield trials with genome-wide markers. *Crop Sci* 54:48–59. <https://doi.org/10.2135/cropsci2013.03.0154>
- Falk DE (2010) Generating and maintaining diversity at the elite level in crop breeding. *Genome* 53:982–991. <https://doi.org/10.1139/G10-081>
- Fernández BJ, Toro MA (1999) The use of mathematical programming to control inbreeding in selection schemes. *J Anim Breed Genet* 116:447–466. <https://doi.org/10.1046/j.1439-0388.1999.00196.x>
- Fernández J, Beatriz V, Ricardo PW, Miguel ÁT (2005) Efficiency of the use of pedigree and molecular marker information in conservation programs. *Genetics* 170:1313–1321. <https://doi.org/10.1534/genetics.104.037325>
- Finkel E (2009) With phenomics plant scientists hope to shift breeding into overdrive. *Science* 325:380–381. [https://doi.org/10.1126/science.325\\_380](https://doi.org/10.1126/science.325_380)
- Forster BP, Heberle-Bors E, Kasha KJ, Touraev A (2007) The resurgence of haploids in higher plants. *Trends Plant Sci* 12:368–375. <https://doi.org/10.1016/j.tplants.2007.06.007>
- Frey KJ, McFerson JK, Branson CV (1988) A procedure for one cycle of recurrent selection per year with spring-sown small grains. *Crop Sci* 28:855–856
- Gaur PM, Srinivasan S, Gowda CLL (2007) Rapid generation advancement in Chickpea. *SAT J* 3(1):3
- Gaynor CR, Gorjanc G, Bentley AR, Ober ES, Howell P, Jackson R (2016) A two-part strategy for using genomic selection to develop inbred lines. *Crop Sci* 56:2372–2386
- Godfrey HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Sherman R, Sandy MT, Camilla T (2010) Food security: the challenge of feeding 9 billion people. *Science* 327:812–818
- Goulden CH (1939) Problems in plant selection. In: Burnett RC (ed) *Proceeding of the seventh genetics congress*. Cambridge University Press, Edinburgh, pp 132–133
- Grafius JE (1965) Shortcuts in plant breeding. *Crop Sci* 5:377
- Grundy B, Villanueva B, Woolliams JA (1998) Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. *Genet Res* 72:159–168
- Guzman PS (1998) Effective population size, genetic variability, and gains from recurrent selection in the BS11 maize population. *Retrospective Theses and Dissertations* 11611, p 128
- Hallander J, Waldmann P (2009) Optimum contribution selection in large general tree breeding populations with an application to scots pine. *Theor Appl Genet* 118:1133–1142
- Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. *Crop Sci* 49:1–12. <https://doi.org/10.2135/cropsci2008.08.0512>
- Henderson CR (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32:69–83
- Hess M, Druet T, Hess A, Garrick D (2017) Fixed-length haplotypes can improve genomic prediction accuracy in an admixed dairy cattle population. *Genet Sel Evol* 49:54. <https://doi.org/10.1186/s12711-017-0329-y>
- Hickey JM, Chiurugwi T, Mackay I, Powell W (2017) Genomic prediction unifies animal and plant breeding programs as platforms for biological discovery. *Nat Genet* 49:1297–1303
- Hill WG (2010) Understanding and using quantitative genetic variation. *Philos Trans R Soc B Biol Sci* 365(1537):73–85. <https://doi.org/10.1098/rstb.2009.0203>
- Hill WG, Michael E, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet* 4:e1000008. <https://doi.org/10.1371/journal.pgen.1000008>
- Humphreys DG, Knox RE (2015) Doubled haploid breeding in cereals advances in plant breeding strategies: breeding, biotechnology and molecular tools. Springer, New York, pp 241–290
- Jain HK (2010) *The green revolution: history, impact and future*. the green revolution: history, impact and future. Stadium Press LLC, Houston
- Javid M, Rosewarne GM, Sudheesh S, Kant P, Leonforte A, Lombardi M, Kennedy RP, Cogan NOI, Slater TA, Kau S (2015) Validation of molecular markers associated with boron tolerance, powdery mildew resistance and salinity tolerance in field peas. *Front Plant Sci* 6:917. <https://doi.org/10.3389/fpls.2015.00917>
- Júnior OPM, Breseghello F, Duarte JB, Morais OP, Rangel PHN, Coelho ASG (2017) Effectiveness of recurrent selection in irrigated rice breeding. *Crop Sci* 57:3043–3058. <https://doi.org/10.2135/cropsci2017.05.0276>
- Kenworthy WJ, Brim CA (1979) Recurrent selection in soybeans. i. seed yield. *Crop Sci* 19:315–318
- Köhl K, Jürgen G (2015) A software tool for the input and management of phenotypic data using personal digital assistants and other mobile devices. *Plant Methods* 11:25. <https://doi.org/10.1186/s13007-015-0069-3>
- Laske CH, Bruno BMT, Nelson JLD, Fernando FC (2012) Breeding objectives and economic values for traits of low input family-based beef cattle production system in the State of Rio Grande Do Sul. *R Bras Zootec* 41:298–305. <https://doi.org/10.1590/S1516-35982012000200010>
- Li L, Tacke E, Hofferbert HR, Lubeck J, Strahwald J, Draffehn AM, Walkemeier B, Gebhardt C (2013) Validation of candidate gene markers for marker-assisted selection of potato cultivars with improved tuber quality. *Theor Appl Genet* 126:1039–1052
- Lind BM, Mitra M, Constance EB, Trevor MF, Andrew JE (2018) The genomics of local adaptation in trees: are we out of the woods yet? *Tree Genet Genom* 14:29
- Liu H, Pamela Z, Haibo W, Liu Liu C, Zhanyuan L, Yanxia W, Yan G (2016) A fast generation cycling system for oat and triticale breeding. *Plant Breed* 135(5):574–579. <https://doi.org/10.1111/pbr.12408>
- Lobos GA, Camargo AV, del Pozo A, Araus JL, Ortiz R, Doonan JH (2017) Editorial: plant phenotyping and phenomics for plant breeding. *Front Plant Sci* 8:2181. <https://doi.org/10.3389/fpls.2017.02181>
- Lopes MS, Reynolds MP, Manes Y, Singh RP, Crossa JJ, Braun HJ (2012) Genetic yield gains and changes in associated traits of CIMMYT Spring bread wheat in a “historic” set representing 30 Years of breeding. *Crop Sci* 52(3):1123–1131. <https://doi.org/10.2135/cropsci2011.09.0467>
- Lopez-Pereira MA, Filippello MP (1995) Emerging roles of the public and private sectors of maize seed industries in the developing world. In: CIMMYT economics program working paper 95-01. Mexico, D.E: CIMMYT, p 94
- Lorenz AJ (2013) Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: a simulation experiment. *G3 Genes Genomes Genet* 3(3):481–491. <https://doi.org/10.1534/g3.112.004911>

- Lush J (1937) Animal breeding. Plans. Iowa State College Press, Ames
- Mackay JJ, Caligari PDS (1999) Major errors in data and their effect on response to selection. *Crop Sci* 39:697–702. <https://doi.org/10.2135/cropsci1999.0011183X003900020016x>
- Maluszynski M, Kasha K, Forster BP, Szarejko I (2003) Doubled haploid production in crop plants: a manual. Springer, Dordrecht. <https://doi.org/10.1007/978-94-017-1293-4>
- Maredia MK, Raitzer DA (2010) Estimating overall returns to international agricultural research in Africa through benefit-cost analysis: a “best-evidence” approach. *Agric Econom* 41(1):81–100. <https://doi.org/10.1111/j.1574-0862.2009.00427.x>
- Masuka B, Atlin GN, Olsen M, Magorokosho C, Labuschagne M, Crossa J, Bänziger M, Pixley KV, Vivek BS, von Biljon A, Macrobert J, Alvarado G, Prasanna BM, Makumbi D, Tareknege A, Das B, Zaman-Allah M, Cairns JE (2017) Gains in maize genetic improvement in eastern and southern Africa: i CIMMYT hybrid breeding pipeline. *Crop Sci*. 57:168–179. <https://doi.org/10.2135/cropsci2016.05.0343>
- Meuwissen THE (1997) Maximizing the response of selection with a predefined rate of inbreeding. *J Anim Sci* 75:934–940. <https://doi.org/10.2527/1997.754934x>
- Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829
- Moghaddar N, Swan AA, van der Werf JHJ (2017) Genomic prediction from observed and imputed high-density ovine genotypes. *Genet Sel Evol* 49:40. <https://doi.org/10.1186/s12711-017-0315-4>
- Moose SP, Mumm RT (2008) Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiol* 147:969–977. <https://doi.org/10.1104/pp.108.118232>
- Payne TS, Stuthman DD, McGraw RL, Bregitzer PP (1986) Physiological changes associated with three cycles of grain yield improvement in oats. *Crop Sci* 26:734–736
- Piepho HP, Laidig F, Drobek T, Meyer U (2014) Dissecting genetic and non-genetic sources of long-term yield trend in German official variety trials. *Theor Appl Genet* 127(5):1009–1018
- Pingali PL (2012) Green revolution: impacts, limits, and the path ahead. *Proc Natl Acad Sci* 109:12302–12308. <https://doi.org/10.1073/pnas.0912953109>
- Platten JD, Cobb JN, Zantua RE (2019) Criteria for evaluating molecular markers: comprehensive quality metrics to improve marker-assisted selection. *PLoS one* 14(1):e0210529
- Prieur V, Clarke SM, Brito LF, McEwan JC, Lee MA, Brauning R, Dodds KG, Auvray B (2017) Estimation of linkage disequilibrium and effective population size in New Zealand sheep using three different methods to create genetic maps. *BMC Genet* 18:68. <https://doi.org/10.1186/s12863-017-0534-2>
- Pryce JE, Johnston J, Hayes BJ, Sahana G, Weigel KA, McParland S, Spurlock D et al (2014) Imputation of genotypes from low density (50,000 markers) to high density (700,000 markers) of cows from research herds in Europe, North America, and Australasia using 2 reference populations. *J Dairy Sci* 97:1799–1811. <https://doi.org/10.3168/jds.2013-7368>
- Ragot M, Bonierbale MW, Weltzien E (2018) From market demand to breeding decisions: a framework. In: Working Paper 2. CGIAR Gender and Breeding Initiative, Lima (Peru)
- Rathore A, Vikas KS, Sarita KP, Chukka SR, Vivek T, Manish KP, Anil KV, Roma RD (2018) Current status and future prospects of next-generation data management and analytical decision support tools for enhancing genetic gains in crops. In: Varshney RK, Pande MK, Chitkineni A (eds) *Plant Genetics and Molecular Biology. Advances in Biochemical Engineering/Biotechnology*. Springer, Cham
- Riedelsheimer C, Endelman JB, Stange M, Sorrells ME, Jannink JL, Melchinger AE (2013) Genomic predictability of interconnected biparental maize populations. *Genetics* 194:493–503. <https://doi.org/10.1534/genetics.113.150227>
- Rife TW, Poland JA (2015) Field book: an open-source application for field data collection on android. *Crop Sci* 54:1624–1627. <https://doi.org/10.2135/cropsci2013.08.0579>
- Ritchie H, Reay DS, Higgins P (2018) Beyond calories: a holistic assessment of the global food system. *Front Sustain Food Syst* 2:57. <https://doi.org/10.3389/fsufs.2018.00057>
- Rizal G, Karki SN, Alcasid M (2014) Shortening the breeding cycle of sorghum, a model crop for research. *Crop Sci* 54:520–529
- Roorkiwal M, Jarquin D, Singh MK, Gaur PM, Bharadwaj C, Rathore A, Howard R et al (2018) Genomic-enabled prediction models using multi-environment trials to estimate the effect of genotype × environment interaction on prediction accuracy in Chickpea. *Sci Rep* 8:11701. <https://doi.org/10.1038/s41598-018-30027-2>
- Rutkoski J (2018) h2s. <https://jerut.shinyapps.io/geneticgain/> Accessed 10 Sept 2018
- Rutkoski J, Jesse P, Suchismita M, Enrique A, Lorena GP, Josá C, Matthew R, Ravi S (2016) Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3 Genes Genomes Genet* 6:2799–2808. <https://doi.org/10.1534/g3.116.032888>
- Saxena K, Saxena RK, Varshney RK (2017) Use of immature seed germination and single seed descent for rapid genetic gains in pigeonpea. *Plant Breed* 136:954–957. <https://doi.org/10.1111/pbr.12538>
- Shelton A, Tracy W, Shelton AC, Tracy WF (2015) recurrent selection and participatory plant breeding for improvement of two organic open-pollinated sweet corn (*Zea mays* L.). *Popul Sustain* 7:5139–5152. <https://doi.org/10.3390/su7055139>
- Singh A, Ganapathysubramanian S, Singh AK, Sarkar S (2016) Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci* 21:110–124
- Smith S, Buback D, Nelson B, Stanek J, Gerke J (2015) Genetic diversity and modern plant breeding. In: Ahuja, Jain SM (eds) *Genetic diversity and erosion in plants*, vol 1. Springer, Cham, pp 55–88
- Snappe JW (1989) Doubled haploid breeding: theoretical basis and practical applications. In: Mujeeb-Kazi A, Sitch LA (eds) *Review of advances in plant biotechnology 1985-88: 2nd international symposium on genetic manipulation in Crop*. CIMMYT Mexico DF and IRRI Manila, Philippines, pp. 19–30
- Sonesson AK, Meuwissen THE (2000) Mating schemes for optimum contribution selection with constrained rates of inbreeding. *Genet Sel Evol* 32:231. <https://doi.org/10.1186/1297-9686-32-3-23>
- Spindel J, McCouch SR (2016) When more is better: how data sharing would accelerate genomic selection of crop plants. *New Phytol* 212:814–826. <https://doi.org/10.1111/nph.14174>
- Sun JJ, Rutkoski JA, Poland JC, Jannink JL, Sorrells ME (2017) Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. *Plant Genome*. <https://doi.org/10.3835/plantgenome2016.11.0111>
- Tadesse W, Inagaki M, Tawkaz S, Baum M, van Ginkel M (2012) Recent advances and application of doubled haploids in wheat breeding Africa. *J Biotechnol* 11:15484–15492. <https://doi.org/10.5897/AJB12.2124>
- Tanaka J, Hayashi T, Iwata H (2016) A practical, rapid generation-advancement system for rice breeding using simplified biotron breeding system. *Breed Sci* 66:542–551. <https://doi.org/10.1270/jsbbs.15038>
- Tattaris M, Reynolds MP (2016) A direct comparison of remote sensing approaches for high-throughput phenotyping in plant breeding. *Front Plant Sci*. 7:1131

- Toro M, Pérez-Enciso M (1990) Optimization of selection response under restricted inbreeding. *Genet Sel Evol* 22:93. <https://doi.org/10.1186/1297-9686-22-1-93>
- Touraev A, Forster BP, Jain SM (2009) Advances in haploid production in higher plants. Springer, Dordrecht. <https://doi.org/10.1007/978-1-4020-8854-4>
- van Ginkel M, Ortiz R (2018) Cross the best with the best, and select the best: HELP in breeding selfing crops. *Crop Sci*. 58:17–30. <https://doi.org/10.2135/cropsci2017.05.0270>
- VanEmon JM (2016) The omics revolution in agricultural research. *J Agric Food Chem* 64:36–44. <https://doi.org/10.1021/acs.jafc.5b04515>
- Vela-Avitúa S, Meuwissen THE, Luan T, Ødegård J (2015) Accuracy of genomic selection for a sib-evaluated trait using identity-by-state and identity-by-descent relationships. *Genet Sel Evol* 47:9. <https://doi.org/10.1186/s12711-014-0084-2>
- Walsh B (2014) Special issues on advances in quantitative genetics: introduction. *Heredity* 112:1–3. <https://doi.org/10.1038/hdy.2013.115>
- Wang J (2016) Pedigrees or markers: which are better in estimating relatedness and inbreeding coefficient? *Theor Popul Biol* 107:4–13
- Wang J, Santiago E, Caballero A (2016) Prediction and estimation of effective population size. *Heredity* 117:193–206. <https://doi.org/10.1038/hdy.2016.43>
- Wang Yu, Bennewitz J, Wellmann R (2017) Novel optimum contribution selection methods accounting for conflicting objectives in breeding programs for livestock breeds with historical migration. *Genet Sel Evol* 49:45. <https://doi.org/10.1186/s12711-017-0320-7>
- Watson A, Ghosh S, Williams MJ, Cuddy WS, Simmonds J, Rey MD, Hatta MAM et al (2018) Speed breeding is a powerful tool to accelerate crop research and breeding. *Nat Plants* 4:23–29. <https://doi.org/10.1038/s41477-017-0083-8>
- Wright S (1931) Evolution in mendelian populations. *Genetics* 16:97–159
- Wright S (1933) Inbreeding and homozygosity. *Proc Natl Acad Sci USA* 19:411–420

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.