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*

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.

© 2015 Z. Published by Elsevier Ireland Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

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

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 $\Delta_G = i r \sigma_A / T$, 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.