REDUCING AND EXPLOITING GENOTYPE BY ENVIRONMENT INTERACTION IN THE CONTEXT OF GENOMEWIDE PREDICTION IN 969 MAIZE BIPARENTAL POPULATIONS

A THESIS

SUBMITTED TO THE FACULTY OF

THE UNIVERSITY OF MINNESOTA

BY

NICHOLAS CARTER AMES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

ADVISOR: DR. REX BERNARDO

MARCH, 2018

ACKNOWLEDGEMENTS

I would like to acknowledge Monsanto for providing the Bernardo lab with one of the most extensive maize data sets public academics has yet to see. Their assistance has been invaluable. I would also like to thank KWS SAAT and Dr. Dietrich Borchardt for taking me on as an intern. My time in Gottingen and with the data provided from them helped me validate my methods. I would also thank Syngenta, Todd Warner, and Joshua Sleper for allowing me to again validate some of the methods shown here and even extend them to internal projects. My time in industry has been invaluable to knowing what is possible and what is not.

I would also like to acknowledge and thank my advisor, Dr. Rex Bernardo, for his guidance, and the opportunity to work with the University of Minnesota. I interviewed with Dr. Bernardo twice, and it was only after I was ready that I joined the lab. I would also like to thank Dr. Gary Muehlbauer, who helped prepare me for my work as a PhD candidate. Every skill I learned as a masters student in his lab got reused in this project.

Finally, I would like to thank all of my colleagues at the University of Minnesota and at MSI for an experience that pushed me and helped me grow.

DEDICATIONS

To my parents, Carter and Marilyn Ames, for making me work as a busser.

To KTN, for keeping me on course.

ABSTRACT

Multi-environment testing remains crucial in genomewide selection, and environmental effects (E_j) complicate selection. We aimed to: 1) determine if past year's data on previous populations can be used to eliminate environments for a current training population; 2) assess if genomewide predictions can reduce the number of environments used in subsequent phenotypic selection; 3) identify which statistical models and environmental factors are best for estimating E_j; and 4) determine the predictive ability in models that include and exclude genotype × environment interaction effects. A total of 969 Monsanto maize (*Zea mays* L.) populations were genotyped and phenotyped at multiple U.S. locations from 2000 to 2008. Environmental data from the National Oceanic and Atmospheric Administration were gathered and interpolated. The data included 154,000 lines, 448 million marker data points, 3.2 million phenotypic observations, 1395 unique environments, and 1.3 million environmental covariable data points. For 27 biparental crosses that we chose as test populations, environmental stability and an index that used genomewide predictions and phenotypic data could replace one out of four environments in phenotypic evaluation. Correlations between predicted and observed E_j were between 0.25 and 0.35 even when only two environmental factors (precipitation and heat units) were used. A nonfactorial model for line performance in a given environment effectively combined both the line genetic effect and E_j, doubling prediction ability for grain yield and test weight. We speculate that this model can be combined with crop modelling for additional prediction ability in predicting plant performance in a given environment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LITERATURE REVIEW

Phenotype is a function of genotype and environment

A breeder's job is to identify and select for genotypes that perform well in a given environment. These environments can be defined in a number of ways, but are generally classified by management decisions, annual weather conditions for the desired growth period, soil qualities, and more broadly as locations, years, and location × year combinations (Löffler et al. 2005). Ultimately, a phenotype is produced through the combination of environmental conditions and genotypes, and these effects cannot be separated completely. While a breeder might choose a genotype that is advantageous in one environment, this genotype might underperform relative to other genotypes in other environments (Allard and Bradshaw 1964; Baker 1988; Singh et al. 1999). This problem complicates efforts to evaluate many crops, as seeds are often limited in early generations and the ability to test in locations may be limited by available testing site support (Bennetzen and Hake 2009). In addition to this, one of the major factors in plant growth, weather conditions (Rosenzweig et al. 2002; Prasad et al. 2008; Parthasaranthi et al. 2013), are ultimately not able to be controlled leading to often undesired testing conditions.

Genotype by environment (GxE) interaction occurs when plants of interest respond to a group of environments in different ways (Hallauer et al. 1988). The most difficult problem that breeders face is that of crossover interaction where, if a breeder was ranking lines or hybrids, the ranking would change depending upon the environment, thus making selection difficult. Alternatively, there is a non-crossover interaction where relative differences among lines change (Muir et al. 1992; Crossa et al. 2004). Non-

crossover interaction does not change breeder selections, but it does may cause inaccurate estimation of traits in non-tested locations leading to lines not performing as well as advertised.

Heritability and loss in selection response

Heritability can be defined as the portion of phenotypic variance that is accounted for by genetic effects (Falconer et al. 1996). More practically, heritability can be thought of in terms of the breeders equation $R = h^2S$ (Falconer et al. 1996), where R is the response to selection, S is the selection differential, and h^2 is narrow sense heritability (proportion of trait selection that is transmitted through additive genetic effects) (Falconer et al. 1996). When seeking to maximize R, h^2 indicates circumstances that have reduced selection gains.

Environmental effects and genotype by environment interaction enter the equation through heritability (Casler 1982). In this case h^2 estimated on an entry mean basis is equal to

$$h^2 = \frac{\sigma_g^2}{\frac{\sigma_e^2}{re} + \frac{\sigma_{ge}^2}{e} + \sigma_g^2}$$

Where σ_g^2 is genetic variance, σ_e^2 is the environmental variance, σ_{ge}^2 is the genotype by environment interaction, r is the number of replications, and e is the number of environments. When a breeder phenotypes plants, he or she does not directly observe the genotypic value. Phenotype consists of additive effects (that can be transmitted through simple selection, as in the above h^2 equation), dominance effects (that while not transmitted in this example, are important to any breeding program that releases hybrids rather than inbreds), epistatic effects, environmental effects, and genotype by environmental interaction effects (Falconer et al. 1996). Often breeders make sure to test

in environments representative of their target region, or test an entire population in the same locations as a way to try to dissipate some of the environmental variance (Allard and Bradshaw 1964). Conditions may also vary across a location this location variance (that is, error arising from disparate testing conditions) can alter the phenotype that a breeder selects on (Cambardella et al. 1994; Hupet and Vanclooster 2005). For example, if some portion of the testing field receives more nitrogen through a graded slope and water leaching than another portion, and nitrogen increases the height of a maize plant, than the breeder may select in part on non-heritable conditions (i.e. field coordinates). Furthermore, the number of locations that a breeder selects at also determines this realized heritability as the ability to estimate genetic effects increases on average with additional locations, with some diminishing returns (Hallauer et al. 1988). As more locations are selected, the probability of a poor selection decreases..

Of the traits that are commonly evaluated in maize (yield, moisture, test weight, days to silking, days to anthesis, plant height, and ear height), yield is troublesome as it commonly has the lowest heritability, but is the most important. Yield narrow-sense heritability estimates can vary widely, ranging from 0.10 to greater than 0.80 depending on the number of replications, trial locations, and level of inbreeding (Lamkey and Hallauer 1987; Weyhrich et al. 1998; Waqar-Ul-Haq et al. 2008). In addition, yield is highly sensitive to environmental conditions like soil quality, fertility, precipitation, and growing degree days, as well as other key growth point conditions such as drought stress or heat during silking causing abortive embryos (Bänziger et al. 2000; Cattivelli et al. 2008; Prasad et al. 2008). Test weight has a moderate narrow sense heritability at 0.56 (Lian et al. 2014). Moisture is a moderate-high heritability trait, with narrow sense

heritability estimated at 0.66 (Lian et al. 2014). Estimated heritability for days to silking and anthesis was 0.94 in a maize nested association mapping panel (Buckler et al. 2009). Estimated narrow sense heritability for plant height and ear height is 0.39 and 0.33 (Lian et al. 2014). Again, all of these are highly dependent on experimental design and should not be construed as being a definitive estimate.

Root architecture in maize is another trait that is impacted by environmental factors and is important due to tolerance of drought and nutrient management issues (Leach et al. 2011; Vadez 2014). Heritability for total root length was similar under well-watered conditions as it was under water stressed conditions (0.41), but heritability for total surface area of maize root systems significantly increased from 0.38 to 0.54 in one experiment (Li et al. 2015). Because root architecture primarily studies water limitations and response, quantification of the precise nature of the drought conditions here is vital as we are studying a very particular genotype by environment effect.

Fertility management is an example of one environmental impact trait that producers can control, but that still must be selected for. Producers may choose to add nitrogen or other fertilizers to their field in order to increase productivity. However, some maize lines may be differentially responsive to fertility treatments (Gallais and Hirel 2004; Gallais and Coque 2005). Generally the selection for fertility has been for nitrogen use efficiency due to cost and environmental considerations (Moll et al. 1982). Comparison of 89 ex-PVP hybrids of both Iowa Stiff SSynthetic (BSSS) and non-BSSS lines under high and low nitrogen fertilizer conditions (Mastrodomenico et al. 2018) estimated broad sense heritability of nitrogen use efficiency (genetic variation due to additive, dominance, and epistatic genetic sources) to be 0.60, but at high expense to

estimate the nitrogen use efficiency directly. Nitrogen uptake efficiency and nitrogen utilization heritability were very low, with uptake efficiency heritability estimated at 0.27 and utilization at 0.11. Interestingly, broad sense heritability for nitrogen use efficiency under high nitrogen was 0.60, and 0.31 under low nitrogen, underscoring the fact that environmental effect and trial condition managements have drastic effects on the ability of the breeder to select desirable phenotypes (Holland et al. 2003).

Strategies to deal with genotype by environment interaction

All these problems of selection and environment have been recognized as an issue since before even DNA was discovered as the genetic transmission material (Woltereck 1909). The first way to deal with GxE interaction is to simply test across a large range of environments and select for the lines that perform best across enough of the environments (Hallauer et al. 1988). This can be combined the use of checks that represent either current standards for the industry competitors, or from current best lines or hybrids within the program (Schaalje et al. 1987; Moehring et al. 2014). While this is the simplest approach from a logistic standpoint, the difference between the producer's environments and the breeder environments is the largest when compared to other ways of dealing with GxE interaction. This means that while the lines that are advanced may be on average the best over the entire region tested, they may not be the best for each individual producer. This is a relatively simple but effective approach to the issue. An analysis of wheat breeding programs from 1979-1980 and 1987-1988 showed that for varieties that were developed nationally, rather than internationally (through CIMMYT) performed better than those developed for other national breeding groups (Maredia et al. 1996). Interestingly though, internationally developed lines performed better than average than

the nationally developed lines in well-irrigated and high rainfall environments, showing that while GxE interaction may be a major factor, it is possible to change an environment to be more like a target environment and relax the locality requirement for this strategy.

The second strategy is to identify a cluster of environments through cluster analysis (Horner and Frey 1957), principal component analysis (Yan et al. 2001), or additive main effects and multiplicative interaction (AMMI) (Gauch et al. 2008), that would minimize duplication of likely environmental conditions and to represent the majority of conditions that a genotype would normally be grown under. A study of the choice of wheat testing locations in Australia showed that while individual locations may not perfectly predict a region, combinations of locations can consistently average out to the mean yield of an entire region (Hamblin et al. 1980). This of course requires that a region be first defined, but serves an effective way to select locations that can discriminate between lines for that defined region.

The third major strategy is to target each environment with lines that most effectively produce the desired phenotype. This however requires either knowledge of the environmental factors that lead to the desired phenotype (Sadras and Calderini 2015), or an approximation by using a phenotype of some known lines evaluated at many locations as a baseline to create an environmental index (Finlay and Wilkinson 1963). Stability analysis seeks to identify lines that perform consistently over most environments (similar to the cluster strategy above) or "stable" lines, but can also identify lines that are niche and specifically capable (or incapable) in some particular environments ("unstable lines")(Lin et al. 1986). Methods to identify line stability were developed early (Yates

and Cochran 1938) and expanded upon later (Finlay and Wilkinson 1963; Eberhart and Russell 1966; Perkins and Jinks 1968).

While targeting lines to maximize desirable GxE interaction is the most favorable outcome to breeding companies and individual producers who purchase cultivars, it is also the most expensive testing process and lengthens the breeding process considerably by introducing additional years of testing before a line can be sufficiently declared to perform well in a location. Even then, there may still be uncertainty due to variable weather conditions. This issue of time in the form of additional growth seasons, and cost in the form of additional plots that are incorporated into the breeding programs testing regime are the key issues that stand in the way of better cultivar identification and deployment (Fehr 1991).

Evaluating the utility of an environment for line selection

There have been significant efforts to get around some of the issues of GxE interaction through the form of general prediction of the crop performance overall (rather than prediction of a specific genotype) and mechanistic physiological understanding. More advanced strategies require the use of more advanced statistical analysis. These models are more complicated than either regression or cluster analysis, with one of the most widely used set of models for GxE interaction analysis is AMMI and genotype plus genotype-by-environment (GGE) biplot models. The most common types of trials used for this data are known as METs, or multi-environment trials, and have been used extensively in stability analysis and GGE biplots (Yan et al. 2001; Yan and Tinker 2006; Fan et al. 2007). Practical applications of the use of AMMI and GGE models have been shown to be useful in identifying environment grouping and effects in maize (de Oliveira

et al. 2016), soybean (Asfaw et al. 2009), rice (Samonte et al. 2005), wheat (Crossa et al. 1991; Yan et al. 2000), and even okra (Ariyo and Ayo-Vaughan 2000). All three of these models aim to dissect the component parts of variation between genotype, environment, and GxE interaction and help identify desirable components. There is much debate over whether AMMI or GGE is more useful, but functionally they operate to find combinations of genotype and environment that lead to positive or negative interaction effects (Yan et al. 2007).

AMMI works by decomposing the residuals of an analysis of variance (ANOVA) through principal components analysis (PCA), plotting on one axis the principal components for each of the lines and environments residuals and the linear effect of each of the environments and lines on the other axis (or alternatively the first two principle components on each axis) (Gauch 1992). GGE plots the first principal component on one axis, and the second principal component on the other axis, where all data have been environment centered (genotype values minus the environmental mean at that environment) (Yan et al. 2000). Both methods work very similarly and many papers have been written on which one is more effective (Crossa et al. 1991; Gauch et al. 1996; Yan et al. 2001; Yan et al. 2007; Yang et al. 2009; Asfaw et al. 2009).

Ideally, all this environmental selection should minimize the number of trials to make the same selection response per unit time. Another direction that can be taken with environmental classification and GxE interaction identification is decomposing environments into a number of smaller component parts (Campbell et al. 2004; Heslot et al. 2014; Jarquín et al. 2014). With the addition of environmental information on each specific site and year, we can begin to classify environments among all environments that

have been encountered, rather than within groups of other environments. This, however, requires understanding the effect of environments on plant growth.

Crop growth models and estimating environmental effects

Crop growth models or crop models have used half a century of physiological and agronomic research to formalize key growth stages for the major crops (Xinyou and Laar 2005; Cao et al. 2009; Sadras and Calderini 2015). With dozens of potential growth models, and at least nine major crop growth simulation models (Rötter et al. 2012), breeders are often faced with a plethora of choices. Crop models can be divided into three different types: statistical, mechanistic, and functional models (Dourado-Neto et al. 1998).

Statistical models simulate large ranges and trends and can estimate variances due to year or county level effects. This kind of model is the one most frequently used for future general trends such as predicted CO_2 increases, or likelihood of yield losses (Makowski et al. 2015). These models are not meant to address site-specific production predictions, but general trends and large-scale problems. Statistical models are also capable of simulating mechanistic models at scale, given a sufficient number of inputs or training data (Lobell and Burke 2010). As might be expected, as the spatial area of the training model expands, the prediction ability increases. Additionally, if locations are kept consistent (i.e. soil and management conditions do not change across the panel of locations) prediction ability increases as year effects can be isolated. This has been most studied with regards to climate change (precipitation, temperature, and carbon dioxide changes) (Rosenzweig et al. 2002; Lobell and Burke 2010), but can apply to any general meteorological trend.

Process based models are complex physiological based models calculated from a series of known genotypes that model plant reactions at the field, canopy, or plant level (Buck-Sorlin 2013). These models are generally created from well-controlled, small-scale trials that extensively phenotype the plant for the desired parameter or output phenotype. Due to the exhaustive phenotyping required this kind of model is generally created once for a specific crop, or a general model is created that can then be adapted to the entire crop (Basso et al. 2013). This problem of scale has long been recognized (de Wit 1965), but serves at least a place to simulate large amounts of outcomes that then may train other models and lead to understanding of the underlying processes.

Functional models aim to bridge the gap between the inability of the statistical model to predict at fine scale producer and field levels, and the difficulty of scaling up the process model due to its need for a large amount of input parameters (Basso et al. 2013). Functional models do this by taking the simple variables required by the statistical model and that are readily available and simple to measure (precipitation, temperature, sunlight, humidity, etc.) and then produce a more complicated measurement that is approximated through these variables (Basso et al. 2013). A prime example of this is the use of the Penman equation (Penman 1956), which can calculate evapotranspiration using temperature and humidity.

Estimating genotype by environment interaction

Environmental effect can be estimated in a number of different ways through these three types of models (the degree of prediction ability depends on the crop, model used, and scale). However, the most difficult part in estimating the specific interaction between a specific cultivar and the environmental fingerprint (Malosetti et al. 2013).

There are three different scenarios for predicting GxE interactions that are different than the above scenario where trials have already been completed: where the environmental effects have been observed, but not with the lines whose phenotypes need estimating; where the lines have been evaluated, but not in the target environments; and where neither environments nor lines have yet been evaluated.

The first prediction scenario, where environmental effects have been estimated but not with the lines that need to be estimated, is a problem that has been heavily studied and become common practice in annual crops and cattle, with high potential in long breeding cycle plants like fruit trees (Boichard et al. 2012; Kumar et al. 2012). Environmental effects can be estimated using the non-target lines and line effects are estimated using a number of standard methods that have been established for genomic prediction (Goddard and Hayes 2007; Solberg et al. 2008; Würschum et al. 2012; Jacobson et al. 2014; Jacobson et al. 2015). Training population size and relatedness to the test population, and the heritability of the trait affect predictive ability have also been discussed heavily (Solberg et al. 2008; Jacobson et al. 2014; Isidro et al. 2015). The second, where the lines have been evaluated but the environments have not has been studied as part of the issue of crop models has been discussed above (Xinyou and Laar 2005).

The third prediction scenario is where prediction ability needs to be the highest in order for prediction to be useful. Poor prediction ability in predicting either the genetic or environmental effect can be devastating to the overall prediction of the genotype-environment interaction and consequently the prediction of the overall phenotype. Multiple models have attempted to predict the most difficult part of this scenario, which

is the GxE interaction (Guo et al. 2013; Malosetti et al. 2013; Heslot et al. 2014; Jarquín et al. 2014). In the previous models listed above, either observed genetic effects or observed environmental effect form a concrete basis for predictions algorithms to work on related material, but in this last scenario there are no trials that replicate either the lines or the environments.

The primary problem confronted by all these papers is that previous procedures estimate the effects of environments or lines by taking the component parts (genetic markers or environmental covariates, such as those used in the statistical crop model), estimate their individual effects and then sum them to create the total effects. This may be feasible with 100, 1,000, 10,000, or 50,000 markers or covariates with the proper model (to prevent over-fitting) (Honarvar and Rostami 2013). However, the number of interactions is a function of both covariates and markers. What might start with 1,000 markers and 500 environmental covariates ends with 500,000 interactions. We can estimate the effects of each of these interactions, but they become vanishingly small and unable to be estimated accurately in addition to being difficult to compute (Jarquín et al. 2014).

The primary approach, suggested by Malosetti et al. (2013) and followed up by Heslot et al. (2014) and Jarquín et al. (2014), is to model the interaction effect not as a summation of an effect, but rather by estimating it as a random effect with a covariance structure defined by the relationship between known observations (i.e. those with an already observed phenotype), and unknown observations (i.e. those observations that have phenotypic information for neither line nor environments). A similar method has been used in genomic Best Linear Unbiased Prediction (BLUP), where genetic markers

(such as single nucleotide polymorphisms or SNPs) estimate relatedness (or alternatively, pedigrees) and create a relationship matrix (Habier et al. 2013). This relationship matrix combined with the phenotypes for known observations then estimates the unknown observations. This procedure requires knowing the SNP states and environmental covariates specific to the observations to be estimated, but given the advances in environmental sensing and sequencing technology, this is becoming a less costly issue than the problem of phenotyping.

The question then becomes how to estimate the relationship between observations that are both a function of genetic and environmental covariates. The simplest approach (Jarquín et al. 2014) is to multiply the genetic covariance (calculated above as the genomic BLUP example) by the environmental covariance (which can be estimated in much the same way). If we assume that the genetic composition in each environment are independent and identically distributed, then this shortcut should approximate the sum of all interactions without necessarily having to calculate each one individually (Jarquín et al. 2014). We then would be able to calculate interaction effects as an already solved gBLUP problem (Henderson 1976; Piepho et al. 2008), with only computational time and the correct composition of environmental covariates and markers to worry about. This has been shown to be effective in a number of studies in adding additional predictive power in populations designed to identify specific interactions (such as drought tolerance) (Beyene et al. 2015; Zhang et al. 2015), or in inbred multiyear trials (Jarquín et al. 2014; Lopez-Cruz et al. 2015). Further study is required to evaluate whether this would be effective in early generation hybrid development or general prediction for breeding programs.

The task of predicting both the effect of environment, genetics, and genotype by environment interaction is at the forefront of our ability to mesh sequencing technology, adequate computational models, precision agriculture, and remote sensing. All of these technologies are required to solve the growing problem of environmental instability in the face of global climate change and target breeding regions which are rapidly changing character.

CHAPTER 2: Reducing Environmental Evaluations in Genomewide Selection in Maize

Synopsis

Multi-environment testing remains crucial in breeding programs that incorporate genomewide selection. Some environments may be subject to substantially larger genotype by environment interaction effects or large within-environment error variance than other environments. Our objectives were to determine if past year's data on previous populations can be used as a basis for eliminating environments for a current training population and if genomewide predictions can reduce the number of environments used in subsequent phenotypic selection in a test population. Phenotypic and marker data for 969 maize biparental populations were provided to us by Monsanto. From these populations, we chose 27 F_2 populations as A/B test populations. The populations were evaluated at 4 to 12 locations in the U.S. Corn Belt from 2000 to 2008. Up to seven environments out of eight could be eliminated without reducing genomewide selection prediction ability. Environmental stability as a criterion for identifying environments for elimination was better than random elimination of environments past two out of seven eliminations. Use of environmental stability combined with an index that used genomewide selection and phenotypic selection data was capable of replacing one out of four environments of phenotypic evaluation. There was no immediate effect on genomewide selection when predicting with reduced environment populations. We conclude that the use of environmental stability performance in conjunction with an index model that combined phenotypic and genomewide selection information allowed for

reduction of tested environments without significant loss of selection response or prediction ability.

Introduction

Genomewide selection involves predicting the performance of a test population, which has been genotyped, from a training population, which has been phenotyped and genotyped (Meuwissen et al., 2001). Multi-environment testing remains crucial in breeding programs that incorporate genomewide selection. First, the training population needs to be phenotyped at a sufficient number and range of environments to help ensure that genomewide predictions can identify the best candidates in the test population. Second, the best candidates in the test population still need to be phenotyped in multiple environments for further selection. Here we focus on the extent to which the number of environments used to phenotype both the training population and the selected candidates in the test population can be reduced, while maintaining prediction ability.

Just as heritability (h^2) determines the accuracy of phenotypic selection, h^2 also determines the prediction ability of genomewide selection (Daetwyler et al. 2008; Goddard 2009; Luan et al. 2009). The expected prediction ability of genomewide selection is $r^2[Nh^2/(r^2Nh^2 + M_e)]^{1/2}$, where r^2 is the level of linkage disequilibrium, N is the size of the training population, and M_e is the effective number of chromosome segments controlling the trait (Daetwyler et al. 2008; Lian et al. 2014). In turn, the expected entry-mean h^2 is equal to $V_g/(V_G + V_{GE}/e + V_e/re)$, where V_G is the genetic variance, V_{GE} is the genotype by environment interaction variance, V_e is the within-environment error variance, e is the number of environments, and r is the number of replications per environment. The traditional way to increase h^2 is to phenotype in more environments or with more

replications per environment. However, increasing the scope of field testing by increasing e or r is laborious and expensive.

An alternative strategy for increasing h^2 is to identify those environments that, when disregarded from the analysis, would actually increase entry-mean h^2. Some environments may be subject to substantially larger genotype by environment interaction effects or large within-environment error variance than other environments. If the target population of environments remains unchanged, removing such environments from the training population could potentially compensate for the lower e and lead to a higher entry-mean h^2. The extent to which h^2 can be increased or maintained by eliminating specific environments in the training population is unknown.

In practice, we may eliminate environments in the training population after the multi-environment phenotypic data become available, but the best scenario is to eliminate such environments prior to phenotyping the training population in those environments. Doing so requires identifying which environments to eliminate based on previous years' data on prior populations. We need to be confident, however, that eliminating environments based on past years' data does not significantly reduce the prediction ability or response to selection. Information is currently lacking on the ability to eliminate training-population environments on the basis of phenotypic data from previous years (Figure 1).

The entry-mean h^2 affects not only the training population, but also the subsequent phenotyping of candidates that have been selected through genomewide predictions. Suppose genomewide selection in maize (*Zea mays* L.) is conducted via a GCA (general combining ability) Model, in which the lines in an A/B biparental cross are the test population and prior crosses that have A as one parent (A/*) or B as one parent (*/B) are

pooled into the training population (Jacobson et al., 2014). Further suppose that in the breeding program, lines that are subjected to phenotypic selection are routinely evaluated at six locations. If genomewide predictions from the GCA Model are already available for the lines, the number of locations used in subsequent phenotyping can possibly be reduced. We currently lack information on the extent to which the GCA Model can reduce the number of locations in subsequent phenotypic selection.

Our objectives were to determine if (1) the number of environments used to phenotype the training population can be reduced without reducing the prediction ability of or response to genomewide selection; (2) past year's data on previous populations can be used as a basis for eliminating environments for a current training population; (3) genomewide predictions can reduce the number of environments used in subsequent phenotypic selection in a test population; and (4) if this reduction in environments has an immediate effect on genomewide selection in the subsequent prediction of a related population.

Materials and methods

Phenotypic and marker data

Phenotypic and marker data for 969 maize biparental populations were provided to us by Monsanto. These populations were the same ones studied by Jacobson et al. (2014) and Lian et al. (2014). From the 969 populations, we chose 27 populations as 27 different A/B test populations on the basis of the following criteria: having at least four corresponding A/* or */B populations; minimum population size of 50 F_3 lines in the A/B cross; h^2 significantly greater than zero; and the A/B, A/*, and */B populations being all crossed to the same inbred tester. The populations were evaluated at 4 to 12 locations in the U.S.

Corn Belt in 2000 to 2008. All of the phenotypic data were for testcrosses. The traits studied were grain yield (Mg ha^{-1}), moisture (g kg^{-1}), and test weight (kg hL^{-1}).

The parents of the A/B, A/*, and */B populations were genotyped with 2911 single nucleotide polymorphism (SNP) markers. The progeny in each cross were genotyped at a low density with 49 to 100 SNP markers polymorphic between A and B (Jacobson et al., 2014). Marker data were imputed from the parents to the progeny as described by (Jacobson et al., 2014).

Experiment 1: Reducing the number of environments in the training population

To test whether the number of testing environments in the training population can be reduced to increase or maintain the prediction ability and selection response, we calculated h^2 with subsets of the environments used to phenotype each A/* and */B population. For a given A/B test population, suppose an A/* population was phenotyped at e = 6 environments whereas a */B population was phenotyped at e = 7 environments. Assuming that more than half of the environments would be retained, we then calculated h^2 in the A/* population for each possible combination of (e – 1) = 5 environments and (e – 2) = 4 environments. The number of combinations of environments was (${}_nC_r = e!/ e_{Sel}(e - e_{Sel})!$) = 6 for ($e$ – 1) = 5 environments and 15 for (e – 2) = 4 environments, where e_{Sel} was the number of selected environments. The same procedure was done for all possible combinations of the (e – 1) = 6, (e – 2) = 5, and (e – 3) = 4 environments used to phenotype the */B cross.

The original h^2 in each A/* and */B population was estimated as $h^2 = V_G/(V_G + V_R/e)$, where V_R was the residual variance and e was the harmonic mean of the number of environments (Holland et al., 2003). The h^2 in each A/* and */B population for each subset

of environments was estimated as $h^2 = V_G/(V_G + V_R/e_{Sel})$, where e_{Sel} was equal to ($e - 1$), ($e - 2$), etc. as described above. Restricted maximum likelihood estimates of V_G and V_R were obtained with the lme4 package (Bates et al., 2015) in R statistical software (Holland et al., 2003; R Foundation for Statistical Computing, 2011). Data available to us were the testcross mean of each line within each environment, and so the intra-location error variance was confounded with the genotype-by-environment interaction variance in V_R.

The subset of e_{Max} ($\leq e$) environments that resulted in the highest h^2 for each trait was identified for each A/* and */B population. We then used only the phenotypic data from the subset of e_{Max} environments in the subsequent genomewide predictions under a GCA model (Jacobson et al., 2014). Genomewide marker effects were estimated for each of the A/* and */B populations by ridge regression-best linear unbiased prediction (RR-BLUP) as implemented in the rrBLUP package (Endelman, 2011). Final marker effects were estimated as the unweighted mean effect across all A/* and */B crosses (Jacobson et al., 2014).

The RR-BLUP marker effects estimated from the A/* and */B crosses were then used to predict the performance of lines in each A/B test population as $\mathbf{y} = \mathbf{Xm}$, where $\mathbf{y}$ was an $N \times 1$ vector of predicted performance; $\mathbf{X}$ was an $N \times N_M$ incidence matrix with elements of 1 (homozygous for parent A allele), –1 (homozygous for parent B allele), and 0 (heterozygous); and $\mathbf{m}$ was an $N_M \times 1$ vector of RR-BLUP marker effects averaged across the A/* and */B crosses. The 10% of lines with the best predicted performance for each trait were identified, and the observed mean of these lines was denoted by $\bar{X}_{0.10}$. Response to selection in the A/B population was calculated as $\bar{X}_{0.10}$ minus the observed mean of the all of the lines in the A/B test population. Predictive ability (r_{MP}) was calculated as the

correlation between the marker-predicted performance and observed performance of the lines in the A/B test population. A t-test (P = 0.05) was done to compare R with *e* versus e_{Max} environments. A z-transformation was used to compare r_{MP} with *e* versus e_{Max} environments. Differences in the mean R and in the mean r_{MP} with *e* versus e_{Max} environments were tested via a t-test (P = 0.05), considering the variation in R and r_{MP} across the 27 test populations.

To determine if h^2 was an effective criterion for eliminating environments, we compared the r_{MP} and R when one or two environments were removed according to h^2 versus at random. Random removal of (e-e_{Sel}) environments across all populations in a training pool was repeated 75 times and h^2, r_{MP}, and R were averaged across repeats.

Experiment 2: Using past phenotypic data to eliminate environments for a current training population

In Experiment 1, environments were eliminated based on phenotypic data on the training population itself (A/* and */B crosses). In Experiment 2, environments were eliminated based on the performance of other crosses evaluated in the same locations where A/* and */B were evaluated (Figure 1). The criterion used to eliminate a location was the portion of V_R attributed to that location, as estimated from past phenotypic data.

Suppose that individuals in population *k* are phenotyped in a set of locations within a given year. The phenotypic value was modeled as $Y_{ij(k)} = \mu_k + g_{i(k)} + l_{j(k)} + r_{ij(k)}$, where $Y_{ij(k)}$ was the phenotypic value of individual *i* (within population *k*) at location *j*; μ_k was the grand mean of population *k*; $g_{i(k)}$ was the genotypic effect of individual *i* (within population *k*); $l_{j(k)}$ was the effect of location *j* that was used to evaluate population *k*; and $r_{ij(k)}$ was the

residual effect. For population k, the V_R attributable to location j was calculated according to the Shukla stability variance parameter (Shukla 1972; Bernardo 1992) as

$V_{R(j|k)} = \{e\Sigma_i[Y_{ij(k)} - Y_{i.(k)} - Y_{.j(k)} + Y_{..(k)}]^2/[(e-2)(n-1)]\} + \{SS(Residual)/[(e-1)(e-2)(n-1)]\}$,

where n was the number of individuals and SS(Residual) was the residual sum of squares from an ANOVA combined across environments. Given that the concept behind Experiment 2 is analyzing prior years' data for the same location, we ignored year effects and estimated $V_{R(j)}$ for each of 431 total unique locations in the data set in the following manner.

First, we considered all locations in which the first population (k = 1, out of 969 populations) was evaluated. The $V_{R(j|k)}$ for each of the locations for population 1 was estimated according to the above formula for the Shukla (1972) stability variance parameter. The analysis was repeated for each of the remaining 968 populations, with the set of locations being different for each level of k. Second, for the jth location, $V_{R(j)}$ was obtained as the mean of all estimated $V_{R(j|k)}$ values. In other words, $V_{R(j)}$ was estimated as the mean within-location Shukla (1972) stability variance across all populations that were evaluated at location j. The averaging of $V_{R(j|k)}$ values did not include $V_{R(j|k)}$ estimated from a cross that was in the training population itself. Suppose A/C is a cross being considered as part of the training population for A/B, and the goal is to determine which (if any) locations used to evaluate A/C should be eliminated. The A/C population was evaluated in locations 10, 20, 100, 200, and 431. In this situation, data on A/C were not used to estimate $V_{R(10)}$, $V_{R(20)}$, $V_{R(100)}$, $V_{R(200)}$, and $V_{R(431)}$ when A/B was the test population. However, data on A/C were used when neither A nor C were parents of the training and test populations. Locations were eliminated separately for each trait, with the locations with the highest $V_{R(j)}$ values being eliminated first. When a location had no previous information, it was assumed

to have the highest $V_{R(j)}$ and was deleted first. Only 6 of the 289 locations involved in the training populations had no previous information.

To help determine if $V_{R(j)}$ was an effective criterion for eliminating environments, we compared the r_{MP} and R when one to 10 environments were removed from each A/* or */B population according to $V_{R(j)}$ versus at random. Random removal of a given number of environments was repeated 75 times and the r_{MP} and selection response were averaged across repeats. The r_{MP} and R were calculated as in Experiment 1.

Experiment 3: Reducing the number of environments used in subsequent phenotypic selection in a test population

Experiment 3 focused on two questions: (1) equivalency between genomewide predictions with the GCA Model and number of environments used in phenotypic selection; and (2) extent to which GCA Model predictions can reduce subsequent phenotyping. First, we estimated the prediction ability of phenotypic selection by slightly modifying the cross-validation procedure described by Jacobson et al. (2014). Within an A/B test population, the testcross performance of the lines in half of the environments (predictor subset) was considered the predictor of the performance of the same lines in the remaining half of the environments (validation subset). In cases where there was an odd number of environments, the extra environment was assigned to the validation subset. Increasing numbers of environments were removed from the predictor subset, and the mean performance of the lines across the remaining environments was calculated. The correlation between the mean performance in the predictor subset and the mean performance in the validation subset was considered as the predictive ability of phenotypic selection (r_P; Jacobson et al., 2014). For a given number of environments in the predictor

subset, the environments that were eliminated were selected at random. The procedure was repeated 250 times and the results for the prediction ability of phenotypic selection were averaged across the repeats.

For each A/B test population, the r_P value (with different numbers of environments) was compared with the r_{MP} value in Experiment 1, with no environments being removed in the GCA model predictions. When the ratio between r_P and r_{MP} was between 0.95 and 1.05, the corresponding number of environments in the predictor subset was declared as equivalent to the GCA Model prediction.

To test the extent to which GCA model predictions can reduce subsequent phenotyping, we calculated weights for the phenotypic data (b_P) and for the genomewide marker predictions (b_M) as described by Riedelsheimer and Melchinger (2013) and Krchov et al. (2015). The phenotypic means (denoted by *P*) and GCA model predictions (denoted by *M*) were combined in a selection index as $I = b_P P + b_M M$, where b_M/b_P was equal to $(1/h^2 - 1)/[\ 1 - (r_{MP}^2/\ h^2)]$ and where b_M and b_P were subsequently scaled to sum to 1. We studied the scenario in which the standard procedure was to phenotype the A/B lines in either three or four environments. We then estimated R and r_{MP} when the A/B lines were phenotyped in fewer environments (predictor group), with the resulting *P* being combined with *M* in the selection index. Out of the 27 A/B populations, six A/B populations had three environments in the predictor group and 19 A/B populations had four environments in the predictor group. The environments were removed from the predictor group on the basis of $V_{R(j)}$ and at randomWhen only one environment remained in the predictor group, genomewide predictions and phenotypic data were each weighted equally ($b_P = b_M = 0.50$) because h^2 could not be estimated. The remaining environments used to evaluate each A/B

population were considered as the validation environments. The correlation between index values and phenotypic values (r_{IP}) in the validation group was calculated. The R was calculated as described in Experiment 1, with the best 10% of individuals being identified from their index values.

The r_{IP} values were tested for significance (P = 0.05) via a z-transformation. Selection response significance was calculated using a t-test comparing the 95% confidence interval for the full population (calculated using mean square error), and the 95% confidence interval for the top 10% of lines as estimated by genomewide selection. To facilitate comparisons, the maximum r_P (with no environments eliminated in phenotypic selection) was subtracted from r_{IP}. Likewise, the maximum response to phenotypic selection was subtracted from the response to index selection.

Experiment 4: Testing the next generation effects of $V_{R(j)}$ based index selection

Reducing the number of environments (on the basis of $V_{R(j)}$) used to phenotype a population may affect the r_{MP} or R in a future test population (Figure 1). For example, if the number of environments used to phenotype populations A/B and Y/Z is reduced from four to two on the basis of $V_{R(j)}$, would that reduction effect the ability of A/B and Y/Z to serve as a training population for A/Z?

To estimate the effects of environmental removal on genomewide selection in a future test population, we used the procedure from Experiment 2 and eliminated the environments in a stepwise fashion on the basis of $V_{R(j)}$. Only environments from the A/* or */B were removed, and no data from the A/B population was removed. We then calculated R and r_{MP} for the GCA Model for each of the 27 A/B populations. Up to ten environments were eliminated.

Results and discussion

Reducing the number of environments in the training population

The h^2 estimates in the training populations were not maximized by including all environments. The mean number of environments used to phenotype each A/* and */B cross in the training population was $e = 7.18$ (Table 1, Supplemental Table 1-3). The mean number of deleted environments that maximized the h^2 estimate was $e_{Del} = 1.75$ for yield, 1.25 for moisture, and 2.13 for test weight. The mean estimated h^2 then slightly increased from 0.41 to 0.47 for yield, 0.67 to 0.70 for moisture, and 0.57 to 0.63 for test weight when the specific sets of e_{Del} environments were eliminated. Increases in mean estimated h^2 for all traits were significant (P = 0.05).

While deleting specific sets of e_{Del} environments increased the estimated h^2, deleting one or two random environments significantly decreased the mean estimated h^2 for each trait (Table 1) as well as the R and r_{MP}. The decrease, while significant, was minimal.

Despite the increases in estimated h^2 when specific sets of e_{Del} environments were deleted, predictive ability (r_{MP}) and selection response (R) remained unchanged. When the e_{Del} environments were eliminated, the mean r_{MP} did not change by more than 0.01 for each trait, and the changes in mean R were small and nonsignificant for yield (0.21 to 0.19 Mg ha^{-1}), moisture (–6.20 to –6.12 g kg^{-1}), and test weight (0.50 to 0.52 kg hL^{-1})(Table 1). These results indicated that deleting specific sets of environments to maximize the estimated h^2 does not make genomewide selection more effective.

Previous research into what factors lead to higher genomewide prediction ability (Daetwyler et al. 2008; Jannink et al. 2010; Riedelsheimer and Melchinger 2013; Lian et

al. 2014) point to h^2 as a major factor. However, even if the increase in estimated h^2 is statistically significant, an increase of between 0.03 and 0.06 is minimal. As indicated in the introduction, the expected prediction ability of genomewide selection is is $r^2[Nh^2/(r^2Nh^2 + M_e)]^{1/2}$ (Lian et al., 2014). Previous estimates for the dataset in this study (Lian et al., 2014) were $h^2 = 0.55$, $r^2 = 0.46$, $M_e \sim 50$, and N ~2,200. The expected genomewide prediction ability for this set of parameters is 0.650. If h^2 increases to 0.60, the expected genomewide prediction ability increases to only 0.652.

A second explanation for the lack of increase the r_{MP} and R was error in the estimates of h^2. Given our h^2 estimations, we might expect that there would be some amount of error in estimating the variance components. This might due to the fact that there was only one replication within environments, or simply that heritability estimation error can be high (Bogyo 1964; Gill and Jensen 1968). While the combinations with maximized h^2 may be high, they may be inflated by error and the true h^2 is somewhat lower than the estimated heritability. This would mean that the combinations we identified may not truly be higher than the h^2 with the e_{Max} environments, but may appear so due to the inability to more precisely estimate heritability (Hirsch 1990; Nyquist and Baker 1991; Lu et al. 1999).

Using past phenotypic data to eliminate environments from a current training population

In the same way that deleting environments on the basis of h^2 did not increase R and r_{MP}, deleting environments on the basis of $V_{R(j)}$ in Experiment 2 did not increase the mean R and mean r_{MP} (Table 2). The $V_{R(j)}$ is arguably a weaker criterion for deleting environments because it is based on past instead of current (for h^2) data. If h^2 did not increase R or r_{MP}, it was therefore unlikely that $V_{R(j)}$ would do so.

Nevertheless, the results showed that $V_{R(j)}$ was an effective criterion for identifying which environments to delete, if any were to be deleted (as in Experiment 3). Compared with deleting e_{Del} environments at random, deleting e_{Del} environments on the basis of $V_{R(j)}$ was superior but only after two out of a mean of 7.1 environments are deleted from each training population (Table 2). We would expect that $V_{R(j)}$ would be less effective, in terms of the effect on R and r_{MP}, if only a small fraction of environments were deleted. Among the A/* and */B crosses for each A/B test population, the total number of environments (across the 27 A/B populations) was 83.4. After two environments were deleted from each A/* and */B cross, the mean of the total number of environments was 60.4 environments. The large number of environments that remained reduced the effectiveness of $V_{R(j)}$. When one or two environments were deleted, the changes in mean r_{MP} (changes of <0.01) and mean R were often nonsignificant (P = 0.05) (Table 2). This suggests that when one or two environments are deleted for each A/* and */B population, it largely does not matter which environments are deleted.

Compared with deleting e_{Del} environments at random, deleting environments on the basis of $V_{R(j)}$ became more effective when more than two environments were deleted from each of the A/* and */B populations (Table 2, and Supplemental Table 4-9). For e_{Del} = 3, the mean R across the three traits was 8-18% larger when the e_{Del} environments were deleted on the basis of $V_{R(j)}$ than at random. For e_{Del} = 7, the mean R across the three traits was 64-80% larger when the e_{Del} environments were deleted on the basis of $V_{R(j)}$ than at random. For context, there were only 8.4 total environments across each set of A/* and */B populations for e_{Del} = 7.

These results altogether suggest that $V_{R(j)}$ is an effective criterion for identifying environments to delete on the basis of past phenotypic data. Furthermore, if R and r_{MP} were unaffected by whether environments were deleted at random versus on the basis of $V_{R(j)}$, this would have indicated that all environments were equally valuable to genomewide predictions. Because this result was not observed (for $e_{Del} > 2$), we surmise that some environments contributed more to r_{MP} and consequently R than others. The inequality of sites and consequently environments is not a given, considering that sites are often culled for poor trial results. Given these results, we can use $V_{R(j)}$ as a criterion to indicate which environments to delete first in Experiment 3 if more than two environments are to be deleted.

Reducing phenotypic selection using genomewide predictions

Our first goal in this experiment was to measure r_{MP} in terms of how many locations of phenotypic testing a genomewide prediction is equivalent to. Results (Supplemental Table 10) indicate that the GCA Model predictions are on average worth at least one environment out of three or four and are specifically worth 1.6 environments for yield, 1.4 environments for moisture, and 1.8 for test weight. To provide context, A/B populations were tested in, on average, seven environments but the environments were split into a predictor set and a validation set. These results suggested that a GCA Model prediction is worth about ~50% of that of a phenotypic observation if first-year phenotyping of lines in an A/B population is conducted in three locations. This ratio of 46-60% can be compared to previous results reported where genomewide selection was equivalent to 3-4 out of 10 environments (Zhao et al. 2012).

The second goal of this particular experiment was to answer the question: "How many locations may be deleted, while retaining r_{MP} and more importantly R, if a selection index that incorporates phenotypic observations and genomewide predictions (in this case, the GCA model) is used?" First, suppose that a breeder normally evaluates lines in four environments to select the best lines to carry forward. The breeder would like to incorporate genomewide predictions to reduce the number of environments used to evaluate lines, without decreasing R (Figure 1). The results indicated that the breeder could delete one environment without affecting the mean R for yield, moisture, or test weight as long as $V_{R(j)}$ is used to select the environment to be deleted (Supplemental Table 11-22, Figure 2-3, Table 3-4).

In particular, the R for yield and moisture became significantly lower with $e_{Del} = 2$ (out of four). For test weight, $e_{Del} = 1$ led to a slightly higher R while $e_{Del} = 3$ led to a lower R. Moisture r_{MP} became significantly different with $e_{Del} = 1$. Test weight r_{MP} had the same patterns as moisture. Given that test weight was the least important trait studied and that R was the final outcome of interest (but is influenced by r_{MP}), we focused on the R results for yield and moisture in concluding that the use of a selection index allows deleting one environment out of four (Table 3-4).

Second, suppose that a breeder normally evaluates lines in three environments to select the best lines to carry forward. In this situation, up to two environments may be deleted and the selection index can be used without affecting R for any of the three traits. However, the mean R for the selection index when one environment was deleted out of three was smaller than the corresponding R when two environments were deleted out of three. Yield r_{MP} at all values of e_{Del} was not significantly different from the r_P with no

environments deleted. Moisture r_{MP} and R were not significantly worse than the r_P and R with phenotypic selection with no environments deleted. Test weight r_{MP} remained better than r_P with no environments deleted at all values of e_{Del}, but R decreased significantly with $e_{Del} = 2$ (Table 3-4).

Furthermore, for the situation in which lines are normally phenotyped in four locations, the mean R with the selection index was higher when e_{Del} environments were deleted on the basis of $V_{R(j)}$ versus at random (Table 3-4). These results were consistent with those from Experiment 2. For r_{MP}, however, there was no clear difference. When the normal situation is to phenotype the lines in three (instead of four) environments, the use of $V_{R(j)}$ made no difference for R.

Effects of deleting environments on subsequent genomewide selection

The results from Experiment 3 indicated that one or two environments may be deleted from the A/* and */B populations as long as a selection index that incorporates genomewide predictions is used for evaluating the lines in an A/B population. The results in Experiment 4 (Table 5, Supplemental Table 23-30) indicated that when one or two environments have been deleted for the A/B population and the A/B population is used to predict the performance of a future test population, such deletion of environments do not lead to a significant reduction in R and r_{MP}. Deleting even larger numbers of environments per A/* and */B cross did not immediately lead to a decrease in mean r_{MP} and mean R. For yield, mean r_{MP} did not decrease significantly until $e_{Del} = 8$ (Table 5). For $e_{Del} = 7$, yield r_{MP} non-significantly decreased only from 0.20 to 0.17, and yield R decreased only from 0.21 to 0.18 Mg ha^{-1}. The r_{MP} and R for both moisture and test weight were likewise non-significant for $e_{Del} = 7$. For $e_{Del} = 7$, the mean total number of environments across the A/*

and */B populations was 8.43. For $e_{Del} > 7$, however, decreases in R and r_{MP} became mostly significant. This result was not surprising given that the mean number of environments in each A/* and */B population was 7.1. The results for $e_{Del} < 7$ deletions suggested that many of the environments in the training population were not contributing (or were contributing very little) to improved r_{MP} or R.

Conclusions

Overall, our conclusions from this study were fourfold. First, using a combination of environments that maximized h^2 did not increase R or r_{MP}. Second, $V_{R(j)}$ calculated from previous years' performance at the same location significantly outperformed random deletion of environments. Third, the number of environments used in phenotyping can be reduced by one (out of three or four) if a selection index that integrates phenotypic data and genomewide predictions is used, and the locations are chosen on the basis of $V_{R(j)}$. Fourth, at least in the short-term, reducing the number of environments based on $V_{R(j)}$ does not diminish a population's ability to serve as a training population for genomewide selection in future test populations.

CHAPTER 3: Leveraging past trial and environmental information to predict future environmental and interaction effects in maize

Synopsis

Maize (*Zea mays* L.) breeders desire to effectively predict the performance of lines in future environments. Our objectives were to: 1) identify which statistical models and environmental factors are best for estimating environmental effects (E_j); 2) assess if using genetically dissimilar trials to estimate E_j results in higher predictive ability; and 3) determine the predictive ability in models that include and exclude genotype × environment (GxE) interaction effects. Phenotypic and marker data for 969 maize biparental populations were provided by Monsanto. We assessed predictive ability in 24 F_2 testcross populations evaluated at 4 to 12 locations in the U.S. Corn Belt from 2000 to 2008. We gathered environmental data from the National Oceanic and Atmospheric Administration and interpolated environmental information on trial locations. Our results showed that the correlations between predicted and observed E_j were as high as 0.25 to 0.35 even when only two environmental factors are used: precipitation and heat units. Predictive ability decreased when E_j effects from the training data set were eliminated, even when those E_j effects were from trials with less genetically related populations. A nonfactorial model for the performance of a line in a given environment effectively combined both the genetic effect of the line and the E_j effect, without needing to estimate them individually, while maintaining the predictive ability obtained with a traditional

factorial model that included GxE interaction. We expect this low-requirement model to gain additional predictive ability when it is merged with a process-driven crop model.

Introduction

In plant breeding, genotypic performance that changes rank from environment to environment (known as crossover interaction) complicates selection of the "best" cultivars and even questions the concept of whether there can even be a "best" cultivar. Evaluating plant performance at multiple locations over multiple years also has the added cost of phenotyping and time constraints, as well as an upper limit on the number of environments that a line can be tested in, as seed quantities may be limited. Climate change has also made the need to address year, location, and genotype by environment interaction effects more urgent. The need to be able to predict these effects is more vital than ever, as past location effects may not hold true under future climate conditions (Katz and Brown 1992; Choi et al. 2017; Hatfield et al. 2017). If breeders were able to effectively predict the performance of lines in future environments, this would allow for more effective selection.

The performance of the *i*th genotype in the *j*th environment is typically modeled as $y_{ij} = u + G_i + E_j + GxE_{ij} + error$, where u is the grand mean, G_i is the genotypic effect, E_j is the environment effect, and GxE_{ij} is genotype by environment interaction effect. Efforts have largely focused on treating GxE and E_j effects as noise in the model and have focused on predicting G_i (Heffner et al. 2009; Jonas and de Koning 2013). To effectively predict the performance of a specific line in a specific environment, training sets for predicting each of these factors must be assembled. Understanding the impact of environmental factors that lead to the E_j effect and GxE_{ij} effect requires three steps.

The first step is that effectively predicting the performance of a line in a specific environment is predicated on an ability to identify the informative variables for both main

effects, G_i and E_j. Prediction of G_i has largely been studied through genomewide prediction through the use of molecular markers, but breaking down the E_j effect into components in the same way has only started recently (Heslot et al. 2014; Jarquín et al. 2014). By identifying the most influential component factors and using different types of statistical models to predict E_j, the subsequent prediction of GxE_{ij} may be reduced in complexity by only including factors that contribute to the E_j component. By establishing which models predict E_j most accurately, we can use only those models and those factors. While in this study we were unable to use our E_j model in the final prediction y_{ij}, this is still valuable information for other studies.

The second step pertains to the balance between genetic relatedness and environmental similarity. Ideally, we need a training population that is genetically related to the individuals in the test population, and that was phenotyped in environments that are similar to the environments for which y_{ij} is to be predicted. In practice, restricting the training population so that genetic relatedness and environmental similarity (with the test population) are both maximized will reduce the size of the training population to a level that is too small for effective prediction. Focusing on genetic relatedness has allowed a non-arbitrary threshold that has previously proved to be effective (Jacobson et al. 2014; Lian et al. 2014). Because lines are evaluated in a limited number of environments, especially in early generation testing, there is a much harsher trade-off between only using trials of lines that are closely related to the predicted line and using more generalized data. For predicting G_i, this is usually not an issue: for example, using 12 populations each with 120 lines that are closely related to a target line leads to 1,440 lines in the training population. The same 12 populations evaluated in eight environments each

would only generate 96 data points for an environmental training set. Attempting to increase the environmental similarity may lead to more accurate predictions of E_j, but the increase in environmental similarity may decrease genetic relatedness and lead to less accurate predictions of G_i. The effect of such trade-offs on the ability to predict y_{ij} needs to be studied.

The final step is incorporating the environmental factors that predict E_j and molecular markers that predict G_i in a way that is meaningful and computationally feasible. For example, suppose 200 lines have been analyzed for 3000 molecular markers. These lines were evaluated in 15 environments that have been characterized for 400 environmental factors. The 3000 molecular markers and 400 environmental factors lead to 1.2 million interactions. Instead of solving for all 1.2 million interactions, relationship matrices for both environments and markers can be used to solve for $200 \times 15 = 3000$ interaction effects through a relationship-based best linear unbiased prediction (BLUP) model (Jarquín et al. 2014). Incorporation of all factors to solve for all three effects (G_i, E_j, and GxE_{ij}) should answer the question of whether predictive ability is increased and, if so, by how much.

Our objectives in this study were to: 1) understand which type of statistical model and number of factors from publically sourced data are best for estimating E_j; 2) test if using genetically dissimilar trials as a training model to estimate E_j results in higher prediction ability; and 3) determine whether prediction ability is gained when more terms (G+E, GxE, G+E+GxE) are included using a relationship-based BLUP model.

Materials and methods

Phenotypic and marker data

Phenotypic and marker data for 969 maize biparental populations were provided to us by Monsanto. These populations were the same ones studied by Jacobson et al. (2014) and Lian et al. (2014). From the 969 populations, we chose 27 F_2 populations as the A/B (a population whose parents are lines A and B) test populations on the basis of the following criteria: having at least four corresponding A/* or */B populations (populations where one of the parents are lines A or B); minimum population size of 50 F_3 lines in the A/B cross; h^2 significantly greater than zero; and the A/B, A/*, and */B populations being all crossed to the same inbred tester. The populations were evaluated at 4 to 12 locations (mean of 7.1) in the U.S. Corn Belt in 2000 to 2008. All of the phenotypic data were from testcrosses. The trials were performed at 1395 unique year-location combinations. Of these, 969 had recorded planting and harvest dates. These 969 year-location combinations included 431 unique locations and contained 5261 unique trials. The traits studied were grain yield (Mg ha^{-1}), moisture (g kg^{-1}), and test weight (kg hL^{-1}).

The parents of the A/B, A/*, and */B populations were genotyped with 2911 single nucleotide polymorphism (SNP) markers. The progeny in each cross were genotyped at a low density with 49 to 100 SNP markers polymorphic between A and B (Jacobson et al., 2014). Marker data were imputed from the parents to the progeny as described by Jacobson et al. (2014).

Environmental data

For the 969 unique year-locations (hereafter referred to as environments), latitude and longitude data were provided to us, but no additional information about the

environment was available. Data on daily minimum temperature (T_{Min}), maximum temperature (T_{Max}), and precipitation from March 2000 to November 2008 were obtained from the National Climate Data Center for all weather stations located within the US. Hourly data for the same period were obtained from the National Solar Radiation database (National Renewable Energy Laboratory 2007) for modeled global horizontal sunlight, which is the modeled direct and diffuse solar radiation received on a horizontal surface, measured in watt-hours per square meter. These hourly data were then summed to give a daily cumulative score, representing the total of sunlight that a plant might receive that day.

Interpolation on each of these four factors (maximum temperature, minimum temperature, precipitation, and modeled global horizontal sunlight) was performed with ArcGIS (Environmental Systems Research Institute (ESRI) 2016) using the inverse distance weighting tool from the spatial analyst package (McCoy and Johnston 2002) and the 12 nearest points for the raster calculations. We did not have environmental information at each of the locations, so interpolation was required in all cases. This was done separately for each of the four factors, and resulted in a raster for every day between March and November in each of the target years. These rasters were then sampled at the latitude and longitude points to obtain daily conditions at each of the locations for the selected factors.

Growing degree days (GDD), crop heat units (CHU), cumulative precipitation, cumulative growing degree days, and cumulative crop heat units were then calculated. The GDD units were calculated as $[(T_{Min} + T_{Max})/2] - T_{Base}$, with T_{Min} and T_{Min} replaced with T_{Base} if T_{Min} or T_{Min} was less than T_{Base} (McMaster and Wilhelm 1997). T_{Base} was set

to 10° Celsius. The CHU (Brown and Bootsma 1993) was calculated in Celsius as Y_{Max} = [3.33 (T_{Max} - 10)] – [0.084(T_{Max} - 10)^2)] Y_{Min} = [1.8 (T_{Min} - 4.4)], and CHU = (Y_{Max} + Y_{Min})/2. If either Y_{Max} or Y_{Min} were less than 0, then they were then set to 0.

The planting date and harvest date of each population was then used to set the bounds of the environmental information that was used for each population. For example, if a population was planted on April 10, 2003, all data for this day were fetched and labeled as day 1 GDD, day 1 precipitation, day 1 sunlight, etc. Data for April 11 were then fetched and all factors for that day were day 2 factors. This procedure was repeated until the date the population was harvested. The populations were grown in the field for a variable number of days (i.e., some were in the field for 120 days, some 130, etc.). The data were aligned and for excess days when a population had already been harvested but another population was still in the field, a 0 was recorded for the population that had been harvested.

Predicting E_j

To predict the E_j for each of the three traits, five models with two different sets of daily factors were used. These five models were: regression without factor selection, Aikaike Information Criterion (AIC) factor selection and subsequent regression, partial least square regression (PLSR), ridge regression-BLUP (RR-BLUP) with shrinkage estimated uniformly across all factors, and RR-BLUP with shrinkage estimated individually for each factor. The two different sets of factors were: precipitation and CHUs measured on a daily basis, and precipitation, CHUs, sunlight, maximum temperature, and minimum temperature measured on a daily basis. We designated these as the two-factor and five-factor models, with each factor comprising not only one predictor variable but multiple

predictor variables that corresponded to the different dates that the crop was grown (Ritchie and NeSmith 1991; Wilson et al. 1995; Riha et al. 1996; Pirttioja et al. 2015). We chose to use CHUs instead of GDDs according to preliminary results (not shown) that indicated CHUs led to a slight but statistically insignificant increase in predictive ability of E_j. Preliminary analysis likewise showed that a seven-factor model, with cumulative precipitation and cumulative CHUs added, was not superior to the five-factor model.

All models were validated using a leave one environment out cross-validation method on all environments that had available planting and harvest dates. Extreme environmental effects (those that were more than 3.5 standard deviations from the mean) were excluded from the training set. Simple regression used all available daily factors to establish a baseline for the effectiveness of subsequent models. The AIC factor selection method used the R MASS library (Venables and Ripley 2002) for stepwise factor selection. Daily factors were selected using a k value of log(number of variables) and regression was subsequently performed on the selected factors.

The PLSR analysis was performed using the R pls library package (Mevik and Wehrens 2007). The optimal number of components in PLSR was calculated using the pls cross-validation function on all data for each trait with the two-factor and five-factor models. Effects of the components were calculated for each leave one out environment cross-validation.

The RR-BLUP analysis was conducted two ways, using shrinkage estimation for all factors uniformly and using a separate shrinkage factor for each of the types of daily factors (i.e. CHU at all days, precipitations at all days, etc.). To estimate the appropriate

level of shrinkage, we searched on a grid with shrinkage tested at intervals of 0.05, 0.10, 0.15, …, 0.95. The shrinkage factor that led to the highest correlation between predicted E_j and observed E_j was chosen. Shrinkage estimation on a per trait basis was done in a stepwise fashion. All factors were set to 0.50 shrinkage, and the shrinkage for one factor was varied from 0.05 to 0.95 at 0.05 intervals. The shrinkage that led to the highest correlation was then chosen between predicted E_j and observed E_j, that shrinkage level for the factor was set, and the next factor type (CHU, precipitation, etc.) was varied. While this may not have led to the best shrinkage estimation for all factors, calculating all combinations of factors was computationally infeasible as there were 17^5 different combinations.

Predicting $G_i + E_j$

To test whether including E_j effects from populations that were genetically dissimilar increased or decreased the ability to predict y_{ij}, we measured the predictive ability of the general combining ability (GCA) model (Jacobson et al. 2014) combined with two of the environmental prediction models previously tested: RR-BLUP (two factor), and AIC (two factor). The GCA model has been previously evaluated within this dataset (Jacobson et al. 2014; Lian et al. 2014) for predicting G_i. For 24 of the 27 different A/B populations (test populations), a set of populations (training pool) that involved either parent A (A/*) or parent B (*/B) were used as the basis of estimating marker effects using RR-BLUP. We separated G_i from E_j in all populations using a linear model, $y_{ij} = u + E_j + G_j + \text{error}$. The marker effects were then subjected to deregression (Garrick et al. 2009) to test whether this would have an effect on the $y_{ij} = u + E_j + G_j + \text{error}$ model. Three of the 27 test populations originally used by Jacobson et al. (2014) had no planting or harvest date

and so the environmental conditions could not be estimated from National Climate Data Center information.

Suppose an A/B test population was evaluated in environments E1, E2, E3, and E4. To predict the effect of a particular environment for the A/B population, we used three different types of training pools: (i) all E_j values from all populations excluding only E1, E2, E3, and E4; (ii) all E_j values from all populations but excluding the environment being predicted (e.g., including E1, E2, and E3 but not E4 if the effect of E4 is being predicted); and (iii) only the E_j values from the A/* and */B populations. This simulated three scenarios: (i) having data on other populations, including those without A or B as a parent, and assuming the A/B test population has not been phenotyped; (ii) having data on other populations, including those without A or B as a parent, and assuming the A/B test population has been phenotyped in a subset of locations; and (iii) using data only from trials of populations that had either the A parent (A/*) or the B parent (*/B). For these three sets of E_j pools, we took the environmental data (the daily factor information previously used, and listed above), and performed AIC regression as well as RR-BLUP using the daily environmental factors.

These effects (from both AIC and RR-BLUP analyses and for all three training population types) were then used to predict the E_j for the environments where the A/B population was evaluated. This led to six predicted E_j values for each trait. These predicted E_j values were then added to either the deregressed or non-deregressed G_i values, and were considered as an estimate of y_{ij}. We calculated the correlation between the observed and predicted y_{ij}, between the predicted (both deregressed and not) and observed G_i and between the predicted and observed E_j.

Predicting y_{ij} in a nonfactorial model

We estimated y_{ij} as being equal to $\mu + t_{ij}$ + error. As described below, the covariance among t_{ij} values was modeled as the Hadamard product of a **G** matrix (a genetic covariance matrix) and an **E** matrix (an environmental covariance matrix). We used as our training population data source for this prediction model the trial data (environmental and genetic) that was included in the GCA model (scenario iii in the previous section). This linear model therefore did not decompose y_{ij} into G_i, E_j, and GxE_{ij}, but instead treated the genetic value as a single effect.

We created an N × N **G** relationship matrix, where N is the number of line-environment observations in the A/B population and its corresponding A/* and */B populations. The **G** matrix established the genetic relationship between each y_{ij} entry, using 2,504 to 2,883 SNP markers. The **G** matrix was calculated using the 'pedigree' packages calcG function (Coster 2012) which calculates a relationship matrix (an adjusted form of genetic covariance between individuals based on allele identity by state). We also created an N × N **E** covariance matrix for environments based on the daily two-factor environmental information. This matrix was calculated using the covariance function in R. The Hadamard product of the **G** and **E** matrices was designated as the **T** matrix, and modeled the covariance among t_{ij} values.

We then used the kinship.BLUP in the rrBLUP package (Endelman 2011) for R to predict y_{ij} in the test population using only the **T** matrix. This was performed for 22 of the original 27 test populations. Of the 27 populations, three did not have any environmental information due to a lack of planting and harvesting data, while two more were not

solvable in reasonable computational time. We calculated the correlation between predicted and observed y_{ij}.

Full Model

We tested the prediction of y_{ij} as the sum of μ, G_i, E_j, and GxE_{ij}, with all effects being included in within the same mixed model. We used the EMMREML package developed for R as well as the DEoptim package for estimating weighting between the terms (Kang et al. 2008; Zhou and Stephens 2012; Akdemir and Godfrey 2015). The covariance matrices **G**, **E**, and **N** (for interaction) were set as the defined covariance structures, and the correlation between the predicted and observed y_{ij} was calculated. The **N** matrix was equal to **T** as described in the above section. Analysis was not done for the same five A/B populations which were excluded in the analysis described in the previous paragraph. We then compared the use of two separate **G** and **E** relationship matrices ($y_{ij} = u + G_i + E_j +$ residual), the use of the **T** covariance matrix ($y_{ij} = u + T_{ij} +$ error), and the use of the **G**, **E**, and **N** matrix ($y_{ij} = u + G_i + E_j + GxE_{ij} +$ residual).

Results and discussion

Predicting E_j

The correlation between predicted and observed environmental effects (E_j) was affected by both the regression method and the whether a two-factor model (precipitation and CHUs measured on a daily basis) or five-factor model (precipitation, CHUs, sunlight, maximum temperature, and minimum temperature measured on a daily basis) were used. For both the two-factor and five-factor model, regression without any intervening model selection led to nonsignificant (P = 0.05) correlations between the predicted and observed E_j (Table 6). These low correlations, which ranged from r = –0.05 to 0.03 for yield,

moisture, and test weight, were likely a result of overfitting. In contrast, regression with model selection (AIC or PLSR) and RR-BLUP led to significant (P = 0.05) correlations between observed E_j.

Regression with AIC selection of individual environmental covariates worked the best for both five-factor model (r = 0.31) and two-factor model (r = 0.29) for yield, and with the two-factor model (r = 0.36) performing significantly better than the two-factor model (r = 0.25) for moisture (Table 6). In general, the two-factor model was as well as or better than the five-factor model for AIC and RR-BLUP. Partial least squares regression using two and five factors had mixed results, with five-factor model being slightly more effective than two-factor models for yield and moisture but not for test weight. The results across different traits and regression methods indicated that having more factors may not lead to better prediction outcomes, in the same way that saturation of markers beyond a certain point in genomewide prediction may not lead to a better prediction ability.

The PLSR approach reduced the large number of factors into 8–10 principal components. Previous analysis (Jarquín et al. 2014) has suggested that the use of covariates derived from the raw data to reduce the dimensions of data might be sufficient to reduce overfitting problems. Our results, particularly for AIC regression and RR-BLUP, suggested that complex covariates may not be needed if the analysis performed is robust against the problems of overfitting. This was seen in the marked improvement of in the correlations for the AIC and RR-BLUP models over multiple regression with either the daily two-factor or five-factor model, when there was no intervening adjustment to reduce shrinkage.

Our results further showed that the correlations between predicted and observed E_j can be as high as 0.25 to 0.35 even when only two environmental factors are used: precipitation and heat units (Riha et al. 1996; Lobell and Burke 2010). On the other hand, these sizes of these correlations leave room for improvement. We speculate that prediction of E_j could be more accurate if crop models were used in conjunction with the AIC approach we used in this study. In this study, the simple data structure obtained from daily raw data effectively predicted E_j without the need to refer to crop models or stress covariates. These definitions may change from crop to crop or from one germplasm pool to the next and often require specific calibration. But by combining both training population data and crop models, we could leverage the best of both approaches. From crop models, we would gain the power of a process model whereas from the trial data we would gain the breadth of a statistical crop model. This topic warrants further investigation.

Breadth of the training data to predict E_j

Previous studies have shown that for predicting G_i, the predictive ability (r_{MP}) was higher if the training population was restricted to biparental crosses that share one of the two parents with the A/B test population (i.e., GCA model), than if all available biparental crosses were used as the training population (Jacobson et al. 2014; Lian et al. 2014; Jacobson et al. 2015). In contrast, results (Table 7, Supplemental Table 31-33) in this current study indicated that for predicting the performance of line *i* in environment *j* (i.e., y_{ij} approximated as $G_i + E_j$) increasing the amount of environmental training data increases r_{MP}, even if it means taking E_j information from biparental crosses that are less genetically related to the test population. While AIC was superior to RR-BLUP for

predicting E_j (Table 6), RR-BLUP was generally superior to AIC selection for predicting $G_i + E_j$. For yield, for example, the mean r_{MP} was 0.39 with RR-BLUP versus 0.20 with AIC in scenario 1, which used all E_j values in all biparental crosses, except for the specific E_j being predicted. Other scenarios were scenario 2, which used all E_j values except for those in the A/B test population, and scenario 3 which used only the E_j values from the A/* and */B populations in the GCA model. Prediction of G_i was constant across all models, so any changes in r_{MP} among models was due to changes in the prediction of E_j due to the different sets of environmental training data in the three scenarios.

With RR-BLUP, the r_{MP} for yield was 0.39 with scenario 1, 0.33 with scenario 2, and 0.23 with scenario 3 (Table 7). The r_{MP} values were as high as 0.59 for moisture with scenario 1 and 0.49 for test weight with scenario 1, and the r_{MP} values were lowest for both traits in scenario 3. The r_{MP} values with AIC selection showed the same pattern for all traits, decreasing at the same rate between each of the scenarios. These results indicated that r_{MP} decreases with the elimination of E_j effects from the training data set, even when those E_j effects are generated from trials with less genetically related populations.

An expanded set of environments (scenarios 1 and 2) is therefore needed if the objective is to predict y_{ij} instead of only G_i. Furthermore, the need for additional data to predict E_j outweighed the need for a high level of genetic relatedness for predicting G_i. While the environments used to evaluate the A/*, */B, and A/B crosses (i.e., in the GCA model) are not selected according to environmental similarity, we speculate that E_j values for each set of trials are partially a function of maturity and general adaptation which, in

turn, are a function of the genetic background. For predicting G_i, the GCA model provides a simple means for identifying which prior populations should be pooled into a training population. For predicting E_j, an obvious criterion analogous to that in the GCA model does not exist, yet a future study would be to identify environments that meet a certain threshold of similarity to the target set of environments and only use those as the environmental training data.

Predicting y_{ij} in a nonfactorial model

A nonfactorial model, in which y_{ij} was modeled as being equal to T_{ij} + error, led to mean r_{MP} values of 0.25 for yield, 0.23 for moisture, and 0.36 for test weight (Table 8-10). These r_{MP} values were comparable to those for predicting G_i in the GCA model (Jacobson et al., 2014a, 2014b). The range in r_{MP} within test populations was extremely high for moisture and test weight, with some test populations having r_{MP} values of 0.77 for yield, 0.81 for moisture, and 0.84 for test weight, and others having highly negative r_{MP} values of –0.20 for yield, –0.59 for moisture, and –0.62 for test weight. The r_{MP} values were most highly correlated (0.55 for yield, 0.30 for moisture, and 0.41 for test weight) with the number of populations in the training set (Table 8-10). The r_{MP} values was less correlated with the number of observations in the training set for yield (0.44) and test weight (0.31), but not for moisture (0.29).

Full model

The mean r_{MP} values did not differ significantly when y_{ij} were predicted as T_{ij} (nonfactorial model) versus when y_{ij} was predicted as $G_i + E_j + GxE_{ij}$ (factorial model). This result was likely due to the high correlation between the **G** and **T** matrices (r = 0.67 on average) and between the **E** and **T** matrices (r = 0.75 on average), as well as the **T**

matrix (covariance of T_{ij} values) and **N** matrix (covariance of GxE_{ij} values) being identical. These high correlations among the covariance matrices indicated multicollinearity and a difficulty in separating individual effects in a factorial model.

For yield and test weight, the nonfactorial model (T_{ij}) and factorial model (with GxE_{ij}) led to mean r_{MP} values that were double the mean r_{MP} compared to when only G_i + E_j were fitted. On average, yield r_{MP} increased from 0.12 when $G_i + E_j$ was fitted to 0.25 with the nonfactorial model and 0.24 with the factorial model. For test weight, the mean r_{MP} increased from 0.18 when $G_i + E_j$ was fitted to 0.36–0.37 with the nonfactorial and factorial models. For moisture, the mean r_{MP} was 0.26 when $G_i + E_j$ was fitted, and 0.23–0.26 with the nonfactorial and factorial models. This difference is supported by previous results on moisture loss and yield, which placed the R^2 of genotype by environment effects for yield in maize hybrids at 0.266, with the same variance of genotype by environment interaction for ear moisture estimated at 0.177 (Rahman et al. 2010).

Overall, the results indicated that the nonfactorial model effectively combined both the G_i and the E_j effects without needing to estimate them individually, while also incorporating G x E interaction in the form of interactive effects. While the r_{MP} values with the nonfactorial model were somewhat low, we should also compare the r_{MP} for the $G_i + E_j$ model (Table 8-10) to those in Table 7, where r_{MP} was much higher for both AIC and RR-BLUP. If we were to combine T_{ij} which was estimated using a form of GBLUP, with either AIC or RRBLUP using all E_j data (which led to a significantly higher r_{MP}), we speculate that the r_{MP} of this model would be higher still than those in Table 7. We were unable to do this due to the difficulties with only having one replication per location. Furthermore, given the extreme computational time required by the factorial model (e.g.,

1000 hours of CPU time on a 1Tb ram Minnesota Supercomputer Institute server on some of the A/* and */B training populations), we recommend the nonfactorial model as a computationally simpler and more accessible alternative for traits that show G x E interaction. This change in computational time did not come from the per se inversion of additional matrices, but the requirement that with three separate terms, repeated solving through a genetic algorithm was required to estimate the kernel weights for each of the terms. Were the specific kernel weightings of the terms known *a priori* then calculation time would have been similar.

The r_{MP} values in Tables 8-10 may not yet be sufficient to make site-specific cultivar recommendations. We speculate that given the success of process-based crop modeling, information may increase the r_{MP} values to sufficient level (Brisson et al. 2003; Liu et al. 2011; Kim and Kaluarachchi 2015). Further refinement with additional on-site information and field level GPS coordinates would allow for plant stage information and soil conditions specific to the trial, which was not available to us in this data set.

ILLUSTRATIONS

Table 3: Mean predictive ability (r_{MP}) and selection response (R) in heritability maximized versus non-maximized environmental combinations

	Yield				Moisture				Test weight			
Approach	e	h^2	r_{MP}	R	e	h^2	r_{MP}	R	e	h^2	r_{MP}	R
Keep all environments	7.18, (9.33, 5.00)	0.41, (0.17, 0.51)	0.20, (-0.05, 0.36)	0.21, (-0.07, 0.44)	7.18, (9.33, 5.00)	0.67, (0.52, 0.80)	0.42, (-0.09, 0.62)	-6.2, (-11.46, 0.48)	7.18, (9.33, 5)	0.57, (0.44, 0.67)	0.34, (-0.03, 0.52)	0.5, (-0.02, 0.88)
Delete environments to increase h^2	5.43, (4.33, 6.57)	0.47*, (0.33, 0.55)	0.20, (-0.06, 0.36)	0.19, (-0.16, 0.43)	5.93, (4, 7.67)	0.70*, (0.58, 0.81)	0.42, (-0.09, 0.62)	-6.12, (-13.45, 0.40)	5.05, (3.6, 6.14)	0.63*, (0.45, 0.76)	0.35, (-0.04, 0.52)	0.52, (0.08, 0.93)
Delete one random environment	6.18, (8.33, 4.00)	0.38*, (0.16,0.48)	0.19*, (-0.06, 0.36)	0.18*, (-0.06, 0.42)	6.18, (8.33, 4.00)	0.64*, (0.5,0.78)	0.41*, (-0.1, 0.61)	-5.90*, (-11.14, 0.50)	6.18, (8.33, 4.00)	0.51*, (0.24,0.63)	0.33*, (-0.03, 0.51)	0.47*, (-0.10, 0.89)
Delete two random environments	5.18, (7.33, 3.00)	0.33*, (0.15,0.44)	0.18*, (-0.05, 0.36)	0.18*, (-0.04, 0.40)	5.18, (7.33, 3.00)	0.60*, (0.36,0.76)	0.39*, (-0.09, 0.6)	-5.78*, (-11.1, 0.36)	5.18, (7.33, 3.00)	0.46*, (0.19,0.59)	0.32*, (-0.03, 0.51)	0.45*, (-0.10, 0.84)

* Indicates that the value is significantly different at a $p<.05$ level than the value when compared to keeping all environments

e indicates the average number of environments a GCA pool contained per training population.

General Combining Ability (GCA) pools contained 11.48, (28, 3) training populations on average.

Table 2: Gain in mean predictive ability (r_{MP}) and selection response (R) when environments are deleted on the basis of $V_{R(j)}$ instead of at random

Environments deleted	Target populations	Yield		Moisture		Test weight	
		r_{MP}	R	r_{MP}	R	r_{MP}	R
1	27	0.00 (ns), (-0.1, 0.07)	0.00 (ns), (-0.11, 0.09)	0.01, (-0.02, 0.04)	-0.31, (-2.37, 0.64)	0.01, (-0.02, 0.04)	0.02 (ns), (-0.15, 0.12)
2	27	0.01, (-0.06, 0.09)	0.02 (ns), (-0.09, 0.21)	0.02, (-0.03, 0.08)	-0.56, (-2.61, 0.74)	0.01, (-0.06, 0.04)	0.04, (-0.13, 0.19)
3	27	0.02, (-0.06, 0.11)	0.03, (-0.10, 0.25)	0.02, (-0.04, 0.12)	-0.43, (-2.92, 1.79)	0.02, (-0.05, 0.06)	0.06, (-0.09, 0.21)
4	27	0.03, (-0.03, 0.14)	0.04, (-0.14, 0.24)	0.04, (-0.02, 0.16)	-0.58, (-2.61, 2.05)	0.02, (-0.05, 0.13)	0.06, (-0.15, 0.34)
5	27	0.05, (-0.05, 0.11)	0.04, (-0.16, 0.26)	0.06, (-0.02, 0.20)	-1.17, (-6.03, 1.40)	0.03, (-0.08, 0.15)	0.10, (-0.42, 0.37)
6	25	0.06, (-0.03, 0.16)	0.05, (-0.11, 0.29)	0.05, (-0.24, 0.17)	-0.83, (-3.67, 5.49)	0.06, (-0.06, 0.19)	0.12, (-0.03, 0.36)
7	23	0.07, (-0.09, 0.26)	0.08, (-0.10, 0.28)	0.10, (0.00, 0.28)	-2.31, (-6.79, 0.00)	0.11, (-0.02, 0.26)	0.21, (-0.15, 0.53)
8	9	0.01 (ns), (-0.13, 0.15)	0.02 (ns), (-0.14, 0.32)	0.07, (-0.02, 0.20)	-0.92, (-2.27, 0.90)	0.07, (-0.01, 0.17)	0.12, (-0.27, 0.31)
9	4	0.04, (-0.07, 0.21)	0.10, (0.03, 0.16)	0.11, (0.06, 0.21)	-1.77, (-2.22, -1.27)	0.07, (0.02, 0.13)	0.21, (0.14, 0.31)
10	3	-0.04, (-0.14, 0.05)	0.02 (ns), (-0.05, 0.08)	0.07, (0.05, 0.10)	-1.57, (-1.97, -1.37)	0.12, (0.08, 0.16)	0.23, (0.20, 0.25)

ns, not significant at P = 0.05. All other values were significant.

Table 3: Mean predictive ability (r_{MP}) across test populations (range in parenthesis) when genomewide predictions are combined with phenotypic data in a selection index

			Predictor environments			
Trait	Criterion for eliminating environments	Maximum environments	1	2	3	4
Yield	$V_{R(j)}$	3	0.00 (-0.15, 0.07)	0.00 (-0.21, 0.11)	-0.01 (-0.19, 0.11)	
Yield	Random	3	-0.01 (-0.17, 0.07)	-0.01 (-0.22, 0.10)	-0.01 (-0.19, 0.11)	
Yield	$V_{R(j)}$	4	-0.03* (-0.17, 0.09)	-0.04* (-0.27, 0.08)	-0.03* (-0.26, 0.08)	-0.02* (-0.24, 0.08)
Yield	Random	4	-0.04* (-0.16, 0.07)	-0.04* (-0.25, 0.07)	-0.03* (-0.25, 0.08)	-0.02* (-0.24, 0.08)
Moisture	$V_{R(j)}$	3	-0.03 (-0.14, 0.09)	-0.06 (-0.26, 0.02)	0.01 (0.00, 0.02)	
Moisture	Random	3	-0.04* (-0.12, 0.07)	-0.04* (-0.14, 0.01)	0.01 (0.00, 0.02)	
Moisture	$V_{R(j)}$	4	-0.10* (-0.17, 0.06)	-0.08* (-0.38, 0.03)	-0.02* (-0.12, 0.07)	0.00 (-0.13, 0.01)
Moisture	Random	4	-0.11* (-0.16, 0.02)	-0.09* (-0.14, -0.04)	-0.04* (-0.14, -0.01)	0.00 (-0.13, 0.01)
Test weight	$V_{R(j)}$	3	-0.06 (-0.15, 0.04)	0.00 (-0.12, 0.07)	0.04 (0.01, 0.08)	
Test weight	Random	3	-0.04* (-0.08, 0.00)	0.01 (-0.05, 0.07)	0.04 (0.01, 0.08)	
Test weight	$V_{R(j)}$	4	-0.06* (-0.35, 0.07)	-0.02 (-0.2, 0.13)	0.02* (-0.02, 0.12)	0.03* (-0.08, 0.09)
Test weight	Random	4	-0.10* (-0.17, 0.00)	-0.04* (-0.12, 0.07)	0.00 (-0.10, 0.08)	0.03* (-0.08, 0.09)

* indicates significant difference from 0 at a p=0.05.

Table 4: Mean selection response (R) across test populations (range in parenthesis) when genomewide predictions are combined with phenotypic data in a selection index.

			Predictor environments			
Trait	Criterion for eliminating environments	Maximum environments	1	2	3	4
Yield	$V_{R(j)}$	3	-0.01 (-0.26, 0.25)	-0.01 (-0.32, 0.27)	0.01 (-0.15, 0.28)	
Yield	Random	3	-0.01 (-0.24, 0.16)	0.00 (-0.24, 0.26)	0.01 (-0.15, 0.28)	
Yield	$V_{R(j)}$	4	-0.04* (-0.24, 0.21)	-0.04* (-0.33, 0.22)	-0.03 (-0.31, 0.20)	-0.03 (-0.30, 0.19)
Yield	Random	4	-0.04* (-0.25, 0.16)	-0.04* (-0.33, 0.18)	-0.04* (-0.31, 0.18)	-0.03 (-0.30, 0.19)
Moisture	$V_{R(j)}$	3	0.25 (-0.28, 1.86)	1.00 (-0.33, 2.76)	-0.21 (-0.41, 0.02)	
Moisture	Random	3	0.44 (-0.08, 1.12)	0.58 (-0.53, 1.43)	-0.21 (-0.41, 0.02)	
Moisture	$V_{R(j)}$	4	1.42* (-2.74, 5.32)	1.13* (-0.59, 6.32)	0.29 (-1.08, 1.61)	0.01 (-0.43, 1.64)
Moisture	Random	4	1.78* (-0.39, 4.62)	1.37* (0.18, 2.65)	0.59* (-0.11, 1.54)	0.01 (-0.43, 1.64)
Test weight	$V_{R(j)}$	3	-0.15* (-0.34, 0.17)	0.02 (-0.18, 0.36)	0.07 (0.04, 0.10)	
Test weight	Random	3	-0.12* (-0.24, 0.01)	-0.01 (-0.12, 0.09)	0.07 (0.04, 0.10)	
Test weight	$V_{R(j)}$	4	-0.13* (-0.37, 0.12)	-0.04 (-0.43, 0.20)	0.03 (-0.11, 0.15)	0.02 (-0.29, 0.09)
Test weight	Random	4	-0.19* (-0.33, 0.01)	-0.08* (-0.26, 0.05)	-0.02 (-0.28, 0.06)	0.02 (-0.29, 0.09)

* indicates significant difference from 0 at a p=0.05.

Table 5: Mean predictive ability (r_{MP}) and selection response (R) across all populations (range in parenthesis) after deleting different numbers of environments based on $V_{R(j)}$.

Environments deleted	Target populations	Yield		Moisture		Test weight	
		r_{MP}	R	r_{MP}	R	r_{MP}	R
0	27	0.20, (-0.05, 0.36)	0.21, (-0.07, 0.44)	0.42, (-0.09, 0.62)	-6.20, (-11.50, 0.50)	0.34, (-0.03, 0.52)	0.50, (-0.02, 0.88)
1	27	0.19, (-0.07, 0.35)	0.18, (-0.09, 0.45)	0.41, (-0.09, 0.61)	-6.21, (-12.30, 0.50)	0.34, (-0.03, 0.51)	0.49, (-0.02, 0.97)
2	27	0.19, (-0.06, 0.37)	0.19, (-0.04, 0.45)	0.41, (-0.08, 0.60)	-6.34, (-13.20, 1.00)	0.33, (-0.02, 0.53)	0.49, (-0.11, 0.98)
3	27	0.20, (-0.06, 0.37)	0.20, (-0.07, 0.45)	0.41, (-0.07, 0.60)	-5.99, (-11.7, 0.70)	0.32, (0.00, 0.52)	0.50, (-0.04, 0.98)
4	27	0.19, (-0.07, 0.38)	0.20, (-0.04, 0.5)	0.41, (-0.07, 0.63)	-5.93, (-13.5, -0.20)	0.32, (0.02, 0.53)	0.48, (-0.05, 0.89)
5	27	0.19, (-0.03, 0.37)	0.18, (-0.02, 0.38)	0.39, (-0.06, 0.61)	-5.82, (-13.30, 0.20)	0.3, (0.02, 0.52)	0.47, (-0.17, 0.85)
6	25	0.18 [.21], (-0.04, 0.33)	0.17 [.22], (-0.07, 0.44)	0.35 [.42], (-0.18, 0.57)	-5.05 [-6.18], (-9.90, 0.80)	0.3 [.34], (0.04, 0.48)	0.46 [.50], (0.05, 0.76)
7	23	0.17 [.20], (-0.04, 0.37)	0.18 [.20], (-0.26, 0.37)	0.36 [.43], (-0.02, 0.57)	-5.93 [-6.45], (-10.20, 0.60)	0.3 [.35], (0.03, 0.48)	0.49 [.50], (0.18, 0.86)
8	9	0.07* [.21], (-0.11, 0.28)	0.09* [.23], (-0.11, 0.36)	0.28* [.46], (0.12, 0.41)	-3.90* [-6.30], (-6.10, -0.40)	0.21* [.34], (0.05, 0.45)	0.29* [.49], (-0.05, 0.55)
9	4	0.07* [.23], (-0.1, 0.23)	0.16 [.26], (0.13, 0.21)	0.31 [.45], (0.19, 0.48)	-5.90 [-7.78], (-8.00, -3.60)	0.22 [.39], (0.04, 0.38)	0.4 [.56], (0.08, 0.55)
10	3	-0.02* [.23], (-0.18, 0.07)	0.07* [.26], (-0.04, 0.15)	0.23* [.43], (0.12, 0.29)	-5.00* [7.90], (-7.00, -3.60)	0.21* [.36], (0.17, 0.24)	0.34* [.53], (0.13, 0.46)

* significant difference (P = 0.05) from the r_{MP} and R measured when no environments were deleted.

Entries are given as: mean [mean of remaining populations when no environments were deleted], (range)

Table 6: Correlations between predicted and observed environmental effects for two-factor models (crop heat units and precipitation) and five-factor models (crop heat units, precipitation, minimum temperature, maximum temperature, and horizontal sunlight received)

Model type	Yield		Moisture		Test weight	
	5 Factor Model	2 Factor Model	5 Factor Model	2 Factor Model	5 Factor Model	2 Factor Model
RRBLUP (Same Shrinkage†)	0.17d (0.13, 0.21)	0.16d (0.13, 0.19)	0.23c (0.19, 0.27)	0.15b (0.11, 0.18)	0.21d (0.17, 0.25)	0.13c (0.1, 0.17)
RRBLUP (Different Shrinkage†)	0.19d (0.15, 0.24)	0.17d (0.13, 0.21)	0.26c (0.21, 0.3)	0.17b (0.13, 0.2)	0.16cd (0.12, 0.21)	0.13c (0.1, 0.17)
Partial Least Square Regression	0.10c (0.07, 0.12)	0.15d (0.12, 0.17)	0.21c (0.18, 0.24)	0.15b (0.12, 0.18)	0.13c (0.1, 0.16)	0.19d (0.16, 0.15)
Multiple Regression with AIC selection	0.29e (0.27, 0.32)	0.31e (0.28, 0.34)	0.36d (0.34, 0.39)	0.25c (0.22, 0.27)	0.31f (0.28, 0.33)	0.26e (0.24, 0.29)
Multiple Regression	0.03b (0, 0.06)	-0.02a (-0.05, 0.01)	-0.05a (-0.08, -0.02)	-0.05a (-0.08, -0.02)	0.04b (0.01, 0.07)	-0.02a (-0.05, 0.01)

Correlations followed by the same letter within a trait were not significantly different (P = 0.05), with 95% confidence intervals in parentheses.

RRBLUP: Ridge regression best linear unbiased predictor

AIC: Aikaike information criterion

†: Shrinkage chosen individually for each of the types of factors (crop heat units, precipitation, etc.) or uniform shrinkage across all factors

Table 7. Predictive ability for the sum of genetic and environment effects ($G_i + E_j$) under three scenarios for the source of environmental training data for an A/B test population: (i) having data on other populations, including those without A or B as a parent, and assuming the A/B test population has not been phenotyped; (ii) having data on other populations, including those without A or B as a parent, and assuming the A/B test population has been phenotyped in a subset of locations; and (iii) using data only from trials of populations that had either the A parent (A/*) or the B parent (*/B)

Trait	*Scenario*	*Model Type*	
		AIC	RR-BLUP
Yield	i	0.20b (0.15, 0.25)	0.39c (0.34, 0.44)
	ii	0.23b (0.18, 0.28)	0.33c (0.28, 0.37)
	iii	0.13a (0.08, 0.18)	0.23b (0.18, 0.28)
Moisture	i	0.51c (0.47, 0.55)	0.59c (0.56, 0.62)
	ii	0.37b (0.33, 0.41)	0.50c (0.46, 0.54)
	iii	0.32b (0.27, 0.36)	0.22a (0.18, 0.26)
Test weight	i	0.29a (0.25, 0.34)	0.49c (0.45, 0.53)
	ii	0.36ab (0.32, 0.4)	0.42bc (0.37, 0.46)
	iii	0.21a (0.16, 0.26)	0.32ab (0.27, 0.36)

Correlations followed by the same letter within a trait were not significantly different (P = 0.05), with 95% confidence intervals in parentheses.

Table 8. Yield predictive ability (r_{MP}) for a model that fitted the sum of genetic and environmental effects ($G_i + E_j$), a nonfactorial model (T_{ij}), and a factorial model ($G_i + E_j + GxE_{ij}$)

TEST POPULATION		**TRAINING POPULATION**			**PREDICTIVE ABILITY**			**CORRELATION BETWEEN COVARIANCE MATRICES**	
NAME	Observations	Crosses	Observations	Environments	$G_i + E_j$	T_{ij}	$G_i + E_j + GxE_{ij}$	E, T	G, T
P37/P38	824	5	3214	25	-0.21a (-0.28, -0.14)	0.27b (0.20, 0.34)	0.30b (0.23, 0.36)	0.82 (0.82, 0.82)	0.68 (0.67, 0.68)
P33/P34	1360	7	3040	51	0.1a (0.05, 0.16)	0.19b (0.14, 0.24)	0.31c (0.25, 0.35)	0.61 (0.61, 0.61)	0.87 (0.87, 0.87)
P25/P22	1350	19	15180	146	0.44a (0.40, 0.49)	0.50a (0.46, 0.54)	0.48a (0.44, 0.52)	0.92 (0.92, 0.92)	0.36 (0.36, 0.36)
P28/P27	1080	10	11829	72	0.52b (0.47, 0.56)	0.36a (0.31, 0.41)	0.35a (0.3, 0.4)	0.88 (0.88, 0.88)	0.57 (0.57, 0.57)
P24/P26	1464	17	16665	127	0.31a (0.26, 0.36)	0.42b (0.37, 0.46)	0.44b (0.4, 0.49)	0.76 (0.76, 0.76)	0.65 (0.65, 0.65)
P26/P27	1472	16	16937	119	0.12b (0.06, 0.17)	0.12b (0.07, 0.17)	-0.04a (-0.09, 0.01)	0.81 (0.81, 0.81)	0.55 (0.55, 0.55)
P41/P42	1281	3	4074	28	-0.1b (-0.15, -0.04)	0.06c (0.01, 0.12)	-0.26a (-0.31, -0.2)	0.67 (0.67, 0.67)	0.81 (0.81, 0.81)

P23/P24	1218	16	15576	120	0.52a (0.47, 0.56)	0.77b (0.75, 0.79)	0.76b (0.74, 0.79)	0.73 (0.73, 0.73)	0.75 (0.75, 0.75)
P29/P27	1215	10	11552	73	-0.09b (-0.14, -0.03)	-0.20a (-0.25, -0.15)	-0.26a (-0.31, -0.21)	0.73 (0.73, 0.73)	0.75 (0.75, 0.75)
P35/P36	1442	4	4982	30	0.03a (-0.02, 0.08)	0.07a (0.01, 0.12)	0.05a (0.00, 0.10)	0.90 (0.90, 0.90)	0.47 (0.47, 0.47)
P13/P14	1412	7	7754	53	-0.15a (-0.20, -0.10)	-0.01b (-0.06, 0.04)	-0.03b (-0.08, 0.02)	0.84 (0.84, 0.84)	0.59 (0.59, 0.59)
P2/P15	921	8	8526	61	0.19a (0.13, 0.25)	0.24a (0.18, 0.3)	0.21a (0.14, 0.27)	0.73 (0.73, 0.73)	0.91 (0.91, 0.91)
P1/P9	867	11	9185	65	0.17a (0.11, 0.24)	0.12a (0.05, 0.18)	0.13a (0.06, 0.19)	0.88 (0.88, 0.88)	0.57 (0.57, 0.57)
P11/P12	1450	7	6852	51	0.09a (0.03, 0.14)	0.30b (0.25, 0.35)	0.35b (0.31, 0.4)	0.49 (0.49, 0.49)	0.92 (0.92, 0.92)
P19/P20	1104	3	2780	17	0.16a (0.10, 0.21)	0.21a (0.15, 0.27)	0.21a (0.15, 0.27)	0.60 (0.60, 0.60)	0.82 (0.82, 0.82)
P9/P10	871	12	8208	68	0.49a (0.44, 0.54)	0.48a (0.42, 0.53)	0.48a (0.42, 0.53)	0.71 (0.71, 0.71)	0.58 (0.58, 0.58)
P17/P18	1465	4	3121	25	-0.02b (-0.07, 0.04)	-0.14a (-0.19, -0.09)	-0.14a (-0.19, -0.09)	0.66 (0.66, 0.66)	0.84 (0.84, 0.84)

P4/P5	1062	18	19656	141	0.04a (-0.02, 0.1)	0.13b (0.07, 0.19)	0.14b (0.08, 0.20)	0.76 (0.76, 0.76)	0.66 (0.66, 0.66)
P16/P13	1423	6	6117	44	0.13a (0.08, 0.19)	0.13a (0.08, 0.19)	0.13a (0.07, 0.18)	0.68 (0.68, 0.68)	0.43 (0.43, 0.43)
P6/P7	1281	14	15488	114	-0.04a (-0.1, 0.01)	0.60b (0.56, 0.63)	0.60b (0.56, 0.63)	0.68 (0.68, 0.68)	0.55 (0.55, 0.55)
P3/P8	1267	9	8495	64	0.31a (0.26, 0.37)	0.35ab (0.30, 0.41)	0.37b (0.32, 0.42)	0.72 (0.72, 0.72)	0.70 (0.70, 0.70)
P3/P4	1308	17	18762	126	-0.29a (-0.34, -0.24)	0.44b (0.40, 0.48)	0.60c (0.57, 0.64)	0.81 (0.81, 0.81)	0.70 (0.70, 0.70)
MEANS	1234	10.14	9909	73.64	0.12a (0.07, 0.18)	0.25b (0.19, 0.30)	0.24b (0.19, 0.29)	0.75 (0.74, 0.75)	0.67 (0.67, 0.67)

Correlations followed by the same letter within a trait were not significantly different (P = 0.05), with 95% confidence intervals in parentheses. Groupings are assigned per row.

G matrix is the covariance matrix for observations created through line single nucleotide polymorphism markers

E matrix is the covariance matrix for observations created through environment covariates

T is the Hadamard product of G and E.

Table 9: Moisture predictive ability (r_{MP}) for a model that fitted the sum of genetic and environmental effects ($G_i + E_j$), a nonfactorial model (T_{ij}), and a factorial model ($G_i + E_j + GxE_{ij}$)

TEST POPULATION		**TRAINING POPULATION**			**PREDICTIVE ABILITY**			**CORRELATION BETWEEN COVARIANCE MATRICES**	
NAME	Observations	Crosses	Observations	Environments	$G_i + E_j$	T_{ij}	$G_i + E_j + GxE_{ij}$	E, T	G, T
P37/P38	824	5	3214	25	0.61b (0.57, 0.65)	0.51a (0.46, 0.56)	0.52a (0.47, 0.57)	0.82 (0.82, 0.82)	0.68 (0.67, 0.68)
P33/P34	1360	7	3040	51	0.23a (0.18, 0.28)	0.69b (0.66, 0.72)	0.72b (0.69, 0.74)	0.61 (0.61, 0.61)	0.87 (0.87, 0.87)
P25/P22	1350	19	15180	146	0.16a (0.11, 0.22)	0.35b (0.30, 0.40)	0.41c (0.36, 0.45)	0.92 (0.92, 0.92)	0.36 (0.36, 0.36)
P28/P27	1080	10	11829	72	0.64a (0.60, 0.67)	0.76b (0.74, 0.79)	0.90c (0.88, 0.91)	0.88 (0.88, 0.88)	0.57 (0.57, 0.57)
P24/P26	1464	17	16665	127	0.67a (0.64, 0.70)	0.73b (0.7, 0.75)	0.80c (0.78, 0.82)	0.76 (0.76, 0.76)	0.65 (0.65, 0.65)
P26/P27	1472	16	16937	119	0.63a (0.60, 0.66)	0.81c (0.8, 0.83)	0.77b (0.75, 0.79)	0.81 (0.81, 0.81)	0.55 (0.55, 0.55)

P41/P42	1281	3	4074	28	0.12c (0.06, 0.17)	0.02b (-0.04, 0.07)	-0.21a (-0.27, -0.16)	0.67 (0.67, 0.67)	0.81 (0.81, 0.81)
P23/P24	1218	16	15576	120	-0.25a (-0.30, -0.20)	-0.09b (-0.15, -0.04)	-0.22a (-0.27, -0.16)	0.73 (0.73, 0.73)	0.75 (0.75, 0.75)
P29/P27	1215	10	11552	73	0.13b (0.07, 0.18)	-0.36a (-0.41, -0.31)	-0.32a (-0.37, -0.27)	0.73 (0.73, 0.73)	0.75 (0.75, 0.75)
P35/P36	1442	4	4982	30	-0.09b (-0.14, -0.04)	-0.32a (-0.37, -0.27)	0.16c (0.11, 0.21)	0.90 (0.90, 0.90)	0.47 (0.47, 0.47)
P13/P14	1412	7	7754	53	0.62a (0.59, 0.66)	0.71c (0.68, 0.74)	0.69b (0.66, 0.72)	0.84 (0.84, 0.84)	0.59 (0.59, 0.59)
P2/P15	921	8	8526	61	0.06a (-0.01, 0.13)	0.05a (-0.02, 0.12)	0.08a (0.01, 0.15)	0.73 (0.73, 0.73)	0.91 (0.91, 0.91)
P1/P9	867	11	9185	65	0.84b (0.82, 0.86)	0.13a (0.06, 0.19)	0.18a (0.11, 0.24)	0.88 (0.88, 0.88)	0.57 (0.57, 0.57)
P11/P12	1450	7	6852	51	0.08a (0.03, 0.13)	0.21b (0.16, 0.26)	0.32c (0.27, 0.36)	0.49 (0.49, 0.49)	0.92 (0.92, 0.92)
P19/P20	1104	3	2780	17	0.20a (0.14, 0.25)	0.2a (0.14, 0.25)	0.21a (0.15, 0.27)	0.60 (0.60, 0.60)	0.82 (0.82, 0.82)
P9/P10	871	12	8208	68	0.24b (0.17, 0.30)	-0.34a (-0.39, -0.28)	-0.31a (-0.37, -0.25)	0.71 (0.71, 0.71)	0.58 (0.58, 0.58)

P17/P18	1465	4	3121	25	-0.37b (-0.42, -0.33)	-0.59a (-0.62, -0.55)	-0.59a (-0.62, -0.55)	0.66 (0.66, 0.66)	0.84 (0.84, 0.84)
P4/P5	1062	18	19656	141	0.3a (0.24, 0.35)	0.64b (0.60, 0.67)	0.62b (0.58, 0.65)	0.76 (0.76, 0.76)	0.66 (0.66, 0.66)
P16/P13	1423	6	6117	44	-0.21a (-0.26, -0.16)	-0.12b (-0.17, -0.07)	-0.20a (-0.25, -0.15)	0.68 (0.68, 0.68)	0.43 (0.43, 0.43)
P6/P7	1281	14	15488	114	0.72c (0.69, 0.74)	0.23a (0.18, 0.28)	0.32b (0.27, 0.37)	0.68 (0.68, 0.68)	0.55 (0.55, 0.55)
P3/P8	1267	9	8495	64	0.63a (0.60, 0.66)	0.72b (0.69, 0.74)	0.71b (0.69, 0.74)	0.72 (0.72, 0.72)	0.70 (0.70, 0.70)
P3/P4	1308	17	18762	126	-0.23a (-0.28, -0.18)	0.02b (-0.03, 0.08)	0.10c (0.04, 0.15)	0.81 (0.81, 0.81)	0.70 (0.70, 0.70)
MEANS	1234	10.14	9909	73.64	0.26a (0.21, 0.31)	0.23a (0.18, 0.27)	0.26a (0.21, 0.3)	0.75 (0.74, 0.75)	0.67 (0.67, 0.67)

Correlations followed by the same letter within a trait were not significantly different (P = 0.05), with 95% confidence intervals in parentheses. Groupings are assigned per row.

G matrix is the covariance matrix for observations created through line single nucleotide polymorphism markers

E matrix is the covariance matrix for observations created through environment covariates

T is the Hadamard product of G and E.

Table 10: Test weight predictive ability (r_{MP}) for a model that fitted the sum of genetic and environmental effects ($G_i + E_j$), a nonfactorial model (T_{ij}), and a factorial model ($G_i + E_j + GxE_{ij}$)

TEST POPULATION		**TRAINING POPULATION**			**PREDICTIVE ABILITY**			**CORRELATION BETWEEN COVARIANCE MATRICES**	
NAME	Observations	Crosses	Observations	Environments	$G_i + E_j$	T_{ij}	$G_i + E_j + GxE_{ij}$	E, T	G, T
P37/P38	824	5	3214	25	0.2b (0.12, 0.28)	0.12a (0.04, 0.20)	0.11a (0.02, 0.19)	0.82 (0.82, 0.82)	0.68 (0.67, 0.68)
P33/P34	1360	7	3040	51	0.27a (0.22, 0.32)	0.76b (0.74, 0.78)	0.77b (0.75, 0.79)	0.61 (0.61, 0.61)	0.87 (0.87, 0.87)
P25/P22	1350	19	15180	146	0.34a (0.29, 0.38)	0.76b (0.73, 0.78)	0.78b (0.75, 0.80)	0.92 (0.92, 0.92)	0.36 (0.36, 0.36)
P28/P27	1080	10	11829	72	0.45a (0.41, 0.50)	0.76c (0.73, 0.78)	0.6b (0.56, 0.64)	0.88 (0.88, 0.88)	0.57 (0.57, 0.57)
P24/P26	1464	17	16665	127	0.69a (0.66, 0.72)	0.74ab (0.71, 0.76)	0.76b (0.73, 0.78)	0.76 (0.76, 0.76)	0.65 (0.65, 0.65)
P26/P27	1472	16	16937	119	0.50a (0.46, 0.55)	0.84b (0.82, 0.85)	0.78b (0.76, 0.81)	0.81 (0.81, 0.81)	0.55 (0.55, 0.55)

P41/P42	1281	3	4074	28	0.23a (0.16, 0.29)	0.54b (0.49, 0.58)	0.49b (0.44, 0.54)	0.67 (0.67, 0.67)	0.81 (0.81, 0.81)
P23/P24	1218	16	15576	120	-0.05a (-0.1, 0.01)	0.30b (0.25, 0.35)	0.52c (0.48, 0.56)	0.73 (0.73, 0.73)	0.75 (0.75, 0.75)
P29/P27	1215	10	11552	73	-0.26c (-0.32, -0.21)	-0.62a (-0.65, -0.58)	-0.46b (-0.51, -0.42)	0.73 (0.73, 0.73)	0.75 (0.75, 0.75)
P35/P36	1442	4	4982	30	0.05a (0.00, 0.11)	0.05a (0.00, 0.11)	0.04a (-0.01, 0.10)	0.90 (0.90, 0.90)	0.47 (0.47, 0.47)
P13/P14	1412	7	7754	53	0.11a (0.06, 0.17)	0.23a (0.18, 0.28)	0.17a (0.11, 0.22)	0.84 (0.84, 0.84)	0.59 (0.59, 0.59)
P2/P15	921	8	8526	61	0.14a (0.07, 0.21)	0.82b (0.79, 0.84)	0.74b (0.71, 0.77)	0.73 (0.73, 0.73)	0.91 (0.91, 0.91)
P1/P9	867	11	9185	65	0.2a (0.13, 0.26)	0.56b (0.52, 0.61)	0.56b (0.51, 0.60)	0.88 (0.88, 0.88)	0.57 (0.57, 0.57)
P11/P12	1450	7	6852	51	0.06a (0.01, 0.11)	0.23b (0.18, 0.28)	0.27b (0.22, 0.32)	0.49 (0.49, 0.49)	0.92 (0.92, 0.92)
P19/P20	1104	3	2780	17	0.08a (0.01, 0.15)	0.24b (0.17, 0.30)	0.26b (0.20, 0.32)	0.60 (0.60, 0.60)	0.82 (0.82, 0.82)
P9/P10	871	12	8208	68	0.33a (0.27, 0.38)	0.54b (0.49, 0.59)	0.56b (0.51, 0.60)	0.71 (0.71, 0.71)	0.58 (0.58, 0.58)

P17/P18	1465	4	3121	25	0.22b (0.17, 0.27)	0.13a (0.08, 0.19)	0.19ab (0.14, 0.25)	0.66 (0.66, 0.66)	0.84 (0.84, 0.84)
P4/P5	1062	18	19656	141	0.56a (0.52, 0.60)	0.56a (0.52, 0.60)	0.57a (0.53, 0.61)	0.76 (0.76, 0.76)	0.66 (0.66, 0.66)
P16/P13	1423	6	6117	44	-0.62a (-0.65, -0.59)	-0.62a (-0.65, -0.58)	-0.63a (-0.66, -0.6)	0.68 (0.68, 0.68)	0.43 (0.43, 0.43)
P6/P7	1281	14	15488	114	0.42b (0.37, 0.46)	0.26a (0.20, 0.31)	0.32a (0.27, 0.37)	0.68 (0.68, 0.68)	0.55 (0.55, 0.55)
P3/P8	1267	9	8495	64	-0.12a (-0.17, -0.06)	0.21b (0.15, 0.26)	0.25b (0.20, 0.30)	0.72 (0.72, 0.72)	0.70 (0.70, 0.70)
P3/P4	1308	17	18762	126	0.20a (0.14, 0.25)	0.51b (0.47, 0.55)	0.59c (0.56, 0.63)	0.81 (0.81, 0.81)	0.70 (0.70, 0.70)
MEANS	1234	10.14	9909	73.64	0.18a (0.13, 0.23)	0.36b (0.32, 0.40)	0.37b (0.33, 0.42)	0.75 (0.74, 0.75)	0.67 (0.67, 0.67)

Correlations followed by the same letter within a trait were not significantly different (P = 0.05), with 95% confidence intervals in parentheses. Groupings are assigned per row.

G matrix is the covariance matrix for observations created through line single nucleotide polymorphism markers

E matrix is the covariance matrix for observations created through environment covariates

T is the Hadamard product of G and E.

Figure 1. Four experiments involving the reduction of the number of environments in genomewide selection. In Experiment 1, the number of environments in the training populations (GCA pool) was reduced to in an attempt to maximize heritability. In Experiment 2, the number of environments in the training population was reduced on the basis of environment instability, $V_{R(j)}$. In Experiment 3, the number of environments used in phenotypic selection was reduced by including genomewide predictions in a selection index. In Experiment 4, tests were conducted to determine whether reducing the number of environments in Experiment 3 negatively affected a future target population.

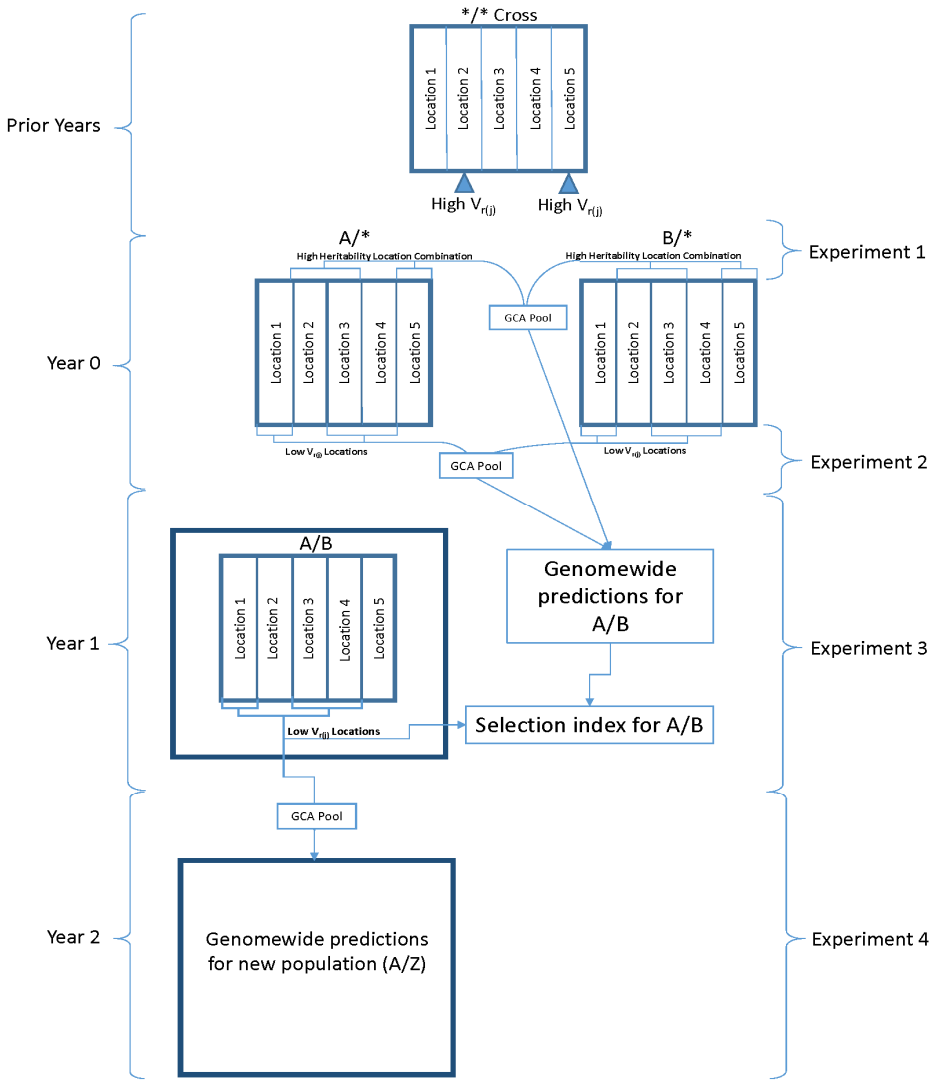


Figure 2: Mean predictive ability (r_{MP}) for A/B test populations with environments deleted (out of three or four) on the basis of $V_{R(j)}$ or at random. Bars indicate the range of values.

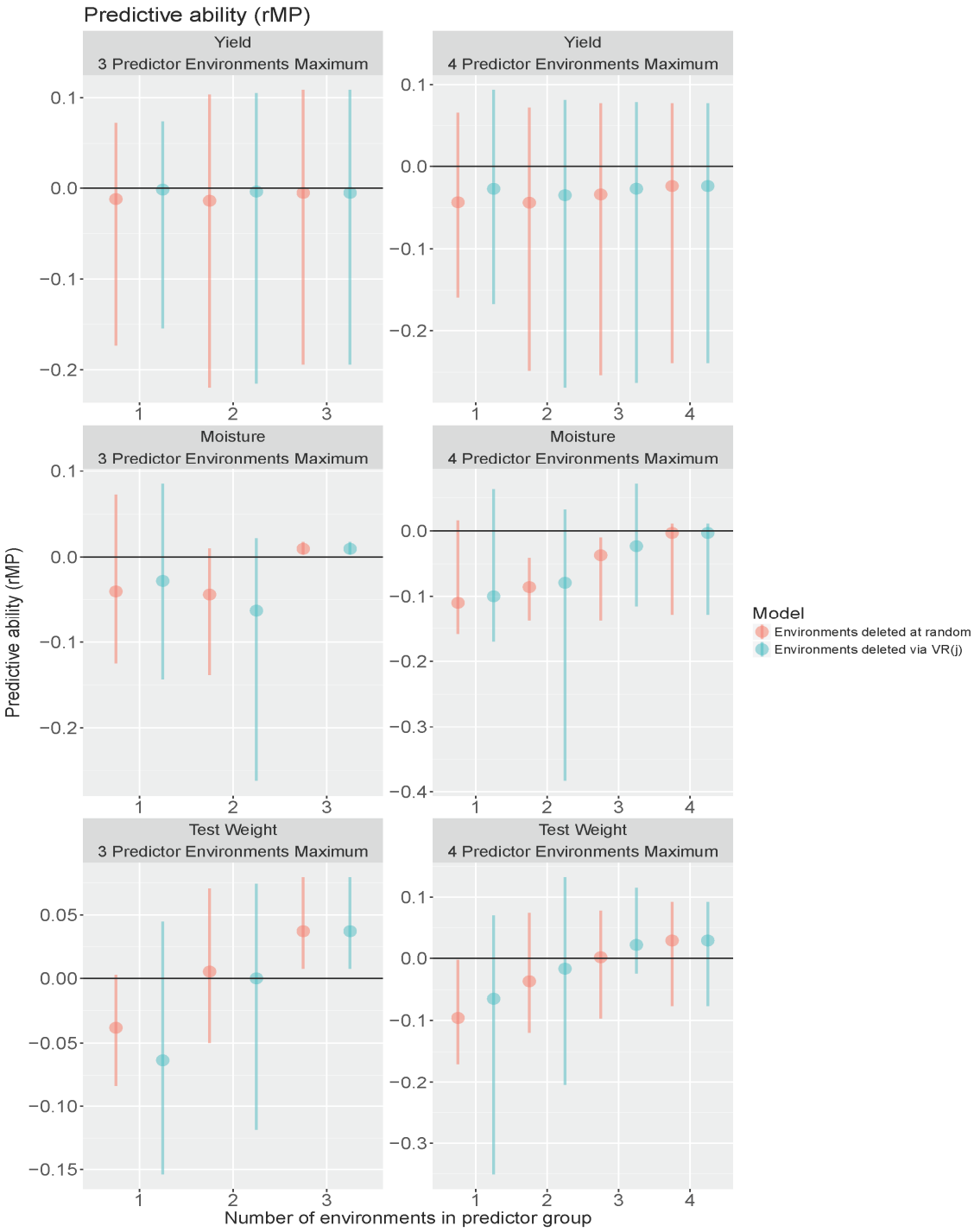
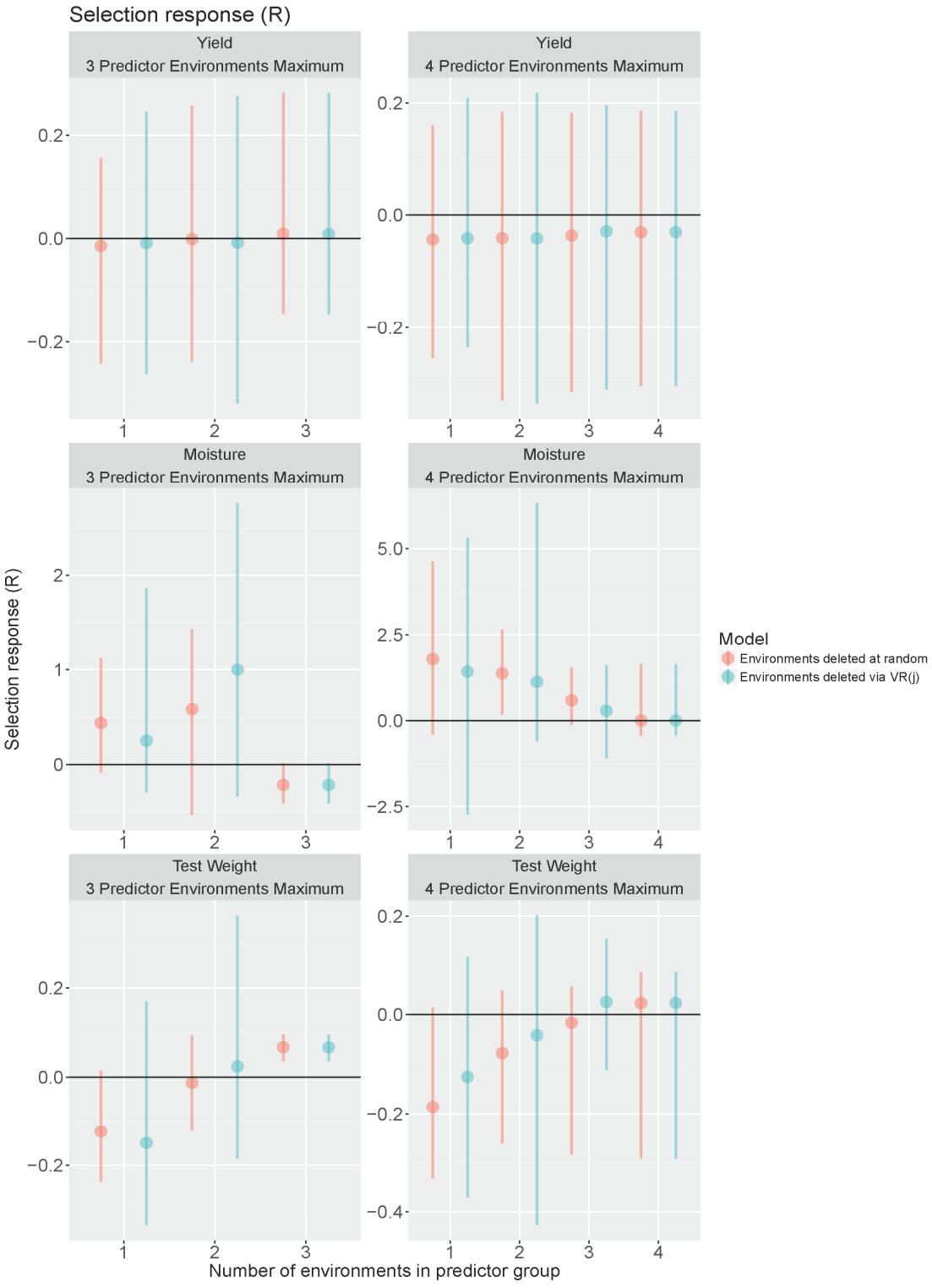


Figure 3: Mean selection response (R) for A/B test populations with environments deleted (out of three or four) on the basis of $V_{R(j)}$ or at random. Bars indicate the range of values.



BIBLIOGRAPHY

Akdemir D, Godfrey OU (2015) EMMREML: Fitting mixed models with known covariance structures.

Allard RW, Bradshaw AD (1964) Implications of genotype-environmental interactions in applied plant breeding. Crop Sci 4:503–508. doi: 10.2135/cropsci1964.0011183X000400050021x

Ariyo O, Ayo-Vaughan M (2000) Analysis of genotype× environment interaction of okra (Abelmoschus esculentus (L) Moench). J Genet Breed 54:35–40.

Asfaw A, Alemayehu F, Gurum F, Atnaf M (2009) AMMI and SREG GGE biplot analysis for matching varieties onto soybean production environments in Ethiopia. Sci Res Essays 4:1322–1330.

Baker RJ (1988) Tests for crossover genotype-environmental interactions. Can J Plant Sci 68:405–410. doi: 10.4141/cjps88-051

Bänziger M, Edmeades GO, Beck D, Bellon M (2000) Breeding for drought and nitrogen stress tolerance in maize: from theory to practice. Mex DF CIMMYT 68. doi: 633.1553

Basso B, Cammarano D, Carfagna E (2013) Review of crop yield forecasting methods and early warning systems. Proceedings of the First Meeting of the Scientific Advisory Committee of the Global Strategy to Improve Agricultural and Rural Statistics. FAO Headquarters, Rome, Italy,

Bates D, Maechler M, Bolker B, Walker S (2015) Package lme4. J Stat Softw 67:1–91. doi: http://lme4.r-forge.r-project.org

Bennetzen JL, Hake S (2009) Handbook of maize: genetics and genomics.

Bernardo R (1992) Weighted vs. unweighted mean performance of varieties across environments. Crop Sci 32:490. doi: 10.2135/cropsci1992.0011183X003200020041x

Beyene Y, Semagn K, Mugo S, et al (2015) Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. Crop Sci 55:154–163. doi: 10.2135/cropsci2014.07.0460

Bogyo TP (1964) Coefficients of variation of heritability estimates obtained from variance analyses. Biometrics 20:122. doi: 10.2307/2527621

Boichard D, Guillaume F, Baur A, et al (2012) Genomic selection in French dairy cattle. Anim. Prod. Sci. 52:115–120.

Brisson N, Gary C, Justes E, et al (2003) An overview of the crop model STICS. European Journal of Agronomy. pp 309–332

Brown D, Bootsma A (1993) Crop heat units for corn and other warm season crops in ontario. Factsheet Minist Agric food Ontario 1–6.

Buck-Sorlin G (2013) Process-based model. Encyclopedia of Systems Biology. Springer New York, New York, NY, pp 1755–1755

Buckler ES, Holland JB, Bradbury PJ, et al (2009) The genetic architecture of maize flowering time. Science (80-) 325:714–718. doi: 10.1126/science.1174276

Cambardella CA, Moorman TB, Parkin TB, et al (1994) Field-scale variability of soil properties in central Iowa soils. Soil Sci Soc Am J 58:1501. doi: 10.2136/sssaj1994.03615995005800050033x

Campbell BT, Baenziger PS, Eskridge KM, et al (2004) Using environmental covariates to explain genotype x environment and QTL x environment interactions for agronomic traits on chromosome 3A of wheat. Crop Sci 44:620–627. doi: 10.2135/cropsci2004.6200

Cao W, White JW, Wang E, International Symposium on Crop Modeling and Decision Support (2008 : Nanjing JS (2009) Crop modeling and decision support. Tsinghua University Press

Casler MD (1982) Genotype × environment interaction bias to parent-offspring regression heritability estimates. Crop Sci 22:540. doi: 10.2135/cropsci1982.0011183X002200030024x

Cattivelli L, Rizza F, Badeck FW, et al (2008) Drought tolerance improvement in crop plants: An integrated view from breeding to genomics. F. Crop. Res. 105:1–14.

Choi YS, Gim HJ, Ho CH, et al (2017) Climatic influence on corn sowing date in the Midwestern United States. Int J Climatol 37:1595–1602. doi: 10.1002/joc.4799

Coster A (2012) Pedigree: pedigree functions.

Crossa J, Fox PN, Pfeiffer WH, et al (1991) AMMI adjustment for statistical analysis of an international wheat yield trial. Theor Appl Genet 81:27–37. doi: 10.1007/BF00226108

Crossa J, Yang R-C, Cornelius PL (2004) Studying crossover genotype × environment interaction using linear-bilinear models and mixed models. J Agric Biol Environ Stat 9:362–380. doi: 10.2307/1400487

Daetwyler HD, Villanueva B, Woolliams JA (2008) Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS One 3:e3395. doi: 10.1371/journal.pone.0003395

de Oliveira LA, da Silva CP, Nuvunga JJ, et al (2016) Bayesian GGE biplot models applied to maize multi-environments trials. Genet Mol Res. doi: 10.4238/gmr.15028612

de Wit CT (1965) Photosynthesis of leaf canopies. Agricultural Research Reports. pp 1–54

Dourado-Neto D, Teruel DA, Reichardt K, et al (1998) Principles of crop modeling and simulation: I. uses of mathematical models in agricultural science. Sci Agric 55:46–50. doi: 10.1590/S0103-90161998000500008

Eberhart SA, Russell WA (1966) Stability parameters for comparing varieties. Crop Sci 6:36–40. doi: 10.2135/cropsci1966.0011183X000600010011x

Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome J 4:250–255. doi: 10.3835/plantgenome2011.08.0024

Environmental Systems Research Institute (ESRI) (2016) ArcGIS Release 10.3.

Falconer DS, Mackay TFC, Frankham R, et al (1996) Introduction to quantitative genetics (4th edn). Trends Genet 12:280. doi: 10.1016/0168-9525(96)81458-2

Fan XM, Kang MS, Chen H, et al (2007) Yield stability of maize hybrids evaluated in multi-environment trials in Yunnan, China. Agron J 99:220–228. doi: 10.2134/agronj2006.0144

Fehr WR (1991) Principles of cultivar development. Volume 1: Theory and technique.

Finlay KW, Wilkinson GN (1963) The analysis of adaptation in a plant-breeding programme. Aust J Agric Res 14:742–754. doi: 10.1071/AR9630742

Gallais A, Coque M (2005) Genetic variation and selection for nitrogen use efficiency in maize: A synthesis. Maydica 50:531–547.

Gallais A, Hirel B (2004) An approach to the genetics of nitrogen use efficiency in maize. Journal of Experimental Botany. pp 295–306

Garrick DJ, Taylor JF, Fernando RL (2009) Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet Sel Evol 41:55. doi: 10.1186/1297-9686-41-55

Gauch HG. J (1992) Statistical analysis of regional yield trials: AMMI analysis of factorial designs.

Gauch HG, Piepho HP, Annicchiarico P (2008) Statistical analysis of yield trials by AMMI and GGE: further considerations. Crop Sci. 48:866–889.

Gauch HG, Zobel, R. W. (1996) AMMI analysis of yield trials. Genotype-by-environment interaction. pp 85–122

Gill JL, Jensen EL (1968) The probability of obtaining a negative estimate of heritability. Biometrics 24:517–526. doi: 10.2307/2528315

Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136:245–257. doi: 10.1007/s10709-008-9308-0

Goddard ME, Hayes BJ (2007) Genomic selection. J Anim Breed Genet 124:323–330. doi: 10.1111/j.1439-0388.2007.00702.x

Guo Z, Tucker DM, Wang D, et al (2013) Accuracy of across-environment henome-wide prediction in maize nested association mapping populations. G3; Genes|Genomes|Genetics 3:263–272. doi: 10.1534/g3.112.005066

Habier D, Fernando RL, Garrick DJ (2013) Genomic BLUP decoded: A look into the black box of genomic prediction. Genetics 194:597–607. doi: 10.1534/genetics.113.152207

Hallauer AR, Russel WA, Lamkey KR (1988) Corn breeding. Corn and Corn Improvement. pp 469–564

Hamblin J, Fisher HM, Ridings HI (1980) The choice of locality for plant breeding when selecting for high yield and general adaptation. Euphytica 29:161–168. doi: 10.1007/BF00037262

Hatfield JL, Wright-Morton L, Hall B (2017) Vulnerability of grain crops and croplands in the Midwest to climatic variability and adaptation strategies. Clim. Change 1–13.

Heffner EL, Sorrells MEM, Jannink J-LJL (2009) Genomic selection for crop improvement. Crop Sci 49:1–12. doi: 10.2135/cropsci2008.08.0512

Henderson CR (1976) A Simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32:69. doi: 10.2307/2529339

Heslot N, Akdemir D, Sorrells ME, Jannink J-L (2014) Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theor Appl Genet 127:463–80. doi: 10.1007/s00122-013-2231-5

Hirsch J (1990) A nemesis for heritability estimation. Behav Brain Sci 13:137–138. doi: 10.1017/S0140525X00077967

Holland JB, Nyquist WE, Cervantes-Martinez CT (2003) Estimating and interpreting heritability for plant breeding. Plant Breed Rev 22:9–112. doi: 10.1002/9780470650202.ch2

Honarvar M, Rostami M (2013) Accuracy of genomic prediction using RR-BLUP and Bayesian LASSO. Eur J Exp Biol 3:42–47.

Horner T, Frey K (1957) Methods for determining natural areas for oat varietal recommendations. Agron J 49:313–315.

Hupet F, Vanclooster M (2005) Micro-variability of hydrological processes at the maize row scale: Implications for soil water content measurements and evapotranspiration estimates. J Hydrol 303:247–270. doi: 10.1016/j.jhydrol.2004.07.017

Isidro J, Jannink JL, Akdemir D, et al (2015) Training set optimization under population structure in genomic selection. Theor Appl Genet. doi: 10.1007/s00122-014-2418-4

Jacobson A, Lian L, Zhong S, Bernardo R (2014) General combining ability model for genomewide selection in a biparental cross. Crop Sci 54:895–905. doi: 10.2135/cropsci2013.11.0774

Jacobson A, Lian L, Zhong S, Bernardo R (2015) Marker imputation before genomewide selection in biparental maize populations. Plant Genome 8:0. doi: 10.3835/plantgenome2014.10.0078

Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics 9:166–77. doi: 10.1093/bfgp/elq001

Jarquín D, Crossa J, Lacaze X, et al (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet 127:595–607. doi: 10.1007/s00122-013-2243-1

Jonas E, de Koning D-J (2013) Does genomic selection have a future in plant breeding? Trends Biotechnol 31:497–504. doi: 10.1016/j.tibtech.2013.06.003

Kang HM, Zaitlen NA, Wade CM, et al (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709–1723. doi: 10.1534/genetics.107.080101

Katz RW, Brown BG (1992) Extreme events in a changing climate: Variability is more important than averages. Clim Change 21:289–302. doi: 10.1007/BF00139728

Kim D, Kaluarachchi J (2015) Validating FAO AquaCrop using Landsat images and regional crop information. Agric Water Manag 149:143–155. doi: 10.1016/j.agwat.2014.10.013

Krchov L-M, Gordillo GA, Bernardo R (2015) Multienvironment validation of the effectiveness of phenotypic and genomewide selection within biparental maize populations. Crop Sci 55:1068–1075. doi: 10.2135/cropsci2014.09.0608

Kumar S, Chagne D, Bink M, et al (2012) Genomic selection for fruit quality traits in apple (Malus x domestica Borkh.). PLoS One. doi: 10.1371/journal.pone.0036674

Lamkey KR, Hallauer AR (1987) Heritability estimated from recurrent selection experiments in maize. Maydica 32:61–78.

Leach KA, Hejlek LG, Hearne LB, et al (2011) Primary root elongation rate and abscisic acid levels of maize in response to water stress. Crop Sci 51:157–172. doi: 10.2135/cropsci2009.12.0708

Li R, Zeng Y, Xu J, et al (2015) Genetic variation for maize root architecture in response to drought stress at the seedling stage. Breed Sci 65:298–307. doi: 10.1270/jsbbs.65.298

Lian L, Jacobson A, Zhong S, Bernardo R (2014) Genomewide prediction accuracy within 969 maize biparental populations. Crop Sci 54:1514. doi: 10.2135/cropsci2013.12.0856

Lin CS, Binns MR, Lefkovitch LP (1986) Stability analysis: where do we stand? Crop Sci 26:894–900. doi: 10.2135/cropsci1986.0011183X002600050012x

Liu HL, Yang JY, Drury CF, et al (2011) Using the DSSAT-CERES-Maize model to simulate crop yield and nitrogen cycling in fields under long-term continuous maize production. Nutr Cycl Agroecosystems 89:313–328. doi: 10.1007/s10705-010-9396-y

Lobell DB, Burke MB (2010) On the use of statistical models to predict crop yield responses to climate change. Agric For Meteorol 150:1443–1452. doi: 10.1016/j.agrformet.2010.07.008

Löffler CM, Wei J, Fast T, et al (2005) Classification of maize environments using crop simulation and geographic information systems. Crop Sci 45:1708–1716. doi: 10.2135/cropsci2004.0370

Lopez-Cruz M, Crossa J, Bonnett D, et al (2015) Increased prediction accuracy in wheat breeding trials using a marker x environment interaction genomic selection model. G3; Genes|Genomes|Genetics 5:569–82. doi: 10.1534/g3.114.016097

Lu PX, Huber DA, White TL (1999) Potential biases of incomplete linear models in heritability estimation and breeding value prediction. Can J For Res 29:724–736. doi: 10.1139/x99-047

Luan T, Woolliams JA, Lien S, et al (2009) The accuracy of genomic selection in Norwegian red cattle assessed by cross-validation. Genetics 183:1119–26. doi: 10.1534/genetics.109.107391

Makowski D, Asseng S, Ewert F, et al (2015) A statistical analysis of three ensembles of crop model responses to temperature and CO2 concentration. Agric For Meteorol 214–215:483–493. doi: 10.1016/j.agrformet.2015.09.013

Malosetti M, Ribaut JM, van Eeuwijk FA (2013) The statistical analysis of multi-environment data: Modeling genotype-by-environment interaction and its genetic basis. Front Physiol 4:44–92. doi: 10.3389/fphys.2013.00044

Maredia MK, Ward R, Byerlee D (1996) Econometric estimation of a global spillover matrix for wheat varietal technology. Agric Econ 14:159–173. doi: 10.1016/0169-5150(96)01188-7

Mastrodomenico A, Hendrix C, Below F (2018) Nitrogen use efficiency and the genetic variation of maize expired plant variety protection germplasm. Agriculture 8:3. doi: 10.3390/agriculture8010003

McCoy J, Johnston K (2002) Using ArcGIS spatial analyst. Anal 238.

McMaster GS, Wilhelm WW (1997) Growing degree-days: One equation, two interpretations. Agric For Meteorol 87:291–300. doi: 10.1016/S0168-1923(97)00027-0

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829. doi: 11290733

Mevik BH, Wehrens R (2007) The pls package: Principal component and partial least squares regression in R. J Stat Softw 18:1–23. doi: 10.1159/000323281

Moehring J, Williams ER, Piepho HP (2014) Efficiency of augmented p‐rep designs in multi‐environmental trials. Theor Appl Genet 127:1049–1060. doi: 10.1007/s00122-014-2278-y

Moll RH, Kamprath EJ, Jackson WA (1982) Analysis and interpretation of factors which contribute to efficiency of nitrogen utilization. Agron J 74:562. doi: 10.2134/agronj1982.00021962007400030037x

Muir W, Nyquist WE, Xu S (1992) Alternative partitioning of the genotype-by-environment interaction. Theor Appl Genet 84:193–200. doi: 10.1007/BF00224000

National Renewable Energy Laboratory (2007) National solar radiation database 1991 – 2005 update : user's manual. Task No PVA76102 472. doi: 10.2172/901864

Nyquist WE, Baker RJ (1991) Estimation of heritability and prediction of selection response in plant populations. CRC Crit Rev Plant Sci 10:235–322. doi: 10.1080/07352689109382313

Parthasaranthi T, Velu G, Jeyakumar P (2013) Impact of crop heat units on growth and developmental physiology of future crop production: a review. Res Rev J Crop Sci Technol 2:11–18.

Penman HL (1956) Estimating evaporation. Eos, Trans Am Geophys Union 37:43–50. doi: 10.1029/TR037i001p00043

Perkins JM, Jinks JL (1968) Environmental and genotype-environmental components of variability. 3. Multiple lines and crosses. Heredity (Edinb) 23:339–356. doi: 10.1038/hdy.1969.11

Piepho HP, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209–228.

Pirttioja N, Carter T, Fronzek S, et al (2015) Temperature and precipitation effects on wheat yield across a European transect: a crop model ensemble analysis using impact response surfaces. Clim Res 65:87–105. doi: 10.3354/cr01322

Prasad PVV, Staggenborg S a., Ristic Z (2008) Impacts of drought and/or heat stress on physiological, developmental, growth, and yield processes of crop plants. Adv Agric Syst Model Ser 1 301–356. doi: 10.2134/advagricsystmodel1.c11

R Foundation for Statistical Computing, Vienna AI 3-900051-07-0 (2011) R Development Core Team. R A Lang Environ Stat Comput 55:275–286.

Rahman H, Durreshawar N, Ali S, et al (2010) Stability analysis of maize hybrids across north west of Pakistan. Pakistan J Bot 42:1083–1091.

Riedelsheimer C, Melchinger AE (2013) Optimizing the allocation of resources for genomic selection in one breeding cycle. TAG 126:2835–48. doi: 10.1007/s00122-013-2175-9

Riha SJ, Wilks DS, Simoens P (1996) Impact of temperature and precipitation variability on crop

model predictions. Clim Change 32:293–311. doi: 10.1007/BF00142466

Ritchie JT, NeSmith DS (1991) Temperature and crop development. Model Plant Soil Syst 5–29. doi: 10.2134/agronmonogr31.c2

Rosenzweig C, Tubiello FN, Goldberg R, et al (2002) Increased crop damage in the US from excess precipitation under climate change. Glob. Environ. Chang. 12:197–202.

Rötter RP, Palosuo T, Kersebaum KC, et al (2012) Simulation of spring barley yield in different climatic zones of Northern and Central Europe: A comparison of nine crop models. F Crop Res 133:23–36. doi: 10.1016/j.fcr.2012.03.016

Sadras V, Calderini D (2015) Crop physiology: applications for genetic improvement and agronomy.

Samonte SOP, Wilson LT, McClung AM, Medley JC (2005) Targeting cultivars onto rice growing environments using AMM and SREG GGE biplot analyses. Crop Sci 45:2414–2424. doi: 10.2135/cropsci2004.0627

Schaalje GB, Lynch DR, Kozub GC (1987) Field evaluation of a modified augmented design for early stage selection involving a large number of test lines without replication. Potato Res 30:35–45. doi: 10.1007/BF02357682

Shukla GK (1972) Some statistical aspects of partitioning genotype-environmental components of variability. Heredity (Edinb) 29:237–245. doi: 10.1038/hdy.1972.87

Singh M, Ceccarelli S, Grando S (1999) Genotype x environment interaction of crossover type: Detecting its presence and estimating the crossover point. Theor Appl Genet 99:988–995. doi: 10.1007/s001220051406

Solberg TR, Sonesson AK, Woolliams JA, Meuwissen THE (2008) Genomic selection using different marker types and densities. J Anim Sci 86:2447–2454. doi: 10.2527/jas.2007-0010

Vadez V (2014) Root hydraulics: The forgotten side of roots in drought adaptation. F Crop Res 165:15–24. doi: 10.1016/j.fcr.2014.03.017

Venables WN, Ripley BD (2002) MASS: modern applied statistics with S. Issues of Accuracy and Scale 868. doi: 10.1198/tech.2003.s33

Waqar-Ul-Haq M, Malik F, Rashid M, et al (2008) Heritability and genetic advance for grain yield and its component characters in maize (Zea Mays L.). Int J Plant Res 2:138–145. doi: 10.5923/j.plant.20120205.01

Weyhrich RA, Lamkey KR, Hallauer AR (1998) Responses to seven methods of recurrent selection in the BS11 maize population. Crop Sci 38:308–321. doi: 10.2135/cropsci1998.0011183X003800020005x

Wilson DR, Muchow RC, Murgatroyd CJ (1995) Model analysis of temperature and solar radiation limitations to maize potential productivity in a cool climate. F Crop Res 43:1–18. doi: 10.1016/0378-4290(95)00037-Q

Woltereck R (1909) Weitere experimentelle untersuchungen uber artveranderung, speziel uber das wesen quantitativer artunterschiede bei daphniden. Verhandlungen der Dtsch Zool Gesellschaft 19:110–173.

Würschum T, Gowda M, Zhao Y, et al (2012) Impact of selective genotyping in the training population on accuracy and bias of genomic selection. Theor Appl Genet 125:707–713. doi:

10.1007/s00122-012-1862-2

Xinyou Y, Laar HH van. (2005) Crop systems dynamics : an ecophysiological simulation model for genotype-by-environment interactions. Wageningen Academic Publishers

Yan W, Cornelius PL, Crossa J, Hunt LA (2001) Two types of GGE biplots for analyzing multi-environment trial data. Crop Sci 41:656–663. doi: 10.2135/cropsci2001.413656x

Yan W, Hunt LA, Sheng Q, Szlavnics Z (2000) Cultivar evaluation and mega-environment investigation based on the GGE biplot. Crop Sci 40:597. doi: 10.2135/cropsci2000.403597x

Yan W, Kang MS, Ma B, et al (2007) GGE biplot vs. AMMI analysis of genotype-by-environment data. Crop Sci 47:643–655. doi: 10.2135/cropsci2006.06.0374

Yan W, Tinker NA (2006) Biplot analysis of multi-environment trial data: Principles and applications. Can J Plant Sci 86:623–645. doi: 10.4141/P05-169

Yang RC, Crossa J, Cornelius PL, Burgueño J (2009) Biplot analysis of genotype × environment interaction: proceed with caution. Crop Sci 49:1564–1576. doi: 10.2135/cropsci2008.11.0665

Yates F, Cochran WG (1938) The analysis of groups of experiments. J Agric Sci 28:556–580. doi: 10.1017/S0021859600050978

Zhang X, Pérez-Rodríguez P, Semagn K, et al (2015) Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. Heredity (Edinb) 114:291–299. doi: 10.1038/hdy.2014.99

Zhao Y, Gowda M, Liu W, et al (2012) Accuracy of genomic selection in European maize elite breeding populations. Theor Appl Genet 124:769–776. doi: 10.1007/s00122-011-1745-y

Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. Nat Genet 44:821–824. doi: 10.1038/ng.2310

APPENDICES

Supplemental Table 4: Yield prediction ability (r_{MP}) and response to selection (R) in heritability maximized versus non-maximized environments

GCA Parent Population	Number of total training populations	Average number of environments in full data set	Average number of environments in heritability maximized populations	Average number of environments removed per population	Heritability using all environmental data	Heritability using heritability maximized environments	Improvement in heritability	Yield r_{MP} using all environment data	Yield r_{MP} using heritability maximized environments	Difference in r_{MP} between heritability maximized and non-maximized environments	Yield R using all environmental data	Heritability maximized yield R	Difference in Yield R between heritability maximized and non-maximized environments
P3/P4	17	7.41	6.06	1.35	0.45	0.50	0.04	0.22	0.22	0.01	0.13	0.12	-0.01
P3/P8	9	7.11	5.67	1.44	0.43	0.47	0.04	0.15	0.15	0.00	0.27	0.16	-0.10
P6/P7	14	8.14	5.93	2.21	0.42	0.46	0.05	0.25	0.25	0.00	0.29	0.38	0.09
P5/P8	10	7.50	5.60	1.90	0.46	0.50	0.04	-0.05	-0.06	0.00	-0.07	0.14	0.21*
P16/P13	6	7.33	5.83	1.50	0.44	0.48	0.05	0.30	0.28	-0.02	0.13	0.11	-0.02
P4/P5	18	7.83	6.22	1.61	0.46	0.50	0.04	0.26	0.28	0.02	0.30	0.26	-0.04
P17/P18	4	6.25	4.75	1.50	0.47	0.51	0.05	0.30	0.28	-0.02	0.36	0.15	-0.21
P9/P10	12	5.67	4.92	0.75	0.51	0.52	0.00	0.32	0.32	0.00	0.44	0.43	0.00
P19/P20	3	5.67	4.33	1.33	0.22	0.33	0.11	0.02	0.06	0.03	0.16	0.20	0.04
P11/P12	7	7.29	5.86	1.43	0.50	0.54	0.04	0.27	0.27	0.00	0.39	0.38	-0.01
P1/P9	11	5.91	4.73	1.18	0.50	0.50	0.00	0.36	0.36	-0.01	0.37	0.35	-0.01
P2/P15	8	7.63	6.38	1.25	0.48	0.52	0.05	0.29	0.29	0.00	0.28	0.27	-0.01
P1/P2	25	6.88	5.36	1.52	0.47	0.51	0.04	0.31	0.31	0.00	0.42	0.37	-0.05
P13/P14	7	7.57	5.86	1.71	0.47	0.51	0.04	0.20	0.27	0.06*	0.02	0.15	0.13
P31/P32	8	6.88	4.88	2.00	0.24	0.34	0.10	0.12	0.13	0.02	0.06	0.04	-0.01
P35/P36	4	7.50	5.50	2.00	0.17	0.39	0.22	0.03	-0.02	-0.05*	0.12	0.07	-0.05
P29/P27	10	7.30	5.00	2.30	0.38	0.45	0.07	0.16	0.17	0.01	0.28	0.31	0.03
P23/P24	16	7.50	4.94	2.56	0.25	0.36	0.12	-0.01	-0.02	-0.01	0.05	-0.09	-0.14
P21/P22	28	7.46	5.64	1.82	0.44	0.50	0.05	0.20	0.16	-0.04	0.28	0.07	-0.21*
P41/P42	3	9.33	6.00	3.33	0.44	0.55	0.11	0.21	0.27	0.05*	0.15	0.32	0.18*
P26/P27	16	7.44	4.94	2.50	0.33	0.44	0.11	0.23	0.21	-0.02	0.08	0.09	0.01
P24/P26	17	7.47	5.00	2.47	0.31	0.42	0.11	0.16	0.16	0.00	0.17	0.11	-0.06
P28/P27	10	7.20	4.80	2.40	0.34	0.44	0.11	0.18	0.16	-0.01	0.15	0.11	-0.03

P25/P22	19	7.68	6.05	1.63	0.49	0.52	0.03	0.27	0.27	0.00	0.34	0.32	-0.02
P33/P34	7	7.29	6.57	0.71	0.48	0.49	0.01	0.18	0.18	0.00	0.15	0.20	0.06
P37/P38	5	5.00	4.40	0.60	0.51	0.52	0.01	0.05	0.05	0.00	-0.06	-0.16	-0.10
P23/P25	16	7.56	5.38	2.19	0.45	0.50	0.06	0.36	0.35	-0.01	0.37	0.26	-0.11

* Indicates that the difference is significantly different at a $p<.05$ level, all other differences were non-significant

Supplemental Table 2: Moisture prediction ability (r_{MP}) and Response to selection (R) in heritability maximized versus non-maximized environments

GCA Parent Population	Number of total training populations	Average number of environments in full data set	Average number of environments in heritability maximized populations	Average number of environments removed per population	Heritability using all environmental data	Heritability using heritability maximized environments	Improvement in heritability	Moisture r_{MP} using all environment data	Moisture r_{MP} using heritability maximized environments	Difference in r_{MP} between heritability maximized and non-maximized environments	Moisture R using all environmental data	Heritability maximized moisture R	Difference in Moisture R between heritability maximized and non-maximized environments
P3/P4	17	7.41	6.06	1.35	0.75	0.77	0.02	0.49	0.49	0	-9.99	-9.99	0.00
P3/P8	9	7.11	5.56	1.56	0.76	0.78	0.03	0.50	0.51	0.01	-5.52	-5.44	0.08
P6/P7	14	8.14	6.93	1.21	0.75	0.77	0.02	0.50	0.48	-0.02	-7.39	-7.81	-0.42
P5/P8	10	7.50	6.30	1.20	0.76	0.78	0.02	0.38	0.38	-0.01	-10.33	-6.03	4.30*
P16/P13	6	7.33	6.17	1.17	0.66	0.68	0.02	0.60	0.58	-0.02	-7.99	-7.56	0.43
P4/P5	18	7.83	6.78	1.06	0.75	0.76	0.01	0.54	0.54	0	-5.34	-5.50	-0.16
P17/P18	4	6.25	5.50	0.75	0.56	0.60	0.04	0.57	0.60	0.03	-10.10	-13.45	-3.35*
P9/P10	12	5.67	4.33	1.33	0.68	0.72	0.03	0.29	0.29	0	-2.04	-2.30	-0.26
P19/P20	3	5.67	4.00	1.67	0.77	0.79	0.03	0.36	0.34	-0.02	-4.51	-4.47	0.04
P11/P12	7	7.29	5.57	1.71	0.60	0.64	0.04	0.62	0.62	0	-7.76	-7.09	0.67
P1/P9	11	5.91	5.00	0.91	0.52	0.58	0.06	0.47	0.48	0.01	-4.07	-3.86	0.21
P2/P15	8	7.63	6.50	1.13	0.72	0.74	0.02	0.51	0.53	0.02	-4.08	-4.22	-0.14
P1/P2	25	6.88	5.76	1.12	0.68	0.72	0.04	0.48	0.47	-0.01	-4.29	-4.29	0.00
P13/P14	7	7.57	6.43	1.14	0.65	0.67	0.02	0.38	0.40	0.01	-2.76	-3.26	-0.50
P31/P32	8	6.88	6.00	0.88	0.64	0.67	0.03	0.19	0.20	0.01	-7.07	-7.20	-0.13
P35/P36	4	7.50	5.75	1.75	0.61	0.65	0.04	-0.09	-0.09	0	0.48	0.40	-0.08
P29/P27	10	7.30	6.70	0.60	0.69	0.72	0.02	0.56	0.56	0	-4.50	-4.63	-0.13
P23/P24	16	7.50	6.38	1.13	0.63	0.66	0.03	0.28	0.29	0.01	-1.89	-2.66	-0.77
P21/P22	28	7.46	5.57	1.89	0.61	0.66	0.05	0.36	0.36	0	-5.65	-5.46	0.19
P41/P42	3	9.33	7.67	1.67	0.80	0.81	0.01	0.55	0.55	0	-11.15	-10.22	0.93
P26/P27	16	7.44	6.19	1.25	0.68	0.71	0.03	0.45	0.46	0.01	-7.08	-6.91	0.17
P24/P26	17	7.47	6.06	1.41	0.65	0.69	0.04	0.42	0.38	-0.03	-7.12	-5.97	1.15
P28/P27	10	7.20	5.90	1.30	0.70	0.72	0.02	0.52	0.54	0.01	-11.46	-11.39	0.07
P25/P22	19	7.68	5.95	1.74	0.62	0.66	0.04	0.39	0.41	0.02	-6.94	-7.74	-0.80
P33/P34	7	7.29	6.57	0.71	0.65	0.68	0.04	0.11	0.09	-0.02	-3.25	-3.30	-0.05
P37/P38	5	5.00	4.20	0.80	0.58	0.59	0.01	0.33	0.33	0	-8.41	-7.05	1.36
P23/P25	16	7.56	6.25	1.31	0.64	0.69	0.05	0.51	0.51	0	-7.24	-7.81	-0.57

* Indicates that the difference is significantly different at a $p<.05$ level, all other differences were non-significant

Supplemental Table 3: Test weight prediction ability (r_{MP}) and response to selection (R) in heritability maximized versus non-maximized environments

GCA Parent Population	Number of total training populations	Average number of environments in full data set	Average number of environments in heritability maximized populations	Average number of environments removed per population	Heritability using all environmental data	Heritability using heritability maximized environments	Improvement in heritability	Test weight r_{MP} using all environment data	Test weight r_{MP} using heritability maximized environments	Difference in r_{MP} between heritability maximized and non-maximized environments	Test weight R using all environmental data	Heritability maximized test weight R	Difference in Test weight R between heritability maximized and non-maximized environments
P3/P4	17	7.41	5.24	2.18	0.54	0.60	0.05	0.36	0.40	0.03	0.56	0.72	0.17*
P3/P8	9	7.11	4.78	2.33	0.54	0.60	0.06	0.46	0.46	0	0.47	0.49	0.02
P6/P7	14	8.14	5.36	2.79	0.67	0.76	0.09	0.46	0.46	0	0.68	0.59	-0.09
P5/P8	10	7.50	4.70	2.80	0.61	0.68	0.07	0.40	0.41	0.01	0.40	0.62	0.22*
P16/P13	6	7.33	5.17	2.17	0.61	0.66	0.05	0.44	0.45	0.01	0.70	0.73	0.03
P4/P5	18	7.83	5.50	2.33	0.59	0.64	0.05	0.31	0.30	-0.01	0.40	0.39	-0.01
P17/P18	4	6.25	4.00	2.25	0.45	0.45	0.01	0.33	0.36	0.03	0.42	0.56	0.14*
P9/P10	12	5.67	4.25	1.42	0.59	0.65	0.05	0.28	0.30	0.02	0.49	0.35	-0.15*
P19/P20	3	5.67	4.33	1.33	0.58	0.55	-0.03	0.28	0.30	0.02	0.38	0.41	0.04
P11/P12	7	7.29	4.57	2.71	0.63	0.69	0.06	0.41	0.35	-0.06	0.44	0.51	0.07
P1/P9	11	5.91	4.36	1.55	0.55	0.61	0.07	0.27	0.29	0.02	0.35	0.28	-0.08
P2/P15	8	7.63	5.50	2.13	0.58	0.64	0.06	0.30	0.27	-0.02	0.40	0.37	-0.03
P1/P2	25	6.88	4.56	2.32	0.59	0.63	0.05	0.32	0.30	-0.02	0.32	0.26	-0.06
P13/P14	7	7.57	6.14	1.43	0.66	0.70	0.04	0.34	0.33	-0.01	0.32	0.27	-0.04
P31/P32	8	6.88	4.75	2.13	0.51	0.57	0.06	0.33	0.37	0.03	0.73	0.75	0.02
P35/P36	4	7.50	4.50	3.00	0.44	0.58	0.13	-0.03	-0.04	-0.01	-0.02	0.08	0.1
P29/P27	10	7.30	5.10	2.20	0.51	0.60	0.09	0.36	0.32	-0.03	0.44	0.44	0
P23/P24	16	7.50	5.38	2.13	0.51	0.57	0.07	0.18	0.27	0.08*	0.31	0.32	0.02
P21/P22	28	7.46	5.39	2.07	0.64	0.70	0.06	0.35	0.35	0	0.45	0.40	-0.04
P41/P42	3	9.33	6.00	3.33	0.48	0.58	0.10	0.37	0.40	0.03	0.49	0.67	0.17*
P26/P27	16	7.44	5.44	2.00	0.58	0.66	0.08	0.52	0.52	0	0.74	0.71	-0.02
P24/P26	17	7.47	5.47	2.00	0.54	0.63	0.09	0.47	0.49	0.02	0.67	0.57	-0.1
P28/P27	10	7.20	5.40	1.80	0.57	0.64	0.07	0.31	0.31	0	0.44	0.50	0.07
P25/P22	19	7.68	5.79	1.89	0.66	0.70	0.05	0.37	0.36	-0.01	0.65	0.75	0.1
P33/P34	7	7.29	5.86	1.43	0.64	0.69	0.05	0.24	0.26	0.01	0.88	0.93	0.05
P37/P38	5	5.00	3.60	1.40	0.61	0.67	0.06	0.48	0.50	0.02	0.58	0.57	-0.02
P23/P25	16	7.56	5.13	2.44	0.59	0.67	0.08	0.34	0.35	0.01	0.69	0.69	0

* Indicates that the difference is significantly different at a $p<.05$ level, all other differences were non-significant

Supplemental Table 4: Yield prediction ability (r_{MP}) using stability variance over a randomized deletion method. Prediction ability and selection response of the stability variance informed model had the average randomized performance subtracted, showing improvement of the informed model over the non-informed model

	Number of Deleted Environments (Yield)									
Population	1	2	3	4	5	6	7	8	9	10
P3/P4	-0.03	-0.01	0.03	0.02	0.03	0.02	0.03	-0.13		
P3/P8	0.01	0.04	0.07	0.14	0.09	0.04	0.11			
P6/P7	0.03	0 (ns)	0.06	0.06	0.1	0.16	0.18	0.15	0.21	
P5/P8	0	-0.01	-0.02	-0.02	-0.01 (ns)	0.01	0.03			
P16/P13	0.01	0.03	0.02	0.06	0.09	0.01 (ns)	0.05			
P4/P5	0.01	0 (ns)	0.01	0.04	0.07	0.08	0.15	0 (ns)		
P17/P18	-0.02	-0.06	-0.06	0.03	0.08	0.06	-0.01 (ns)			
P9/P10	0	-0.01	0.03	0.07	0.1	0.04				
P19/P20	-0.1	-0.02	-0.03	0.03	0 (ns)					
P11/P12	-0.01	-0.04	-0.03	0.03	0.09	0.14	0.15			
P1/P9	-0.01	0.01	0.02	0.05	0.06	0.16				
P2/P15	-0.03	-0.02	-0.03	-0.02	-0.03	0.04	0.16	0.04		
P1/P2	0.02	0.02	0.03	0 (ns)	0.03	-0.01	0.02			
P13/P14	0.02	0.07	0.11	0.09	0.11	0.16	0.26			
P31/P32	-0.01	0.05	0.07	0.09	0.1	0.11	0.09			
P35/P36	0 (ns)	-0.02	-0.03	-0.02	0.03	0.03	0.05			
P29/P27	0.03	0 (ns)	0.05	-0.03	-0.03	0.06	0.1	0.09		
P23/P24	0	0.02	0	-0.01	0 (ns)	-0.03	0.07	-0.13		
P21/P22	0	0.01	0.02	0.01	0.04	0.1	0.02	0.12	0.03	0.05
P41/P42	0.07	0.09	0.07	0.09	0.09	0.1	0.07	0.05	-0.02	-0.02
P26/P27	0.01	0 (ns)	-0.01	0 (ns)	0.01	-0.03	0.01 (ns)			
P24/P26	0.02	0.05	0.06	0.06	0.08	0.08	-0.01 (ns)			
P28/P27	0.01	0.01	0.05	0.06	0.03	0.01 (ns)	-0.09			
P25/P22	0 (ns)	0.02	0	-0.01	-0.05	0.02	0.05	-0.05	-0.07	-0.14
P33/P34	0.02	-0.01	0.01	0 (ns)	0.05	0.04	0.05			
P37/P38	-0.01	0.04	0.11	0.11	0.09					

P23/P25	-0.01	-0.01	-0.04	-0.03	0 (ns)	0.06	0.14			

(ns) indicates that the entry is not significant at a $p<.05$ level, all others were significant

Supplemental Table 5: Yield selection response (R) using stability variance over a randomized deletion method. Prediction ability and selection response of the stability variance informed model had the average randomized performance subtracted, showing improvement of the informed model over the non-informed

	Number of Deleted Environments (Yield)									
Population	1	2	3	4	5	6	7	8	9	10
P3/P4	-0.01	-0.06	-0.03	-0.05	0.05	-0.05	0 (ns)	-0.04		
P3/P8	0.08	-0.05	0.09	0.24	0.15	0.07	0.12			
P6/P7	0.05	-0.03	0.05	0.15	0.06	0.13	0.13	0.05	0.11	
P5/P8	-0.03	0.01 (ns)	0.02	0.05	0.01 (ns)	0.03	-0.1			
P16/P13	-0.05	0.02	-0.03	0.04	0.05	0.08	-0.02 (ns)			
P4/P5	0.02	0.08	0.04	0.04	-0.01 (ns)	0.1	0.03	-0.07		
P17/P18	-0.02	-0.03	-0.1	0 (ns)	0.01 (ns)	0.19	0.07			
P9/P10	-0.05	-0.04	0 (ns)	-0.06	0.06	0.01 (ns)				
P19/P20	-0.11	0.05	0.07	0.12	0.02					
P11/P12	-0.06	-0.07	-0.02	-0.05	0.13	0.13	0.07			
P1/P9	0.08	-0.02	0.06	0 (ns)	-0.03	-0.1				
P2/P15	-0.01	0.01	-0.02	-0.03	-0.07	-0.03	0.24	-0.11		
P1/P2	0.05	0.07	0.08	0.12	0.03	-0.11	0.02			
P13/P14	0.04	0.12	0.25	0.14	0.26	0.19	0.23			
P31/P32	-0.01	0.1	0.18	0.15	0.12	0.11	0.28			
P35/P36	0.05	0.03	0.02	0 (ns)	0.13	0.04	0.12			
P29/P27	0.05	0.02	-0.06	-0.14	-0.16	0.16	0.27	0.32		
P23/P24	-0.06	-0.06	-0.09	-0.01	0.08	-0.04	-0.05	-0.14		
P21/P22	-0.02	-0.09	-0.03	0.12	0.07	0.02	0.07	0.03	0.03	0.04
P41/P42	0.03	0.21	0.19	0.14	0.13	0.29	0.17	0.08	0.16	0.08
P26/P27	0.01	-0.01	-0.06	0.09	0.07	-0.08	-0.01 (ns)			
P24/P26	-0.1	0.07	0.05	0.06	-0.05	0.09	0.04			
P28/P27	-0.02	0.04	0.03	-0.02	0.01	-0.08	-0.05			
P25/P22	0.09	0.1	0.03	-0.06	-0.05	0 (ns)	0.02 (ns)	0.06	0.11	-0.05
P33/P34	0.04	0.01 (ns)	0.02	0.09	0.13	0.06	0.04			
P37/P38	-0.06	-0.01 (ns)	0.02 (ns)	0.19	0.07					

P23/P25	-0.01	-0.05	-0.04	-0.12	-0.1	0 (ns)	0.11			

(ns) indicates that the entry is not significant at a $p<.05$ level, all others were significant

Supplemental Table 6: Moisture prediction ability (r_{MP}) using stability variance over a randomized deletion method. Prediction ability and selection response of the stability variance informed model had the average randomized performance subtracted, showing improvement of the informed model over the non-informed model

	Number of Deleted Environments (Moisture)									
Population	1	2	3	4	5	6	7	8	9	10
P3/P4	-0.01	0	0	0.04	0.01	0.05	0.08	0.2		
P3/P8	-0.01	0.02	-0.01	-0.02	0.12	0.07	0.19			
P6/P7	0	-0.01	0.01	0.02	0.04	0.09	0.04	0.09	0.21	
P5/P8	0.01	0.03	0.01	0.06	0.05	-0.13	0 (ns)			
P16/P13	0.04	0.01	0 (ns)	0.09	-0.02	0.17	0.06			
P4/P5	0	0.01	0.05	0.05	0.09	0.14	0.08	0.09		
P17/P18	0.01	0.08	-0.04	0.14	0.2	0.14	0.23			
P9/P10	-0.02	-0.01	-0.01	0.13	0.01 (ns)	-0.24				
P19/P20	-0.01	0 (ns)	0 (ns)	-0.02	0.06					
P11/P12	-0.01	0 (ns)	0.03	0.06	0.16	0.09	0.28			
P1/P9	0.01	0.01	0.02	0.02	0.1	0.17				
P2/P15	0.02	0.05	0.04	0.06	0.02	0.03	0.2	0.03		
P1/P2	0	0.01	0.02	-0.01	0.02	0.11	0.14			
P13/P14	0	0.05	0.04	0.04	0.07	-0.03	0.17			
P31/P32	-0.01	0.02	-0.01	-0.01	0.02	-0.02	0.05			
P35/P36	0.01	0.01	0.02	0.02	0.03	0.06	0.07			
P29/P27	0.03	0.03	0.03	0.04	0.12	0.04	0.07	0.08		
P23/P24	0.02	0.07	0.12	0.13	0.09	0.09	0.14	-0.02		
P21/P22	-0.01	0.04	0.04	0.01	0.05	0.12	0.1	0.12	0.06	0.1
P41/P42	0.01	0.01	0.01	0.01	0.06	0.12	0 (ns)	-0.01	0.08	0.05
P26/P27	0.04	0.07	0.05	0.09	0.08	0.04	0.07			
P24/P26	0	-0.02	-0.01	-0.02	0 (ns)	0.01 (ns)	0.06			
P28/P27	0.02	0.03	0.05	0.02	0.05	0.02	0.14			
P25/P22	0	0.01	0.03	0.04	0.08	0.06	0.07	0.07	0.1	0.05
P33/P34	-0.01	-0.03	0 (ns)	0.01	-0.01	0.03	0.05			
P37/P38	0.03	0.05	0.08	0.16	0.19					

P23/P25	0.01	0.02	0.01	0.03	0.06	0.04	0.07			

(ns) indicates that the entry is not significant at a $p<.05$ level, all others were significant

Supplemental Table 7: Moisture selection response (R) using stability variance over a randomized deletion method. Prediction ability and selection response of the stability variance informed model had the average randomized performance subtracted, showing improvement of the informed model over the non-informed model

	Number of Deleted Environments (Moisture)									
Population	1	2	3	4	5	6	7	8	9	10
P3/P4	-0.32	-1.31	-0.96	0.01 (ns)	0.15 (ns)	-1.81	-3.42	-0.68		
P3/P8	-0.13	-1.19	-0.76	-0.98	-0.95	0.6	-2.54			
P6/P7	-0.48	0.74	-0.76	0.5	0.55	-0.25	-1.56	-1.38	-1.97	
P5/P8	-2.37	-1.24	-0.31 (ns)	-1.35	0.72	5.49	-2.19			
P16/P13	-0.21	0.64	0.74	-0.44	-0.18	-1.12	-1.37			
P4/P5	-0.38	0.03 (ns)	0.01 (ns)	-0.64	-0.31	-1.16	-1.21	-1.32		
P17/P18	-0.32	-2.36	1.79	-2.61	-6.03	-1.83	-6.79			
P9/P10	0.53	-0.37	0.64	-0.48	0.74	1.4				
P19/P20	-0.46	-0.19	-0.98	-1.17	-2.92					
P11/P12	0.2	-0.08	0.31	-0.53	-1.56	-3.67	-4.51			
P1/P9	0.25	0.1	-0.46	-0.33	-1.3	0.34				
P2/P15	-0.07	-0.36	-0.45	-0.81	-1.58	-0.43	-2.26	-2.27		
P1/P2	-0.17	0.27	0.71	-0.18	-0.72	-0.82	0 (ns)			
P13/P14	0.41	-0.12	-0.28	0.09	-0.9	-0.34	-0.48			
P31/P32	-0.36	-1.35	-1.08	2.05	1.4	-3.38	-5.03			
P35/P36	0.01 (ns)	0.64	0.13	-0.75	-0.41	-1.02	-0.09 (ns)			
P29/P27	0.09	-0.2	-0.35	-0.64	-1.21	-1.11	-1.39	-0.82		
P23/P24	-0.39	-0.85	-0.35	-0.89	-1.18	0.26	-1.2	0.9		
P21/P22	-0.14	-0.99	-0.8	-0.45	-2.25	-2.02	-1.72	-1.42	-1.27	-1.97
P41/P42	-1.45	-0.33	-1.55	0.5	-0.74	-1.74	-2.01	0.43	-2.22	-1.37
P26/P27	-0.38	-0.28	-1.18	0.4	-0.84	-0.25	-2.07			
P24/P26	-0.62	-0.71	-0.17	-0.79	-0.35	-0.54	-2.39			
P28/P27	0.64	-0.4	-0.81	-1.27	-2.52	-2.27	-0.95			
P25/P22	-1.63	-2.12	-0.88	-0.74	-1.99	-1.51	-3.52	-1.68	-1.63	-1.37
P33/P34	0.22	-0.16 (ns)	-0.96	-0.48	-0.37	-0.95	-3.53			
P37/P38	-0.52	-2.61	-2.92	-2.36	-5.52					

P23/P25	-0.26	-0.34	0.13	-1.28	-1.25	-2.53	-2.79			

(ns) indicates that the entry is not significant at a $p<.05$ level, all others were significant

Supplemental Table 8: Test weight prediction ability (r_{MP}) using stability variance over a randomized deletion method. Prediction ability and selection response of the stability variance informed model had the average randomized performance subtracted, showing improvement of the informed model over the non-informed model

	Number of Deleted Environments (Test weight)									
Population	1	2	3	4	5	6	7	8	9	10
P3/P4	0.03	0.03	0.04	0.05	0.06	0.04	0.09	0.08		
P3/P8	-0.02	0.02	0.03	0.1	0.12	0.09	0.09			
P6/P7	0.01	0.01	0.02	0.03	0.05	0.13	0.11	0.12	0.13	
P5/P8	0	0.01	0.03	0.03	0.05	0.12	0.03			
P16/P13	-0.01	-0.01	-0.03	0.05	0.09	0.05	0.16			
P4/P5	0 (ns)	0	0.01	-0.01	0 (ns)	0.01	0.05	0.06		
P17/P18	0 (ns)	0.01 (ns)	0.05	0.02	-0.08	0.19	0.26			
P9/P10	0	0.02	0.04	0	-0.01	-0.06				
P19/P20	0.01	-0.06	0.02	0.03	0.04 (ns)					
P11/P12	0	0.02	0.03	0.06	-0.02	-0.02	0.13			
P1/P9	0	0.02	0.06	0 (ns)	0.03	0.11				
P2/P15	0.01	0.01	0.01	-0.02	0	0.06	0.22	0.02		
P1/P2	0.01	0	0.01	-0.02	0.05	0.09	0.07			
P13/P14	0.02	0	0.02	0.06	-0.02	0.07	0.01 (ns)			
P31/P32	0.01	0.03	0.02	-0.01 (ns)	0.02	0.08	-0.02			
P35/P36	0	0.01	0.03	0.04	0.04	0.05	0.03			
P29/P27	0.02	0.03	0.04	0.13	0.13	0 (ns)	0.21	0.11		
P23/P24	0.03	0.03	-0.05	-0.05	0.03	0.04	0.16	-0.01 (ns)		
P21/P22	0.01	0	0	0 (ns)	-0.04	-0.04	0.04	0.07	0.02	0.16
P41/P42	0.01	0.03	0.02	-0.04	0.08	0.17	0.22	0.17	0.06	0.12
P26/P27	0 (ns)	0.02	0.01	0.06	0.05	0.09	0.04			
P24/P26	-0.01	0	0.01	0.05	0.07	0.02	0.16			
P28/P27	0.01	0.02	0.02	0.01	0 (ns)	0.03	0.11			
P25/P22	-0.01	-0.01	0	0.01	0.01	0.02	0.09	0.02	0.06	0.08
P33/P34	0 (ns)	0	0.02	-0.01	0.01	0.06	0.02			
P37/P38	0.04	0.04	0.06	-0.02	0.15 (ns)					

P23/P25	0.01	-0.01	0.02	0.01	0.04	0.04	0.16			

(ns) indicates that the entry is not significant at a $p<.05$ level, all others were significant

Supplemental Table 9: Test weight selection response (R) using stability variance over a randomized deletion method. Prediction ability and selection response of the stability variance informed model had the average randomized performance subtracted, showing improvement of the informed model over the non-informed model

	Number of Deleted Environments (Test weight)									
Population	1	2	3	4	5	6	7	8	9	10
P3/P4	0.02	0.1	0.06	0.02	0.09	0.07	0.24	-0.27		
P3/P8	0.02	0.03	0.07	0.16	0.09	0.13	-0.05			
P6/P7	0.01	0.03	0.08	0.11	0.17	0.24	0.17	0.16	0.31	
P5/P8	-0.07	0.06	0.11	0.07	0.25	0.21	0.31			
P16/P13	0.01	0.03	0.03	0.17	0.37	0.18	0.53			
P4/P5	0.03	0.07	-0.09	-0.02	0.05	0.07	0.07	0.18		
P17/P18	-0.01 (ns)	0 (ns)	0.12	-0.02 (ns)	-0.42	0.36	0.52			
P9/P10	0.01	0.05	-0.04	0.03	0.05	0.12				
P19/P20	0.01 (ns)	-0.09	0.21	-0.13	0.03					
P11/P12	0.03	0.01	0.09	0.08	0.12	0.04	0.16			
P1/P9	-0.02	-0.01 (ns)	0.08	-0.05	0.05	0.04				
P2/P15	0.1	0.11	0.05	-0.11	0.16	0.2	0.3	0.07		
P1/P2	-0.03	-0.02	0 (ns)	0.01 (ns)	0.1	0.08	0.02 (ns)			
P13/P14	0.01	0.02	0.04	0.14	0.01 (ns)	-0.02	0.06			
P31/P32	0.03	-0.13	0.11	0.09	-0.05	0.11	0.22			
P35/P36	0.08	-0.01	0.05	-0.01 (ns)	0.09	0.07	0.17			
P29/P27	0.05	-0.01 (ns)	0.05	0.14	0.19	0.04	0.28	0.11		
P23/P24	0.06	-0.06	-0.03	0.02	-0.02	0.07	0.27	0.1		
P21/P22	-0.15	-0.03	0 (ns)	0.06	-0.14	-0.03	-0.15	0.16	0.14	0.2
P41/P42	-0.01 (ns)	0.01	-0.06	0.11	0.24	0.27	0.38	0.31	0.17	0.24
P26/P27	-0.04	0.05	0.02	0.12	0.07	0.04	0.25			
P24/P26	0.03	0.11	-0.07	0.01 (ns)	0.06	0.19	0.2			
P28/P27	0.09	0.14	0.17	0.2	0.13	0.19	0.18			
P25/P22	-0.02	0.08	0.2	-0.15	0.11	0.19	0.36	0.21	0.24	0.25
P33/P34	0.08	0.14	0.21	0.34	0.36	0.26	0.16			
P37/P38	0.01	0.19	0.12	0.28	0.37					

P23/P25	0.12	0.09	0.11	0.06	0.1	-0.03	0.18			

(ns) indicates that the entry is not significant at a p<.05 level , all others were significant

Supplemental Table 10: Number of phenotypic selection environments a GCA prediction is equivalent to on an prediction ability basis

Populations	Moisture environments replaced	Test weight environments replaced	Yield environments replaced	Average number of Environments Replaced
P3/P4	2	4	4	3.33
P3/P8	2	2	0	1.33
P6/P7	1.5	1	3	1.83
P5/P8	0	4	0	1.33
P16/P13	3	1.5	3	2.50
P4/P5	1	3	1.5	1.83
P17/P18	2	3	3	2.67
P9/P10	1	2	3	2.00
P19/P20	2	1	2	1.67
P11/P12	1	4	4	3.00
P1/P9	2	1.5	2	1.83
P2/P15	2	2	2	2.00
P1/P2	3	3	1	2.33
P13/P14	1.5	2	0	1.17
P31/P32	0	1	1	0.67
P35/P36	0	0	0	0.00
P29/P27	2	2	3	2.33
P23/P24	0	0	0	0.00
P21/P22	4	2	1.5	2.50
P41/P42	1.5	0	2	1.17
P26/P27	1	0	3	1.33
P24/P26	1	1.5	0	0.83
P28/P27	1	3	1	1.67
P25/P22	2	2.5	3	2.50
P33/P34	0	0	0	0.00
P37/P38	0	2	0	0.67
P23/P25	1.5	1.5	1	1.33
Average number of environments replaced	1.41	1.83	1.63	

Supplemental Table 11: Index model yield selection response (R). Phenotyping R with the maximum number of predictor environments subtracted was subtracted from index R, showing improvement of the index model over the maximum phenotypic evaluation

		Number of Predictor Environments					
Population	Number of Validation Locations	1	2	3	4	5	6
P31/P32	2	0.08	0.07				
P2/P15	3	0.01	-0.03	-0.05			
P4/P5	3	-0.17*	-0.05	-0.06*			
P9/P10	3	0.06	0.06*	0.05			
P19/P20	3	0.07*	0.02	-0.02			
P1/P9	2	0.25*	0.27*	0.28*			
P37/P38	2	-0.26*	-0.32*	-0.15*			
P5/P8	4	-0.23*	-0.3*	-0.31*	-0.27*		
P1/P2	3	0.01	-0.04*	0	-0.03		
P13/P14	4	-0.24*	-0.33*	-0.31*	-0.3*		
P3/P4	4	-0.02	-0.12*	-0.13*	-0.12*		
P3/P8	3	0.21*	0.21*	0.2*	0.19*		
P16/P13	3	-0.07*	-0.04	-0.04*	-0.09*		
P17/P18	4	-0.02	-0.02	-0.03*	-0.04*		
P11/P12	4	0.12*	0.19*	0.18*	0.18*		
P25/P22	4	-0.04*	0.1*	0.12*	0.12*		
P33/P34	4	-0.19*	-0.2*	-0.17*	-0.13*		
P23/P25	3	-0.01	-0.03	-0.04	-0.03		
P35/P36	4	-0.1*	-0.17*	-0.15*	-0.16*		
P29/P27	3	0.02	0.02	0.13*	0.13*		
P23/P24	3	0	0.22*	0.15*	0.17*		
P21/P22	4	-0.06*	-0.04*	-0.02	-0.05*		
P41/P42	3	-0.1*	-0.18*	-0.16*	-0.14*		
P26/P27	4	-0.06*	0.06*	0.06*	0.07*		
P24/P26	3	0	-0.09*	0	-0.06*		
P28/P27	4	0.01	-0.01	-0.04*	-0.02		
P6/P7	6	-0.01*	0.09*	0.12*	0.11*	0.11*	0.1*

* indicates that the score is significantly different from 0

Supplemental Table 12: Index model yield prediction ability (r_{MP}). Phenotyping r_{MP} with the maximum number of predictor environments subtracted was subtracted from index r_{MP}, showing improvement of the index model over the maximum phenotypic evaluation

		Number of Predictor Environments					
Population	Number of Validation Locations	1	2	3	4	5	6
P31/P32	2	0.1*	0.08*				
P2/P15	3	0.07*	0.11*	0.11*			
P4/P5	3	-0.03	0.01	0.03*			
P9/P10	3	0.05*	0.08*	0.07*			
P19/P20	3	0.01	-0.06*	-0.11*			
P1/P9	2	0.03*	0.06*	0.06*			
P37/P38	2	-0.15*	-0.21*	-0.19*			
P5/P8	4	-0.17*	-0.27*	-0.26*	-0.24*		
P1/P2	3	-0.06*	-0.07*	-0.05*	-0.01		
P13/P14	4	-0.06*	-0.1*	-0.08*	-0.07*		
P3/P4	4	0.01	-0.01	0.01	0		
P3/P8	3	-0.08*	-0.08*	-0.08*	-0.05*		
P16/P13	3	0.08*	0.08*	0.08*	0.07*		
P17/P18	4	-0.06*	-0.02*	-0.02*	0		
P11/P12	4	0.02*	0.02*	0.02*	0.02*		
P25/P22	4	-0.02*	0.04*	0.04*	0.03*		
P33/P34	4	-0.05*	-0.08*	-0.07*	-0.07*		
P23/P25	3	0	0	0.01	0		
P35/P36	4	-0.07*	-0.16*	-0.17*	-0.17*		
P29/P27	3	0.09*	0.06*	0.06*	0.06*		
P23/P24	3	-0.01	0	-0.01	0		
P21/P22	4	-0.01	-0.02*	0.01	0		
P41/P42	3	-0.07*	-0.07*	-0.06*	-0.05*		
P26/P27	4	-0.04*	0.06*	0.08*	0.08*		
P24/P26	3	0	-0.05*	-0.03	-0.03		
P28/P27	4	-0.03*	0	0.01	-0.01		
P6/P7	6	0.01*	-0.01*	-0.02*	-0.03*	-0.03*	-0.03

* indicates that the score is significantly different from 0

Supplemental Table 13: Index model moisture selection response (R). Phenotyping R with the maximum number of predictor environments subtracted was subtracted from index R, showing improvement of the index model over the maximum phenotypic evaluation

		Number of Predictor Environments					
Population	Number of Validation Locations	1	2	3	4	5	6
P31/P32	2	0.33	0				
P2/P15	3	0.22	0.63	-0.12			
P4/P5	3	-0.23	2.76*	-0.26			
P9/P10	3	0.04	0.33	0.02			
P19/P20	3	-0.1	0.65	-0.11			
P1/P9	2	-0.28	-0.33	-0.41			
P37/P38	2	1.86*	1.94*	-0.39			
P5/P8	4	5.32*	2.16*	0.15	-0.1		
P1/P2	3	1.17*	2.2*	1.04	-0.01		
P13/P14	4	-0.35*	-0.59*	-0.6*	-0.07		
P3/P4	4	4.47*	1.43*	0.98*	-0.03		
P3/P8	3	0.94*	0.6*	0.28	-0.09		
P16/P13	3	1.75*	0.76*	-0.01	-0.07		
P17/P18	4	-2.74*	0.75	1.61*	-0.14		
P11/P12	4	1.08*	1.47*	0.64*	0		
P25/P22	4	2.52*	0.78*	0.81*	-0.08		
P33/P34	4	1.48*	2.09*	1.37*	-0.11		
P23/P25	3	1.34*	-0.1	-0.44	-0.43		
P35/P36	4	0.7*	-0.42*	-0.22*	-0.04		
P29/P27	3	0.44*	0.12	-0.6*	-0.11		
P23/P24	3	0.16	0.66	-0.04	0.02		
P21/P22	4	0.55*	-0.41*	-1.08*	1.64*		
P41/P42	3	1.79*	6.32*	0.32	-0.13		
P26/P27	4	0.84*	1.17*	0.17	-0.06		
P24/P26	3	1.11*	0.51*	0.12	-0.07		
P28/P27	4	4.42*	1.93*	1*	0.06		
P6/P7	6	4.87*	3.44*	3.38*	2.12*	0.45*	-0.04

* indicates that the score is significantly different from 0

Supplemental Table 14: Index model moisture prediction ability (r_{MP}). Phenotyping r_{MP} with the maximum number of predictor environments subtracted was subtracted from index r_{MP}, showing improvement of the index model over the maximum phenotypic evaluation

		Number of Predictor Environments					
Population	Number of Validation Locations	1	2	3	4	5	6
P31/P32	2	-0.01	0				
P2/P15	3	0.09*	0.02	0.02			
P4/P5	3	-0.01	-0.26*	0.01			
P9/P10	3	0.05*	0.01	0			
P19/P20	3	-0.09*	-0.07*	0			
P1/P9	2	-0.06*	0	0.01			
P37/P38	2	-0.14*	-0.07*	0			
P5/P8	4	-0.16*	-0.1*	-0.04*	0		
P1/P2	3	-0.12*	-0.19*	-0.12*	0.01		
P13/P14	4	-0.08*	-0.03*	0	0.01		
P3/P4	4	-0.13*	-0.1*	-0.05*	0		
P3/P8	3	0	-0.02*	0	0.01		
P16/P13	3	-0.13*	-0.06*	0.01	0		
P17/P18	4	0.06*	-0.04	-0.04*	0		
P11/P12	4	-0.13*	-0.09*	-0.02*	0		
P25/P22	4	-0.17*	-0.08*	-0.03*	0		
P33/P34	4	-0.06*	-0.03*	0.01	0		
P23/P25	3	-0.14*	-0.01	-0.01	0.01		
P35/P36	4	-0.06*	-0.07*	-0.04*	0		
P29/P27	3	-0.07*	-0.03*	0	0		
P23/P24	3	-0.06*	-0.1*	-0.04*	0		
P21/P22	4	-0.06*	0.03*	0.07*	-0.13*		
P41/P42	3	-0.15*	-0.38*	-0.05*	0		
P26/P27	4	-0.17*	-0.09*	-0.03*	0		
P24/P26	3	-0.16*	-0.07*	-0.03*	0		
P28/P27	4	-0.1*	-0.06*	-0.03*	0		
P6/P7	6	-0.14*	-0.11*	-0.1*	-0.05*	-0.02*	0

* indicates that the score is significantly different from 0

Supplemental Table 15: Index model test weight selection response (R). Phenotyping R with the maximum number of predictor environments subtracted was subtracted from index R, showing improvement of the index model over the maximum phenotypic evaluation

		Number of Predictor Environments					
Population	Number of Validation Locations	1	2	3	4	5	6
P31/P32	1	0.04	0.01				
P2/P15	2	-0.12	-0.02	0.05			
P4/P5	3	-0.26*	0.04	0.1*			
P9/P10	3	-0.15*	-0.18*	0.04			
P19/P20	2	0.17*	0.36*	0.07			
P1/P9	2	-0.34*	0.01	0.06			
P37/P38	2	0.04	0				
P5/P8	4	-0.09*	-0.19*	0.02	0.07		
P1/P2	3	-0.17*	-0.03	0.04	0.02		
P13/P14	4	0	0.05*	0.03*	0.04*		
P3/P4	4	-0.23*	-0.24*	-0.07*	0.02		
P3/P8	4	-0.34*	-0.01	0.08*	0.03*		
P16/P13	4	-0.08*	-0.05	0.05*	0.04*		
P17/P18	3	-0.07*	0.06	0.03	0.04		
P11/P12	4	0.11*	0.1*	0.09*	0.08*		
P25/P22	4	-0.02	0.12*	0.09*	0.06*		
P33/P34	4	-0.13*	0.02	0.04	0.05*		
P23/P25	3	-0.01	0.2*	0.15*	0.09*		
P35/P36	3	-0.13*	-0.04	-0.11*	-0.29*		
P29/P27	3	-0.14*	-0.03	0.02	0.01		
P23/P24	3	-0.25*	-0.06	-0.08*	0.03		
P21/P22	4	-0.09*	0.07*	0.09*	0.02		
P41/P42	3	-0.37*	-0.32*	-0.04	0		
P26/P27	3	-0.21*	-0.06	0.09*			
P24/P26	3	-0.35*	-0.43*	-0.09*	0.03		
P28/P27	4	0.12*	0.04	0.1*	0.08*		
P6/P7	6	-0.31*	-0.09*	-0.02*	0.02*	0.02*	0.02*

* indicates that the score is significantly different from 0

Supplemental Table 16: Index model test weight prediction ability (r_{MP}). Phenotyping r_{MP} with the maximum number of predictor environments subtracted was subtracted from index r_{MP}, showing improvement of the index model over the maximum phenotypic evaluation

		Number of Predictor Environments					
Population	Number of Validation Locations	1	2	3	4	5	6
P31/P32	1	0.02*	0.03				
P2/P15	2	-0.03	0.07*	0.03			
P4/P5	3	-0.02	0.02	0.08*			
P9/P10	3	-0.15*	-0.12*	0.02			
P19/P20	2	0.04	0.07*	0.01			
P1/P9	2	-0.14*	0.02	0.04			
P37/P38	2	-0.02	0.05*				
P5/P8	4	-0.01	-0.03*	0.03*	0.07*		
P1/P2	3	-0.09*	0.06*	0.08*	0.07*		
P13/P14	4	-0.01	-0.02*	0.01	0.03*		
P3/P4	4	-0.06*	-0.04*	0.01	0.04*		
P3/P8	4	-0.12*	-0.02*	0.03*	0.03*		
P16/P13	4	-0.06*	-0.07*	0.01	0.03*		
P17/P18	3	-0.04	-0.03*	0.01	0.02*		
P11/P12	4	0.07*	0.13*	0.12*	0.09*		
P25/P22	4	-0.08*	0.01	0.03*	0.03*		
P33/P34	4	-0.06*	0.02*	0.03*	0.01		
P23/P25	3	-0.02	0.06*	0.04*	0.04*		
P35/P36	3	0.01	-0.01	-0.01	-0.08*		
P29/P27	3	-0.1*	-0.03*	0.01	0.03*		
P23/P24	3	-0.07*	-0.02	-0.01	0.01		
P21/P22	4	-0.06*	-0.01	0.02*	0.02*		
P41/P42	3	-0.35*	-0.08*	0	0.02		
P26/P27	3	-0.09*	-0.07*	0.05*			
P24/P26	3	-0.15*	-0.2*	-0.02	0.02*		
P28/P27	4	0.04*	-0.02	0.01	0.04*		
P6/P7	6	-0.11*	-0.06*	-0.01*	0.02*	0.02*	0.01

* indicates that the score is significantly different from 0

Supplemental Table 17: Null index model yield selection response (R). Phenotyping R with the maximum number of predictor environments subtracted was subtracted from index R, showing improvement of the index model over the maximum phenotypic evaluation

		Number of Predictor Environments					
Population	Number of Validation Locations	1	2	3	4	5	6
P31/P32	2	0.04	0.07				
P2/P15	3	0.07*	-0.01	-0.05			
P4/P5	3	-0.09*	-0.04*	-0.06*			
P9/P10	3	0.05*	0.04*	0.05			
P19/P20	3	-0.02	-0.01	-0.02			
P1/P9	2	0.16*	0.26*	0.28*			
P37/P38	2	-0.24*	-0.24*	-0.15*			
P5/P8	4	-0.22*	-0.28*	-0.29*	-0.27*		
P1/P2	3	0.01	-0.02*	-0.02	-0.03		
P13/P14	4	-0.25*	-0.33*	-0.31*	-0.3*		
P3/P4	4	-0.06*	-0.12*	-0.14*	-0.12*		
P3/P8	3	0.16*	0.14*	0.17*	0.19*		
P16/P13	3	-0.09*	-0.12*	-0.1*	-0.09*		
P17/P18	4	0.03*	-0.02*	-0.04*	-0.04*		
P11/P12	4	0.16*	0.18*	0.18*	0.18*		
P25/P22	4	0.02*	0.13*	0.11*	0.12*		
P33/P34	4	-0.18*	-0.18*	-0.16*	-0.13*		
P23/P25	3	-0.01	-0.01	-0.02	-0.03		
P35/P36	4	-0.07*	-0.14*	-0.15*	-0.16*		
P29/P27	3	-0.03*	0.1*	0.15*	0.13*		
P23/P24	3	0	0.18*	0.17*	0.17*		
P21/P22	4	-0.09*	-0.07*	-0.06*	-0.05*		
P41/P42	3	-0.07*	-0.16*	-0.15*	-0.14*		
P26/P27	4	0.01	0.07*	0.08*	0.07*		
P24/P26	3	-0.1*	-0.1*	-0.09*	-0.06*		
P28/P27	4	-0.03*	-0.03*	-0.03*	-0.02		

P6/P7	6	-0.03*	0.06*	0.08*	0.1*	0.1*	0.1*

* indicates that the score is significantly different from 0

Supplemental Table 18: Null index model yield prediction ability (r_{MP}). Phenotyping r_{MP} with the maximum number of predictor environments subtracted was subtracted from index r_{MP}, showing improvement of the index model over the maximum phenotypic evaluation

		Number of Predictor Environments					
Population	Number of Validation Locations	1	2	3	4	5	6
P31/P32	2	0.09*	0.08*				
P2/P15	3	0.07*	0.1*	0.11*			
P4/P5	3	-0.01*	0.02*	0.03*			
P9/P10	3	0.05*	0.07*	0.07*			
P19/P20	3	-0.06*	-0.12*	-0.11*			
P1/P9	2	0.04*	0.06*	0.06*			
P37/P38	2	-0.17*	-0.22*	-0.19*			
P5/P8	4	-0.16*	-0.25*	-0.25*	-0.24*		
P1/P2	3	-0.07*	-0.07*	-0.04*	-0.01*		
P13/P14	4	-0.08*	-0.11*	-0.09*	-0.07*		
P3/P4	4	-0.02*	-0.01*	0.01	0.01		
P3/P8	3	-0.09*	-0.09*	-0.07*	-0.05*		
P16/P13	3	0.07*	0.05*	0.06*	0.07*		
P17/P18	4	-0.02*	-0.02	-0.02	0		
P11/P12	4	0.02*	0.02*	0.02*	0.02*		
P25/P22	4	-0.03*	0.01*	0.02*	0.03*		
P33/P34	4	-0.11*	-0.09*	-0.08*	-0.07*		
P23/P25	3	-0.04*	-0.03*	-0.02*	0		
P35/P36	4	-0.12*	-0.17*	-0.17*	-0.17*		
P29/P27	3	0.01	0.05*	0.06*	0.06*		
P23/P24	3	0	-0.01	-0.01	0		
P21/P22	4	-0.08*	-0.05*	-0.02*	0		
P41/P42	3	-0.06*	-0.08*	-0.07*	-0.05*		
P26/P27	4	0.04*	0.07*	0.08*	0.08*		
P24/P26	3	-0.07*	-0.05*	-0.05*	-0.03*		
P28/P27	4	-0.04*	-0.02*	-0.02*	-0.01*		

P6/P7	6	-0.05*	-0.06*	-0.06*	-0.05*	-0.04*	-0.03

* indicates that the score is significantly different from 0

Supplemental Table 19: Null index model moisture selection response (R). Phenotyping R with the maximum number of predictor environments subtracted was subtracted from index R, showing improvement of the index model over the maximum phenotypic evaluation

		Number of Predictor Environments					
Population	Number of Validation Locations	1	2	3	4	5	6
P31/P32	2	0.25	0				
P2/P15	3	0.09	0.71	-0.12			
P4/P5	3	0.58	1.2*	-0.26			
P9/P10	3	0.01	0.11	0.02			
P19/P20	3	0.91*	0.59*	-0.11			
P1/P9	2	-0.08	-0.53	-0.41			
P37/P38	2	1.12*	1.43*	-0.39			
P5/P8	4	4.62*	2.65*	1.14*	-0.1		
P1/P2	3	1.58*	1.82*	0.45*	-0.01		
P13/P14	4	0.04	0.28*	0.07	-0.07		
P3/P4	4	2.74*	2.07*	0.88*	-0.03		
P3/P8	3	2.44*	1.51*	0.81*	-0.09		
P16/P13	3	2.22*	1.43*	0.74*	-0.07		
P17/P18	4	-0.39	1.8*	1.39*	-0.14		
P11/P12	4	1.28*	1.24*	0.51*	0		
P25/P22	4	2.92*	1.44*	0.58*	-0.08		
P33/P34	4	2.63*	2.22*	0.71*	-0.11		
P23/P25	3	1.41*	0.96*	-0.11	-0.43		
P35/P36	4	0.14	0.18*	0.01	-0.04		
P29/P27	3	0.8*	0.49*	-0.11	-0.11		
P23/P24	3	2.08*	1.34*	0.54*	0.02		
P21/P22	4	0.82*	0.85*	1.54*	1.64*		
P41/P42	3	0.71*	1.76*	0.54*	-0.13		
P26/P27	4	1.83*	1.35*	0.36*	-0.06		
P24/P26	3	1.86*	0.99*	0.36*	-0.07		
P28/P27	4	4.14*	1.61*	0.79*	0.06		

P6/P7	6	4.26*	3.54*	2.43*	1.41*	0.57*	-0.04

* indicates that the score is significantly different from 0

Supplemental Table 20: Null index model moisture prediction ability (r_{MP}). Phenotyping r_{MP} with the maximum number of predictor environments subtracted was subtracted from index r_{MP}, showing improvement of the index model over the maximum phenotypic evaluation

		Number of Predictor Environments					
Population	Number of Validation Locations	1	2	3	4	5	6
P31/P32	2	-0.05*	0				
P2/P15	3	0.07*	0.01	0.02			
P4/P5	3	-0.06	-0.14*	0.01			
P9/P10	3	0	-0.02	0			
P19/P20	3	-0.12*	-0.05*	0			
P1/P9	2	-0.04*	0	0.01			
P37/P38	2	-0.09*	-0.05*	0			
P5/P8	4	-0.14*	-0.09*	-0.04*	0		
P1/P2	3	-0.1*	-0.11*	-0.04*	0.01		
P13/P14	4	-0.08*	-0.06*	-0.02*	0.01		
P3/P4	4	-0.14*	-0.1*	-0.03*	0		
P3/P8	3	-0.11*	-0.07*	-0.03*	0.01		
P16/P13	3	-0.14*	-0.09*	-0.03*	0		
P17/P18	4	0.02*	-0.04*	-0.02*	0		
P11/P12	4	-0.11*	-0.08*	-0.03*	0		
P25/P22	4	-0.15*	-0.08*	-0.03*	0		
P33/P34	4	-0.1*	-0.06*	-0.03*	0		
P23/P25	3	-0.11*	-0.06*	-0.01	0.01		
P35/P36	4	-0.07*	-0.1*	-0.04*	0		
P29/P27	3	-0.1*	-0.09*	-0.03*	0		
P23/P24	3	-0.12*	-0.09*	-0.04*	0		
P21/P22	4	-0.08*	-0.08*	-0.14*	-0.13*		
P41/P42	3	-0.1*	-0.14*	-0.04*	0		
P26/P27	4	-0.14*	-0.09*	-0.04*	0		
P24/P26	3	-0.16*	-0.09*	-0.04*	0		
P28/P27	4	-0.14*	-0.1*	-0.04*	0		

P6/P7	6	-0.19*	-0.16*	-0.1*	-0.06*	-0.02*	0

* indicates that the score is significantly different from 0

Supplemental Table 21: Null index model test weight selection response (R). Phenotyping R with the maximum number of predictor environments subtracted was subtracted from index R, showing improvement of the index model over the maximum phenotypic evaluation

		Number of Predictor Environments					
Population	Number of Validation Locations	1	2	3	4	5	6
P31/P32	1	-0.09	0.01				
P2/P15	2	-0.07	-0.05	0.05			
P4/P5	3	-0.24*	0.03	0.1*			
P9/P10	3	-0.14*	-0.12*	0.04			
P19/P20	2	0.01	0.09	0.07			
P1/P9	2	-0.14*	-0.01	0.06			
P37/P38	2	0.06	0				
P5/P8	4	-0.28*	-0.21*	-0.03	0.07		
P1/P2	3	-0.16*	-0.04*	-0.04*	0.02		
P13/P14	4	0.01	0.05*	0.04*	0.04*		
P3/P4	4	-0.29*	-0.13*	-0.01	0.02		
P3/P8	4	-0.2*	-0.01	0.03*	0.03*		
P16/P13	4	-0.2*	-0.04*	0.03*	0.04*		
P17/P18	3	-0.12*	-0.05*	-0.01	0.04		
P11/P12	4	-0.05*	0.03*	0.05*	0.08*		
P25/P22	4	-0.08*	0	0.04*	0.06*		
P33/P34	4	-0.17*	0.03*	0.04*	0.05*		
P23/P25	3	-0.15*	-0.06*	0.06*	0.09*		
P35/P36	3	-0.32*	-0.26*	-0.28*	-0.29*		
P29/P27	3	-0.13*	-0.06*	-0.03*	0.01		
P23/P24	3	-0.33*	-0.19*	-0.08*	0.03		
P21/P22	4	-0.3*	-0.12*	-0.03*	0.02		
P41/P42	3	-0.22*	-0.15*	-0.07*	0		
P26/P27	3	-0.16*	-0.02	0.09*			
P24/P26	3	-0.33*	-0.15*	-0.06*	0.03		
P28/P27	4	-0.01	-0.02	0.05*	0.08*		

P6/P7	6	-0.25*	-0.25*	-0.17*	-0.11*	-0.04*	0.02*

* indicates that the score is significantly different from 0

Supplemental Table 22: Null Index model test weight Prediction Accuracy (r_{MP}). Phenotyping r_{MP} with the maximum number of predictor environments subtracted was subtracted from index r_{MP}, showing improvement of the index model over the maximum phenotypic evaluation

		Number of Predictor Environments					
Population	Number of Validation Locations	1	2	3	4	5	6
P31/P32	1	-0.01	0.03				
P2/P15	2	0	0.02	0.03			
P4/P5	3	-0.01	0.07*	0.08*			
P9/P10	3	-0.08*	-0.05*	0.02			
P19/P20	2	-0.01	-0.01	0.01			
P1/P9	2	-0.05*	0.02	0.04			
P37/P38	2	-0.03*	0.05*				
P5/P8	4	-0.11*	-0.06*	0.03*	0.07*		
P1/P2	3	-0.05*	0.03*	0.04*	0.07*		
P13/P14	4	-0.06*	-0.03*	0.01*	0.03*		
P3/P4	4	-0.09*	-0.03*	0.02*	0.04*		
P3/P8	4	-0.11*	-0.03*	0.01*	0.03*		
P16/P13	4	-0.15*	-0.06*	0	0.03*		
P17/P18	3	-0.06*	-0.06*	-0.01	0.02*		
P11/P12	4	0	0.07*	0.08*	0.09*		
P25/P22	4	-0.13*	-0.03*	0.01*	0.03*		
P33/P34	4	-0.11*	-0.05*	-0.02*	0.01		
P23/P25	3	-0.1*	-0.03*	0.01	0.04*		
P35/P36	3	-0.1*	-0.12*	-0.1*	-0.08*		
P29/P27	3	-0.12*	-0.03*	0	0.03*		
P23/P24	3	-0.11*	-0.06*	-0.02*	0.01		
P21/P22	4	-0.09*	-0.03*	-0.01*	0.02*		
P41/P42	3	-0.17*	-0.05*	-0.02*	0.02		
P26/P27	3	-0.07*	-0.02	0.05*			
P24/P26	3	-0.14*	-0.08*	-0.01	0.02*		
P28/P27	4	-0.03*	-0.02*	0.01*	0.04*		

P6/P7	6	-0.16*	-0.13*	-0.07*	-0.04*	-0.01*	0.01

* indicates that the score is significantly different from 0

Supplemental Table 23: Yield mean predictive ability (r_{MP}) across all populations after deleting different numbers of environments based on increased stability variance

	Number of Environments Deleted (Yield)										
Population	0	1	2	3	4	5	6	7	8	9	10
P5/P8	-0.05	-0.06	-0.06	-0.06	-0.07	-0.03	-0.03	-0.04			
P2/P15	0.29	0.26	0.26	0.25	0.25	0.21	0.21	0.30	0.16		
P1/P2	0.31	0.33	0.33	0.32	0.29	0.32	0.26	0.22			
P13/P14	0.20	0.22	0.26	0.29	0.25	0.29	0.31	0.37			
P3/P4	0.22	0.18	0.19	0.24	0.20	0.20	0.17	0.14	-0.08*		
P3/P8	0.15	0.14	0.14	0.21	0.19	0.13	0.07	0.14			
P16/P13	0.30	0.30	0.31	0.30	0.32	0.32	0.25	0.19			
P4/P5	0.26	0.26	0.25	0.26	0.26	0.28	0.28	0.25	0.06		
P17/P18	0.30	0.27	0.21	0.16	0.22	0.23	0.15	0.11			
P9/P10	0.32	0.32	0.30	0.32	0.30	0.27	0.16				
P19/P20	0.02	-0.07	-0.02	-0.01	0.04	0.02					
P11/P12	0.27	0.25	0.22	0.21	0.24	0.30	0.27	0.27			
P1/P9	0.36	0.35	0.37	0.37	0.38	0.37	0.32				
P31/P32	0.12	0.11	0.15	0.18	0.19	0.18	0.17	0.12			
P25/P22	0.27	0.27	0.29	0.26	0.24	0.19	0.20	0.21	-0.07*	-0.10*	-0.18*
P33/P34	0.18	0.18	0.16	0.14	0.13	0.15	0.11	0.09			
P23/P25	0.36	0.34	0.32	0.29	0.28	0.29	0.31	0.35			
P35/P36	0.03	0.02	0.00	-0.01	-0.02	0.03	0.03	0.04			
P29/P27	0.16	0.17	0.14	0.14	0.07	0.07	0.12	0.14	0.10		
P23/P24	-0.01	-0.01	0.01	-0.01	-0.02	0.00	-0.04	0.07	-0.11		
P21/P22	0.20	0.20	0.19	0.17	0.16	0.17	0.22	0.12	0.13	0.08	0.06
P41/P42	0.21	0.26	0.28	0.27	0.26	0.27	0.27	0.21	0.15	0.07	0.07
P26/P27	0.23	0.23	0.22	0.21	0.20	0.21	0.17	0.17			
P24/P26	0.16	0.18	0.21	0.21	0.21	0.21	0.17	0.09			
P28/P27	0.18	0.17	0.17	0.17	0.18	0.12	0.07	-0.01			
P37/P38	0.05	0.03	0.08	0.11	0.10	0.06					
P6/P7	0.25	0.27	0.25	0.30	0.28	0.31	0.33	0.31	0.28	0.23	

* indicates significant difference ($p<.05$) from 0 environments deleted

Supplemental Table 24: Yield selection response (R, using the top 10% of predicted lines) across all populations after deleting different numbers of environments based on increased stability variance

	Number of Environments Deleted (Yield)										
Population	0	1	2	3	4	5	6	7	8	9	10
P5/P8	-0.07	-0.09	-0.03	0.01	-0.01	-0.02	-0.03	-0.26			
P2/P15	0.28	0.23	0.27	0.24	0.24	0.15	0.10	0.35	0.00		
P1/P2	0.42	0.45	0.45	0.45	0.50	0.36	0.23	0.27			
P13/P14	0.02	0.04	0.09	0.24	0.10	0.32*	0.29*	0.37*			
P3/P4	0.13	0.09	0.06	0.10	0.10	0.17	0.08	0.10	-0.01		
P3/P8	0.27	0.26	0.07	0.19	0.35	0.15	0.08	0.14			
P16/P13	0.13	0.09	0.16	0.14	0.22	0.19	0.24	0.07			
P4/P5	0.30	0.31	0.33	0.29	0.22	0.18	0.25	0.12	-0.03*		
P17/P18	0.36	0.26	0.23	0.14	0.23	0.18	0.29	0.20			
P9/P10	0.44	0.37	0.36	0.40	0.29	0.35	0.20				
P19/P20	0.16	-0.02	0.11	0.12	0.16	-0.01					
P11/P12	0.39	0.29	0.26	0.26	0.20	0.38	0.28	0.23			
P1/P9	0.37	0.44	0.33	0.41	0.34	0.28	0.16				
P31/P32	0.06	0.05	0.17	0.28	0.22	0.21	0.17	0.30			
P25/P22	0.34	0.42	0.42	0.34	0.25	0.21	0.16	0.15	0.08	0.13	-0.04*
P33/P34	0.15	0.19	0.19	0.19	0.22	0.23	0.16	0.12			
P23/P25	0.37	0.29	0.26	0.25	0.16	0.16	0.22	0.29			
P35/P36	0.12	0.09	0.06	0.02	0.00	0.13	0.02	0.10			
P29/P27	0.28	0.22	0.16	0.11	-0.04*	-0.02	0.22	0.30	0.36		
P23/P24	0.05	-0.03	-0.04	-0.07	0.02	0.10	-0.07	0.01	-0.11		
P21/P22	0.28	0.16	0.09	0.10	0.29	0.24	0.20	0.22	0.06	0.16	0.10
P41/P42	0.15	0.19	0.39	0.37	0.33	0.28	0.44*	0.32	0.18	0.21	0.15
P26/P27	0.08	0.12	0.10	0.07	0.16	0.16	0.05	0.11			
P24/P26	0.17	0.06	0.23	0.22	0.25	0.15	0.25	0.19			
P28/P27	0.15	0.11	0.14	0.13	0.04	0.06	-0.02	0.01			
P37/P38	-0.06	-0.05	0.02	0.02	0.16	0.07					
P6/P7	0.29	0.35	0.28	0.32	0.42	0.29	0.36	0.35	0.24	0.15	

* indicates significant difference ($p<.05$) from 0 environments deleted

Supplemental Table 25: Moisture mean predictive ability (r_{MP}) across all populations after deleting different numbers of environments based on increased stability variance

	Number of Environments Deleted (Moisture)										
Population	0	1	2	3	4	5	6	7	8	9	10
P5/P8	0.38	0.37	0.39	0.38	0.38	0.30	0.09*	0.25			
P2/P15	0.51	0.52	0.54	0.53	0.53	0.48	0.47	0.46	0.34		
P1/P2	0.48	0.48	0.49	0.48	0.46	0.47	0.47	0.41			
P13/P14	0.38	0.38	0.41	0.40	0.40	0.39	0.34	0.35			
P3/P4	0.49	0.48	0.48	0.47	0.47	0.43	0.40	0.44	0.33		
P3/P8	0.50	0.49	0.50	0.48	0.45	0.47	0.43	0.45			
P16/P13	0.60	0.61	0.54	0.57	0.55	0.52	0.50	0.46			
P4/P5	0.54	0.53	0.52	0.53	0.55	0.54	0.52	0.44	0.26*		
P17/P18	0.57	0.57	0.60	0.46	0.61	0.57	0.50	0.43			
P9/P10	0.29	0.25	0.19	0.19	0.28	0.12	-0.18*				
P19/P20	0.36	0.34	0.30	0.30	0.20	0.29					
P11/P12	0.62	0.60	0.60	0.60	0.63	0.61	0.57	0.57			
P1/P9	0.47	0.47	0.47	0.46	0.47	0.46	0.30				
P31/P32	0.19	0.17	0.19	0.16	0.14	0.17	0.13	0.15			
P25/P22	0.39	0.38	0.38	0.40	0.41	0.39	0.37	0.35	0.18	0.19	0.12*
P33/P34	0.11	0.10	0.08	0.11	0.11	0.08	0.12	0.14			
P23/P25	0.51	0.51	0.50	0.49	0.49	0.49	0.49	0.47			
P35/P36	-0.09	-0.09	-0.08	-0.07	-0.07	-0.06	-0.02	-0.02			
P29/P27	0.56	0.56	0.57	0.57	0.56	0.50	0.48	0.41	0.36		
P23/P24	0.28	0.29	0.30	0.34	0.35	0.33	0.28	0.22	0.12		
P21/P22	0.36	0.35	0.37	0.34	0.32	0.30	0.32	0.31	0.26	0.24	0.28
P41/P42	0.55	0.54	0.55	0.54	0.51	0.52	0.50	0.47	0.29*	0.33*	0.29*
P26/P27	0.45	0.47	0.49	0.47	0.45	0.44	0.41	0.41			
P24/P26	0.42	0.42	0.39	0.40	0.39	0.38	0.37	0.34			
P28/P27	0.52	0.54	0.53	0.54	0.52	0.53	0.43	0.47			
P37/P38	0.33	0.33	0.37	0.36	0.36	0.34					
P6/P7	0.50	0.50	0.49	0.48	0.47	0.47	0.45	0.39	0.41	0.48	

* indicates significant difference ($p<.05$) from 0 environments deleted

Supplemental Table 26: Moisture selection response (R, using the top 10% of predicted lines) across all populations after deleting different numbers of environments based on increased stability variance

	Number of Environments Deleted (Moisture)										
Population	0	1	2	3	4	5	6	7	8	9	10
P5/P8	-10.3	-10.2	-8.3	-7.6	-7.8	-4.9*	0.8*	-6.0*			
P2/P15	-4.1	-4.1	-4.3	-4.4	-4.4	-4.3	-3.7	-4.1	-3.5		
P1/P2	-4.3	-4.1	-3.7	-3.4	-4.3	-4.9	-4.9	-3.0			
P13/P14	-2.8	-2.1	-2.6	-2.6	-2.3	-2.9	-2.5	-2.1			
P3/P4	-10.0	-9.5	-10.0	-10.0	-8.5	-8.5	-9.0	-9.7	-5.1*		
P3/P8	-5.5	-5.2	-6.2	-5.9	-5.2	-5.2	-4.2	-6.2			
P16/P13	-8.0	-8.2	-7.2	-7.2	-7.6	-7.7	-6.9	-6.4			
P4/P5	-5.3	-5.6	-5.2	-5.2	-5.8	-5.1	-5.0	-5.3	-2.7*		
P17/P18	-10.1	-10.4	-13.2	-9.1	-13.5	-13.3	-9.9	-10.2			
P9/P10	-2.0	-1.5	-1.9	-0.9	-2.0	-0.6	0.7*				
P19/P20	-4.5	-5.0	-4.7	-4.8	-3.3	-5.5					
P11/P12	-7.8	-7.1	-7.4	-6.4	-7.5	-8.0	-8.6	-8.2			
P1/P9	-4.1	-3.7	-3.8	-4.4	-4.2	-4.4	-2.2				
P31/P32	-7.1	-6.2	-7.0	-5.7	-2.7	-2.3	-6.2	-7.3			
P25/P22	-6.9	-8.1	-8.7	-7.7	-7.2	-7.4	-6.8	-6.7	-4.3	-3.6*	-3.6*
P33/P34	-3.2	-2.8	-2.8	-3.3	-3.3	-3.2	-3.5	-5.6			
P23/P25	-7.2	-7.7	-7.7	-7.3	-8.2	-8.2	-7.3	-8.2			
P35/P36	0.5	0.5	1.0	0.7	-0.2	0.2	-0.7	0.6			
P29/P27	-4.5	-4.4	-4.6	-4.6	-4.6	-4.4	-3.9	-3.5	-3.3		
P23/P24	-1.9	-2.9	-3.2	-2.9	-2.8	-2.9	-1.4	-2.6	-0.4		
P21/P22	-5.7	-6.2	-7.2	-5.9	-6.3	-6.0	-6.4	-5.3	-4.3	-4.3	-4.4
P41/P42	-11.1	-12.3	-11.3	-11.3	-9.7	-10.7	-9.1	-10.0	-5.4*	-7.7*	-7.0*
P26/P27	-7.1	-7.4	-7.3	-7.5	-6.3	-6.4	-6.1	-7.0			
P24/P26	-7.1	-7.0	-6.6	-6.4	-6.4	-5.5	-5.7	-5.7			
P28/P27	-11.5	-10.5	-11.5	-11.7	-11.1	-10.9	-8.1*	-8.1*			
P37/P38	-8.4	-7.7	-8.9	-8.5	-8.2	-8.2					
P6/P7	-7.4	-8.4	-6.9	-7.8	-6.6	-6.0	-5.7	-5.8	-6.1	-8.0	

* indicates significant difference ($p<.05$) from 0 environments deleted

Supplemental Table 27: Test weight mean predictive ability (r_{MP}) across all populations after deleting different numbers of environments based on increased stability variance

	Number of Environments Deleted (Test weight)										
Population	0	1	2	3	4	5	6	7	8	9	10
P5/P8	0.40	0.39	0.39	0.39	0.39	0.37	0.37	0.24			
P2/P15	0.30	0.28	0.29	0.25	0.23	0.24	0.25	0.32	0.11		
P1/P2	0.32	0.32	0.31	0.32	0.29	0.34	0.36	0.30			
P13/P14	0.34	0.35	0.33	0.31	0.36	0.28	0.31	0.23			
P3/P4	0.36	0.38	0.38	0.39	0.39	0.38	0.34	0.33	0.24*		
P3/P8	0.46	0.43	0.45	0.45	0.49	0.47	0.42	0.31			
P16/P13	0.44	0.42	0.42	0.37	0.41	0.44	0.43	0.47			
P4/P5	0.31	0.31	0.31	0.30	0.29	0.28	0.25	0.22	0.18		
P17/P18	0.33	0.31	0.28	0.27	0.21	0.04*	0.30	0.35			
P9/P10	0.28	0.28	0.29	0.29	0.28	0.27	0.18				
P19/P20	0.28	0.26	0.19	0.22	0.23	0.09					
P11/P12	0.41	0.41	0.41	0.40	0.40	0.31	0.28	0.29*			
P1/P9	0.27	0.27	0.28	0.30	0.24	0.23	0.21				
P31/P32	0.33	0.32	0.34	0.31	0.29	0.28	0.24	0.13			
P25/P22	0.37	0.35	0.34	0.35	0.35	0.33	0.27	0.27	0.14	0.18	0.21
P33/P34	0.24	0.24	0.24	0.25	0.23	0.24	0.25	0.18			
P23/P25	0.34	0.33	0.31	0.34	0.33	0.34	0.30	0.32			
P35/P36	-0.03	-0.03	-0.02	0.00	0.02	0.02	0.04	0.03			
P29/P27	0.36	0.36	0.35	0.35	0.41	0.41	0.23	0.34	0.23		
P23/P24	0.18	0.20	0.17	0.10	0.10	0.15	0.16	0.30	0.05		
P21/P22	0.35	0.36	0.34	0.34	0.31	0.28	0.23	0.26	0.16	0.04	0.17
P41/P42	0.37	0.36	0.35	0.32	0.28	0.38	0.41	0.45	0.30	0.27	0.24
P26/P27	0.52	0.51	0.53	0.52	0.53	0.52	0.47	0.44			
P24/P26	0.47	0.45	0.45	0.44	0.46	0.47	0.45	0.42			
P28/P27	0.31	0.32	0.33	0.32	0.30	0.24	0.27	0.29			
P37/P38	0.48	0.49	0.48	0.37	0.30	0.29					
P6/P7	0.46	0.45	0.44	0.45	0.44	0.44	0.48	0.48	0.45	0.38	

* indicates significant difference ($p<.05$) from 0 environments deleted

Supplemental Table 28: Test weight selection response (R, using the top 10% of predicted lines) across all populations after deleting different numbers of environments based on increased stability variance

	Number of Environments Deleted (Test weight)										
Population	0	1	2	3	4	5	6	7	8	9	10
P5/P8	0.40	0.34	0.47	0.53	0.47	0.63	0.51	0.60			
P2/P15	0.40	0.45	0.45	0.37	0.18	0.43	0.43	0.46	0.15		
P1/P2	0.32	0.26	0.26	0.26	0.29	0.38	0.36	0.24			
P13/P14	0.32	0.31	0.31	0.31	0.40	0.28	0.23	0.20			
P3/P4	0.56	0.59	0.69	0.64	0.58	0.58	0.52	0.61	-0.05		
P3/P8	0.47	0.46	0.48	0.49	0.58	0.56	0.46	0.29			
P16/P13	0.70	0.70	0.70	0.71	0.76	0.85	0.76	0.86			
P4/P5	0.40	0.42	0.47	0.30	0.36	0.44	0.36	0.35	0.36		
P17/P18	0.42	0.40	0.34	0.34	0.24	-0.17*	0.53	0.58			
P9/P10	0.49	0.43	0.45	0.32	0.41	0.29	0.44				
P19/P20	0.38	0.42	0.27	0.48	0.23	0.21					
P11/P12	0.44	0.48	0.48	0.55	0.48	0.48	0.32	0.33			
P1/P9	0.35	0.28	0.27	0.30	0.18	0.20	0.14				
P31/P32	0.73	0.69	0.56	0.66	0.67	0.53	0.59	0.41			
P25/P22	0.65	0.60	0.71	0.81	0.48	0.58	0.62	0.67	0.41*	0.51	0.43
P33/P34	0.88	0.97	0.98	0.98	0.89	0.84	0.70	0.62			
P23/P25	0.69	0.78	0.73	0.69	0.59	0.61	0.49	0.55			
P35/P36	-0.02	-0.02	-0.11	-0.04	-0.05	0.03	0.05	0.23			
P29/P27	0.44	0.46	0.38	0.41	0.47	0.48	0.38	0.39	0.23		
P23/P24	0.31	0.33	0.19	0.24	0.25	0.16	0.21	0.41	0.12		
P21/P22	0.45	0.26	0.28	0.43	0.51	0.35	0.33	0.18	0.37	0.08*	0.13
P41/P42	0.49	0.52	0.49	0.40	0.48	0.65	0.57	0.74	0.49	0.45	0.46
P26/P27	0.74	0.71	0.79	0.75	0.83	0.80	0.70	0.71			
P24/P26	0.67	0.68	0.75	0.56	0.58	0.65	0.65	0.58			
P28/P27	0.44	0.52	0.61	0.65	0.70	0.60	0.54	0.49			
P37/P38	0.58	0.53	0.55	0.53	0.56	0.58					
P6/P7	0.68	0.67	0.69	0.74	0.74	0.70	0.70	0.71	0.55	0.55	

* indicates significant difference ($p<.05$) from 0 environments deleted

Supplemental Table 29: Average predictive ability (r_{MP}) across all populations after deleting different numbers of environments based on increased stability

	Environments Deleted										
Trait	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Yield	0.20	0.19	0.19	0.20	0.19	0.19	0.18	0.17	0.07	0.07	-0.02*
Moisture	0.42	0.41	0.41	0.41	0.41	0.39	0.35	0.36	0.28	0.31	0.23*
Test weight	0.34	0.34	0.33	0.32	0.32	0.30	0.30	0.30	0.21	0.22	0.21

* indicates significant difference from deleting 0 environments

Supplemental Table 30: Average selection response (R, using the top 10% of predicted lines) across all populations after deleting different numbers of environments based on increased stability variance

	Environments Deleted										
Trait	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Yield	0.21	0.18	0.19	0.20	0.20	0.18	0.17	0.18	0.09	0.16	0.07
Moisture	-6.20	-6.21	-6.34	-5.99	-5.93	-5.81	-5.06	-5.93	-3.90*	-5.90	-5.00
Test weight	0.50	0.49	0.49	0.50	0.48	0.47	0.46	0.49	0.29	0.40	0.34

* indicates significant difference from delete 0

Supplemental Table 31: Predictive ability for the sum of genetic and environment effects ($G_i + E_j$) for yield under three scenarios for the source of environmental training data for an A/B test population: (i) having data on other populations, including those without A or B as a parent, and assuming the A/B test population has not been phenotyped; (ii) having data on other populations, including those without A or B as a parent, and assuming the A/B test population has been phenotyped in a subset of locations; and (iii) using data only from trials of populations that had either the A parent (A/*) or the B parent (*/B)

Populations	AIC			RR-BLUP		
	Scenario i	Scenario ii	Scenario iii	Scenario i	Scenario ii	Scenario iii
P2/P15	-0.36 (-0.42, -0.3)	-0.21 (-0.28, -0.14)	0.43 (0.37, 0.48)	0.51 (0.46, 0.56)	0.63 (0.59, 0.67)	0.31 (0.24, 0.37)
P1/P2	0.04 (-0.02, 0.10)	0.08 (0.02, 0.14)	-0.09 (-0.15, -0.03)	0.28 (0.22, 0.33)	0.27 (0.22, 0.33)	0.66 (0.62, 0.69)
P13/P14	0.32 (0.28, 0.37)	0.11 (0.06, 0.16)	0.11 (0.06, 0.16)	0.35 (0.30, 0.39)	0.34 (0.29, 0.39)	-0.11 (-0.16, -0.06)
P3/P4	0.66 (0.63, 0.69)	0.75 (0.73, 0.78)	-0.04 (-0.09, 0.02)	0.74 (0.72, 0.77)	0.79 (0.77, 0.81)	0.49 (0.45, 0.53)
P3/P8	0.12 (0.06, 0.17)	0.29 (0.23, 0.34)	0.42 (0.37, 0.46)	-0.05 (-0.11, 0.00)	0.34 (0.28, 0.39)	0.21 (0.15, 0.27)
P16/P13	0.50 (0.46, 0.54)	0.37 (0.32, 0.42)	0.02 (-0.03, 0.08)	0.45 (0.40, 0.49)	0.49 (0.45, 0.53)	0.49 (0.45, 0.53)
P4/P5	0.15 (0.09, 0.21)	0.28 (0.23, 0.34)	0.07 (0.01, 0.13)	0.23 (0.17, 0.29)	0.10 (0.05, 0.16)	0.09 (0.03, 0.15)
P17/P18	0.14 (0.08, 0.19)	-0.02 (-0.07, 0.04)	0.45 (0.40, 0.49)	0.48 (0.44, 0.52)	0.58 (0.54, 0.61)	0.14 (0.09, 0.20)
P9/P10	0.41 (0.36, 0.47)	0.27 (0.20, 0.33)	0.10 (0.04, 0.17)	0.28 (0.21, 0.34)	0.31 (0.25, 0.37)	0.12 (0.05, 0.18)
P19/P20	0.57 (0.53, 0.61)	0.35 (0.29, 0.40)	0.12 (0.06, 0.18)	0.61 (0.57, 0.65)	0.71 (0.67, 0.74)	0.04 (-0.03, 0.10)
P11/P12	-0.02 (-0.07, 0.03)	-0.03 (-0.08, 0.02)	-0.47 (-0.5, -0.42)	0.55 (0.51, 0.59)	0.60 (0.56, 0.63)	0.12 (0.07, 0.17)
P1/P9	0.61 (0.56, 0.65)	0.33 (0.27, 0.39)	0.22 (0.16, 0.29)	0.41 (0.36, 0.47)	0.47 (0.42, 0.52)	0.59 (0.55, 0.64)
P25/P22	-0.19 (-0.24, -0.14)	0.14 (0.08, 0.19)	0.21 (0.16, 0.26)	0.38 (0.33, 0.42)	0.54 (0.50, 0.58)	0.45 (0.41, 0.49)
P33/P34	0.04 (-0.02, 0.09)	0.00 (-0.05, 0.05)	0.36 (0.31, 0.41)	0.24 (0.19, 0.29)	0.22 (0.17, 0.27)	0.08 (0.03, 0.14)
P35/P36	0.41 (0.36, 0.46)	0.18 (0.13, 0.23)	0.16 (0.11, 0.22)	0.45 (0.41, 0.49)	0.51 (0.47, 0.55)	0.19 (0.14, 0.25)
P29/P27	0.37 (0.33, 0.42)	-0.09 (-0.15, -0.04)	0.11 (0.05, 0.16)	0.47 (0.42, 0.51)	0.24 (0.19, 0.29)	0.33 (0.28, 0.38)
P23/P24	-0.13 (-0.18, -0.07)	-0.05 (-0.11, 0.00)	0.78 (0.76, 0.80)	-0.23 (-0.29, -0.18)	-0.09 (-0.15, -0.04)	0.21 (0.16, 0.27)
P21/P22	0.00 (-0.06, 0.05)	-0.04 (-0.09, 0.01)	-0.42 (-0.46, -0.38)	0.64 (0.61, 0.67)	0.65 (0.61, 0.68)	-0.1 (-0.15, -0.05)
P41/P42	0.13 (0.07, 0.20)	0.24 (0.18, 0.30)	0.01 (-0.05, 0.08)	0.12 (0.05, 0.18)	0.20 (0.14, 0.26)	-0.55 (-0.59, -0.50)
P26/P27	0.50 (0.46, 0.55)	0.55 (0.50, 0.59)	0.67 (0.63, 0.70)	-0.18 (-0.24, -0.13)	0.04 (-0.02, 0.10)	0.63 (0.59, 0.67)
P24/P26	0.32 (0.26, 0.37)	0.58 (0.54, 0.62)	0.15 (0.10, 0.21)	-0.06 (-0.12, 0.00)	0.15 (0.09, 0.20)	0.49 (0.45, 0.54)
P28/P27	0.49 (0.44, 0.53)	0.47 (0.42, 0.51)	-0.34 (-0.39, -0.28)	0.42 (0.37, 0.47)	0.43 (0.38, 0.48)	0.59 (0.55, 0.63)

P37/P38	0.35 (0.27, 0.42)	0.19 (0.11, 0.27)	-0.35 (-0.42, -0.28)	0.48 (0.41, 0.54)	0.58 (0.52, 0.63)	-0.11 (-0.19, -0.03)
P6/P7	0.09 (0.04, 0.15)	0.10 (0.05, 0.16)	0.40 (0.36, 0.45)	0.26 (0.21, 0.31)	0.26 (0.21, 0.31)	0.12 (0.07, 0.18)
Mean	0.23 (0.18, 0.28)	0.20 (0.15, 0.25)	0.13 (0.08, 0.18)	0.33 (0.28, 0.37)	0.39 (0.34, 0.44)	0.23 (0.18, 0.28)

95% Confidence intervals are given in parentheses.

Supplemental Table 32: Predictive ability for the sum of genetic and environment effects ($G_i + E_j$) for moisture under three scenarios for the source of environmental training data for an A/B test population: (i) having data on other populations, including those without A or B as a parent, and assuming the A/B test population has not been phenotyped; (ii) having data on other populations, including those without A or B as a parent, and assuming the A/B test population has been phenotyped in a subset of locations; and (iii) using data only from trials of populations that had either the A parent (A/*) or the B parent (*/B)

Populations	AIC			RR-BLUP		
	Scenario i	Scenario ii	Scenario iii	Scenario i	Scenario ii	Scenario iii
P2/P15	0.84 (0.82, 0.86)	0.35 (0.29, 0.41)	-0.60 (-0.65, -0.56)	0.94 (0.93, 0.95)	0.95 (0.95, 0.96)	0.69 (0.65, 0.72)
P1/P2	0.35 (0.29, 0.40)	0.61 (0.57, 0.64)	0.90 (0.89, 0.91)	0.48 (0.43, 0.53)	0.53 (0.49, 0.57)	0.91 (0.90, 0.92)
P13/P14	0.71 (0.68, 0.74)	0.82 (0.81, 0.84)	-0.49 (-0.53, -0.45)	0.83 (0.81, 0.85)	0.89 (0.88, 0.90)	0.17 (0.12, 0.22)
P3/P4	0.31 (0.26, 0.36)	0.70 (0.68, 0.73)	0.30 (0.25, 0.35)	0.25 (0.20, 0.30)	0.46 (0.42, 0.50)	0.02 (-0.04, 0.07)
P3/P8	0.83 (0.82, 0.85)	0.85 (0.83, 0.86)	0.77 (0.74, 0.79)	0.51 (0.47, 0.55)	0.74 (0.72, 0.77)	-0.04 (-0.09, 0.02)
P16/P13	0.53 (0.49, 0.57)	0.45 (0.41, 0.50)	-0.33 (-0.38, -0.28)	0.46 (0.42, 0.51)	0.55 (0.51, 0.59)	0.14 (0.08, 0.19)
P4/P5	-0.12 (-0.18, -0.06)	-0.35 (-0.40, -0.30)	0.43 (0.38, 0.48)	0.50 (0.46, 0.55)	0.45 (0.40, 0.50)	0.80 (0.78, 0.82)
P17/P18	0.33 (0.29, 0.38)	0.31 (0.26, 0.35)	0.67 (0.64, 0.70)	0.69 (0.66, 0.72)	0.68 (0.65, 0.71)	0.34 (0.29, 0.39)
P9/P10	0.34 (0.28, 0.40)	0.31 (0.25, 0.37)	0.00 (-0.07, 0.06)	0.60 (0.56, 0.64)	0.67 (0.64, 0.71)	0.24 (0.18, 0.30)
P19/P20	-0.09 (-0.15, -0.02)	-0.15 (-0.21, -0.09)	0.10 (0.04, 0.16)	-0.63 (-0.67, -0.59)	-0.65 (-0.68, -0.61)	-0.35 (-0.41, -0.3)
P11/P12	0.47 (0.43, 0.51)	0.54 (0.50, 0.58)	-0.24 (-0.29, -0.20)	0.67 (0.65, 0.70)	0.78 (0.76, 0.80)	0.22 (0.18, 0.27)
P1/P9	0.88 (0.86, 0.89)	0.91 (0.89, 0.92)	-0.23 (-0.29, -0.17)	0.90 (0.89, 0.91)	0.93 (0.92, 0.94)	-0.65 (-0.69, -0.61)
P25/P22	0.56 (0.52, 0.6)	0.59 (0.56, 0.63)	0.12 (0.06, 0.17)	0.63 (0.59, 0.66)	0.59 (0.55, 0.62)	0.17 (0.12, 0.22)
P33/P34	0.79 (0.77, 0.81)	0.75 (0.73, 0.78)	0.81 (0.79, 0.82)	0.82 (0.81, 0.84)	0.79 (0.77, 0.81)	0.79 (0.77, 0.81)
P35/P36	0.70 (0.67, 0.72)	0.78 (0.75, 0.80)	0.68 (0.64, 0.70)	0.74 (0.72, 0.77)	0.85 (0.84, 0.87)	-0.44 (-0.48, -0.39)
P29/P27	0.22 (0.17, 0.27)	0.24 (0.19, 0.30)	0.27 (0.22, 0.32)	0.48 (0.44, 0.52)	0.56 (0.53, 0.60)	-0.14 (-0.20, -0.09)
P23/P24	0.37 (0.32, 0.42)	0.74 (0.72, 0.77)	0.14 (0.09, 0.20)	0.57 (0.53, 0.60)	0.70 (0.67, 0.73)	0.04 (-0.02, 0.09)
P21/P22	-0.27 (-0.31, -0.22)	0.18 (0.13, 0.23)	0.33 (0.28, 0.38)	-0.12 (-0.17, -0.07)	-0.06 (-0.11, 0.00)	-0.68 (-0.70, -0.65)
P41/P42	-0.02 (-0.08, 0.05)	0.37 (0.31, 0.43)	0.50 (0.45, 0.55)	0.51 (0.46, 0.55)	0.76 (0.73, 0.78)	-0.65 (-0.69, -0.61)
P26/P27	-0.40 (-0.44, -0.34)	0.62 (0.59, 0.66)	0.92 (0.91, 0.93)	-0.04 (-0.1, 0.02)	0.41 (0.36, 0.46)	0.92 (0.91, 0.93)
P24/P26	-0.37 (-0.42, -0.31)	0.64 (0.61, 0.68)	0.80 (0.77, 0.82)	0.04 (-0.02, 0.1)	0.29 (0.24, 0.35)	0.91 (0.90, 0.92)
P28/P27	0.54 (0.50, 0.58)	0.84 (0.83, 0.86)	0.71 (0.68, 0.74)	0.75 (0.72, 0.77)	0.83 (0.81, 0.85)	0.77 (0.74, 0.79)

P37/P38	0.69 (0.65, 0.73)	0.80 (0.76, 0.83)	0.76 (0.73, 0.80)	0.67 (0.63, 0.72)	0.76 (0.73, 0.80)	0.57 (0.51, 0.62)
P6/P7	0.72 (0.69, 0.74)	0.40 (0.35, 0.44)	0.26 (0.21, 0.31)	0.72 (0.70, 0.75)	0.75 (0.72, 0.77)	0.54 (0.50, 0.57)
Mean	0.37 (0.33, 0.41)	0.51 (0.47, 0.55)	0.32 (0.27, 0.36)	0.50 (0.46, 0.54)	0.59 (0.56, 0.62)	0.22 (0.18, 0.26)

95% Confidence intervals are given in parentheses.

Supplemental Table 33: Predictive ability for the sum of genetic and environment effects ($G_i + E_j$) for test weight under three scenarios for the source of environmental training data for an A/B test population: (i) having data on other populations, including those without A or B as a parent, and assuming the A/B test population has not been phenotyped; (ii) having data on other populations, including those without A or B as a parent, and assuming the A/B test population has been phenotyped in a subset of locations; and (iii) using data only from trials of populations that had either the A parent (A/*) or the B parent (*/B)

Populations	AIC			RR-BLUP		
	Scenario i	Scenario ii	Scenario iii	Scenario i	Scenario ii	Scenario iii
P2/P15	0.90 (0.88, 0.91)	0.83 (0.81, 0.85)	0.04 (-0.03, 0.11)	0.68 (0.64, 0.72)	0.91 (0.90, 0.93)	0.74 (0.71, 0.77)
P1/P2	0.39 (0.34, 0.44)	0.25 (0.20, 0.31)	0.72 (0.69, 0.74)	0.30 (0.25, 0.36)	0.26 (0.20, 0.31)	0.83 (0.81, 0.85)
P13/P14	0.43 (0.39, 0.48)	0.58 (0.54, 0.61)	0.30 (0.25, 0.35)	0.45 (0.41, 0.49)	0.69 (0.66, 0.71)	0.66 (0.62, 0.68)
P3/P4	0.64 (0.6, 0.67)	0.64 (0.60, 0.67)	-0.31 (-0.36, -0.26)	0.82 (0.80, 0.83)	0.66 (0.63, 0.69)	0.26 (0.21, 0.31)
P3/P8	-0.23 (-0.29, -0.17)	-0.30 (-0.36, -0.25)	-0.37 (-0.42, -0.32)	0.30 (0.25, 0.35)	0.32 (0.26, 0.37)	0.65 (0.62, 0.69)
P16/P13	-0.12 (-0.17, -0.06)	-0.48 (-0.52, -0.44)	0.54 (0.50, 0.58)	0.37 (0.32, 0.42)	0.44 (0.40, 0.49)	-0.33 (-0.38, -0.28)
P4/P5	0.36 (0.30, 0.41)	0.39 (0.33, 0.44)	0.18 (0.12, 0.24)	0.25 (0.20, 0.31)	0.33 (0.27, 0.38)	0.20 (0.15, 0.26)
P17/P18	0.42 (0.37, 0.46)	0.45 (0.41, 0.49)	-0.30 (-0.35, -0.25)	0.59 (0.55, 0.62)	0.73 (0.70, 0.76)	-0.08 (-0.13, -0.02)
P9/P10	0.11 (0.04, 0.17)	0.20 (0.13, 0.26)	0.70 (0.66, 0.73)	-0.08 (-0.14, -0.01)	0.03 (-0.03, 0.10)	0.73 (0.70, 0.76)
P19/P20	0.77 (0.74, 0.79)	0.64 (0.61, 0.68)	-0.47 (-0.52, -0.42)	0.87 (0.85, 0.88)	0.88 (0.86, 0.89)	0.31 (0.25, 0.37)
P11/P12	0.74 (0.72, 0.76)	0.66 (0.63, 0.69)	0.10 (0.05, 0.16)	0.56 (0.52, 0.59)	0.73 (0.71, 0.75)	0.28 (0.24, 0.33)
P1/P9	-0.31 (-0.37, -0.25)	-0.33 (-0.39, -0.27)	-0.25 (-0.31, -0.19)	-0.28 (-0.34, -0.22)	0.18 (0.11, 0.24)	-0.23 (-0.3, -0.17)
P25/P22	0.73 (0.71, 0.76)	0.67 (0.64, 0.70)	0.72 (0.69, 0.74)	0.61 (0.58, 0.65)	0.73 (0.70, 0.75)	0.78 (0.75, 0.80)
P33/P34	0.56 (0.52, 0.59)	0.54 (0.50, 0.58)	0.63 (0.60, 0.66)	0.73 (0.71, 0.76)	0.79 (0.77, 0.81)	0.73 (0.70, 0.75)
P35/P36	-0.19 (-0.25, -0.14)	-0.04 (-0.10, 0.01)	0.00 (-0.06, 0.05)	0.23 (0.17, 0.28)	0.32 (0.27, 0.37)	-0.07 (-0.12, -0.01)
P29/P27	0.05 (0.00, 0.11)	-0.01 (-0.07, 0.04)	0.13 (0.08, 0.19)	0.31 (0.26, 0.36)	0.13 (0.07, 0.18)	-0.59 (-0.63, -0.55)
P23/P24	0.25 (0.20, 0.31)	0.06 (0.00, 0.11)	0.03 (-0.03, 0.09)	-0.22 (-0.27, -0.17)	0.16 (0.10, 0.21)	0.06 (0.00, 0.11)
P21/P22	0.29 (0.24, 0.34)	0.20 (0.15, 0.25)	0.54 (0.50, 0.57)	0.74 (0.72, 0.77)	0.86 (0.84, 0.87)	0.11 (0.06, 0.16)
P41/P42	0.43 (0.38, 0.49)	0.14 (0.08, 0.20)	0.36 (0.30, 0.41)	0.71 (0.67, 0.74)	0.64 (0.60, 0.68)	0.63 (0.59, 0.67)
P26/P27	0.70 (0.67, 0.73)	0.40 (0.35, 0.45)	0.69 (0.65, 0.72)	0.36 (0.31, 0.41)	0.38 (0.32, 0.43)	0.86 (0.84, 0.87)
P24/P26	0.55 (0.51, 0.59)	0.30 (0.25, 0.36)	0.33 (0.27, 0.38)	0.32 (0.26, 0.37)	0.34 (0.28, 0.39)	0.66 (0.63, 0.69)
P28/P27	0.70 (0.67, 0.73)	0.68 (0.65, 0.71)	0.74 (0.71, 0.76)	0.63 (0.59, 0.66)	0.60 (0.56, 0.63)	0.11 (0.05, 0.17)

P37/P38	0.07 (-0.01, 0.16)	-0.02 (-0.10, 0.07)	-0.06 (-0.14, 0.03)	-0.08 (-0.17, 0.00)	-0.06 (-0.14, 0.02)	0.05 (-0.03, 0.13)
P6/P7	0.42 (0.37, 0.46)	0.60 (0.56, 0.63)	0.01 (-0.05, 0.06)	0.83 (0.81, 0.85)	0.79 (0.76, 0.81)	0.26 (0.21, 0.31)
Mean	0.36 (0.32, 0.40)	0.29 (0.25, 0.34)	0.21 (0.16, 0.26)	0.42 (0.37, 0.46)	0.49 (0.45, 0.53)	0.32 (0.27, 0.36)

95% Confidence intervals are given in parentheses.