Multi-environment Validation and Efficiency of Genomewide Selection among Maize Doubled Haploids

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Lisa-Marie Krchov

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Rex Bernardo

November 2014

ACKNOWLEDGEMENTS

I am very grateful to my academic advisor, Dr. Rex Bernardo, for giving me the opportunity to work with him and for his guidance, mentorship and support throughout my research and writing the thesis.

I am also thankful to my advisory committee, Drs. James Anderson, Yang Da, Andrés Gordillo, and James Orf, who provided helpful advice and productive discussion. I am grateful for the assistance and friendship of fellow graduate students Emily Combs, Amy Jacobson, Lian Lian, Chris Schaefer, and Cathrine Ziyomo.

I would like to express my sincere thanks to Dr. Andrés Gordillo for the revision of my thesis and for his support throughout the whole time of my studies. His criticism and his patient explanations were always a very important advice to my work.

I am grateful to KWS SAAT AG for providing financial support for my fellowship and opportunities to have fruitful interactions with plant breeders and researchers. I am very thankful to Dr. Günter Seitz who gave me the opportunity to work on my Ph.D. in the USA.

I am also very thankful to AgReliant Genetics for their assistance with field experiments and data collection. I am deeply grateful for the opportunities to gain a lot of experience in the fields and practical breeding and to have productive interactions with plant breeders and researches. Special thanks to Harry Brokish, José Osorio, and Jialiang Chen.

Finally, I take this opportunity to express my deep gratitude to my beloved parents and my brother for their moral support throughout my whole life and patience during my study. I am deeply indebted to them for their continued support and unwavering faith in me. I also thank my boyfriend Henning for his patience and continuous support, during my stay in the USA as well as during the writing period.

ABSTRACT

Here I present two empirical studies on the effectiveness of genomewide selection within biparental populations in maize (*Zea mays* L.). To my knowledge, these are the first reported empirical studies on the effectiveness of genomewide selection compared to phenotypic selection, across multiple years and locations within individual biparental populations. In the first study, I report the results of a comparison of phenotypic selection and genomewide selection based on multiple years (2008–2012), multiple locations (six per year), and two testers. My objectives were to compare the accuracy of phenotypic selection, genomewide selection and an index for combined phenotypic and marker information. Phenotypic selection was always as accurate as or more accurate than genomewide selection. Selection based on marker and phenotypic information was slightly more accurate than genomewide selection or phenotypic selection alone. I concluded that for genomewide selection to be superior to phenotypic selection, the gains must not be measured in terms of the per-generation response with equal population sizes and selection intensities. In the second study I considered different total budgets and per-sample costs for phenotyping and genotyping in maize and assessed the observed and expected gains from the schemes for phenotypic selection and genomewide selection. My objectives were to determine whether or not genomewide selection is more efficient than phenotypic selection under a fixed budget and to give recommendations for implementing genomewide selection in a commercial maize breeding program. Whether or not genomewide selection was more efficient than phenotypic selection depended on the following factors: (1) accuracy of phenotypic

selection and of genomewide selection for the trait in a given cross; (2) size of the training population; (3) total budget; (4) costs of producing and genotyping a DH line; and (5) number of selected lines. The results indicated that in general prediction accuracies above 0.50 usually lead to higher efficiency in genomewide selection. In general, the relative efficiencies decreased as size of the training population decreased. A larger total budget and lower costs of DH production and of genotyping would enable larger test populations, thus leading to higher efficiencies of genomewide selection.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

CHAPTER 2

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 1

MULTI-ENVIRONMENT VALIDATION OF THE ACCURACY OF PHENOTYPIC AND GENOMEWIDE SELECTION WITHIN MAIZE BREEDING POPULATIONS

Published information is lacking on whether genomewide selection, based on a single tester and a single year of testing, can identify maize (*Zea mays* L.) lines that would perform well in multiple subsequent years and with multiple testers. Our objectives were to determine: (i) if phenotypic selection or genomewide selection is more predictive of maize performance in future environments and different testers; (ii) if combining both marker and phenotypic information is advantageous in selection; and (iii) the upwards bias in correlations between marker-predicted values and phenotypic values (r_{MP}) when cross validation across individuals and across environments is not performed. We evaluated four elite populations; each with 150 or 250 doubled haploid (DH) lines, in 18 environments in the U.S. Corn Belt. The DH lines were genotyped with 3072 single nucleotide polymorphism markers. The accuracy of genomewide selection was r_{MP} = 0.14 to 0.66 for grain yield and r_{MP} = 0.49 to 0.66 for moisture. Phenotypic selection was always as accurate as or more accurate than genomewide selection. The r_{MP} was lower when different testers were used for the training and test populations. Selection based on marker and phenotypic information was slightly more accurate than genomewide selection or phenotypic selection alone. The prediction accuracies were

higher when cross-validation across individuals and years was not performed than when cross-validation across individuals and years was performed. We conclude that for genomewide selection to be superior to phenotypic selection, the gains must not be measured in terms of the per-generation response with equal population sizes and selection intensities.

INTRODUCTION

Genomewide selection (or genomic selection) (Meuwissen et al., 2001) has emerged as a promising method for routine use in commercial maize (*Zea mays* L.) breeding. The effectiveness of genomewide selection is supported by sufficiently high correlations between marker-predicted and observed performance for yield and agronomic traits in maize (Lorenzana and Bernardo, 2009; Schulz-Streeck et al., 2012; Technow et al., 2012; Massman et al., 2013a; Jacobson et al., 2014; Lian et al., 2014) and in other plant species (Heffner et al., 2011; Asoro et al., 2011; Lorenz et al., 2012). Observed responses to genomewide selection were positive in a maize biparental cross (Massman et al., 2013a), in an adapted × exotic maize cross (Combs and Bernardo, 2013), and in a diverse collection of oat (*Avena sativa* L.) lines (Asoro et al., 2013). Furthermore, simulations have shown that the predicted response to selection based on both phenotypic and marker data was larger than the predicted response to phenotypic selection alone or to genomewide selection alone (Riedelsheimer and Melchinger, 2013).

The accuracy of genomewide prediction, which is measured as the correlation between marker-predicted values and phenotypic values (r_{MP}), depends on the quality of the phenotypic data in the training population. In most of the previous studies mentioned above, phenotypic data in the training populations were obtained from evaluations in only a single year and in a limited number of locations. For example, Massman et al. (2013b) conducted two cycles of genomewide selection after phenotyping the training population at four Minnesota locations in 2007. Combs and Bernardo (2013) conducted four cycles of genomewide selection after phenotyping the training population at four Minnesota locations in 2010. In addition, most previous studies of r_{MP} have considered the performance of maize lines when crossed with only one inbred tester (Albrecht et al., 2011; Zhao et al., 2011; Schulz-Streeck et al., 2012).

With phenotypic selection, initial yield evaluations of maize lines typically involve testcrossing the lines to only one tester and evaluating the testcrosses at multiple locations during a single year (Bernardo, 2010). Commercial maize breeders have become confident, from experience, that such level of testing is sufficient for the first year of phenotypic selection. However, published information is lacking to show that genomewide selection—based on a single tester and a single year of testing—can identify maize lines that would perform well in multiple subsequent years and with multiple testers. This lack of published information hinders the routine and wide use of genomewide selection in maize inbred and hybrid development programs.

Several studies have investigated the effectiveness of genomewide selection versus phenotypic selection. Empirical data from maize, barley (*Hordeum vulgare* L.),

wheat (*Triticum aestivum* L.), and *Arabidopsis* populations showed that phenotypic selection was expected to be more accurate than genomewide selection (Lorenzana and Bernardo, 2009; Heffner et al., 2011; Zhao et al., 2011). However, all of these previous studies did not involve empirical validation in independent years. To our knowledge, there are no published reports on multi-year validation of the accuracy of genomewide selection compared to phenotypic selection. This article reports the results of a comparison of phenotypic selection and genomewide selection based on multiple years (2008–2012), multiple locations (six per year), and two testers. The comparisons were made for four doubled haploid populations that were undergoing non-recurrent testcross selection within the AgReliant Genetics breeding program.

The availability of multiple-year and multiple-tester data enabled us to study the extent to which genomewide selection is predictive of maize performance in subsequent years and with other testers compared to phenotypic selection. Single-year data sets allow the estimation of r_{MP} among independent sets of individuals (via cross validation) but not among independent sets of years. In contrast, the multi-year data set in this study allowed us to assess r_{MP} with cross validation across individuals as well as across years. Our objectives in this study were to determine: (i) if phenotypic selection or genomewide selection is more predictive of maize performance in future environments and with different testers; (ii) if combining both marker and phenotypic information is more predictive of future performance compared with phenotypic selection alone and genomewide selection alone; and (iii) the upwards bias in r_{MP} when cross-validation across individuals and across environments is not performed.

MATERIALS AND METHODS

Phenotypic and Genotypic Data

We studied four biparental crosses (designated A/B, C/B, A/D, and E/F) between elite lines that belonged to the Iowa Stiff Stalk Synthetic heterotic group (Table 1). The numbers of doubled haploid (DH) lines developed in each cross were 150 in crosses A/B, C/D, and E/F and 250 in C/B. The DH lines in each of the four biparental crosses were testcrossed to one or two tester inbreds that belonged to an opposite heterotic group. All plant materials used in this study were proprietary to AgReliant Genetics, LLC (Westfield, IN).

Testcrosses in the A/B, C/B, and C/D populations were evaluated in yield trials in 2008, 2011, and 2012. Testcrosses in the E/F population were evaluated in 2009, 2011, and 2012. The experiments in 2008 and 2009 (referred to as year 1) corresponded to the first stage of testcross evaluation (TC_1) at AgReliant Genetics, whereas the experiments in 2011 and 2012 (designated as year 2 and year 3, respectively) were special validation experiments to evaluate the same set of lines previously evaluated in year 1 at the TC_1 level. The A/B and C/B populations were both evaluated in crosses with two testers in years 2 and 3 [designated as (1) for tester T1 and (2) for tester T2], whereas C/D and E/F were evaluated in crosses with only one tester in years 2 and 3 (Table 1).

In year 1, all testcrosses were evaluated at six locations in the U.S. Corn Belt in unreplicated incomplete block designs. In years 2 and 3, all testcross progenies of the four populations were evaluated in the U.S. Corn Belt in an alpha lattice design at six

locations. Crosses A/B(1), C/B(1), and E/F were evaluated in all three years in Hopedale and Farmersville, IL, Ft. Branch and Greentown, IN, Oskaloosa, IA, and Kearney, NE. Crosses A/B(2), C/B(2), and A/D were evaluated in all three years in Minonk, Mendota, and St. Anne, IL, Ft. Wayne, IN, Jamaica, IA, and Stanton, NE.

At each of the six locations, trials were performed with two replications and consisted of 44 blocks each with eight entries. The trials for C/B consisted of 68 blocks with eight entries. Two-row-plots each 6.40 m long and spaced 0.76 m apart were grown. Grain yield (t ha^{-1}) and grain moisture at harvest (g kg^{-1}) were recorded in each plot.

All DH lines were genotyped, using the Illumina iScan platform, with 3072 SNP markers distributed across the genome. Excluding markers with more than 10% missing values or with more than 5% heterozygosity resulted in numbers of polymorphic markers (N_M) that ranged from N_M = 467 for cross C/B to N_M = 1125 for cross C/D (Table 1). Nine to 14 DH lines per cross were discarded from the analysis due to more than 10% missing values, more than 5% heterozygosity, or more than 5% of nonparental marker alleles in the DH line. The genetic similarity between the parental lines and the testers was calculated as the simple matching coefficient (Sokal and Michener, 1958).

Analysis of Variance Components

Variance components were computed considering all effects in the linear model (other than the grand mean) as random. Estimates of the following variance components were computed as described by Searle (1971): σ^2_g (genotypic variance), σ^2_{gl} (genotype-by-

location interaction variance), σ^2_{gy} (genotype-by-year interaction variance), σ^2_{gly} (genotype-by-location-by-year interaction variance), and σ^2_e (error variance). Variance components were estimated across all three years. Within each of the four crosses and for each tester, heritability (h^2) on an entry-mean basis was calculated for each trait as

$$h^2 = \frac{\sigma^2_g}{\sigma^2_g + \frac{\sigma^2_{gl}}{L} + \frac{\sigma^2_{gy}}{Y} + \frac{\sigma^2_{gly}}{LY} + \frac{\sigma^2_e}{LYR}} \quad [\mathbf{1}]$$

where L was the number of locations, Y was the number of years, and R was the number of replications. All data analyses were done in SAS version 9.3 (SAS Institute Inc., Cary, NC, 2004).

Accuracy of Phenotypic and Genomewide Selection

The phenotypic dataset for the *N* individuals in each cross was divided into a training population and a test population. As described later, five-fold cross validation was used. The training population corresponded to the marker and phenotypic data for individuals 1 to *n* in year i, where *n* was equal to (4/5) *N*. The test population corresponded to the phenotypic data for individuals *n*+1 to *N* in years j+k (Fig.1). The data from each year were considered as the training population and the data from the two remaining years were considered as the corresponding test population. In other words, the data for the training and test populations (in parentheses) were from the following years: Year 1 (Year 2+3); Year 2 (Year 1+3); and Year 3 (Year 1+2). Analyses were carried out for all three combinations and the mean prediction accuracies were calculated across all three combinations. As described later, each of the five subsets of $n = N/5$ lines was used as

the test population. The number of years was Y = 1 for calculating h^2 in the training population, and Y = 2 for calculating h^2 in the test population.

All genomewide predictions were based on the training population (i.e., individuals 1 to *n* in year i). To evaluate the effectiveness of phenotypic selection and genomewide selection, we compared five correlations (Fig.1):

(i) For phenotypic selection, r_P was the Pearson correlation between the observed performance of individuals *n*+1 to *N* in year i and the observed performance of individuals *n*+1 to *N* in years j+k.

(ii) For genomewide selection, r_{MP} was the correlation between the marker-predicted performance of individuals *n*+1 to *N* in year i and the observed performance of individuals *n*+1 to *N* in years j+k.

(iii) For genomewide selection without cross-validation across individuals, $r_{MP(N)}$ was the correlation between the marker-predicted performance of individuals 1 to *n* in year i and the observed performance of individuals 1 to *n* in years j+k. The difference between $r_{MP(N)}$ and r_{MP} indicated the upward bias when the same individuals were used to train and test the prediction model.

(iv) For genomewide selection without cross validation across years, $r_{MP(Y)}$ was the correlation between the marker-predicted performance of individuals *n*+1 to *N* in year i and the observed performance of individuals *n*+1 to *N* in year i. The difference between $r_{MP(Y)}$ and r_{MP} indicated the upward bias in r_{MP} when the same years were used to train and test the prediction model.

(v) For genomewide selection without cross validation across individuals and years, $r_{MP(YN)}$ was the correlation between the marker-predicted performance of individuals 1 to *N* in year i and the observed performance of individuals 1 to *N* in year i.

While Fig.1 illustrates the correlations with one partitioning of the *N* individuals into the training population and test population, the correlations were obtained by repeated partitioning of the data. The *N* DH lines were randomly divided into five subsets of *N*/5 lines each in a five-fold cross validation procedure. The r_P, r_{MP}, $r_{MP(N)}$, $r_{MP(Y)}$, and $r_{MP(YN)}$ values were calculated using each of the five subsets as the *N/5* individuals in the test population. The remaining $n = (4/5)N$ individuals were used as the training population. This procedure was repeated 20 times for each partitioning of the years into the training and test populations. The mean correlations were calculated across all repeats.

Genomewide Selection Model

Genomewide marker effects for calculating r_{MP}, $r_{MP(N)}$, $r_{MP(Y)}$, and $r_{MP(YN)}$ were obtained by ridge-regression best linear unbiased prediction (RR-BLUP), which assumes that all markers have the same variance and the predictions of all marker effects are equally shrunken toward zero by the penalty term λ (Whittaker et al., 2000). We used a mixed-model formulation of ridge regression according to Piepho (2009), which allows estimating the components of variance and, hence, the penalty parameter λ by restricted maximum likelihood. The penalty parameter λ was defined as $\lambda = V_R/V_M$, where V_M is

the variance explained by the markers and V_R was the residual variance associated with the entry means. The linear model for RR-BLUP was

$$\mathbf{y} = \mu\mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \qquad [2]$$

where **y** was an $n \times 1$ vector of testcross phenotypic means of the DH lines; $\mathbf{1}_n$ was an $n \times 1$ vector with all elements equal to 1; μ was the overall testcross mean of the DH lines; **X** was an $n \times N_M$ design matrix with elements equal to 2 if the DH line was homozygous for the marker allele from the first parental inbred and 0 if the DH line was homozygous for the marker allele from the second parental inbred; **β** was an $N_M \times 1$ vector of marker effects; and **e** was an $n \times 1$ vector of residual effects.

Genomewide Selection across Different Testers

To investigate the effect of different testers in genomewide selection, we compared the r_{MP} and r_P when using the same tester in the training and test population versus a different tester in the training population and test population. This procedure was done for crosses A/B and C/B (Table 1).

In addition, we evaluated the prediction accuracy for general combining ability (GCA) across the two testers (T1 and T2) in crosses A/B and C/B. The GCA values of the DH lines were estimated as the mean performance of each DH line when crossed with both T1 and T2. The training population had data for both testers, and the r_{MP} values were calculated in two different ways: (1) for a training population in which lines 1 to n were crossed to T1 and lines n +1 to N were crossed to T2; and (2) for a training

population in which lines 1 to *n* were crossed to T1 and as well as to T2. Both training populations were therefore of the same size. The r_{MP} was then the correlation between the predicted and observed GCA in the test population.

Selection with both Phenotypic and Genomewide Marker Data

Marker and phenotypic information were combined in an index (Lande and Thompson (1990). Each DH line had a mean phenotypic value (P) as well as a marker-predicted value (M) obtained from Eq.1. Two weights that summed to 1 were assigned to P (b_P) and M (b_M), and the index was calculated as $I = b_P P + b_M M$. The weights were calculated according to Lande and Thompson (1990) as $b_M/b_P = (1/h^2 - 1)/(1 - p)$, where p was the proportion of the genetic variance explained by the markers and was calculated as $(r_{MP})^2/h^2$.

To assess the efficiency of this index we calculated the Pearson correlation (r_{IP}) between the index of individuals n+1 to N in year i and the observed performance of individuals n+1 to N in years j+k. To compare the accuracy of phenotypic selection, genomewide selection, and selection based on the above index, we compared r_P, r_{MP} and the r_{IP}. All pairwise differences between the correlation coefficients r_P, r_{MP}, $r_{MP(N)}$, $r_{MP(YN)}$, and r_{IP} were tested for significance via a Fisher z-transformation for correlation coefficients.

RESULTS AND DISCUSSION

Heritability and Genotype-Environment Interactions

For both grain yield and moisture, the estimates of genotypic variance were significant (P = 0.05) for all six combinations of crosses and testers (Table 2). For grain yield, all other variance components were likewise significant except for the variance component for genotype × year interaction for cross A/B in combination with tester T1, the genotype × location interaction for cross C/B in combination with tester T1, and the genotype × year × location interaction for cross A/B in combination with tester T2. For moisture, all variance component estimates were significant.

The genotype-by-year interaction variance (σ^2_{gy}) was generally larger than the genotype-by-location interaction variance (σ^2_{gl}) for both traits. The large genotype-by-year interaction was probably due to the differences in temperature and precipitation among the three years. The 2012 season had much higher temperatures and received less precipitation than average, whereas the 2008 and 2011 seasons had temperatures and precipitation that were close to normal.

The heritability (h^2) in the training population (single year) ranged from 0.41 to 0.71 for grain yield and from 0.70 to 0.87 for moisture across the six combinations between biparental crosses and testers (Table 3). The h^2 in the test population (two years) ranged from 0.39 to 0.78 for grain yield and from 0.68 to 0.86 for moisture.

Accuracy of Genomewide Selection and Phenotypic Selection

The r_{MP} values ranged from 0.15 to 0.66 for grain yield and from 0.49 to 0.66 for moisture (Table 3). The r_P values ranged from 0.21 to 0.74 for grain yield and from 0.53 to 0.76 for moisture. All r_{MP} and r_P values were significantly different from zero (P = 0.05). For grain yield, only one cross (E/F) had a significant difference between r_{MP} (0.19) and r_P (0.34). For moisture, the differences between r_{MP} and r_P were significant for crosses A/B and C/B in combination with tester T1, and for cross E/F. In these three instances, r_P was larger than r_{MP}.

Overall, our results showed that phenotypic selection was always as good as or better than genomewide selection. In other words, genomewide selection was never better than phenotypic selection. Our findings are in agreement with previous studies on genomewide selection. In biparental populations of maize, barley, and Arabidopsis, the accuracy of phenotypic selection was 0.1 to 0.5 times as large as the accuracy of genomewide selection (Lorenzana and Bernardo, 2009). In six European maize populations, phenotypic selection for grain yield and moisture was more accurate than genomewide selection across all six populations (Zhao et al., 2011). Massman et al. (2013a) showed that phenotypic selection was more accurate than genomewide selection across four traits in 14 maize populations. In two wheat biparental populations, phenotypic selection was more accurate than genomewide selection for nine traits (Heffner et al., 2011).

Yet, prediction accuracy is not the only factor that determines the selection gain. The response to selection must not necessarily be measured in terms of the per-

generation response with equal population sizes and selection intensities, but in terms of the response per unit of time and budget. Off-season nurseries, low-cost genotyping and greenhouses can allow genomewide selection to outperform phenotypic selection on the basis of response per unit time and cost (Bernardo and Yu, 2007; Wong and Bernardo, 2008; Heffner et al., 2010). In their study, Lorenzana and Bernardo (2009) suggested that the response from three cycles of genomewide selection would be around 1.5 times the gain from one cycle of phenotypic selection. On the basis of cost and time, Heffner et al. (2010) concluded a prediction accuracy of 0.20 would make genomewide selection superior to phenotypic selection. In contrast, Riedelsheimer and Melchinger (2013) cautioned that under certain situations, the gain from selection for genomewide selection can be less than for phenotypic selection. They illustrated that the budget and the cost for genotyping are the factors with the most influence on whether or not genomewide selection is superior to phenotypic selection.

Genomewide Selection with Different Testers

The r_{MP} and r_P values were smaller when the tester differed between the training and test population (Table 4). Compared to having the same tester, having different testers in the training and test population led to a significant decrease in r_{MP} and r_P for both crosses and traits. The reduction in accuracy due to different testers was significantly higher for phenotypic selection than for genomewide selection in cross A/B for grain yield and in cross C/B for moisture.

The germplasm used likely contributed to the reduction in accuracy when the testers differed between training and test populations. The A/B and C/B crosses

belonged to the Iowa Stiff Stalk Synthetic heterotic group and had been previously selected for hybrid performance when crossed with a line from the Lancaster SureCrop heterotic group. Tester T1 was closely related to this Lancaster SureCrop line. This implied that the r_{MP} and r_P values would be higher with T1 as the tester than with T2 as the tester. We found this to be true by calculating r_{MP} and r_P for crosses A/B and C/B both in combination with T2 with year 2 as the training population and year 3 as the test population. In both crosses, the correlations were lower when T2 was the tester than when T1 was the tester. Our results depended on the testers used, and a different set of testers may lead to different results regarding the relative values of r_{MP} with the same and with different testers.

If the goal in testcross evaluation is to select lines that perform well with a particular tester, using this particular tester in the training population leads to the highest r_{MP} (Table 4). But if the main goal is to identify lines with good GCA, the use of multiple testers may be preferable. The r_{MP} values for GCA effects were intermediate to the r_{MP} values for the same tester and for different testers (Table 4). The r_{MP} values did not differ when training population comprised lines 1 to n testcrossed to T1 and lines n+1 to N testcrossed to T2, versus lines 1 to n testcrossed to both T1 and T2.

Index with Combined Marker and Phenotypic Information

Using an index combining phenotypic and marker information lead to numerically higher prediction accuracies than genomewide selection alone and phenotypic selection alone in all crosses and for both traits. However, index selection was statistically

superior in only two cases: moisture in cross A/B in combination with tester T2, and moisture in cross C/D (Table 3). The differences between r_{IP} values and $r_{MP(N)}$ or r_P values were not significant (P = 0.05) for grain yield for all four crosses. The differences between r_{IP} and $r_{MP(N)}$ values ranged from 0.0 to 0.04 for grain yield and from 0.0 to 0.1 for moisture. The differences between r_{IP} and r_P values ranged from 0.00 to 0.09 for grain yield and from 0.01 to 0.16 for moisture.

While the differences were statistically significant in only two cases, the values of r_{IP} were numerically higher than or equal to the values of r_P or $r_{MP(N)}$ across all traits and all populations [except for moisture in A/B (T1)] (Table 3). In agreement with these results, Riedelsheimer and Melchinger (2013) found in their model calculations that the use of an index combining phenotypic and predicted values was especially beneficial under limited resources and large genotype × environment interactions. Since both phenotypic and genotypic data are available for the training dataset, we recommend an index that integrates both types of data when selecting the best candidates from the training population.

Overestimation of Prediction Accuracy due to the Lack of Cross-Validation across Environments or Individuals

In calculating genomewide prediction models, cross-validation is typically done across individuals but not across environments. Our results showed that a lack of cross-validation across environments could lead to an upwards bias in the estimate of prediction accuracy. The correlation between marker-predicted values and observed values was significantly greater when cross-validation across years was not performed

[$r_{MP(Y)}$] than when cross-validation across years was performed (r_{MP}). This result was observed for both grain yield and moisture in all crosses, except for cross A/B in combination with tester T1 for both traits (Table 3).

A difference between r_{MP} and $r_{MP(Y)}$ is expected if genotype-by-year interaction (σ^2_{gy}) is important. Otherwise, if σ^2_{gy} is small or nonsignificant, it should not matter whether or not cross-validation is performed across years. The A/B (T1) cross had a nonsignificant σ^2_{gy} for grain yield and the smallest σ^2_{gy} for moisture among the crosses (Table 2), and these two cases corresponded to those for which r_{MP} and $r_{MP(Y)}$ did not differ significantly (Table 3). Overall, our results indicate if genotype-by-year interaction is important and cross-validation across years is not performed, we obtain an overly optimistic estimate of prediction accuracy.

As expected, a lack of cross-validation across years and individuals led to inflated prediction accuracies [$r_{MP(YN)}$]. The $r_{MP(YN)}$ values were greater than the $r_{MP(Y)}$ values for grain yield and moisture, and the differences between $r_{MP(YN)}$ and $r_{MP(Y)}$ were significant for both traits across all crosses (Table 3). The reduction from $r_{MP(YN)}$ to $r_{MP(Y)}$ ranged from 0.13 to 0.29 for grain yield and from 0.07 to 0.22 for moisture. Likewise, a lack of cross-validation across individuals led to higher prediction accuracies [$r_{MP(N)}$]. As expected, the r_{MP} values were always smaller than the $r_{MP(N)}$ values for grain yield and moisture (Table 3). For grain yield, the difference between r_{MP} and $r_{MP(N)}$ was significant for three out of six combinations between crosses and testers (crosses A/B and C/B in combination with tester T1 and cross E/F). For moisture, the difference between r_{MP} and $r_{MP(N)}$ was significant for four out of the six combinations

between crosses and testers (for crosses A/B and C/B in combination with tester T1, and for cross E/F). The reduction from $r_{MP(N)}$ to r_{MP} ranged from 0.03 to 0.15 for grain yield and from 0.05 to 0.13 for moisture. In a simulation study, Lorenz (2013) likewise found that r_{MP} was lower than $r_{MP(N)}$. He reasoned that the reduction of prediction accuracy for lines not included in the training population was due to the generation of spurious linkage disequilibrium between unlinked markers. This spurious linkage disequilibrium is generated by the sampling effect when drawing individuals from finite population sizes.

In practice, a breeder wishes to select the best lines within the population as a whole. The different prediction accuracies within the training population [$r_{MP(N)}$] and within the test population (r_{MP}) suggest that selection should be done independently in the two sets of lines. The larger differences between $r_{MP(YN)}$ and $r_{MP(Y)}$ than between $r_{MP(N)}$ and r_{MP} indicated that if cross-validation across years is not done (which is the case in most studies), the predicted gains from selection within the training population are much too optimistic. Regardless of whether or not cross-validation across years is done, the effects of the difference between $r_{MP(N)}$ and r_{MP} or between $r_{MP(YN)}$ and $r_{MP(Y)}$ should be less pronounced if selection among the lines in the training population is independent from selection among the lines in the test population. Moreover, the difference between $r_{MP(Y)}$ and $r_{MP(YN)}$ was consistently larger than the difference between $r_{MP(Y)}$ and r_{MP}, which indicates that the relative merit of predicting untested lines is underestimated when cross-validations are not performed across years. The latter increases the relative merit of predicting untested lines. In a follow up study, we are

investigating the gains per unit cost when lines within each of the population in the current study are portioned into training populations and test populations of different sizes.

Table 1. Maize biparental crosses used for multi-year validation of genomewide selection.

Cross	Marker similarity between parental lines	Tester			Marker similarity between testers	Number of doubled haploids	Number of markers
		Year 1	Year 2	Year 3			
A/B	0.69	T1	T1,T2	T1,T2	0.77	135	467
C/B	0.78	T1	T1,T2	T1,T2	0.77	237	627
C/D	0.53	T3	T1	T1	0.72	139	1125
E/F	0.69	T4	T1	T1	0.87	141	508

Table 2. Variance component estimates, heritability, and overall means for grain yield and moisture in four maize biparental crosses crossed to one or two testers (T1 and T2).

	Cross					
Parameter	A/B (T1)	A/B (T2)	C/B (T1)	C/B (T2)	C/D	E/F
	Grain yield (t ha^{-1})					
σ^2_g †	2.79*	1.51*	0.64*	0.49*	2.30*	0.79*
σ^2_{gl}	0	0.53*	0.02	0	0	0.33*
σ^2_{gy}	0.04	1.18*	0.51*	1.36*	1.46*	0.30*
σ^2_{yl}	11.54*	4.76*	6.90*	0.18*	32.58*	15.30*
σ^2_{gyl}	1.31*	0.23	1.68*	2.04*	2.37*	0
σ^2_e	9.60*	10.84*	8.83*	10.76*	11.15*	12.45*
h^2	0.94	0.72	0.69	0.45	0.78	0.77
Mean	12.01	11.01	12.48	9.89	9.24	12.20
	Moisture (g kg^{-1})					
σ^2_g	5.30*	3.40*	2.40*	2.40*	20.50*	6.50*
σ^2_{gl}	0.50*	2.10*	0.40*	1.20*	2.70*	1.80*
σ^2_{gy}	0.90*	1.70*	2.30*	4.00*	3.00*	1.20*
σ^2_{yl}	71.00*	46.7*	64.90*	23.80*	49.10*	75.20*
σ^2_{gyl}	0.90*	1.40*	1.80*	2.50*	5.70*	2.30*
σ^2_e	8.4*	7.80*	9.00*	10.50*	13.00*	10.30*
h^2	0.92	0.80	0.72	0.62	0.93	0.91
Mean	199	197	200	203	194	231

* Significant at P = 0.05.

† σ^2_g, genotypic variance; σ^2_{gl}, genotype-by-location interaction variance; σ^2_{gy}, genotype-by-year interaction variance; σ^2_{yl}, year-by-location interaction variance; σ^2_{gyl}, genotype-by-year-by-location interaction variance; σ^2_e, within-location error variance; h^2, entry-mean heritability across all environments.

Table 3. Heritabilities and prediction accuracies for grain yield and moisture in four maize biparental crosses crossed to one or two testers (T1 and T2).

Parameter	Grain yield						Moisture					
	A/B (T1)	A/B (T2)	C/B (T1)	C/B (T2)	C/D	E/F	A/B (T1)	A/B (T2)	C/B (T1)	C/B (T2)	C/D	E/F
h^2 (Training)[†]	0.71	0.62	0.51	0.41	0.57	0.45	0.81	0.72	0.79	0.70	0.73	0.87
h^2 (Test)	0.78	0.65	0.59	0.39	0.55	0.51	0.86	0.74	0.82	0.68	0.76	0.88
r_P[‡]	0.74	0.43	0.50	0.21	0.39	0.34	0.74	0.53	0.68	0.55	0.66	0.76
r_{MP}	0.66	0.49	0.46	0.15	0.32	0.19	0.64	0.49	0.60	0.55	0.66	0.64
$r_{MP(N)}$	0.73	0.52	0.54	0.20	0.41	0.34	0.77	0.56	0.69	0.60	0.72	0.76
$r_{MP(Y)}$	0.64	0.59	0.55	0.43	0.52	0.37	0.66	0.66	0.72	0.7	0.77	0.78
$r_{MP(YN)}$	0.85	0.83	0.68	0.63	0.81	0.68	0.88	0.87	0.79	0.78	0.89	0.90
r_{IP}	0.74	0.52	0.57	0.23	0.45	0.38	0.76	0.66	0.69	0.60	0.82	0.80
r_P vs. r_{MP}	NS	NS	NS	NS	NS	*	*	NS	*	NS	NS	*
r_{MP} vs. $r_{MP(N)}$	*	NS	*	NS	NS	*	*	NS	*	NS	*	*
r_{MP} vs. $r_{MP(Y)}$	NS	*	*	*	*	*	NS	*	*	*	*	*
$r_{MP(Y)}$ vs. $r_{MP(YN)}$	*	*	*	*	*	*	*	*	*	*	*	*
r_P vs. r_{IP}	NS	NS	NS	NS	NS	NS	NS	*	NS	NS	*	NS
r_{MP} vs. r_{IP}	NS	NS	NS	NS	NS	NS	NS	*	NS	NS	*	NS

* Significant at P = 0.05; NS, nonsignificant at P = 0.05.

† h^2 (Training), mean entry-mean heritability in the training population; h^2 (Test), mean entry-mean heritability in the test population.

‡ r_P, Pearson correlation between the observed performance of lines n+1 to N in year i and the observed performance of lines n+1 to N in years j+k; r_{MP}, Pearson correlation between the marker-predicted performance of lines n+1 to N in year i and the observed performance of lines n+1 to N in years j+k; $r_{MP(N)}$, Pearson correlation between the marker-predicted performance of lines 1 to n in year i and the observed performance of lines 1 to n in years j+k; $r_{MP(Y)}$, Pearson correlation between the marker-predicted performance of lines n+1 to N in year i and the observed performance of lines n+1 to N in year i; $r_{MP(YN)}$, Pearson correlation between the marker-predicted performance of lines 1 to N in year i and the observed performance of lines 1 to N in year i; r_{IP}, Pearson correlation between the index of lines n+1 to N in year i and the observed performance of lines n+1 to N in years j+k.

Table 4. Prediction accuracies (r_P and r_{MP}) for tester 1, tester 2, and both testers pooled in the training population for crosses A/B and C/B.

Correlation	Prediction of:	Grain yield		Moisture	
		A/B	C/B	A/B	C/B
r_P †	Performance with the same tester	0.74a‡	0.5a	0.74a	0.68a
	Performance with a different tester	0.43b	0.21b	0.53b	0.55b
r_{MP}	Performance with the same tester	0.66a	0.46a	0.64a	0.60a
	Performance with a different tester	0.49a	0.15b	0.49a	0.55a
	General combining ability (1)§	0.56a	0.25b	0.56a	0.61a
	General combining ability (2)¶	0.57a	0.28b	0.54a	0.59a

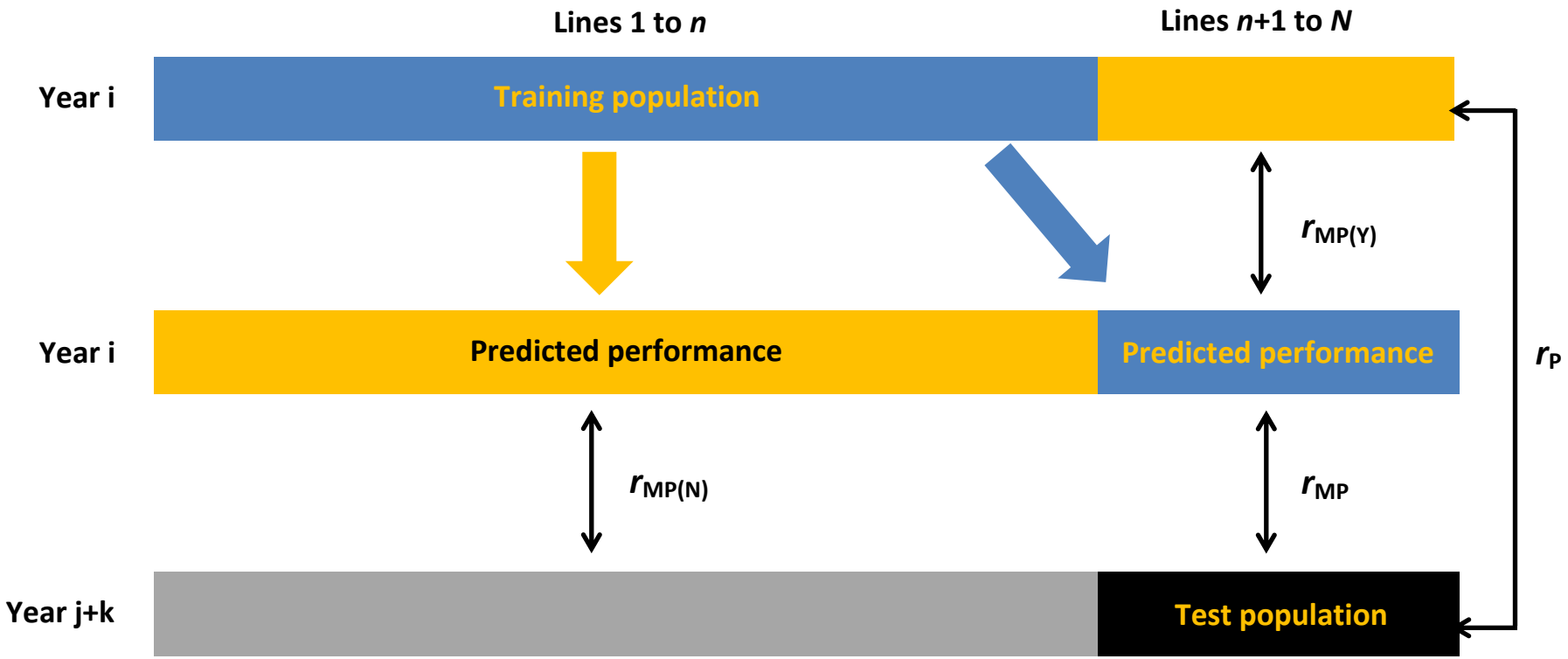
† r_P, Pearson correlation between the observed performance of lines n+1 to N in year i and the observed performance of lines n+1 to N in years j+k; r_{MP}, Pearson correlation between the marker-predicted performance of lines n+1 to N in year i and the observed performance of lines n+1 to N in years j+k.

‡ Within each correlation and within each column, correlations with a common letter were not significantly different (P = 0.05).

§ Training population consisted of lines 1 to *n* crossed to T1 and lines *n* to *N* crossed to T2, where *n* was half of the total number of lines.

¶ Training population consisted of lines 1 to *n* crossed to both T1 and T2, where *n* was half of the total number of lines.

Figure 1. Scheme for calculating r_P, r_{MP}, $r_{MP(N)}$, and $r_{MP(Y)}$. r_P was the correlation between the observed performance of individuals n+1 to N in year i and the observed performance of individuals n+1 to N in years j+k, r_{MP} was the correlation between the marker-predicted performance of individuals n+1 to N in year i and the observed performance of individuals n+1 to N in years j+k, $r_{MP(N)}$ was the correlation between the marker-predicted performance of individuals 1 to n in year i and the observed performance of individuals 1 to n in years j+k, and $r_{MP(Y)}$ was the correlation between the marker-predicted performance of individuals n+1 to N in year i and the observed performance of individuals n+1 to N in year i.


Lines 1 to n
Lines n+1 to N
Year i
Training population
$r_{MP(Y)}$
Year i
Predicted performance
Predicted performance
r_P
$r_{MP(N)}$
r_{MP}
Year j+k
Test population

CHAPTER 2

EFFICIENCY OF GENOMEWIDE SELECTION FOR TESTCROSS PERFORMANCE OF DOUBLED HAPLOID LINES IN A MAIZE BREEDING PROGRAM

When genomewide selection is less accurate than phenotypic selection, this lower accuracy could be offset by increasing the selection intensity in genomewide selection. In maize (*Zea mays* L.), this increase in selection intensity is possible because genotyping is cheaper than phenotyping. Our objectives were to (i) determine whether or not genomewide selection is more efficient than phenotypic selection under a fixed budget, and (ii) give recommendations for implementing genomewide selection in a line development program in maize. On the basis of empirical prediction accuracies for genomewide selection across multiple populations, testers, years, and locations, we assessed the observed and expected gains when a subset of doubled haploid (DH) lines from a biparental cross are genotyped and phenotyped in testcross combination, and are used as a training population to predict the performance of a remaining set of DH lines that have been genotyped but not phenotyped. For different total budgets and per-sample costs of phenotyping and genotyping, we identified the conditions that led to the maximum relative efficiency (RE_{Max}) of genomewide selection over phenotypic selection. We found that a training population of 60 to 80 lines most often led to $RE_{Max} > 1$. For RE_{Max} to exceed 1, prediction accuracies typically needed to be at least 0.50.

The RE_{Max} values were often the highest when selection was not stringent and when the test population included around 125 lines or more. However, crosses differed in their RE_{Max} values and our general recommendations do not guarantee that genomewide selection will always be efficient.

INTRODUCTION

In a commercial maize (*Zea mays* L.) breeding program, new lines are usually developed by crossing two elite inbred lines from the same heterotic group, inducing haploids from F_1 plants, and doubling the chromosomes of haploids to produce doubled haploid (DH) lines. Suppose a total of *N* DH lines have been induced in a biparental cross. In phenotypic selection, the DH lines are crossed with a tester from an opposite heterotic group and the testcrosses are evaluated in field trials for multiple traits (Fig. 1).

An alternative breeding approach is to apply genomewide selection (GWS) among a portion of the DH lines. In this scheme, only *n* of the DH lines are evaluated in field trials and these lines serve as the training population for predicting the performance of the remaining *N*-*n* DH lines (test population) that are not phenotyped (Fig. 1). In other words, genomewide predictions are substituted for phenotyping for a portion of the population (Heslot et. al, 2014). This simple scheme for GWS during line development will be efficient if two conditions are met: (1) genotyping is cheaper and more convenient than phenotyping, and (2) genomewide predictions are accurate.

These two conditions have been largely met in maize. The costs of genotyping continue to decrease whereas the costs of field testing in maize are either stagnant or

continue to increase. In maize, genotyping for several hundred singe nucleotide polymorphism (SNP) markers currently costs around $20-30 (http://epigenome.usc.edu/services/microarray.htm) whereas the costs of field trials at 6-8 locations is around $90-120 (http://www.techservicespro.com/Sales/Corn.htm). In addition, genotyping is more convenient than phenotyping because it is not subject to seasonal variation and it requires less effort than field experiments. Empirical studies in maize have indicated sufficiently high prediction accuracies for genomewide selection for traits such as grain yield and moisture (Lorenzana and Bernardo, 2009; Schulz-Streeck et al., 2012; Technow et al., 2012; Massman et al., 2013; Krchov et al., 2014). When genomewide selection is less accurate than phenotypic selection, this lower accuracy could be offset by increasing the selection intensity with genomewide selection. This increase in selection intensity is made possible by the lower per-sample costs with genotyping than with phenotyping, thereby allowing a larger number of DH lines that can be genotyped (rather than phenotyped) under a fixed budget. Different sizes of the training population (n) and of the test population (N-n) can be accommodated under a fixed budget, and the relative sizes of the training and test populations would affect the gains from the genomewide selection in Fig.1.

Previous studies (Lorenz, 2013; Riedelsheimer and Melchinger, 2013) have investigated the optimum resource allocation during genomewide selection, but these previous studies did not utilize empirical data. In this study, we considered the empirical prediction accuracies for genomewide selection across multiple populations, testers, years, and locations (Krchov et al., 2014) as well as different total budgets and per-

sample costs for phenotyping and genotyping in maize. We then assessed the observed and expected gains from the schemes for phenotypic selection and genomewide selection depicted in Fig. 1. Our objectives were to (i) determine whether or not genomewide selection (Fig. 1) is more efficient than phenotypic selection under a fixed budget, and (ii) give recommendations for implementing genomewide selection in a commercial line development program in maize.

MATERIALS AND METHODS

Phenotypic and Genotypic Data

The field experiments were described in detail by Krchov et al. (2014). We studied four biparental crosses (designated A/B, C/B, A/D, and E/F) each with 150 or 250 doubled haploid (DH) lines that belonged to the Iowa Stiff Stalk Synthetic group. The DH lines in each of the four biparental crosses were testcrossed to one or two tester inbreds that belonged to an opposite heterotic group. All plant materials used in this study were proprietary to AgReliant Genetics, LLC (Westfield, IN). Testcrosses in the A/B, C/B, and C/D populations were evaluated in yield trials in 2008, 2011, and 2012. Testcrosses in the E/F population were evaluated in 2009, 2011, and 2012. Analyses were carried out for the traits grain yield (t ha^{-1}) and grain moisture (g kg^{-1}).

All DH lines were genotyped with 3072 random SNP markers on the Illumina iScan platform. Markers with more than 10% missing values or with more than 5% heterozygosity were excluded from further analysis.

Accuracy of Phenotypic and Genomewide Selection

The phenotypic dataset for the *N* lines in each cross was divided into a training population and a test population. The training population refers to the marker and phenotypic data for lines 1 to *n* in year i, where *n* was equal to 100, 80, 60, 40, and 20 lines (as reference please see Fig. 1 in Krchov et al. (2014). The test population refers to the phenotypic data for lines *n*+1 to *N* in years j+k. The data from each year were considered as the training population and the data from the two remaining years were considered as the corresponding test population. In other words, the data for the training and test populations (in parentheses) were from the following years: Year 1 (Year 2+3); Year 2 (Year 1+3); and Year 3 (Year 1+2). Analyses were carried out for all three cases and the results were averaged across all three cases. All genomewide predictions were based on the training population (i.e., lines 1 to *n* in year i) and genomewide marker effects were calculated by ridge regression best linear unbiased prediction as described by Krchov et al. (2014).

The accuracy of phenotypic selection, (denoted by r_P) was calculated as the Pearson correlation between the observed performance of lines *n*+1 to *N* in year i and the observed performance of lines *n*+1 to *N* in years j+k. The accuracy of genomewide selection, (denoted as r_{MP}) was the correlation between the marker-predicted

performance of lines *n*+1 to *N* in year i and the observed performance of lines *n*+1 to *N* in years j+k. The accuracy of genomewide selection without cross-validation across lines, (denoted as $r_{MP(N)}$) was the correlation between the marker-predicted performance of lines 1 to *n* in year i and the observed performance of lines 1 to *n* in years j+k.

The r_P, r_{MP}, and $r_{MP(N)}$ values were calculated for each of 100 repeats, with each repeat having a random assignment of the N lines into the training population (n = 20, 40, 60, 80, or 100 lines) and test population (*N*-*n* lines). The mean value of r_P, r_{MP}, and $r_{MP(N)}$ were calculated across the 100 repeats.

Observed Selection Gain

Observed selection gains were determined by identifying the best lines (by phenotypic selection or genomewide selection) based on the data in year i and calculating the deviation of the mean of these selected lines from the overall mean in years j+k. Observed gains from genomewide selection were calculated as follows:

(1) Among lines 1 to *n* (i.e. training population), the 5, 10 and 20 lines with the best predicted performance based on year i were selected. The observed gain for the training population was calculated as the difference between the mean of these selected lines in years j+k and the overall mean in years j+k.

(2) Among lines *n*+1 to *N* (test population) the 5, 10 and 20 lines with the best predicted performance based on year i were selected. The observed gain for the test population was calculated as the difference between the mean of these selected lines in years j+k and the overall mean in years j+k. The overall observed gain from genomewide selection

was then obtained as the weighted (by the number of lines) mean of the observed gains in the training population and in the test population.

(3) For phenotypic selection, the best 5, 10, and 20 lines were selected based on the phenotypic data in year i. The observed gain from phenotypic selection was calculated as the mean of these selected lines in years j+k minus the overall mean in years j+k. The analysis was performed for 100 random repeats, and the observed gains were averaged across all 100 repeats.

Expected Selection Gain

The expected genetic gain among testcrosses can be generally expressed as $R = ir\sigma_G$,where R is the response to selection, i is the standardized selection differential when p% of the lines are selected, r is the selection accuracy, and σ_G is the square root of the testcross genetic variance. The values of σ_G for the populations studied were obtained from the estimates of the testcross genetic variances reported by Krchov et al. (2014). The number of lines selected is denoted by N_{sel}. The standardized selection differential (i) for a given proportion of selected lines ($p=N_{sel}/N$) was calculated following an approximation from Burrows (1972) for finite populations:

$$i = i_p - \frac{N-N_{sel}}{2N_{sel}\,(N+1)i_p} \qquad [1]$$

where $i(p)$ was the standardized selection differential in infinite populations.

For genomewide selection, the selection accuracy is equal to $r_{MP(N)}$ in the training population and r_{MP} in the test population. The expected selection gain for genomewide selection was calculated as

$$R = \left(i_{Training} r_{MP(N)} \frac{n}{N} + i_{Test} r_{MP} \frac{(N-n)}{N} \right) \sigma_G \quad [2]$$

where $i_{Training}$ was the standardized selection differential for the training population and i_{Test} was the standardized selection differential for the test population. For phenotypic selection, the expected selection gain was calculated as

$$R = i\, r_P \sigma_G \quad [3]$$

We used r_P instead of the square root of heritability for consistency between how expected gains were calculated for phenotypic selection and genomewide selection.

To compare the observed and expected gains, we calculated the latter based on different sizes of the training population (n = 20, 40, 60, 80, and 100) and considered all of the remaining lines as part of the test population. Because the six crosses differed in their total numbers of DH lines (as reference see Krchov et al. 2014), the sizes of the test population therefore varied among the crosses even though n was constant. These comparisons were made to determine the correspondence between the observed and expected gains, and such comparisons were not based on the same total budget.

For the comparisons of the expected gains from phenotypic versus genomewide selection, the results are expressed as the relative efficiency (RE). RE values were calculated as the ratio of expected gain from genomewide selection divided by the expected gain from phenotypic selection. Thus, any RE values greater than 1 indicate a higher efficiency of genomewide selection. RE_{Max} values were the maximum RE values that could be achieved among different allocations regarding the number of lines in the training population and the resulting number of lines in the test population.

All pairwise differences between the expected gains from phenotypic selection and genomewide selection were tested for significance via a t-test. All pairwise differences between the expected gains and observed gains were also tested for significance via a t-test.

Total Budget and Per-Unit Costs

We also compared the expected gains while keeping to a fixed budget. We compared the following budgets for selection within a single population: $12,000, $14,000, $16,000, $18,000, $20,000 and $22,000. These budgets represent the budget of a high investment maize breeding program, excluding advanced field testing and commercialization. We considered the following costs on the basis of information within a commercial breeding program (G.Seitz, personal communication, 2013) or from other sources of information: (1) $25 for making the initial cross between two inbred lines within one heterotic group; (2) $15 or $50 (http://www.plantbreeding.iastate.edu/DHF/DHF.htm) to produce a DH line; (3) $12 to increase the seeds of a DH line; (4) $15 to produce testcross seeds for a

DH line; (5) $15 or $30 to genotype each line with 3072 single nucleotide polymorphism (SNPs) markers; and (6) $90 [i.e., $15 per plot (http://www.techservicespro.com/Sales/Corn.htm) at six locations] to phenotype the testcross of each DH line.

For phenotypic selection, the total per-unit costs (initial cross, production and increase of a DH line, production of testcross seed, and yield trials), was $157 per DH line. The number of DH lines that can be tested within a given budget was equal to the total budget ($12,000 to $22,000); divided by $157.

For genomewide selection, the per-line cost for the training population was equal to the costs for phenotyping ($157) plus the costs for genotyping ($15 or $30). The per-line costs for the test population was equal to the costs of producing a DH line plus the cost of genotyping ($15 or $30). The total costs for the training population was equal to the per-line costs ($172 or $187) multiplied by the size of the training population (n). The available budget for the test population was equal to the total budget minus the total costs for the training population. The number of lines in the test population was then obtained as the available budget for the test population divided by the per-line cost ($30 or $45) for the test population. Only those schemes that did not exceed the total budget were considered.

RESULTS AND DISCUSSION

Observed versus Expected Gains from Genomewide Selection

Overall, the expected and observed gains from genomewide selection corresponded very well for both grain yield and moisture (Fig. 2). The R^2 value between the observed and expected gains was 0.73 for grain yield and 0.70 for moisture. Within each of the six crosses, the correlations between observed and expected gains ranged from 0.51 in cross A/B (T1) to 0.89 in cross E/F for grain yield, and from 0.11 in cross C/D to 0.74 in cross E/F for moisture. All differences between the observed and expected gains for grain yield and moisture were not significant (P = 0.05) for each trait. However, larger deviations between the observed and expected gains occurred as their values increased. Such larger values occurred when fewer lines were selected (i.e., larger selection differentials and responses), and the selection of fewer lines was expected to increase the standard error of the mean of the selected lines. Our recommendations are largely based on expected rather than observed gains. However, the good correspondence between expected and observed gains allows drawing conclusions from the expected gains.

Relative Efficiency of Genomewide Selection

Whether or not the genomewide selection scheme shown in Fig. 1 was more efficient than phenotypic selection (i.e., $RE_{Max} > 1$) depended on the following factors: (1) accuracy of phenotypic selection (r_P) and of genomewide selection [$r_{MP(N)}$ and r_{MP}] for the trait in a given cross (Table 1); (2) size of the training population (Table 1); (3) total

budget; (4) costs of producing and genotyping a DH line; and (5) number of selected lines (N_{sel}). The RE_{Max} values ranged from 0.57 to 1.40 for grain yield (Table 2) and from 0.88 to 1.19 for moisture (Table 3). There were crosses and resource allocations for which genomewide selection scheme depicted in Fig. 1 was always efficient, as well as crosses and resource allocations for which the scheme was always inefficient.

The $r_{MP(N)}$ and r_{MP} values for grain yield were highest in the A/B (T1) cross and lowest in the C/B (T2) cross (Table 1). The RE_{Max} values for grain yield were generally greater than 1 in the A/B (T1) cross, and were always less than 1 in the C/B (T2) cross (Table 2). However, the highest $r_{MP(N)}$ and r_{MP} values did not necessarily lead to the highest RE_{Max} values because RE_{Max} also depends on the accuracy of phenotypic selection (r_P) relative to $r_{MP(N)}$ and r_{MP}. For example, the RE_{Max} values for grain yield were higher in the A/B (T2) cross than in the A/B (T1) cross; the former had higher $r_{MP(N)}$ and r_{MP} values but the latter had a lower r_P value, thereby compensating for its lower $r_{MP(N)}$ and r_{MP}. Our results indicated that in general, $r_{MP(N)}$ and r_{MP} values above 0.50 usually lead to $RE_{Max} > 1$.

In general, the RE_{Max} values decreased as size of the training population (n) decreased. The r_{MP} and $r_{MP(N)}$ values decreased as (n) decreased (Table 1). But in agreement with the theoretical findings of Daetwyler et al. (2008), the relationship between n and r_{MP} and $r_{MP(N)}$ was not linear (Table 1). The $r_{MP(N)}$ values were fairly stable across the different values of n, and did not decrease drastically even with $n = 20$ (Table 1). However, the r_{MP} values decreased rapidly as n decreased to 20. The maximum size of the training population was not always $n = 100$ because of budget

constraints. With the lowest total budget ($12,000), for example, training populations of n = 80 or 100 were not possible. The RE_{Max} values were not always attained by maximizing n within a given budget. Because of the high prediction accuracies in the A/B (T1) and A/B (T2) crosses, RE_{Max} in these two crosses was reached with training populations smaller the maximum n that was possible within a given budget (Tables 3 and 4), i.e. the relative increase in $r_{MP(N)}$ and r_{MP} were smaller than the increase of selection intensity in the test population. But in the four other crosses, the RE_{Max} values were reached when n was maximum.

A larger total budget not only accommodates a larger training population, but it also allows a larger test population. The RE_{Max} values increased when the total budget increased from $12,000 to $22,000 (Tables 2 and 3). When the total budget was $12,000, the maximum size of the training population was 60 whereas the maximum size of the test population was 56 DH lines (Table 4). When the total budget was increased to $22,000, the same size of the training population (n = 60) allows a test population of up to 389 DH lines. Additionally, the larger test populations possible under a larger total budget led to a higher standardized selection differential when the same number of lines were selected (N_{sel} = 5, 10, or 20) from test populations of increasing size.

Such results are consistent with the simulation results of Riedelsheimer and Melchinger (2013), who showed that once a certain budget is allocated to the training population and a high prediction accuracy is achieved, further selection gains can be obtained only by expanding the test population substantially, so that the selection

intensity is increased. Reduced genotyping costs ($15 instead of $30 per line) and DH production costs ($15 instead of $50 per line) likewise led to larger training and test populations, which in turn led to higher RE_{Max} values for both grain yield (Table 2) and moisture (Table 3).

The RE_{Max} values for both grain yield and moisture increased as the number of selected lines (N_{sel}) increased from 5 to 20 (Tables 2 and 3). When the genotyping cost was $15 per line, the RE_{Max} values always exceeded 1 when N_{sel} was 20, except for two cases [grain yield in C/B (T1) and E/F] out of the 12 combinations between crosses and traits. These two cases coincided with those for which $r_{MP(N)}$ and r_{MP} were the lowest (Table 1). In this case, a breeder might compensate low prediction accuracies with higher number of selected lines.

Recommendations

Our results indicated that the genomewide selection scheme shown in Fig. 1 was not always more effective than phenotypic selection. However, we offer four recommendations that would enhance the efficiency of the genomewide selection scheme in Fig. 1 in maize breeding programs. First, we recommend a training population size of 60 to 80 lines within a biparental cross. A training population size of 20 to 40 lines leads to low prediction accuracies that are not compensated by a larger test population. Second, r_{MP} and $r_{MP(N)}$ values should be greater than 0.50 because they lead to relative efficiencies greater 1. Unfortunately, a breeder cannot directly control the

prediction accuracy and can estimate the prediction accuracy only after the population has been phenotyped and genotyped. Third, selection should not be stringent and the number of selected lines (N_{sel}) should be around 20 regardless of the size of the training and test populations. Fourth, the test population should include around 125 or more. We base this recommendation on the size of the test population (Table 4) that corresponded to the minimum recommended size of the training population (60 lines) and all of the total budgets, except for the lowest ($12,000). A larger total budget and lower costs of DH production and of genotyping would enable larger test populations, thus leading to higher efficiencies of genomewide selection.

Our findings are in agreement with those of Riedelsheimer and Melchinger (2013), who showed from simulation studies that the cost of genotyping is a main determinant of whether or not genomewide selection is superior to phenotypic selection, especially under a small budget. We calculated genotyping costs with $30, which is about the current price (http://epigenome.usc.edu/services/microarray.htm). But over the last few years genotyping costs have dropped dramatically and we speculate that the costs will still be decreasing in the future. We also calculated genotyping costs with $15, but this price might be even lower in the future and genomewide selection will be even more efficient.

In addition, genomewide selection can still be more efficient than phenotypic selection for two reasons that do not pertain to gain per unit cost. First, the genomewide selection scheme in Fig. 1 allows the evaluation of DH lines in the first (D_1) generation that do not have enough seeds for testcrossing and phenotyping yet. When producing

DH lines, the number of viable seed per ear varies between 5 and 20 kernels after artificial doubling (Geiger and Gordillo, 2009). Depending on the efficiency of the DH production system, the proportion of lines with not enough seeds may vary between 50% and 70%. Thus, these lines with limited seeds could then be included in the test population, and the training population would comprise DH lines with enough seeds for testcrossing and phenotyping. Second, genotyping is more convenient than phenotyping. A breeder has a given set of resources in terms of people, equipment, and hectares to phenotype the breeding germplasm. Fixed calendar days for planting and harvesting naturally limit for the amount of field testing. Thus, there is a fixed capacity for phenotyping in a breeding program. The reduced amount of field testing in the genomewide selection scheme in Fig. 1 may lead to a better quality of the field data, thereby enhancing the effectiveness of selection.

Table 1. Prediction accuracies for grain yield and moisture in four maize biparental crosses crossed to one or two testers (T1 and T2).

Cross	Parameter	Grain yield					Moisture					
			Training population size						Training population size			
			100	80	60	40	20		100	80	60	40	20
A/B (T1)	r_{MP} †		0.66	0.65	0.62	0.57	0.46		0.63	0.61	0.57	0.53	0.43
	$r_{MP(N)}$		0.74	0.74	0.73	0.72	0.70		0.77	0.76	0.76	0.75	0.75
	r_P	0.68						0.74					
	h^2	0.94						0.92					
A/B (T2)	r_{MP}		0.48	0.48	0.44	0.41	0.33		0.58	0.56	0.53	0.49	0.41
	$r_{MP(N)}$		0.53	0.52	0.51	0.46	0.48		0.67	0.66	0.66	0.66	0.64
	r_P	0.44						0.64					
	h^2	0.72						0.80					
C/B (T1)	r_{MP}		0.42	0.39	0.35	0.31	0.23		0.44	0.42	0.41	0.37	0.29
	$r_{MP(N)}$		0.54	0.52	0.52	0.50	0.48		0.53	0.53	0.53	0.53	0.52
	r_P	0.50						0.53					
	h^2	0.69						0.72					
C/B (T2)	r_{MP}		0.1	0.09	0.07	0.06	0.03		0.40	0.38	0.35	0.30	0.23
	$r_{MP(N)}$		0.14	0.14	0.15	0.14	0.14		0.44	0.43	0.43	0.42	0.42
	r_P	0.16						0.42					
	h^2	0.45						0.62					

C/D	r_{MP}		0.45	0.37	0.34	0.30	0.18		0.70	0.66	0.62	0.53	0.41
	$r_{MP(N)}$		0.53	0.52	0.49	0.51	0.48		0.79	0.78	0.77	0.76	0.75
	r_P	0.47						0.73					
	h^2	0.78						0.93					
E/F	r_{MP}		0.21	0.18	0.14	0.11	0.06		0.64	0.62	0.59	0.56	0.46
	$r_{MP(N)}$		0.35	0.35	0.33	0.32	0.32		0.76	0.77	0.76	0.76	0.77
	r_P	0.34						0.76					
	h^2	0.77						0.91					

† r_{MP}, Pearson correlation between the marker-predicted performance of individuals n+1 to N in year i and the observed performance of individuals n+1 to N in years j+k; $r_{MP(N)}$,Pearson correlation between the marker-predicted performance of individuals 1 to n in year i and the observed performance of individuals 1 to n in years j+k; $r_{P,}$ Pearson correlation between the observed performance of individuals n+1 to N in year i and the observed performance of individuals n+1 to N in years j+k; h^2, entry-mean heritability calculated across three years.

Table 2. Maximum values for relative efficiency (RE_{Max}) of genomewide selection over phenotypic selection for grain yield for four maize biparental crosses crossed to two testers.

			Standard costs		Lower genotyping cost †		Higher DH production cost ‡	
Cross	Number selected N_{sel}	Total budget ($)	RE_{Max}	Training population	RE_{Max}	Training population	RE_{Max}	Training population
A/B (T1)	5	12,000	1.00	40	1.06	60	0.99	40
		14,000	1.03	60	1.08	60	1.00	60
		16,000	1.04	60	1.09	60	1.01	60
		18,000	1.05	80	1.10	80	1.02	60
		20,000	1.06	80	1.11	80	1.03	80
		22,000	1.07	100	1.12	80	1.04	80
	10	12,000	1.05	40	1.13	40	1.02	40
		14,000	1.06	60	1.13	60	1.01	60
		16,000	1.08	60	1.15	60	1.03	60
		18,000	1.08	60	1.15	60	1.04	60
		20,000	1.09	80	1.16	80	1.04	80
		22,000	1.10	80	1.17	80	1.06	80
	20	12,000	1.14	40	1.26	40	1.07	40
		14,000	1.14	40	1.24	40	1.07	40
		16,000	1.13	60	1.24	60	1.08	40
		18,000	1.14	60	1.24	60	1.07	60
		20,000	1.15	60	1.24	60	1.08	60
		22,000	1.15	60	1.24	80	1.09	60
A/B (T2)	5	12,000	1.09	60	1.14	60	1.02	40
		14,000	1.11	60	1.16	60	1.10	60
		16,000	1.14	80	1.20	80	1.11	60
		18,000	1.17	80	1.23	80	1.11	60
		20,000	1.19	80	1.25	80	1.14	80
		22,000	1.20	80	1.26	80	1.15	80
	10	12,000	1.11	40	1.21	40	1.05	40
		14,000	1.14	60	1.21	60	1.09	60
		16,000	1.15	60	1.24	80	1.12	60
		18,000	1.19	80	1.28	80	1.14	60
		20,000	1.22	80	1.30	80	1.15	60
		22,000	1.23	80	1.31	80	1.17	80

	20	12,000	1.21	40	1.35	40	1.10	40
		14,000	1.21	40	1.34	40	1.12	40
		16,000	1.22	60	1.33	60	1.14	60
		18,000	1.23	80	1.36	80	1.18	60
		20,000	1.26	80	1.39	80	1.19	80
		22,000	1.28	80	1.40	80	1.20	80
C/B (T1)	5	12,000	0.94	60	0.94	60	0.85	40
		14,000	0.90	60	1.00	80	0.97	60
		16,000	0.97	80	0.98	80	0.92	60
		18,000	0.95	80	1.03	100	0.90	60
		20,000	1.00	100	1.02	100	0.97	80
		22,000	1.00	100	1.04	120	0.96	80
	10	12,000	0.95	60	0.97	60	0.88	40
		14,000	0.93	60	1.00	80	0.96	60
		16,000	0.98	80	1.00	80	0.94	60
		18,000	0.97	80	1.04	100	0.92	60
		20,000	1.01	100	1.05	100	0.98	80
		22,000	1.01	100	1.05	120	0.97	80
	20	12,000	0.95	60	1.03	60	0.92	40
		14,000	0.97	60	1.03	60	0.94	60
		16,000	0.99	80	1.06	80	0.95	60
		18,000	1.00	80	1.07	80	0.95	60
		20,000	1.02	100	1.09	100	0.99	80
		22,000	1.03	100	1.10	100	0.99	80
C/B (T2)	5	12,000	0.82	60	0.76	60	0.66	40
		14,000	0.73	60	0.88	80	0.87	60
		16,000	0.83	80	0.79	80	0.78	60
		18,000	0.78	80	0.83	100	0.73	60
		20,000	0.81	100	0.80	100	0.83	80
		22,000	0.79	100	0.84	120	0.79	80
	10	12,000	0.82	60	0.79	60	0.68	40
		14,000	0.75	60	0.88	80	0.86	60
		16,000	0.84	80	0.82	80	0.79	60
		18,000	0.79	80	0.84	100	0.74	60
		20,000	0.82	100	0.82	100	0.84	80
		22,000	0.78	100	0.86	120	0.81	80
	20	12,000	0.83	60	0.84	60	0.72	40

		14,000	0.78	60	0.88	80	0.84	60
		16,000	0.84	80	0.86	80	0.80	60
		18,000	0.82	80	0.85	100	0.77	60
		20,000	0.82	100	0.85	100	0.84	80
		22,000	0.82	100	0.87	120	0.82	80
C/D	5	12,000	0.95	60	0.95	60	0.90	40
		14,000	0.92	60	1.06	80	0.98	60
		16,000	1.02	80	1.01	80	0.94	60
		18,000	0.99	80	1.09	100	0.91	60
		20,000	1.07	100	1.11	100	1.02	80
		22,000	1.07	100	1.14	120	1.00	80
	10	12,000	0.96	60	0.98	60	0.92	40
		14,000	0.94	60	1.06	80	0.97	60
		16,000	1.03	80	1.05	80	0.95	60
		18,000	1.01	80	1.10	100	0.93	60
		20,000	1.08	100	1.13	100	1.03	80
		22,000	1.09	100	1.16	120	1.01	80
	20	12,000	0.96	60	1.05	60	0.97	40
		14,000	0.98	60	1.06	80	0.95	60
		16,000	1.03	80	1.10	80	0.97	60
		18,000	1.04	80	1.12	100	0.96	60
		20,000	1.08	100	1.17	100	1.03	80
		22,000	1.11	100	1.17	120	1.03	80
E/F	5	12,000	0.57	60	0.75	60	0.66	40
		14,000	0.63	60	0.97	80	0.90	60
		16,000	0.74	80	0.84	80	0.79	60
		18,000	0.82	80	0.95	100	0.72	60
		20,000	0.88	100	0.88	100	0.91	80
		22,000	0.97	100	0.97	120	0.85	80
	10	12,000	0.59	60	0.78	60	0.68	40
		14,000	0.65	60	0.97	80	0.89	60
		16,000	0.76	80	0.86	80	0.80	60
		18,000	0.84	80	0.96	100	0.74	60
		20,000	0.90	100	0.91	100	0.91	80
		22,000	0.98	100	0.98	120	0.86	80
	20	12,000	0.64	60	0.83	60	0.72	40
		14,000	0.69	60	0.97	80	0.87	60

	16,000	0.78	80	0.90	80	0.81	60
	18,000	0.87	80	0.97	100	0.76	60
	20,000	0.93	100	0.94	100	0.92	80
	22,000	1.00	100	1.00	120	0.88	80

† Genotyping cost of $15 instead of $30 per line.

‡ Doubled haploid (DH) production cost of $50 instead of $15 per line.

Table 3. Maximum values for relative efficiency (RE_{Max}) of genomewide selection over phenotypic selection for moisture for four maize biparental crosses crossed to two testers.

Cross	Number selected N_{sel}	Total budget ($)	Standard costs		Lower genotyping cost †		Higher DH production cost ‡	
			RE_{Max}	Training population	RE_{Max}	Training population	RE_{Max}	Training population
A/B (T1)	5	12,000	0.95	60	0.95	60	0.90	40
		14,000	0.94	60	0.98	80	0.96	60
		16,000	0.93	60	0.98	80	0.94	60
		18,000	0.98	80	1.01	100	0.93	60
		20,000	0.98	100	1.00	100	0.97	80
		22,000	0.98	100	1.00	100	0.96	80
	10	12,000	0.96	60	0.99	60	0.93	40
		14,000	0.96	60	0.99	80	0.96	60
		16,000	0.96	60	1.00	80	0.96	60
		18,000	1.00	80	1.05	100	0.95	60
		20,000	0.99	80	1.02	100	0.97	80
		22,000	0.99	100	1.04	100	0.97	80
	20	12,000	1.01	40	1.05	40	0.98	40
		14,000	1.01	60	1.08	40	0.97	40
		16,000	1.02	60	1.07	60	0.97	60
		18,000	1.02	80	1.08	100	0.98	60
		20,000	1.03	80	1.08	80	0.98	80
		22,000	1.03	80	1.09	100	0.98	80
A/B (T2)	5	12,000	0.97	60	1.00	60	0.93	40
		14,000	0.97	60	1.01	60	0.97	60
		16,000	0.98	80	1.02	80	0.96	60
		18,000	0.99	80	1.03	80	0.96	60
		20,000	0.99	100	1.04	100	0.98	80
		22,000	1.00	100	1.05	100	0.98	80
	10	12,000	0.97	40	1.04	40	0.96	40
		14,000	1.00	60	1.06	60	0.96	60
		16,000	1.00	60	1.06	60	0.98	60
		18,000	1.01	80	1.07	80	0.98	60

		20,000	1.02	80	1.08	80	0.98	80
		22,000	1.03	80	1.08	100	0.99	80
	20	12,000	1.07	40	1.17	40	1.02	40
		14,000	1.05	40	1.14	40	1.01	40
		16,000	1.06	60	1.15	60	0.99	40
		18,000	1.06	60	1.15	60	1.01	60
		20,000	1.06	60	1.15	80	1.02	60
		22,000	1.07	80	1.15	80	1.02	60
C/B (T1)	5	12,000	0.93	60	0.95	60	0.88	40
		14,000	0.92	60	0.96	80	0.94	60
		16,000	0.94	80	0.96	80	0.92	60
		18,000	0.94	80	0.96	80	0.91	60
		20,000	0.95	100	0.97	100	0.94	80
		22,000	0.95	100	0.98	100	0.93	80
	10	12,000	0.93	60	0.99	60	0.91	40
		14,000	0.94	60	0.96	60	0.94	60
		16,000	0.95	80	0.99	60	0.94	60
		18,000	0.96	80	0.98	60	0.94	60
		20,000	0.96	100	1.00	100	0.95	80
		22,000	0.97	100	1.01	100	0.94	80
	20	12,000	0.99	40	1.07	40	0.96	40
		14,000	0.98	60	1.08	60	0.92	40
		16,000	1.00	60	1.08	60	0.95	60
		18,000	1.00	60	1.08	60	0.97	60
		20,000	1.00	80	1.08	60	0.95	60
		22,000	1.00	80	1.07	60	0.96	60
C/B (T2)	5	12,000	0.96	60	1.00	60	0.89	40
		14,000	0.97	60	1.01	60	0.96	60
		16,000	0.98	80	1.03	80	0.96	60
		18,000	1.00	80	1.05	80	0.96	60
		20,000	1.02	100	1.06	100	0.98	80
		22,000	1.03	100	1.08	100	0.99	80
	10	12,000	0.97	60	1.04	60	0.92	40
		14,000	0.99	60	1.06	60	0.76	40
		16,000	1.00	60	1.07	60	0.63	40
		18,000	1.02	80	1.09	80	0.55	40
		20,000	1.04	80	1.09	80	0.99	80

		22,000	1.04	100	1.11	100	1.00	80
	20	12,000	1.01	60	1.10	60	0.97	40
		14,000	1.04	60	1.14	60	0.74	40
		16,000	1.06	80	1.16	60	0.64	40
		18,000	1.07	80	1.15	80	0.57	40
		20,000	1.08	100	1.17	80	0.99	80
		22,000	1.09	100	1.18	80	1.02	80
C/D	5	12,000	0.99	60	1.02	60	0.92	40
		14,000	0.99	60	1.03	60	0.99	60
		16,000	1.01	80	1.05	80	0.99	60
		18,000	1.02	80	1.07	80	0.98	60
		20,000	1.03	80	1.08	100	1.01	80
		22,000	1.04	80	1.10	120	1.01	80
	10	12,000	1.00	40	1.06	60	0.95	40
		14,000	1.02	60	1.08	60	0.99	60
		16,000	1.03	60	1.09	60	1.00	60
		18,000	1.05	60	1.10	80	1.00	60
		20,000	1.06	80	1.12	80	1.02	80
		22,000	1.06	80	1.13	100	1.03	80
	20	12,000	1.03	40	1.13	60	1.00	40
		14,000	1.02	60	1.04	60	0.97	60
		16,000	1.09	60	1.18	60	1.02	60
		18,000	1.09	60	1.18	60	1.04	60
		20,000	1.09	80	1.19	80	1.04	60
		22,000	1.10	80	1.19	80	1.05	80
E/F	5	12,000	0.93	60	0.96	60	0.90	40
		14,000	0.93	60	0.97	80	0.94	60
		16,000	0.95	80	0.98	80	0.93	60
		18,000	0.95	80	0.98	80	0.92	60
		20,000	0.95	100	0.98	80	0.95	80
		22,000	0.95	100	0.99	100	0.95	80
	10	12,000	0.94	60	0.99	40	0.93	40
		14,000	0.95	60	1.00	60	0.94	60
		16,000	0.96	60	1.01	60	0.94	60
		18,000	0.97	80	1.02	80	0.94	60
		20,000	0.98	80	1.02	80	0.96	80
		22,000	0.98	80	1.02	80	0.96	80

20	12,000	1.02	40	1.06	40	0.98	40
	14,000	1.01	40	1.08	40	0.98	40
	16,000	1.01	60	1.09	60	0.97	40
	18,000	1.01	60	1.09	60	0.97	60
	20,000	1.01	80	1.09	80	0.97	60
	22,000	1.02	80	1.09	80	0.98	80

† Genotyping cost of $15 instead of $30 per line.

‡ Doubled haploid (DH) production cost of $50 instead of $15 per line.

Table 4. Number of lines in the training and test populations for different total budgets and costs.

Total budget ($)	Training population	Test population		
		Standard costs	Lower genotyping costs [†]	Higher DH production costs [‡]
12,000	100			
	80			
	60	17	56	
	40	100	171	36
	20	184	285	93
	Lines in phenotypic selection	76	76	60
14,000	100			
	80		8	
	60	62	123	4
	40	145	237	61
	20	228	352	118
	Lines in phenotypic selection	89	89	71
16,000	100			
	80	23	74	
	60	106	189	29
	40	189	304	86
	20	272	418	143
	Lines in phenotypic selection	102	102	81
18,000	100		26	
	80	68	141	
	60	151	256	54
	40	234	370	111
	20	317	485	168
	Lines in phenotypic	115	115	91

	selection			
20,000	100	29	93	
	80	112	208	23
	60	195	322	79
	40	278	437	136
	20	361	552	193
	Lines in phenotypic selection	127	127	101
	120		45	
22,000	100	73	160	
	80	156	274	48
	60	240	389	104
	40	323	504	161
	20	406	618	218
	Lines in phenotypic selection	140	140	111

† Genotyping cost of $15 instead of $30 per line.

‡ Doubled haploid (DH) production cost of $50 instead of $15 per line.

Fig. 1. Hybrid maize breeding schemes for phenotypic and genomewide selection among maize doubled haploid (DH) lines.

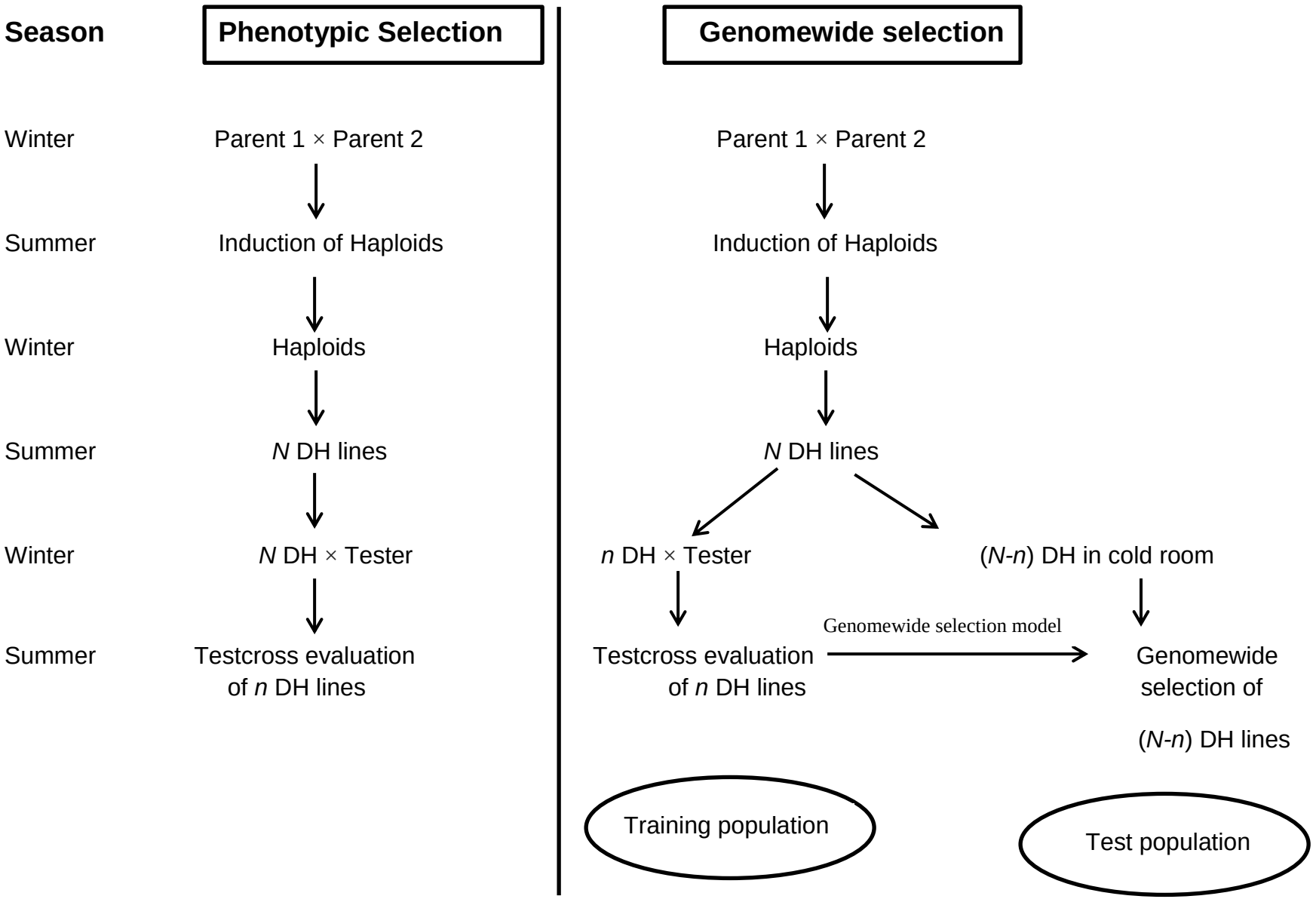
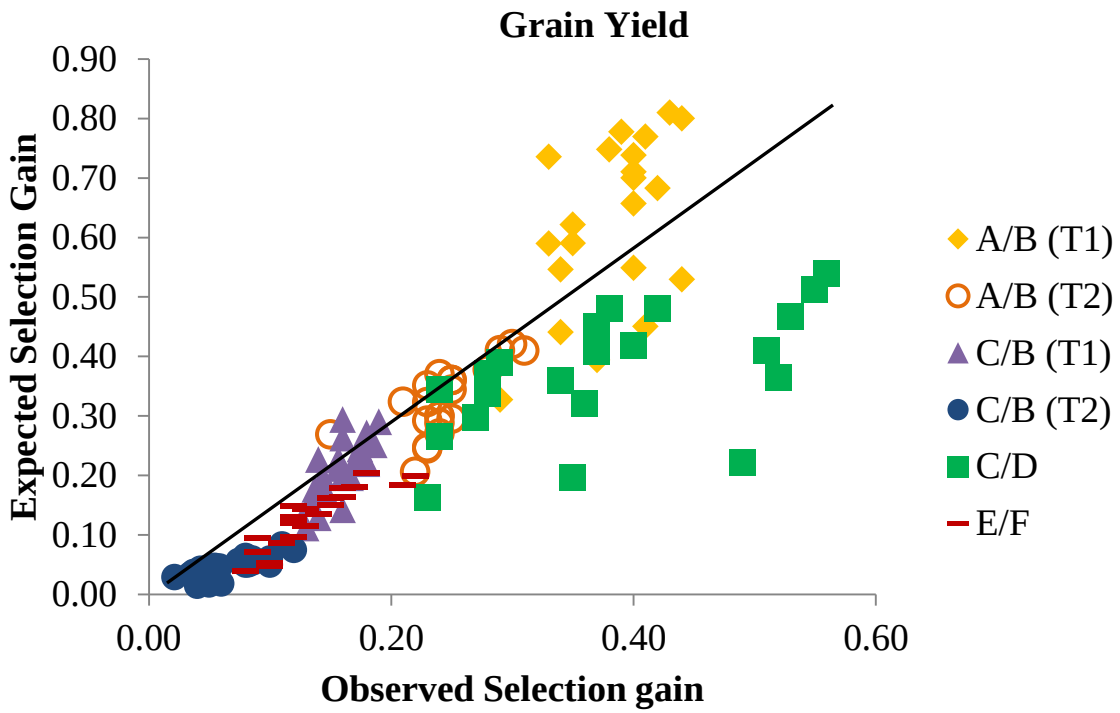
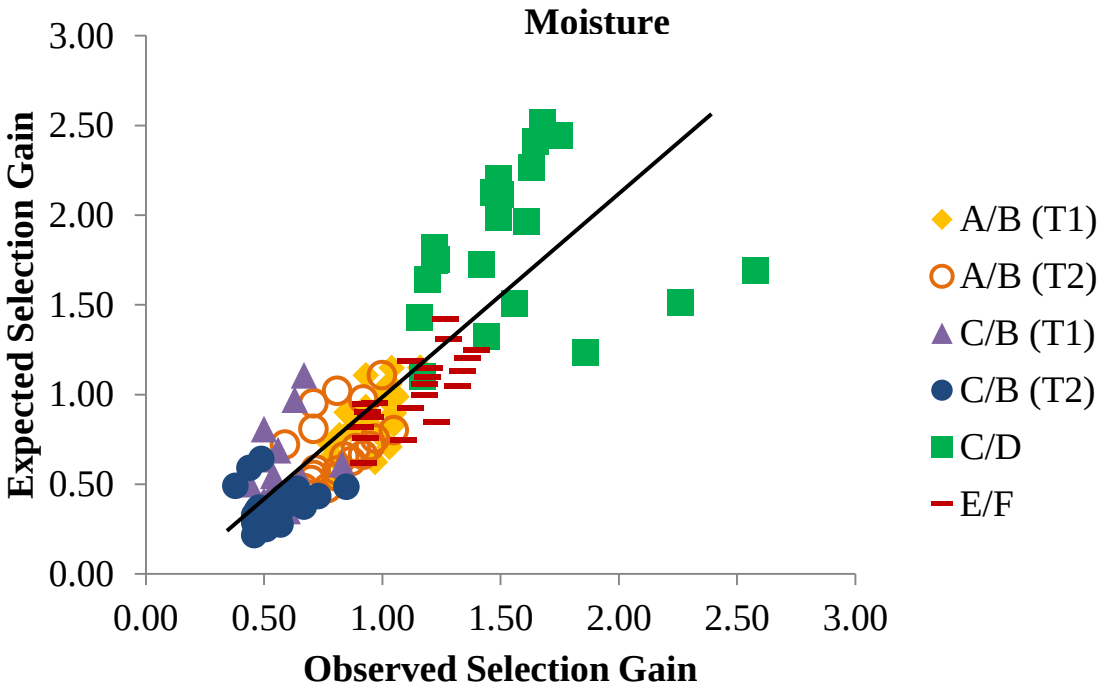

Season
Phenotypic Selection
Genomewide selection
Winter
Parent 1 × Parent 2
Summer
Induction of Haploids
Winter
Haploids
Summer
N DH lines
Winter
N DH × Tester
Summer
Testcross evaluation of n DH lines
Parent 1 × Parent 2
Induction of Haploids
Haploids
N DH lines
n DH × Tester
(N-n) DH in cold room
Genomewide selection model
Testcross evaluation of n DH lines
Genomewide selection of (N-n) DH lines
Training population
Test population

Figure 2. Expected selection gains versus observed selection gains for four biparental crosses crossed to two testers for grain yield (t/ha^{-1}) and moisture ($g\ kg^{-1}$).


Grain Yield
Expected Selection Gain
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.00
0.20
0.40
0.60
Observed Selection gain
A/B (T1)
A/B (T2)
C/B (T1)
C/B (T2)
C/D
E/F


Moisture
Expected Selection Gain
3.00
2.50
2.00
1.50
1.00
0.50
0.00
0.00
0.50
1.00
1.50
2.00
2.50
3.00
Observed Selection Gain
A/B (T1)
A/B (T2)
C/B (T1)
C/B (T2)
C/D
E/F

BIBLIOGRAPHY

Albrecht, T., V. Wimmer, H.-J. Auinger, M. Erbe, C. Knaak, M. Ouzunova, H. Simianer, and C.-C. Schön. 2011. Genome-based prediction of testcross values in maize. Theor. Appl. Genet. 123:339-350.

Asoro, F.G., M.A. Newell, W.D. Beavis, M.P. Scott, and J.-L. Jannink. 2011. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. The Plant Genome 4:132-144.

Asoro, F.G., M.A. Newell, W.D. Beavis, M.P. Scott, N.A. Tinker, and J.-L. Jannink. 2013. Comparison of genomic, marker-assisted, and pedigree-BLUP selection methods to increase ß-glucan concentration in elite oat germplasm. Crop Sci. 53:1894-1906.

Bernardo, R. 2010. Breeding for quantitative traits in plants. 2nd ed. Stemma Press, Woodbury, MN.

Bernardo, R., and J. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. Crop Sci. 47:1082-1090.

Burrows, P.M. 1972. Expected selection differentials for directional selection. Biometrics 28: 1091-1100.

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

Geiger, H.H., and G.A. Gordillo. 2009. Doubled Haploids in Hybrid Maize Breeding. Maydica 54:485-499.

Hallauer, A.R., and J.B. Miranda Fo. 1981. Quantitative genetics in maize breeding. Iowa State Univ. Press, Ames.

Heffner, E.L., A.J. Lorenz, J.-L. Jannink, and M.E. Sorrells. 2010. Plant breeding with genomic selection: Gain per unit time and cost. Crop Sci. 50:1681-1690.

Heffner, E.L., J. Jannink, and M.E. Sorrells. 2011. Genomic selection accuracy using multi-family prediction models in a wheat breeding program. The Plant Genome 1:1-39.

Heslot, N., D. Akdemir, M.E. Sorrells, J.J. Jannink. 2014. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theor. Appl. Genet. 127:463-480.

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

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

Lian, L., A. Jacobson, S. Zhong, and R. Bernardo. 2014. Genomewide prediction accuracy within 969 maize biparental populations. Crop Sci. 54: 1514-1522.

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

Lorenz, A.J., K.P. Smith, and J.-L. Jannink. 2012. Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. Crop Sci. 52:1609-1621.

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

Massman, J.M., A. Gordillo, R.E. Lorenzana, and R. Bernardo. 2013a. Genomewide predictions from maize single-cross data. Theor. Appl. Genet. 126:13-22.

Massman, J.M., H.-J.G. Jung, and R. Bernardo. 2013b. Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. Crop Sci. 53:58-66.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819-1829.

Piepho, H.P. 2009. Ridge regression and extensions for genome-wide selection in maize. Crop Sci. 49:1165-1176.

Riedelsheimer, C., and A.E. Melchinger. 2013. Optimizing the allocation of resources for genomic selection in one breeding cycle. Theor. Appl. Genet. 126:2835-2848.

Schulz-Streeck, T., J.O. Ogutu, Z. Karaman, C. Knaak, H.P. Piepho. 2012. Genomic selection using multiple populations. Crop Sci. 52:2453-2461.

Searle, S.R. 1971. Topics in variance component estimation. Biometrics 27:1-76.

Sokal, R.R., and C.D. Michener. 1958. A statistical method for evaluating systematic relationships. Univ. Kansas Sci. Bull. V.38, p. 1409-1438.

Sprague, G. F., and W. T. Federer. 1951. A comparison of variance components in corn yield trials. II. Error, year x variety, location x variety and variety components. Agron. J. 43:535-541.

Technow, F., C. Riedelsheimer, T.A. Schrag, and A.E. Melchinger. 2012. Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. Theor. Appl. Genet. 125:1181-1194.

Whittaker, J.C., R. Thompson, and M.C. Denham. 2000. Marker-assisted selection using ridge regression. Genet. Res. 75:249-252.

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

Zhao, Y., M. Gowda, W. Liu, T. Würschum, H.P. Maurer, F.H. Longin, N. Ranc, and J.C. Reif. 2011. Accuracy of genomic selection in European maize elite breeding populations. Theor. Appl. Genet. 124:769-776.