ARTICLE

Crop Breeding and Applied Biotechnology
21(S): e394621S15, 2021
Brazilian Society of Plant Breeding.
Printed in Brazil
http://dx.doi.org/10.1590/1984-70332021v21Sa28

Implementation of genomic selection in public-sector plant breeding programs: Current status and opportunities

Cleiton Antonio Wartha[1] and Aaron Joel Lorenz[1*]

Abstract: *The usefulness of genomic selection (GS) for improvement of complex traits has been demonstrated in plant and animal breeding. While fully adopted in the livestock sector and commercial plant breeding, the implementation of GS in public-sector plant breeding programs lags. The goal of this review is to discuss advancements in GS implementation and opportunities for near-term and long-term adoption in public-sector plant breeding programs. We also highlighted specific applications of GS for cultivar development ordered by what we believe to be their feasibility, where feasibility is defined by cost, disruption to current breeding practices, and risk.*

Keywords: *Genomic selection implementation, plant breeding pipeline, realized genetic gain, crop improvement, quantitative genetics*

INTRODUCTION

The wide availability of high-density molecular markers has been opening new opportunities to plant breeding programs for at least the past decade while enabling higher genetic gains from selection. Genome-wide markers can provide more accurate estimates of genetic relatedness among individuals than pedigree relationships by capturing Mendelian sampling effects (Habier et al. 2007). In fact, Bernardo (1994) successfully used background markers to capture relationships and predict the yield of maize single-crosses using best linear unbiased prediction. The use of ridge regression as an alternative to marker subset selection was first proposed for marker-assisted selection (MAS) by Whittaker et al. (2000) within the context of biparental populations. The concept of genomic selection (GS), however, was fully introduced by Meuwissen et al. (2001) in a landmark article proposing statistical models for marker-based prediction that could take advantage of high-density molecular marker data. The application of genomic selection to plant breeding has been studied since the late 2000s (Bernardo and Yu 2007) and has been considered one of the key post-1990 technological advances in crop improvement, accompanied by genetic engineering, quantitative trait locus (QTL) mapping, phenomics, envirotyping, and gene editing (Bernardo 2016).

Marker-assisted selection was proposed and successfully used for selecting traits controlled by genes with large effects (Young 1999). However, many economically important traits are complex, being controlled by many small-effect loci that are strongly influenced by the environment. The MAS procedure includes QTL identification followed by estimation of marker effects - a two-step

***Corresponding author:**
E-mail: lore0149@umn.edu
ORCID: 0000-0002-4361-1683

Received: 25 August 2021
Accepted: 31 August 2021
Published: 03 September 2021

[1] University of Minnesota, Department of Agronomy and Plant Genetics, 411 Borlaug Hall, 1991 Upper Buford Circle, Saint Paul, MN 55108, USA

approach that can result in biased marker effects (Moreau et al. 2004, Heffner et al. 2009). Meanwhile, small effect loci are often excluded from the prediction model due to stringent significance thresholds used for QTL detection. A major advantage of GS is that it circumvents the individual QTL detection and validation. This also circumvents the need to develop specific mapping populations that are costly and time-consuming. Multiple researchers have indeed confirmed that the genomic prediction approach is more accurate than a MAS approach (Lorenzana and Bernardo 2009, Lorenz 2013), and have reported improved genetic gains in both simulated and real breeding populations (Bernardo and Yu 2007, Wong and Bernardo 2008, Asoro et al. 2013, Massman et al. 2013).

The fundamental goal of GS in cultivar development is to achieve higher rates of genetic gain at a lower cost and in less time than with conventional phenotypic methods. While the number of publications on GS for plant breeding has greatly increased since 2007 (Figure 1), and GS has been widely adopted by multinational seed companies, challenges remain and thus there are still relatively few examples of smaller public programs that have fully adopted GS. Hence, the objective of this review is to discuss advancements in GS implementation and opportunities for near-term and long-term adoption in public-sector plant breeding programs.

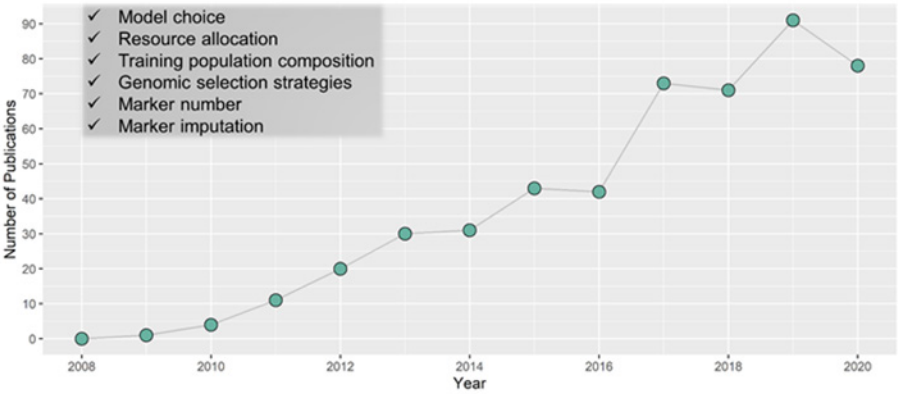

Figure 1. Number of publications on genomic selection in plant breeding obtained from a search on “genomic selection” AND “plant breeding” in the Scopus Database on 22 August 2021.

PROMISING APPLICATIONS OF GENOMIC PREDICTION FOR CULTIVAR IMPROVEMENT

Breeding programs are oriented towards obtaining high rates of genetic gain while maintaining useful genetic diversity for the traits of interest. The expected response to selection per unit time can be elegantly summarized by the breeder’s equation: $\Delta G = \frac{ih\sigma_A}{L}$, where ΔG is the gain per year, i is the selection intensity, h is the selection accuracy, σ_A is the square root of additive genetic variance, and L is the time in years to complete a cycle of breeding (Lush 1937, Falconer and Mackay 1996). While this formula can predict genetic gain only under many assumptions in a highly structured and systematic recurrent selection program, it nevertheless has been invaluable to plant breeders running cultivar development pipelines by providing a framework to assist them in identifying numerous ways in which genetic gain can be enhanced (Cobb et al. 2019). Identifying ways to enhance genetic gain in breeding programs is nothing new, but more accurate predictions of complex traits made possible by genomic prediction has opened new opportunities. We briefly review some of these potential applications.

Reducing cycle time

A main advantage of GS over phenotypic selection (PS) is the potential reduction in cycle time. Opportunities for recombination and allele frequency changes that only occurred every five or more years are now possible within months using GS. Phenotypic selection is limited by the evaluation of performance in the target population of environments (TPE), which is only one opportunity for field evaluation per year during the growing season of major crops in temperate regions. Meanwhile, GS can be performed in a greenhouse or year-round nursery. Under the assumption of equal additive genetic variance and selection intensity for GS and PS, a reduced selection accuracy of GS compared to PS can be compensated for by greatly reduced cycle time (Heffner et al. 2010).

Increasing selection intensity

Resources for testing in breeding programs are, of course, limited, and thus the allocation of testing resources to different testing stages (preliminary versus advanced) and between replication and population size is a topic of great interest to plant breeders (Mi et al. 2011, Riedelsheimer and Melchinger 2013, Longin et al. 2015). Genomic prediction can act as a replacement for phenotyping in advancing individuals to the next stage of testing. In terms of resource allocation, an estimated cost of a two-row field plot in one environment is $14 (in U.S. dollars), while the per-sample cost of genotyping for a few thousand single nucleotide polymorphism (SNP) markers is the same (Bernardo 2021). The potential for strong genotype x environment (GxE) interaction effects for complex traits necessitates that phenotyping be performed in many environments, thus increasing the cost of phenotyping for each selection candidate.

The cost advantage of genotyping over phenotyping is especially pronounced in perennial species for which field plots need to be maintained for several years. If genotyping is less expensive than phenotyping, genomic prediction allows for increased population size by generating and testing more candidates for the same cost, effectively increasing the selection intensity. This GS application is particularly useful in the early stages of a breeding program where early-generation material often consumes the most testing resources because of large population sizes before selection (Bassi et al. 2016).

Increased selection accuracy for traits that are difficult to phenotype

Genomic prediction has the potential to increase the selection accuracy for traits that are difficult to accurately phenotype because it effectively shares information between individuals in the form of genetic relationships. From the individual marker perspective, genomic prediction leverages “hidden replication” of alleles, and therefore the sum of predicted marker effects could be a more accurate predictor of genetic value than an individual phenotype. Examples of potential traits for this use are resistance to insect pests (Badji et al. 2021), nematodes (Ravelombola et al. 2020) and diseases (Adeyemo et al. 2020), root traits (Wolfe et al. 2017), postharvest storage (Roth et al. 2020), and end-product quality traits (Dreisigacker et al. 2021). This approach still requires the upfront investment in design and laborious phenotyping of the training population (TP), but the investment may decrease future costs by performing GS in several populations.

Genomic prediction for single-cross performance in hybrid cultivars

Genomic selection can be deployed in different stages of hybrid breeding to predict the performance of all possible hybrids made between all parent candidates at any stage in the breeding program. The prediction of performance *per se* of lines targets traits that exhibit mainly an additive genetic architecture, such as disease resistance traits. The topcross performance consists of predicting the general combining ability (GCA) of inbred progenies to advance the superior ones to additional crosses with more testers and evaluation in more environments (Albrecht et al. 2011, Albrecht et al. 2014, Riedelsheimer et al. 2012, Jacobson et al. 2014). The major benefits of GS in this stage are better allocation of field testing resources and cost savings if genotyping is cheaper than creating and evaluating the topcrosses. Lastly, GS can be expanded to predict the specific combining ability (SCA) effect of a specific parental combination in addition to GCA effects of the parents to predict total single-cross performance. This approach changes the scope of genetic evaluation from narrower set of field tested candidate lines to the in silico evaluation of all possible single-cross combinations (Technow et al. 2012, Massman et al. 2013, Kadam et al. 2016, Kadam et al. 2021).

ACADEMIC STUDIES HAVE PROVIDED NEW KNOWLEDGE ON OPTIMIZATION, BUT IMPLEMENTATION LAGS

The number of publications on optimizing the efficiency of GS for plant breeding has greatly increased during the last decade (Figure 1). Factors such as size of the TP, heritability, relatedness between TP and target population, linkage disequilibrium (LD) and effective number of factors play important roles in determining the prediction accuracy and thus the selection response (Daetwyler et al. 2008, Combs and Bernardo 2013a). For instance, the expected prediction accuracy increases as the product of the TP size and the trait heritability (Nh^2) (Daetwyler et al. 2008). Increased TP size improves marker effect estimation accuracy, with more pronounced effects when h^2 is low and diminishing returns as N becomes larger (Lorenzana and Bernardo 2009, Albrecht et al. 2011, Heffner et al. 2011, Guo et al. 2012). The genetic relationship between the TP and the target population is another key factor (Wientjes et al. 2013, Cooper et al. 2014, Lorenz and Smith 2015). High relatedness helps to ensure that the loci generating variation in the target population are the same as in the TP, and that the marker and QTL alleles are in the same linkage phase so that marker effects can be translated between the TP and target population (Habier et al. 2010, Riedelsheimer et al. 2013). Other effects such as allele-by-genetic background effects could be ameliorated by selecting training and target populations that are genetically similar. A good guiding principle is to design the TP with the target population in mind.

Several prediction models based on whole-genome marker data have been proposed to predict the genetic value of selection candidates for complex traits and they differ on their assumptions and treatment of marker effects (Jannink et al. 2010). The 'large predictor effects *p*, small available observations *n'* problem has been tackled with parametric (e.g. genomic best linear unbiased prediction, ridge regression, Bayesian regression), and non-parametric approaches (e.g. kernel regression, support vector regression, random forest, neural networks) (Gianola et al. 2009, de los Campos et al. 2013). Even though the prediction accuracy of different parametric models depends on the genetic architecture of the trait and LD structure of the population (Lorenz et al. 2011), empirical studies have shown reasonably good performance of these models for complex traits with different genetic architectures (Crossa et al. 2017). The ridge regression – best linear unbiased prediction (RR-BLUP) is the most commonly utilized model due to computational efficiency, with marker effects treated as random effects, normally distributed, and equally shrunken towards zero (Endelman 2011). Various studies have examined other relevant factors for the success of GS in a plant breeding program such as modeling $\mathrm{G{\times}E}$ interaction and non-additive effects, resource allocation, and multiple-trait models. These factors have been extensively reviewed recently elsewhere (Crossa et al. 2017, Voss-Fels et al. 2019, Xu et al. 2020).

Research on the development of strategies to implement GS in breeding programs needs to account for the reproductive biology of the crop, genetic architecture underlying the target traits, current program design, and resource availability. Computer simulations that examine the reorganization of conventional breeding programs around GS with distinct population improvement and product development components have predicted larger genetic gains (Gaynor et al. 2017) while selection can be balanced with maintenance of genetic diversity by means of a specific algorithm (Gorjanc et al. 2018). In a perennial ryegrass program, stochastic simulation comparing GS and conventional breeding strategies have shown significantly larger genetic gain utilizing GS due to a four-year reduction in cycle time (Lin et al. 2016).

Despite an abundance of studies generating the knowledge summarized above, there is a limited number of reported cases of actual implementation of GS in public breeding programs (Voss-Fels et al. 2019). Converting promising publications into practical application involves the resolution of many tactical logistics and practical program constraints that are not often addressed in journal publications. For instance, careful consideration of reviewed aspects in resource allocation, optimal TP design and required phenotyping efforts, choice of genotyping platform, redesign of the breeding pipeline, and detailed cost-benefit analysis are required for the adequate implementation of GS.

EXAMPLES OF GENOMIC SELECTION IMPLEMENTATION IN LIVESTOCK BREEDING AND COMMERCIAL PLANT BREEDING

There are empirical results showing the effectiveness of GS in the livestock sector and commercial plant breeding. Genomic selection was initially proposed in the context of animal breeding and subsequently rapidly adopted due to the reproductive biology, production system and economics of these species (Georges et al. 2019). Genomic selection has long been feasible in livestock due to higher investment return from higher individual value and more benefits from early selection and reduced generation interval (Meuwissen et al. 2016). Genomic selection was first adopted in dairy cattle in 2008, where it replaced the progeny testing and enabled a drastic reduction in generation interval (Wiggans et al. 2017). The generation interval for sires of bulls was reduced from seven years to less than 2.5 years and the rate of genetic gain increased 50-100% for milk, fat, and protein yield (García-Ruiz et al. 2016). In beef cattle breeding, GS provided benefits for traits that are very expensive or difficult to measure, such as feed conversion efficiency and beef quality (Bolormaa et al. 2013). An advantage was also observed in swine breeding with increased selection accuracy for maternal traits (Lillehammer et al. 2011). Lastly, GS enabled increased efficiency in early selection for low heritability traits in poultry breeding (Wolc et al. 2016).

In plant breeding, multinational breeding companies are routinely using GS for major crops (Bernardo 2016, Hickey et al. 2017). Commercial maize breeding deployed GS along with precision phenotyping to develop maize hybrids for increased yield in drought-stress conditions in the western U.S. Corn Belt. The training data sets entailed drought-managed environments and target environments managed over multiple years that enabled predictions for all stages of the breeding programs (Cooper et al. 2014, Cooper et al. 2014). The key agronomic traits were yield performance under both drought and well-watered conditions, seedling emergence under stressful growing conditions, and resistance to various diseases (Gaffney et al. 2015).

Commercial seed companies rely on large-scale operations performed in well-equipped and centralized platforms with standardized protocols to implement GS. Cost efficiency is obtained by sharing genotyping platforms and workflows among the numerous breeding teams within the company. Additionally, the large number of running samples gives a cost advantage of 50-70% savings in the overall genotyping cost in comparison to individual breeding programs in underdeveloped nations (Xu et al. 2020). Besides economies of scale, a large amount of accumulated historical data stored in structured relational databases is leveraged by dedicated data analysts to develop large training sets representing many target environments, ultimately providing high prediction accuracy (Xu et al. 2020, Zhao et al. 2021).

WHY SHOULD GENOMIC SELECTION BE IMPLEMENTED IN PUBLIC-SECTOR PLANT BREEDING PROGRAMS?

The commercial sector is the main source of new cultivars under situations where the market and other conditions are favorable, mainly for commodity and vegetable crops in major production areas. The public sector holds a key role in the education of future plant breeders and scientists; advancement and communication of basic and applied research; and germplasm improvement. These reasons for the implementation of GS in the public-sector plant breeding programs are discussed below.

The public university is the main source of education for the next generation of agricultural scientists. Education and training in an academic and field setting is combined with research focusing on a specific breeding objective or methodology. Most of the plant breeding students in developed countries will pursue a career in the commercial sector (Baenziger 2019). According to Baenziger (2019), experiential learning needs to occur at a similar scale and level of automation as in the commercial sector. Furthermore, public breeding programs must use state-of-the-art methods, tools, and techniques to allow seamless transition for technically skilled professionals to act in multiple career paths. The methods and tools are the same and updated but usually used at a smaller scale due to the costs of the new technologies.

The public sector has a role in methodology development and research that requires long-term investments. While the commercial sector needs to make a profit to sustainably support the ongoing research and development of new products, the public sector has more freedom to pursue higher risk research endeavors. Publication in peer-reviewed journals is valuable to disseminate findings to other researchers working in the field. Making advancements in breeding methodology freely available to everyone helps to further the field and its application.

Public-sector breeding institutions have an essential role for smallholder farmers and minor crops. International breeding centers belonging to the Consortium of International Agricultural Research Centers (CGIAR) are the main source of new improved germplasm for smallholder farmers. The CGIAR is an excellent global open-source breeding network that applies knowledge and world-class research to address challenges in food, land, and water systems. The public sector has also been a main provider of new barley, oat, and pearl millet cultivars (Baenziger 2019). For many other crops, the main role evolved towards germplasm preservation and development by means of “pre-breeding” that creates parental lines with specific trait combinations that are used by the commercial sector for cultivar development (Carena 2013, USDA 2015).

Public breeding programs are also pivotal to breed new varieties of so-called “orphan crops”, i.e., some small grains, roots, tubers, trees, etc. (Frey 1996). In this context, the development of improved germplasm has extensive importance and social return to the respective cropping region of the species. Furthermore, public institutions are important contributors to applied research for cultivar development targeting geographic locations outside the major production areas, alternative management systems, and small markets that are not profitable enough to justify significant investments from the commercial sector, yet are still economically important for some regions (Shelton and Tracy 2017). For instance, the University of Minnesota (UMN) soybean breeding program develops new food-type soybean cultivars adapted to Minnesota. The targeted specialty markets are small-seeded soybeans for natto and soybean sprouts, large-seeded soybeans for tofu and soymilk production, and black-seeded soybeans for specialty food products. These market types can command a premium in the market place and provide an important source of revenue through export markets to some specific growing areas.

GENOMIC SELECTION IMPLEMENTATION IN THE PUBLIC SECTOR FOR CULTIVAR DEVELOPMENT

The implementation of GS for crop improvement in public-sector plant breeding programs has been slow with a few programs exploring this approach or partially using it (Hickey et al. 2017). One could notice that most of the GS studies in plant breeding focus on the factors influencing the predictive ability in the same manner there were numerous studies to map QTLs. As pointed out by Bernardo (2016), there have been far few published reports that deployed the identified QTLs to improve germplasm, likewise there have been only a limited number of articles published to-date describing realized genetic gains from GS in crop improvement (Table 1).

Previous studies have compared PS and/or traditional MAS to GS using empirical data involving crops and traits with different genetic architecture. The first empirical field results comparing GS to MARS targeted testcross performance for stover and yield indices in a biparental maize population of recombinant inbred lines, with larger realized gains with GS compared to MARS (Massman et al. 2013). Studies on biparental populations of tropical maize exhibited realized gains under drought conditions attained through GS (Beyene et al. 2015, Vivek et al. 2017). Other studies demonstrated similar responses to selection per cycle with GS and PS, but with a shorter cycle and reduced costs under GS (Table 1). With the goal of conserving genetic diversity while achieving high genetic gains, Zhang et al. (2017) implemented GS in multi-parental tropical maize populations originating from crossing among 18 CIMMYT (International Center for Maize and Wheat Improvement) elite lines (Zhang et al. 2017) and CIMMYT-Asia synthetic populations (Das et al. 2020, Das et al. 2021). In this sense, the CIMMYT maize breeding program is developing a phased GS implementation with a test-half-predict-half (THPH) strategy that has been cost effective since 2018 (Beyene et al. 2019, Atanda et al. 2021). Moreover, the program is currently adapting approaches for the use of across-year and across-breeding programs data to eventually eliminate the need of the first-year yield testing (Atanda et al. 2021).

Besides maize, most applied GS experiments to date have been in self-pollinated small grains crops. One example of routine implementation of GS for cultivar development is the barley breeding program at the UMN. The program routinely relies on genomic predictions of Fusarium head blight (FHB) resistance, yield, winter hardiness, and malting quality, and has done so since 2010 (Bernardo 2016). In small grains, Asoro et al. (2013) reported increased genetic gain for β-glucan concentration from GS implementation in elite oats. Meanwhile in wheat, Rutkoski et al. (2015) published an empirical selection experiment over two cycles of GS for quantitative adult plant resistance to stem rust using a historical population derived from CIMMYT stem rust screening nurseries. Genomic selection for processing and end-

Table 1. Summary of the main features and results from published experimental studies reporting realized genetic gains from GS

References	Species	Trait	Number of genomic selection cycles	Realized genetic gain (per cycle, in t ha^{-1} or %)	Genomic/phenotypic selection ratio (%)
(Massman et al. 2013)	Maize	Yield + stover index	2	0.300 *	+3.9 †
(Combs and Bernardo 2013b)	Maize	Yield	4	-0.181 *‡	-2.9 ‡
(Asoro et al. 2013)	Oat	β-glucan concentration	2	1.12%	+1.2
(Beyene et al. 2015)	Maize	Yield under drought stress	2	0.086 *	+7.3
(Rutkoski et al. 2015)	Wheat	Stem rust severity	2	-5.84%	-20.4
(Vivek et al. 2017)	Maize	Yield under drought stress	1	0.450 *	+24.5
(Zhang et al. 2017)	Maize	Yield	3	0.225 *	§
(Das et al. 2020, Das et al. 2021)	Maize	Yield under drought stress	2	0.245 *	+7.5

* Testcross values reported instead of population performance *per se*; † Realized genetic gains from genomic selection were compared to marker-assisted recurrent selection (MARS) instead of phenotypic selection; ‡ Lower reported values in the later genomic selection cycles were caused by an atypical short growing season due to late planting and early harvest; §Ratio not estimated because only realized gains from genomic selection were reported.

use quality traits has been routinely implemented since 2013 at the CIMMYT wheat breeding program (Dreisigacker et al. 2021). Since 2017, the UMN wheat breeding program has implemented GS for FHB at the F_5 stage to eliminate highly susceptible lines from entering the yield trials (Adeyemo et al. 2020). Similarly, the University of Nebraska wheat breeding program implemented GS for FHB resistance in the breeding nursery and for grain yield in the preliminary yield trials ($F_{3:6}$) (Belamkar et al. 2018).

In perennial species, GS is being deployed to accelerate the domestication of intermediate wheatgrass and improve domestication traits, grain yield, seed-related traits, and disease resistance (Bajgain et al. 2020). Genomic selection has also been implemented at three Next Generation Cassava Breeding institutions in Africa to reduce cycle times in cassava breeding (Wolfe et al. 2017). Cassava is a highly heterozygous non-inbred crop that is vegetatively propagated and higher rates of genetic gain are expected from prediction across generations and across TPs from different breeding populations or programs.

CHALLENGES OF GENOMIC SELECTION IMPLEMENTATION IN PUBLIC-SECTOR BREEDING PROGRAMS

While the programs and publications highlighted above are a source of optimism for future GS applications, challenges still remain and gaps exist between research and practical application in public programs.

Redesign of the breeding program

Taking full advantage of GS often requires redesigning conventional breeding programs through shortening the entire pipeline by skipping an initial year of field testing, or recycling breeding lines as parents earlier. Even more aggressive approaches require complete reorganization of the pipeline around GS under a two-part strategy with distinct components of population development through rapid cycling GS and product development through a series of field trials (Gaynor et al. 2017). It should be recognized that plant breeding programs are dynamic entities that evolved to their current state through trial and error over long periods of time. Because of this, there is often inertia that can slow their redesign. Plant breeders only change their pipelines under two main conditions: novelty – being able to do something that they were previously not able to; and efficiency – do something that was already possible but more efficiently (Baenziger 2019). The pipeline redesign is a disruptive process and should only be performed with strategies that allow GS to surpass the genetic gain rate achieved by PS or when GS is sufficiently less expensive than PS to ensure return on investment. Moreover, there are differences in the skill set of a typical breeder with a field-based background and the skills required to implement GS. Successful GS implementation depends on the resourcefulness of the breeder and ability to efficiently deploy genotypic, phenotypic, and environmental information based on detailed evaluation of distinct breeding strategies.

Cost of genotyping

Dramatic advances in sequencing technologies have allowed collection of high-density molecular marker information at less cost. Currently, SNP-based arrays have robust allele calling, cost-effectiveness per data point and high-throughput analysis. However, the estimated cost per sample for at least several hundred SNP markers is still $14 (in U.S. dollars) (Bernardo 2021). Additionally, there is the cost of the DNA extraction. The routine implementation of GS requires the genotyping of large populations for both the TP and selection candidates, with thousands of lines per year. For example, a typical public soybean or small grains program in the U.S. can include 10,000 progeny rows per year. Genotyping these at this price point would require an investment of hundreds of thousands of U.S. dollars per year. This can be a major limitation for small-scale breeding programs within developed and developing countries.

Turnaround time

Tight seasonal deadlines are a ubiquitous feature of agricultural research programs, especially plant breeding programs, that often dictate methodology. These deadlines are paramount to decisions in adapting new technologies and taking advantage of new opportunities. For example, plant breeding programs that began to use winter nurseries decades ago needed to alter planting and harvest times, as well as streamline data analyses to meet the new even tighter deadlines imposed by this new opportunity to advance generations more quickly (Duvick 1996). The same consideration needs to be taken with the turnaround time and complex bioinformatics involved in obtaining genomic predictions. For breeding programs that rely on third-party providers for genotyping, an estimate of turnaround time between tissue sampling and genomic predictions is about three months. It can be partitioned in one week for tissue sampling, one week for DNA extraction, one week for shipping, six weeks for genotyping by service provider, one week for quality control and statistical analysis, and one week to interpret the results and finalize selections (Bassi et al. 2016). The turnaround time can be shortened by the use of in-house genotyping. In-house genotyping allows for more flexibility and shorter turnaround time, but should be balanced with the entry expenses for the acquisition of the genotyping platform and availability of staff. Thus, deployment of GS has to be tailored to ensure enough turnaround time from tissue sampling in the seedling stage to the use of genomic predictions before pollination and crossing or decision on advancements to the next stage in the pipeline.

Human resources and temporary specialized personnel

In a recent survey on plant breeding capacity in U.S. public institutions, the majority of breeding programs (71.6%) reported the lack of advanced scientific personnel (allied scientists, statisticians, bioinformatics specialists, etc.) employed within the program (Coe et al. 2020). The absence of personnel with advanced knowledge and skills in breeding-related sciences to aid in GS implementation can be attributed to a lack of stable long-term funding. Public breeding program leaders are typically researchers with publication, advising, teaching, and outreach responsibilities on top of the responsibilities related to planning, managing, and conducting breeding activities. Consequently, breeding programs rely on post-doctoral associates and graduate students with the technical knowledge to carry out the data analyses. Coe et al. (2020) indicated that post-doctoral associates and graduate students spent more time on plant breeding research than on germplasm enhancement or variety development. These individuals are temporary by definition as they depend on short-term funding and move to other institutions after completion of their studies. Even though the constant turnover of students is expected, planning is essential to avoid detrimental effects on the routine implementation of GS and associated turnaround time and seasonal deadlines.

Risk of failure and assessment of success rate

Assessing success or failure of GS is difficult in the short term. The use of prediction accuracies in academic studies is adequate to evaluate the performance of a prediction model. However, the breeder is mostly interested in the percentage of the top individuals that are correctly selected or the correct culling of the low performing individuals. Low to moderate prediction accuracies might entail a risk of failure that is too high for a breeding program. Pondering that the possibility of failure is existent, the breeder might only find out the selections made were inadequate when a considerable amount of resources were already invested. Under PS, field observations and data are available yearly to support steering the selection decisions in the correct direction. For illustration, if a breeder makes poor PS decisions in a given stage (e.g. preliminary yield trials), the breeder will observe them right away in the next stage (e.g. advanced yield trials) and can adjust parental selection and breeding populations accordingly. In GS, on the other hand, the result of poor selection decisions may not be apparent until considerable resources, time, and effort have been expended.

OPPORTUNITIES AND TOOLS TO ADDRESS THESE CHALLENGES

In order to address the challenges described above, there are several relevant tools that can be assembled in a framework to enhance the rate of genetic gain. The optimal strategy to compile the technological options depends on the crop and particular features of the breeding program. We outline below some of the tools helpful for breeding programs to implement GS.

Genotyping platform and imputation

The efficient implementation of GS requires a high-throughput SNP genotyping platform that enables genotyping of large populations of progenies at reduced cost while generating clean and repeatable marker data. With remarkable advancements in next-generation DNA sequencing technologies and ongoing cost reduction of high-density marker panels, the cost of genotyping may be of minimal concern in future years. Current strategies such as reduced-representation sequencing and targeted genotyping-by-sequencing (GBS) have become one of the most promising approaches (Niedzicka et al. 2016). In the case of biparental populations with limited recombination events, a low-density (several hundred polymorphic and representative markers) and therefore low-cost genotyping platform could be enough to ensure that at least one marker is in LD with each causal polymorphism.

Another opportunity to reduce the genotyping costs is the use of marker imputation (Bassi et al. 2016). The parental lines are genotyped with high-density whole-genome or skim sequencing technology while the progenies are genotyped with an inexpensive low-density platform (Howie et al. 2011, Torkamaneh et al. 2018). Under the assumption of absence of double crossovers, the alleles present in the low-density marker set are combined with the high-density markers from the parental lines to impute the alleles at all remaining loci. This procedure can yield high-density marker information using inexpensive genotyping on large number of progenies (Jacobson et al. 2015, Gorjanc et al. 2017).

Reliable and streamlined data storage and decision support workflow

A workflow comprised of software for data storage, curation, and analysis is essential for the seamless implementation of GS under the tight seasonal deadlines that are a hallmark of cultivar development programs. The use of high-throughput genotyping and phenotyping platforms is expected to largely increase the amount of data available for breeding activities. The currently used tools to manage breeding programs might not be suitable to handle the amount of data required for GS and other techniques, with the need of improved software.

The specialized workflow should combine breeding data sources such as phenotypic, genotypic, pedigree and environmental data to inform selection decisions. In a nutshell, it should enable consistent storage, mining, and analysis of information from multiple genotypes, years, and locations in an interoperable and user-friendly manner. In order to automate the breeding routine processes, all data collection should also be digitalized.

Status and future prospects of next-generation data management analysis and decision support tools for crop improvement have been reviewed in detail by Rathore et al. (2018). Currently, there are very few systems meeting the described requirements and which are available to public-sector plant breeding programs, with no single system in widespread use (Cobb et al. 2019). BreedBase is a free comprehensive breeding management and analysis software available as a web application tool. The software is capable of designing field layouts, support the collection of phenotypic information, field tissue sampling for genotyping, storage of high-density genotypic information, and GS-related analytics (Mueller et al. 2019). Another free web-based tool 'Genotype investigator for genome-wide analyses' (Gigwa) is capable of storage, filtering, graphical visualization, and sharing capabilities for multiple users to access, query, and analyze genotypic datasets (Sempéré et al. 2019). However, it does not perform analysis and store pedigree information. Both listed softwares follow the standards of the public plant Breeding Application Programming Interface (BrAPI), created to ease data exchange and interaction between distinct systems (Selby et al. 2019). The adoption of BrAPI across platforms was a great step forwards, enhancing prospects for future developments in public breeding database and tool development. We feel that further development and support of such systems that seamlessly tie together relational marker and phenotype data storage along with analysis and visualization will be a key element to full-scale GS implementation in the public sector.

Publicly available training data

The cost to develop a reliable TP of sufficient population size and enough environments could be a limiting factor for small and medium-size breeding programs. The establishment of a shareable and standardized genotyping platform enables the partnership of different breeding programs towards the goal of GS implementation through easy sharing of training data. One successful case of standardization and shareable genotypic, phenotypic, and environmental information in a large network is the public institutions in the U.S. that established the Maize Genome to Fields Initiative (McFarland et al. 2020). Another option to reduce costs with the initial TP development is the incorporation of publicly available data sets (Spindel and McCouch 2016). Candidate individuals are selected from the larger set to compose the TP based on optimization algorithms (Akdemir et al. 2015) and the internal TP is built using a sliding window approach with gradual replacement of the publicly available data set by the genotypic and phenotypic information collected within the program. The contribution of the external data set decreases as more information is collected on the individuals of the program across the cycles.

FORMS OF GENOMIC SELECTION FOR CULTIVAR DEVELOPMENT AND POTENTIAL FOR NEAR AND LONG-TERM ADOPTION

When GS is implemented in plant breeding, there is a dynamic shift between markers and phenotypic information and the definition of a new role for phenotyping. Likewise, GS should be considered a tool in the breeding program in a similar manner to MAS rather than an entire replacement of PS. Therefore, phenotypic evaluations serve the purpose of calibrating statistical models for GS in addition to PS.

Any implementation of GS for cultivar development involves using a predictive model to identify the high-performing and low-performing individuals. The optimal pipeline stage at which to implement GS varies due to the differences among programs that make them unique, and differences among key target traits in terms of trait heritability and cost of phenotyping (Figure 2). Breeders have the opportunity to predict the expected genetic value of individuals in distinct stages of a breeding program. Different GS breeding schemes should be evaluated for each specific situation to understand the trade-offs in cycle time and selection accuracy and to optimize genetic gain. We illustrate below different forms of GS deployment for cultivar development ordered by what we believe to be their feasibility in the near-term, where feasibility is defined by cost, disruption to current breeding practices, and risk.

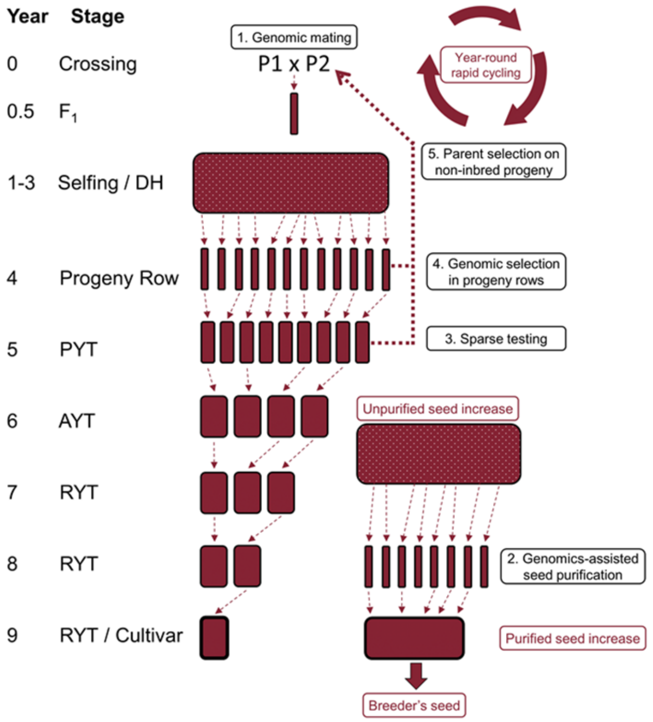

Figure 2. Layout of a conventional pure line breeding program adapted from Gaynor et al. (2017) and prospective forms to implement genomic selection ordered by ability for near-term adoption. DH: double haploid; PYT: preliminary yield trial; AYT: advanced yield trial; RYT: regional yield trial.

Genomic mating

Choosing the parents and parental combinations is perhaps the most critical decision a breeder makes, establishing the foundation of the entire program for years to come. The ultimate goal is to identify the specific parental combinations that would generate progeny with higher genotypic values, recognizing that the value of a cross is really defined by the performance of its best progeny rather than the mean progeny performance. Schnell and Utz (1975) proposed the usefulness criterion ($U_p = \mu + ih\sigma_A$) to quantify the expected mean of superior progeny as a function of the expected overall population mean, selection intensity, square root of the trait heritability, and the standard deviation of breeding values among the progeny resulting from the cross. Assuming that most elite lines are fixed for the same superior alleles, identifying those that can give rise to breeding populations of high genetic variance becomes critical. The conventional way to identifying crosses giving rise to high genetic variance is through using pedigree information to ensure unrelated breeding lines are crossed with one another. However, markers have opened up opportunities to identify specific parental combinations that complement one another based on actual allelic effects (Akdemir and Sánchez 2016).

Methods that rely on marker effects estimated from a TP using genome-wide prediction models and in silico simulations to predict the progeny genetic variance of a prospective cross were reported (Bernardo 2014, Mohammadi et al. 2015). Briefly, the progenies are simulated using the genotypic information of genotyped parents and modeling of recombination frequency and distribution with a genetic map. The simulated marker information of the progenies is combined with the marker effects obtained from a prediction model (e.g., RR-BLUP) to calculate the genomic estimated breeding values (GEBV) and the progeny variance is obtained as the sample variance of the GEBVs. In addition, deterministic equations were developed to predict the genetic variance of a prospective cross based on modeling the segregation and recombination of genome-wide marker effects (Zhong and Jannink 2007, Lehermeier et al. 2017, Osthushenrich et al. 2017), which match the simulated outcomes very well (Neyhart et al. 2019). The major advantage of the deterministic equations approach compared to in silico simulations is computational speed. A potential downward bias with underestimation of the true genetic variance was described when using the sample variance of GEBVs due to shrinkage of RR-BLUP marker effects when h^2 is <1 (Lian et al. 2015). In this sense, Lehermeier et al. (2017) suggested an analytical derivation for the genetic variance with a fully Bayesian estimate based on Markov chain Monte Carlo (MCMC) samples.

The issue of parental selection is relatively more important than questions related to number versus size of breeding populations. For instance, Bernardo (2003) obtained large selection responses under a wide range of combinations between number and size of breeding populations. Nevertheless, knowledge on the magnitude of the predicted genetic variance of a potential cross could be used to weight the population size. Breeders could optimize their allocation of resources by adjusting the size of the populations that will be created.

Cross selection based on predicted mean and genetic variance can be further expanded to multiple-trait improvement by predicting the expected genetic correlation (r_G) between two traits (Allier et al. 2019, Neyhart et al. 2019). If the prediction of the genetic correlation is accurate, ideal crosses could be distinguished based on the correlated progeny mean to improve two unfavorably correlated traits. Genetic correlation among traits is caused by shared genetic influence (i.e., pleiotropy) or non-random association of alleles (i.e., LD) (Falconer and Mackay 1996). Breeding efforts are functionally constrained in the first case but unfavorable correlations due to LD can be disrupted by successive recombination events. Neyhart et al. (2019) reported increased multi-trait genetic gain by 11–27% when compared to selection on the predicted cross mean when choosing crosses based on predicted r_G for unfavorably correlated traits FHB severity and plant height in barley. In a retrospective analysis for soybean seed yield and maturity, Jean et al. (2021) compared predicted performance of crosses and progeny persistence during selection. The authors observed that most crosses predicted to have superior performance were retained by the breeders and crosses with below-average predicted yield produced no superior advanced lines, demonstrating the utility of identifying crosses informed by r_G predictions.

One foreseen limitation for programs working on crops with less investment, and possibly complex polyploid genomes, is the lack of a genetic map. This investment in a genetic map may be advantageous because it can increase genetic gains by enabling a tool to design crosses that are more likely to produce improved lines while increasing the efficiency in resource allocation within the program. The empirical validation of predicted r_G requires phenotypic data on families with large population sizes due to greater error in estimation of higher order statistics, and therefore the empirical validation of such predictions may not be feasible on a regular basis.

Genomics-assisted seed purification

Pure lines are often derived from a single plant in an early generation of inbreeding (around F_3 or F_4) from a biparental cross. While the selection candidates undergo multi-location yield trials, unpurified seed increases occur simultaneously through generations of selfing and bulk harvesting to ensure seed availability for advanced yield trials. The bulk harvesting often starts at a generation in which not all loci are fixed (e.g., F_4 or F_5) and therefore genetic heterogeneity within breeding lines is common. To attain a cultivar of high genetic purity, a purification step is often undertaken on the basis of easily observable and highly heritable phenotypic traits (e.g., flowering time, maturity date, plant height, flower color, trichome color, etc.). This purification step creates a bottleneck that can shift allele frequencies in the genetically heterogeneous breeding line that was derived and tested. Genomic selection holds potential to leverage this genetic bottleneck in the subline selection to the breeder's advantage by selecting those sublines with best predicted performance and thus shifting allele frequencies in favorable directions during purification. Selection of new sublines outperforming the original heterogeneous "mother cultivar" was demonstrated by (Sebastian et al. 2010) using MAS, which could easily be extended using GS.

Reallocation of yield trial resources around Genomic Selection: Sparse testing designs

Genomic selection can be particularly useful for sparse testing in preliminary yield trials (PYTs) when there is a large number of selection candidates that need to be reduced for more extensive evaluation in replicated multi-environment trials (MET). The PYT is a prime stage for this application because of the large number of untested genotypes and limited amount of seed for field plots. Reducing replication of breeding lines within and across environments allows for increased selection intensity under fixed costs. Therefore, increased selection intensity or increased selection accuracy of phenotyped lines at low heritability by broader sampling of the TPE is expected. Under the supposition that the prediction might be less accurate than the observed phenotype, an increase in selection intensity with the low-cost screening of large population sizes can outweigh the loss in accuracy and result in higher genetic gain (Jarquin et al. 2020).

This approach proposes a shift from the phenotype being the selection criterion and the genotype as the unit of evaluation to a prediction based on allele effects, which means that the allele is now the unit of evaluation (Lorenz et al. 2011). The purpose of the field trial becomes the replication of alleles, with hidden replication of alleles increasing with population size (Lorenz 2013). With the new role of phenotyping being to provide useful data to train a genomic prediction model, field trial design with allocation of plot resources to genotypes and which individuals to evaluate in the trials becomes an important topic.

Individual lines can be evaluated without replication or partial replication across environments, but relatives are randomized among environments. The genomic predictions benefit from borrowing information through genomic similarity among relatives and from genotypes within and across environments by means of correlated environments (Burgueño et al. 2012, Jarquin et al. 2020). Sharing of information in this way effectively allows the evaluation of some lines in environments in which they were never tested, saving overall testing resources.

In light of this, Belamkar et al. (2018) performed a retrospective analysis and investigated the TP design for PYT in wheat breeding with varying numbers of lines from the same year as well as lines from previous years. They reported a substantially greater number of lines correctly selected for advancement when predictions were made with 50% of the lines from the testing year added to the TP. Verges and van Sanford (2020) corroborated this TP design result and also investigated the meaning of different prediction accuracies from a breeder's standpoint. With a 20% proportion of selected individuals common to the PYT stage, a prediction accuracy >0.4 could bring successful results in terms of identifying the top performing lines for a complex trait such as grain yield.

Endelman et al. (2014) found an advantage to spreading the PYT genotypes across multiple locations with limited replication compared to phenotyping a smaller number of genotypes with more replicates. This approach leverages the hidden replication of alleles by increasing the population size and testing more lines at different environments instead of all lines in one environment or fewer lines with higher replication in all environments. Lado et al. (2016) observed that the use of marker information from related lines evaluated in multiple environments was useful to predict the performance of new genotypes with high accuracy (0.5). To predict genotypes before phenotyping, the authors found the best strategy to be borrowing information from relatives evaluated in multiple environments and modeling the correlation matrix across environments to exploit the GxE interaction. Furthermore, Jarquin et al. (2020) found that the GS model accounting for GxE interaction yielded higher prediction accuracies and captured more phenotypic variability than the main effects models in different sparse testing allocation designs with varying levels of overlapping genotypes in environments.

Spatial models that account for field spatial heterogeneity in large PYT trials can be incorporated to reduce plot error variance within trials and obtain the best estimates of phenotype to be used to train the GS model (Gilmour et al. 1997, Rodríguez-Álvarez et al. 2018). Likewise, the use of linear mixed models accounting for spatial variation in field trials combined with the GS framework has potential to increase the efficiency of PYT selections (Lado et al. 2013, Belamkar et al. 2018).

Genomic selection in single plants or progeny rows

The benefits of GS can be exploited by increasing the selection accuracy for complex traits in early-generation screening. The progeny row stage involves screening of lines derived from a single plant and consequently a small amount of available seed planted in a single row or small plot, resulting in low heritability of trait measurements. Only observational selection for traits with high heritability are often carried out in this stage.

Genomic prediction can be deployed to increase the selection accuracy for complex traits by using the predictions as an additional year of yield selection. Advanced lines from the breeding program that were highly replicated in METs can be leveraged to design the TP. Under low to moderate prediction accuracies, GS still works as a pretest with reduced selection intensity by culling the low performing lines (Longin et al. 2015). Even without a reduction in breeding cycle time, applying GS among lines can enable higher selection accuracy and the early exclusion of low performing lines allows for more resources to be targeted to more promising lines at advanced testing stages.

Smallwood et al. (2019) compared context-specific GS strategies and PS in a F_5-derived recombinant inbred line (RIL) soybean population grown in a single plot in one location to simulate the progeny row stage. A subset of the RILs with similar maturity was further evaluated in replicated METs to validate the predictions in field conditions and test the effect of selection across generations for soybean yield and seed composition traits. Genomic selection strategies outperformed PS and showed strong potential for advancing fatty acid traits at this stage. No major differences in selection accuracy were found for yield, protein, and oil content but selections using GS can be made prior to harvest, improving the efficiency in comparison to PS. Similarly, no major differences in selection accuracy between GS and PS for seed oil concentration were observed by Hemingway et al. (2021) with biparental context-specific GS targeting the progeny row stage in four biparental soybean populations ($BC_1F_{4:5}$ lines) using a low-density marker panel.

Genotyping in early breeding generations is expected to be costly due to the large number of selection candidates. The number of lines in the progeny row stage varies among breeding programs but it can easily range from a few thousands to 20,000. These numbers mean that genotyping could be cost prohibitive for the budget of a public-sector plant breeding program. Genomic selection could be combined with a step of PS on a secondary trait to cull the lowest potential lines, allowing a reduction in the number of candidates to be genotyped. In this context, low-cost high-throughput phenotyping (HTP) could be used to evaluate lines for early season traits that are highly heritable and genetically correlated to yield to reduce the number of lines to be sampled and genotyped (Araus et al. 2018). For example, canopy coverage obtained from an unmanned aerial system has been reported to be correlated to seed yield in soybeans and thus is a good candidate for indirect selection (Xavier et al. 2017). Besides increased selection accuracy from the conventional visual selection method, the successful pairing of GS and HTP could increase the population size evaluated in the progeny row stage, allowing for higher selection intensity.

Parental selection based solely on the prediction of genotypic value

Parental selection is considered one of the most critical decisions breeders routinely face. The goal of GS here is to accurately predict breeding values to perform parental selection on progenies at early stages such as the progeny rows or first year of yield testing. This strategy is powerful but requires careful planning and consideration when applied in the early stages of the pipeline. Advanced elite lines with proven performance based on replicated field trials are currently used as new parents. The inbred recycling procedure aims to develop new and improved versions of elite inbreds. This task enables the use of GS without changes to the conventional breeding pipeline. Higher genetic gains through reduced cycle time are expected from early recycling of progenies as parents as quickly as the genotype is determined to have an above average breeding value for a quantitative trait (Heffner et al. 2010, Atlin et al. 2017, Cobb et al. 2019, Atanda et al. 2021). A breeding cycle can be determined as the time when a cross is made until selection of new parents.

This method was incorporated in a simulation of different breeding programs implementing GS at different stages of the pipeline (Gaynor et al. 2017). A given number of parental lines were selected as the ones with the highest GEBVs from a set of candidates that comprised the parents in the previous crossing block as well as the entries in the PYT and later stages of yield trials. The approach enabled a cycle time reduction from four years in the conventional program to three years in the alternative program that incorporated GS. The cycle time to selection of new parents can be further reduced to two years by selecting non-inbred progeny at the progeny row stage and yield trials.

An extreme form of parental selection based solely on predicted genotypic values is year-round rapid cycling GS (RCGS) (Asoro et al. 2013, Combs and Bernardo 2013b, Massman et al. 2013, Beyene et al. 2015, Rutkoski et al. 2015, Vivek et al. 2017, Zhang et al. 2017, Das et al. 2020, Das et al. 2021). Conventional recurrent selection is amenable to GS by using genome-wide markers to select the individuals that will be recombined instead of phenotypic measurements in the target environment. At least one generation of reliable phenotypic evaluation in the TPE is conducted in cycle 0, which also serves as the TP (Bernardo and Yu 2007). From calculated marker effects in cycle 0, the performance of individual plants is predicted prior to flowering and the best individuals are intermated in a year-round nursery or greenhouse to form the next cycle. Two cycles per year of RCGS mean that cycle time using GS is reduced by as much as eight-fold compared to phenotypic selection (Gaynor et al. 2017).

The effectiveness of RCGS was initially assessed in a temperate maize biparental mapping population, with higher realized genetic gains for indices of stover and grain yield from RCGS in comparison to MARS (Massman et al. 2013). Superior response was also reported in tropical maize multiparental populations with cycles of RCGS for drought and waterlogging stress (Zhang et al. 2017, Das et al. 2020, Das et al. 2021). For example, Zhang et al. (2017) performed four cycles and obtained genetic gain for grain yield of 0.225 t ha^{-1} per cycle, which is equivalent to 0.100 t ha^{-1} year^{-1} over 4.5 years from the initial cross to the last cycle.

Studies deploying computer simulations demonstrated that breeding cycles of population development can be separated from the product development component as early generation parents are selected solely on the basis of GEBVs (Gaynor et al. 2017). This strategy is a special case of RCGS and requires the reorganization of the breeding program in two distinctive components: 1) the population improvement component based on RCGS and 2) the product development component to identify new breeding lines using phenotypic evaluation. The population improvement through RCGS in early breeding generations relies on predicted breeding values to identify the parents of subsequent breeding cycles. This approach is expected to increase the frequency of favorable alleles and drive genetic gain in the base population. Meanwhile, individuals with the highest GEBVs are selected to derive lines and ensure a constant supply of improved germplasm. Inbreeding through single-seed descent or double haploidization is used to derive inbred lines that follow the conventional product development component, which might entail GS. The update of the TP and prediction models used in RCGS occurs with phenotypic and genotypic data gathered from the line testing stages in the product development pipeline. The two-part breeding strategy involves substantial restructuration of the breeding program taking into account costs, infrastructure, logistics, number, and size of populations.

Depending on the biology of the crop, genetic gain could be further increased by increasing the number of cycles per year used in RCGS. Rapid generation advance or “speed breeding” relies on intensive greenhouse or growth chamber use to accelerate plant development under prolonged photoperiod and controlled temperatures and greatly shortens generation time (Watson et al. 2018, Jähne et al. 2020). Under these conditions, phenotyping or use of MAS for basic adult plant traits is possible. The two-part breeding strategy and a modified speed breeding approach are currently being combined in the CIMMYT wheat program to accelerate the development of high-Zn and competitive yield varieties (Dreisigacker et al. 2021).

While implementing RCGS may be possible in commercial or large public breeding programs, there are still challenges for breeding programs with less capacity. Practical implementation will face challenges in turnaround time, cost of genotyping, and access to state-of-the-art growth facilities for flowering synchrony. Long turnaround time between sample tissue collection and genomic predictions can be overcome with streamlined software workflow for data storage and predictive analysis besides in-house genotyping. One way to mitigate the cost of genotyping is to adjust the number of selection candidates per cycle for varying levels of selection intensity.

CONCLUSION AND PERSPECTIVES

Genomic selection has considerable potential for improving complex traits by reducing breeding cycle time and costs all while possibly increasing selection accuracy. Although genomic selection has been widely adopted in commercial settings and extensively studied in academic settings, actual full implementation in public plant breeding programs lags. The integration of genomic selection with conventional breeding in public plant breeding programs will impact current and future efforts in crop improvement. In addition, implementation would benefit the education of future plant breeders that are likely to be employed in multinational companies fully immersed in genomic selection. Moreover, breeding methodology research - widely communicated by public programs and thus critical to progress in the discipline of plant breeding as a whole – would benefit from full genomic selection adoption to help ensure research questions and findings are as close to state-of-the-art as possible. Gaps exist between research and practical application because challenges such as redesigning of the breeding program, cost of genotyping, turnaround time, temporary human resources, and risk of failure remain. We provided a review of pertinent findings reported in the literature along with a roadmap of applications of genomic selection for cultivar development across the different stages of the breeding pipeline. Advancements and new approaches for implementation of genomic selection will continue to emerge in applied breeding programs, and adoption in public-sector breeding programs will help to ensure that this methodology is not only used by a few select programs, but rather is available to many programs diverse in their objectives and size.

ACKNOWLEDGMENTS

The authors would like to acknowledge funding from the North Central Soybean Research Program (NCSRP). Cleiton Antonio Wartha was supported by the grant entitled “SOYGEN 2: Increasing soybean genetic gain for yield and seed composition by developing tools, know-how and community among public breeders in the north central US”. In addition, we thank Lucas Roberts for proofreading and editing a draft of the manuscript.

REFERENCES

Adeyemo E, Bajgain P, Conley E, Sallam AH and Anderson JA (2020) Optimizing training population size and content to improve prediction accuracy of FHB-related traits in wheat. **Agronomy 10:** 543.

Akdemir D and Sánchez JI (2016) Efficient breeding by genomic mating. **Frontiers in Genetics 7**: 1-12.

Akdemir D, Sanchez JI and Jannink JL (2015) Optimization of genomic selection training populations with a genetic algorithm. **Genetics Selection Evolution 47**: 1-10.

Albrecht T, Auinger HJ, Wimmer V, Ogutu JO, Knaak C, Ouzunova M, Piepho HP and Schön CC (2014) Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. **Theoretical and Applied Genetics 127**: 1375-1386.

Albrecht T, Wimmer V, Auinger HJ, Erbe M, Knaak C, Ouzunova M, Simianer H and Schön CC (2011) Genome-based prediction of testcross values in maize. **Theoretical and Applied Genetics 123**: 339-350.

Allier A, Moreau L, Charcosset A, Teyssèdre S and Lehermeier C (2019) Usefulness criterion and post-selection parental contributions in multi-parental crosses: Application to polygenic trait introgression. **G3: Genes, Genomes, Genetics 9**: 1469-1479.

Araus JL, Kefauver SC, Zaman-Allah M, Olsen MS and Cairns JE (2018) Translating high-throughput phenotyping into genetic gain. **Trends in Plant Science 23**: 451-466.

Asoro FG, Newell MA, Beavis WD, Scott MP, Tinker NA and Jannink JL (2013) Genomic, marker-assisted, and pedigree-BLUP selection methods for β-glucan concentration in elite oat. **Crop Science 53**: 1894-1906.

Atanda SA, Olsen M, Burgueño J, Crossa J, Dzidzienyo D, Beyene Y, Gowda M, Dreher K, Zhang X, Prasanna BM, Tongoona P, Danquah EY, Olaoye G and Robbins KR (2021) Maximizing efficiency of genomic selection in CIMMYT’s tropical maize breeding program. **Theoretical and Applied Genetics 134**: 279-294.

Atlin GN, Cairns JE and Das B (2017) Rapid breeding and varietal replacement are critical to adaptation of cropping systems in the developing world to climate change. **Global Food Security 12**: 31-37.

Badji A, Machida L, Kwemoi DB, Kumi F, Okii D, Mwila N, Agbahoungba S, Ibanda A, Bararyenya A, Nghituwamhata SN, Odong T, Wasswa P, Otim M, Ochwo-Ssemakula M, Talwana H, Asea G, Kyamanywa S and Rubaihayo P (2021) Factors influencing genomic prediction accuracies of tropical maize resistance to fall armyworm and weevils. **Plants 10**: 1-22.

Baenziger PS (2019) Modern plant breeding: a perspective from the public sector in the United States. In Zeigler R (ed) **Sustaining global food security: The nexus of science and policy**. CSIRO Publishing, Melbourne, p. 148-161.

Bajgain P, Zhang X and Anderson JA (2020) Dominance and G×E interaction effects improve genomic prediction and genetic gain in intermediate wheatgrass (*Thinopyrum intermedium*). **The Plant Genome 13**: e20012.

Bassi FM, Bentley AR, Charmet G, Ortiz R and Crossa J (2016) Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). **Plant Science 242**: 23-36.

Belamkar V, Guttieri MJ, Hussain W, Jarquín D, El-basyoni I, Poland J, Lorenz AJ and Baenziger PS (2018) Genomic selection in preliminary yield trials in a winter wheat breeding program. **G3: Genes, Genomes, Genetics 8**: 2735-2747.

Bernardo R (1994) Prediction of maize single-cross performance using RFLPs and information from related hybrids. **Crop Science 34**: 20-25.

Bernardo R (2003) Parental selection, number of breeding populations, and size of each population in inbred development. **Theoretical and Applied Genetics 107**: 1252-1256.

Bernardo R (2014) Genomewide selection of parental inbreds: Classes of loci and virtual biparental populations. **Crop Science 54**: 2586-2595.

Bernardo R (2016) Bandwagons I, too, have known. **Theoretical and Applied Genetics 129**: 2323-2332.

Bernardo R (2021) Upgrading a maize breeding program via two-cycle genomewide selection: Same cost, same or less time, and larger gains. **Crop Science** 1-12.

Bernardo R and Yu J (2007) Prospects for genomewide selection for quantitative traits in maize. **Crop Science 47**: 1082-1090.

Beyene Y, Gowda M, Olsen M, Robbins KR, Pérez-Rodríguez P, Alvarado G, Dreher K, Gao SY, Mugo S, Prasanna BM and Crossa J (2019) Empirical comparison of tropical maize hybrids selected through genomic and phenotypic selections. **Frontiers in Plant Science 10**: 1502.

Beyene Y, Semagn K, Mugo S, Tarekegne A, Babu R, Meisel B, Sehabiague P, Makumbi D, Magorokosho C, Oikeh S, Gakunga J, Vargas M, Olsen M, Prasanna BM, Banziger M and Crossa J (2015) Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. **Crop Science 55**: 154-163.

Bolormaa S, Pryce JE, Kemper K, Savin K, Hayes BJ, Barendse W, Zhang Y, Reich CM, Mason BA, Bunch RJ, Harrison BE, Reverter A, Herd RM, Tier B, Graser HU and Goddard ME (2013) Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in Bos taurus, Bos indicus, and composite beef cattle. **Journal of Animal Science 91**: 3088-3104.

Burgueño J, de los Campos G, Weigel K and Crossa J (2012) Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. **Crop Science 52**: 707-719.

Carena MJ (2013) Challenges and opportunities for developing maize cultivars in the public sector. **Euphytica 191**: 165-171.

Cobb JN, Juma RU, Biswas PS, Arbelaez JD, Rutkoski J, Atlin G, Hagen T, Quinn M and Ng EH (2019) Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. **Theoretical and Applied Genetics 132**: 627-645.

Coe MT, Evans KM, Gasic K and Main D (2020) Plant breeding capacity in U.S. public institutions. **Crop Science 60**: 2373-2385.

Combs E and Bernardo R (2013a) Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. **The Plant Genome 6**: 1-7.

Combs E and Bernardo R (2013b) Genomewide selection to introgress semidwarf maize germplasm into U.S. Corn Belt inbreds. **Crop Science 53**: 1427-1436.

Cooper M, Gho C, Leafgren R, Tang T and Messina C (2014) Breeding drought-tolerant maize hybrids for the US corn-belt: Discovery to product. **Journal of Experimental Botany 65**: 6191-6194.

Cooper M, Messina CD, Podlich D, Totir LR, Baumgarten A, Hausmann NJ, Wright D and Graham G (2014) Predicting the future of plant breeding: Complementing empirical evaluation with genetic prediction. **Crop and Pasture Science 65**: 311-336.

Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de los Campos G, Burgueño J, González-Camacho JM, Pérez-Elizalde S, Beyene Y, Dreisigacker S, Singh R, Zhang X, Gowda M, Roorkiwal M, Rutkoski J and Varshney RK (2017) Genomic selection in plant breeding: Methods, models, and perspectives. **Trends in Plant Science 22**: 961-975.

Daetwyler HD, Villanueva B and Woolliams JA (2008) Accuracy of predicting the genetic risk of disease using a genome-wide approach. **PLoS ONE 3**: e3395.

Das RR, Vinayan MT, Patel MB, Phagna RK, Singh SB, Shahi JP, Sarma A, Barua NS, Babu R, Seetharam K, Burgueño JA and Zaidi PH (2020) Genetic gains with rapid-cycle genomic selection for combined drought and waterlogging tolerance in tropical maize (*Zea mays* L.). **The Plant Genome 13**: e20035.

Das RR, Vinayan MT, Seetharam K, Patel M, Phagna RK, Singh SB, Shahi JP, Sarma A, Barua NS, Babu R and Zaidi PH (2021) Genetic gains with genomic versus phenotypic selection for drought and waterlogging tolerance in tropical maize (*Zea mays* L.). **The Crop Journal** (in press).

de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD and Calus MPL (2013) Whole-genome regression and prediction methods applied to plant and animal breeding. **Genetics 193**: 327-345.

Dreisigacker S, Crossa J, Pérez-Rodríguez P, Montesinos-López OA, Rosyara U, Juliana P, Mondal S, Crespo-Herrera L, Govindan V, Singh RP and Braun H-J (2021) Implementation of genomic selection in the CIMMYT global wheat program, findings from the past 10 years. **Crop Breeding, Genetics and Genomics 3**: e210004.

Duvick DN (1996) Plant breeding, an evolutionary concept. **Crop Science 36**: 539-548.

Endelman JB, Atlin GN, Beyene Y, Semagn K, Zhang X, Sorrells ME and Jannink JL (2014) Optimal design of preliminary yield trials with genome-wide markers. **Crop Science 54**: 48-59.

Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. **The Plant Genome 4**: 250-255.

Falconer DS and Mackay TFC (1996) **Introduction to quantitative genetics**. Longman, New York, 464p.

Frey KJ (1996) **National plant breeding study-I**. Iowa State University/ USDA, 25p. (Special Report 98).

Gaffney J, Schussler J, Löffler C, Cai W, Paszkiewicz S, Messina C, Groeteke J, Keaschall J and Cooper M (2015) Industry-scale evaluation of maize hybrids selected for increased yield in drought-stress conditions of the US corn belt. **Crop Science 55**: 1608-1618.

García-Ruiz A, Cole JB, Paul M, Wiggans GR, Ruiz-lópez FJ and Curtis P (2016) Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. **Proceedings of the National Academy of Sciences of the United States of America 113**: E3995-E4004.

Gaynor RC, Gorjanc G, Bentley AR, Ober ES, Howell P, Jackson R, Mackay IJ and Hickey JM (2017) A two-part strategy for using genomic selection to develop inbred lines. **Crop Science 57**: 2372-2386.

Georges M, Charlier C and Hayes B (2019) Harnessing genomic information for livestock improvement. **Nature Reviews Genetics 20**: 135-156.

Gianola D, de los Campos G, Hill WG, Manfredi E and Fernando R (2009) Additive genetic variability and the Bayesian alphabet. **Genetics 183**: 347-363.

Gilmour AR, Cullis BR, Verbyla Arūnas P and Verbyla Arunas P (1997) Accounting for natural and extraneous variation in the analysis of field experiments. **Journal of Agricultural, Biological, and Environmental Statistics 2**: 269-293.

Gorjanc G, Battagin M, Dumasy JF, Antolin R, Gaynor RC and Hickey JM (2017) Prospects for cost-effective genomic selection via accurate within-family imputation. **Crop Science 57**: 216-228.

Gorjanc G, Gaynor RC and Hickey JM (2018) Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. **Theoretical and Applied Genetics 131**: 1953-1966.

Guo Z, Tucker DM, Lu J, Kishore V and Gay G (2012) Evaluation of genome-wide selection efficiency in maize nested association mapping populations. **Theoretical and Applied Genetics 124**: 261-275.

Habier D, Fernando RL and Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. **Genetics 177**: 2389-2397.

Habier D, Tetens J, Seefried FR, Lichtner P and Thaller G (2010) The impact of genetic relationship information on genomic breeding values in German Holstein cattle. **Genetics Selection Evolution 42**: 1-12.

Heffner EL, Jannink J and Sorrells ME (2011) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. **The Plant Genome 4**: 65-75.

Heffner EL, Lorenz AJ, Jannink JL and Sorrells ME (2010) Plant breeding with genomic selection: Gain per unit time and cost. **Crop Science 50**: 1681-1690.

Heffner EL, Sorrells ME and Jannink JL (2009) Genomic selection for crop improvement. **Crop Science 49**: 1-12.

Hemingway J, Schnebly SR and Rajcan I (2021) Accuracy of genomic prediction for seed oil concentration in high oleic soybean populations using a low-density marker panel. **Crop Science** (https://doi.org/10.1002/csc2.20607).

Hickey JM, Chiurugwi T, Mackay I and Powell W (2017) Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. **Nature Genetics 49**: 1297-1303.

Howie B, Marchini J and Stephens M (2011) Genotype imputation with thousands of genomes. **G3: Genes, Genomes, Genetics 1**: 457-470.

Jacobson A, Lian L, Zhong S and Bernardo R (2014) General combining ability model for genomewide selection in a biparental cross. **Crop Science 54**: 895-905.

Jacobson A, Lian L, Zhong S and Bernardo R (2015) Marker imputation before genomewide selection in biparental maize populations. **The Plant Genome 8**: 1-9.

Jähne F, Hahn V, Würschum T and Leiser WL (2020) Speed breeding short-day crops by LED-controlled light schemes. **Theoretical and Applied Genetics 133**: 2335-2342.

Jannink JL, Lorenz AJ and Iwata H (2010) Genomic selection in plant breeding: From theory to practice. **Briefings in Functional Genomics and Proteomics 9**: 166-177.

Jarquin D, Howard R, Crossa J, Beyene Y, Gowda M, Martini JWR, Pazaran GC, Burgueño J, Pacheco A, Grondona M, Wimmer V and Prasanna BM (2020) Genomic prediction enhanced sparse testing for multi-environment trials. **G3: Genes, Genomes, Genetics 10**: 2725-2739.

Jean M, Cober E, O'Donoughue L, Rajcan I and Belzile F (2021) Improvement of key agronomical traits in soybean through genomic prediction of superior crosses. **Agronomy Journal** (Doi: 10.1002/csc2.20583).

Kadam DC, Potts SM, Bohn MO, Lipka AE and Lorenz AJ (2016) Genomic prediction of single crosses in the early stages of a maize hybrid breeding pipeline. **G3: Genes, Genomes, Genetics 6**: 3443-3453.

Kadam DC, Rodriguez OR and Lorenz AJ (2021) Optimization of training sets for genomic prediction of early-stage single crosses in maize. **Theoretical and Applied Genetics 134**: 687-699.

Lado B, Barrios PG, Quincke M, Silva P and Gutiérrez L (2016) Modeling genotype × Environment interaction for genomic selection with unbalanced data from a wheat breeding program. **Crop Science 56**: 2165-2179.

Lado B, Matus I, Rodríguez A, Inostroza L, Poland J, Belzile F, del Pozo A, Quincke M, Castro M and von Zitzewitz J (2013) Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data. **G3: Genes, Genomes, Genetics 3**: 2105-2114.

Lehermeier C, Teyssèdre S and Schön CC (2017) Genetic gain increases by applying the usefulness criterion with improved variance prediction in selection of crosses. **Genetics 207**: 1651-1661.

Lian L, Jacobson A, Zhong S and Bernardo R (2015) Prediction of genetic variance in biparental maize populations: Genomewide marker effects versus mean genetic variance in prior populations. **Crop Science 55**: 1181-1188.

Lillehammer M, Meuwissen THE and Sonesson AK (2011) Genomic selection for maternal traits in pigs. **Journal of Animal Science 89**: 3908-3916.

Lin Z, Cogan NOI, Pembleton LW, Spangenberg GC, Forster JW, Hayes BJ and Daetwyler HD (2016) Genetic gain and inbreeding from genomic selection in a simulated commercial breeding program for perennial ryegrass. **The Plant Genome 9**: 1-12.

Longin CFH, Mi X and Würschum T (2015) Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. **Theoretical and Applied Genetics 128**: 1297-1306.

Lorenz AJ (2013) Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: A simulation experiment. **G3: Genes, Genomes, Genetics 3**: 481-491.

Lorenz AJ and Smith KP (2015) Adding genetically distant individuals to training populations reduces genomic prediction accuracy in Barley. **Crop Science 55**: 2657-2667.

Lorenz AJ, Chao S, Asoro FG, Heffner EL, Hayashi T, Iwata H, Smith KP, Sorrells ME and Jannink JL (2011) Genomic selection in plant breeding. knowledge and prospects. **Advances in Agronomy 110**: 77-123.

Lorenzana RE and Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. **Theoretical and Applied Genetics 120**: 151-161.

Lush JL (1937) **Animal breeding plans**. Iowa State College Press, Ames, 31p.

Massman JM, Gordillo A, Lorenzana RE and Bernardo R (2013) Genomewide predictions from maize single-cross data. **Theoretical and Applied Genetics 126**: 13-22.

Massman JM, Jung HJG and Bernardo R (2013) Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. **Crop Science**

53: 58-66.

McFarland BA, Alkhalifah N, Bohm M, Bubert J, Buckler ES, Ciampitti I, Edwards J, Ertl D, Gage JL, Falcon CM, Flint-Garcia S, Gore MA, Graham C, Hirsch CN, Holland JB, Hood E, Hooker D, Jarquin D, Kaeppler SM, Knoll J, Kruger G, Lauter N, Lee EC, Lima DC, Lorenz A, Lynch JP, McKay J, Miller ND, Moose SP, Murray SC, Nelson R, Poudyal C, Rocheford T, Rodriguez O, Romay MC, Schnable JC, Schnable PS, Scully B, Sekhon R, Silverstein K, Singh M, Smith M, Spalding EP, Springer N, Thelen K, Thomison P, Tuinstra M, Wallace J, Walls R, Wills D, Wisser RJ, Xu W, Yeh CT and De Leon N (2020) Maize genomes to fields (G2F): 2014-2017 field seasons: Genotype, phenotype, climatic, soil, and inbred ear image datasets. **BMC Research Notes 13**: 71.

Meuwissen T, Hayes B and Goddard M (2016) Genomic selection: A paradigm shift in animal breeding. **Animal Frontiers 6**: 6-14.

Meuwissen THE, Hayes BJ and Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. **Genetics 157**: 1819-1829.

Mi X, Wegenast T, Utz HF, Dhillon BS and Melchinger AE (2011) Best linear unbiased prediction and optimum allocation of test resources in maize breeding with doubled haploids. **Theoretical and Applied Genetics 123**: 1-10.

Mohammadi M, Tiede T and Smith KP (2015) Popvar: A genome-wide procedure for predicting genetic variance and correlated response in biparental breeding populations. **Crop Science 55**: 2068-2077.

Moreau L, Charcosset A and Gallais A (2004) Experimental evaluation of several cycles of marker-assisted selection in maize. **Euphytica 137**: 111-118.

Mueller LA., Menda N, Strickler SR, Saha S, Morales N, Flores M, Tecle IY, Fernandez-Pozo N, Bauchet G, Ogbonna A, York T, Foerster H, Ellerbrock B and Hosmani P (2019) Breedbase. Boyce Thompson Institute for Plant Research, Cornell University. Available at < https://data.nal.usda.gov/dataset/breedbase>. Accessed in August 2021.

Neyhart JL, Lorenz AJ and Smith KP (2019) Multi-trait improvement by predicting genetic correlations in breeding crosses. **G3: Genes, Genomes, Genetics 9**: 3153-3165.

Niedzicka M, Fijarczyk A, Dudek K, Stuglik M and Babik W (2016) Molecular Inversion Probes for targeted resequencing in non-model organisms. **Scientific Reports 6**: 1-9.

Osthushenrich T, Frisch M and Herzog E (2017) Genomic selection of crossing partners on basis of the expected mean and variance of their derived lines. **PLoS ONE 12**: e0188839.

Rathore A, Singh VK, Pandey SK, Rao CS, Thakur V, Pandey MK, Anil Kumar V and Das RR (2018) Current status and future prospects of next-generation data management and analytical decision support tools for enhancing genetic gains in crops. In Varshney R, Pandey M and Chitikineni A (eds) **Plant genetics and molecular biology**. **Advances in Biochemical/Engineering/Biotechnology.** Volume 164, Springer, Cham, p. 277-292.

Ravelombola WS, Qin J, Shi A, Nice L, Bao Y, Lorenz A, Orf JH, Young ND and Chen S (2020) Genome-wide association study and genomic selection for tolerance of soybean biomass to soybean cyst nematode infestation. **PLoS ONE 15**: e0235089.

Riedelsheimer C and Melchinger AE (2013) Optimizing the allocation of resources for genomic selection in one breeding cycle. **Theoretical and Applied Genetics 126**: 2835-2848.

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L and Melchinger AE (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. **Nature Genetics 44**: 217-220.

Riedelsheimer C, Endelman JB, Stange M, Sorrells ME, Jannink JL and Melchinger AE (2013) Genomic predictability of interconnected biparental maize populations. **Genetics 194**: 493-503.

Rodríguez-Álvarez MX, Boer MP, van Eeuwijk FA and Eilers PHC (2018) Correcting for spatial heterogeneity in plant breeding experiments with P-splines. **Spatial Statistics 23**: 52-71.

Roth M, Muranty H, Di Guardo M, Guerra W, Patocchi A and Costa F (2020) Genomic prediction of fruit texture and training population optimization towards the application of genomic selection in apple. **Horticulture Research 7**: 148.

Rutkoski J, Singh RP, Huerta-Espino J, Bhavani S, Poland J, Jannink JL and Sorrells ME (2015) Genetic gain from phenotypic and genomic selection for quantitative resistance to stem rust of wheat. **The Plant Genome 8**: 1-10.

Schnell FW and Utz HF (1975) F1-leistung und elternwahl euphyder züchtung von selbstbefruchtern. In **Bericht über die arbeitstagung der vereinigung österreichischer pflanzenzüchter**. BAL Gumpenstein, Gumpenstein, p. 243-248.

Sebastian SA, Streit LG, Stephens PA, Thompson JA, Hedges BR, Fabrizius MA, Soper JF, Schmidt DH, Kallem RL, Hinds MA, Feng L and Hoeck JA (2010) Context-specific marker-assisted selection for improved grain yield in elite soybean populations. **Crop Science 50**: 1196-1206.

Selby P, Abbeloos R, Backlund JE, Basterrechea Salido M, Bauchet G, Benites-Alfaro OE, Birkett C, Calaminos VC, Carceller P, Cornut G, Vasques Costa B, Edwards JD, Finkers R, Yanxin Gao S, Ghaffar M, Glaser P, Guignon V, Hok P, Kilian A, König P, Lagare JEB, Lange M, Laporte MA, Larmande P, Lebauer DS, Lyon DA, Marshall DS, Matthews D, Milne I, Mistry N, Morales N, Mueller LA, Neveu P, Papoutsoglou E, Pearce B, Perez-Masias I, Pommier C, Ramírez-González RH, Rathore A, Raquel AM, Raubach S, Rife T, Robbins K, Rouard M, Sarma C, Scholz U, Sempéré G, Shaw PD, Simon R, Soldevilla N, Stephen G, Sun Q, Tovar C, Uszynski G, Verouden M and Wren J (2019) BrAPI - An application programming interface for plant breeding applications. **Bioinformatics 35**: 4147-4155.

Sempéré G, Pételt A, Rouard M, Frouin J, Hueber Y, De Bellis F and Larmande P (2019) Gigwa v2-Extended and improved genotype investigator. **GigaScience 8**: giz051.

Shelton AC and Tracy WF (2017) Cultivar development in the U.S. public sector. **Crop Science 57**: 1823-1835.

Smallwood CJ, Saxton AM, Gillman JD, Bhandari HS, Wadl PA, Fallen BD, Hyten DL, Song Q and Pantalone VR (2019) Context-specific genomic selection strategies outperform phenotypic selection for soybean quantitative traits in the progeny row stage. **Crop Science 59**: 54-67.

Spindel JE and McCouch SR (2016) When more is better: how data sharing would accelerate genomic selection of crop plants. **New Phytologist 212**: 814-826.

Technow F, Riedelsheimer C, Schrag TA and Melchinger AE (2012) Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. **Theoretical and Applied Genetics 125**: 1181-1194.

Torkamaneh D, Boyle B and Belzile F (2018) Efficient genome-wide genotyping strategies and data integration in crop plants. **Theoretical and Applied Genetics 131**: 499-511.

USDA (2015) **USDA roadmap for plant breeding**. USDA, Washington, DC, 36p.

Verges VL and van Sanford DA (2020) Genomic selection at preliminary yield trial stage: Training population design to predict untested lines. **Agronomy 10**: 1-16.

Vivek BS, Krishna GK, Vengadessan V, Babu R, Zaidi PH, Kha LQ, Mandal SS, Grudloyma P, Takalkar S, Krothapalli K, Singh IS, Ocampo ETM, Xingming F, Burgueño J, Azrai M, Singh RP and Crossa J (2017) Use of genomic estimated breeding values results in rapid genetic gains for drought tolerance in maize. **The Plant Genome 10**: 1-8.

Voss-Fels KP, Cooper M and Hayes BJ (2019) Accelerating crop genetic gains with genomic selection. **Theoretical and Applied Genetics 132**: 669-686.

Watson A, Ghosh S, Williams MJ, Cuddy WS, Simmonds J, Rey MD, Asyraf Md Hatta M, Hinchliffe A, Steed A, Reynolds D, Adamski NM, Breakspear A, Korolev A, Rayner T, Dixon LE, Riaz A, Martin W, Ryan M, Edwards D, Batley J, Raman H, Carter J, Rogers C, Domoney C, Moore G, Harwood W, Nicholson P, Dieters MJ, Delacy IH, Zhou J, Uauy C, Boden SA, Park RF, Wulff BBH and Hickey LT (2018) Speed breeding is a powerful tool to accelerate crop research and breeding. **Nature Plants 4**: 23-29.

Whittaker JC, Thompson R and Denham MC (2000) Marker-assisted selection using ridge regression. **Genetical Research 75**: 249-252.

Wientjes YCJ, Veerkamp RF and Calus MPL (2013) The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. **193**: 621-631.

Wiggans GR, Cole JB, Hubbard SM and Sonstegard TS (2017) Genomic selection in dairy cattle: The USDA experience. **Annual Review of Animal Biosciences 5**: 309-327.

Wolc A, Kranis A, Arango J, Settar P, Fulton JE, O'Sullivan NP, Avendano A, Watson KA, Hickey JM, de los Campos G, Fernando RL, Garrick DJ and Dekkers JCM (2016) Implementation of genomic selection in the poultry industry. **Animal Frontiers 6**: 23-31.

Wolfe MD, Del Carpio DP, Alabi O, Ezenwaka LC, Ikeogu UN, Kayondo IS, Lozano R, Okeke UG, Ozimati AA, Williams E, Egesi C, Kawuki RS, Kulakow P, Rabbi IY and Jannink J (2017) Prospects for genomic selection in cassava breeding. **The Plant Genome 10**: 2017.03.0015.

Wong CK and Bernardo R (2008) Genomewide selection in oil palm: Increasing selection gain per unit time and cost with small populations. **Theoretical and Applied Genetics 116**: 815-824.

Xavier A, Hall B, Hearst AA, Cherkauer KA and Rainey KM (2017) Genetic architecture of phenomic-enabled canopy. **Genetics 206**: 1081-1089.

Xu Y, Liu X, Fu J, Wang H, Wang J, Huang C, Prasanna BM, Olsen MS, Wang G and Zhang A (2020) Enhancing genetic gain through genomic selection: from livestock to plants. **Plant Communications 1**: 100005.

Young ND (1999) A cautiously optimistic vision for marker-assisted breeding. **Molecular Breeding 5**: 505-510.

Zhang X, Pérez-Rodríguez P, Burgueño J, Olsen M, Buckler E, Atlin G, Prasanna BM, Vargas M, Vicente FS and Crossa J (2017) Rapid cycling genomic selection in a multiparental tropical maize population. **G3: Genes, Genomes, Genetics 7**: 2315-2326.

Zhao Y, Thorwarth P, Jiang Y, Philipp N, Schulthess AW, Gils M, Boeven PHG, Longin CFH, Schacht J, Ebmeyer E, Korzun V, Mirdita V, Dörnte J, Avenhaus U, Horbach R, Cöster H, Holzapfel J, Ramgraber L, Kühnle S, Varenne P, Starke A, Schürmann F, Beier S, Scholz U, Liu F, Schmidt RH and Reif JC (2021) Unlocking big data doubled the accuracy in predicting the grain yield in hybrid wheat. **Science Advances 7**: eabf9106.

Zhong S and Jannink JL (2007) Using quantitative trait loci results to discriminate among crosses on the basis of their progeny mean and variance. **Genetics 177**: 567-576.