PREDICTING GENETIC VALUE OF BREEDING LINES USING GENOMIC SELECTION IN A WINTER WHEAT BREEDING PROGRAM

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Elliot Lee Heffner

January 2011

PREDICTING GENETIC VALUE OF BREEDING LINES USING GENOMIC SELECTION IN A WINTER WHEAT BREEDING PROGRAM

Elliot Lee Heffner, Ph.D.

Cornell University 2011

The use of marker-assisted selection (MAS) to predict genetic value of breeding lines is increasing in private and public plant breeding. MAS is an attractive alternative to phenotypic selection because MAS can be performed on a single plant or seed and decrease selection cycle duration. Advancements in genotyping are rapidly decreasing marker costs so that genotyping is becoming cheaper than phenotyping. Thus, the potential of MAS to achieve greater gains from selection per unit time and cost than phenotypic selection is growing. The ability to achieve genome-wide genotyping, however, may not be best utilized by conventional-MAS methods that have proven to be largely ineffective for improving the complex quantitative traits that dictate the success of new crop varieties.

An emerging alternative to MAS is a technique termed genomic selection (GS) that uses a random-effects statistical modeling approach to jointly estimate all marker effects. This method does not require significance testing and has the goal of capturing small-effect QTL that are excluded by significance thresholds used in conventional-MAS. The use of GS is becoming a popular tool in animal breeding and is garnering the attention of plant breeders; however, evidence regarding the performance and the best methodology for applying GS in plant breeding is currently limited.

In this research, GS was compared to conventional-MAS and phenotypic selection (PS) by deterministic simulation and empirical evaluations in plant breeding.

Performance of these methods was empirically tested in two biparental wheat populations and in an advanced wheat breeding population comprised of multiple families derived from many different crosses. These studies showed that GS was superior to conventional-MAS in predicting the genetic value of breeding lines and that GS was competitive with PS in terms of accuracy. Furthermore, results indicate that GS could significantly reduce the selection cycle duration and achieve prediction accuracies that would enable plant breeders to achieve greater gains per unit time and cost than are possible with current MAS strategies.

BIOGRAPHICAL SKETCH

Elliot Lee Heffner was born in Reading, PA on February 17, 1984 to Martha J. Pool and Gary L. Heffner. He grew up on his parents' dairy farm in Berks County, PA and graduated from Conrad Weiser High School in 2002. He attended The Pennsylvania State University where he received B.A. in Agroecology and was a member of the university's rowing team. Elliot began his Ph.D. in Plant Breeding in Genetics at Cornell University in 2006, and his advisor was Dr. Mark Sorrells. While at Cornell, he co-taught three genetics courses through the Cornell Prison Education Program at the Auburn and Cayuga Correctional Facilities. Elliot will begin his post-graduate career as a Research Scientist for Pioneer-Hi Bred in Des Moines, IA. There, he will conduct a maize breeding program and focus on developing innovative breeding methodologies that integrate new technologies and genomic tools.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to:

My entire family for their endless support and giving me the opportunity to pursue my own directions in academics, travel, and life.

Dr. Richard Bitner for being a mentor and close friend.

Mr. Stephen Miller for inspiring me to pursue an advanced education in agriculture.

Dr. Mark Sorrells and Dr. Jean-Luc Jannink for their mentorship, support, and encouragement throughout my Ph.D. They were both invaluable to my development as a scientist and plant breeder.

Dr. Edward Buckler, Dr. Elizabeth Mannix, and Dr. Ronnie Coffman for serving as members on my Ph.D. committee and for their valuable contributions to my education.

Past and present members of the Sorrells lab for their tireless efforts in the field and for their support and friendship.

The Plant Breeding and Genetics Department and the Cornell community for providing an exciting and positive academic and research environment.

I am also grateful for funding from the USDA National Needs Graduate Fellowship Competitive Grant No. 2005-38420-15785, the Cornell University Provost's Diversity Fellowship, and the Pioneer Hi-Bred Graduate Student Fellowship.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

Chapter 1 was published as a review and interpretation in *Crop Science:*

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

Chapter 2 was published as an original research article in *Crop Science*:

Heffner EL, Lorenz AJ, Jannink JL, Sorrells ME (2010) Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. Crop Sci 50:1681-1690

Chapter 3 was submitted as an original research article to *Theor. and App. Genetics*:

Heffner EL, Iwata H, Souza EJ, Jannink JL, Sorrells ME (submitted) Genomic selection across environments for grain quality traits in biparental wheat populations.

Chapter 4 will be submitted as an original research article to *(undecided)*:

Heffner EL, Jannink JL, Sorrells ME (in prep) Genomic selection accuracy using multiple-family data in a winter wheat breeding program.

CHAPTER ONE

GENOMIC SELECTION FOR CROP IMPROVEMENT

ABSTRACT

Despite important strides in marker technologies, the use of marker-assisted selection has stagnated for the improvement of quantitative traits. Bi-parental mating designs for the detection of loci affecting these traits (QTL) impede their application, and the statistical methods used are ill-suited to the traits' polygenic nature. Genomic selection (GS) has been proposed to address these deficiencies. Genomic selection predicts the breeding values of lines in a population by analyzing their phenotypes and high-density marker scores. A key to the success of GS is that it incorporates all marker information in the prediction model, thereby avoiding biased marker effect estimates and capturing more of the variation due to small effect QTL. In simulations, the correlation between true breeding value and the genomic estimated breeding value has reached levels of 0.85 even for polygenic low heritability traits. This level of accuracy is sufficient to consider selecting for agronomic performance using marker information alone. Such selection would substantially accelerate the breeding cycle, enhancing gains per unit time. It would dramatically change the role of phenotyping, which would then serve to update prediction models and no longer to select lines. While research to date shows the exceptional promise of GS, work remains to be done to validate it empirically and to incorporate it into breeding schemes.

Introduction

The use of marker-assisted selection (MAS) in plant breeding has continued to increase in the public and private sector. Most applications, however, have been constrained to simple, monogenic traits (reviewed by Xu and Crouch 2008). While MAS has had significant impacts in backcrossing of major genes into elite varieties (Holland 2004), backcrossing is regarded as the most conservative of breeding methods because improvement occurs through the pyramiding of only a few target genes (Lee 1995). Gene pyramiding is inefficient for quantitative traits that are often controlled by many small-effect quantitative trait loci (QTL; Kearsy and Farquhar 1998).

Current MAS methods are better suited for manipulating a few major effect genes than many small effect genes (Dekkers and Hospital 2002). Unfortunately, these small effect genes underly the complex polygenic traits that are crucial for the success of new crop varieties (Crosbie et al. 2003). Two primary limitations to MAS are 1) the biparental mapping populations used in most QTL studies do not readily translate to breeding applications and 2) statistical methods used to identify target loci and implement MAS have been inadequate for improving polygenic traits controlled by many loci of small effect. The application of *Genomic Selection* (*GS*), proposed by Meuwissen et al. (2001), to breeding populations using high marker densities is emerging as a solution to both of these deficiencies. We review here current GS methods and their performance. Furthermore, we present future directions for GS research and some exciting opportunities GS provides that could revolutionize crop improvement.

Current MAS Limitations

The most common method of QTL detection is the use of a biparental mapping population. While these studies are important to the understanding of genetic architecture, building mapping populations distinct from breeding populations often strains the resources of a breeding program. Available resources limit the size of mapping populations and consequently, the accuracy of QTL position and effect estimates (Dekkers and Hospital 2002; Schön et al. 2004). Also, allelic diversity and genetic background effects that are present in a breeding program will not be captured with a single biparental population. Therefore, multiple mapping populations are needed, QTL positions require validation, and QTL effects must be re-estimated by breeders in their specific germplasm. The validation in locally adapted germplasm is important because poor estimates of the numerous small effect QTL will lead to gains from MAS that are inferior to traditional phenotypic selection (Bernardo 2001). Therefore, the resources required for QTL detection coupled with validation and effect re-estimation limit the effectiveness of biparental population derived QTL for MAS in plant breeding populations (reviewed by Holland 2004).

To avoid this disconnect between biparental and breeding populations, linkage disequilibrium (LD) based mapping can be used for dissection of complex traits in breeding populations that already have extensive phenotypic data across locations and years (Jannink et al. 2001; Rafalski 2002). This strategy avoids the need to develop special mapping populations that impose an additional burden on breeding programs. Also, mapping within breeding populations will allow for QTL identification and allelic value estimates that can be directly utilized by MAS without the need for extensive validation (Breseghello and Sorrells 2006; Holland 2004). However, low

heritability, small population sizes, few large-effect QTL, confounding population structure, and arbitrary significance thresholds found in current association mapping efforts allow identification of only a few QTL with overestimated effects (Beavis 1998; Schön et al. 2004; Xu 2003a).

To minimize the limitations for successful MAS, Lande and Thompson (1990) proposed a visionary two-step approach: 1) select significant markers from large marker sets and 2) combine phenotypic information with significant markers in a selection index that would explain a significant proportion of additive genetic variance. In the first step, they were unable to estimate all marker effects simultaneously with simple regression due to the lack of degrees of freedom. Therefore, they proposed selecting the most significant markers from the previous generation via multiple linear regressions and then re-estimating effects of the selected markers in the current generation with independent multiple regressions (Lande and Thompson 1990).

Lande and Thompson (1990) introduced this two-step approach to handle large marker sets because they estimated that hundreds of molecular markers would be needed to capture a significant proportion of the additive genetic variance. In the early 1990's, genome-wide marker coverage was a limiting factor for MAS, but in recent years, plant breeders have encountered a major shift in the amount of genomic information that is available due to the rapid advances in marker technologies. Although genotyping is still a major expense, the declining costs per marker data point have facilitated large scale genotyping efforts in breeding programs. For example, the Monsanto Company reports that from 2000 to 2006, they experienced a six-fold decrease in cost per marker data point and have increased the volume of their marker

data by forty-fold (Eathington et al. 2007). The availability of abundant markers and the reduction of genotyping costs will present new tools for plant breeders only if statistical methodologies for the utilization of genomewide marker coverage are developed.

The two-step process by Lande and Thompson (1990) has been criticized as an inefficient use of available data (Meuwissen et al. 2001): one would rather want to use all available data in a single step to get maximally accurate estimates of marker effects. Genomic selection (GS) is a form of MAS that *simultaneously estimates all* locus, haplotype, or marker effects across the *entire* genome to calculate genomic estimated breeding values (GEBVs; Meuwissen et al. 2001). This approach contrasts greatly with traditional MAS because there is not a defined subset of significant markers used for selection. Instead, GS analyzes jointly *all markers* on a population attempting to explain the *total genetic variance* with dense genomewide marker coverage through summing marker effects to predict breeding value of individuals (Meuwissen et al. 2001).

The central process of GS is the calculation GEBVs for individuals having only genotypic data using a model that was "trained" from individuals having both phenotypic and genotypic data (Fig. 1.1; Meuwissen et al. 2001). The population of individuals with both phenotypic and genotypic data is known as the "training population" as it is used to estimate model parameters that will subsequently be used to calculate GEBVs of selection candidates (e.g. breeding lines) having only genotypic data (Fig. 1.1). These GEBVs are then used to select the individuals for advancement in the breeding cycle. Therefore, selection of an individual *without phenotypic* data can be performed by using a model to predict the individual's breeding value

(Meuwissen et al. 2001). To maximize GEBV accuracy, the training population must be representative of selection candidates in the breeding program to which GS will be applied.

Historically, estimated breeding values (EBVs) for quantitative traits have been calculated by best linear unbiased prediction (BLUP) based only on phenotypic data of individuals and their relatives (Henderson 1984). The use of EBVs via BLUP has been popular in animal breeding and in recent years has been utilized by plant breeders (reviewed by Piepho et al. 2007). However, data on markers linked to known QTL can also be used for calculation of EBVs (Fernando and Grossman 1989) and this method was predicted to increase gains from selection in animal breeding up to 38% (Meuwissen and Goddard 1996). These results were encouraging but they require extensive prior QTL discovery efforts in non-breeding populations.

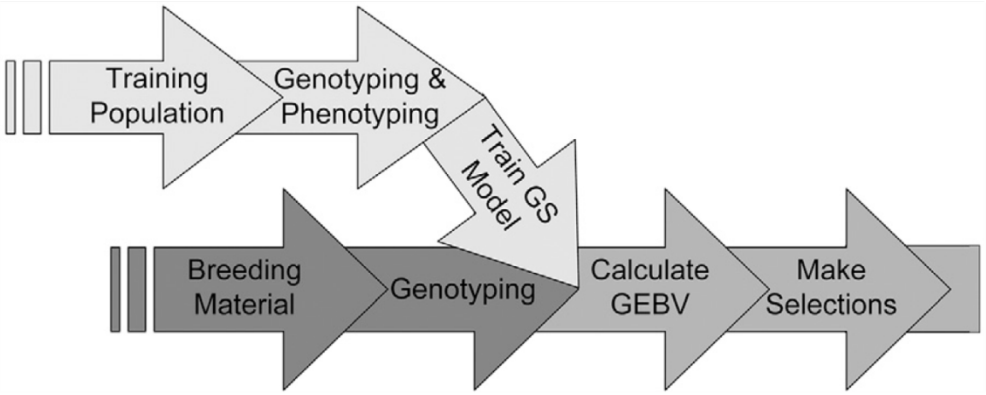


Figure 1.1 Diagram of genomic selection (GS) processes starting from the training population and selection candidates continuing through to GEBV-based selection. Note that while we show here a single occurrence of model training, training can be performed iteratively as new phenotype and marker data accumulate.

Marker Density and Linkage Disequilibrium

Genomic selection differs from current MAS strategies because instead of only using markers that have a predefined significant correlation with a trait, all markers are used to estimate breeding values for each genotype. Consequently, dense marker coverage is needed to maximize the number of QTL in LD with at least one marker thereby also maximizing the number of QTL whose effects will be captured by markers. Target marker density will be dictated by the rate of LD decay across the genome, as assessed by the relationship between inter-marker coefficient of determination, r^2, and genetic distance.

Rate and pattern of LD decay are affected by population characteristics such as evolutionary history, mating system, population size, admixture, recombination rate, and selection effects (Gaut and Long 2003). Therefore, LD decay rates are highly variable among species, populations, and genomic regions. Examples of this variability in LD decay rates include: 75-500 kb in a diversity panel of rice (*Oryza sativa*; Mather et al. 2007), 10-20 cM (roughly 50-100 Mb) in elite cultivars of wheat (*Triticum aestivum;* Chao 2007; Maccaferri et al. 2005), 0.1 to 1.5 kb in diverse inbred lines of maize (*Zea mays* ssp. *mays;* Remington et al. 2001; Tenaillon et al. 2001), and 15-20 kb in a diversity panel of sorghum (*Sorghum bicolor*; Hamblin et al. 2005). Examples from diversity panels may give rough predictions of LD decay in a species, but because many factors affect LD, individual breeding programs will need to determine LD decay rates on a case by case basis in their specific breeding populations.

Linkage disequilibrium estimates can be used to determine target marker densities for GS. For example, Calus and Veerkamp (2007) used the average r^2 between adjacent markers as a measure of their marker density relative to the decay of LD. They found that for a high heritability trait, average adjacent marker r^2 of 0.15 was sufficient, but for a low heritability trait, increasing the r^2 to 0.2 improved the accuracy of GEBV predictions. These marker densities may still be out of reach for some crops or populations. Looking to the near future, however, high throughput sequencing has made marker discovery affordable for most crop species and the continued reduction of genotyping costs will facilitate dense genomewide marker coverage for all crop species (reviewed in Zhu et al. 2008). Note that the conditions of complete genome saturation and of at least one marker in LD with each QTL need not be met in order to derive useful prediction models for GEBV. While it is tempting to surmise a minimum number of markers needed to obtain useful GEBVs, the many factors affecting this number and the lack of empirical results currently available would make any guess meaningless. Clearly, this subject requires urgent attention.

Statistical Models and Performance

The challenge of QTL analysis is the selection of the appropriate statistical model to identify QTL and estimate their effects (Broman and Speed 2002). In breeding programs, statistical methods for GS will need to simultaneously estimate many marker effects from a limited number of phenotypes. A greater number of explanatory variables (markers) than observations (phenotyped lines) leads to a lack of degrees of freedom that must be handled through the selection and use of the most appropriate statistical model, i.e. the model that results in the highest GEBV accuracy with consideration of model complexity and computation requirements. In the

assessment of model performance, GEBV accuracy has a precise definition, namely, the Pearson correlation between the GEBV and the true breeding value (TBV). Accuracy defined in this way is directly proportional to gain from selection when selecting on the GEBV, that is, $R=ir\sigma_A$, where R is the response, i is the selection intensity, r is the accuracy defined above, and σ_A is the square-root of the additive genetic variance of TBV (Falconer and Mackay 1996, p. 189). We briefly describe here three models: stepwise regression, ridge regression, and Bayesian estimation.

Stepwise Regression for MAS

Traditional MAS considers marker effects as fixed requiring stepwise regression (SR) approaches that avoid the lack of degrees of freedom problem by fitting markers singly or in small groups. After the model selection process during which markers are added or removed from the model on the basis of arbitrary significance thresholds, non-significant markers are assigned an effect of zero and significant marker effects are simultaneously tested to estimate their effects. This stepwise approach to set non-significant marker effects to zero is critical for maintaining model estimability (Lande and Thompson 1990). Significance thresholds that may maximize response to selection cannot be determined analytically, though guidelines have been established through simulation (Hospital et al. 1997; Moreau et al. 1998). The general guideline is that liberal p-value thresholds improve selection gain (Hospital et al. 1997; Moreau et al. 1998). Nevertheless, when only significant marker effects are estimated, only a portion of the genetic variance will be captured (Goddard and Hayes 2007) and effects retained in the model can be greatly overestimated (Beavis 1998; Hayes 2007), particularly when many effects are tested.

Limitations of SR for MAS in practice were reported by (Moreau et al. 2004). In 300 test-crossed maize progenies evaluated in 14 trials over 11 locations for dry grain yield and grain moisture, they discovered 16 QTL for dry grain yield and 12 QTL for grain moisture explaining 50% of the total phenotypic variance of both traits. When using an index combining phenotypic and marker information for a single cycle followed by two cycles of marker-only selection, they observed no genetic gain from the two cycles of marker selection (Moreau et al. 2004). They suggested that this inefficiency of MAS could be caused by fixation of major effect loci in the first cycle of selection and inaccurate estimation of remaining effects resulting in no gain from the cycles of marker selection (Moreau et al. 2004). These complications were probably consequences of SR that detects only large effects and that overestimates effects.

In a GS simulation by Meuwissen et al. (2001), SR resulted in low GEBV accuracy due to limited detection of QTL. The simulated outcrossing population had an effective population size of 100 with a trait heritability of 0.5. After 1,000 generations of random mating to establish mutation-drift equilibrium, generation 1001 had a population size of 200 (100 males; 100 females). Two generations (1002 and 1003) of size 2000 with 20 half-sib families of size 100 individuals were then simulated. Generations 1001 and 1002 were used to train the model while GEBV accuracy was calculated on generation 1003. Genotypic data consisted of 101 multi-allelic markers on each of 10 chromosomes of length 100 cM. Adjacent pairs of markers were considered haplotypes such that 50,000 haplotype effects were estimated. The accuracy of GEBV for SR (0.318) was less than that expected for strictly phenotype-based BLUP (about 0.4; Meuwissen et al. 2001). In agreement with Lange and Whittaker (2001), Meuwissen et al. (2001) concluded that SR's procedure

to identify marker subsets is suboptimal for MAS in situations where the majority of the additive genetic variance is generated by many QTL. Note, however, that the GEBV accuracy of SR depends on the details of the analysis: using the Meuwissen et al. (2001) simulation design, Habier et al. (2007) found that SR produced an accuracy of 0.61. Habier et al. (2007) attributed this difference to the use of a less stringent significance threshold than was used by Meuwissen et al. (2001). This conclusion was supported by simulations showing prediction accuracy changed with changes in significance thresholds (Piyasatian et al. 2007).

Ridge Regression-BLUP for GS

The ridge regression-BLUP (*RR*-BLUP) method can simultaneously estimate all marker effects for GS (Meuwissen et al. 2001; Whittaker et al. 2000). Rather than categorizing markers as either significant or as having no effect, ridge regression shrinks all marker effects towards zero (Breiman 1995; Whittaker et al. 2000). The method makes the assumption that markers are random effects with a common variance (Meuwissen et al. 2001; Table 1.1). Equal variance does not assume all markers have the same effect (Bernardo and Yu 2007), but that marker effects are all equally shrunken toward zero. Nevertheless, the assumption that individual markers have the same variance is unrealistic and therefore *RR*-BLUP incorrectly treats all effects equally (Xu 2003b). Despite the incorrect assumption of equal marker variance, *RR*-BLUP is superior to SR because it is able to simultaneously estimate effects for all markers: by avoiding marker selection, it avoids the biases that go with that selection (Whittaker et al. 2000). Also, a ridge regression approach is more appropriate than SR for instances where there are few or no large effects and many small effects (Breiman 1995), as is the case with most quantitative traits.

In the simulation by Meuwissen et al. (2001), *RR*-BLUP had a GEBV accuracy of 0.732, which was 41% and 33% greater than SR and phenotype-based BLUP, respectively. With higher SR significance thresholds, Habier et al. (2007) reported *RR*-BLUP resulted in 4% and 11% increase in GEBV accuracy compared to SR and traditional BLUP, respectively. In addition to these studies, Muir (2007) simulated 512 genotypes with a low heritability trait (h^2=0.1) in each of 4 training generations. These conditions resulted in an even higher *RR*-BLUP GEBV accuracy of 0.83 despite the lower heritability. This gain in GEBV accuracy was attributed to the four training generations used by Muir (2007), as opposed to two generations used in previous studies (Habier et al. 2007; Meuwissen et al. 2001).

In a GS simulation on a population derived from a biparental cross of maize inbreds, Bernardo and Yu (2007) found that relative to phenotypic selection, the increase in selection gain from *RR*-BLUP was 18% greater than that from SR for a highly heritable trait (h^2=0.8) controlled by twenty QTL. For a trait with low heritability (h^2=0.2) controlled by 100 QTL, the increase in selection gain from *RR*-BLUP was 43% greater than that from SR (Bernardo and Yu 2007). Similar results were observed by Piyasatian et al. (2007) who found that in the first round of selection in a simulated cross between two inbred parents, gain from selection from *RR*-BLUP was 109% and 32% greater than that of traditional BLUP and SR, respectively.

Bayesian Estimation

The simplifying assumption of equal and fixed marker effect variances allows *RR*-BLUP parameters to be efficiently computed using maximum likelihood methods (Meuwissen et al. 2001). While *RR*-BLUP can provide a conservative EBV by

shrinking all marker effects equally (Muir 2007), the presumably incorrect assumption that underlies it can lead to over-shrinking of large effects (Table 1.1; Meuwissen et al. 2001; Xu 2003b). Bayesian methods have been adopted in order to relax this assumption and better model marker effects of differing sizes (Hayes 2007). Here, a separate variance is estimated for each marker, and the variances are assumed to follow a specified prior distribution (Meuwissen et al. 2001).

Meuwissen et al. (2001) proposed two types of prior distribution for the marker variance. The first type of prior (BayesA) uses an inverted chi-square distribution with degrees of freedom and scale parameters chosen so that the mean and variance of the distribution match the expected mean and variance of the marker variances. In the simulation design described above, BayesA outperformed both SR and *RR*-BLUP with a GEBV accuracy of 0.798. Different parameter values for the BayesA inverted chi-square prior distribution have also been proposed that place much higher density on marker variances close to zero, thereby forcing more marker effect estimates close to zero (ter Braak et al. 2005; Xu 2003b).

The BayesA method of Xu (2003b) was applied to data from a doubled haploid barley (*Hordeum vulgare*) population of 145 lines with 127 SNP markers covering 1500 cM for yield, heading date, maturity, test weight, lodging, and kernel weight. Xu (2003b) reported that SR and BayesA both found large effect QTL, but that BayesA provided better QTL location and effect estimation. Also, in simulation of a population derived from a biparental inbred cross, ter Braak et al. (2005) found that BayesA prior parameters forcing more marker effect shrinkage gave better estimates of QTL effects than did the Meuwissen et al. (2001) parameters. A comparison of

these different prior parameterizations in an association genetics rather than linkage mapping context has not been done.

The second type of prior distribution Meuwissen et al. (2001) proposed (BayesB) contrasts with BayesA by having a prior mass at zero, thereby allowing for markers with no effects. The inverted chi-square prior of BayesA may be set to strongly regress variances towards zero, but it does not permit the value of zero itself. BayesB thus presents a more realistic prior because we expect that some regions of the genome will carry no QTL so that some markers should have estimates of zero effect. The results from Meuwissen et al. (2001) showed that BayesB had a GEBV accuracy of 0.848, greater than all other methods tested. Of the Bayesian methods, BayesB was not only more accurate, but was also less computationally demanding. Meuwissen et al. (2001) concluded that Bayesian methods outperformed *RR*-BLUP through better estimation of large effect QTL by allowing for unequal variances.

de Roos et al. (2007) used Bayesian modeling as described by Meuwissen and Goddard (2004) in actual dairy cattle data for a single chromosome containing 32 markers with one being a known causal mutation for fat percentage. They compared Bayesian GS that used all marker information to regression on the genotype at the known causal mutation and to traditional BLUP with no markers. Using a cross validation population of 1,135, they concluded that Bayesian GS and regression on the causal mutation had similar accuracies (0.752 and 0.746, respectively), with both being superior to traditional BLUP (EBV accuracy of 0.508). Interestingly, the GS analysis often did not place the causal mutation in the correct marker bracket but was nevertheless able to calculate accurate GEBV. This robustness of GEBV accuracy

provides evidence that GS can perform well for breeders in the absence of the discovery of QTL (de Roos et al. 2007).

In the future, genotyping costs will decrease, but it is unlikely that phenotyping costs will also decrease thus shifting goals towards reducing phenotyping and increasing genotyping. Bernardo and Yu (2007) suggested this shift would be feasible when the cost of a marker data point is 5,000 times less than the cost of phenotyping a single entry. Regardless of the threshold, it is desirable to decrease the number of phenotypic records needed for training models for accurate GEBVs. Simulations by Meuwissen et al. (2001) showed that with 2200 phenotypic records, *RR*-BLUP and BayesB had GEBV accuracies of 0.732 and 0.848, respectively. When the number of phenotypic records was reduced to 500, *RR*-BLUP and BayesB GEBV accuracies decreased to 0.579 and 0.708, respectively (Meuwissen et al. 2001). Thus the effect of low numbers of phenotypic records was less severe for BayesB than for *RR*-BLUP. In addition, Fernando (2007) found that in contrast to *RR*-BLUP, BayesB's GEBV accuracy did not decline as the number of markers increased. These findings suggest that Bayesian methods may be better suited to handling situations with increased colinearity between markers caused by extremely large markers sets and limited phenotypic records (Table 1.1). Computational issues may arise for Bayesian methods under high marker densities and collinearities, and these will need to be resolved by improved statistical methods (ter Braak et al. 2005).

Table 1.1 General characteristics and trends of performance for traditional BLUP and GS methods. It is important to note these are general summaries based on current understanding of model performance.

Method	Marker effect; variance assumptions	Proportion of markers fitted in model	Performance with increased		Large Effect QTL	Small Effect QTL	Inbreeding depression; loss of diversity
			Marker density	QTL number			
Traditional BLUP	N/A	N/A	N/A	N/A	Captured only by phenotype	Captured only by phenotype	Yes
Stepwise Regression	Fixed	Subset	Reduced	Reduced	Over-estimated	Excluded	Marginally Reduced
RR-BLUP	Random; Equal	All	Reduced †	Increased	Under-estimated	Captured	Reduced
BayesA	Random; Unique All > 0	All	?	Reduced	More accurately estimated	Captured	Reduced
BayesB	Random; Unique Some=0	All	Insensitive †	Reduced	More accurately estimated	Captured	Reduced

† Source: Fernando (2007)

Inclusion of a Polygenic Effect Term Accounting for Kinship

Phenotypic information from relatives contribute to an individual's EBV because EBVs vary according to the additive relationship (*A*) matrix, i.e., a matrix that contains, for each pair of individuals, the proportion of alleles for which they are identical by descent (van Arendonk et al. 1994; Lynch and Walsh 1998, p. 751). When markers are introduced into the analysis, some genetic effects will be captured by markers in LD with QTL, but residual genetic effects will still be assumed to vary according to the *A*-matrix. These residual effects can be captured by including a

polygenic term in the model. In association mapping, the inclusion of this matrix has been popularized as a statistical control for population structure and familial relatedness (Yu et al. 2006; Zhao et al. 2007).

The *A*-matrix can be calculated on the basis of the pedigree or the marker data, with pedigree information providing exact expected relationships and markers providing estimated realized relationships. When marker number is high enough that marker sampling plays a minor role (i.e., relationship estimates on the basis of markers are accurate), marker-estimated relationships will better reflect true relationships than will pedigree-expected relationships. In particular, four mechanisms lead realized relationships to diverge from their expectation: random Mendelian segregation, segregation distortion, selection, and pedigree recording errors. For example, parental contributions to inbreds vary from their expected 50% because of random Mendelian segregation during selfing. For the genomes of maize and wheat, there is a 10% probability that single seed decent derived inbreds will have less than 38% and 43% genome contribution from one parent, respectively (Frisch and Melchinger 2007).

The value of including a polygenic effect term in the model will depend strongly on marker density available in the study for two reasons. First, if density is such that all QTL are in strong LD with a marker, all genetic effects will be absorbed by markers and none will be left for the polygenic term to capture (Bernardo and Yu 2007; Meuwissen et al. 2001; Zhong and Jannink 2007). Second, even markers that are in linkage *equilibrium* with all QTL carry information about relationships among individuals, and this information contributes to the accuracy of GEBV (Habier et al. 2007). Indeed, this contribution depends on the number of markers included in the GS

method and, because SR uses only a subset of markers, it benefits least from genetic relationship contributions to GEBV accuracy (Habier et al. 2007).

Research to look explicitly at the value of including a polygenic effect term used adjacent-marker r^2 as a measure of marker density. For a high heritability trait (h^2=0.5), the polygenic effect term increased GEBV up to an adjacent-marker r^2 of 0.14, while for a low heritability trait (h^2=0.1), the term made no difference already at an r^2 of 0.11 (Calus and Veerkamp 2007). At lower adjacent-marker r^2 the polygenic term fulfills its role of explaining genetic variance not absorbed by markers and it therefore contributes to GEBV accuracy (Calus and Veerkamp 2007; Villanueva et al. 2005).

Selection Index Theory Applied to Genomic Selection

A selection index integrates and weights multiple traits to achieve greater gains than if traits with independent thresholds are individually or collectively selected (Hazel and Lush 1942; Hazel 1943). Selection indices can incorporate marker data as indirect selection traits (Lande and Thompson 1990; Neimann-Sorensen and Robertson 1961; Smith 1967). However, current MAS applied to loci selected by SR violates the selection index assumptions of multivariate normality and small changes in allele frequencies because selection is based on only few large effect loci (Dekkers 2007; Lande and Thompson 1990). Because GS is based on many markers distributed throughout the genome, index selection assumptions are met providing an opportunity to use index selection theory to predict response to GS (Dekkers 2007).

Dekkers (2007) used selection index theory by adding marker derived breeding values as a separate correlated trait to the selection index (Lande and Thompson 1990). In a simulated swine breeding program, selection on only marker data could outperform phenotypic selection for low heritability traits (0.1) even with moderate GEBV accuracy (0.55). When marker and phenotypic data were both used for a single trait, even greater accuracies were observed. This increase was due to marker information that allowed for within family selection (Dekkers 2007). For two negatively correlated traits with heritabilities of 0.3 and 0.1, Dekkers (2007) found using only markers increased gains from selection over phenotypic selection by 8.5% for the index of the two traits and 66% for the low heritability trait alone. Using both markers and phenotype increased gains from selection over phenotypic selection by 21% for the index of the two traits and 80.5% for the low heritability trait alone. These results show the potential of GS to increase gains for multiple traits especially in cases where phenotypic data is available on selection candidates and traits have low heritability.

Maintaining Genetic Diversity and Reducing Inbreeding Depression

Gains from selection can be increased by raising the selection intensity or the accuracy of EBV of breeding lines. Increased selection intensity reduces the number of lines selected thus lowering the effective population size thereby increasing the loss of genetic variability. Traditional BLUP increases EBV accuracy by incorporating ancestor and collateral relative phenotypes in the calculation (Henderson 1984). But including family information in EBV calculation increases the correlation between EBV of family members, making it more likely that multiple sibs will be selected (Wray and Thompson 1990). Sibling co-selection, in turn also reduces effective

population size. Therefore, while increased selection intensity and a higher EBV accuracy lead to greater short term gains from selection, they both may reduce long term gains by decreasing genetic variation and increasing rates of inbreeding (Quinton et al. 1992).

Daetwyler et al. (2007) reviewed these issues and determined that GS differs from simple phenotypic selection and traditional BLUP by using markers to more accurately estimate Mendelian sampling variation, i.e., deviations between siblings within families. Mendelian sampling variation, generated by random segregation, is created anew each generation. Selecting strictly on this variation therefore enables sustained genetic progress by decreasing co-selection of sibs and thus reducing inbreeding and the loss of genetic variation (Woolliams et al. 1999). Optimized selection schemes have been proposed where parent combinations are restricted by their level of coancestry to limit the loss of genetic variation and the rate of inbreeding (Grundy et al. 1998; Meuwissen 1997). In these schemes an individual's selective advantage depends largely on the Mendelian sampling term, i.e. on its performance relative to its siblings (Avendaño et al. 2004). Unlike traditional BLUP based on pedigree data that account for average relationships, tracking markers enables GS to also track the random segregation that makes up the Mendelian sampling term. The benefit is both more accurate EBVs and decreased correlation between EBVs within families, countering the mechanism whereby the use of family information increases loss of genetic diversity (Daetwyler et al. 2007). Note that the greater emphasis placed by GS on the Mendelian sampling term does not completely negate variable long-term genetic contributions among individuals and its consequent increase in inbreeding rate. In particular, superior individuals carry superior alleles and selection of those alleles will, in turn, lead their carriers to leave more offspring behind (Daetwyler et al. 2007).

Thus, it still may be advisable to manage rates of inbreeding (e.g., Avendaño et al. 2004) even in the context of GS. Nevertheless, the advantages of GS in regard to inbreeding and the maintenance of genetic diversity should prove valuable for crops such as alfalfa (*Medicago sativa*) that suffer from inbreeding depression and for maintaining genetic variation in advanced cycle breeding programs.

Gains from Selection Per Unit Time

MAS strategies increase gain mainly through gain per unit time, rather than gain per cycle (Bernardo and Yu 2007; Edwards and Johnson 1994; Hospital et al. 1997; Koebner and Summers 2003; Meuwissen et al. 2001; Muir 2007). To look at GS's impact on gains per unit time, Schaeffer (2006) suggested a plan for implementation of GS into a dairy breeding program. Through reduction in time and costs needed to prove the value of a bull, assuming a GEBV accuracy of 0.75, Schaeffer (2006) determined that GS could provide a twofold increase in rate of genetic gain and save 92% of the costs of the current progeny test based breeding program.

In plants, the importance of generation time varies between crops, but the goal of reducing cycle time remains. In maize, a crop that uses doubled haploids and off-season nurseries, test cross performance selection still requires at least two years (Bernardo and Yu 2007) providing an opportunity for GS to reduce unit time per selection cycle by reducing the need for progeny test data in every cycle. In the more extreme case of oil palm, which takes 19 years to complete a cycle of selection, Wong and Bernardo (2008) reported that GS reduced the selection cycle to 6 years. Even with small population sizes (N=50) that adversely effected GEBV accuracy, their

simulations indicated that GS would outperform MARS and phenotypic selection when considering gain per unit cost and time.

Genotype by Environment Interactions and Epistasis

Genotype by environment (G × E) interaction is a challenge in plant breeding because the large number of experimental lines and environments i.e. locations and years make it impossible to test a line in all possible environmental conditions of a breeding program's target region (Allard and Bradshaw 1964). Consider, however, that the genotype of any line is composed of alleles that, over time, will have been evaluated in a larger sample of target environments than would be feasible for any particular line. Thus, it may be possible to accurately predict GEBV even in the presence of high G × E. As an extreme example, for winter annual crops, a severe winter may only come around once a decade. Variety releases for the region need to be hardy to such winters because crop failure even once per decade is too frequent. With GS, a given generation of experimental lines need never experience a test winter if the alleles they carry were characterized during a severe winter. Similar cases include the infrequent but devastating conditions caused by severe drought, flooding, disease pressure, and insect infestation. The broader insight that these examples illustrate is that when using GS, lines are not evaluated solely on the basis of their own phenotypic performance, but on the basis of information shared across other lines, other years and locations, and even possibly other breeding programs. This information sharing should provide GS with stability in the face of G × E.

Anticipating the effect of epistasis on the potential of GS is difficult. Almost all GS prediction accuracy evaluations derive from simulations that adopted additive genetic models. There is current debate, at both theoretical and empirical levels, of the likely importance of epistasis in the architecture of quantitative traits (Carlborg and Haley 2004; Hill et al. 2008; Holland 2007; Mackay 2009). To discuss this issue, it is essential to distinguish between the genotypic value versus the breeding value of a line (Falconer and Mackay 1996). The genotypic value is the expected phenotype of the line given its genotype and includes additive and non-additive genetic effects. The breeding value is the expected phenotype of line's progeny and includes only additive effects. The additive models used by GS should predict the breeding value rather than genotypic value (Goddard and Hayes 2007). Consequently, correlations between GEBVs and line phenotypes may well be lower than those obtained in additive effect simulations, but they should nevertheless reflect a line's value as a parent. For cases where estimates of genotypic value are desired in the presence of epistasis, methods are currently being developed and tested (e.g., Gianola et al. 2006; Gianola and van Kaam 2008; Gonzalez-Recio et al. 2008). Further empirical evaluation of the prediction accuracies of these methods should help address the ongoing debate over the importance of epistasis in the mapping of genotype to phenotype. Because of the small contribution that epistasis makes to breeding value (Holland 2001), genomic selection using simpler additive models should be effective for maximizing gain from selection.

Future Directions

Statistical Methods

A statistical model will more faithfully capture QTL information as its assumptions about the underlying genetic architecture, made explicit in the prior distributions of QTL effects or variance, are more correct (Meuwissen et al. 2001). There are two obstacles to translating this fact into improved models. First, GS may gain in accuracy not just by capturing more QTL information but also by better capturing relationship information (Habier et al. 2007). There may be a tradeoff between the kinds of prior distributions of effects that promote the use of these two information sources (Habier et al. 2007). Second, we simply do not know, for any complex trait, what the underlying genetic architecture is, and therefore we do not have adequate prior knowledge at our disposal. Therefore, statistical models that are relatively insensitive to the underlying architecture may be optimal for most populations; although, identifying those models remains challenging.

Finally, the marker technologies upon which GS methods depend are constantly changing. Next generation sequencing technologies and improvement of genotyping platforms present breeders with powerful tools for characterizing the genetic composition of their germplasm. As these technologies continue to evolve, they will provide quantitatively and qualitatively different information (e.g., copy number and epigenetic variation; Stranger et al. 2007; Zhang et al. 2008), and statistical machinery will also need to evolve to use this information efficiently to increase prediction accuracy.

Software and Database Development

While statistical methods of prediction must be continually advanced, an integral part of their performance will be the software packages used to implement them. In conjunction with this software, robust databases that can efficiently link breeding lines, testing environments, genotypic data, phenotypic data, and breeding programs will need to be developed to simplify flow and use of information. While private breeding companies have invested heavily in data management systems that will likely be efficient in executing GS (e.g., Eathington et al. 2007), public sector breeding programs also need database software that integrates the wide variety of data they generate (Heckenberger et al. 2008; Tinker and Yan 2006). Recent developments in the public sector are promising, e.g., the barley coordinated agricultural project hordeum toolbox (http://hordeumtoolbox.org/); the GDPDM database schema that links with the association analysis software TASSEL (http://www.maizegenetics.net); the German GABI-BRAIN project (http://brain.uni-hohenheim.de/eng/indexeng.html), and the Canadian COOL-DUDE (Yan and Tinker 2007). Adaptation of these tools to link with GS and development of user-friendly GS analyses themselves are needed to take GS from theory to practice.

Changes to Breeding Program Structure

The accuracies of GEBV observed in research offer the possibility that future elite and parental lines will be selected on their GEBV rather than on their phenotypic records from extensive field testing. The most immediate impact of this circumstance would be a great increase in the speed of the breeding cycle (Fig. 1.2; Wong and Bernardo 2008), thereby increasing selection gains per unit time. This shift would also

fundamentally alter the role of phenotyping in plant breeding (Fig. 1.2). Note that Figure 1.2 offers a somewhat futuristic view of the use of GS, contingent on its validation in practice. We do not, at this point, advocate dispensing with phenotypic evaluation prior to parent selection.

The purpose of phenotyping now is to select the best lines from a segregating population and to evaluate fewer lines with greater replication in each cycle of selection. But, in a GS driven breeding cycle, the purpose of phenotyping is to estimate or re-estimate marker effects. It is far from clear, at this point, whether it will be advantageous to evaluate only the best lines or to evaluate few lines with high replication. Figure 1.2 therefore separates the germplasm improvement cycle from the prediction model improvement cycle. Indeed, if we use the guidelines for optimal QTL linkage mapping, evaluation should include not just the best, but the best and the worst lines (Darvasi and Soller 1992; Lander and Botstein 1989) and many unreplicated lines instead of few replicated lines (Knapp and Bridges 1990). Figure 1.2 also emphasizes the need for model updating and re-evaluation. Marker effects may change as a result of allele frequency changes (Muir 2007) or of epistatic gene action. Model updating with each breeding cycle should mitigate reduced gains from GS caused by these mechanisms. Thus, GS could radically change the practice of field evaluation for breeders. Of course, regardless of the breeding method used, final field evaluations of varieties across the target environments will be needed before they are distributed to farmers.

GS may also diminish the need for breeders to select parents strictly from the set of lines evaluated in their target environments (Goddard and Hayes 2007). Once a predictive linear model is established for their target environments, any genotype with

high target environment specific GEBV will become a candidate. Thus, GS should facilitate germplasm exchange and increase the probability of selecting useful germplasm.

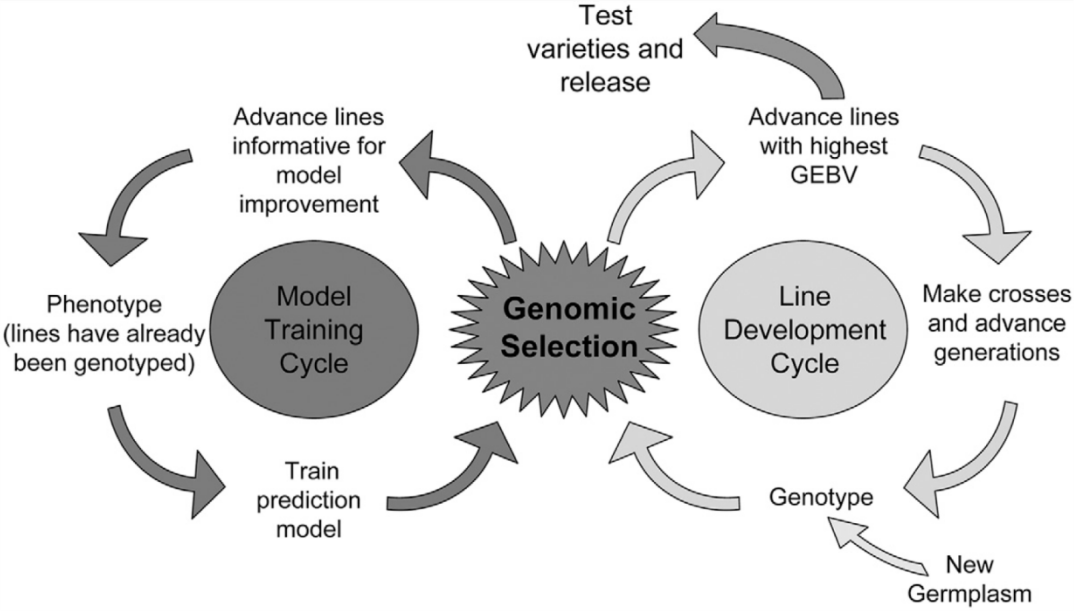


Figure 1.2 Flow diagram of a GS breeding program. Breeding cycle time is shortened by removing phenotypic evaluation of lines prior to selection as parents for the next cycle. Model training and line development cycle length will be crop and breeding program specific.

Conclusion

It has been predicted for over two decades that molecular marker technology would reshape breeding programs and facilitate rapid gains from selection (Stuber et al. 1982; Tanksley et al. 1989). The failure of current MAS to significantly improve polygenic traits has thwarted this prediction. Genomic Selection looks to fulfill it by using genomewide marker coverage to accurately estimate breeding values, accelerate the breeding cycle, and introduce greater flexibility in the relationship between phenotypic evaluation and selection. To do so, however, GS must shift from theory to practice. As evident in this review and interpretation, GS has almost exclusively been tested through simulation, and, therefore, its potential value should be assessed with cautious optimism. The accuracy of GS and its cost effectiveness must now be evaluated in breeding programs to provide the empirical evidence needed to warrant the addition of GS to the plant breeders' toolbox.

Acknowledgements

The authors thank Adam Famoso, Michael Gore, and Jesse Munkvold for their insights and critical reviews. Work of Elliot Heffner was funded by the USDA National Needs Fellowship Grant 2005-38420-15785. Work of Jean-Luc Jannink was partially funded by the USDA-NRI Grant Numbers 2008-55301-18746 and 2006-55606-16722. Additional funding for this research was provided by Hatch 140-149.

REFERENCES

Allard RW, Bradshaw AD (1964) Implications of genotype-environmental interactions in applied plant breeding. Crop Sci 4:503-508

Avendaño S, Woolliams J, Villanueva B (2004) Mendelian sampling terms as a selective advantage in optimum breeding schemes with restrictions on the rate of inbreeding. Genet Res 83:55-64

Beavis WD (1998) QTL analyses: Power, precision, and accuracy. Molecular Dissection of Complex Traits. 145–162

Bernardo R (2001) What if we knew all the genes for a quantitative trait in hybrid crops? Crop Sci 41:1-4

Bernardo R, Yu J (2007) Prospects for genomewide selection for quantitative traits in maize. Crop Sci 47:1082

Breiman L (1995) Better subset regression using the nonnegative garrote. Technometrics 37:373-384

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics 172:1165-1177

Broman KW, Speed TP (2002) A model selection approach for the identification of quantitative trait loci in experimental crosses. Journal of the Royal Statistical Society Series B 64:641-656

Calus M, Veerkamp R (2007) Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. J Anim Breed Genet 124:362-368

Carlborg O, Haley CS (2004) Epistasis: Too often neglected in complex trait studies. Nat Rev Genet 5:618-625

Chao SM, Zhang WJ, Dubcovsky J, Sorrells ME (2007) Evaluation of genetic diversity and genome-wide linkage disequilibrium among US wheat (*Triticum aestivum* L.) germplasm representing different market classes. Crop Sci 47:1018-1030

Crosbie TM, Eathington SR, Johnson GR, Edwards M, Reiter R, Stark S, Mohanty RG, Oyervides M, Buehler RE, Walker AK (2003) Plant breeding: Past, present, and future. Plant Breeding: The Arnel R.Hallauer Int.Symp., Mexico City 17–22

Daetwyler HD, Villanueva B, Bijma P, Woolliams JA (2007) Inbreeding in genome-wide selection. J Anim Breed Genet 124:369-376

Darvasi A, Soller M (1992) Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. TAG Theoretical and Applied Genetics 85:353-359

de Roos AP, Schrooten C, Mullaart E, Calus MP, Veerkamp RF (2007) Breeding value estimation for fat percentage using dense markers on *Bos taurus* autosome 14. J Dairy Sci 90:4821-4829

Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. Nat Rev Genet 3:22-32

Dekkers JCM (2007) Prediction of response to marker-assisted and genomic selection using selection index theory. J Anim Breed Genet 124:331-341

Eathington SR, Crosbie TM, Edwards MD, Reiter RS, Bull JK (2007) Molecular Markers in a Commercial Breeding Program. Crop Sci Crop Sci. 47: S154-S163

Edwards M, Johnson L (1994) RFLP for rapid recurrent selection. analysis of molecular marker data. American Society of Horticultural Science and Crop Science Society of America., Corvallis, OR33-40

Falconer D, Mackay T (1996) Quantitative genetics. Longman, Harrow, United Kingdom

Fernando RL (2007) Genomic selection. Acta Agriculturae Scandinavica, Section A-Animal Sciences 57:192-195

Fernando R, Grossman M (1989) Marker assisted selection using best linear unbiased prediction. Genet Sel Evol 21:467-477

Frisch M, Melchinger AE (2007) Variance of the parental genome contribution to inbred lines derived from biparental crosses. Genetics 176:477

Gaut BS, Long AD (2003) The lowdown on linkage disequilibrium. The Plant Cell Online 15:1502-1506

Gianola D, van Kaam JB (2008). Reproducing kernel hilbert spaces regression methods for genomic assisted prediction of quantitative traits. Genetics 178:2289

Gianola D, Fernando RL, Stella A (2006) Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics 173:1761

Goddard M, Hayes B (2007) Genomic selection. J Anim Breed Genet 124:323-330

Gonzalez-Recio O, Gianola D, Long N, Weigel KA, Rosa GJM, Avendano S (2008) Nonparametric methods for incorporating genomic information into genetic evaluations: An application to mortality in broilers. Genetics 178:2305

Grundy B, Villanueva B, Wooliams J (1998) Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. Genet Res 72:159-168

Habier D, Fernando R, Dekkers J (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389

Hamblin MT, Salas Fernandez MG, Casa AM, Mitchell SE, Paterson AH, Kresovich S (2005) Equilibrium processes cannot explain high levels of short-and medium-range linkage disequilibrium in the domesticated grass sorghum bicolor. Genetics 171:1247-1256

Hayes B (2007) QTL mapping, MAS, and genomic selection. A Short-Course Organized by Animal Breeding & Genetics Department of Animal Science Iowa State University June 4-8. (Available on-line at http://www.ans.iastate.edu/section/abg/ shortcourse/notes.pdf.) Verified 28 Aug. 2008

Hazel L (1943) The genetic basis for constructing selection indexes. Genetics 28:476-490

Hazel L, Lush JL (1942) The efficiency of three methods of selection. J Hered 33:393-399

Heckenberger M, Maurer HP, Melchinger AE, Frisch M (2008) The plabsoft database: A comprehensive database management system for integrating phenotypic and genomic data in academic and commercial plant breeding programs. Euphytica 161:173-179

Henderson C (1984) Application of linear models in animal breeding. University of Guelph, Ontario

Hill WG, Goddard ME, Visscher PM, Mackay TFC (2008) Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet 4:e1000008

Holland JB (2004) Implementation of molecular markers for quantitative traits in breeding programs—challenges and opportunities. New Directions for a Diverse Planet: Proceedings for the 4th International Crop Science Congress, Brisbane, Australia 26

Holland JB (2001) Epistasis and plant breeding. Plant Breed Rev 21:27-92

Holland JB (2007) Genetic architecture of complex traits in plants. Curr Opin Plant Biol 10:156-161

Hospital F, Moreau L, Lacoudre F, Charcosset A, Gallais A (1997) More on the efficiency of marker-assisted selection. Theor App Genet 95:1181-1189

Jannink JL, Bink MC, Jansen RC (2001) Using complex plant pedigrees to map valuable genes. Trends Plant Sci. 6:337

Kearsey M, Farquhar A (1998) QTL analysis in plants; where are we now? Heredity 80:137-142

Knapp S, Bridges W (1990) Using Molecular Markers to Estimate Quantitative Trait Locus Parameters: Power and Genetic Variances for Unreplicated and Replicated Progeny. Genetics 126:769-777

Koebner RMD, Summers RW (2003) 21st century wheat breeding: Plot selection or plate detection? Trends Biotechnol. 21:59-63

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743-756

Lander E, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199

Lange C, Whittaker JC (2001) On prediction of genetic values in marker-assisted selection. Genetics 159:1375-1381

Lee M (1995) DNA markers and plant breeding programs. Adv Agron 55:265-344

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Sunderland, Ma

Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Mol Breed 15:271-290

Mackay TFC (2009) The genetic architecture of complex behaviors: lessons from Drosophila. Genetica 136, 295–302

Mather KA, Caicedo AL, Polato NR, Olsen KM, McCouch S, Purugganan MD (2007) The extent of linkage disequilibrium in rice (Oryza sativa L.). Genetics 177:2223

Meuwissen T, (1997) Maximizing the response of selection with a predefined rate of inbreeding. J Anim Sci 75:934-940

Meuwissen THE, Goddard ME (1996) Marker-assisted selection in animal breeding schemes. Proc Int Soc Anim Genet Tours, France, 21–25 July160

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

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

Moreau L, Charcosset A, Hospital F, Gallais A (1998) Marker-assisted selection efficiency in populations of finite size. Genetics 148:1353-1365

Muir WM (2007) Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J Anim Breed Genet 124:342-355

Neimann-Sorensen A, Robertson A (1961) The association between blood groups and several production characteristics in three Danish cattle breeds. Acta Agric Scand 11:163-196

Piepho HP, Möhring J, Melchinger AE, Büchse A (2007) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 1-20

Piyasatian N, Fernando RL, Dekkers JCM (2007) Genomic selection for marker-assisted improvement in line crosses. Theor Appl Genet 115:665-674

Quinton M, Smith C, Goddard M (1992) Comparison of selection methods at the same level of inbreeding. J Anim Sci 70:1060

Rafalski, A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler IV ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479

Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet 123:218-223

Schön C, Utz S, Groh B, Truberg S, Openshaw S, Melchinger, A (2004) QTL mapping based on resampling in a vast maize testcross experiment confirms the infinitesimal model of quantitative genetics for complex traits. Genetics 167:485-498

Smith C (1967) Improvement of metric traits through specific genetic loci. Anim Prod 9:349-358

Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 315:848

Stuber C, Goodman M, Moll R (1982) Improvement of yield and ear number resulting from selection at allozyme loci in a maize population. Crop Sci 22:737

Tanksley S, Young N, Paterson A, Bonierbale M (1989) RFLP mapping in plant breeding: New tools for an old science. Biotechnology 7:257-264

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (zea mays ssp. mays L.). Proc. Natl. Acad. Sci. USA 98:9161

ter Braak CJF, Boer MP, Bink MCAM (2005) Extending Xu's Bayesian model for estimating polygenic effects using markers of the entire genome. Genetics 170:1435-1438

Tinker N, Yan W (2006) Information systems for crop performance data. Canadian Journal of Plant Science 86:647-662

van Arendonk J, Tier B, Kinghorn B (1994) Use of multiple genetic markers in prediction of breeding values. Genetics 137:319-329

Villanueva B, Pong-Wong R, Fernandez J,Toro M (2005) Benefits from marker-assisted selection under an additive polygenic genetic model. J. Anim. Sci. 83:1747-1752

Whittaker JC, Thompson R, Denham MC (2000) Marker-assisted selection using ridge regression. Genet Res 75:249-252

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

Woolliams J, Bijma P, Villanueva B (1999) Expected genetic contributions and their impact on gene flow and genetic gain. Genetics 153:1009-1020

Wray NR, Thompson R (1990) Prediction of rates of inbreeding in selected populations. Genet. Res. 55:41-54

Xu S (2003a) Theoretical basis of the Beavis effect. Genetics 165:2259

Xu S (2003b) Estimating polygenic effects using markers of the entire genome. Genetics 163:789-801

Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: From publications to practice. Crop Sci 48:391

Yan W, Tinker NA (2007) DUDE: A user-friendly crop information system. Agron. J. 99:1029

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zhang W, Lee HR, Koo DH, Jiang J (2008) Epigenetic modification of centromeric chromatin: Hypomethylation of DNA sequences in the CENH3-associated chromatin in arabidopsis thaliana and maize. The Plant Cell Online 20:25

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P (2007) An arabidopsis example of association mapping in structured samples. PLoS Genet 3:e4

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

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. The Plant Genome 1:5-20

CHAPTER TWO

PLANT BREEDING WITH GENOMIC SELECTION: GAIN PER UNIT TIME AND COST

ABSTRACT

Advancements in genotyping are rapidly decreasing marker costs and increasing genome coverage. This is facilitating the use of marker-assisted selection (MAS) in plant breeding. Commonly employed MAS strategies, however, are not well suited for agronomically important complex traits, requiring extra time for field-based phenotyping to identify agronomically superior lines. Genomic selection (GS) is an emerging alternative to MAS that uses all marker information to calculate genomic estimated breeding values (GEBVs) for complex traits. Selections are made directly on GEBV without further phenotyping. We developed an analytical framework to: 1) compare gains from MAS and GS for complex traits and 2) provide a plant breeding context for interpreting results from studies on GEBV accuracy. We designed MAS and GS breeding strategies with equal budgets for a high-investment maize program and a low-investment winter wheat program. Results indicate that GS can outperform MAS on a per year basis even at low GEBV accuracies. Using a previously reported GEBV accuracy of 0.53 for net merit in dairy cattle, expected annual gain from GS exceeded that of MAS by about 3-fold for maize and 2-fold for winter wheat. We conclude that if moderate selection accuracies can be achieved, GS could dramatically accelerate genetic gain through its shorter breeding cycle.

INTRODUCTION

Marker-assisted selection (MAS) has been a useful tool for plant breeders, but has had limited success in improving complex traits due, in part, to its inability to capture small-effect quantitative trait loci (QTL; Bernardo 2008; Xu and Crouch 2008). A promising approach, termed *genomic selection* (GS), attempts to avoid this deficiency by capturing both large and small-effect QTL with dense genome-wide molecular marker coverage to predict complex trait values (Meuwissen et al. 2001). Prediction accuracies reported by GS studies, coupled with the continued advances in high-throughput genotyping technologies, make GS a promising tool to increase plant breeding efficiency (reviewed by Heffner et al. 2009).

Genomic selection is already revolutionizing the dairy cattle breeding industry (Hayes et al. 2009). Empirical results from several dairy cattle breeding programs have shown prediction accuracies of genomic estimated breeding values (GEBVs) to be 2% - 20% greater than those of estimates using pedigree information (Hayes et al. 2009). While empirical GS results from plant breeding programs are not yet available, several studies have shown promising results. Bernardo and Yu (2007) showed through simulation that GS produced up to 43% greater genetic gain than marker-assisted recurrent selection for polygenic traits of low heritability in maize. Using empirical barley (*Hordeum vulgare* L) marker data and simulated phenotypes, Zhong et al. (2009), found that GEBV accuracy was similar to that of phenotype-based estimates. Finally, Lorenzana and Bernardo (2009) analyzed empirical data from maize, barley, and *Arabidopsis* biparental populations and showed GS response per cycle would be at least half that of phenotypic selection for nearly all traits studied.

In the above studies, gains from GS on a per cycle basis are not particularly impressive. However, by replacing time-intensive phenotypic evaluation of highly complex traits with GEBVs, GS can shorten breeding cycle length and thereby increase gains per unit time. This is especially true for perennial crops that require many years before phenotypic evaluations can be performed (Wong and Bernardo 2008). Based on GEBV accuracies found in their study on biparental plant populations, Lorenzana and Bernardo (2009) suggest GS gains per year would approach 1.5 times that of phenotypic selection in a case where 3 cycles of GS could be completed to each phenotypic selection cycle. Likewise, Schaeffer (2006) reported that using GEBVs in place of progeny testing in dairy cattle breeding could reduce costs by 92% and increase genetic gain per year by two-fold. In a further study using different GS breeding schemes, König et al. (2009) projected that GS could reduce breeding program costs by 22.4%.

The reallocation of breeding program resources needed to implement MAS or GS affects the overall budget, the selection pressure at different stages, and the length of the breeding cycle. Natural tradeoffs arise between allocations for phenotyping versus genotyping and for numbers of selection candidates versus the thoroughness of their evaluation. Determination of the relative value of MAS and GS will be conditional on certain allocation decisions. Our first objective was to compare gains from MAS and GS for breeding programs of annual crops while accounting for cost and cycle time of each strategy. Our second objective was to provide a plant breeding context for interpreting the potential impact of GEBV accuracy on gains from GS. To meet these objectives, we designed example GS and MAS plant breeding programs with equivalent budgets and compared them on the basis of gains per cycle and per unit time. To extend the range of application of our results, we designed two distinct

programs: 1) a high-investment maize inbred development program resembling commercial programs and 2) a low-investment winter wheat program resembling public programs.

MATERIALS AND METHODS

Selection Criteria

The economic value of a cultivar includes several traits of different value that are often genetically correlated with one another. The selection criterion used in this study was *net merit*, which is an index encompassing the relative importance of all traits beneficial to growers and consumers. The term net merit is predominantly used in the animal breeding literature (VanRaden 2004). The definition of net merit and its calculation by way of an index will differ between breeding programs due to differences in breeding goals (e.g., drought tolerance, disease pressure, forage quality versus grain yield). Nevertheless, net merit is expected to have a highly complex genetic architecture and low heritability in all situations.

Maize Breeding Program Structure

For maize, the MAS breeding program (MAS-BP) consisted of one stage of marker-based selection followed by two stages of phenotypic evaluation prior to parent selection (Fig. 2.1). The maize GS breeding program (GS-BP) consisted of a single stage of GS prior to parent selection (Fig. 2.1). Both programs were designed using doubled haploids (DHs) to reduce the time required for inbred development because DHs are routinely used in commercial, high-investment maize breeding

programs (Seitz 2005). We assumed that both programs have access to state-of-the art DH conversion technologies and employ off-season nurseries to achieve three growing cycles per year. Therefore, the four stages of the DH process – crossing parental lines, crossing F_1 plants by haploid inducer, double chromosomes of haploid plants, and selfing of doubled haploids (DH_0) – would take 1 year. During this process, DH_0 seedlings are genotyped for early generation selection via GS or MAS.

The first stage of selection in the MAS-BP consists of genotyping 4500 DH_0 lines for well-characterized QTL (e.g., disease resistance) and subsequent marker-based selection. Of these 4500 DH_0 lines, 20% are selected, selfed, and advanced to general combining ability (GCA) testing. The optimum number of testers and selection intensities for two stages of GCA testing were adapted from Longin et al. (2007). In the first GCA testing stage, 919 DH_1 lines are selfed, evaluated for *per se* performance, crossed to a single tester, and testcross progeny are evaluated at three locations. Of these 919 DH_1 lines, 45 are advanced to the next stage of GCA testing where they are selfed, evaluated for *per se* performance, crossed to five testers. Testcross progeny are evaluated at eight locations. Finally, 10 DH_2 lines are selected as parents to constitute the next breeding cycle and also enter advanced testing to evaluate specific combining ability (SCA) of inbreds prior to commercial development. Cycle length of the designed maize MAS-BP is three years.

Each maize GS-BP cycle consists of generating 6600 DH_0 lines, genotyping, GEBV calculation, and selecting ten DH_0 lines based on their net merit predictions. The ten lines selected are used as parents for the next breeding cycle and also enter advanced testing. An additional 56 lines (66 total) are selected, advanced through two stages of seed increase and inbred *per se* evaluation. Seed quantities from the selfed

DH_0 would likely be sufficient for parent recombination, but not for extensive advanced testing; therefore, two cycles of selfing are used to increase seed quantity. The additional 56 lines are included in this stage to increase the inbred *per se* performance data available for updating the GS prediction model, as this is an important component of maize inbred line net merit. The data from additional inbred testing, advanced testing, and historical records are used for training GS models. Cycle length for the designed maize GS-BP is one year.

Winter Wheat Breeding Program Structure

The general structure of the winter wheat breeding program (Fig. 2.2) was modeled after the Cornell University Winter Wheat Breeding Program. The winter wheat MAS- and GS-BP were designed to be identical for the first five stages. Inbred (F_5) lines are created by advancing selected individuals through single seed descent (SSD). Greenhouses are used to reduce generation time from 1 to 0.5 years. F_2 and F_3 plants are genotyped for 10 well-characterized QTL (e.g., disease, milling quality) and undergo marker-based selection. This two-stage enrichment strategy is used to increase the frequency of desired alleles because of the improbability of obtaining progeny homozygous for all target QTL in small populations (Bonnett et al. 2005). While these markers could be included in whole-genome profiling for the GS-BP, we presumed the enrichment step is still used in a GS-BP to eliminate unnecessary costly genome-wide marker scoring and greenhouse space for lines not carrying essential QTL alleles. The number of lines in genotyping stages was set to a multiple of 96 to match genotyping plate size for efficiency and cost savings, which are not trivial when genotyping on a small scale.

In the winter wheat MAS-BP, 288 F_5-derived lines are planted in single-row plots for seed increase and visual evaluation of agronomic traits. Twenty-five percent of the lines are culled for being visually deficient in agronomic performance and plant type. Once inbred lines have been developed with sufficient amounts of seed, three stages of field evaluation are conducted prior to selection of 10 parental lines for recombination and advancement to regional testing prior to variety release. Cycle length for the designed winter wheat MAS-BP is seven years.

In the winter wheat GS-BP, field evaluation prior to parent selection is replaced by conducting GS on 288 F_5-derived lines. Ten F_5-derived lines are selected on the basis of their net merit GEBV and recombined to begin the next cycle. An additional 206 F_5-derived lines are selected (which includes those 10 used for recombination) to be grown in the field for seed increase. Similar to the MAS-BP, a fraction of the lines (33%) are culled during seed increase due to visual agronomic deficiencies. The remaining 144 lines are phenotyped, used for updating GS models, and serve as candidates for advanced testing and subsequent variety release. We assumed that this additional data would be necessary to supplement data from advanced testing for updating the GS prediction model. Cycle length for this winter wheat GS-BP is three years.

Budgets

Budgets for the maize and winter wheat breeding programs (Table 2.1) are represented by maize yield plot units (YPUs), i.e., the cost of growing and evaluating a single maize yield trial plot where 1 YPU=US$20 (Bernardo and Yu 2007). The budget of the high-investment maize inbred development program (excluding

advanced testing and commercialization) was set to 10,000 YPUs (US$200,000). The budget of the low-investment winter wheat breeding program (excluding advanced testing and commercialization) was set to 3,800 YPUs (US$76,000), which is slightly greater than one-third the budget of the maize program.

The budgets of the GS and MAS breeding programs within each crop species were set to be equivalent allowing easy comparison between the two breeding strategies in terms of expected genetic gain at equal investment. Minor differences in final budgets resulted from rounding to whole numbers for field plots and population sizes (Table 2.1). Budgets were calculated on a per cycle basis by totaling the costs for each stage of selection. In practice, plant breeding programs are operated as pipelines where each stage occurs once per year so that new selection candidates, parents, and varieties are produced each year. Therefore, despite differing selection cycle lengths for the MAS- and GS-BPs, cost per cycle is equivalent to cost per year. This allowed gains per cycle and per year to be compared between GS- and MAS-BPs on the basis of equivalent budgets.

Assuming a highly efficient DH production system, the cost of maize DH line production was 0.5 YPU (Longin et al. 2007). The GS-BP produces 6,600 DHs whereas the MAS-BP produces only 4,500 DHs because of extra phenotyping costs. Genotyping is performed using single nucleotide polymorphisms (SNPs), with the cost per SNP being between US$0.03-0.15 (Bernardo 2008). Cost of genotyping and DNA extraction was 0.5 YPU for MAS and 1.0 YPU for GS. It was assumed genotyping costs do not increase linearly with marker number because of fixed costs for DNA extraction and economies of scale. Genotyping for MAS was budgeted for 50-100 markers with the assumption that high-input maize breeding programs would be

selecting many loci identified in previous linkage and association mapping efforts. Genome-wide genotyping for GS was budgeted for several hundred or more markers. More precise estimates of marker number and cost for private sector maize MAS- and GS-BP were not publicly available.

For the winter wheat programs, costs were approximated using current data from the Cornell Winter Wheat Breeding Program. Cost for a yield plot trial was 2.0 YPUs (US$40), a single greenhouse cycle was 0.4 YPUs (US$8), and a field seed increase was 1.0 YPU (US$20). The cost of DNA extraction and MAS genotyping was 1 YPU for 10 microsatellites (≈$1.50 per marker; Wong and Bernardo 2008). DNA extraction and genome-wide genotyping for GS with several hundred or more markers was 2.0 YPUs, which is currently possible with Diversity Arrays Technology (Akbari et al. 2006).

We assumed maize and winter wheat MAS- and GS-BPs had been ongoing rather than account for all the variable expenses in launching these programs. This includes a trained prediction model and identified and validated marker-QTL associations for MAS. Currently, it is unknown how the yearly costs for QTL discovery and validation for MAS and the costs of model training for GS would compare in practice, partly because of the diverse situations encountered in plant breeding programs. To give the MAS-BPs the benefit of this uncertainty, resources for these activities were not budgeted in the MAS program, but were included in the GS-BPs by allocating resources to phenotyping solely for GS model updating.

Table 2.1 Budgets of the winter wheat (3,800 YPUs) and maize (10,000 YPUs) MAS- and GS-BPs. GS-BP figures are lightly shaded and MAS-BP figures are heavily shaded. Tstrs=number of GCA testers, Locs=number of test locations; Geno=genotyping; TC=testcross; Pop=population.

Maize		GS-BP		MAS-BP			
Stage	Cost	Units	Cost	Units	Tstrs	Locs	Cost
DH	0.5	6600	3300	4500			2250
MAS-Geno	0.5			4500			2250
GS-Geno	1	6600	6600				
DH1 Plots	1	66	66	919			919
DH2 Plots	1	66	66	45			45
TC1 Plots	1			919	1	3	2757
TC2 Plots	1			45	5	8	1800
Total			10032				10021

Winter Wheat		GS-BP			MAS-BP		
Stage	Cost	Units	Locs	Cost	Units	Locs	Cost
GH1-F1	0.4	40		16	40		16
GH2-F2	0.4	768		307.2	768		307.2
MAS-Geno-F2	1	768		768	768		768
GH3-F3	0.4	576		230.4	576		230.4
MAS-Geno-F3	1	576		576	576		576
GH4-F4	0.4	288		115.2	288		115.2
GH5-F5	0.4	288		115.2	288		115.2
GS-Geno-F5	2	288		576			
Seed Increase	1	216	1	216	288	1	288
Yield Trial 1	2	144	3	864	216	1	432
Yield Trial 2	2				108	3	648
Yield Trial 3	2				54	3	324
Total				3784			3820

Calculating response to selection

Univariate and multivariate (Cochran 1951; Utz 1969) forms of the classical breeder's equation were used to determine the expected genetic gain for each program outlined above. The univariate breeder's equation was used for the GS-BPs because they include only one stage of selection. The expected genetic gain can be expressed as $R=ir_A\sigma_A$, where R is the response to selection, i is the intensity of selection (mean deviation of selected individuals in units of phenotypic standard deviation), r_A is the selection accuracy, and σ_A is the standard deviation of breeding values (Falconer and Mackay 1996). Selection accuracy is equal to the correlation between selection criteria and breeding value (i.e. correlation between phenotypes or GEBVs and true breeding values (TBVs). In the context of mass selection on the phenotype, r_A is equal to the square root of the narrow-sense heritability. Selection accuracy (r_A) will be used herein to describe the ability of phenotypes, GEBVS, or their combination to predict TBVs.

To calculate expected genetic gain of the MAS-BPs, involving multiple stages of selection, exact formulas originally derived by Cochran (1951) for two stages and extended to three stages by Utz (1969), as described in Tomerius (2001), were used. In addition to the parameters that determine expected response from a single stage of selection, multi-stage selection is dependent upon the correlation between selection criteria employed in each stage (Tomerius 2001). Wricke and Weber (1986) provided a detailed description for calculating expected gain from multiple stages of selection. We wrote an R program (R Development Core Team 2009) involving the R package *mvtnorm* (Genz et. al. 2009) to numerically determine the truncation points of the multivariate distribution for stages two and three. Bulmer's recursive equation

(Bulmer 1971; Falconer and Mackay 1996) was iterated until genetic variance reached equilibrium. The equilibrium genetic variance was used to calculate genetic gain.

Quantitative genetic parameters for the maize breeding programs

For the maize model, the relative values of variance components were taken from the reference scenario (VC2.2) of Longin et al. (2007). $\sigma^2_{GCA} = 0.40$, $\sigma^2_{GCAxy} = 0.20$, $\sigma^2_{GCAxl} = 0.20$, $\sigma^2_{GCAxyxl} = 0.40$, $\sigma^2_{SCA} = 0.20$, $\sigma^2_{SCAxy} = 0.10$, $\sigma^2_{SCAxl} = 0.10$, $\sigma^2_{SCAxyxl} = 0.20$, $\sigma^2_e = 2$, where the subscripts *GCA* = general combining ability, *GCA*×*y* = GCA by year interaction, *GCA*×*l* = GCA by location interaction, *GCA*×*y*×*l* = GCA by year by location interaction, *SCA* = specific combining ability, and *e* = residual. The interactions involving SCA correspond to those of GCA. These variance component values produce a $h^2 = 0.11$ on a plot basis. Longin et al. (2007) based these variance components on results from DH testcross populations of commercial breeding programs and elite material from the University of Hohenheim maize breeding program.

In the maize MAS-BP, it was assumed that markers known to be tightly linked to or within well-characterized QTL are available. Considering such a resource, the accuracy of MAS (h_1) for predicting net merit was set to 0.40. A lack of published MAS accuracies for net merit in private, high-investment maize breeding programs forced this approximation. An accuracy of 0.40 on net merit using MAS is undoubtedly an overestimate, but provides a conservative comparison of GS-BP to MAS-BP.

Selection accuracies in stages two and three were calculated as $h_i = \sqrt{\sigma^2_{GCA} / \sigma^2_{\bar{x}(i)}}$ where $\sigma^2_{\bar{x}(i)}$ is the variance of the DH testcross mean in stage *i*. For stage two,

$$\sigma^2_{\bar{x}(2)} = \sigma^2_{GCA} + \sigma^2_{GCA\times y} + \frac{\sigma^2_{GCA\times l}}{L_2} + \frac{\sigma^2_{GCA\times y\times l}}{L_2} + \frac{\sigma^2_{SCA}}{T_2} + \frac{\sigma^2_{SCA\times y}}{T_2} + \frac{\sigma^2_{SCA\times l}}{T_2 L_2} + \frac{\sigma^2_{SCA\times y\times l}}{T_2 L_2} + \frac{\sigma^2_e}{T_2 L_2}$$

where L_2 is the number of locations and T_2 is the number of testers used in stage two. For stage three, $\sigma^2_{\bar{x}(i)}$ is the variance of index scores calculated by combining testcross performance in stage two with average testcross performance in stage three (Wricke and Weber 1986):

$$\sigma^2_{\bar{x}(3)} = \sigma^2_{GCA} + \frac{\sigma^2_{GCA\times y}}{2} + \frac{\sigma^2_{GCA\times l}}{L_2 + L_3} + \frac{\sigma^2_{GCA\times y\times l}}{L_2 + L_3} + \frac{\sigma^2_{SCA}}{T_2 + T_3} + \frac{\sigma^2_{SCA\times y}}{T_2 + T_3} + \frac{\sigma^2_{SCA\times l}}{T_2 L_2 + T_3 L_3} + \frac{\sigma^2_{SCA\times y\times l}}{T_2 L_2 + T_3 L_3} + \frac{\sigma^2_e}{T_2 L_2 + T_3 L_3}$$

where L_3 is the number of locations and T_3 is the number of testers used in stage three. This set of assumptions will be referred to as the reference scenario, and the resulting accuracy will be referred to as the reference heritability.

Correlations between selection criteria in different stages were calculated as $r_{12} = h_1 \times h_2$, $r_{13} = h_1 \times h_3$, and $r_{23} = \sigma_{23} / \sqrt{\sigma^2_{\bar{x}(2)} \times \sigma^2_{\bar{x}(3)}}$, where σ_{23} is the covariance between the selection criteria of stages two and three:

$$\sigma_{23} = \sigma^2_{GCA} + \frac{L_c \sigma^2_{GCA\times l}}{L_2 L_3} + \frac{T_c \sigma^2_{SCA}}{T_2 T_3} + \frac{L_c T_c \sigma^2_{SCA\times l}}{L_2 L_3 T_2 T_3}$$

(Longin et al. 2007; Wricke and Weber 1986). These correlations were used in the three-stage selection formulas.

To investigate the effect of our assumptions on the results, we artificially varied the accuracies of the MAS-BP. Accuracies for stages two and three were doubled, resulting in $h_2 = 0.93$ and $h_3 = 1$. This scenario will be referred to as the high-heritability scenario. In this scenario, correlations between selection criteria in the different stages were calculated as the product of the accuracies of each stage, as r_{12} and r_{13} above.

Quantitative genetic parameters of a the winter wheat breeding programs

Relative values of the variance components for the winter wheat breeding program were set to $\sigma^2_G = 0.40$, $\sigma^2_{Gxy} = 0.20$, $\sigma^2_{Gxl} = 0.20$, $\sigma^2_{Gxyxl} = 0.40$, and $\sigma^2_e = 2$, where σ^2_G is the additive genetic variance, σ^2_{Gxy} is the interaction between breeding values and years, σ^2_{Gxl} is the interaction between breeding values and locations, σ^2_{Gxyxl} is the interaction between all three aforementioned factors, and σ^2_e is the residual variance. These variance component values produce a plot basis $h^2 = 0.13$. These were chosen to be similar to the maize variance components, excluding the SCA variance components.

The marker-based enrichment stage is equivalent between the winter wheat MAS- and GS-BP's (Fig. 2.1, Table 2.1), and therefore, expected genetic gains were calculated for all stages after marker-based enrichment. Three stages of field evaluation and selection were included in calculating the expected gain of the MAS-BP. Accuracies in the second and third stages of field evaluation were calculated assuming performances from previous years were combined with the present year performance into an index. Accuracy for stage i was calculated as $h_i = \sqrt{\sigma^2_G / \sigma^2_{\bar{x}(i)}}$, where $\sigma^2_{\bar{x}(i)} = \sigma^2_G + \frac{\sigma^2_{G\times y}}{Y_i} + \frac{\sigma^2_{G\times l}}{L_i} + \frac{\sigma^2_{G\times y\times l}}{L_i} + \frac{\sigma^2_e}{L_i}$

where Y_i is the number of years used in calculating stage *i* index performance (e.g., in stage 2, Y=2) and L_i is the sum of location-year combinations (e.g., $L_3 = 7 = 1$ location in year 1 + 3 locations in year 2 + 3 locations in year 3). Correlations between selection criteria in different stages were calculated as r_{23}, as they were for the maize program.

RESULTS

Expected genetic gain per cycle of the maize MAS-BP was 1.34 genetic standard deviation units (hereafter abbreviated to "units") assuming the reference variance components (Fig. 2.3). Under the reference scenario assumptions, a GEBV accuracy of 0.55 or greater would be needed for the maize GS-BP to exceed the MAS-BP in genetic gain *per cycle*. The winter wheat GS-BP is expected to exceed the MAS-BP in genetic gain *per cycle* with a GEBV accuracy of 0.75 or greater. The maize GS-BP had a lower "break-even accuracy" than the wheat GS-BP because the maize program allocation of resources allowed the generation of more DH lines and thus greater selection intensity in the GS-BP than the MAS-BP. Obviously, the expected genetic gain per cycle under the high-heritability scenario was higher, requiring GEBV accuracies to be near 1 for the GS-BPs to achieve gains similar to those of the MAS-BPs.

A more relevant basis on which to compare genetic gain from different breeding schemes is, however, on a unit time and cost basis (Fehr 1987). Budgets of the GS-BP and MAS-BP were set to be approximately equal. For maize, the impact of being able to achieve 3 cycles of the GS-BP within the time required to achieve 1 cycle of the MAS-BP is clearly illustrated in Fig. 2.4. A similar situation exists for

winter wheat: the GS-BP can achieve 2.33 cycles to 1 cycle of the MAS-BP (Fig. 2.4). For maize, a GEBV accuracy of only 0.20 is needed for the GS-BP's expected genetic gain *per year* to surpass that of the MAS-BP under either heritability assumption. A slightly higher threshold of 0.30 was found for winter wheat. If GEBV accuracies of 0.50 could be achieved, assuming the reference heritabilities, genetic gain *per year* for GS-BP would exceed that of MAS-BP by about 3-fold for maize and 2-fold for winter wheat. Even under the high-heritability scenario, GS would be expected to provide about 2.5- and 1.5-fold more genetic gain for maize and wheat, respectively.

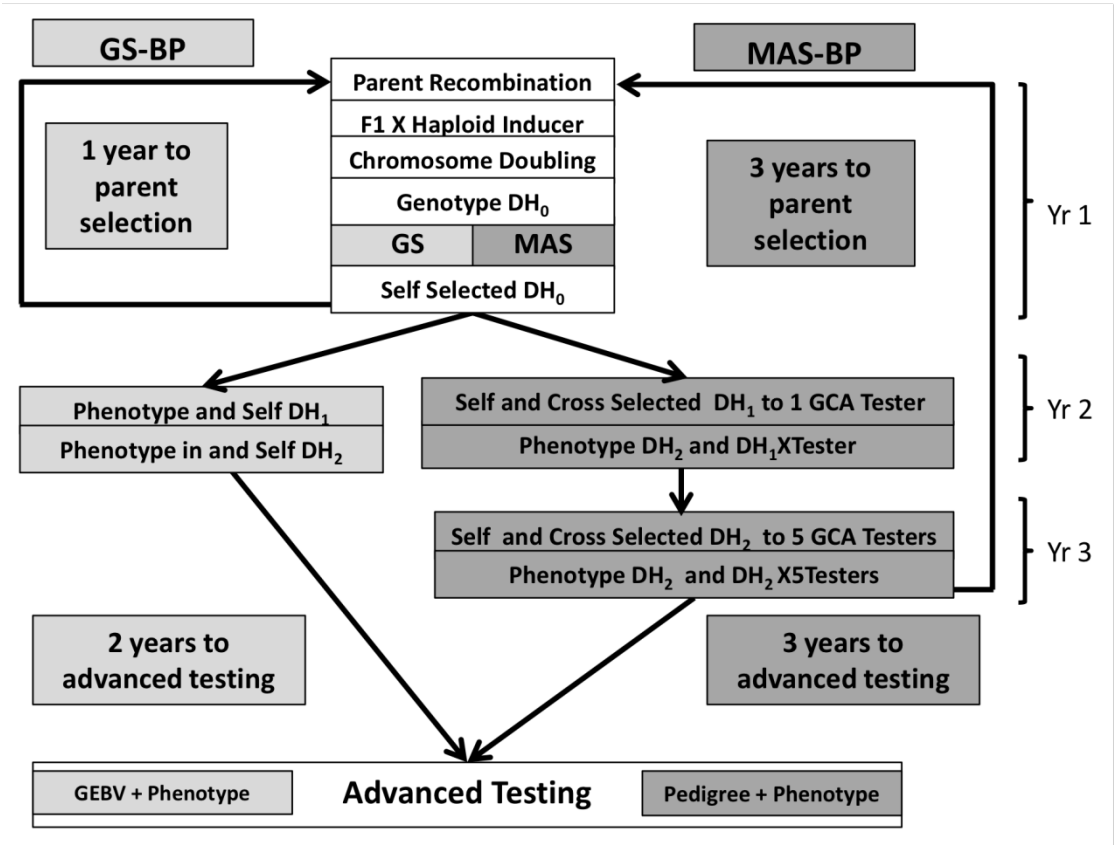


Figure 2.1 The maize MAS- and GS-BP schemes. The GS-BP selection cycle length is 1 year; whereas, the MAS-BP selection cycle length is 3 years. GS-BP stages are lightly shaded and MAS-BP stages are heavily shaded, and stages common to both programs are not shaded.

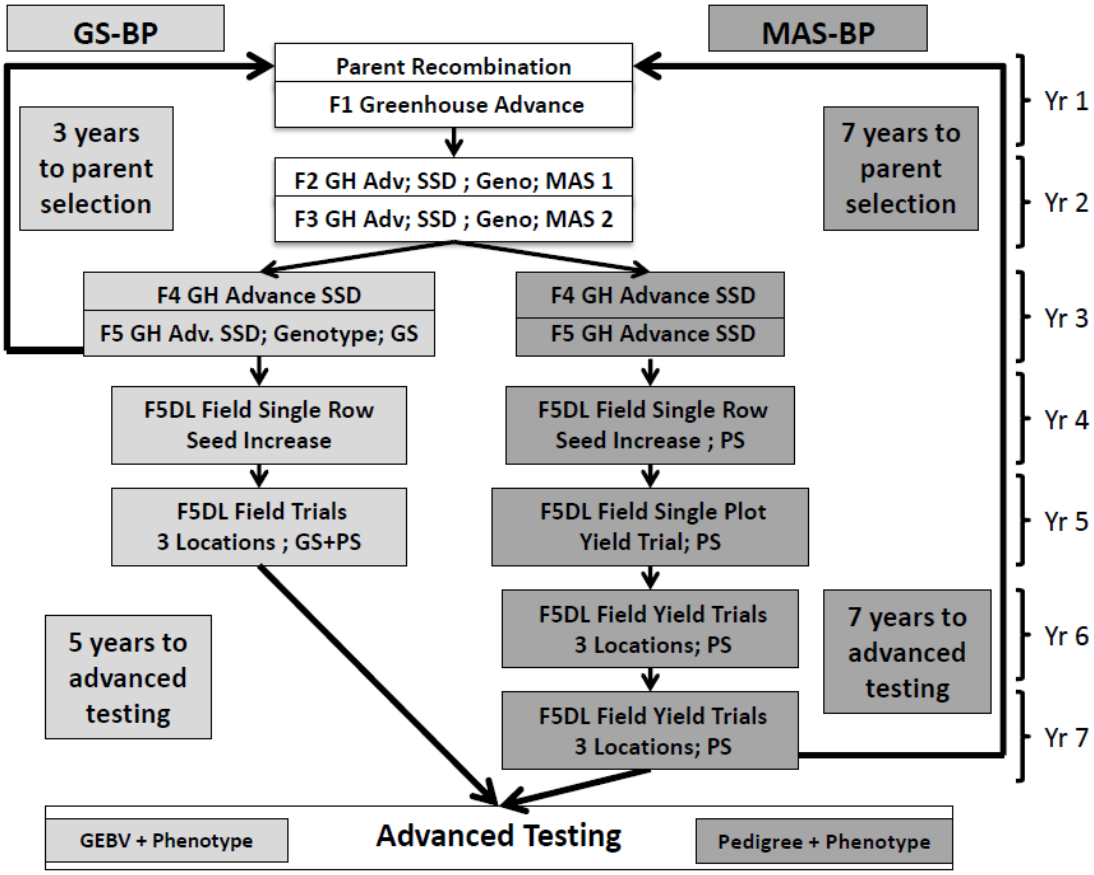


Figure 2.2 The winter wheat MAS- and GS-BP schemes. The GS-BP selection cycle length is 3 years; whereas, the MAS-BP selection cycle length is 7 years. GS-BP stages are lightly shaded, MAS-BP stages are heavily shaded, and stages common to both programs are not shaded. GH= greenhouse; Adv= advance; Geno=genotyping; PS=phenotypic selection; F5DL= F_5 derived line.

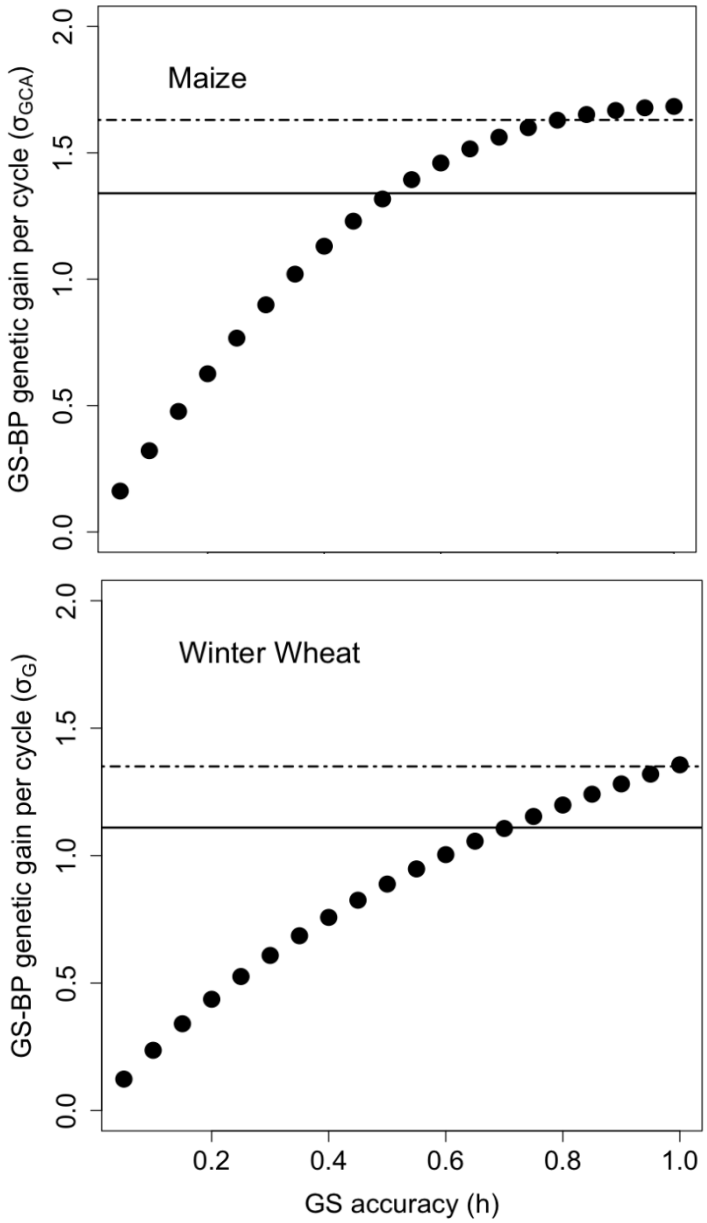


Figure 2.3 Expected genetic gain per cycle of the GS-BP plotted against the accuracy of GEBVs. Solid line indicates expected genetic gain of the MAS-BP using the reference heritability, while the dashed line indicates the expected genetic gain using the high-heritability scenario. Units for maize are GCA standard deviation units. Units for wheat are genetic standard deviation units.

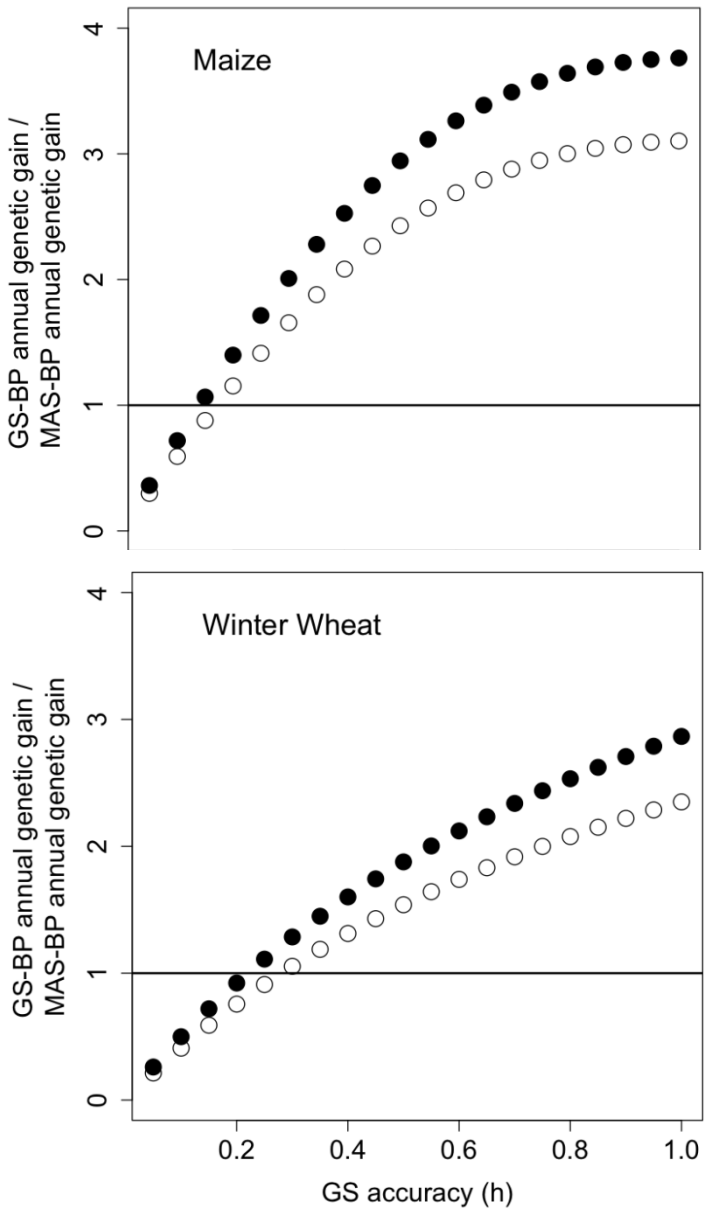


Figure 2.4 Ratio of annual genetic gain expected to be achieved by the GS-BP to that of the MAS-BP. Ratios were calculated using the reference-heritability (closed circles) and the high-heritability scenario (open circles).

DISCUSSION

Our results show that it is feasible to design a GS breeding program that achieves greater genetic gain per year with only low to moderate GEBV accuracies and a budget that is equivalent to a MAS breeding program (Fig. 2.4). At high GEBV accuracies, genetic gain using GS is expected to be several fold higher than MAS. This result holds even when we assume heritabilities in the MAS-BP that are unrealistically high. Since the efficiency of selecting on markers relative to selecting on phenotypes increases as heritability decreases (Lande and Thompson 1990; Hospital et al. 1997), the advantage of GS over MAS should be even greater if lower heritabilities are assumed.

Computer simulation studies have found GEBV accuracies between 0.62 and 0.85 using simulated (Habier et al. 2007; Meuwissen et al. 2001) or empirical marker data (Zhong et al. 2009) and simulated breeding values and phenotypes. Also, an empirical study of biparental plant populations using cross-validation found the GEBV accuracies for grain yield averaged 0.54 and 0.61 for three maize and two barley populations, respectively (Lorenzana and Bernardo 2009). Clearly, if these accuracies hold for actual breeding programs, our results indicate GS to be a clear winner over MAS in terms of genetic gain per unit time and cost.

Empirical GEBV accuracies from plant breeding programs are not yet publicly available, but high-quality data is available from livestock studies, particularly dairy cattle. VanRaden et al. (2009) was able to predict net merit of a validation set with 0.53 accuracy using 38,416 SNPs and a training population of 3,576 Holstein bulls with breeding values measured by progeny testing. The validation set consisted of

1,759 progeny tested bulls independent of the training population. Using this accuracy, our results showed gain per year for GS exceeded that of MAS by about 3-fold for maize and 2-fold for winter wheat. The differences in population dynamics and breeding objectives make it difficult to directly extend GEBV accuracies from cattle to plants; however, some of these differences show promise for high GEBV accuracies in plants. For example, levels of LD across cattle are much lower (e.g. de Roos et al. 2008) than LD within plant breeding programs, especially self-pollinating species (e.g. Chao et al. 2007). Thus, fewer markers should be needed in plants than animals to have all QTL in LD with at least one marker. Also, as cattle GS models have been trained with highly accurate phenotypes from a cooperative database consisting of extensive bull progeny testing (VanRaden et al. 2009), plant GS models can be trained with highly accurate phenotypes obtained through sound experimental design and replication in time and space. These attributes, along with the previously discussed simulation results, suggest that GEBV accuracies needed for GS to significantly outperform MAS in gain per year will be attainable in plant breeding.

The shorter breeding cycle of the GS-BPs resulted in greater annual gains than the MAS-BP under low to moderate GEBV accuracies. Another benefit of accelerating cycle time is the concentration of resources on a narrower germplasm pool. This can be best illustrated by the maize breeding programs. Because we assumed all phases of the breeding cycle occur each year, the annual budget was equal to the budget of a single cycle for each program. Thus, over the three-year maize MAS-BP cycle, three times the budget of a single cycle will have been spent, in effect spread out over three different sets of germplasm all going through the pipeline. In contrast, in the GS-BP during that same time, all of those three budgets will have been spent advancing the same set of germplasm. This allows for resources in the GS-BP to

be concentrated on advancing a more elite, though narrower, germplasm pool. Gains on that pool, cumulating over cycles, are therefore greater. This difference, however, would logically also result in a greater loss of genetic diversity over time. Therefore, we believe that research on the maintenance of genetic diversity within GS programs will be important.

The commercial maize and public winter wheat breeding programs were modeled to provide a contrast of breeding strategies and operating budgets to allow more general application of findings. A major difference in assumptions between the two types of breeding programs involved the training populations used for GEBV estimation. In the maize GS-BP, we assumed that inbred testing and large scale advanced testing data would be available for training a robust GS prediction model. The winter wheat GS-BP, on the other hand, allocated greater resources to additional phenotyping to supplement data from regional advanced trials that are much smaller and less intensive than a typical commercial maize advanced testing program. These extra resources are also allocated in the public winter wheat GS-BP because such a program would probably lack access to extensive genotype and phenotype databases. Moreover, less intensive phenotyping is used for training the winter wheat GS model (unreplicated plots at three locations) compared to a typical commercial maize advanced testing program, resulting in lower heritability and thus requiring a greater number of lines for model updating.

Another important difference between the maize and winter wheat GS-BP's is the frequency of GS model updating. In the case of the winter wheat GS-BP, three years are required for generation of the F_5-derived lines genotyped for selection and two years are required to obtain phenotypes on lines selected for model updating.

Therefore, the training population includes individuals derived from the cycle just previous to that of the selection candidates. Selection candidates of the maize GS-BP, on the other hand, are separated from individuals in the training population by two cycles. This larger separation occurs because only 1 year is required to develop DH lines (i.e., selection candidates), and 2 years are needed for selected DHs to produce advanced testing results for GS model updating. Model accuracy is highest when the training population includes individuals of the same generation as the selection candidates. Accuracy declines as generation number between the last model update and selection candidates increases (Habier et al. 2007; Meuwissen et al. 2001; Muir 2007) because selection causes changes in variances, allele frequencies, and LD relationships between markers and QTL (Bulmer 1971; Muir 2007). Under random-mating, simulations have shown model accuracy to decrease by about 5% per generation (Meuwissen et al. 2001; Habier et al. 2007), but accuracy decrease was much more rapid under selection (Muir 2007). Therefore, the greater minimum GEBV accuracies required for winter wheat in comparison to maize (Fig. 2.4) could be compensated by potentially higher GEBV accuracies in wheat caused by more frequent model updating.

Our results show potentially enormous benefits from conducting a GS breeding program for crops. These findings are in line with similar studies on livestock breeding economics and expected genetic gain from GS compared to conventional programs (König et al. 2009; Schaeffer, 2006). Despite the potential benefits, high startup costs required for amassing a large enough training population and fear of low accuracies are possible hindrances to transitioning to a GS breeding program. A key finding of this study is that even at low GEBV accuracies, GS-BPs were able to perform at least as well as MAS-BPs due to faster cycles of selection and

recombination. While low GEBV accuracy may also raise concern about variability in selection response, selection response at low accuracy is generally less variable than response at high accuracy (Hill 1974). Nevertheless, GS-BPs will go through more cycles of selection and variability accumulates with each cycle so that GS-BP variability will likely exceed that of MAS-BP. The deterministic methods used in this study analytically predict expected rates of gain, not variation around the expectation. Stochastic simulation and perhaps most importantly empirical results are needed on genetic gain over time using GS for complex traits, such as net merit, in dynamic plant breeding programs.

As for any discussion on the impact and use of technology, our assumptions and results will quickly be outdated. Genotyping and sequencing technology is advancing at an extremely rapid rate, which is reducing the cost of dense marker data. For instance, human geneticists are looking forward to completing ambitious projects – $1000 human genome sequence, 1000 Genomes Project, and Personal Genome Project – that were nearly unthinkable just a few years ago (von Bubnoff 2008). Similar advances are being made in crop genotyping, as more species are being sequenced more quickly and new marker technologies are being applied to crops (Varshney et al. 2009). Also, continued advancement in computational techniques for predicting GEBVs holds great potential for increasing accuracy at little to no extra cost (e.g. Gianola et al. 2009; Habier et al. 2009). All the while, phenotyping costs are stagnant or increasing.

We conclude that GS could significantly increase genetic gain per year and that results from this study warrant more research on integrating GS in plant breeding programs. The continued advancement of high-throughput genotyping, statistical

models for calculating GEBVs, and GS breeding methodologies will only strengthen this conclusion.

Acknowledgements

Support for the work of Elliot Heffner was provided by USDA National Needs Graduate Fellowship Competitive Grant No. 2005-38420-15785 from the National Institute of Food and Agriculture. Support for the work of Aaron Lorenz and Jean-Luc Jannink was provided by USDA-NIFA grant No. 2009-85606-05701 and No. 2009-65300-05661. Additional funding for this research was provided by USDA – NIFA National Research Initiative CAP grant No. 2005-05130 and by Hatch 149-402.

REFERENCES

Akbari M, Wenzl P, Caig P, Carling J, Xia L, Yang SY, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes N, Sharp P, Vaughan P, Rathmell B, Huttner E, Kilian A (2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor App Genet 113:1409-1420

Bernardo R (2008) Molecular markers and selection for complex traits in plants: Learning from the last 20 years. Crop Science 48:1649-1664

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

Bonnett, DG, Rebetzke GJ, Spielmeyer W (2005) Strategies for efficient implementation of molecular markers in wheat breeding. Molecular Breeding 15:75-85

Bulmer MG (1971) Effect of selection on genetic variability. American Naturalist 105:201-211

Chao SM, Zhang WJ, Dubcovsky J, Sorrells ME (2007) Evaluation of genetic diversity and genome-wide linkage disequilibrium among US wheat (*Triticum aestivum* L.) germplasm representing different market classes. Crop Science 47:1018-1030

Cochran WG (1951) Improvement by means of selection. *In* J. Neyman (ed.) Proc. 2nd Berkeley Symp. On Mathematical Statistics and Probability, Berkeley, CA. 31 July- 12 Aug. 1950. Univ. of California Press, Berkeley

de Roos APW, Hayes BJ, Spelman RJ, Goddard ME (2008) Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. Genetics 179:1503-1512

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. 4th ed. Longmans Green, Harlow, Essex, UK

Fehr WR (1987) Principles of cultivar development. Volume 1. Theory and technique. Macmillan Publishing Company, New York

Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T (2009) mvtnorm: Multivariate Normal and t Distributions. R package version 0.9-7. URL http://CRAN.R-project.org/package=mvtnorm

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

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

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

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009a) Invited review: Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci 92:433

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

Hill WG (1974) Variability of Response to Selection in Genetic Experiments. Biometrics 30: 363-366

Hospital F, Moreau L, Lacoudre F, Charcosset A, Gallais A (1997) More on the efficiency of marker-assisted selection. Theoretical and Applied Genetics 95:1181-1189

König S, Simianer H, Willam A (2009) Economic evaluation of genomic breeding programs. J. Dairy Sci. 92:382-391

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743-756

Longin CFH, Utz HF, Melchinger AE, Reif JC (2007) Hybrid maize breeding with doubled haploids. II. Optimum type and number of testers in two-stage selection for general combining ability. Theor App Genet 114:393-402

Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor and App Genet 120:151-161

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

Muir, W M (2007) Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J Anim Breed Genet 124:342-355

R Development Core Team (2009) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet 123:218-223

Seitz G (2005) The use of doubled haploids in corn breeding. p. 1-7. *In* Proc. Ill. Corn Breeders School. 41st. Univ. of Illinois. 7-8 Mar. 2005. Univ. of Illinois. Urbana-Champaign

Tomerius AM (2001) Optimizing the development of seed-parent lines in hybrid rye breeding. Ph.D. diss. Univ. of Hohenheim, Stuttgart, Germany

Utz HF (1969) Mehrstufenselektion in der Pflanzenzüchtung. (In German.) Ph.D. diss. Univ. of Hohenheim, Stuttgart, Germany

VanRaden PM (2004) Invited review: Selection on net merit to improve lifetime profit. J Dairy Sci 87:3125-3131

VanRaden P, Van Tassell C, Wiggans G, Sonstegard T, Schnabel R, Taylor J, Schenkel F (2009) Invited review: Reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92:16-24

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27:522-530

von Bubnoff A (2008) Next-generation sequencing: The race is on. Cell 132:721-723

Wong CK, Bernardo R (2008) Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. Theor Appl Genet 116:815-824

Wricke G, Weber WE (1986) Quantitative genetics and selection in plant breeding. Walter De Gruyter and Co., Berlin

Xu YB, Crouch JH (2008) Marker-assisted selection in plant breeding: From publications to practice. Crop Science 48:391-407

Zhong SQ, Dekkers JCM, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics 182:355-364

CHAPTER 3

GENOMIC SELECTION ACROSS ENVIRONMENTS FOR GRAIN QUALITY IN BIPARENTAL WHEAT POPULATIONS

ABSTRACT

Genomic selection (GS) is a promising tool for plant and animal breeding that uses genome-wide molecular marker data to capture small and large effect quantitative trait loci and predict the genetic value of selection candidates. GS has been shown previously to have higher prediction accuracies than conventional marker-assisted selection (MAS) for quantitative traits. Challenges in modeling genotype-by-environment interactions, however, reduce GS accuracies. In plant breeding, the ability to produce large numbers of progeny per cross and replicate them across many environments presents opportunities to meet these challenges. In this study, we compared phenotypic and marker-based prediction accuracy of genetic value for nine different grain quality traits within two biparental soft winter wheat (*Triticum aestivum* L.) populations. We used a cross-validation approach that trained and validated prediction accuracy across years to evaluate effects of model training population size, training population replication, and marker density in the presence of GxE. Results showed that prediction accuracy was significantly greater using GS versus MAS for all traits studied, and that accuracy for GS reached a plateau at low marker densities (128-256). Despite the moderate to high heritabilities of the traits studied, the average ratio of GS accuracy to phenotypic selection accuracy was 0.66, 0.54, and 0.42 for training population sizes of 96, 48, and 24, respectively. These results provide further empirical evidence that GS could produce greater genetic gain per unit time and cost than both phenotypic selection and conventional-MAS in plant

breeding with use of year-round nurseries and inexpensive, high-throughput genotyping technology.

INTRODUCTION

The use of molecular marker data to predict the genetic value of selection candidates is an important tool used in both plant and animal breeding programs. Effective marker-assisted selection (MAS) has largely been based on predictions derived from a few markers that are linked to large effect quantitative trait loci (QTL; Holland 2004). Genomic selection (GS) aims to improve MAS accuracy for quantitative traits by capturing both large and small QTL effects with genome-wide marker coverage (Meuwissen et al. 2001). In short, GS uses phenotypic and genotypic data from breeding lines, i.e. the training population (TP), to estimate marker effects that are then used to predict the genetic value of selection candidates having only genome-wide marker data, reviewed by (Heffner et al. 2009). With rapid reduction of high-throughput genotyping costs, GS is now being implemented widely in dairy cattle breeding (Hayes et al. 2009). The current status of GS in private sector plant breeding is not publically available; however, several simulation studies (Wong and Bernardo 2008; Bernardo and Yu 2007; Zhong et al. 2009; Heffner et al. 2010) and an empirical study in biparental maize (*Zea mays* L.), barley (*Hordeum vulgare* L.), and *Arabidopsis thailiana* (L.) populations suggest that GS will outperform previous MAS methods in plant breeding programs (Lorenzana and Bernardo 2009).

Unlike animal breeders, plant breeders have the ability to create large biparental populations that can be replicated within and across environments. Biparental populations also have extensive linkage disequilibrium (LD), allowing for

complete genome coverage with only a few hundred markers. These features enable "context-specific" MAS, where MAS is conducted within each cross by using genotypic and phenotypic information from target environments. Marker effects are thus relative to the genetic background and testing environments which improves prediction accuracy by minimizing error caused by epistasis and genotype by environment interaction (GxE; Podlich et al. 2004; Sebastian et al. 2010). Similarly, GS can be conducted within biparental populations, herein referred to as biparental-GS, where a subset of the progeny constitutes the TP to estimate marker effects. The resulting prediction models are then used for predicting genetic value of remaining progeny and/or for subsequent cycles of marker-assisted recurrent selection (MARS; Bernardo and Yu 2007).

A context specific approach can address GxE by growing TPs in target environments (Podlich et al. 2004; Sebastian et al. 2010); however, two major questions remain: 1) how many environments should be used for training, and 2) should TP lines be replicated within and across environments or be unreplicated and distributed among environments. To maximize GS accuracy, it seems optimal to use many training environments and the largest TP possible by not replicating lines. This is because the ability to capture QTL effects through LD is improved by using more individuals at the expense of replication (Knapp and Bridges 1990). This strategy may, however, not always be best as smaller, replicated TPs can result in higher heritability and better predictions with GS models that rely more on estimating genetic relationships rather than QTL effects (Zhong et al. 2009). The importance of estimating relationships or QTL effects will depend strongly on which method maximizes GS accuracy and the number of selection cycles that will occur between marker-effect re-estimation. This is due to the fact that accuracy from estimating

genetic relationships will deteriorate faster from recombination than will accuracy from estimating QTL effects if marker-QTL LD is strong (Zhong et al. 2009; Muir 2007; Habier et al. 2007). Consideration of these issues, along with the cost of genotyping many unreplicated lines and testing in many environments, will be important in implementing GS in plant breeding.

In any marker-based selection strategy, selection response should increase as heritability increases. But, scenarios of high heritability will also result in high phenotypic selection accuracy, resulting in little benefit of using MAS (Holland 2004; Hospital et al. 1997; Lande and Thompson 1990). MAS can, however, compare favorably to phenotypic selection for traits with high heritability if MAS cycles are shorter and less expensive than phenotypic selection cycles. In the case of oil palm (*Elaeis guineensis* Jacq.) breeding, GS with small TPs (*Ntp*=50) could reduce the selection cycle from 19 yr to 6 yr and increase gains from selection per unit time and cost (Wong and Bernardo 2008). This remarkable reduction in cycle time and cost favors marker-based prediction even if mediocre prediction accuracies result from using small TPs. Despite a less dramatic reduction in cycle time for field crops, Lorenzana and Bernardo (2009) calculated GS accuracies and suggested a MARS scheme using biparental-GS would approach 1.5 times more gain than phenotypic selection for maize and barley.

In hexaploid wheat (*Triticum aestivum* L.), there are several important grain quality traits that, despite being highly heritable, are strong targets for GS as they are polygenic (e.g. Munkvold et al. 2009; Smith 2008) and require significant resources for accurate phenotyping. One important grain quality trait is resistance to pre-harvest sprouting (PHS), the premature germination of seeds while still attached to the mother

plant. PHS prior to harvest causes breakdown of starch and decreases seed quality, test weight, and grain value. In addition to test weight, traits used to evaluate overall milling and baking quality include: flour yield, flour protein, softness, gluten strength, and water absorption. Reliable phenotyping methods have been developed for PHS (Anderson et al. 1993) and milling and baking quality (Guttieri et al. 2001; Guttieri and Souza 2003; Walker et al. 2008); nevertheless, these phenotypes are costly, time-consuming, and destructive, making early-generation testing of large populations difficult.

The objective of this research was to compare the accuracy of phenotypic and marker-based prediction of genetic value for nine different grain quality traits within two different biparental wheat populations. To meet this objective, a cross-validation approach that trained and validated prediction accuracy across years to evaluate selection strategies in the presence of GxE was used. Three marker-based prediction methods were tested to compare conventional-MAS, using multiple linear regression (*MLR*), and GS, using ridge regression (*RR*) and Bayes-Cπ (*BC*). Prediction accuracy of these methods was evaluated for three different training population sizes (*Ntp*= 24, 48, and 96) to determine accuracies possible for traits, such as wheat grain quality, that are expensive to phenotype and will thereby greatly limit *Ntp*. Finally, the effects of marker number (*Nm*), number of model training environments (*Nenv*), and replication of TP lines across environments on GS prediction accuracy were evaluated.

MATERIALS AND METHODS

Populations

Two doubled-haploid (DH) biparental hexaploid winter wheat populations were analyzed: 1) Cayuga x Caledonia (CC) and 2) Foster x KanQueen (FKQ). The CC population contained 209 soft white winter wheat lines and was previously used in a PHS QTL study by Munkvold et al. (2009). The female parent, a PHS susceptible line, Caledonia, is an off-type selection from Geneva (Sorrells et al. 2004). The male parent, a PHS resistant line, Cayuga, is derived from a Geneva backcross to a cross of Geneva and Clark's Cream (Sorrells and Anderson 1998). The FKQ population contained 174 soft red winter wheat lines differing for milling quality characteristics. The female parent, Foster, is an Agripro Company variety originating from Kentucky (VanSanford et al. 1997) and has very good milling quality, ranking 14th of 768 soft red wheat cultivars (Guttieri et al. 2008). The male parent, KanQueen, is a semi-hard red public variety originating from Kansas in 1949 (Bayles and Clark 1954). KanQueen has very poor milling quality ranking 764th of 768 soft red wheat cultivars (Guttieri et al. 2008).

Phenotypic Data

Data for nine quantitative traits were analyzed with seven milling and baking quality traits common to both populations, PHS only for CC, and test weight only for FKQ. PHS phenotyping was conducted as described by Anderson et al. (1993) and Munkvold et al. (2009). Harvested grain was tempered to 15% moisture and measured after milling on a modified Brabender Quadramat Junior mill as described by Finney

and Andrews (1986). The milling quality traits measured were flour yield (the percentage of flour obtained from milling), and softness (percentage of fine flour obtained i.e. that which can pass through a 94-mesh (180μm) screen). The two main components of baking quality, gluten strength and water absorption, were measured by flour protein concentration and four solvent retention capacity (SRC) tests. Flour protein was measured using a near infared analyzer (Unity Spectrastar 2200, Columbia MD). SRC was the measured as amount of solvent retained by the flour after centrifugation and draining. The four SRC solvents analyzed each predict different components of baking quality: water (H_2O-SRC) for global water absorption, sodium carbonate (NaCO-SRC) for damaged starch, sucrose (Suc-SRC) for arabinoxylan and partially hydrated gliadin content, and lactic acid (LA-SRC) for gluten strength. All milling and baking quality tests were done by the USDA-ARS Soft Wheat Quality Laboratory in Wooster, Ohio as described by Guttieri et al. (2008).

All phenotypic data for CC were collected from locations near Ithaca, NY, USA. Milling and baking quality phenotypes were collected on 50g samples from 1.26m x 3m, 6 row plots grown in three years (2005, 2006, and 2008) in one location each. PHS data were collected on samples from 1m rows in a randomized complete block design with two replications. Data were collected for six years (2001-2006) with two locations in 2002 and 2003 and three locations in 2001 and 2004-2006 for a total of 16 environments (Munkvold et al. 2009). Phenotypic data for FKQ was collected on 50g samples from 1.26m x 3m, six row plots for two years in Ithaca, NY, USA (2005, 2006) and for one year in Wooster, Ohio, USA (2006). All milling and baking and test weight data were collected on a single replicate and raw scores from each environment were used for the analysis. For the comparison of phenotypic accuracy to marker-based prediction, PHS was analyzed on a yearly basis using best linear

unbiased predictors (BLUPs) for each line in each year by fitting a random effects linear model in R (R Development Core Team 2009) that accounted for location, replicate, harvest date, and line effects. For the comparison of marker-based prediction accuracy when varying *Ntp*, replication, and number of environments, BLUPs for each line in each environment were calculated by fitting a random effects linear model in R that accounted for replicate, harvest date, and line effects.

Genotypic Data

The total number of markers available for CC was 484: 215 simple sequence repeats (SSRs), 147 Diversity Array Technology markers (DArT; Triticarte Pty. Ltd., Yaralumla, Australia), 72 amplified fragment length polymorphisms, 31 target region amplification polymorphisms, 16 restriction fragment length polymorphisms, three expressed sequence tag-SSRs, and one sequence tagged site (Munkvold et al. 2009). The FKQ was genotyped with 5,000 DArT markers (Triticarte Pty. Ltd., Yaralumla, Australia) of which 1481 were polymorphic. Marker sets were filtered to 399 markers for CC and 574 markers for FKQ by removing redundant or skewed markers (α=0.01). Linkage groups were determined by using the Map Manager QTXb20 computer program (Manly et al. 2001) using the Kosambi mapping function with a linkage threshold significance of α=0.001. Missing marker data was then imputed based on the observed multipoint marker data using the R/qtl package (Broman et al. 2003).

Marker Effect Estimation

Three methods were used to estimate marker effects: 1) multiple linear regression (*MLR*), 2) ridge-regression BLUP (*RR*), and 3) a Bayesian approach called BayesCπ (*BC*). Each of these methods was executed using R (R Development Core Team 2009).

Mutliple Linear Regression (*MLR*)

Multiple regression of trait values and marker alleles was conducted using a forward-backward variable selection approach where markers were modeled as fixed effects and significant markers were determined by forward (α=0.2) and backward (α=0.2) selection. Relaxed significance thresholds were used to achieve higher selection responses than those found using more stringent thresholds (e.g. Hospital et al. 1997; Lorenzana and Bernardo 2009). Regression coefficients of markers included in the final model were used as marker effects to predict the GEBV of each selection candidate.

Ridge-Regression Best Linear Unbiased Prediction (*RR*)

A *RR* model was used to simultaneously estimate marker effects through modeling markers as random effects with a common variance (Meuwissen et al. 2001; Whittaker et al. 2000). The *RR* model thereby shrinks each marker effect equally toward zero, but does allow for markers to have unequal effects. Goddard (2009) and Piepho (2009) showed that *RR* is equivalent to a model where a realized-relationship matrix is determined from marker information in order to estimate marker effects

(Habier et al. 2007; VanRaden 2008). Variance components to solve mixed-model equations (Henderson 1984) and the additive realized-relationship matrix where calculated using R package 'emma' (Kang et al. 2008).

Bayesian Estimation: BayesCπ (*BC*)

To avoid the presumably incorrect assumption of equal marker variances, over-shrinking of large effect loci, and not allowing markers to have zero effects, several Bayesian models have been proposed (Gianola et al. 2009). In Bayesian GS models that allow for markers with no effect, if the proportion of markers with zero effect (π) is assumed known, an incorrect π can negatively affect prediction accuracy (Verbyla et al. 2010; Gianola et al. 2009). Therefore, we used *BC*, which is an extension of the BayesC (Kizilkaya et al. 2010), that jointly estimates π from the training data (Dekkers et al. 2009, Jannink 2010). Like BayesB (Meuwissen et al. 2001), the *BC* method allows for markers to have no effect; however, markers that are included in the model are assumed to have a common variance (Kizilkaya et al. 2010). We adapted *BC* code written by R.L. Fernando (Dekkers et al. 2009), and for each analysis we used starting π parameter of 0.5 and 2,000 iterations with 1,000 burn-in iterations, which was sufficient to reach approximate convergence (stabilization of π) for each analysis.

Prediction Accuracy and Cross-validation

For each validation line GEBV was calculated as $y_i = X_i g$: where y_i was the validation line phenotype, X_i was the vector of the marker scores for that line, and g was the vector of marker effects obtained from TP using *MLR*, *RR*, or *BC*. Prediction

accuracy (*r*) was calculated for each model as the correlation of the GEBV and the "true" genetic value (TGV) of the selection candidate, divided by the square root of the broad-sense heritability (H) of the TGV on a progeny-mean basis (*r*=cor(GEBV:TGV)/H). The TGV was determined by calculating the BLUP for each selection candidate across all years not used in the TP by fitting a random effects linear model in R that accounted for year and line effects. The correction factor, H, was used to account for the estimation error of the TGV (Dekkers 2007). For comparison, phenotypic accuracy (*rP*) was calculated similarly, but the GEBV was replaced with a phenotypic estimated genetic value (PEGV): the observed phenotype of a selection candidate in the environments used for training the model. Thus, PEGV is composed of both additive and non-additive effects that can contribute to phenotypic *r*; whereas, the marker-based prediction models used in this study only capture additive effects, i.e. breeding value.

The impact of *Ntp* on *r* was investigated by using *Ntp*=24, 48, and 96 to correspond with phenotyping limitations of grain quality and current 96 or 384-well DNA sample plates. Therefore, the validation population (VP) size was the total population size minus *Ntp*. To avoid bias introduced by genotype by environment interactions (GxE), cross-validation was done across environments i.e., training data came from a single year and validation data came from all other years. FKQ data collected from NY and OH in 2006 were considered unique environments for cross-validation, as this should not significantly bias accuracy because the correlation between NY 2006 and OH 2006 (0.78) was similar to that of NY 2005 and NY 2006 (0.75).

The impact of *Nm* on *r* was investigated by using the maximum *Nm* available (CC=399; FKQ=574) and four subsets of *Nm*=64, 128, 256, and 384 to correspond with current custom high-throughput genotyping capabilities. Marker subsets were created using K-means clustering (Hartigan and Wong 1979) in R where each marker was treated as an explanatory variable, the number of clusters equaled the *Nm*, the number of random starts equaled 1000, and the marker closest to the centroid of each cluster was chosen. This procedure was considered important to select informative marker subsets by minimizing LD between selected markers and maximizing genome marker coverage.

For cross-validation described above, 30 TPs were randomly selected for each *Ntp* and PEGV was calculated for each. GEBV accuracy was also determined for all method-*Nm*-*Ntp* combinations. Therefore, each marker-based prediction method was used for 900 analyses for test weight (FKQ) and all milling and baking quality traits (FKQ and CC) and for 2700 analyses for PHS (CC). The reported *r* for phenotypic and marker-based selection was that average *r* for all 30 TPs, and prediction methods were compared using a paired *t*-test (α=0.05) across the 30 TPs.

To investigate the effect of the replication of TP lines across environments on *r*, the CC-PHS dataset was analyzed as it contained a large number of environments. As the number of locations per year in this dataset was unbalanced, an equal number of environments for both the training and validation was achieved by dividing the dataset into two "year groups" (2001, 2003, and 2005; 2002, 2004, 2006). This odd-even year grouping should still represent a random sample, as the year number should be not predictive of the overall environmental conditions and GxE. We assumed a maximum of 96 field plots and two different scenarios were tested: 1) 96 unreplicated

lines could all be grown in the same environment, or lines could be replicated across environments, i.e. 48, 24, 16, or 12 lines could be replicated across two, four, six, or eight environments, respectively, and 2) 96 unreplicated lines could all be grown in the same environment or, while still being unreplicated, lines could be split evenly across two, four, six, or eight environments. For each scenario, 30 TPs and eight TP-environment combinations were randomly selected. Calculation of *r* was done as previously described where the TGV of a selection candidate was a BLUP calculated using data from all eight environments in the validation data. This was repeated for both of the "year groups" and for optimal marker number for each prediction method (*Nm*=64 for *MLR* and *Nm*=256 for *RR* and *BC*) that was determined by the other analyses conducted in this study.

RESULTS

Phenotypic and Marker-based Prediction

Marker-based prediction accuracies (rM) for all methods were greatest for the largest TP used (TP=96), with accuracy of *RR* (*rRR*) and *BC* (*rBC*) being greater than the accuracy of *MLR* (*rMLR*) for all traits across both populations (Table 3.1). The mean *rRR* (0.52) and mean *rBC* (0.53) were more than 1.4 times greater than the mean *rMLR* (0.36). For traits shared by both populations, FKQ had greater rM than CC for all three methods. In CC, the mean *rRR* (0.49) was greater than the mean *rBC* (0.47) with *rBC* being significantly greater than *rRR* only once (softness). In contrast, FKQ's mean *rRR* (0.53) was less than its mean *rBC* (0.58), with *rRR* never being significantly greater than *rBC*.

Table 3.1 Phenotypic and marker-based prediction accuracy for each trait and population (pop). Marker-based prediction was based on *Ntp*=96 and *Nm* that lead to highest accuracy for each trait-population-method combination

Trait	Pop	Vg	H	rP	rMLR	rRR	rBC	rMLR/rP	rRR/rP	rBC/rP
PHS	CC	0.41	0.95	0.82[a]	0.50[b]	0.68[c]	0.67[d]	0.61	0.84	0.82
Test weight	FKQ	1.41	0.76	0.85[a]	0.38[b]	0.50[c]	0.49[d]	0.45	0.58	0.57
Flour yield	CC	0.62	0.80	0.66[a]	0.37[b]	0.56[c]	0.51[d]	0.56	0.84	0.77
	FKQ	2.17	0.95	0.91[a]	0.39[b]	0.56[c]	0.60[d]	0.43	0.61	0.66
Flour protein	CC	0.09	0.75	0.60[a]	0.29[b]	0.39[c]	0.38[c]	0.49	0.66	0.64
	FKQ	0.49	0.93	0.89[a]	0.36[b]	0.48[c]	0.52[d]	0.41	0.54	0.59
Softness	CC	6.17	0.92	0.85[a]	0.18[b]	0.27[c]	0.31[d]	0.22	0.32	0.37
	FKQ	11.69	0.94	0.89[a]	0.27[b]	0.37[c]	0.43[d]	0.31	0.41	0.49
LA-SRC	CC	26.45	0.86	0.80[a]	0.51[b]	0.64[c]	0.62[c]	0.65	0.81	0.78
	FKQ	52.07	0.90	0.87[a]	0.36[b]	0.54[c]	0.58[d]	0.41	0.62	0.66
NaCo-SRC	CC	2.45	0.89	0.78[a]	0.24[b]	0.42[c]	0.39[d]	0.30	0.54	0.50
	FKQ	12.59	0.94	0.94[a]	0.42[b]	0.62[c]	0.74[d]	0.45	0.66	0.79
Suc-SRC	CC	1.51	0.67	0.51[a]	0.17[b]	0.41[c]	0.38[c]	0.34	0.81	0.74
	FKQ	17.90	0.91	0.87[a]	0.33[b]	0.54[c]	0.60[d]	0.38	0.62	0.69
H20-SRC	CC	0.48	0.67	0.63[a]	0.27[b]	0.53[c]	0.46[d]	0.42	0.84	0.73
	FKQ	3.08	0.92	0.87[a]	0.48[b]	0.63[c]	0.64[c]	0.56	0.73	0.74
Mean			**0.86**	**0.80**	**0.36**	**0.52**	**0.53**	**0.45**	**0.66**	**0.66**

[abcd] Within each population and trait, prediction accuracies that share the same letter were not significantly different at $\alpha = 0.05$

The phenotypic prediction accuracy (*rP*) was significantly greater for all traits than rMs for both biparental wheat populations (Table 3.1). The mean *rP* across all traits and populations was 0.80, with a maximum *rP* of 0.94 (FKQ:NaCO-SRC) and a minimum *rP* of 0.51 (CC:Suc-SRC). For the seven milling and baking quality traits

shared by both populations, the mean *rP* was 0.79 with FKQ (mean *rP*=0.89) having a 1.3 times greater *rP* than CC (mean *rP*=0.69). Accordingly, the genetic variance component (*Vg*) and the H^2 for each trait were also greater for FKQ than for CC (Table 3.1).

The rM to *rP* ratio (rM/*rP*) across all traits (*Ntp*=96) for both *RR* and *BC* was 0.66, which was 1.47 times greater than *MLR* (0.45; Table 3.1). When using the best marker-based prediction method for each trait-population combination, the mean rM/*rP* among shared traits was greater for CC (0.70) than for FKQ (0.66). The highest rM/*rP* for CC was 0.84 (PHS, flour yield, and H_2O-SRC) and for FKQ was 0.79 (Suc-SRC). The lowest rM/*rP* was for softness in both CC (0.37) and FKQ (0.49).

Prediction Accuracy vs. Training Population Size and Marker Number

Reducing the *Ntp* used to predict TGV had a large negative effect on rM (Fig. 3.1, Supplementary Table 3.1 and 3.2). The mean rM across all traits and methods was 0.30, 0.42, and 0.59 for *Ntp* of 24, 48, and 96, respectively. The reduction of rM with *Ntp* was less severe for *RR* than the other methods. While *RR* (0.52) and *BC* (0.53) had similar mean rMs at *Ntp*=96, when *Ntp* was reduced to 48 and 24, *rRR* decreased by 17% and 36% whereas *rBC* decreased by 33% and 61%, respectively. The reduction of *rMLR* was largest for *Ntp*=48 (39%) and was similar to *rBC* for *Ntp*=24 (59%).

The mean rM across all trait-population-*Ntp* combinations was highest for *Nm*=256 for both *RR* and *BC* (Fig. 3.2). In contrast, *MLR* showed the highest rM when *Nm* was smaller than *Ntp*, i.e. *Nm*=64 and *Ntp*=96. For all methods, rM was

significantly different between *Nm*=64 and 256, but *Nm*=64 was not significantly different from *Nm*=128, and *Nm*=256 was not significantly different from 128 and 384. These results suggest rM reached a plateau between *Nm* =128 and 256 for *RR* and *BC*, and a maximum *rMLR* was achieved when *Ntp* was greater than *Nm*. As *Ntp* decreased, *rRR* and *rBC* reached plateaus at smaller *Nm*, with maximum rM achieved with $Nm \leq 128$ markers for *Ntp*=24.

Number of Environments and Replications used for Marker-based Prediction

With the limit of possible field plots set to 96, decreasing *Ntp* allowed for increased replication of each TP line across environments; however, reducing *Ntp* had an overall negative effect on rM for PHS in CC (Fig. 3.3 and Supplementary Table 3.3). *MLR* showed a consistently significant decrease in rM as *Ntp* decreased, except for *Ntp*= 16 and 12. For *RR* and *BC*, there was no significant difference in rM between *Ntp*=96 and 48 or *Ntp*=16 and 12. In the contrasting scenario where 96 TP lines were unreplicated and distributed evenly across one, two, four, six, or eight environments, the overall differences rM were negligible for each prediction method (Supplementary Table 3.4). For *rRR* and *rBC*, TPs with two, four, and six environments ($rM \approx 0.60$) were significantly higher than one or eight environments ($rM \approx 0.59$); whereas, none of the scenarios were significantly different for *MLR*. There were also small differences in the standard error (SE) of rM, with training in a single environment for each method being having the highest standard error for each method tested Supplementary Table 3.4).

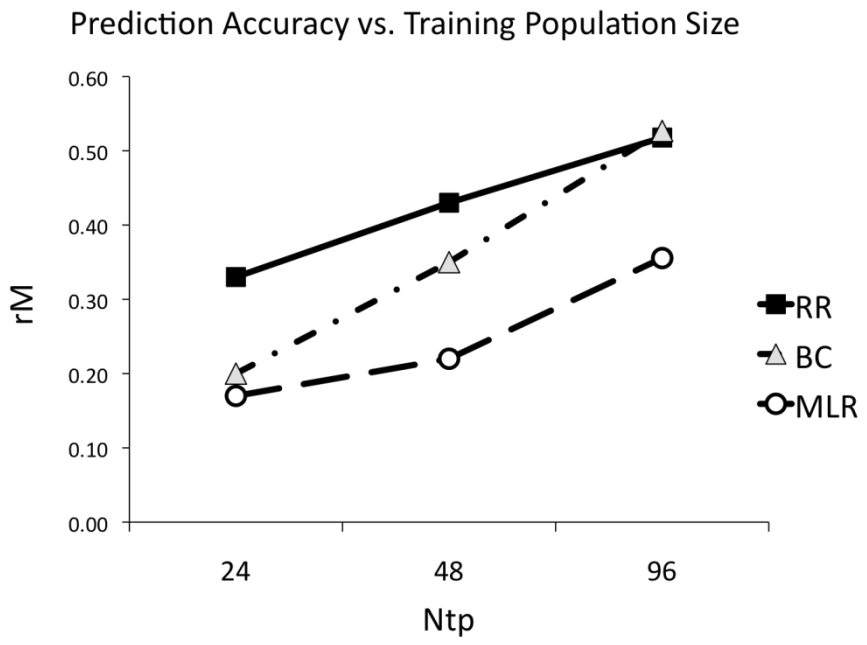


Figure 3.1 Effect of training population size (*Ntp*) on mean marker-based prediction accuracy (rM) for all trait-population-optimal marker set combinations

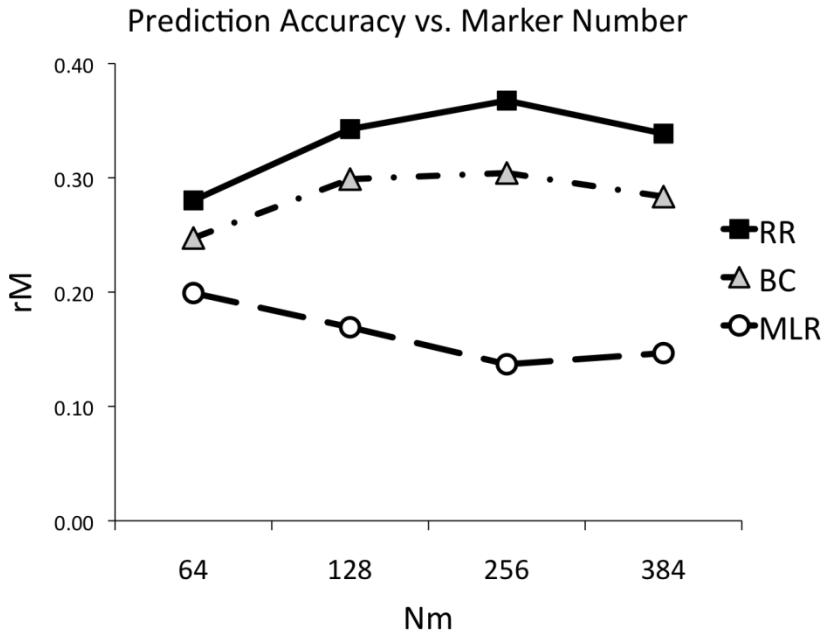


Figure 3.2 Effect of marker number (*Nm*) on the mean marker-based prediction accuracy (rM) for all trait-population-training population size (*Ntp*) combinations

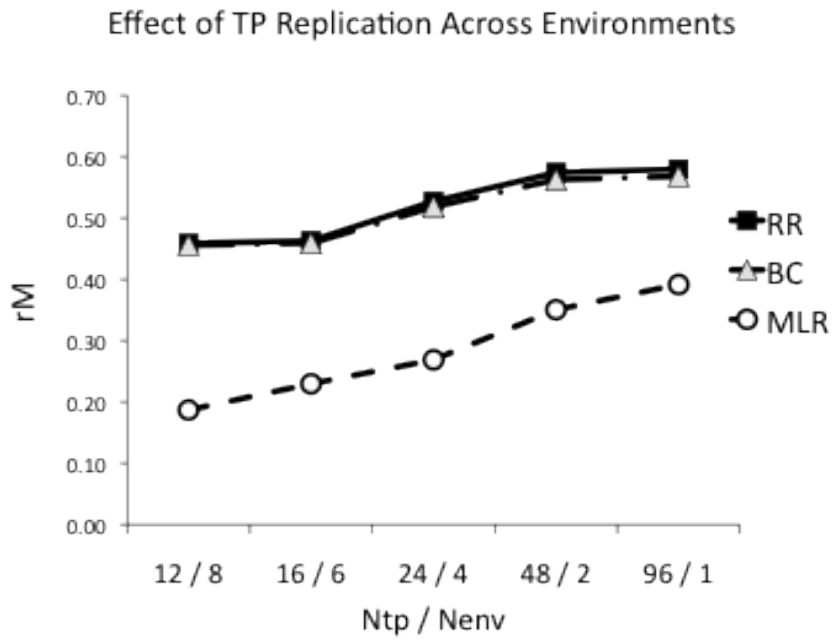


Figure 3.3 Effect of training population size (*Ntp*) and replication across different numbers of environments (*Nenv*) on mean marker-based prediction accuracy (rM)

Supplementary Table 3.1 Phenotypic and marker-based prediction accuracy for each trait and population (pop). Marker-based prediction was based on *Ntp*=48 and *Nm* that lead to highest accuracy for each trait-population-method combination

Trait	Pop	Vg	H	rP	rMLR	rRR	rBC	rMLR/rP	rRR/rP	rBC/rP
PHS	CC	0.41	0.95	0.81[a]	0.32[b]	0.60[c]	0.48[d]	0.39	0.74	0.59
Test weight	FKQ	1.41	0.75	0.85[a]	0.22[b]	0.41[c]	0.40[c]	0.26	0.49	0.47
Flour yield	CC	0.62	0.80	0.66[a]	0.22[b]	0.46[c]	0.33[d]	0.33	0.71	0.51
	FKQ	2.17	0.95	0.91[a]	0.24[b]	0.47[c]	0.38[d]	0.26	0.52	0.42
Flour protein	CC	0.09	0.75	0.59[a]	0.17[b]	0.30[c]	0.25[d]	0.29	0.52	0.42
	FKQ	0.49	0.93	0.88[a]	0.18[b]	0.36[c]	0.36[c]	0.20	0.41	0.40
Softness	CC	6.17	0.92	0.85[a]	0.13[b]	0.18[c]	0.16[bc]	0.15	0.21	0.19
	FKQ	11.69	0.93	0.89[a]	0.15[b]	0.24[c]	0.25[c]	0.16	0.27	0.28
LA-SRC	CC	26.45	0.86	0.79[a]	0.30[b]	0.56[c]	0.40[d]	0.38	0.70	0.51
	FKQ	52.07	0.90	0.87[a]	0.25[b]	0.48[c]	0.43[d]	0.28	0.55	0.49
NaCo-SRC	CC	2.45	0.89	0.79[a]	0.14[b]	0.33[c]	0.24[d]	0.18	0.42	0.30
	FKQ	12.59	0.94	0.94[a]	0.30[b]	0.53[c]	0.50[d]	0.32	0.56	0.53
Suc-SRC	CC	1.51	0.68	0.50[a]	0.12[b]	0.33[c]	0.20[d]	0.23	0.65	0.39
	FKQ	17.90	0.92	0.87[a]	0.22[b]	0.42[c]	0.39[c]	0.25	0.48	0.44
H20-SRC	CC	0.48	0.68	0.62[a]	0.18[b]	0.43[c]	0.26[d]	0.29	0.70	0.42
	FKQ	3.08	0.92	0.86[a]	0.27[b]	0.56[c]	0.47[d]	0.31	0.64	0.55
Mean			**0.86**	**0.80**	**0.22**	**0.43**	**0.35**	**0.27**	**0.54**	**0.44**

[abcd] Within each population and trait, prediction accuracies that share the same letter were not significantly different at $\alpha = 0.05$

Supplementary Table 3.2 Phenotypic and marker-based prediction accuracy for each trait and population (pop). Marker-based prediction was based on *Ntp*=24 and *Nm* that lead to highest accuracy for each trait-population-method combination

Trait	Pop	Vg	H	rP	rMLR	rRR	rBC	rMLR/rP	rRR/rP	rBC/rP
PHS	CC	0.41	0.95	0.81[a]	0.26[b]	0.51[c]	0.30[d]	0.32	0.62	0.37
Test weight	FKQ	1.41	0.75	0.85[a]	0.15[b]	0.29[c]	0.20[b]	0.18	0.34	0.24
Flour yield	CC	0.62	0.80	0.65[a]	0.17[b]	0.39[c]	0.21[b]	0.26	0.60	0.32
	FKQ	2.17	0.95	0.91[a]	0.18[b]	0.38[c]	0.24[d]	0.20	0.41	0.27
Flour protein	CC	0.09	0.75	0.58[a]	0.12[b]	0.23[c]	0.14[b]	0.21	0.40	0.24
	FKQ	0.49	0.93	0.89[a]	0.14[b]	0.24[c]	0.17[b]	0.16	0.27	0.20
Softness	CC	6.17	0.92	0.85[a]	0.06[b]	0.12[c]	0.10[c]	0.07	0.14	0.12
	FKQ	11.69	0.94	0.89[a]	0.12[b]	0.18[c]	0.13[b]	0.13	0.20	0.15
LA-SRC	CC	26.45	0.86	0.80[a]	0.22[b]	0.46[c]	0.27[b]	0.28	0.57	0.34
	FKQ	52.07	0.90	0.87[a]	0.17[b]	0.41[c]	0.25[d]	0.20	0.47	0.29
NaCo-SRC	CC	2.45	0.89	0.78[a]	0.11[b]	0.26[c]	0.14[b]	0.14	0.34	0.18
	FKQ	12.59	0.94	0.94[a]	0.25[b]	0.39[c]	0.26[b]	0.27	0.41	0.27
Suc-SRC	CC	1.51	0.68	0.50[a]	0.08[b]	0.23[c]	0.12[b]	0.16	0.45	0.23
	FKQ	17.90	0.92	0.87[a]	0.18[b]	0.32[c]	0.20[b]	0.21	0.36	0.23
H20-SRC	CC	0.48	0.68	0.62[a]	0.14[b]	0.33[c]	0.17[b]	0.22	0.53	0.28
	FKQ	3.08	0.92	0.86[a]	0.22[b]	0.46[c]	0.28[d]	0.25	0.54	0.32
Mean			**0.86**	**0.80**	**0.17**	**0.33**	**0.20**	**0.21**	**0.42**	**0.26**

[abcd] Within each population and trait, prediction accuracies that share the same letter were not significantly different at $\alpha = 0.05$

Supplementary Table 3.3 Effect of *Ntp* and replication across different numbers of environments (*Nenv*) on mean marker-based prediction accuracy (*r*) and standard error (SE)

Ntp-Nenv	rMLR	rMLR-SE	rRR	rRR-SE	rBC	rBC-SE
96-1	0.39[a]	0.0062	0.58[a]	0.0037	0.57[a]	0.0038
48-2	0.35[b]	0.0056	0.57[a]	0.0035	0.56[a]	0.0038
24-4	0.27[c]	0.0065	0.53[b]	0.0044	0.52[b]	0.0046
16-6	0.23[d]	0.0062	0.46[c]	0.0055	0.46[c]	0.0056
12-8	0.19[d]	0.0223	0.46[c]	0.0175	0.46[c]	0.0177

[abcd] Within each method, accuracies that share the same letter were not significantly different at $\alpha = 0.05$.

Supplementary Table 3.4 Effect of distributing training population lines across different numbers of environments (*Nenv*) on marker-based prediction accuracy (*r*) and standard error (SE)

Nenv	rMLR	rMLR-SE	rRR	rRR-SE	rBC	rBC-SE
1	0.39[a]	0.0064	0.59[a]	0.0037	0.58[a]	0.0038
2	0.39[a]	0.0056	0.60[b]	0.0033	0.59[b]	0.0035
4	0.38[a]	0.0058	0.60[b]	0.0033	0.58[ac]	0.0034
6	0.37[a]	0.0059	0.60[b]	0.0035	0.59[bc]	0.0037
8	0.37[a]	0.0056	0.59[a]	0.0035	0.58[a]	0.0037

[abcd] Within each method, accuracies that share the same letter were not significantly different at $\alpha = 0.05$.

Supplementary Table 3.5 Restricted maximum likelihood estimate (REML) of variance components for PHS in CC

Random Effect	Variance Ratio	Variance Component	Percent of Total
Environment (E)	0.604	0.606	21.138
Genotype (G)	0.695	0.698	24.337
Replicate	0.070	0.071	2.459
Harvest Date	0.320	0.321	11.203
GxE	0.168	0.168	5.867
Error		1.004	34.995
Total			100

DISCUSSION

Marker-based Prediction Accuracy

Marker-based prediction accuracy using GS was clearly superior to using conventional-MAS for all levels of *Ntp* and *Nm* for each grain quality trait studied. The observed advantage of using a random effects approach (*RR* and *BC*) versus a fixed effects approach (*MLR*) for situations of large *Nm* and small *Ntp* is consistent with results found for other biparental populations in simulation (Wong and Bernardo 2008; Bernardo and Yu 2007; Piyasatian et al. 2007) and empirically (Lorenzana and Bernardo 2009).

Despite the benefit of a low marker density being adequate to cover the genome in a biparental-GS approach, the issue of small *Ntp* and large *Nm* will still be present due to the practical limitations of *Ntp* when having a separate TP for each

biparental population. This will be especially true for traits, like wheat grain quality, that are expensive to phenotype. Accordingly, we evaluated rM for small Ntps (96, 48, and 24), and, as expected, rM decreased as *Ntp* decreased. Notably, *RR* showed considerably less reduction in rM than *MLR* (Lorenzana and Bernardo 2009) and *BC*. This indicates that predictions based on marker-based relationships are less affected by the lack of statistical power to estimate specific marker effects caused by small *Ntp* and large *Nm* (Zhong et al. 2009). The rapidly increasing trend of the *rBC* with increase of *Ntp* (Fig. 3.1) suggests that *BC* may have outperformed *RR* if larger Ntps were used, but it is unlikely that significantly larger TPs will be feasible for each biparental population in most breeding programs.

Bayesian models have outperformed *RR* previously; however, only small differences have been reported between them for polygenic traits (Hayes et al. 2009; Zhong et al. 2009; VanRaden et al. 2009; Lorenzana and Bernardo 2009). Despite *BC* having a more realistic assumption of at least some markers having zero effect, the average accuracies across both populations also showed little difference between *RR* and *BC* when *Ntp*=96 (Fig. 3.1). Interestingly, this trend did not hold true when looking at the milling and baking quality traits for each population independently. For FKQ, *BC* was generally more accurate than *RR*, but for CC, *RR* was significantly more accurate for four traits and *BC* was significantly more accurate only for softness (*Ntp*=96; Fig. 3.1).

The advantage of *BC* over *RR* for FKQ, but not for CC, was likely influenced by size of the marker effects present in each population. FKQ is a typical population used for biparental QTL mapping: the two parents were chosen for large phenotypic differences to increase genetic variance for the trait of interest, and thereby increase

the magnitude of QTL effects and the power to detect them. In contrast, CC was made from a cross between two elite parents with good milling and baking characteristics resulting in a population with a smaller genetic variance for those quality traits. As *BC* allows for markers with no effect, the markers that remain in the model can have a greater contribution to the predicted breeding values than would be possible using *RR*. Therefore, it was not surprising that *BC* performed better than *RR* in FKQ, where larger QTL effects would be expected. In practice, most breeding crosses will be made between elite material and, therefore, be more similar to CC than FKQ. Consequently, as seen in CC and previously mentioned empirical studies, *RR* will likely be comparable or even better than Bayesian models for highly polygenic traits in biparental-GS because power for QTL detection and effect estimation will be restricted by limited genetic variance and small Ntps.

Prediction Accuracy in the Presence of Genotype by Environment Interaction

All cross-validation procedures were performed such that training and validation data came from distinct environments to attain prediction accuracies that were not inflated by confounding GxE. In addition to the cross-validation across environments, we investigated both the effects of replicating lines across environments and distributing unreplicated lines among environments. Reducing *Ntp* through replicating TP lines across environments had a negative effect on rM (Fig. 3.3 and Supplementary Table 3.3). An *Ntp* less than 48 showed significantly reduced rM, that was likely due to adverse effects of small *Ntp* (as seen in Fig. 3.1); however, a TP with *Ntp*=48 grown in two environments was not significantly different for *RR* and *BC* than a TP of *Ntp*=96 grown in a single environment. This suggests that increasing replication at the expense of *Ntp* could be beneficial in cases where costs of

genotyping and generating TP lines are greater than increasing seed and replicating TP lines across environments. Increased replication by reducing *Ntp* may, however, have further effects on rM in a recurrent GS scheme as selection candidates will become less related to the TP with each cycle of selection. That is, prediction accuracy achieved using smaller, replicated TPs will likely decrease more each generation because accuracy will largely depend on estimating genetic relationships that rapidly breakdown with recombination (Zhong et al. 2009).

Distributing the TP across more than one environment provided only negligible improvements in rM over training all TP lines tested in a single environment despite previously reported QTL by environment interactions (Munkvold et al. 2009) and GxE explaining 6% of the phenotypic variation for PHS in CC (Supplementary Table 3.4 and 3.5). This is not necessarily surprising as we compared each scenario by its average rM across all environmental combinations. Therefore, a more informative statistic should be rM standard error, and while differences were small, rM standard error was highest for all methods when only one environment was used to train the model (Supplementary Table 3.4). This result supports the intuition that rM stability should increase by training with more environments as multi-environment training has the advantage of spreading the risks of unpredictable weather conditions that can lead to poor data quality, or, in some cases, no phenotypic variance (e.g. complete lack of or very extreme incidence of disease, lodging, or drought). In addition, traits that exhibit greater GxE than was observed for PHS in CC may increase the value of distributing TPs across multiple environments to capture GxE. Clearly, GxE will differ for each breeding program, population, and trait. Thus, more research is needed to be able to predict the best allocation of TP resources across

target environments in order to achieve maximum rM while also considering the cost tradeoff of training with more environments.

Marker-based vs. Phenotypic Selection per Unit Time and Cost

The accuracy of predicting genetic value was significantly greater using phenotypic data than marker data, regardless of marker-based prediction method (Fig. 3.1, Supplementary Table 3.1 and 3.2). The overall high *rP* was not surprising as PHS has been shown to have a moderate heritability (H^2=0.44; Munkvold et al 2009) and the other eight quality traits have all been shown to have high heritability ($H^2 > 0.70$; Huang et al 2006; Smith 2008). High H^2 should also translate into high rM; however, statistical power to estimate marker effects is also heavily influenced by *Ntp* and the number of QTL controlling the trait (Beavis 1998). Therefore, the inferiority of marker-based prediction to phenotypic selection was expected as all traits studied were polygenic (Munkvold et al. 2009; Smith, 2008; Huang et al 2006;) with medium to high H^2 and *Ntp* was limited to 96 individuals to represent a feasible maximum *Ntp* for each cross in a wheat breeding program. Furthermore, phenotypic prediction captures both additive and non-additive effects; whereas, the GS and *MLR* models used only capture additive effects, i.e. breeding value. So, while appropriate for assessing genetic value prediction, *rP* may be an inflated estimate in terms of breeding value prediction accuracy in cases where non-additive effects are present.

Even with lower accuracies from marker-based than phenotypic selection, greenhouses, off-season nurseries, and low-cost genotyping can allow MAS or GS to outperform phenotypic selection on a gain per unit time and cost basis (e.g. Bernardo and Yu 2007; Hospital et al. 1997; Wong and Bernardo 2008; Heffner et al 2010). In

barley and maize biparental populations, Lorenzana and Bernardo (2009) showed that rM was generally at least half that of *rP* for all traits studied. Assuming the possibility of growing three generations per year, Lorenzana and Bernardo (2009) concluded the annual gain from GS would approach 1.5 times that of phenotypic selection for maize and barley biparental-GS. The prediction accuracies for the grain quality traits across environments achieved in this study are consistent with Lorenzana and Bernardo (2009), with the average *rRR*/*rP* for *Ntp*=96, *Ntp*=48, and *Ntp*=24 equaling 0.64, 0.54, and 0.42, respectively. Using their same rough approximation, with only two generations per year for winter wheat, our results suggest that GS with *Ntp*>48 would outperform phenotypic selection for wheat grain quality traits. It should be noted, however, that each GS cycle without marker effect re-estimation will result in decreases in rM from changes in marker effects, gene frequency, and QTL-marker LD with each cycle of selection (Bernardo and Yu 2007; Muir 2007).

In addition to enabling more cycles per year, marker-based selection can raise selection intensity by increasing the number of selection candidates. This is possible when high-throughput genotyping is cheaper than phenotypic selection. Considering the nine grain-quality traits we analyzed, cost of inbreeding a line, increasing seed, growing field plots, and phenotyping (~US$60) is at least 3 times the cost of genome-wide marker coverage on a single plant (~US$20-US$25; cost of 384 SNP genotyping; S. Chao and S. McCouch, pers. comm.). Of course, the training cycle will be more expensive per line than phenotypic selection alone; therefore, *Ntp* size will need to be balanced with population sizes and genotyping costs of subsequent GS cycles.

Future Considerations for Biparental-GS Approaches

The interest in MAS has largely been centered on its ability to decrease the length of the selection cycle; therefore, a major caveat of biparental-GS is that it requires phenotyped TP lines from each cross prior to conducting GS. Even in the case of maize, where DH lines can be created in a year with the use of winter nurseries, it would be at least two years before GS could be implemented for each inbred cross (Bernardo and Yu 2007). In contrast, a multi-family-GS approach, as used in cattle (e.g. Hayes et al. 2009), uses predictions generated from a TP comprised of advanced breeding lines from many families that have already gone through the breeding program. This would eliminate the need to wait for phenotypes from a new cross, thereby facilitating immediate application of GS to newly generated lines and populations and a further reduction of cycle time in plant breeding (Heffner et al. 2010).

The application of MAS strategies within each biparental cross will allow inexpensive genome-wide genotyping, as LD will be extensive. However, it is unlikely that genome-wide genotyping will be a major limitation with the steady advancements in high-throughput genotyping (e.g. Deschamps and Campbell 2010). Also, biparental-GS is a population-specific approach, which is useful in attaining accurate marker estimates as the confounding effects of genetic background, and rare allele frequencies are avoided (Podlich et al. 2004; Sebastian et al. 2010). But, regardless of the number of environments used, a single season of phenotyping for model training, as is typical of MARS (e.g. Bernardo and Yu 2007), could lead to inaccurate allele effect estimates if GxE is largely due to genotype by year effects as is common in many regions. Increasing the number of seasons of training for each cross

maybe advantageous, but this would increase the length of the training cycle. A multi-family-GS approach that utilizes data from many biparental crosses that extend over environments and years (Heffner et al, 2010) may therefore be more attractive for capturing GxE as allelic effects would be more robust across time and space.

Conclusion

Marker-based prediction accuracy achieved using GS was clearly superior to conventional-MAS for the nine wheat grain quality traits investigated in this study. The observed prediction accuracies, coupled with the ability to reduce breeding costs, shorten selection cycles, and increase selection intensity, support the use of GS for many traits, including high-heritability traits where phenotypic selection is already effective. Looking forward, comparisons between biparental-GS and multi-family-GS approaches will be important for making decisions on how to best implement GS in plant breeding and maximize gains from selection per unit time and cost.

Acknowledgements

The authors thank USDA/ARS Soft Wheat Quality Laboratory in Wooster, Ohio for their careful evaluation of milling and baking quality that provided accurate data used for this study. Support for the work of Elliot Heffner was provided by USDA National Needs Graduate Fellowship Competitive Grant No. 2005-38420-15785 from the National Institute of Food and Agriculture. Support for the work Jean-Luc Jannink was provided by USDA-NIFA grant No. 2009-85606-05701 and No. 2009-65300-05661. Additional funding for this research was provided by USDA – NIFA National Research Initiative CAP grant No. 2005-05130 and by Hatch 149-402.

REFERENCES

Anderson JA, Sorrells ME, Tanksley SD (1993) RFLP analysis of genomic regions associated with resistance to preharvest sprouting in wheat. 33:453-459

Bayles BB, Clark JA (1954) Classification of wheat varieties grown in the United States in 1949. Washington, U.S. Dept. of Agriculture, pp 173

Beavis WD (1998) QTL analyses: power, precision, and accuracy. In: Paterson AH (eds) Molecular Dissection of Complex Traits. Boca Raton, FL, CRC Press, pp 145–162

Bernardo R, Yu J (2007) Prospects for Genomewide Selection for Quantitative Traits in Maize. Crop Sci 47:1082

Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. 19:889-890

Dekkers JCM (2007) Prediction of response to marker-assisted and genomic selection using selection index theory. J Anim Breed Genet 124:331-341

Dekkers JCM, Garrick DJ, Fernando RL (2009) Use of high-density snp genotyping for genetic improvement of livestock. A short course organized by the Animal Breeding & Genetics Department of Animal Science Iowa State University June 1-10. http://www.ans.iastate.edu/stud/courses/short/2009/. Verified 17 June 2010

Deschamps S, Campbell MA (2010) Utilization of next-generation sequencing platforms in plant genomics and genetic variant discovery. Mol Breed 25:553-570

Edwards M, Johnson L (1994) RFLPs for rapid recurrent selection.p. 33–40. *In* Analysis of molecular marker data. Joint Plant Breeding Symposia Series. ASA, Madison, WI

Finney PL, Andrews LC (1986) Revised microtesting for soft wheat quality evaluations. Cereal Chem 63:177-182

Gianola D, De Los Campos G, Hill WG, Manfredi E, Fernando R (2009) Additive genetic variability and the Bayesian alphabet. Genetics 183:347

Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136:245-257

Guttieri MJ, Bowen D, Gannon D, O'Brien K, Souza E (2001) Solvent retention capacities of irrigated soft white spring wheat flours. Crop Sci 41:1054

Guttieri MJ, Souza EJ (2003) Sources of variation in the solvent retention capacity test of wheat flour. Crop Sci 43:1628

Guttieri MJ, Souza EJ, Sneller C (2008) Nonstarch polysaccharides in wheat flour wire-cut cookie making. J Agric Food Chem 56:10927-10932

Habier D, Fernando R, Dekkers J (2007) The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. Genetics 177:2389

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci 92:433

Hartigan J, Wong M (1978) Algorithm AS: a K-mean clustering algorithm. Appl Stat 100-108

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

Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci 49:1

Henderson C (1984) Application of linear models in animal breeding. University of Guelph, Ontario

Holland JB (2004) Implementation of molecular markers for quantitative traits in breeding programs: Challenges and opportunities. p. 26. In T. Fischer et al. (ed.) New Directions for a Diverse Planet: Proc. for the 4th Int. Crop Science Congress, Brisbane, Australia. 26 Sept.–1 Oct. 2004. Regional Institute, Gosford, Australia

Hospital F, Moreau L, Lacoudre F, Charcosset A, Gallais A (1997) More on the efficiency of marker-assisted selection. Theor Appl Genet 95:1181-1189

Huang X, Cloutier S, Lycar L, Radovanovic N, Humphreys D, Noll J, Somers D, Brown P (2006) Molecular detection of QTLs for agronomic and quality traits in a doubled haploid population derived from two Canadian wheats (Triticum aestivum L.). Theor Appl Genet 113:753-766

Jannink JL (2010) Dynamics of long-term genomic selection. Genet Sel Evol 42:35doi:10.1186/1297-9686-42-35

Johnson R (2004) Marker-assisted selection. Plant Breed Rev 24:293-310

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709

Kizilkaya K, Fernando R, Garrick D (2010) Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. J Anim Sci 88:544

Knapp S, Bridges W (1990) Using Molecular Markers to Estimate Quantitative Trait Locus Parameters: Power and Genetic Variances for Unreplicated and Replicated Progeny. Genetics 126:769-777

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743-756

Lorenzana RE, Bernardo R (2009) Accuracy of genetic value predictions for marker based selection in biparental plant populations. Theor Appl Genet 120:151-161

Manly KF, Cudmore RH, Meer JM (2001) Map Manager QTX, cross-platform software for genetic mapping. Mammalian Genome 12:930-932

Meuwissen THE, Hayes B, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819

Muir WM (2007) Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J Anim Breed Genet 124:342-355

Munkvold JD, Tanaka J, Benscher D, Sorrells ME (2009) Mapping quantitative trait loci for preharvest sprouting resistance in white wheat. Theor Appl Genet 119:1223-1235

Piepho H (2009) Ridge regression and extensions for genomewide selection in maize. Crop Sci 49:1165

Piyasatian N, Fernando RL, Dekkers JC (2007) Genomic selection for marker-assisted improvement in line crosses. Theor Appl Genet 115:665-674

Podlich DW, Winkler CR, Cooper M (2004) Mapping as you go: an effective approach for marker-assisted selection of complex traits. Crop Sci 44:1560

R Development Core Team (2009) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

Sebastian S, Streit L, Stephens P, Thompson J, Hedges B, Fabrizius M, Soper J,Schmidt D, Kallem R, Hinds M (2010) Context-specific marker-assisted selection for improved grain yield in elite soybean populations. Crop Sci 50:1196

Smith N (2008) Identification and validation of quantitative trait loci affecting the milling and baking quality of soft red winter wheat. Dissertation, The Ohio State University

Sorrells M, Anderson J (1998) Registration of 'Cayuga' wheat. Crop Sci 38:551-552

Sorrells M, Benscher D, Cox W (2004) Registration of 'Caledonia' wheat. Crop Sci 44:1471

VanRaden P, Van Tassell C, Wiggans G, Sonstegard T, Schnabel R, Taylor J, Schenkel F (2009) Invited review: Reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92:16

VanRaden P (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414

Van Sanford D, Swanson C, Pearce W, Tutt C, Tomes L, Hershman D (1997) Registration of 'Foster' wheat. Crop Sci 37:627

Verbyla KL, Bowman PJ, Hayes BJ, Goddard ME (2010) Sensitivity of genomic selection to using different prior distributions. BMC Proc 4 Suppl 1:S5

Walker C, Campbell KG, Carter B, Kidwell K (2008) Using the solvent retention capacity test when breeding wheat for diverse production environments. Crop Sci 48:495

Whittaker JC, Thompson R, Denham MC (2000) Marker-assisted selection using ridge regression. Genet Res 75:249-252

Wong CK, Bernardo R (2008) Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. Theor Appl Genet 116:815-824

Zhong S, Dekkers JCM, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: A barley case study. Genetics 182:355-364

CHAPTER 4

GENOMIC SELECTION ACCURACY USING MULTI-FAMILY PREDICTION MODELS IN A WINTER WHEAT BREEDING PROGRAM

ABSTRACT

Genomic selection (GS) uses genome-wide molecular marker data to predict the genetic value of selection candidates in breeding programs. In plant breeding, the ability to produce large numbers of progeny per cross allows GS to be conducted within each family. However, this approach requires phenotypes of lines from each cross prior to conducting GS. This will prolong the selection cycle and may result in lower gains per year than approaches that estimate marker-effects with multiple families from previous selection cycles. In this study, phenotypic, conventional marker-assisted selection (MAS), and GS prediction accuracy of genetic values were compared for 13 agronomic traits in a population of 374 winter wheat (*Triticum aestivum* L.) breeding lines from an advanced-cycle winter wheat breeding program. A cross-validation approach that trained and validated prediction accuracy across years was used to evaluate effects of model selection, training population size, and marker density in the presence of GxE. Prediction accuracies using GS were 28% greater than with conventional-MAS and were 95% as accurate as phenotypic selection (PS) when averaged across all 13 traits studied. For net merit, the average accuracy across six selection indices for GS was 14% greater than for PS. These results provide empirical evidence that multi-family-GS could produce greater genetic gain per unit time and cost than both phenotypic selection and conventional-MAS in plant breeding.

INTRODUCTION

Quantitative traits such as grain yield have proven difficult to improve with marker-assisted selection (MAS). The main limitations are: 1) small population sizes and conventional statistical methods that have inadequate power to detect and accurately estimate effects of small-effect quantitative trait loci (QTL), and 2) gene-by-gene interactions (epistasis) and genotype-by-environment interactions (GxE) that have limited the transferability of QTL effect estimates across populations and environments (reviewed by Bernardo 2008; Xu and Crouch 2008). These limitations can be mitigated in plant breeding with improved marker-based breeding methods like genomic selection (GS; Meuwissen et al. 2001) and with "mapping-as-you-go" approaches that continually re-estimate marker effects in breeding populations and target environments in parallel with the selection process (Podlich et al. 2004).

Genomic selection addresses the first limitation by using a random-effects approach to jointly estimate all marker effects without significance testing to capture small-effect QTL that are excluded by conventional-MAS (Meuwissen et al. 2001). Marker estimates for GS are derived from a "training population" (TP), composed of breeding material with both phenotypic and genome-wide marker data. Marker estimates are then used to calculate genomic estimated breeding values (GEBVs) of new breeding lines in the "selection population" (SP). The combination of affordable, high-throughput genotyping and GS prediction methods has resulted in marker-based prediction accuracies that are revolutionizing cattle breeding (reviewed by Hayes et al. 2009a and Calus 2010) and show great promise for increasing gains from selection in plant breeding (reviewed by Heffner et al. 2009 and Jannink et al. 2010).

A “mapping-as-you-go” approach addresses the second limitation by re-estimating marker-effects in new breeding populations across target environments to capture changes in epistasis and GxE that will result from shifts in genetic backgrounds caused by selection (Podlich et al. 2004). Well-funded breeding programs are able to maximize the context-specificity of marker effect estimates by conducting MAS within each new cross (e.g. a biparental population), as they have the resources to generate large progeny numbers for each cross and have extensive multi-environment testing regimens (Sebastian et al. 2010). Unfortunately, this approach will also require the phenotyping of a subset of progeny from each cross before performing marker-based selection and will fail to leverage data generated from previous breeding cycles. Alternatively, MAS cycle times can be reduced to increase gains by avoiding this phenotyping step through estimating marker effects with data across multiple families in a breeding program (Jannink et al. 2001; Rafalski 2002; Breseghello and Sorrells 2006; Heffner et al. 2010). While a multi-family approach will be less “population-specific” and may increase error due to epistasis, this approach should reduce error due to GxE as it can leverage multi-year data thereby providing a greater sample of target environmental conditions (Podlich et al. 2004; Heffner et al. 2009).

Genomic selection within each cross, herein referred to as biparental-GS, has been shown to achieve higher prediction accuracies than conventional-MAS in biparental populations both in simulations (Bernardo and Yu 2007; Wong and Bernardo 2008) and in empirical studies (Lorenzana and Bernardo 2009; Heffner et al. submitted). Biparental-GS also has been found to compare favorably to phenotypic selection (PS). Lorenzana and Bernardo (2009) reported biparental-GS accuracies for maize (*Zea mays* L.), barley (*Hordeum vulgare* L.), and *Arabidopsis thailiana* (L.) that

would approach 1.5 times more gain than PS when using year-round nurseries capable of three GS cycles per year. Heffner et al. (submitted) reported GS prediction accuracies (r_{GS}) for nine quality traits in winter wheat (*Triticum aestivum* L.). Even when using TPs as small as 48 inbred lines and a maximum of only two GS cycles per year for winter wheat, their results also suggest that biparental-GS would outperform PS.

Performance of GS in plant populations using marker effects estimated from multiple families, herein called multi-family-GS, is limited. In a simulation study that used empirical marker data and simulated phenotypes for two-row barley, Zhong et al. (2009) reported a r_{GS} of ~0.60 for a trait controlled by 80 QTL with a heritability (h^2) of 0.40. Using two years of phenotypic data from 1,700 maize hybrids to predict 288 new hybrid combinations grown in two different years, van Eeuwijk et al. (2009) reported a r_{GS} of ~ 0.70 for ear height in maize (h^2=0.36). Finally, Crossa et al. (2010) used cross-validation to evaluate GS in 599 historical wheat lines and 284 maize inbreds from the International Maize and Wheat Improvement Center (CIMMYT). Using multiple GS models and environments, r_{GS} for wheat grain yield ranged from 0.36 to 0.61, for maize flowering time ranged from 0.46 to 0.79, and for maize grain yield ranged from 0.42 to 0.53 (Crossa et al. 2010). Results from these studies strongly support the utility of GS in plant breeding because deterministic simulation has shown that if r_{GS} for net merit (i.e. overall performance) exceeds 0.50, GS could greatly outperform conventional-MAS in terms of gain per unit time and cost (Heffner et al. 2010).

Empirical comparisons between conventional-MAS using markers identified by association mapping and multi-family-GS in plant breeding programs are currently

unavailable. However, association mapping of human height provides an empirical example of the difficulty of capturing genetic variance for highly polygenic traits with fixed effect models that have stringent thresholds. Despite a high h^2 (~0.80) and a TP of tens of thousands of individuals, only ~5% of the phenotypic variance for human height has been accounted for with ~50 significant markers (Gudbjartsson et al. 2008; Lettre et al. 2008; Weedon et al. 2008; Visscher 2008). In contrast, Yang et al. (2010) fit all 300,000 markers simultaneously as random effects and found that using all markers explained 45% of phenotypic variation for human height in a population of ~4,000 unrelated individuals. Thus, they concluded a large proportion of the genetic variance was explained with small-effect markers that do not pass stringent thresholds. Relaxing these thresholds can improve the amount of genetic variance explained with significant markers (Hospital et al. 1997; Moreau et al. 1998); however, small population sizes, low heritability, and confounding population structure will still cause small-effect markers to be below significance thresholds. Consequently, genetic variance will go uncaptured and significant QTL effects will be overestimated (Beavis 1998; Schön et al. 2004; Xu 2003). This suggests that GS models should outperform conventional-MAS models in plant populations composed of multiple families.

The objective of this study was to empirically compare phenotypic prediction accuracy (r_P), conventional-MAS accuracy (r_{MAS}), and r_{GS} when marker effects were estimated with multi-family data from a breeding program. To meet this objective, cross-validation across years was performed in a population of 374 elite wheat breeding lines using genome-wide marker data and several marker-based prediction models to predict performance of 13 agronomic traits. For each trait, training population size (N_{TP}) and marker number (N_M) were varied to investigate their effects

on prediction accuracy. Finally, r_{GS} and r_P for net merit were compared using index selection.

METHODS AND MATERIALS

Data

A population of 374 soft winter wheat varieties and F_5-derived advanced breeding lines resulting from many different crosses in the Cornell University Wheat Breeding Program were analyzed in this study. Lines were genotyped with 5,000 Diversity Array Technology markers (DArT; Triticarte Pty. Ltd., Yaralumla, Australia), resulting in 1544 polymorphic markers. Some markers were perfectly correlated to each other due to complete LD ($r^{2=}1$). Therefore, the data set was trimmed to 1158 markers by selecting the marker with the least missing data from each pair or group of markers that were in complete LD. Missing data was imputed as the mean marker score for each marker because precise map position was unknown for many of the markers.

Phenotypic data for 13 traits were analyzed: grain yield, plant height, heading date (i.e. days to heading), lodging, pre-harvest sprouting (PHS), flour yield, flour protein, softness, sucrose solvent retention capacity (Suc-SRC), water SRC (H_2O-SRC), lactic acid SRC (LA-SRC), and sodium carbonate SRC (NaCO-SRC). PHS is the premature germination of seeds while still attached to the mother plant that decreases grain value and was measured as described by Anderson et al. (1993) and Munkvold et al. (2009). Milling and baking quality traits were measured as follows: flour yield - percentage of flour obtained from milling, softness - percentage of fine

flour obtained i.e. that which can pass through a 94-mesh (180μm) screen, protein - percent protein of flour measured using a near infrared analyzer (Unity Spectrastar 2200, Columbia MD), and SRC - amount of solvent retained by the flour after centrifugation and draining. The four SRC tests were used to predict overall baking quality: H_2O-SRC for global water absorption, NaCO-SRC for damaged starch, Suc-SRC for arabinoxylan and partially hydrated gliadin content, and LA-SRC for gluten strength. The USDA-ARS Soft Wheat Quality Laboratory in Wooster, Ohio performed all milling and baking quality tests as described by Guttieri et al. (2008).

Phenotypic data were collected from field trials in two years, 2008 and 2009, with three locations per year near Ithaca, NY. Each year, two locations were yield plots (1.26 m by 4 m) and one location was single 1 m rows. All traits were measured in yield trials and PHS, height, and heading date were also measured in single row trials. Each location was arranged in an unreplicated augmented design (Federer, 1956) with 6 check varieties replicated 10 times each. A two-stage analysis was used to calculate line best linear unbiased predictions (BLUPs) because it is less computationally demanding than a one-stage analysis and has been shown to generate similar results (Möhring and Piepho 2009). First, BLUPs were calculated for each trait in each location with a two-dimensional, first-order autoregressive (AR1 x AR1) spatial model with lines as random effects in ASReml (Gilmour et al. 2009). For PHS, an additional random effect of harvest date was included. Second, line BLUPs were calculated for each year with random effects of line and location. Only the first stage was used for H_2O-SRC and NaCO-SRC because they were only measured in two locations in 2008.

Prediction Models

Six methods were used to estimate marker effects for marker-based prediction: association analysis (*AA*), association analysis including kinship as a covariate (*AK*), ridge-regression BLUP (*RR*), BayesA (*BA*), BayesB (*BB*), and Bayes-Cπ (*BC*). All statistical procedures herein were executed using R (R Development Core Team 2009).

Conventional-MAS Models

A two-stage approach using association analysis and multiple linear regression (*MLR*) was used to represent conventional-MAS using multi-family data. That is, association analysis first reduced the number of markers (predictor variables), and *MLR* then selected markers to be included in the final prediction model and estimated the marker effects (regression coefficients) used to calculate GEBVs. *AA* and *AK* modeled environments and markers as fixed effects, and *AK* had an additional random covariate, a simple identity-by-state allele sharing kinship matrix (*K*), to account for genetic covariance among individuals to reduce the number of false positive marker-trait associations caused by population structure and genetic relatedness (Zhao et al. 2007; Kang et al 2008). Calculation of *K* and detection of marker-trait associations were performed with the R package 'emma' (Kang et al. 2008). A significance threshold of 0.05 was used for *AA* and *AK* because relaxed thresholds have been shown to increase marker-based prediction accuracy (Hospital et al. 1997; Moreau et al. 1998). Relaxed thresholds were also used in *MLR* for forward (0.2) and backward (0.2) variable selection.

Genomic Selection Models

Four GS models were used in this study. *RR* assumes all markers have a common variance (Meuwissen et al. 2001; Whittaker et al. 2000), and thus shrinks each marker effect equally toward zero. *RR* is equivalent to estimating markers effects with a realized-relationship matrix determined from markers (Habier et al. 2007; Goddard 2009; Piepho 2009). The additive realized-relationship matrix was estimated in R, and the R package 'emma' (Kang et al. 2008) was used to estimate the variance components to solve mixed-model equations (Henderson 1984).

Three Bayesian models were used to address the simple, but likely unrealistic *RR* assumptions of all markers having non-zero effects and equal marker variances. *BA* fits all markers but allows each marker to have its own variance (Meuwissen et al. 2001). In addition to allowing for unique marker variance, *BB* also specifies that a portion of the markers (π) have no effect (Meuwissen et al. 2001). Thus, *BB* is equivalent *BA* when $\pi = 0$. Finally, *BC* assumes common marker variances and allows for some markers to have no effect (Dekkers et al. 2009, Jannink 2010). Additionally, *BC* jointly estimates π from the training data to avoid an incorrect π that can negatively affect prediction accuracy (Verbyla et al. 2010; Gianola et al. 2009). We adapted *BA*, *BB*, and *BC* code written by R.L. Fernando (Dekkers et al. 2009). For *BC*, a starting π=0.50 was used. For *BB*, π=0.90 was used because preliminary results showed that π values of 0.95 and 0.975 generally decreased accuracy. Each method was run for 2,000 iterations, had a burn-in period of 200 iterations. This was considered sufficient for approximate convergence, as the average correlation of results from two independent runs of 40 random TPs for each trait and each model was greater than 0.99.

Prediction Accuracy and Cross-validation

Phenotypic prediction accuracy (r_P) was the correlation of the observed phenotypes from 2008 and 2009. Marker-based prediction accuracy for conventional-MAS (r_{MAS}) and r_{GS} was the correlation of GEBVs from one year and the observed phenotypes on the other year. GEBVs were calculated as $GEBV_i = X_i g$: where $GEBV_i$ was the GEBV of line i, X_i was the vector of the marker scores for that line, and g was the vector of marker effects obtained from TP using a marker-based prediction model.

Three different training population sizes (N_{TP}=288, 192, and 96) were used for marker-based prediction with N_m=1158. Multiples of 96 were used to correspond with current 96 or 384-well DNA sample plates. As overall population size was 374 and maximum N_{TP} was 288, GEBVs were calculated for 86 lines marker effects estimated from the TP. The observed phenotypes and GEBVs of the 86 lines from one year were correlated to observed phenotypes of the other year to obtain prediction accuracies and to avoid bias introduced by genotype by year interactions. To achieve adequate sampling of the genetic diversity both for training and validation, TP lines were randomly and proportionally sampled from six genetic clusters of size 48, 79, 95, 38, 50, and 64. Cluster assignment and selection of optimal cluster number and model ("VEI": diagonal, varying volume, equal shape) using the Bayesian information criterion were done using the R package ‘mclust’ (Fraley and Raferty 2002; Fraley and Raferty 2006).

Four marker densities (N_M=1158, 768, 384, and 192) were used to assess the impact of N_M on prediction accuracy when using a N_{TP}=288. K-means clustering (Hartigan and Wong 1979) was used to select informative marker subsets to minimize

LD between selected markers and maximize genome coverage. Markers were selected using the function 'kmeans' in R where: each marker was treated as an explanatory variable, the number of clusters equaled the N_M, the number of random starts equaled 1000, and the marker closest to the centroid of each cluster was chosen. Significance thresholds for *AA* and *AK* were relaxed from 0.5 (N_M=1158) to 0.1, 0.2, and 0.3 for N_M=768, 384, and 192, respectively.

All six marker-based prediction methods were evaluated for N_{TP}=288 and N_M=1158 for 100 TPs using training data in 2008 and 2009 for a total of 1,200 analyses for each trait. Because the Bayesian models were computational intensive for cross-validation and a preliminary analysis showed all GS methods produced similar trends, only two GS models (*RR* and *BC*) were used for investigation of the effects of N_{TP} and N_M. Therefore, four models (*AA*, *AK*, *RR*, and *BC)* were used for N_{TP}=192 and 96 with N_M=1158 and N_M=768, 384, and 192 with N_{TP}=288. As before, each scenario was analyzed each model for 100 TPs using training data in 2008 and 2009, totaling an additional 4,000 analyses for each trait. Reported r_P, r_{MAS}, and r_{GS} for each trait was the average accuracy for all 100 TPs across both years, and prediction methods were compared using a paired *t*-test (α=0.01).

Net Merit Prediction Accuracy

To predict net merit, trait predictions were combined using weighting determined by the "Smith-Hazel index" (Smith 1936; Hazel 1943) and by the "base index" (Panse 1946, Brim et al. 1959; Williams 1962). The estimated Smith-Hazel index is $\hat{b} = \hat{P}^{-1}\hat{G}a$, where $\hat{b}$ is the vector of estimated trait-weights, a is the vector of relative economic trait-weights, $\hat{P}$ is the estimated phenotypic covariance matrix,

and $\hat{G}$ is the estimated additive-genetic covariance matrix. The base index ignores phenotypic and genetic covariances; thus, trait predictions are weighted simply by their relative economic trait-weights.

Three economic weighting indices were used: 1) emphasis on yield, 2) emphasis on milling and baking quality traits, and 3) a "balanced" index representing current breeding goals of the Cornell University Wheat Breeding Program (Table 4.2). H_2O-SRC and NaCO-SRC were excluded from the indices, as they were only measured in two locations in 2008. The phenotypic covariance matrix ($\hat{P}$) was estimated using line BLUPs calculated using phenotypes for each line and each trait from all four locations. The additive-genetic covariance matrix ($\hat{G}$) was estimated using GEBVs for each trait and each line that were calculated from genotypic data and trait BLUPs from all 374 lines using *RR*. Phenotypic prediction accuracy and *RR* were used for comparing prediction accuracy for each index. For the Smith-Hazel index, $r_P = cor(Ph_1\hat{b} : Ph_2\hat{b})$ and $r_{RR} = cor(GEBV_1 a : Ph_2\hat{b})$ where: Ph_1 is a vector of observed phenotypes from one year, $GEBV_1$ is a vector of GEBVs from the one year, and Ph_2 is a vector of observed phenotypes from the other year. For the base index, $r_P = cor(Ph_1 a : Ph_2 a)$ and $r_{RR} = cor(GEBV_1 a : Ph_2 a)$. All phenotypes were standardized to mean zero with a standard deviation of one prior to index analyses. Net merit accuracies were also divided by the square root of the broad-sense heritability (*H*) of net merit in validation data on a line-mean basis to account for the validation phenotypes not being equal to the "true genetic value" (Dekkers 2007).

RESULTS

Marker-based and Phenotypic Prediction Accuracy

The r_{GS} was greater than r_{MAS} for all 13 traits studied (Table 4.1). For the maximum N_M (1158) and N_{TP} (288), mean r_{GS} (0.58) across all methods and traits was 28% greater than r_{MAS} (0.46). A large range of r_{GS} was observed, ranging from 0.17 (grain yield; *BC*) to 0.76 (LA-SRC; *BA*). The range for r_{MAS} was 0.18 (grain yield; *AK*) to 0.63 (Suc-SRC; *AA*). Only slight differences were detected between the GS models, with *BA* having the highest mean accuracy across all traits. Accuracy of *BA* and *RR* were most similar, as *BA* was significantly different from *RR* only for grain yield (*BA*=0.22 versus *RR*=0.20). In most cases where *BB* and *BC* were significantly different than *BA* and *RR*, their r_{GS} were marginally lower than *RR* and *BA*, but again, differences were quite small. The best conventional-MAS method was *AA*, which was significantly greater than *AK* for six of the 13 traits.

A wide range of r_P was also observed, ranging from 0.21 (grain yield) to 0.89 (heading date), reflecting the wide range of H^2 for the traits in this study (Supplemental Table 4.3). When comparing the r_P to the highest r_{GS} achieved across all four GS models, r_P was greater for nine traits, not significantly different for three traits (test weight, H_2O-SRC and grain yield), and less than r_{GS} for one trait (lodging; Table 4.1). The ratio of r_{GS} and r_P (r_{GS} / r_P) ranged from 0.84 (heading date) to 1.08 (lodging) with a mean ratio of 0.95. When comparing the r_P to the highest r_{MAS} achieved across both conventional-MAS models, r_P was greater than r_{MAS} for all traits with r_{MAS}/r_P ranging from 0.56 (height) to 0.91 (grain yield) with a mean r_{MAS}/r_P of 0.76. Finally, the slope from linear regression of r_{GS} by r_P (0.81) and

Table 4.1 Phenotypic and marker-based prediction accuracy for 13 traits. Marker-based prediction was based on N_{TP}=288

Trait	r_P	r_{AA}	r_{AK}	r_{RR}	r_{BA}	r_{BB}	r_{BC}	r_{GS}/r_P
Flour Yield	0.831	0.648	0.544	0.762^{A}	0.761^{A}	0.740	0.761	0.917
H_2O-SRC	0.607^{A}	0.484^{B}	0.475^{B}	0.577^{C}	0.585^{AC}	0.570^{CD}	0.556^{D}	0.963
Heading Date	0.891	0.551	0.431	0.748^{A}	0.750^{AB}	0.724	0.746^{B}	0.842
Height	0.851	0.521	0.467	0.740^{A}	0.746^{AB}	0.719	0.739^{B}	0.877
LA-SRC	0.786	0.626^{A}	0.625^{A}	0.750^{BC}	0.757^{B}	0.754^{C}	0.751^{BC}	0.963
Lodging	0.262^{A}	0.190^{B}	0.186^{B}	0.278^{AC}	0.281^{AC}	0.284^{C}	0.232	1.073
NaCO-SRC	0.726	0.592	0.559	0.658^{AB}	0.680^{A}	0.671^{B}	0.656^{B}	0.937
PHS	0.578	0.465^{A}	0.449^{A}	0.553^{B}	0.554^{B}	0.552^{B}	0.553^{B}	0.958
Protein	0.493	0.377	0.303	0.449^{AB}	0.453^{A}	0.446^{B}	0.450^{B}	0.919
Softness	0.774	0.529^{A}	0.507^{A}	0.664^{BC}	0.670^{B}	0.649^{C}	0.664^{BC}	0.866
Suc-SRC	0.799	0.627	0.568	0.736^{AB}	0.741^{AC}	0.727^{BD}	0.734^{CD}	0.927
Test Weight	0.568^{AB}	0.470^{B}	0.455^{B}	0.560^{AB}	0.571^{A}	0.560^{BC}	0.554^{C}	1.005
Grain Yield	0.205^{AB}	0.186^{CD}	0.180^{C}	0.199^{AD}	0.223^{B}	0.222^{B}	0.174^{C}	1.088
Mean	**0.644**	**0.482**	**0.442**	**0.590**	**0.598**	**0.586**	**0.582**	**0.949**

^{ABCD} Within each trait, accuracies that share the same letter were not significantly different for $\alpha = 0.01$

r_{MAS} by r_P (0.64) showed that r_{GS}/r_P and r_{MAS}/r_P decreased as r_P increased (Supplementary Fig. 4.1).

Effects of Training Population Size and Marker Number

Decreasing N_{TP} had a strong negative effect on the mean accuracy across all traits for each of the four prediction models tested (Fig. 4.1; Supplemental Table 4.1). Decreasing N_{TP} from 288 to 198 and 96 reduced the average r_{GS} by 11% and 30% and

r_{MAS} by 24% and 35%, respectively (Fig. 4.1). Reducing N_M from 1158 to 768 and 384 resulted in a small decrease in r_{GS} and a small increase r_{MAS} (Fig. 4.2), while reducing N_M from 1158 to 192 reduced the average r_{GS} by 10% and r_{MAS} by less than 4%.

Prediction Accuracy for Net Merit

Prediction accuracy using the base index and the "yield", "balanced", and "quality" economic weights (Table 4.2) was greater for GS (*RR*) than for PS (Table 4.3). For both GS and PS, the balanced weights had the lowest prediction accuracy prior to adjusting for error in the validation data. After adjusting the validation data for error by dividing by the square root of the H^2 (Dekkers 2007), yield and balanced weights had equivalent accuracies and the quality weights had the highest accuracy. The Smith-Hazel index resulted in higher r_{GS} than r_P for the quality weights, equal accuracy for the balanced weights, and lower accuracy for the yield weights. The Smith-Hazel index using the yield weights had the highest accuracy while the quality weights had the lowest accuracy before and after correction using *H*. The mean r_{GS}/H for all indices (0.54) was greater than the mean r_P/H (0.47) as only one index (Smith Hazel – yield) resulted in a r_P/H greater than r_{GS}/H.

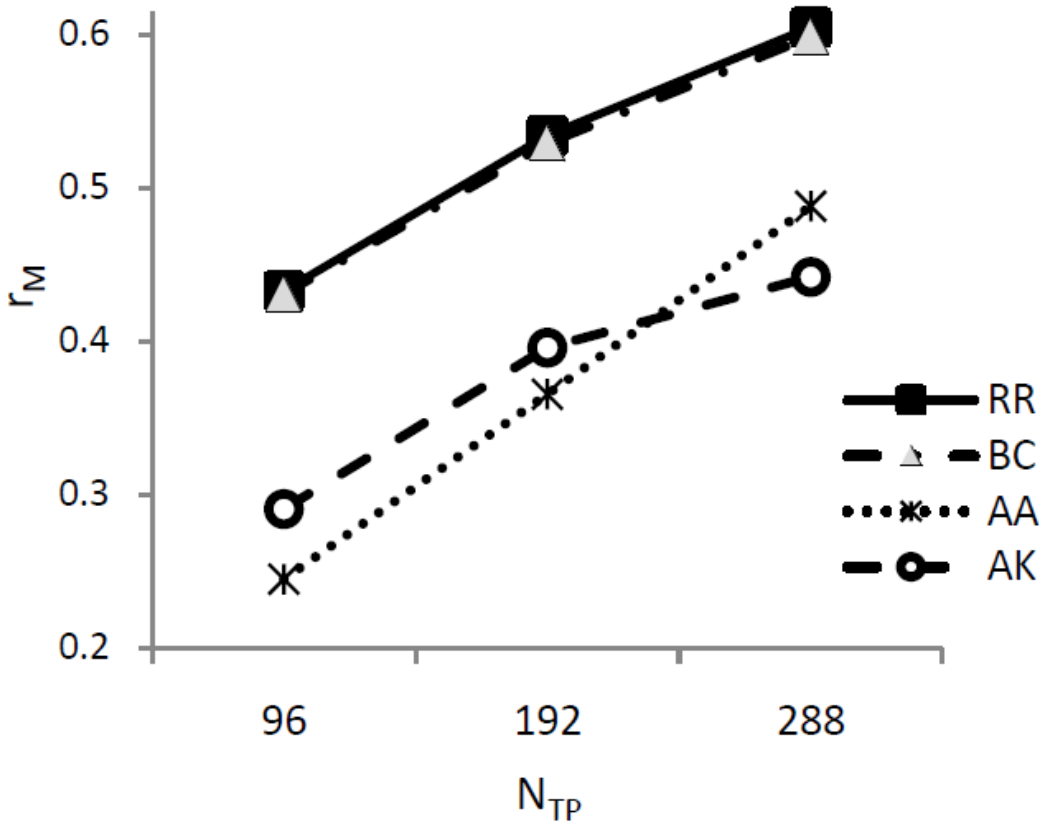


Figure 4.1 The effect of training population size (N_{TP}) on marker-based prediction accuracy (r_M)

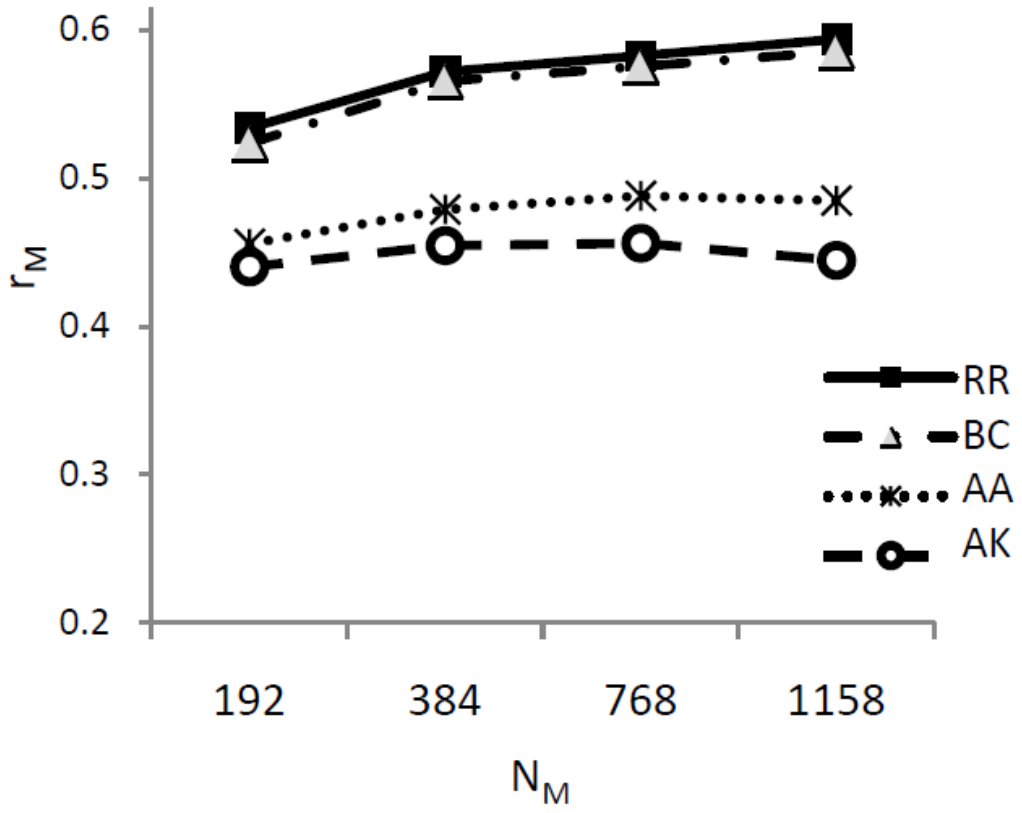


Figure 4.2 The effect of marker number (N_M) on marker-based prediction accuracy (r_M)

Table 4.2 Economic weight indices for index selection

Trait	Index		
	Yield	Balanced	Quality
Flour Yield	1.1	2.1	4.2
Heading Date	-1.0	-1.0	-1.0
Height	-4.0	-4.0	-4.0
LA-SRC	-1.1	-2.1	-4.2
Lodging	-20.0	-20.0	-20.0
PHS	-5.0	-10.0	-10.0
Protein	-0.8	-1.6	-3.2
Softness	1.1	2.1	4.2
Suc-SRC	-1.1	-2.1	-4.2
Test Weight	15.0	15.0	15.0
Grain Yield	50.0	40.0	30.0

Table 4.3 Phenotypic and GS prediction accuracy for net merit using the base and Smith-Hazel indices for each economic weight index

Index	r_P	r_{GS}	H^A	r_P/H^B	r_{GS}/H^B
Base					
Yield	0.16	0.21	0.49	0.32	0.44
Balanced	0.14	0.20	0.45	0.32	0.44
Quality	0.17	0.21	0.44	0.38	0.48
Smith-Hazel					
Yield	0.33	0.31	0.45	0.73	0.68
Balanced	0.29	0.28^{NS}	0.46	0.62	0.62
Quality	0.23	0.28	0.51	0.45	0.55
Mean	0.22	0.25	0.47	0.47	0.54

[A] Square root of the broad-sense heritability (H^2) for net merit in validation data

[B] Accuracy divided by H to adjust for error in the validation data

[NS] r_{GS} and r_P were not significantly different for $\alpha = 0.01$

Supplementary Table 4.1 Effect of training population size (N_{TP}) on prediction accuracy. Comparison of GS accuracy and phenotypic prediction accuracy (r_{GS} / r_P) used the highest accuracy of the two GS models.

Trait	N_{TP}	r_P	r_{AA}	r_{AK}	r_{RR}	r_{BC}	r_{GS}/r_P
Flour Yield	288	0.831	0.648	0.544	0.762	0.761	0.917
	192	0.831	0.469^{A}	0.493^{A}	0.682	0.681	0.821
	96	0.831	0.324	0.401	0.573^{A}	0.571^{A}	0.690
H_2O-SRC	288	0.607	0.484^{A}	0.475^{A}	0.577	0.556	0.951
	192	0.607	0.413^{A}	0.427^{A}	0.533	0.513	0.878
	96	0.607	0.267^{A}	0.279^{A}	0.405	0.346	0.667
Heading Date	288	0.891	0.551	0.431	0.748	0.746	0.840
	192	0.891	0.385	0.359	0.607^{A}	0.605^{A}	0.681
	96	0.891	0.173	0.219	0.404^{A}	0.405^{A}	0.453
Height	288	0.851	0.521	0.467	0.740	0.739	0.870
	192	0.851	0.352	0.416	0.610	0.608	0.717
	96	0.851	0.240	0.294	0.481^{A}	0.480^{A}	0.565
LA-SRC	288	0.786	0.626^{A}	0.625^{A}	0.750^{B}	0.751^{B}	0.954
	192	0.786	0.490	0.566	0.695^{A}	0.695^{A}	0.884
	96	0.786	0.387	0.438	0.626	0.626	0.796
Lodging	288	0.262^{A}	0.190^{B}	0.186^{B}	0.278^{A}	0.232	1.061
	192	0.262	0.166	0.170	0.243	0.203	0.927
	96	0.262	0.125	0.127	0.189	0.175	0.721
NaCO-SRC	288	0.726	0.592	0.559	0.658^{A}	0.656^{A}	0.906
	192	0.726	0.485	0.480	0.601	0.601	0.828
	96	0.726	0.287	0.342	0.474	0.400	0.653

^{AB} Within each trait and N_{TP}, accuracies that share the same letter were not significantly different for $\alpha = 0.01$

Supplementary Table 4.1 Continued Effect of training population size (N_{TP}) on prediction accuracy. Comparison of GS accuracy and phenotypic prediction accuracy (r_{GS}/r_P) used the highest accuracy of the two GS models.

Trait	N_{TP}	r_P	r_{AA}	r_{AK}	r_{RR}	r_{BC}	r_{GS}/r_P
PHS	288	0.578	0.465^{A}	0.449^{A}	0.553^{B}	0.533^{B}	0.957
	192	0.578	0.368	0.407	0.509	0.507	0.881
	96	0.578	0.255	0.294	0.429	0.423	0.742
Protein	288	0.493	0.377	0.303	0.449^{A}	0.450^{A}	0.913
	192	0.493	0.287	0.271	0.406	0.404	0.824
	96	0.493	0.189	0.191	0.336	0.334	0.682
Softness	288	0.774	0.529^{A}	0.507^{A}	0.664^{B}	0.664^{B}	0.858
	192	0.774	0.374	0.455	0.577	0.576	0.745
	96	0.774	0.214	0.283	0.406	0.407	0.525
Suc-SRC	288	0.799	0.627	0.569	0.736	0.734	0.921
	192	0.799	0.471	0.507	0.668	0.666	0.836
	96	0.799	0.344	0.415	0.573	0.573	0.717
Test Weight	288	0.568^{A}	0.470^{B}	0.455^{B}	0.560^{A}	0.554	0.986
	192	0.568	0.383	0.420	0.530	0.522	0.933
	96	0.568	0.283	0.342	0.480	0.472	0.845
Grain Yield	288	0.205	0.186^{A}	0.180^{A}	0.199^{B}	0.174^{B}	0.971
	192	0.205	0.152^{ABC}	0.156^{ABC}	0.165^{B}	0.147^{C}	0.805
	96	0.205	0.088^{A}	0.100^{AB}	0.112^{B}	0.113^{B}	0.546
Mean	288	0.644	0.482	0.442	0.590	0.582	0.931
	192	0.644	0.369	0.394	0.525	0.518	0.828
	96	0.644	0.244	0.287	0.422	0.410	0.662

^{ABC} Within each trait and N_{TP}, accuracies that share the same letter were not significantly different for $\alpha = 0.01$

Supplementary Table 4.2 Effect of marker number (N_M) on prediction accuracy. Comparison of GS accuracy and phenotypic prediction accuracy (r_{GS}/r_P) used the highest accuracy of the two GS models.

Trait	N_M	r_P	r_{AA}	r_{AK}	r_{RR}	r_{BC}	r_{GS}/r_P
Flour Yield	1158	0.831	0.648	0.544	0.762	0.761	0.917
	768	0.831	0.649	0.573	0.749	0.747	0.901
	384	0.831	0.636	0.587	0.733	0.731	0.882
	192	0.831	0.621	0.568	0.701	0.700	0.844
H_2O-SRC	1158	0.607	0.484^A	0.475^A	0.577	0.556	0.951
	768	0.607	0.458	0.427	0.551	0.543	0.908
	384	0.607	0.456	0.435	0.559^A	0.559^A	0.921
	192	0.607	0.454^A	0.444^A	0.527	0.506	0.868
Heading Date	1158	0.891	0.551	0.431	0.748	0.746	0.840
	768	0.891	0.551	0.466	0.744^A	0.743^A	0.835
	384	0.891	0.535	0.470	0.694	0.692	0.779
	192	0.891	0.464^A	0.460^A	0.612	0.611	0.687
Height	1158	0.851	0.521	0.467	0.740	0.739	0.870
	768	0.851	0.576	0.530	0.720^A	0.718^A	0.846
	384	0.851	0.589	0.539	0.698	0.696	0.820
	192	0.851	0.558	0.531	0.634^A	0.634^A	0.745
LA-SRC	1158	0.786	0.626^A	0.625^A	0.750^B	0.751^B	0.954
	768	0.786	0.648^A	0.639^A	0.744^B	0.745^B	0.948
	384	0.786	0.657	0.646	0.733^A	0.733^A	0.933
	192	0.786	0.663^A	0.661^A	0.711^B	0.712^B	0.906
Lodging	1158	0.262^A	0.190^B	0.186^B	0.278^A	0.232	1.061
	768	0.262	0.206	0.204	0.285^A	0.242	1.088
	384	0.262^A	0.226	0.226	0.306	0.248	1.168
	192	0.262^A	0.177	0.180	0.255^A	0.209	0.973
NaCO-SRC	1158	0.726	0.592	0.559	0.658^A	0.656^A	0.906
	768	0.726	0.572	0.543	0.631^A	0.631^A	0.869
	384	0.726	0.550	0.536	0.633^A	0.632^A	0.872
	192	0.726	0.525	0.517	0.584^A	0.580^A	0.804

^{AB} Within each trait and N_M, accuracies that share the same letter were not significantly different for $\alpha = 0.01$

Supplementary Table 4.2 Continued Effect of marker number (N_M) on prediction accuracy. Comparison of GS accuracy and phenotypic prediction accuracy (r_{GS}/r_P) used the highest accuracy of the two GS models.

Trait	N_M	r_P	r_{AA}	r_{AK}	r_{RR}	r_{BC}	r_{GS}/r_P
PHS	1158	0.578	0.465^{A}	0.449^{A}	0.553^{B}	0.533^{B}	0.957
	768	0.578	0.478	0.455	0.551^{A}	0.550^{A}	0.953
	384	0.578	0.448^{A}	0.448^{A}	0.549^{B}	0.550^{B}	0.952
	192	0.578	0.466^{A}	0.460^{A}	0.525^{B}	0.523^{B}	0.908
Protein	1158	0.493	0.377	0.303	0.449^{A}	0.450^{A}	0.913
	768	0.493	0.370	0.328	0.448^{A}	0.448^{A}	0.909
	384	0.493	0.380	0.348	0.445^{A}	0.445^{A}	0.903
	192	0.493	0.317	0.286	0.404	0.408	0.828
Softness	1158	0.774	0.529^{A}	0.507^{A}	0.664^{B}	0.664^{B}	0.858
	768	0.774	0.528^{A}	0.515^{A}	0.661	0.660	0.854
	384	0.774	0.490	0.466	0.636^{A}	0.635^{A}	0.822
	192	0.774	0.440	0.413	0.566^{A}	0.564^{A}	0.731
Suc-SRC	1158	0.799	0.627	0.569	0.736	0.734	0.921
	768	0.799	0.642	0.586	0.720	0.719	0.901
	384	0.799	0.613	0.570	0.712^{A}	0.711^{A}	0.891
	192	0.799	0.600	0.568	0.681	0.683	0.855
Test Weight	1158	0.568^{A}	0.470^{B}	0.455^{B}	0.560^{A}	0.554	0.986
	768	0.568	0.479^{A}	0.477^{A}	0.560	0.557	0.986
	384	0.568	0.475	0.461	0.550	0.546	0.968
	192	0.568	0.483	0.472	0.539	0.534	0.949
Grain Yield	1158	0.205^{A}	0.186^{BC}	0.180^{C}	0.199^{AB}	0.174^{C}	0.971
	768	0.205	0.161^{AB}	0.166^{A}	0.182	0.147^{B}	0.888
	384	0.205	0.155^{AB}	0.163^{AB}	0.167^{A}	0.148^{B}	0.815
	192	0.205	0.132^{A}	0.135^{A}	0.173	0.106	0.844
Mean	1158	0.644	0.482	0.442	0.590	0.582	0.931
	768	0.644	0.486	0.455	0.580	0.573	0.914
	384	0.644	0.472	0.453	0.570	0.564	0.902
	192	0.644	0.454	0.438	0.532	0.521	0.842

^{ABC} Within each trait and N_M, accuracies that share the same letter were not significantly different for $\alpha = 0.01$

Supplementary Table 4.3 Phenotypic and marker-based prediction accuracy for N_{TP}=288 and N_M=1158. Accuracies were corrected for error in the validation data by dividing be the square root of the broad-sense heritability of the validation data (H_{yr}). Comparison of GS accuracy and phenotypic prediction accuracy (r_{GS}/r_P) used the highest accuracy of the four GS models.

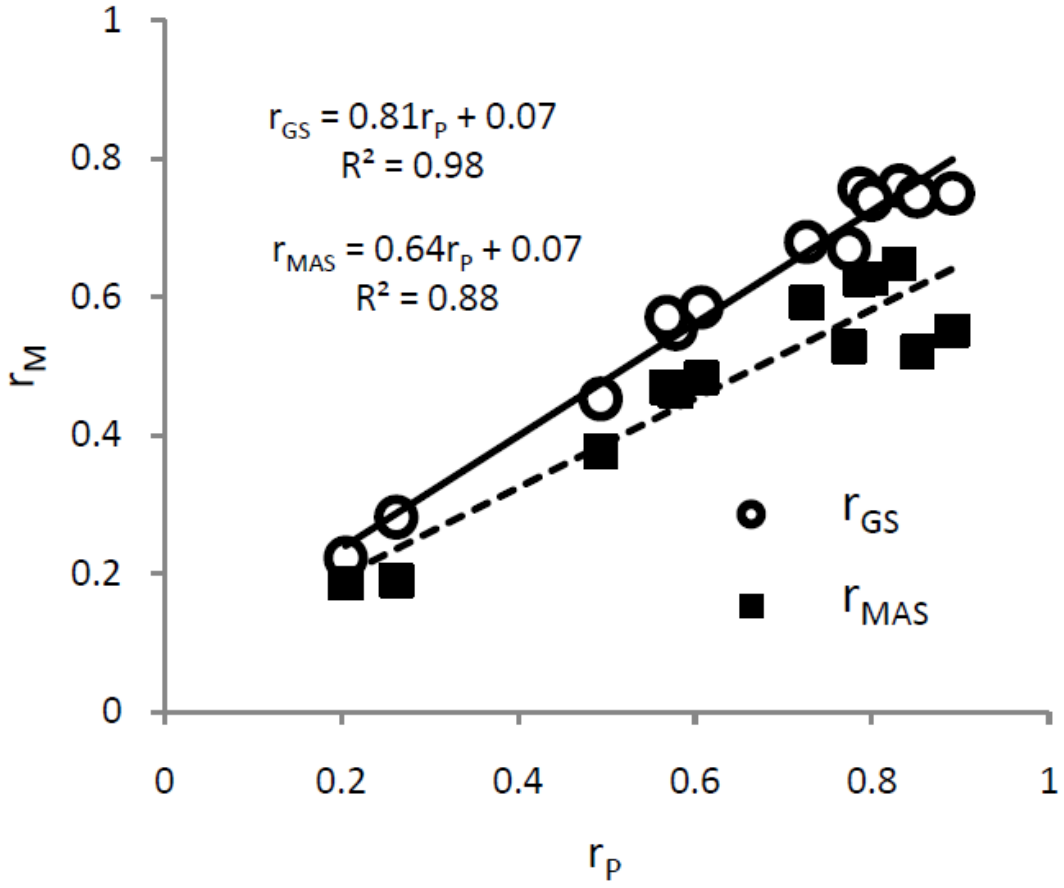
Trait	H^2	H^2_{yr}	H_{yr}	r_P/H_{yr}	r_{AA}/H_{yr}	r_{AK}/H_{yr}	r_{RR}/H_{yr}	r_{BA}/H_{yr}	r_{BB}/H_{yr}	r_{BC}/H_{yr}	r_{GS}/r_P
Flour Yield	0.93	0.87	0.93	0.89	0.69	0.58	0.82^{A}	0.86^{A}	0.79	0.82	0.92
Heading Date	0.92	0.85	0.92	0.97	0.60	0.47	0.81^{A}	0.81^{AB}	0.79	0.81^{B}	0.84
Height	0.92	0.86	0.93	0.92	0.56	0.50	0.80^{A}	0.81^{AB}	0.78	0.80^{B}	0.88
LA-SRC	0.91	0.83	0.91	0.86	0.69^{A}	0.69^{A}	0.83^{BC}	0.83^{B}	0.83^{C}	0.83^{BC}	0.96
Lodging	0.37	0.23	0.47	0.55^{A}	0.40^{B}	0.39^{B}	0.59^{AC}	0.60^{AC}	0.60^{C}	0.49	1.08
PHS	0.71	0.55	0.74	0.78	0.63^{A}	0.61^{A}	0.75^{B}	0.75^{B}	0.75^{B}	0.75^{B}	0.96
Protein	0.72	0.56	0.75	0.66	0.50	0.40	0.60^{AB}	0.61^{A}	0.60^{B}	0.60^{B}	0.92
Softness	0.91	0.84	0.92	0.84	0.58^{A}	0.55^{A}	0.72^{BC}	0.73^{B}	0.71^{C}	0.72^{BC}	0.87
Suc-SRC	0.90	0.82	0.91	0.88	0.69	0.63	0.81^{AB}	0.82^{AC}	0.80^{BD}	0.81^{CD}	0.93
Test Weight	0.72	0.56	0.75	0.76^{A}	0.63^{B}	0.61^{B}	0.75^{AC}	0.77^{A}	0.75^{CD}	0.74^{D}	1.01^{NS}
Grain Yield	0.35	0.21	0.46	0.45^{AB}	0.41^{CD}	0.39^{C}	0.43^{AD}	0.49^{AB}	0.48^{B}	0.38^{C}	1.09^{NS}
Mean	**0.76**	**0.65**	**0.79**	**0.82**	**0.61**	**0.56**	**0.75**	**0.76**	**0.74**	**0.74**	**0.95**

^{ABCD} Within each trait, pairwise comparisons that were not significantly different for $\alpha = 0.01$ share the same letter

^{NS} r_{GS} and r_P were not significantly different for $\alpha = 0.01$

Supplementary Table 4.4 Correlations of phenotypes (top triangle) and GEBVs (bottom triangle) for each trait used in the selection indices. Correlations between phenotypes and GEBVs for each trait are represented by the diagonal (bold) values. GEBVs were calculated with *RR* using all 374 lines and environments.

	Flour Yield	Heading	Height	LA-SRC	Lodging	PHS	Protein	Softness	Suc-SRC	Test Weight	Grain Yield
Flour Yield	**0.89**	0.11	-0.10	0.15	-0.21	0.23	-0.36	0.06	-0.60	0.02	0.23
Heading	0.12	**0.81**	0.10	-0.03	-0.02	-0.04	-0.09	0.03	-0.16	0.04	-0.01
Height	-0.37	0.22	**0.76**	-0.03	0.49	-0.20	0.21	-0.01	0.20	0.31	-0.22
LA-SRC	0.28	-0.08	-0.07	**0.92**	-0.05	-0.12	0.08	-0.15	0.34	0.22	-0.04
Lodging	-0.53	-0.07	0.51	-0.18	**0.67**	-0.25	0.21	-0.10	0.24	0.08	-0.36
PHS	0.32	0.12	-0.48	-0.21	-0.57	**0.83**	-0.38	0.10	-0.31	-0.54	0.30
Protein	-0.44	-0.12	0.53	0.26	0.44	-0.67	**0.80**	-0.39	0.41	0.39	-0.40
Softness	0.07	0.06	0.03	-0.25	-0.10	0.03	-0.23	**0.89**	-0.09	-0.18	0.19
Suc-SRC	-0.60	-0.23	0.40	0.36	0.48	-0.54	0.65	-0.21	**0.90**	0.18	-0.29
Test Weight	0.02	0.02	0.55	0.36	0.21	-0.73	0.62	-0.05	0.35	**0.83**	-0.16
Grain Yield	0.30	0.02	-0.48	-0.09	-0.60	0.59	-0.62	0.26	-0.50	-0.46	**0.72**



Supplementary Figure 4.1 Simple linear regression and linear fits of r_{GS} by r_P and r_{MAS} by r_P.

DISCUSSION

Comparison of Marker-based Prediction Methods

In this study, prediction accuracy using GS was superior to conventional-MAS in a wheat breeding population composed of multiple families. Averaged across the 13 traits, the average r_{GS} (0.59) was 28% higher than the average r_{MAS} (0.46; Table 4.1). Clearly, the GS approach of jointly estimating all marker effects was able to capture more of the genetic variance than the two-stage conventional-MAS approach that first selected significant markers and then estimated their effects. This result is consistent with previous empirical studies using biparental plant populations (Lorenzana and Bernardo 2009; Heffner et al. submitted) and multiple-family animal populations (Moser et al. 2009) and provides additional empirical evidence that GS will increase the accuracy of marker-based selection in plant breeding.

Four GS models that each had different prior assumptions for marker effect and variance distributions were used in this study to investigate the effect of these assumptions on r_{GS}. Despite these model differences, the mean r_{GS} across all traits ranged from only 0.58 to 0.60 (Table 4.1). This similarity in model performance is consistent with other empirical GS studies (Verbyla et al. 2009; VanRaden et al. 2009; Hayes et al. 2009a; Hayes et al. 2009b; Moser et al. 2009; Luan et al. 2009; Lorenzana and Bernardo 2009; Su et al. 2010; Heffner et al. submitted). Some differences between GS models; however, were significant (α=0.01) as r_{GS} standard errors were small because 100 TPs were sampled for each method. Nevertheless, differences were small in this study, and it was concluded that r_{GS} was generally not influenced by GS model choice.

The observed similarity in model performance is likely caused by two key factors: 1) effective population size (N_e) and 2) trait architecture (Daetwyler et al. 2010). First, a population's N_e determines the number of independent chromosomal segments (M_e), where each independent segment can be traced back to a single ancestor. Second, the trait architecture refers to the number of QTL (N_{QTL}) and the distribution of their effects. The M_e is important for *RR* because *RR* models genetic relationships by estimating the proportion of the genome that is identical between individuals. Thus, N_{QTL} does not impact *RR* performance unless N_{QTL} is very small (e.g. $N_{QTL} < 10$, Daetwyler et al. 2010). In contrast, the Bayesian models used in this study assign a portion of the marker effects equal to zero (*BC*), model unique marker variances (*BA*), or both (*BB*) to calculate GEBVs by targeting QTL. Consequently, the Bayesian models are favored over *RR* when either $N_{QTL} < M_e$ or a few QTL control a large portion of the genetic variance. This is because they heavily shrink or remove segments with no effect and/or differentially weight segments that contain QTL with small to large effects. When $N_{QTL} \geq M_e$ and all QTL effects are small (i.e. the infinitesimal model), these models are not expected to outperform *RR* because of the high probability that every chromosomal segment will contain a QTL. Therefore, the similarity in GS model performance suggests that the traits in this study were likely controlled by many small-effect QTL (Daetwyler et al. 2010).

Assuming $N_{QTL} \geq M_e$, an estimate of M_e can then be used to predict N_{QTL} underlying the traits analyzed in this study (Daetwyler et al. 2010). In this population, LD decayed below an r^2=0.2 at ~1.5 centiMorgans (cM) suggesting that the N_e was ~65 individuals using $1/(1 + 4N_ec)$, where c is the recombination frequency (Sved 1971). Using Goddard's (2009) theoretical approximation of $M_e = 1/4\ 2N_eL/\log(4N_eL)$, where L was the genome length of wheat in Morgans (~30M), M_e was estimated to be

~1,000. While this estimate of M_e is a rough approximation, this estimate along with the observed similarities of model performance suggests: 1) M_e was at least several hundred, 2) the 13 traits studied here are highly polygenic, 3) these traits will be best predicted by GS models that capture the effects of a large number of QTL.

Model selection should also consider relatedness between the TP and SP because genetic relationships deteriorate with each generation. Models that rely more on marker-QTL LD, e.g. Bayesian models like those used here, should produce higher accuracies than *RR* in scenarios were the TP and SP are separated by multiple generations (Habier et al. 2007; Zhong et al. 2009, Meuwissen 2009). To conduct marker-based selection on selection candidates that do not have phenotypes, the TP and SP will be separated by at least one selection cycle. However, in practice, the lag between TP and SP may be greater because several GS cycles may occur while selected lines go through seed increases and/or inbreeding cycles before entering the TP to update the model (e.g. Heffner et al. 2010). Also, TPs that span several cycles of selection and have greater genetic diversity would result in larger TPs, N_e, and, consequently M_e. In such cases, N_{QTL} may be significantly less than M_e, even for highly polygenic traits, suggesting that Bayesian models would be more accurate than *RR*. Nevertheless, updating the GS model each selection cycle should maintain genetic similarity between the TP and SP. Thus, a significant portion of r_{GS} in plant breeding programs may still result from capturing genetic relationships with markers.

Effects of Training Population Size and Marker Number

The rapid decrease in accuracy with reductions in N_{TP} for both conventional-MAS and GS (Fig. 4.1; Supplemental Table 4.1) was expected because it is well

known that increasing N_{TP} improves the estimation of marker effects (e.g. Knapp and Bridges, 1990) and accuracy (e.g. VanRaden et al. 2009, Hayes et al. 2009, Zhong et al. 2009). Meuwissen (2009) predicted that the TP size must approach 10*Ne*L in order to reach r_{GS}≈0.90 and 1*Ne*L in order to reach r_{GS}≈0.70-0.80 when TP and SP are unrelated, e.g. lines in SP come from a different breeding population or the TP and SP are separated by many generations. Using this approximation, N_{TP} for this population would need to be 9,000 to 15,000 for r_{GS}≈0.90 and 900 to 1,500 for r_{GS}≈0.70-0.80. The latter is more feasible for public plant breeding programs, but both seem possible for well-funded programs with extensive testing regimens (e.g. Eathington et al 2007; Sebastian et al 2010). In most cases, plant breeders will retrain models frequently for calculating GEBVs; thus, TP and SP will be closely related and high r_{GS} may be achieved with N_{TP} far smaller than 10*Ne*L (Meuwissen 2009). In addition to the effect of N_e and L, N_{TP} requirements will be affected by trait heritability, particularly when trait heritabilities are low (i.e. $h^2 < 0.40$; Hayes et al. 2009c). In this study, many traits had low H^2, N_{TP} was clearly below the requirements estimated above, and accuracy showed a near linear increase with increased N_{TP}. Therefore, considerable improvements in r_{GS} should have been achieved in this study if N_{TP} was increased.

Genome coverage is considered optimal when every QTL is in complete LD with at least one marker. This is becoming feasible for breeding programs as high-density SNP platforms and genotyping-by-sequencing are becoming affordable. The benefit of increasing marker densities was supported by this study as the highest r_{GS} was observed at the maximum N_M. Increases in r_{GS} were, however, gradual after N_M was increased beyond 384. The diminishing return from increasing N_M has been seen in other empirical studies (Lorenzana and Bernardo 2009; VanRaden et al. 2009;

Lorenz et al. 2011; Heffner et al. submitted) and suggests the advantage of high N_M will be realized only if N_{TP} scales with N_M (Muir 2007). Scaling N_{TP} with N_M should also improve r_{MAS}, but conventional-MAS is still expected to be suboptimal to GS for high N_M because conventional-MAS is less efficient for situations with few observations (small N_{TP}) many predictor variables (large N_M) and multicollinearity (e.g. Meuwissen et al. 2001). Accordingly, in this study the highest r_{MAS} was not achieved with the maximum N_M. Thus, scaling N_{TP} with N_M and using GS to manage situations of few observations and many predictors will be important for plant breeders to capture the benefits of affordable, high-density genotyping.

Phenotypic vs. Marker-based Prediction

Phenotypic prediction generally outperformed marker-based prediction, with r_P being 39% greater than conventional-MAS and 9% greater than GS when averaged across traits and prediction methods. Using the highest accuracy achieved for each trait, GS was, on average, 95% as accurate as PS (Table 4.1). Traits with high heritability should have higher accuracy than those with lower heritability, but the same is true for r_P. Therefore, there is little room for improving upon r_P when H^2 is high and PS is relatively inexpensive (Holland 2004, Hospital et al. 1997; Lande and Thompson 1990). The decreased benefit of r_{GS} and r_{MAS} as r_P increases was observed here, as the slopes of from linear regression for r_{GS} by r_P (0.81) and r_{MAS} by r_P (0.64) were both less than one (Supplementary Fig. 4.1). Nevertheless, GS was competitive with PS suggesting that GS will compare favorably to PS for many traits, especially because GS can shorten selection cycles (e.g. Schaeffer 2006; Wong and Bernardo 2008; Heffner et al. 2010), genotyping is becoming cheaper than phenotyping (e.g.

Bernardo, 2008), and N_{TP} become larger as data is accumulated each GS training cycle.

Accuracy of Predicting Net Merit

In this study, GS was comparable to PS on an individual trait basis; however, a breeder's primary goal is improving net merit – a single character determined by the sum of all economically important traits (e.g. VanRaden 2004). Accordingly, net merit can be estimated by an index of genetic values for all traits of interest where each trait is weighted by its relative economic importance (e.g. Lin 1978). Three different economic weighting indices ("yield", "quality", and "balanced") combining 11 traits (Table 4.3) and two selection index methods ("Smith-Hazel" and "base") were used to compare r_P and r_{GS} (Table 4.4). The Smith-Hazel index (Smith 1936; Hazel 1943) was used because it is considered an optimal index. It accounts for the genetic and phenotypic correlations between traits that would cause a simple phenotypic index to be an imperfect predictor of the actual breeding goal- additive genetic net merit (Lynch and Walsh 2008). A base index that ignores these parameters and simply weights traits by their economic values (Panse 1946; Brim et al. 1959; Williams 1962) was also used. While theoretically inferior to the Smith-Haze index, the base index can be favorable when large data sets are not available for accurate estimation of the phenotypic and genetic trait correlations (Williams 1962; Harris 1964). It was unknown which method was best for this population; therefore, accuracies were averaged across the six indices tested. This resulted in a major finding in this study: r_{GS} was 14% greater than r_P for net merit (Table 4.3). Only one index resulted in r_P being greater than r_{GS}: the Smith-Hazel index for yield where r_P =0.33 and r_{GS} =0.31. It was expected that GS would compare well with PS because r_{GS} was competitive

with r_P for low H^2 traits; however, this was easily the largest r_{GS}/r_P observed. As improving net merit is the primary goal for breeders, the high relative accuracy of GS for net merit was a very interesting result and more research on GS performance for net merit is needed.

Biparental-GS or Multi-family-GS

Empirical studies of GS in biparental populations have shown that biparental-GS will likely be superior to conventional-MAS and phenotypic selection in terms of gain per unit time and cost (Lorenzana and Bernardo 2009; Heffner et al. submitted). Biparental-GS has two clear attributes: 1) relatively low genotyping costs because extensive LD should make genome-wide marker coverage achievable with a few hundred markers, and 2) marker effect estimates will be "population specific" and should mitigate error caused by epistasis and rare alleles. Biparental-GS, however, requires phenotypes of lines from each cross prior to conducting GS, which may prolong the selection cycle and result in lower gains per year than multi-family-GS. Also, biparental-GS may not maximize r_{GS} because N_{TP} will likely be limited because a separate TP is created for each cross. This may explain why, for the same nine wheat quality traits, r_{GS}/H in this multi-family-GS study (~0.7) was greater than r_{GS}/H in a biparental-GS study (~0.5) by Heffner et al. (submitted). Such comparisons of accuracies between multi-family and biparental-GS studies, however, should be made with caution because differences in genetic variances could make comparisons misleading.

The benefit of using multi-family-GS to reduce cycle time was shown by a recent simulation study for wheat and maize by Heffner et al. (2010). Their results

suggest that if r_{GS} for net merit approaches 0.5, as reported here, multi-family-GS could increase gain per unit time and per unit cost by more than two to threefold in plant breeding. Of course, breeders will use TPs to predict untested progeny rather than perform cross-validation. Therefore, it remains unclear how r_{GS} reported here will compare to those achieved in actual multi-family-GS, biparental-GS, and conventional-MAS breeding programs.

Two main features of this study should make these results relevant to actual plant breeding programs: 1) the population consisted of current advanced breeding lines of an advanced-cycle breeding program, and 2) predictions were made using training data from one year and validated using lines that were not in the TP and phenotypes from another year to avoid inflation of r_{GS} caused by common GxE deviations in the TP and SP. If the results observed here and by Heffner et al. (2010) hold true, GS will dramatically increase gains from selection in plant breeding programs.

Conclusion

Advances in high-throughput genotyping, statistical models, and breeding methodology are making GS a promising tool for substantially increasing gains in animal and plant breeding. This was strongly supported by this study, as the observed prediction accuracies suggest that GS will be superior to both conventional-MAS and phenotypic selection in terms of gain per unit time and cost. Further research and software development is needed to enable widespread adoption of GS in plant breeding programs.

Acknowledgements

The authors express deep gratitude to the late Walter Federer for his advice on field designs and his many contributions to plant breeding during his career. The authors also thank Edward Souza and the USDA/ARS Soft Wheat Quality Laboratory in Wooster, Ohio for their evaluations of milling and baking quality and Aaron Lorenz, Jesse Poland, and Adam Famoso for their helpful suggestions. Support for the work of Elliot Heffner was provided by USDA National Needs Graduate Fellowship Competitive Grant No. 2005-38420-15785 from the National Institute of Food and Agriculture. Support for the work Jean-Luc Jannink was provided by USDA-NIFA grant No. 2009-85606-05701 and No. 2009-65300-05661. Additional funding for this research was provided by USDA – NIFA National Research Initiative CAP grant No. 2005-05130 and by Hatch 149-402.

REFERENCES

Anderson JA, Sorrells ME, Tanksley SD (1993) RFLP analysis of genomic regions associated with resistance to preharvest sprouting in wheat. Crop Science 33:453-459

Beavis WD (1998) QTL analyses: power, precision, and accuracy. In: Paterson AH (eds) Molecular Dissection of Complex Traits. Boca Raton, FL, CRC Press, pp 145–162

Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci 48:1649

Bernardo R, Yu J (2007) Prospects for Genomewide Selection for Quantitative Traits in Maize. Crop Sci 47:1082

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics 172:1165-1177

Brim C, Johnson HW, Cockerham CC (1959) Multiple selection criteria in soybeans. Agron J 51:42

Butler D, Cullis B, Gilmour A, Gogel B (2009) ASReml-R reference manual. Queensland Department of Primary Industries and Fisheries. Brisbane, Queensland

Calus M (2010) Genomic breeding value prediction: methods and procedures. Animal 4:157-164

Crossa J, de los Campos G, Pérez P, Gianola D, Atlin G, Burgueño J, Araus JL, Makumbi D, Yan J, Arief V (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics. doi: 10.1534/genetics.110.118521.

Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. Genetics 185:1021-1031

Dekkers JCM (2007) Prediction of response to marker-assisted and genomic selection using selection index theory. J Anim Breed Genet 124:331-341

Dekkers JCM, Garrick DJ, Fernando RL (2009) Use of high-density snp genotyping for genetic improvement of livestock. A short course organized by the Animal Breeding & Genetics Department of Animal Science Iowa State University June 1-10. http://www.ans.iastate.edu/stud/courses/short/2009/. Verified 17 June 2010

Eathington SR, Crosbie TM, Edwards MD, Reiter RS, Bull JK (2007) Molecular Markers in a Commercial Breeding Program. Crop Sci Crop Sci. 47: S154-S163

Federer WT (1956) Augmented (or hoonuiaku) designs. Hawaiian Planter's Records 55:191-208

Fraley C, Raftery AE (2006) MCLUST version 3 for R: Normal mixture modeling and model-based clustering. Technical Report No. 504, Department of Statistics, University of Washington. http://CRAN.R-project.org/package=mclust

Fraley C, Raftery AE (2002) Model-based clustering, discriminant analysis, and density estimation. Journal of the American Statistical Association 97:611-631

Gianola D, De Los Campos G, Hill WG, Manfredi E, Fernando R (2009) Additive genetic variability and the Bayesian alphabet. Genetics 183:347

Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136:245-257

Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson BV, Zusmanovich P, Sulem P, Thorlacius S, Gylfason A, Steinberg S (2008) Many sequence variants affecting diversity of adult human height. Nat Genet 40:609-615

Guttieri MJ, Souza EJ, Sneller C (2008) Nonstarch polysaccharides in wheat flour wire-cut cookie making. J Agric Food Chem 56:10927-10932

Habier D, Fernando R, Dekkers J (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389

Harris DL (1964) Expected and predicted progress from index selection involving estimates of population parameters. Biometrics 20, 46-72

Hartigan J, Wong M (1978) Algorithm AS: a K-mean clustering algorithm. Appl Stat 100-108

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009a) Invited review: Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci 92:433

Hayes BJ, Bowman PJ, Chamberlain AJ, Verbyla K, Goddard ME (2009b) Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet Sel Evol 41:51

Hayes BJ, Visscher PM, Goddard ME. (2009c). Increased accuracy of artificial selection by using the realized relationship matrix. Genetics Research 91, 47-60.

Hazel L (1943) The genetic basis for constructing selection indexes. Genetics 28:476

Heffner EL, Iwata H, Souza EJ, Jannink JL, Sorrells ME (submitted) Genomic selection across environments for grain quality traits in biparental wheat populations.

Heffner EL, Lorenz AJ, Jannink JL, Sorrells ME (2010) Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. Crop Sci 50:1681-1690

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

Holland JB (2004) Implementation of molecular markers for quantitative traits in breeding programs—challenges and opportunities. New Directions for a Diverse Planet: Proceedings for the 4th International Crop Science Congress, Brisbane, Australia 26

Henderson C (1984) Application of linear models in animal breeding. University of Guelph, Ontario

Hospital F, Moreau L, Lacoudre F, Charcosset A, Gallais A (1997) More on the efficiency of marker-assisted selection. Theor Appl Genet 95:1181-1189

Jaccoud D, Peng K, Feinstein D, Kilian A, Journals O (2001) Diversity Arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29:e25

Jannink JL, Bink MC, Jansen RC (2001) Using complex plant pedigrees to map valuable genes. Trends Plant Sci. 6:337

Jannink JL (2010) Dynamics of long-term genomic selection. Genetics Selection Evolution 42:35

Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Briefings in Functional Genomics 9:166

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709

Knapp S, Bridges W (1990) Using Molecular Markers to Estimate Quantitative Trait Locus Parameters: Power and Genetic Variances for Unreplicated and Replicated Progeny. Genetics 126:769-777

Lin CY (1978) Index selection for genetic improvement of quantitative characters. Theor Appl Genet 73:556-562

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743-756

Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, Eyheramendy S, Voight BF, Butler JL, Guiducci C et al. (2008) Identification of ten loci associated with height highlights new biological pathways in human growth. Nat. Genet. 40:584–591

Lorenz AJ, Chao S, Asoro FG, Heffner EL, Hayashi T, Iwata H, Smith KP, Sorrells ME, Jannink JL (2011) Genomic Selection in Plant Breeding: Knowledge and Prospects. Advances in Agronomy 110:In Press

Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor Appl Genet 120:151-161

Luan T, Woolliams JA, Lien S, Kent M, Svendsen M, Meuwissen THE (2009) The accuracy of genomic selection in Norwegian red cattle assessed by cross-validation. Genetics 183:1119

Lynch M, Walsh B (2008). Theory of selection index. Evolution and seletion of quantitative traits: II. Advanced topics in breeding and evolution. http://nitro.biosci.arizona.edu/zbook/NewVolume_2/newvol2.html. Verified 3 Sept 2010

Meuwissen THE (2009) Accuracy of breeding values of'unrelated' individuals predicted by dense SNP genotyping. Genetics Selection Evolution 41:35

Meuwissen THE, Hayes B, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819

Moreau L, Charcosset A, Hospital F, Gallais A (1998) Marker-Assisted Selection Efficiency in Populations of Finite Size. Genetics 148:1353-1365

Mohring J, Piepho HP (2009). Comparison of Weighting in Two-Stage Analysis of Plant Breeding Trials. Crop Sci 49(6): 1977-1988.

Moser G, Tier B, Crump RE, Khatkar MS, Raadsma HW (2009) A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. Genet Sel Evol 41:56

Munkvold JD, Tanaka J, Benscher D, Sorrells ME (2009) Mapping quantitative trait loci for preharvest sprouting resistance in white wheat. Theor Appl Genet 119:1223-1235

Panse V (1946) An application of the discriminant function for selection in poultry. Journal of Genetics 47:242-248

Piepho H (2009) Ridge regression and extensions for genomewide selection in maize. Crop Sci 49:1165

Podlich DW, Winkler CR, Cooper M (2004) Mapping as you go: an effective approach for marker-assisted selection of complex traits. Crop Sci 44:1560

R Development Core Team (2009) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94

Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet 123:218-223

Schön C, Utz S, Groh B, Truberg S, Openshaw S, Melchinger A (2004) QTL mapping based on resampling in a vast maize testcross experiment confirms the infinitesimal model of quantitative genetics for complex traits. Genetics 167:485-498

Sebastian S, Streit L, Stephens P, Thompson J, Hedges B, Fabrizius M, Soper J, Schmidt D, Kallem R, Hinds M (2010) Context-specific marker-assisted selection for improved grain yield in elite soybean populations. Crop Sci 50:1196

Smith HF (1936) A discriminant function for plant selection. Ann.Eugen 7:0-25O

Su G, Guldbrandtsen B, Gregersen V, Lund M (2010) Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population. J Dairy Sci 93:1175-1183

van Eeuwijk FA, Boer M, Totir LR, Bink M, Wright D, Winkler CR, Podlich D, Boldman K, Baumgarten A, Smalley M, Arbelbide M, ter Braak CJF, Cooper M (2010) Mixed model approaches for the identification of QTLs within a maize hybrid breeding program. TAG Theor. Appl. Genet. 120:429-440

VanRaden P, Van Tassell C, Wiggans G, Sonstegard T, Schnabel R, Taylor J, Schenkel F (2009) Invited review: Reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92:16

VanRaden P (2004) Invited Review: Selection on net merit to improve lifetime profit. J Dairy Sci 87:3125-3131

Verbyla KL, Bowman PJ, Hayes BJ, Goddard ME (2010) Sensitivity of genomic selection to using different prior distributions. BMC Proc 4 Suppl 1:S5

Visscher PM (2008) Sizing up human height variation. Nat Genet 40:489-490

Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, Freathy RM, Perry JRB, Stevens S, Hall AS (2008) Genome-wide association analysis identifies 20 loci that influence adult height. Nat Genet 40:575-583

Whittaker JC, Thompson R, Denham MC (2000) Marker-assisted selection using ridge regression. Genet Res 75:249-252

Williams J (1962) The evaluation of a selection index. Biometrics 18:375-393

Wong CK, Bernardo R (2008) Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. Theor Appl Genet 116:815-824

Xu S (2003) Theoretical basis of the Beavis effect. Genetics 165:2259

Xu Y, Crouch JH (2008) Marker-Assisted Selection in Plant Breeding: From Publications to Practice. Crop Sci 48:391

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42:565-569

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P (2007) An Arabidopsis example of association mapping in structured samples. PLoS Genet 3:e4

Zhong S, Dekkers J, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics 182:355