Plant Science 242 (2016) 23–36

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.)

Filippo M. Bassi[a], Alison R. Bentley[b], Gilles Charmet[c], Rodomiro Ortiz[d,*], Jose Crossa[e]

[a] International Center for Agricultural Research in the Dry Areas (ICARDA), Av. Mohamed Alaoui, Rabat 10000, Morocco
[b] The John Bingham Laboratory, NIAB, Huntingdon Road, Cambridge CB3 0LE, United Kingdom
[c] Institut national de la Recherche Agronomique, INRA UMR1095 GDEC Clermont-Ferrand, 5 chemin de Beaulieu, 63039 Clermont-Ferrand, France
[d] Swedish University of Agricultural Sciences (SLU), Sundsvagen 14 Box 101, SE23453, Alnarp, Sweden
[e] Biometrics and Statistics Unit, International Maize and Wheat Improvement Center (CIMMYT), Apdo. Postal 6-641, 06600 Mexico DF, Mexico

ARTICLE INFO

Article history:
Received 16 April 2015
Received in revised form 23 August 2015
Accepted 27 August 2015
Available online 6 September 2015

Keywords:
Breeding value
Genetic gain
Genotype × environment interaction
Marker-aided breeding
Quantitative trait loci

ABSTRACT

In the last decade the breeding technology referred to as 'genomic selection' (GS) has been implemented in a variety of species, with particular success in animal breeding. Recent research shows the potential of GS to reshape wheat breeding. Many authors have concluded that the estimated genetic gain per year applying GS is several times that of conventional breeding. GS is, however, a new technology for wheat breeding and many programs worldwide are still struggling to identify the best strategy for its implementation. This article provides practical guidelines on the key considerations when implementing GS. A review of the existing GS literature for a range of species is provided and used to prime breeder-oriented considerations on the practical applications of GS. Furthermore, this article discusses potential breeding schemes for GS, genotyping considerations, and methods for effective training population design. The components of selection intensity, progress toward inbreeding in half- or full-sibs recurrent schemes, and the generation of selection are also presented.

1. Introduction

Classical breeding of wheat (*Triticum aestivum* L. and *T. durum* Desf.) has evolved dramatically in the last century. This has been the combined result of implementation of accurate experimental field designs, statistical methods, the development of doubled haploids (DH), the application of the concepts of quantitative and population genetics, and the integration of various plant sciences disciplines such as pathology, entomology, and physiology. This evolution has pushed the yearly genetic gain obtained through selective breeding (Δ_G) to a near-linear increase of 1% in potential grain yield [1,2]. Unfortunately, this impressive rate of gain is not sufficient to cope with the 2% yearly increase in the world population, which relies heavily on wheat products as source of food [3]. A solution is needed for this estimated 1% gap between production and demand.

In recent years, the deployment of molecular tools has been used as a means to accelerate yield gain. In particular, marker-assisted selection (MAS) to improve breeding efficiency has become commonplace in breeding programs [4]. Numerous MAS strategies have been developed, including marker assisted backcrossing [5–7] with foreground and background selection [8,9], enrichment of favorable alleles in early generations [10,11], selection for quantitative traits using markers at multiple loci [12,13], and across multiple cycles of selection [14]. Frisch and Melchinger [15] provide the selection theory for marker-aided backcrossing. Their research indicates that selection response depend on marker linkage map and parents' marker genotypes. Furthermore, the number of required marker data points will be reduced 50% by increasing population sizes from generation BC_1 to BC_3 and without affecting the proportion of the recurrent parent genome [16,17]. A 3-stage strategy for combining recombinant selection at markers flanking target gene with single-marker assays and genome-wide selection with high -throughput markers in BC_1 was more efficient than genome-wide background selection with high-throughput

Abbreviations: ΔG, genetic gain; BP, breeding population; CIMMYT, Centro Internacional de Mejoramiento de Maíz y Trigo; CP, coefficient of determination; DH, doubled-haploid; EC, environmental co-variable; GBLUP, genomic best linear unbiased predictor; GBS, genotyping-by-sequencing; GE, genotype × environment interaction; GEBV, genomic estimated breeding value (GEBV); GS, genomic selection; LD, linkage disequilibrium; ICARDA, International Center for Agricultural Research in the Dry Areas; MARS, marker-aided recurrent selection; MAS, marker-assisted selection; PS, phenotypic selection; QTL, quantitative trait loci; RILs, recombinant inbred lines; SNPs, single-nucleotide polymorphisms; TP, training population; TBV, true breeding value; VP, validation population.

* Corresponding author.
E-mail address: Rodomiro.Ortiz@slu.se (R. Ortiz).

http://dx.doi.org/10.1016/j.plantsci.2015.08.021

markers alone [18]. While this breeding technology has helped in keeping the yield gains from plateauing, there are no reports of wheat breeding programs achieving yearly gains above 1%. Some of the reasons why this technology has not led to step changes in genetic gain include affordability, marker availability, and the quantitative nature of many traits.

Wheat is an inbred cereal that generates small farm revenues, thus limiting its investment in research, and therefore the affordability of large-scale MAS programs. The detection of quantitative trait loci (QTL) for quantitative target traits (such as grain yield) is also limited by the precision of estimating QTL effects [19]. Bernardo [20] surmised that, of the 10,000 QTL identified in mapping studies in 12 major crop species, only a handful had been deployed for MAS in breeding. From the understanding of these limitations, and taking advantage of the ever-reducing cost of molecular markers [21], the concept of genomic selection (GS) was derived [22,23], with the specific intent of employing genome-wide marker data to effectively select for multi-genic quantitative traits early in the breeding cycle [24,25].

Genomic selection uses genome-wide markers to predict the breeding value of individuals. To perform GS, a population that has been both genotyped and phenotyped, referred to as the training population (TP), is used to train or calibrate a statistical model, which is then used to predict breeding or genotypic values of non-phenotyped selection candidates. This second set of individuals, which are genotyped but not phenotyped, is referred to as the breeding population (BP). The performances for various traits of the BP are therefore predicted using allelic identity with loci that were found associated with the phenotype in the TP. Intensively phenotyped and genotyped diverse lines from a breeding program provided a potential TP for robust calibration models [26].

The genomic estimated breeding value (GEBV) is derived on the combination of useful loci that occur in the genome of each individual of the BP and it provides a direct estimation of the likelihood of each individual to have a superior phenotype (i.e., high breeding value). Selections of new breeding parents are made based on the GEBV. This leads to shorter breeding cycle duration as it is no longer necessary to wait for late filial generations (i.e., usually F_6 or following in the case of wheat) to phenotyping quantitative traits such as yield and its components. A third set of individuals, defined as the validation population (VP), is genotyped and phenotyped. The GEBV is calculated for the VP, and its correlation to the actual phenotypic value is used to estimate the 'accuracy' of the GS model.

The expected gain from GS per unit time is defined as $\Delta_G = ir\sigma_A/T$, where i is the selection intensity, r is the selection accuracy, σ_A is the square root of the additive genetic variance, and T is the length of time to complete one breeding cycle [27]. Assuming equal selection intensities and equal genetic variance for both GS and phenotypic selection (PS), greater gain per unit time can be achieved as long as the reduction in breeding cycle duration by GS compensates for the reduction in selection accuracy. Given realistic assumptions of selection accuracies, breeding cycle times, and selection intensities, GS can increase the genetic gain per year compared to PS in both animal and crop breeding [28–32]. Moreover, for those traits that have a long generation time or are difficult to evaluate (i.e., insect resistance, bread making quality, and others) GS becomes cheaper or easier than PS so that more candidates can be characterized for a given cost, thus enabling an increase in selection intensity.

Here, we review the current knowledge accumulated for GS in various species, and use this to deliver practical recommendations on how to conduct wheat breeding using GS. Many of the topics presented in this article are still pending validation, and it will be stated throughout the text when unconfirmed results were used for deriving recommendations. This reflects the fact that many innovative wheat breeders are initiating GS today, before protocols are optimized, and thus we think would welcome a set of practical recommendations.

2. Lessons from animal breeding using genomic selection

In livestock, breeding values rank animals on genetic merit. Those sires and dams with the highest scores are the breeding stocks for the next generation. Genomic prediction has been used extensively in livestock breeding, particularly in dairy cattle [33], as a tool for predicting breeding values for quantitative traits using dense DNA markers throughout the genome [34]. It improves reliability by accounting for the inheritance of genes with small effects. The accuracy of prediction depends on the TP features such as size, marker number, trait heritability and relationship to the BP. For example, a higher accuracy and lower bias were noted in Norwegian Red Cattle for production traits with high heritability than for low-heritability health traits, which will require more records to achieve similar accuracy [35]. The accuracy of estimating breeding values in livestock may ensue solely from the ability of DNA markers to capture genetic relationships [36]. Likewise, this type of selection seems to be more accurate than phenotypic selection for low-heritability traits in juvenile animals, particularly when lacking phenotypic records, and may lead to reducing breeding costs [37].

Dairy cattle's breeding is particularly suited to the application of GS for two reasons. Firstly, breeding selection is more intense on males (bulls or sires), for which no phenotypic record is available (i.e., no milk production). Traditionally, dairy bull breeding values are estimated based on progeny testing, which takes time (until bulls have daughters and daughters produce milk). In contrast, genotyping and subsequent GS can be done at birth. Secondly, thanks to the global effort in recording the results of progeny testing for milk production, large phenotypic datasets were already available, and the addition of genotypes led to a comprehensive TP at marginal cost.

Simulation has been very useful for comparing methods with the aim of increasing the accuracy of estimating breeding values in livestock breeding [38]. Some private dairy breeding programs, particularly in Holstein cattle, are already marketing bull teams based on their GEBV when just two years old. Such an approach may lead to doubling the rate of genetic gain in dairy cattle breeding [39].

Thus genome-wide prediction of breeding values has become a standard method for selecting animals as parents for the next generation in livestock breeding. Still, GS is today a predominant reality only for those species where a single animal, like the sire, is sold at a high price. However, some of its concepts remain very relevant for the genetic enhancement of crops. The use of predicted breeding values in crop breeding, unlike livestock breeding, may further benefit from generating larger populations in a short time, by the various mating designs that can be implemented, and for easily producing pure lines, hybrids or clones [40]. In the case of crops, inbred lines or F_1 hybrids allow breeders to replicate a given genotype as many times as needed. Since no relative can be more related to an individual than itself, plant breeders rarely recur to progeny testing, which is instead common practice in animal breeding. Thus, when adapting GS approaches from animals to plants it is critical to understand that plant breeders can rely on replicated trials that ensure high accuracy in estimating the actual breeding value and in a relatively short amount of time, making PS quite efficient in crops. Furthermore, the existence of strong genotype × environment interactions and of complex population structure among plant populations, make the use of GS more challenging in plant breeding than in livestock.

3. Lessons from plant breeding using GS

After the initial work of Meuwissen et al. [23], Bernardo and Yu [24] were the first to show the impact of GS for plant breeding. The authors performed a computer simulation study showing that using the whole set of markers used for genotyping gave better prediction accuracy of breeding values than using only subsets of markers significantly associated with QTL (i.e., MAS). A few years later, de los Campos et al. [41], and Crossa et al. [42,43] showed the first genomic-enabled predictions in real plant breeding scenarios, showing it is possible to reach relatively high genomic predictions in several maize and wheat data sets. These authors were the first to demonstrate wheat predictions using pedigree as well as genomic relationship information as implemented in several parametric and non-parametric statistical models. After these initial results, a large amount of scientific work has been conducted and published to study prediction accuracy in several crop species.

The first proof that genomic selection works in plant breeding was given by Massman et al. [25], whose results from a breeding experiment demonstrated real genetic gains achieved through GS in a bi-parental temperate maize population derived from a cross between B73 and Mo17. This study involved genotyping 233 recombinant inbred lines (RILs) with 284 markers, evaluating the testcrosses under well-watered conditions and advancing the population using GS and marker-aided recurrent selection (MARS). The authors reported superior response to GS for stover yield, as well as stover and grain yield indices by 14 to 50% over MARS. Combs and Bernardo [44] performed five cycles of GS to introgress semi dwarf maize germplasm into US Corn Belt inbred lines and reported consistent gains from Cycle 0 (C_0) to Cycle 1 (C_1), but not beyond. More recently, Beyene et al. [45] showed the application of GS to improve tropical maize populations for grain yield under drought stress conditions. The authors compared GS with pedigree selection across eight biparental tropical maize populations, and reported that the average gain from GS per cycle across eight populations evaluated in drought environmental conditions in sub-Saharan Africa was 0.086 Mg ha^{-1}. They also reported that average grain yield of C_3-derived hybrids was significantly higher than that of hybrids derived from C_0. Hybrids derived from C_3 produced 7.3% higher grain yield than those developed through the conventional pedigree breeding method.

Hofheinz et al. [46] indicate that ridge regression based on preliminary estimates of trait heritability give an approximation of best linear unbiased prediction without losing accuracy. However, they noted that one cycle of a breeding program may not be suitable as an indicator for the accuracy of predicting lines of the next cycle unless traits show a high heritability. Ridge regression with shrinkage factors that were proportional to single-marker analysis of variance estimates of variance components or a modification of the expectation-maximization algorithm that yields heteroscedastic marker variances are alternatives to Bayesian methods, particularly when computational feasibility or accuracy of effect estimates are important [47]. Zhao et al. [48] used weighted best linear unbiased prediction (W-BLUP), which treats the effects of known functional markers more appropriately, to increase accuracy of prediction for model traits such as heading time and plant height in wheat.

Genomic selection seems to be a promising approach for hybrid breeding in self- pollinating crops such as wheat [49,50], particularly when heterotic pools are unknown. Longin et al. [51] give a perspective on hybrid wheat breeding according to quantitative genetics and indicate various factors affecting selection gains in hybrid and inbred line breeding. Expected selection gains are smaller in the former than in the latter, and also depend significantly on the hybrid seed production costs and the genetic variance available in hybrid versus line breeding [52].

4. What to keep in mind when attempting GS in wheat

The success of a GS breeding approach is determined by its ability to increase the rate of gain per unit of time, while maintaining affordable costs. The most cost-effective way to ensure the success of GS is to implement accurate selection in early generations. Hence, improving the 'accuracy' of GS has been the focus of many studies. When markers and QTL are in perfect linkage disequilibrium (LD), accuracy is determined by the TP size (N), heritability of the trait (h^2) in the TP, and the effective number of loci Me [53,54]:

$$r = \sqrt{\frac{Nh^2}{Nh^2 + Me}}$$

When QTL and markers are not closely linked, accuracy will be low unless the TP and selection candidates are closely related [55]. Thus, the relationship between the TP and the selection candidates is a key factor affecting accuracy [36,39,56–58]. In fact, simulation studies have shown that genetic architecture affects the relative performance of different GS models [54,59]. Other factors that can sometimes affect accuracies include (i) genotype-by-environment interaction between the TP and breeding target environments [60,61], (ii) choice of statistical model [62], (iii) marker platform [63,64], and (iv) genotype imputation method [65].

With so many variables to consider, it becomes nearly impossible to predict the accuracy of GS in advance. Ultimately GS strategies will need to be tested 'the hard way' by putting them into practice in breeding programs. Some initial recommendations can be, however, derived on the basis of results in other species.

5. Training population

First and foremost, the composition of the TP, its size, and its relatedness to the BP are key elements in determining the prediction accuracy of GS. The choice of individuals to include in the TP is one of the most difficult variables to optimize but it is pivotal to achieving high prediction accuracies. In general, the GS approaches that have shown the highest levels of accuracy have been those which had a TP with a large number of individuals, all highly related to the BP, and with limited population structure [66]. Doubling the size of the TP always increased accuracy when simulated mating designs based on 2-row spring barley lines [32], although the increase was dependent on whether or not QTL were observed (0.06–0.12 and 0.03–0.06 accuracy increase, respectively) and the analysis method used. Furthermore, the accuracy of GEBVs was reduced when predicting across breeds of cattle, lacking parental relatedness [67]. It seems that marker effects can be inconsistent due to the presence of different alleles, allele frequencies, linkage phases, and background effects (including epistasis; i.e., the interaction between two or more genes controlling a single phenotype) if the TP and BP are unrelated. The ideal TP would be therefore composed of full-sibs (or half-sibs) of the BP, and it should therefore be a priority of plant breeders to develop ad hoc TP for each BP. It is also important to keep this relatedness throughout the GS cycles. At each cycle of recombination and selection, the progenies of the BP could accumulate genetic diversity and gene frequencies could change to the point that the TP diverges from the BP. Therefore the breeder should be prepared to update the TP at each cycle [28], or use closed recurrent selection schemes with crosses occurring only between full-sibs or half-sibs. The first option has been widely discussed [68–73] and it has found the widest application thus far because of its simplicity of implementation, but as discussed it presents several limitations in terms of selection accuracies. Instead, a hybrid method merging these two options has been selected for developing the breeding schemes

presented here to ensure high accuracies in the second and third cycles of recurrent GS. In addition, inter-mating of full-sibs and half-sibs guarantees a rapid movement toward inbreeding and fixing of advantageous alleles. Moreover, as the TP in closed recurrent selection cycles remain related to the BP, it becomes possible to accumulate several phenotypic scorings for the BP over time and locations, and further improve the accuracy of the predictions [28]. To maximise the prediction accuracy of the model, only 'within' population (closed recurrent) selection is proposed here, while crosses occurring among populations (open recurrent selection) are considered a C_0 of a new GS breeding scheme. The new populations created at each cycle by artificial crosses (i.e., sub-populations) can be inter-mated to maintain genetic diversity, or not intermated to rapidly narrow the frequency of alleles segregating.

The level of LD should be similar in the TP and BP, and it has been shown that predictions are generally higher when LD is high, such as in heterozygous segregating filial generation (i.e., F_2 and F_3). However, low marker density can cause artificial overestimates of LD and when coupled with the near homozygosity of late filial generations (i.e. F_5 and following) causes a substantial decrease in prediction accuracy [32]. Hence, when TP and BP are not derived from the same parents, or are not at the same level of inbreeding, marker density must be increased to account for the increased effective population size and recombination rate [28].

There is a need for more research on the use of TPs that are not sibs of the BP and on the impact of the LD phase of the QTL effects across less related individuals [28]. Overall, GEBV accuracy decreases as more unrelated the TP is to the BP due to inconsistent QTL effects. However, since the key advantage of GS over phenotypic selection is the rate of gain per unit time, it is not realistic to slow the process by delaying the selection in the BP until phenotypic data can be obtained from a fixed (i.e., F_5 or later) TP, generated from the same cross as the BP. The use of historical data as TP is common practice in animal breeding, and has been tested with some success in plant breeding as well [74]. The advantage is that the historical data can be derived from field trials conducted on germplasm, which was then used as parents of the BPs under selection, and therefore a TP with low levels of relatedness to the BP is available for prediction from the start. These types of TP have been worded as 'far related' to indicate that are not derived from sibs of the BP, but rather by genotypes that contributed to the BP in some ways and share a certain degrees of parentage. Still, the allelic combinations within a given BP will only be sparsely represented in far related TPs, preventing the model from properly assessing epistasis and ultimately reducing the prediction accuracy. Various efforts are underway by the wheat breeding community to assemble large international phenotypic datasets that could be later used as universal TPs (e.g., by the Wheat Initiative's Expert Working Group on Wheat Breeding Methods and Strategies). TPs derived from historical datasets can be deployed in GS breeding schemes as long as it is clear that the accuracies will be low and therefore indexes should be used to weight the significance of the results. Here four GS breeding approaches are presented to take advantage of far related TPs in the early recurrent cycles, but that aim at achieving highly related TP by the later cycles.

Another critical issue concerning the TP is its size. In this sense, a general consensus does not exist in the literature, other than the TP should be as large as possible. In maize bi-parental populations, acceptable accuracy was achieved in bi-parental progenies using as few as 60 individuals [30]. Riedelsheimer et al. [71] showed least-squares estimates of accuracy of 0.59 based on a TP of 84 individual for full sib maize doubled haploid populations. Bentley et al. [75] tested the effect of TP size on prediction of key agronomic traits in elite European wheat using stratified bootstrap resampling [76] of TPs of 50, 100 and 200 individuals. Although this showed improved prediction as the TP size increased for key traits (height, flowering time, thousand kernel weight and protein content) the prediction of yield did not show marked improvement when increasing from 50 to 200 bootstrap samples in the TP. On the other hand, several hundred individuals would be required to achieve similar accuracy with a historical (less related) TP. The ideal size of a TP depends on several factors (heritability of the trait, level of relatedness, population structure, level of accuracy desired, and several others). To achieve accuracies above 0.5, based on an empirical estimation from previous works we suggest using a TP of at least 50 individuals that are full-sibs of the BP, 100 individuals for half-sibs, and at least 1000 individuals for a less related TP. In most cases the availability of multiple observations for each trait (e.g., multi-environments testing) can supplement the reduction of the number of individuals. Low heritability traits pose a problem for both phenotypic and genomic selection, and thus require larger TPs and number of test locations in order to maintain high accuracies [67]. A TP of 10,000 less related individuals is necessary to achieve an accuracy of 0.7 for a trait of heritability 0.3 [67]. GS outperformed both phenotypic and MARS when trait heritability was 0.2, 0.5 and 0.8, but GEBV accuracy decreased with decreasing heritability [24].

The strategy used for selecting the individuals of the TP is also important. The general rule of breeding applies also to GS predictions: if a population does not segregate for a trait it is not possible to improve it with selection. When extrapolating this concept to the TP, it is evident that a TP that does not have the same alleles segregating in the BP cannot be used to train a model for its prediction. More importantly, a TP that has been fixed for specific traits has lost the ability to predict for these traits in the BP. It is therefore important to select a TP that is related to the BP (i.e., with the same alleles), but also that has not been biased for any specific trait, unless the same bias is also imposed on the BP. There are examples in which a TP that had undergone PS against lodging and late flowering habitus was used to predict the performance of a BP that had not undergone any PS, and it resulted in a BP that was still segregating for these two undesirable traits [77].

An optimization method was recently proposed when a large set of candidates have been genotyped and the resources for phenotyping (i.e., TP size) are limited [78]. From the theoretical formula for coefficient of determination (R^2), Rincent et al. [78] developed an algorithm for optimally sampling the subset of lines that maximizes the crossbred performance. The usefulness of their method was subsequently demonstrated on both simulated and real maize data, wheat and rice data [66].

6. Genomic-enabled prediction incorporating genotype × environment interactions

In plant breeding, multi-environment trials for assessing genotype × environment interactions (GE) play an important role in selecting stable and high performance phenotypes. Despite the importance of GE in plant breeding trials, most GS research used single-environment or averaged means for prediction models. It was very recently that research demonstrated that multi-environment linear mixed models can account for correlated environmental structures within the genomic best linear unbiased predictor (GBLUP) framework and thus can predict performance of unobserved phenotypes using pedigree and molecular markers [79].

The approaches used to model GE have evolved over time with changes in the information available (e.g., molecular markers and the increased availability of environmental data) and advances in statistical and computational methods. With genetic information such as that given by molecular markers, several approaches have been developed [80,81] that allow incorporation of environmental (EC) and genotypic (GC) covariates, as first proposed by

Denis [82]. The quantity of markers and EC information that can be incorporated into these models is, however, limited. Recently, the GS models first introduced by Meuwissen et al. [23] were further extended into multi-environment models that can handle large numbers of individuals genotyped with large numbers of markers and evaluated in multiple environments. The first to use genomic prediction incorporating dense molecular markers as well as pedigree information was Burgueño et al. [79], who proposed analysis of multi-environment data using a multi-variate version of the genomic best linear unbiased predictor (GBLUP) and pedigree best linear unbiased predictor. They found that the multi-environment GBLUP outperformed the multi-environment pedigree-based mixed model, as well as that of single-environment pedigree or genomic models. Heslot et al. [83] presented genomic GE models that estimate marker effects in each environment separately.

Modern information systems can capture large volumes of environmental information such as, inter alia, temperature, radiation, soil fertility, or moisture, thus coinciding with the developmental phases of a crop. In principle, this information should be useful for incorporating GE into the analysis; however, modeling interactions between high-dimensional markers and high-dimensional environmental co-variables (EC) can be a very difficult task. Recently, Jarquín et al. [84] proposed dealing with interactions between markers and EC using random effects models, where main and interaction effects are modeled using Gaussian processes with covariance functions based on genetic and environmental similarity among entries. They used the structure induced by a reaction norm model as a covariance function, and applied the proposed approach to wheat data evaluated over multiple years and locations. Furthermore, 130 ECs were defined based on five phases of the phenology of the crop. The interaction terms accounted for a sizable proportion (15%) of the within-environment yield variance, and the prediction accuracy of models including interaction terms was substantially higher (20%) than that of models based on main effects only. Hence, methods that can capitalize upon the wealth of genomic and environmental information available are likely to become increasingly important. Likewise, when attempting to develop a practical implementation of a GS scheme for breeding of superior wheat cultivars, it remains compulsory to rely on multi-environment trials in order to have solid phenotypic values for the TP. In this sense where defined the terms preliminary yield trials (PYT) when the TP is only tested in one or few environments in yield plots of reduced size, and advanced yield trials (AYT) when the amount of seeds for the TP is sufficient to conduct multiple environments trials in yield plots of proper size and without the confounding effect of heterozygosity.

This reaction norm model [84] was used by Zhang et al. [85] with 19 tropical maize bi-parental populations evaluated in multi-environment trials used to assess prediction accuracy of different quantitative traits using low-density (~200 markers) single-nucleotide polymorphisms (SNPs) and genotyping-by-sequencing (GBS, ~2000 markers). Results showed that low-density SNPs (~200 markers) were largely sufficient to get good prediction in bi-parental maize populations for simple traits with moderate-to-high heritability, but GBS outperformed low-density SNPs for complex traits and simple traits evaluated under stress conditions with low-to-moderate heritability. Moreover, heritability and genetic architecture of target traits affected prediction performance. Prediction accuracy of complex traits such as grain yield were consistently lower than those of simple traits, e.g. anthesis date and plant height. Prediction accuracy under stress conditions was consistently lower and more variable than under well-watered conditions for all the target traits because of their poor heritability under stress conditions. Another important result of this study was that the prediction accuracy of GE models was found to be superior to that of non-GE models for complex traits and marginal for simple traits.

Recently Heslot et al. [83] described cases where crop modeling together with ECs can be used for genomic prediction incorporating GE. These authors used a crop model to derive stress covariates from daily weather data for predicted crop development stages and proposed an extension of the factorial regression model to genomic prediction. Lopez-Cruz et al. [86] proposed GS models that accommodate GE by explicitly modeling interactions between all available markers and environments. This M×E model can be easily implemented using existing software for GS and the model can be implemented using both shrinkage methods as well as variable selection methods. Also the M×E decomposes marker effects into components that are common across environments (stability) and environment-specific deviations. This information, which is not provided by standard multi-environment mixed models, can be used to identify genomic regions whose effects are stable across environments and others that are responsible for M×E.

7. Costs associated with genomic selection

Wheat breeders are faced with several practical considerations when deciding how and if to apply GS in their program. One of these is the cost of genotyping the TP and BP as compared to the cost of PS and the associated financial return on increased genetic gain via GS. One hectare of properly managed land can accommodate 1000 experimental plots of 4.5 m^2, which are sufficient to estimate differences in yield performance. The price for land management varies drastically from country to country. The two CGIAR centres that have received the global wheat breeding, namely Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT) and the International Center for Agricultural Research in the dry Areas (ICARDA), have an average cost of US$ 10,000 per ha; i.e., equal to US$ 10 per plot. Assuming that at least three environments and two replications are needed for controlling the error, the analysis of each individual for agronomical performances is estimated at US$ 60. To this cost can be added the price of conducting bread baking analysis, which is typically performed only on a subset of superior material, and the cost for storing the seeds and conducting artificial inoculations. All things considered, the PS cost reaches averages of approximately US$ 100 per individual.

Table 1 provides a calculation of the cost of various genotyping platforms that can be used for GS. In terms of genotyping, a few considerations can be derived from the work conducted in other species. The number of markers used has been demonstrated to only minimally affect the accuracy of the predictions, as long as at least one marker exists in LD with each QTL [71]. Rather than number, even marker distribution across the genome, their ability to tag important loci, and the generation at which they are used become important factors. In fact, in the early generations (F_2 or F_3) a large number of loci are in LD, meaning that a lower number of markers would be sufficient to predict the alleles at all loci. In later generations, this is no longer true and the number of markers deployed should rise accordingly. However, big steps forward have been made for in silico computation of the alleles of loci that are located in-between markers. This approach is defined as “imputation”. Practically, the parents of a population are genotyped at high density using specific platforms, while their offspring are only scanned at low resolution. Starting from the allele calling of the low-resolution genotyping, and assuming the absence of double crossovers, the alleles at all other loci can be imputed starting from the parental data. This allows the use of cheaper genotyping of progeny while still obtaining high-density genotyping matrices. Other considerations to be kept in mind when selecting the ideal marker platform for conducting GS are: (i) the necessity to

Table 1
Comparison of genotyping platforms for genomic selection of wheat.

Platform	Number of markers[a]	Cost per individual (US $)[b]	Advantage	Disadvantages	Provider
KASPAr	90	1	No missing data Co-dominant Scalable number of markers	Bi-allelic	LGC Genomics
Illumina 15K	10,000	35	No missing data Co-dominant	Bi-allelic	Various
Genotyping by Sequencing	10,000	12	Multi-allelic Co-dominant	Lots of missing data Massive data	Universities
DArT Seq	10,000	25	Multi-allelic Co-dominant	Missing data	Triticarte
Illumina 90K	15,000	50	No missing data Co-dominant	Bi-allelic	Various
Axiom 35K	20,000	50	No missing data Co-dominant	Bi-allelic	Affymetrix
Axiom 850K	300,000	250	No missing data Co-dominant	Bi-allelic Massive data	Affymetrix

[a] Expected number of polymorphic markers based on literature.
[b] Derived from quotes received in the past 12 months; each provider might change the price based on the population size or established collaborations. These estimates do not include the cost of DNA extraction nor of shipping plates to providers.

Table 2
Comparisons between various genomic selection (GS) schemes, their gains (Δ_G) and inbreeding levels among full-sibs (FS) and half-sibs (HS), assuming selection for yield (details and assumptions used for developing this table are presented in the section 'comparison of the proposed GS schemes').

	F_2 recurrent mass GS									
	F	N	FS	HS	G	S	TP	S.A.	S.I. (k)	Δ_G/h^2
On-season 2	F_2	1000	0.500	0.500	1000	40	Unrelated	0.30	2.66	0.80
Off-season 2	F_1	20	0.375	0.425						
On-season 3	F_2	500	0.188	0.213	500	80	Unrelated	0.30	1.52	0.46
Off-season 3	F_1	40	0.141	0.181						
On-season 4	F_2	400	0.070	0.090	400 + 500	80	Unrelated	0.30	1.40	0.42
Off-season 4	F_1	40	0.053	0.077						
On-season 5	F_2	400	0.026	0.038	400	80	PYT DH3	0.50	1.40	0.70
Off-season 5	F_1	40	0.020	0.033						
Total		2440	0.020	0.033	2800	280		0.35	1.74	2.37
	F_3 recurrent GS									
	F	N	FS	HS	G	S	TP	S.A.	S.I. (k)	Δ_G/h^2
On-season 2	F_2	20,000	0.500	0.500		325				
Off-season 2	F_3	750	0.250	0.250	750	40	Unrelated	0.30	2.06	0.41
On-season 3	F_1	20	0.188	0.213		20				
Off-season 3	F_2	10,000	0.094	0.106		150				
On-season 4	F_3	300	0.047	0.053	300 + 500	80	PYT $F_{4:5}$	0.40	1.22	0.33
Off-season 4	F_1	40	0.035	0.045		40				
On-season 5	F_2	10,000	0.018	0.023		150				
Off-season 5	F_3	300	0.009	0.011	300	80	AYT $F_{6:8}$	0.50	1.22	0.41
On-season 6	F_1	60	0.007	0.010						
Total		41,470	0.009	0.011	1850	560		0.40	1.50	1.15
	F_4 recurrent GS									
	F	N	FS	HS	G	S	TP	S.A.	S.I. (k)	Δ_G / h_2
On-season 2	F_2	20,000	0.500	0.500		1500				
Off-season 2	F_3	1500	0.250	0.250		325				
On-season 3	F_4	750	0.125	0.125	750	40	Unrelated	0.30	2.06	0.31
Off-season 3	F_1	20	0.094	0.106		20				
On-season 4	F_2	10,000	0.047	0.053		750				
Off-season 4	F_3	750	0.023	0.027		100				
On-season 5	F_4	200	0.012	0.013	200 + 500	80	AYT $F_{6:8}$	0.50	0.97	0.24
Off-season 5	F_1	40	0.009	0.011		40				
On-season 6	F_2	10,000	0.004	0.006		600				
Off-season 6	F_3	600	0.002	0.003		80				
On-season 7	F_4	160	0.001	0.001	160	80	$AYT_{multi\text{-}env.}$	0.60	0.80	0.24
Off-season 7	F_1	60	0.001	0.001						
Total		44,080	0.009	0.011	1610	1730		0.47	1.28	0.79

F = generation, N = number of individuals, G = number of genotypes, S = number of selected individuals, TP = training population, S.A. = selection accuracy, S.I. = selection intensity, h^2 = heritability.

be co-dominant in nature, as dominance effects in the heterozygotes cannot be distinguished by dominant markers; (ii) the overall absence of missing data as it can cause noise in the models, even though it can be partially controlled by imputation; (iii) the ability to discriminate two or more alleles. This last aspect is important in multi-parental schemes, where multiple alleles are segregating. Depending on the budget available and the computational capacities, different platforms can be employed. It is not the scope of this article to discuss which genotyping system is the best. Rather, the number of individuals to be phenotyped and genotyped has been presented for each GS breeding scheme (Table 2), together with the cost of genotyping an individual for each marker platform. The reader is then free to compute its cost of conducting GS based on the preferred platform and specific land charges. In fact, to the prices reported in Table 1 need to be added approximately US$ 3 per sample for DNA extraction. Also, the size of each scheme has been fixed to 50 individuals in each TP, with 10 TPs to be used to select from 10 BPs. It can be derived that at least 50 plots, replicated twice in at least four locations are needed for proper GS modelling that accounts for GE. This is a total of 4000 plots or approximately 4 ha of land. Furthermore, the BP will also need several cycles of normal PS in the field before reaching cultivar release consideration. It appears therefore that GS schemes in wheat will have similar costs to PS. Hence, unless the price per ha of land is significantly higher than US$ 10,000 or if the number of locations used for preliminary yield trials is much higher than three, wheat breeders should probably not consider switching to GS as a method to reduce the costs of their program. Instead, the potential of increasing the gain per unit time should be the real driver for a GS revolution in wheat breeding.

8. Wheat breeding practices missing from GS research

Gaps exist between some of the academic analyses designed to date to evaluate GS for wheat and the practical activities conducted in wheat breeding programs. In this section some of these incongruences have been listed together with suggestions on how to integrate them into GS schemes.

All wheat breeding programs derive their early segregating populations from several crosses of various parents. These crosses are then selected on the basis of 'among' population performances, meaning that a large number of crosses get discarded at each cycle. This aspect of 'among' populations selection is often given low importance in academic research [6,87]. In the schemes presented here this aspect has been integrated, both including steps of PS, or by using GEBV as indicators of overall performance of the populations. Colloquially, wheat breeders deploying GS need to understand that the use of GEBV will replace the phenotypic value in all its functions. Therefore, the selection of individuals can be made as 'indices' of GEBV from various traits can be made to provide selection priorities, e.g., on agronomics, host plant resistance, grain quality, and so on [88]; but also, a selection 'among' BPs can be made on the basis of their overall average and maximum GEBV. Hence, a BP with lower GEBV could be moved out of GS the same way that an individual with lower GEBV would be discarded over one with better performances.

Wheat breeding programs have to adapt their selection schemes to consider mere logistics. The implementation of shuttle breeding with two field seasons per year has become a reality for most spring wheat breeding programs, but this has required the optimization of seed movement, time of harvest and type of field experiments conducted. These considerations have also been included in the GS breeding schemes presented here. It can be estimated for wheat breeding programs that rely on service providers for genotyping that the turnaround time between tissue sampling and complete GEBV predictions is approximately three months, divided as: one week for leaf sampling, one week for DNA extraction, one week for shipping, six weeks for genotyping by service provider, one week for imputation and running the models, and one week is left to the wheat breeder for completing the selection. Based on these calculations, it needs to be ensured that the growing conditions will allow for three months to pass between the stages of two weeks old seedlings that can be sampled for tissue and the time of flowering. In most regions with cold or mildly cold winters this would not be a problem. In other instances, it could be considered to use instead artificially controlled environments where the temperature can be kept between 16 °C and 18 °C and daylight length is reduced to 6 h. Once the crosses have been completed, then the artificial environment can be accelerated again for the wheat to mature in time for the next sowing season. An additional consideration is the use of DH to accelerate inbreeding. This approach is rarely used for spring-types as in comparison with shuttle-breeding only provides minor acceleration of inbreeding, while completely loses the possibility of field selection. Instead, DH represents a clear advantage for winter-types. A mention has been made to indicate what GS breeding schemes would be more suitable for the use in winter-types and DH technology. However, in rapid-cycle recurrent programs as those described here it would be logistically very challenging to insert a step of DH. Rather it would be more convenient to utilize artificial environments like greenhouses or growing chambers to also ensure an off-season for winter-types.

In general, a rapid-cycle recurrent GS scheme aims at matching the 'generation of genotyping' with the 'generation of prediction' and the 'generation of crossing'. When these conditions are met the gain per cycle will be maximised. On the other hand, the 'generation of genotyping' of the TP might not match that of the BP, while the 'generation of prediction' of the BP is always after the 'generation of phenotyping' of the TP. These values can therefore be used in the future to summarize a breeding scheme for GS in wheat.

Another incongruence that exists between the plant breeders' approach and some of the conclusions presented in various articles is the use of accuracy alone as a method to determine the success of a GS approach. While accuracy is a good system to determine the performances of a model, it does not directly respond to the most important question a breeder has to ask of the data: how many of the top 5% individuals for a given trait were correctly selected? In fact, an accuracy of 0.4 seems to suggest that 40% (i.e., 8 individuals) were properly selected, which is a level of failure too high for a plant breeding program. However, this is not necessarily the case and often models with accuracies of 0.4 or 0.5 were instead capable of picking 60% or 70% of the top individuals [89]. This is probably due to better prediction ability when estimating the extreme values, and higher noise in estimating the average performances. Empirical gain from selection is therefore the only true measure, and predictions must be validated. Hence, it would be preferable if future studies of GS performances in wheat also include the percentage of top individuals correctly selected. However, it should be kept in mind that the phenotype is only a predictor of true breeding value (TBV), just as GEBV is. So selecting the very same individuals in GS and PS is not necessarily the grail to seek, since the proportion of the top 5% TBV selected by PS may not be higher than that selected by GS.

One final consideration is the progress toward inbreeding that occurs as progenies derived from the same parents (full- or half-sibs) are inter-crossed in the cycles 1, 2, and 3 of a GS recurrent program. The crossing of full-sibs will move the progenies toward inbreeding with a frequency of $\frac{1}{3}$ of the loci per cycle, while half-sibs at a rate of ¼. This is particularly important in wheat breeding where inbred lines are released as cultivars to farmers. Table 2 provides a summary of the inbreeding level at each step of GS closed recurrent schemes. In the following section, the terminology "sub" has been added to the word "population" to count the number of

subsequent crosses among full- or half- sibs from the same population; for instance: a sub-sub-population is derived by first crossing two parents, then mating two of their sibs, and then again mating progenies of those sibs.

The consideration of increasing inbreeding also has an effect on another component of the genetic gain: the selection intensity. In fact, wheat breeders apply different selection intensities depending on the inbreeding level reached at each cycle. Similarly, the breeding schemes for GS presented here have variable levels of selection intensity depending on the generation, inbreeding value, and improved accuracies due to updating of the TP.

9. Breeding schemes for genomic selection in wheat

9.1. F_2 recurrent mass GS

This scheme (Fig. 1) takes advantage of the increase in gain generated by shortening the length of each recurrent cycle. As no PS is conducted (see other schemes below), a larger number of lines have to be genotyped. However, due to the high LD that exists at the F_2 generation, platforms that mark as little as 100 loci can be employed to drive down the cost of genotyping in combination with imputation methods. However, this is true only for those cycles with low levels of inbreeding (Table 2). Also, nearly all steps can be conducted in the greenhouse, making this a method of choice for those regions that do not benefit of a field off-season or for germplasm that requires a long vernalization step. Due to the speed of cycling, the TP is mainly 'far related', but using the DH technology full- or half- sib TPs are generated in the fourth recurrent cycle.

C_0: 20 bi-parental populations are derived from artificial crossing of 40 parental lines, and 50 F_2 individuals each are grown in the greenhouse. All 1000 individuals are used for genotyping. On the basis of known markers-trait associations (i.e., association genetics or QTL analysis) the genes with major effects are identified and selected (i.e., MAS). Additionally, the same set of markers is used to estimate the genetic diversity within each population. Furthermore, on the basis of a 'far related' TP, the GEBV for various traits is predicted. Compiling the MAS data, and ensuring maximum GEBV and genetic diversity, four individuals from each population are selected and used in pairs for artificial crossing. Ten populations are discarded on the basis of their overall low GEBV values and lack of major genes segregating. To generate a full-sib training population, all the individuals that have been genotyped get harvested as single seed and grown out as F_3 to produce double haploids (DH_1) in on-season 3.

C_1: 20 sub-populations, derived from 10 initial populations (each represented by 2 full-sibs crosses, from 4 selected individuals), are grown in season 3 in the greenhouse. A total of 25 F_2 each are used for genotyping (500 individuals). In addition, all the 50 DH_1 produced for each of the 10 populations are also genotyped. In total, 1000 individuals are genotyped. The selection procedures follow the same directives as for C_0 and eight individuals are selected from each of the 10 sub-populations. Again, 10 sub-populations are discarded on the basis of overall GEBV values.

C_2: 40 sub-sub-populations, derived from selected 10 sub-populations (each represented by 4 full-sibs crosses, from 8 selected individuals) are grown in season 4 in the greenhouse. A total of 10 F_2 each are used for genotyping (400 individuals). Selection follows the same principle of C_0, with 20 sub-sub-populations discarded, and four individuals selected from each of the remaining 20 sub-sub-populations and used for crossing in pairs. Simultaneously, the training population has been increased sufficiently (DH_3) to perform preliminary yield trials at one location in 4.5 m^2 plots.

C_3: Since this scheme is particularly fast, a full forth cycle of recombination can be performed before entering normal breeding selection. In season 5, 40 sub-sub populations of 10 F_2 individuals each are grown in the greenhouse. These individuals are genotyped to undergo GS. At this stage, the GEBV are calculated on the basis of the PYT characterization of full-sib training populations (DH_3). Using these values, a variable number of individuals are selected from each sub-sub-sub-populations to recombine, depending on how many sub-sub-sub-sub-populations are required to enter normal breeding in season 6.

9.2. F_3 recurrent GS with phenotypic selection in F_2

The goal is to couple a step of PS, which allows to reduce the number of individuals to be genotyped, while taking advantage of the rapid gain per year of GS (Fig. 2). As per the F_2 scheme, the importance that is given to the GEBV changes at each cycle of selection based on the accuracy of the phenotyping and relatedness of the TP (i.e., indices).

C_0: 40 parents are crossed in pairs to generate 20 bi-parental populations. Approximately 1000 F_2 individuals are grown in the normal on-season in the field. Based on lodging, flowering time, and overall visual scoring a strong 'among' population selection is applied to discard one quarter (i.e., 5) of the crosses. For the remaining 15 crosses, 2.5% selection intensity (25 individuals) is applied 'within' populations. Two seeds from each selected F_2 individuals are grown under artificially controlled conditions, for a total of 50 individuals per population and 750 individuals total. The DNA is extracted from all individuals and used for genotyping. In this cycle, a full-sib TP has yet to be created and therefore it would not be possible to calculate GEBV with high-accuracy. Instead, major effect genes are fixed at this cycle by employing MAS: a subset of 100 or 200 selected markers can be used for this purpose. Also, this set of markers can be used to determine the level of genetic diversity between the full-sib progenies. Finally, GEBV can be estimated using a 'far related' TP. Two pairs of full-sibs (four individuals in total) are selected from each of the populations using an index: *i*) presence of major alleles based on MAS, *ii*) good genetic diversity, *iii*) larger GEBV. The five populations that have the smallest number of available major alleles, and the lowest overall GEBV will be discarded at this stage. The two pairs of full-sibs are artificially crossed to generate two cycle 1 sub-populations, each derived from one C_0 population. All 50 F_3 individuals from each population are harvested to later become the full-sibs TP. One F_4 seed from each individual is moved forward by means of single seed descent, 50 individuals per population total.

C_1: 20 sub-populations enter the first recurrent cycle, derived from 10 initial populations (each represented by 2 full-sibs crosses, from 4 selected individuals). Approximately 500 F_2 individuals are sown in the field during the 3rd off-season. Selection for flowering time, lodging and other visual scorings are applied to discard 5 sub-populations (from 5 different populations) and select 10 individuals from each of the remaining 15 sub-populations. Simultaneously, preliminary yield data are collected from $F_{4:5}$ training populations as PYT. Once yield data are collected, one single seed is propagated from each plot to generate F_6 individuals for the TP. During the 3rd on-season, 2 seeds for each of the 10 selected F_2 are sown in the greenhouse for a total of 300 individuals (15 sub-populations derived from 10 populations). Additionally, all the F_6 individuals from the 10 training populations are sown as single seeds to become the individuals of the TP. The DNA is extracted and large-scale genotyping performed on all individuals. Major effects are further fixed by means of MAS, while the GEBV are calculated using the PYT data. Using these values, only one sub-population from each initial population is selected, and within sub-populations eight individuals are used for crossing.

C_2: 40 sub-sub-populations (derived from 10 initial populations, each represented by 4 full-sibs crosses, from 8 selected individuals)

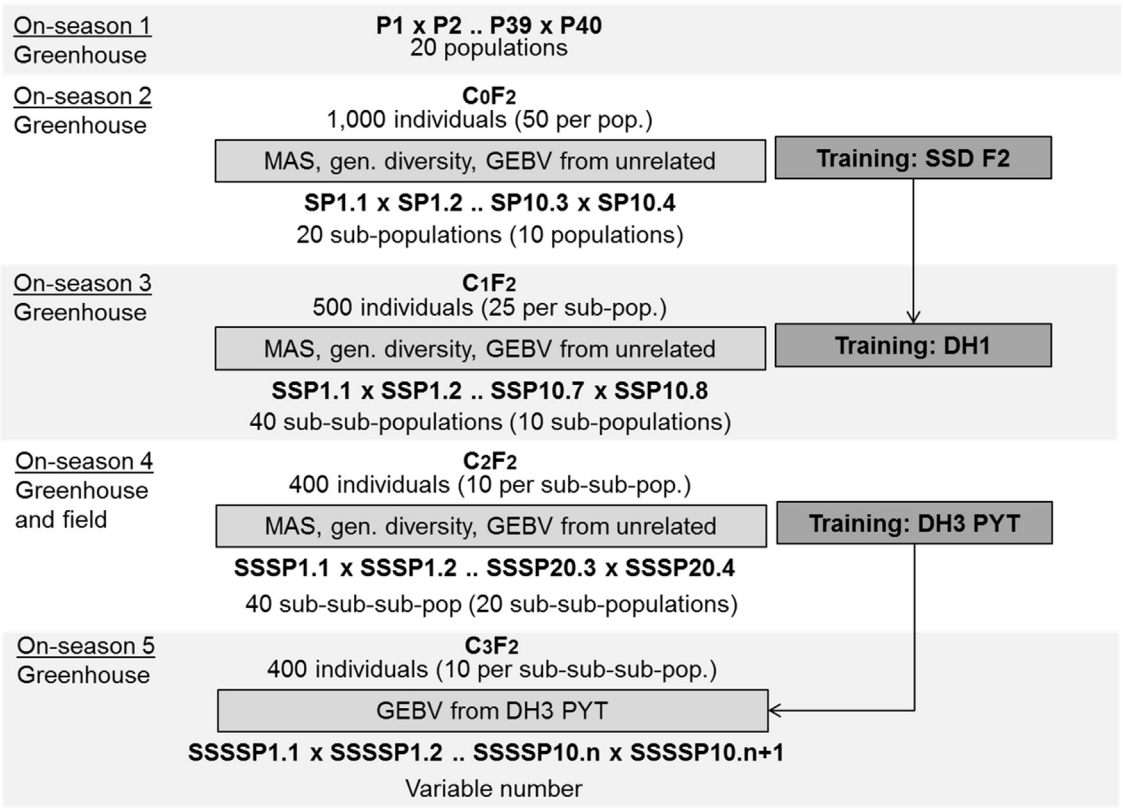

Fig. 1. Schematic representation of recurrent mass genomic selection in F_2. [P = parental line, C = recurrent cycle, SSD = single seed descent, SSP = sub-sub-population, DH = doubled haploid].

enter this cycle. Approximately 250 F_2 individuals each are sown in the field during the on-season 4 and will undergo phenotypic selection. One quarter of the sub-sub-populations is discarded, while for the selected ones a total of 5 F_2 individuals are identified and two F_3 seeds each are moved forward. Simultaneously, the individuals of the TP have been harvested as $F_{6:8}$ seeds in off-season 4 and the seeds have been increased sufficiently to conduct AYT. Additionally, baking quality data will be collected from at least two locations, and the response to diseases will be recorded during the on- or the off-season. One single plant will be harvested from each plot at one location to produce F_9 seeds. During the 5th off-season, the 300 selected F_3 individuals from the 30 sub-sub-populations and the 500 F_9 of the 10 training populations will be sown in the greenhouse. DNA is extracted and large-scale genotyping is conducted. At this cycle, the GEBV is calculated on the basis of the F_9 genotyping data and the field, quality, and disease scorings from the $F_{6:8}$. Ten sub-sub-populations are discarded on the basis of small average GEBV values. In the remaining 20 sub-sub-populations, six individuals are selected and used for mating.

C_3: a total of 60 sub-sub-sub-populations (derived from 10 initial populations, each represented by 6 full-sibs crosses, from 12 selected individuals). This material is now ready to enter normal breeding selection.

9.3. F_4 recurrent GS with phenotypic selection in F_2 and F_3

The aim is to use PS to keep the number of genotyping individual low. Also, postponing the genotyping generation compared to the F_3 scheme creates the advantage that a full-sib TP can be developed by C_1, while by C_2 large amounts of phenotypic data can be collected for the training population before doing the last cycle of recurrent selection. However, two additional years are required to complete the approach as compared to the F_3 scheme (Fig. 3).

Briefly, this breeding approach uses the same concepts of the F_3 cycling, but a second phenotypic selection step is conducted at the F_3 level. This second selection can be conducted in disease nurseries so that 'among' and 'within' populations selection can be done with higher accuracy. In C_0, the crossing is performed between individuals that are selected on the basis of MAS, genetic diversity and GEBV from an unrelated training population, as described for the F_3 scheme. The training population is derived from single seed descent of the F_4 individuals grown at C_0. However, the GEBV values at the end of C_1 are derived from genotypic data from F_7 training individuals and from PYT of $F_{5:6}$, while the GEBV in cycle 2 are predicted on the basis of multi years and multi locations trials of $F_{7:8}$ and $F_{7:9}$ individuals.

9.4. F7 recurrent GS with PS until F_6

The breeding PS is performed as normal until the first PYT in the 4th on-season as $F_{5:6}$. At this stage, one single F_6 plant from all populations that undergo field trials is also genotyped (Fig. 4). The combination of genotyping and phenotyping is used to make a GEBV prediction and determine the additive genetic component of each trait. The material with the highest GEBV is then intercrossed in the following season (off-season 4) and enters C_0. At the same time, only the selection is moved forward to the next stage of AYT. C_1 can follow the same procedure or take advantage of the large set of lines already genotyped and phenotyped. These would then become the TP for the following cycles and any of the previously described schemes can be deployed to accelerate the gain per year. Further, this is the ideal scheme to be included as first recurrent cycle for winter wheat types. The main difference would be that seasons 1 to 3 are replaced by two years for the development of a sufficient amount of DH seeds to conduct PYT assessment. There-

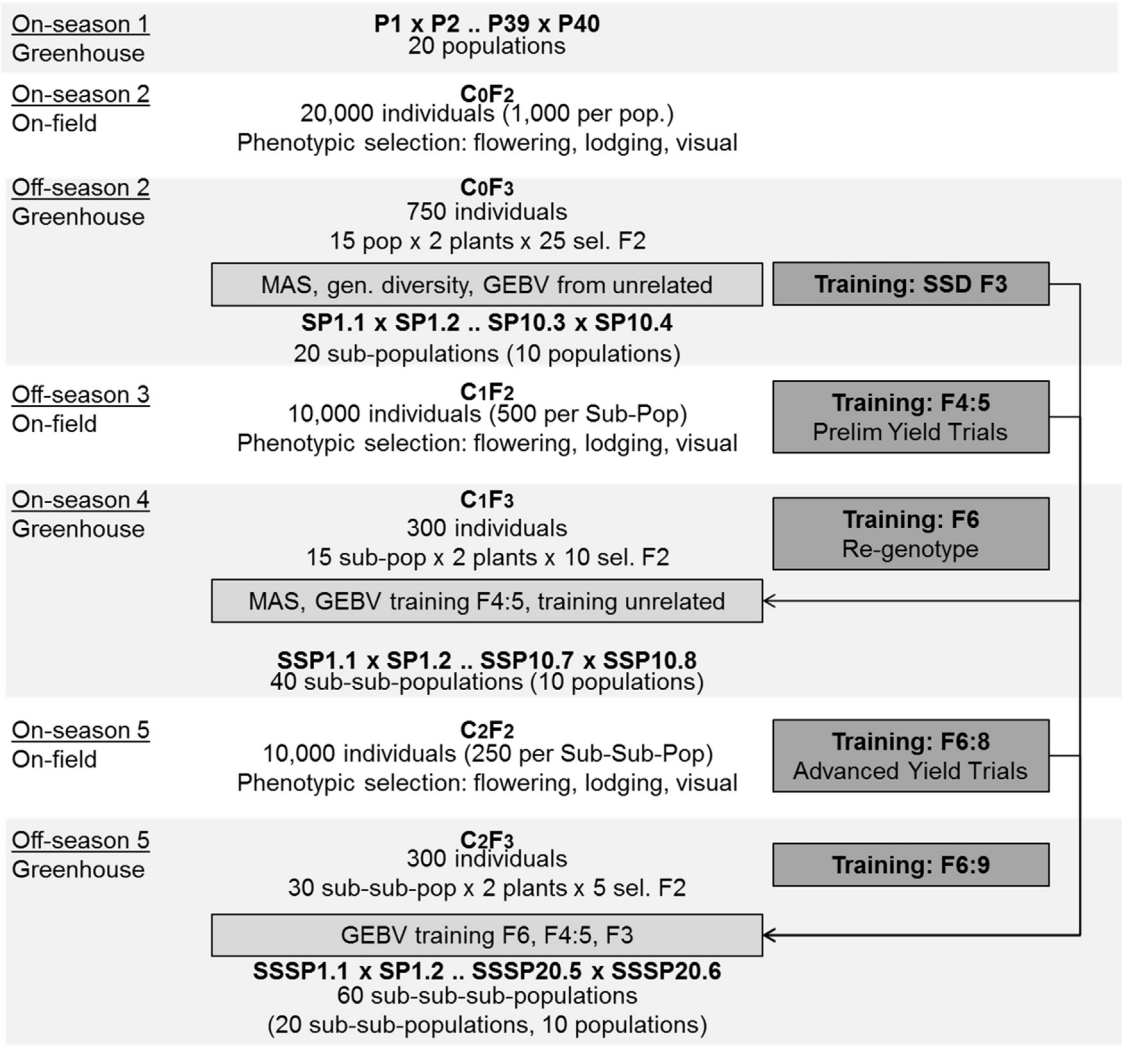

Fig. 2. Schematic representation of F_3 recurrent genomic selection. [P = parental line, C = recurrent cycle, SSD = single seed descent, SSP = sub-sub-population].

fore reducing the time per cycle by one year, but without imposing PS before the PYT.

10. Comparison of the proposed GS schemes

In order to compare the four GS schemes described, Table 2 shows their total gains, inbreeding levels when exiting GS, and total number of individuals to be phenotyped and genotyped. The values of Table 2 have been provided to adapt to any trait as the heritability values were not estimated, but all considerations have been made thinking about yield as the primary trait. The level of inbreeding is calculated considering hybridization occurring only within families (FS) or related families (HS). Self-fertilization increases inbreeding level by ½ at each cycle, FS mating by 1/3, and HS mating by ¼. The value for inbreeding relates only to those loci not under selection, PS nor GS. The alleles under marker selection can be maintained as heterozygous or fixed depending on the desire of the wheat breeder. The alleles of highly heritable traits can also be fixed by PS.

The values of selection accuracy are self-generated, using estimations from previous work. When the TP is unrelated or distantly related to the BP the accuracy tends to be low (between 0.2 and 0.4). Here, the value 0.3 was assigned. When the TP is fully related to the BP, but the field data are not sufficient to properly estimate GE (i.e., PYT) or there is still dominance (i.e., in the early generations) the accuracy tends to be between 0.3 and 0.5 and therefore an average value (0.4) was used. When the TP is fully related to the BP and at least 4 environments (AYT) are available, the accuracy is normally between 0.4 and 0.6 and the average value (0.5) was used. In cases where more than 4 environments are available the accuracy value used in the Table was 0.6.

Selection intensity value was calculated using the formula $i = S/\sigma P$ [90] where S is the selection differential and σP is the standard deviation of the trait. Based on Hallauer and Miranda [91], i can be estimated as k, function of p, which is the rate of selected individuals over the total number of individuals, when the population size is greater than 50 and it is independent of the trait under selection. The rate of genetic gain per year (Δ_G) is presented as a ratio of the heritability (h^2), ($\Delta G/h^2$) because different populations and different traits will have different heritability values, but this value is independent from PS vs. GS. In generating Table 2 it was assumed that no gain toward superior yield could be made by PS in early generations. The gain per year was calculated as

$$\frac{\text{Selection accuracy} \times \text{Selection intensity}}{\text{Years for one cycle}}$$

with the value of 'years for one cycle' set at 1, 1.5, and 2 for recurrent selection in F_2, F_3, and F_4 generations, respectively. Table 2 is not made for F_7 recurrent GS as its value will depend on the single wheat breeder decisions, but an estimate is provided in the section below.

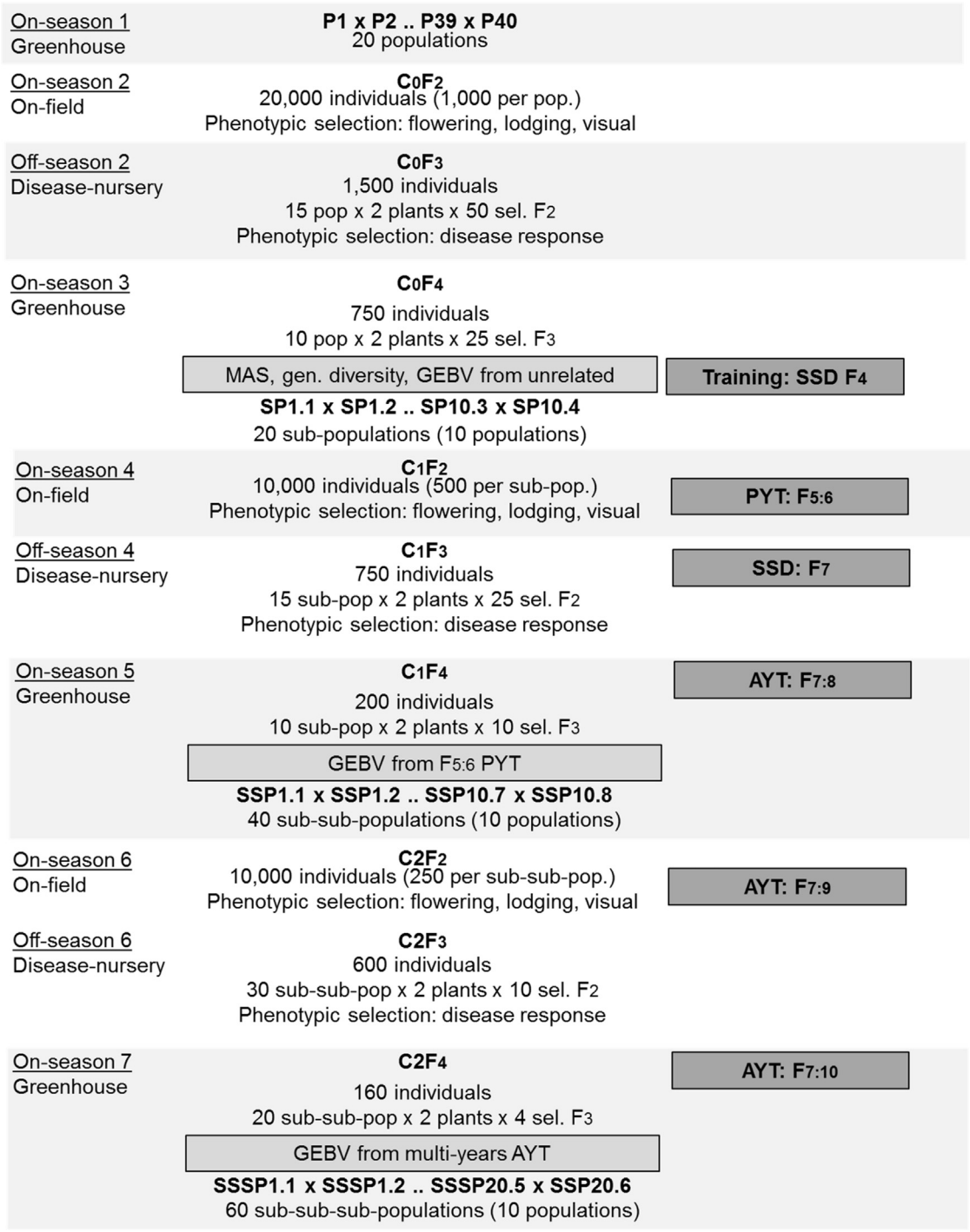

Fig. 3. Schematic representation of F_4 recurrent selection scheme. [P = parental line, C = recurrent cycle, SSD = single seed descent, SSP = sub-sub-population].

A classical breeding program for wheat applies an average selection intensity of 2.00 [89] and can assume a selection accuracy of 1 (as it is based on PS), and requires 5 years (F_8 generation) before undergoing recurrent crossing. Therefore the $\Delta G/h^2$ for a period of 10 years can be estimated as 0.8, equal to a rate of genetic gain of 0.08 per year. Operating F_7 GS recurrent scheme should guarantee similar outcomes as PS, with the reduction of the cycle's length by one year, but at lower accuracy than PS. The gain should be therefore similar. Since large scale PS is used, a very large number of populations can enter the initial stages of GS, while approximately only 10 or 20 populations should reach the phenotyping and genotyping stage in order to make the approach financially feasible. The main advantages are that it does not require significant changes to any normal breeding schemes, while still exploiting the power of genomic prediction to identify the true additive genetic value of the progenies under selection. The advantage of this scheme resides in its simplicity of implementation, the low cost of genotyping, and the possibility of using GS to replace MAS for complex and simple traits. Further, since larger populations can be handled through PS for simple traits, the effective size of the F_2 and following generations can be maintained larger, ensuring enough allelic combinations to pyramid several traits. Finally, this is an ideal scheme for winter types through the implementation of DH technology.

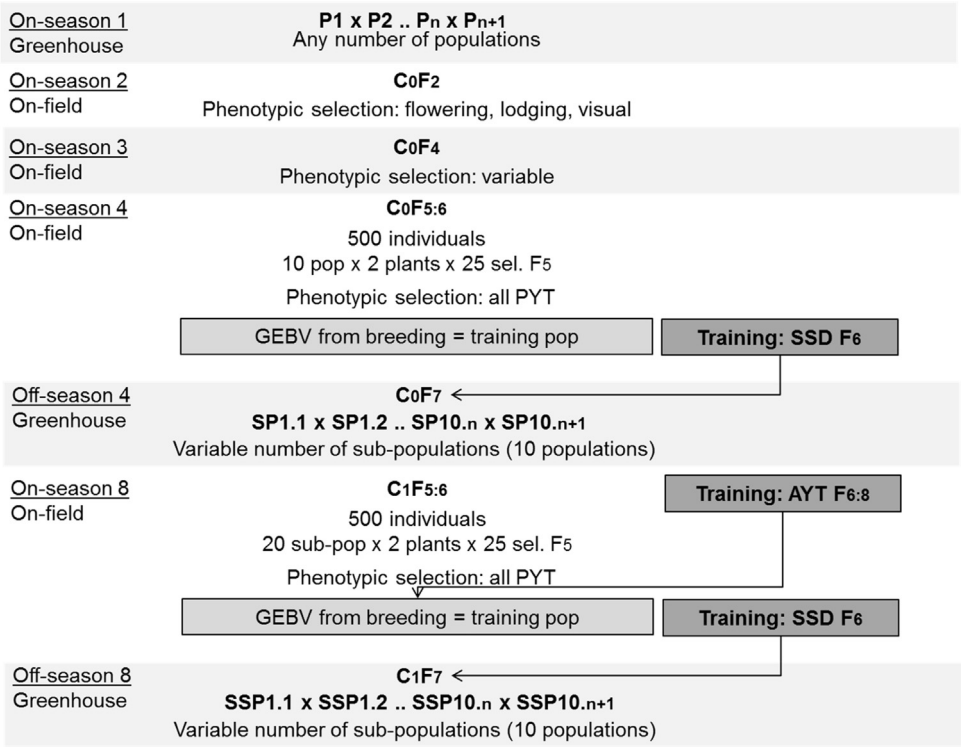

Fig. 4. Schematic representation of recurrent genomic selection in F_7. [P = parental line, C = recurrent cycle, SSD = single seed descent, SSP = sub-sub-population].

The F_2 recurrent scheme benefits from the highest accumulated total gain in a period of five years (2.37) equal to a rate of 0.47 per year, which is nearly six times higher than what achievable by classical breeding. The material selected this way will enter breeding after four full cycles of recurrent selection, which guarantees fixing of all major alleles and increased frequency of advantageous alleles of minor effect genes. Full-sib mating ensures that high levels of inbreeding are also reached. Also, since there is no PS, the population size has to be kept small, which would make very difficult to find progenies harbouring all positive alleles, but the use of recurrent cycles counters this negative effect and ensure selection of individuals with combinations of all superior alleles from the two initial parents. However, the cost of genotyping remains the highest, with 2800 individuals to be screened in a period of five years. Furthermore, costly DH needs to be generated, the average accuracies (0.35) are low, and the first time that the material undergoes PS is once it exits the GS scheme.

The F_3 and F_4 schemes are relatively similar, with a rate of genetic gain per year of 0.19 and 0.11, respectively. Compared to classical breeding, the F_3 scheme offers more than doubling of the yield gain, while the F_4 recurrent approach increases it by 1.4-fold. The average accuracies are higher in the F_4 (0.47) compared to the F_3 scheme (0.4), and similarly the number of individuals to be genotyped is lower in the F_4 scheme (1610) compared to the F_3 (1850). The three cycles of recurrent selection applied to various traits ensures that highly performing material exits the GS approach and then enters the final stages of breeding. Finally, the step of PS ensures that larger population sizes can be handled.

Apart from these specific considerations, one of the most interesting aspect of F_2, F_3, and F_4 closed recurrent GS schemes is that the material that exits these selection procedures is highly fixed in homozygosity, with inbreeding values well above 98%. Therefore the material can rapidly be multiplied and used for large-scale field evaluations, and eventually considered for variety release.

Conclusive data on the true gain as a result of GS in wheat remain elusive, although numerous papers and pieces of supporting evidence for this approach are starting to gather. In this article we have endeavoured to present sound guidelines for wheat breeders, allowing them to embark on the deployment of this new selection method. Each of the schemes presented has its advantages and disadvantages. Ultimately, it remains the wheat breeders' priority to select the scheme that best suit their needs, and transform the words reported here in more food for humanity.

Acknowledgments

Alison Bentley and Gilles Charmet are members of the Wheat Initiative (www.wheatinitiative.org) Expert Working Group on Breeding Methods and Strategies. Dr. Ian Mackay (NIAB) contributed to the calculation of gains per cycle from GS compared to PS. Dr. Jessica Rutkosky (Cornell Univ.) and Sandra Dunckel (Kansas State Univ.) provided insightful comments on an earlier version of this paper. Likewise, Professor M.E. Sorrells (Cornell Univ.) provided insightful comments, suggestions, and feedback of an earlier version of this manuscript. Filippo M. Bassi and Rodomiro Ortiz were partially funded by Vetenskapsrådet (VR, Sweden) Development Research during the writing of this manuscript. This work also benefits from funding by Mistra–Stiftelsen för miljöstrategisk forskning and SLU to Rodomiro Ortiz.

References

[1] F.-X. Oury, C. Godin, A. Mailliard, A. Chassin, O. Gardet, A. Giraud, E. Heumez, J.-Y. Morlais, B. Rolland, M. Rousset, M. Trottet, G. Charmet, A study of genetic progress due to selection reveals a negative effect of climate change on bread wheat yield in France, Eur. J. Agron. 40 (2012) 28–38.
[2] T. Fischer, D. Byerlee, G. Edmeades, Crop Yields and Global Food Security, ACIAR, Canberra, Australia, 2014.
[3] FAO, The State of the World's Land and Water Resources for Food and Agriculture: Managing Systems at Risk, Food and Agriculture Organization of the United Nations, Rome, Italy, 2011.
[4] S.D. Tanksley, N.D. Young, A.H. Paterson, RFLP mapping in plant breeding: new tools for and old science, Nat. Biotechnol. 7 (1989) 257–264.
[5] S.D. Tanksley, Molecular markers in plant breeding, Plant Mol. Biol. Rep. 1 (1983) 3–8.

[6] R. Bernardo, Parental selection, number of breeding populations, and size of each population in inbred development, Theor. Appl. Genet. 107 (2003) 1252–1256.
[7] Y. Xu, J.H. Crouch, Marker-assisted selection in plant breeding: from publications to practice, Crop Sci. 48 (2008) 391–407.
[8] J. Hillel, T. Schaap, A. Haberfeld, A.J. Jeffreys, Y. Plotzky, A. Cahaner, U. Lavu, DNA fingerprints applied to gene introgression in breeding programs, Genetics 789 (1990) 783–789.
[9] F. Hospital, C. Chevalet, P. Mulsant, Using markers in gene introgression breeding programs, Genetics 132 (1992) 1199–1210.
[10] D. Bonnett, G.J. Rebetzke, W. Spielmeyer, Strategies for efficient implementation of molecular markers in wheat breeding, Mol. Breed. 15 (2005) 75–85.
[11] N.K. Howes, S.M. Woods, Simulations and practical problems of applying multiple marker assisted selection and doubled haploids to wheat breeding programs, Eur. J. Plant Pathol. 100 (1998) 225–230.
[12] R.L. Fernando, M. Grossman, Marker assisted selection using best linear unbiased prediction, Genet. Sel. Evol. 21 (1989) 246–277.
[13] R. Lande, R. Thompson, Efficiency of marker-assisted selection in the improvement of quantitative traits, Genetics 124 (1990) 743–756.
[14] W. Zhang, C. Smith, Computer simulation of marker-assisted selection utilizing linkage disequilibrium, Theor. Appl. Genet. 83 (1992) 813–820.
[15] M. Frisch, A.E. Melchinger, Selection theory for marker-assisted backcrossing, Genetics 170 (2005) 909–917.
[16] M. Frisch, M. Bohn, A.E. Melchinger, Comparison of selection strategies for marker-assisted backcrossing of a gene, Crop Sci. 39 (1999) 1295–1301.
[17] V. Prigge, A.E. Melchinger, B.S. Dhillon, M. Frisch, Efficiency gain of marker-assisted backcrossing by sequentially increasing marker densities over generations, Theor. Appl. Genet. 119 (2009) 23–32.
[18] E. Herzog, M. Frisch, Selection strategies for marker-assisted backcrossing with high-throughput marker systems, Theor. Appl. Genet. 123 (2011) 251–260.
[19] J.C.M. Dekkers, F. Hospital, The use of molecular genetics in the improvement of agricultural populations, Nat. Rev. Genet. 3 (2002) 22–32.
[20] R. Bernardo, Molecular markers and selection for complex traits in plants: learning from the last 20 years, Crop Sci. 48 (2008) 1649–1664.
[21] K.A. Wetterstrand, (2014). DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP), Retrieved 12 June 2014, Available at: www.genome.gov/sequencingcosts
[22] C.S. Haley, P.M. Visscher, Strategies to utilize marker-quantitative trait loci associations, J. Dairy Sci. 81 (Suppl. 2) (1988) 85–97.
[23] T.H.E. Meuwissen, B.J. Hayes, M.E. Goddard, Prediction of total genetic values using genome-wide dense marker maps, Genetics 157 (2001) 1819–1829.
[24] R. Bernardo, J. Yu, Prospects for genome-wide selection for quantitative traits in maize, Crop Sci. 47 (2007) 1082–1090.
[25] J.M. Massman, H.-J.G. Jung, R. Bernardo, Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize, Crop Sci. 53 (2013) 58–66.
[26] T. Würschum, J.C. Reif, T. Kraft, G. Janssen, Y. Zha, Genomic selection in sugar beet breeding populations, BMC Genet. 14 (2013) 85, http://dx.doi.org/10.1186/1471-2156-14-85.
[27] D.S. Falconer, T.F.C. Mackay, Introduction to Quantitative Genetics, fourth ed., Longman, New York, 1996.
[28] E.L. Heffner, A.J. Lorenz, J.-L. Jannink, M.E. Sorrells, Plant breeding with genomic selection: gain per unit time and cost, Crop Sci. 50 (2010) 1681–1690.
[29] R.E. Lorenzana, R. Bernardo, Accuracy of genotypic value predictions for marker-based selection in biparental plant populations, Theor. Appl. Genet. 120 (2009) 151–161.
[30] L.R. Schaeffer, Strategy for applying genome-wide selection in dairy cattle, J. Animal Breed. Genet. 123 (2006) 218–223.
[31] C.K. Wong, R. Bernardo, Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations, Theor. Appl. Genet. 116 (2008) 815–824.
[32] S. Zhong, J.C.M. Dekkers, R.L. Fernando, J.-L. Jannink, Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study, Genetics 182 (2009) 355–364.
[33] P.M. Van Raden, C.P. Van Tassell, G.R. Wiggans, T.S. Sonstengard, R.D. Schnabel, J.F. Taylor, F.S. Schenkel, Reliability of genomic predictions for North American Holstein bulls, J.Dairy Sci. 92 (2009) 16–24.
[34] M.P.L. Calus, Genomic breeding value prediction: methods and procedures, Animal 4 (2010) 157–164.
[35] T.L. Luan, J.A. Woolliams, S. Lien, M. Kent, M. Svendsen, T.H.E. Meuwissen, The accuracy of genomic selection in Norwegian Red Cattle assessed by cross-validation, Genetics 183 (2009) 1119–1126.
[36] D. Habier, R.L. Fernando, J.C.M. Dekkers, The impact of genetic relationship information on genome-assisted breeding values, Genetics 177 (2007) 2389–2397.
[37] M.P.L. Calus, T.H.E. Meuwissen, A.P.W. de Roos, R.F. Veerkamp, Accuracy of genomic selection using different methods to define haplotypes, Genetics 178 (2008) 553–561.
[38] J.M. Hickey, G. Gorjanc, Simulated data for genomic selection and genome-wide association studies using a combination of coalescent and gene drop methods, G3 Genes Genomes Genet. 2 (2012) 425–427.
[39] B.J. Hayes, P.J. Bowman, A.J. Chamberlain, M.E. Goddard, Genomic selection in dairy cattle: progress and challenges, J. Dairy Sci. 92 (2009) 433–443.
[40] Z. Abera Desta, R. Ortiz, Genomic selection: genome-wide breeding value prediction in plant improvement, Trends Plant Sci. 19 (2014) 592–601.
[41] G. de los Campos, H. Naya, D. Gianola, J. Crossa, A. Legarra, E. Manfredi, K. Weigel, J.M. Cotes, Predicting quantitative traits with regression models for dense molecular markers and pedigree, Genetics 182 (2009) 375–385.
[42] J. Crossa, G. de los Campos, P. Pérez, D. Gianola, J. Burgueño, J.L. Araus, D. Makumbi, R.P. Singh, S. Dreisigacker, J. Yan, V. Arief, M. Bänziger, H.-J. Braun, Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers, Genetics 186 (2010) 713–724.
[43] J. Crossa, P. Pérez, G. de los Campos, G. Mahuku, S. Dreisigacker, et al., Genomic selection and prediction in plant breeding, J. Crop Improv. 25 (2011) 239–261.
[44] E. Combs, R. Bernardo, Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers, Plant Genome J. 6 (2013) 1–7.
[45] Y. Beyene, K. Semagn, S. Mugo, A. Tarekegne, R. Babu, B. Meisel, P. Sehabiague, D. Makumbi, C. Magorokosho, S. Oikeh, J. Gakunga, M. Vargas, M. Olsen, B.M. Prasanna, M. Bänziger, J. Crossa, Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress, Crop Sci. 55 (2015) 154–163.
[46] N. Hofheinz, D. Borchardt, K. Weissleder, M. Frisch, Genome-based prediction of test cross performance in two subsequent breeding cycles, Theor. Appl. Genet. 125 (2012) 1639–1645.
[47] N. Hofheinz, M. Frisch, Heteroscedastic ridge regression approaches for genome-wide prediction with a focus on computational efficiency and accurate effect estimation, G3 Genes Genomes Genet. 4 (2014) 539–546.
[48] Y. Zhao, M.F. Mette, M. Gowda, C.F. Longin, J.C. Reif, Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat, Heredity 112 (2014) 638–645.
[49] Y. Zhao, M.F. Mette, J.C. Reif, Genomic selection in hybrid breeding, Plant Breed. 134 (2015) 1–10.
[50] C.F.H. Longin, J.C. Reif, Redesigning the exploitation of wheat genetic resources, Trends Plant Sci. 19 (2014) 631–636.
[51] C.F.H. Longin, J.C. Reif, T. Würschum, Long-term perspective of hybrid versus line breeding in wheat based on quantitative genetic theory, Theor. Appl. Genet. 127 (2014) 1635–1641.
[52] C.F. Longin, X. Mi, A.E. Melchinger, J.C. Reif, T. Würschum, Optimum allocation of test resources and comparison of breeding strategies for hybrid wheat, Theor. Appl. Genet. 127 (2014) 2117–2226.
[53] M. Goddard, Genomic selection: prediction of accuracy and maximisation of long term response, Genetica 136 (2009) 245–257.
[54] H.D. Daetwyler, R. Pong-Wong, B. Villanueva, J.A. Woolliams, The impact of genetic architecture on genome-wide evaluation methods, Genetics 185 (2010) 1021–1031.
[55] G. de Los Campos, A.I. Vazquez, R. Fernando, Y.C. Klimentidis, D. Sorensen, Prediction of complex human traits using the genomic best linear unbiased predictor, PLoS Genet. 9 (2013) e1003608.
[56] A.P.W. de Roos, B.J. Hayes, M.E. Goddard, Reliability of genomic predictions across multiple populations, Genetics 183 (2009) 1545–1553.
[57] N. Long, D. Gianola, G.J.M. Rosa, K.A. Weigel, Long-term impacts of genome-enabled selection, J. Appl. Genet. 52 (2011) 467–480.
[58] M. Pszczola, T. Strabel, H.A. Mulder, M.P.L. Calus, Reliability of direct genomic values for animals with different relationships within and to the reference population, J. Dairy Sci. 95 (2012) 389–400.
[59] V. Wimmer, C. Lehermeier, T. Albrecht, H.-J. Auinger, Y. Wang, C.-C. Schön, Genome-wide prediction of traits with different genetic architecture through efficient variable selection, Genetics 195 (2013) 573–587.
[60] J.C. Dawson, J.B. Endelman, N. Heslot, J. Crossa, J. Poland, S. Dreisigacker, et al., The use of unbalanced historical data for genomic selection in an international wheat breeding program, Field Crops Res. 154 (2013) 12–22.
[61] D. Ly, M.T. Hamblin, I. Rabbi, G. Melaku, M. Bakare, H.G. Gauch, et al., Relatedness and genotype-by-environment interaction affect prediction accuracies in genomic selection: a study in cassava, Crop Sci. 53 (2013) 1312–1325.
[62] N. Heslot, H.-P. Yang, M.E. Sorrells, J.-L. Jannink, Genomic selection in plant breeding: a comparison of models, Crop Sci. 52 (2012) 146–160.
[63] J. Poland, J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, et al., Genomic selection in wheat breeding using genotyping-by-sequencing, Plant Genome J. 5 (2012) 103–113.
[64] T.R. Solberg, A.K. Sonesson, J.A. Woolliams, T.H.E. Meuwissen, Genomic selection using different marker types and densities, J. Anim. Sci. 86 (2008) 2447–2454.
[65] J.E. Rutkoski, J. Poland, J.-L. Jannink, M.E. Sorrells, Imputation of unordered markers and the impact on genomic selection accuracy, G3 Genes Genomes Genet. 3 (2013) 427–439.
[66] J. Isidro, J.-L. Jannink, D. Akdemir, J. Poland, N. Heslot, M.E. Sorrells, Training set optimization under population structure in genomic selection, Theor. Appl. Genet. 128 (2015) 145–158.
[67] B.J. Hayes, P.M. Visscher, M.E. Goddard, Increased accuracy of selection by using the realized relationship matrix, Genet. Res. 91 (2009) 47–60.
[68] T.H.E. Meuwissen, Accuracy of breeding values of unrelated individuals predicted by dense SNP genotyping, Genet. Sel. Evol. 41 (2009) 35.
[69] E.L. Heffner, J.-L. Jannink, H. Iwata, E. Souza, M.E. Sorrells, Genomic selection accuracy for grain quality traits in biparental wheat populations, Crop Sci. 51 (2011) 2597–2606.

[70] E.L. Heffner, J.-L. Jannink, M.E. Sorrells, Genomic selection accuracy using multifamily prediction models in a wheat breeding program, Plant Genome 4 (2011) 65–75.
[71] C. Riedelsheimer, J.B. Endelman, M. Stange, M.E. Sorrells, J.-L. Jannink, A.E. Melchinger, Genomic predictability of interconnected biparental maize populations, Genetics 194 (2013) 493–503.
[72] A.J. Lorenz, K.P. Smith, J.-L. Jannink, Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley, Crop Sci. 52 (2012) 1609–1621.
[73] F.G. Asoro, M.A. Newell, W.D. Beavis, M.P. Scott, J.-L. Jannink, Accuracy and training population design for genomic selection on quantitative traits in elite North American oats, Plant Genome 4 (2011) 132–144.
[74] J. Rutkoski, R.P. Singh, J. Huerta-Espino, S. Bhavani, J. Poland, J.-L. Jannink, M.E. Sorrells, Efficient use of historical data for genomic selection: a case study of stem rust resistance in wheat, Plant Genome J. 8 (2015) 1–10.
[75] A.R. Bentley, M. Scutari, N. Gosman, S. Faure, F. Bedford, P. Howell, J. Cockram, G.A. Rose, T. Barber, J. Irigoyen, R. Horsnell, C. Pumfrey, E. Winnie, J. Schacht, K. Beauchêne, S. Praud, A. Greenland, D. Balding, i, J. Mackay, Applying association mapping and genomic selection to the dissection of key traits in elite European wheat, Theor. Appl. Genet. 127 (2014) 2619–2633.
[76] C.A. Davison, D.V. Hinkley, Bootstrap Methods and their Applications, Cambridge Univ. Press, Cambridge, United Kingdom, 1997.
[77] J. Spindel, H. Begum, D. Akdemir, P. Virk, C. Collard, E. Redoña, G. Atlin, J.-L.-Jannink, S.R. McCouch, Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines, PLoS Genet. 11 (2015) e1004982.
[78] R. Rincent, D. Laloë, S. Nicolas, T. Altmann, D. Brunel, P. Revilla, V.M. Rodríguez, J. Moreno-Gonzalez, A. Melchinger, E. Bauer, C.-C. Schoen, N. Meyer, C. Giauffret, C. Bauland, P. Jamin, J. Laborde, H. Monod, P. Flament, A. Charcosset, L. Moreau, Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.), Genetics 192 (2012) 715–728.
[79] J. Burgueño, G. de-los-Campos, K. Weigel, J. Crossa, Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers, Crop Sci. 52 (2012) 707–719.
[80] M. Malosetti, J. Voltas, I. Romagoza, S.E. Ullrich, F.A. van Eeuwjik, Mixed models including environmental covariables for studying QTL by environment interaction, Euphytica 137 (2004) 139–145.
[81] M. Malosetti, M.P. Boers, M.C.A.M. Bink, F.A. van Eeuwjik, Multi-trait QTL analysis based on mixed model with parsimonious covariance matrices, In: Proc. 8th World Congress Genetics Applied to Livestock Production, 13–18 August 2006, Belo Horizonte, Minas Gerais, Brazil, 2015, http://www.wcgalp8.org.br/wcgalp8, Article 25-04.
[82] J.B. Denis, Analyse de regression factorielle, Biom. Praxim. 20 (1980) 1–34.
[83] N. Heslot, D. Akdemir, M.E. Sorrells, J.-L. Jannink, Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions, Theor. Appl. Genet. 127 (2014) 463–480.
[84] D. Jarquín, J. Crossa, X. Lacaze, P. Cheyron, J. Daucourt, J. Lorgeou, F. Piraux, et al., A reaction norm model for genomic selection using high-dimensional genomic and environmental data, Theor. Appl. Genet. 127 (2014) 595–607.
[85] X. Zhang, P. Pérez-Rodríguez, K. Semagn, Y. Beyene, R. Babu, M.A. López-Cruz, F. San Vicente, M. Olsen, E. Buckler, J.-L. Jannink, B.M. Prasanna, J. Crossa, Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs, Heredity 114 (2015) 291–299.
[86] M. Lopez-Cruz, J. Crossa, D. Bonnett, S. Dreisigacker, J. Poland, J.-L. Jannink, R.P. Singh, E. Autrique, G. de los Campos, Increased prediction accuracy in wheat breeding trials using a marker environment interaction genomic selection model, G3 Genes Genomes Genet. (2015), http://dx.doi.org/10.1534/g3.114.016097.
[87] P.B. Cregan, R.H. Busch, Early generation bulk hybrid yield testing of adapted hard red spring wheat crosses, Crop Sci. 17 (1977) 887–891.
[88] J.B. Endelman, G.N. Atlin, Y. Beyene, K. Semagn, X. Zhang, M.E. Sorrells, J.-L. Jannink, Optimal design of preliminary yield trials with genome-wide markers, Crop Sci. 54 (2014) 48–59.
[89] L. Ornella, P. Pérez, E. Tapia, J.M. González-Camacho, J. Burgueño, X. Zhang, S. Singh, F. San Vicente, D. Bonnett, S. Dreisigacker, R. Singh, N. Long, J. Crossa, Genomic-enabled prediction with classification algorithms, Heredity 112 (2014) 616–626.
[90] A. Robertson, A theory of limits in artificial selection, Proc. R. Soc. B 153 (1960) 234–249.
[91] A.R. Hallauer, J.B. Miranda, Quantitative Genetics in Maize Breeding, Iowa State University Press, Ames, Iowa, 1988, pp. 166–167.