

Applications of Genomewide Selection in a New Plant Breeding Program

A THESIS
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Jeffrey L. Neyhart

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Kevin P. Smith

July 2019

Copyright © 2019 Jeffrey L. Neyhart

ABSTRACT

Newly established breeding programs must undergo population improvement and determine superior germplasm for deployment in diverse growing environments. More rapid progress towards these goals may be made by incorporating genomewide selection, or the use of genomewide molecular markers to predict the merit of unphenotyped individuals. Within the context of a new two-row barley (*Hordeum vulgare* L.) breeding program, my objectives were to i) investigate various methods of updating training population data and their impact on long-term genomewide recurrent selection, ii) assess genomewide prediction accuracy with informed subsetting of data across diverse environments, and iii) validate genomewide predictions of the mean, genetic variance, and superior progeny mean of potential breeding crosses. My first study relied on simulations to examine the impact on prediction accuracy and response to selection when updating the training population each cycle with lines selected based on predictions (best, worst, both best and worst), model criteria (PEVmean and CDmean), random sampling, or no selections. In the short-term, we found that updating with the best or both best and worst predicted lines resulted in high prediction accuracy and genetic gain; in the long-term, all methods (besides not updating) performed similarly. In an actual breeding program, a breeder may want phenotypic data on lines predicted to be the best and our results suggest that this method may be effective for long-term genomewide selection and practical for a breeder. In my second study, a 183-line training population and 50-line offspring validation population were phenotyped in 29 location-year environments for grain yield, heading date, and plant height. Environmental relationships were measured using phenotypic data, geographic distance, or environmental covariables. When adding data from increasingly distant environments to a training set, we observed diminishing gains in prediction accuracy; in some cases, accuracy declined with additional data. Clustering environments led to a small, but non-significant gain in prediction accuracy compared to simply using

data from all environments. Our results suggest that informative environmental subsets may improve genomewide selection within a single population, but not when predicting a new generation under realistic breeding circumstances. Finally, my third study used genomewide marker effects from the same training population above to predict the mean (μ), genetic variance (V_G), and superior progeny mean (μ_{SP} ; mean of the best 10% of lines) of 330,078 possible crosses for *Fusarium* head blight (FHB) severity, heading date, and plant height. Twenty-seven of these crosses were developed as validation populations. Predictions of μ and μ_{SP} were moderate to high in accuracy ($r_{MP} = 0.46 - 0.69$), while predictions of V_G were less accurate ($r_{MP} = 0.01 - 0.48$). Predictive ability was likely a function of trait heritability, as r_{MP} estimates for heading date (the most heritable) were highest and r_{MP} estimates for FHB severity (the least heritable) were lowest. Accurate predictions of V_G and μ are feasible, but, like any implementation of genomewide selection, reliable phenotypic data is critical.

Table of Contents

Abstract	i
List of Tables	iv
List of Figures	v
List of Abbreviations	vi
Preface	vii
1 Chapter 1: Evaluating methods of updating training data in long-term genomewide selection	1
Introduction	3
Methods and Materials	6
Results	14
Discussion	19
Figures	28
2 Chapter 2: Strategies for subdividing environments in inter-generational genomewide selection	33
Introduction	34
Methods and Materials	37
Results	46
Discussion	51
Figures	59
Tables	66
Supplemental Figures	67
Supplemental Tables	72
3 Chapter 3: Validating genomewide predictions of genetic variance in a contemporary breeding program	76
Introduction	78
Methods and Materials	81
Results and Discussion	88
Figures	97
Tables	101
Supplemental Figures	103
Supplemental Tables	104
Epilogue: Future Research Directions	105
Bibliography	120

List of Tables

2.1	Descriptions of distance measures.	66
2.2	Intra-environment heritability.	72
2.3	Environmental covariate descriptions	73
2.4	Summary of variance components per population.	74
2.5	Significant environmental covariables	75
3.1	Attributes of validation families	101
3.2	Prediction accuracy and bias	102
3.3	Validation family heritability.	104

List of Figures

1.1	Typical breeding cycle timeframe.	28
1.2	Diagram of simulated breeding cycle.	29
1.3	Prediction accuracy over breeding cycles.	30
1.4	Genetics variance and response to selection over breeding cycles	31
1.5	Genomic relationship, inbreeding, number of QTL fixed for an allele, and persistence of linkage disequilibrium phase over breeding cycles.	32
2.1	Map of multi-environment trials.	59
2.2	Predictions schemes using environmental distance measures.	60
2.3	Biplots from AMMI analysis.	61
2.4	Leave-one-environment-out ranked prediction results.	62
2.5	Time-forward ranked prediction results.	63
2.6	Leave-one-environment-out cluster prediction results.	64
2.7	Time-forward cluster prediction results.	65
2.8	Phenotypic correlation heatmap	67
2.9	Cluster comparison	68
2.10	Cluster heritability estimates	69
2.11	Prediction accuracies using all data.	70
2.12	Genomic relatedness between populations	71
3.1	Comparison of predicted family means and genetic variances.	97
3.2	Distribution of three trait phenotypes on a per-family basis.	98
3.3	Predictive ability of family mean, genetic variance, and superior progeny mean.	99
3.4	Bias when predicting genetic variance.	100
3.5	Prediction accuracy of Fusarium head blight resistance in two locations.	103

List of Abbreviations

ABR, USDA-ARS Aberdeen, ID; AMMI, additive main effects and multiplicative interactions; BARI, Busch Agricultural Resources, Inc.; BCAP, Barley Coordinated Agricultural Project; BLUE, best linear unbiased estimate; BLUP, best linear unbiased prediction; CRD, completely randomized design; CV, cross-validation; EC, environmental covariable; FHB, Fusarium head blight; GBS, genotyping-by-sequencing; GEI, genotype-environment interaction; GGE, genotype main effects and genotype-by-environment interactions; GWS, genomewide selection; IBD, incomplete block design; IPCA, interaction principal component axis; LD, linkage disequilibrium; LOEO, leave-one-environment-out; MSU, Montana State University; NDSU, North Dakota State University; NOAA, National Oceanic and Atmospheric Administration; PGV, predicted genotypic value; POV, parent-offspring validation; QTL, quantitative trait loci; RIL, recombinant inbred line; RR-BLUP, ridge regression best linear unbiased prediction; SNP, single nucleotide polymorphism; T3, The Triticeae Toolbox; TF, time-forward; TP, training population; TS, training set; UMN, University of Minnesota; USDA-ARS, United States Department of Agriculture - Agricultural Research Service; VP, validation population; WSU, Washington State University

Preface

In 2013, the University of Minnesota barley breeding program initiated efforts to improve two-row barley for the Upper Midwest region. Much of the motivation behind this new program stemmed from demand in the malting and brewing industries for superior cultivars. The craft brewing segment of this industry, a source of much of this demand, has experienced rapid growth in the last decade. In that time, the number of breweries in the United States increased nearly five-fold from about 1,500 to nearly 7,500, almost twice the pre-Prohibition high (Brewers Association, 2018). In Minnesota, the proliferation of breweries has been equally robust, and the state now boasts 178 such establishments, with an economic impact of more than \$2 billion (Brewers Association, 2018).

The introduction of this breeding program coincided with a time of rapid technological change in the field of plant breeding and genetics. First, the type and size of data has shifted. The declining cost of genome sequencing has made molecular markers abundant and ubiquitous, while the presence of hand-held computers and unmanned aerial vehicles permits more rapid and error-free collection of phenotypic data. Second, the changing landscape of data has necessitated new computational tools to make this data useful for breeders. For instance, an understanding of bioinformatics analysis pipelines is now necessary to identify molecular markers from genome sequencing data. Open-source tools for statistical analysis, such as the R programming language, have become standard for students and practitioners of breeding, and there is a constant output of new extensions for the language.

With any dissertation project, the technological capacity and contemporary trends dictated the focus of this thesis. The three chapters are all centered around the use of genomewide molecular markers to make predictions of – and subsequently selections on – breeding material that has not yet been phenotyped, in a process called genomewide selection (Meuwissen

et al., 2001). Much research has been conducted on the theoretical or empirical accuracy of genomewide predictions of simple genotypic means (Lorenzana and Bernardo, 2009; Heffner et al., 2011; Asoro et al., 2011; Iwata and Jannink, 2011; Combs and Bernardo, 2013; Lorenz et al., 2012; Sallam et al., 2015). So, while all chapters in this thesis address questions related to the implementation of genomewide selection in a newly established breeding program, their focus is aimed at ancillary topics such as logistical breeding decisions, advanced modeling, and extensions of marker effects. Chapter 1 describes a simulation study that explored different methods of updating genomewide selection training data to maintain gain from selection in a long-term breeding program. Chapter 2 documents the results of a project to strategically subset phenotypic data from a multi-environment dataset to make predictions for locally-adapted breeding material. Chapter 3 presents an analysis of a new use of genomewide markers: predicting the genetic variance in breeding populations. Finally, I close with a short epilogue that frames future directions and new questions that may stem from this work.

I have the utmost gratitude to my advisor, Dr. Kevin P. Smith, for the years of mentorship. He has consistently created an atmosphere of respect, collaboration, and positivity, and there are few people who are better-able to maintain perspective and a vision for long-term goals, especially when best laid plans inevitably go awry.

Many thanks go to Karen Beaubien, Ed Schiefelbein, Dimitri von Ruckert, and Guillermo Velasquez for their technical assistance in the laboratory and in the field. They have helped avert disaster on more than one occasion. I thank past and present postdocs and graduate students of the barley group, including Mohsen Mohammadi, Ana Poets, Jo Heuschele, Austin Case, Ahmad Sallam, Celeste Falcon, Tyler Tiede, Alexandria Hemshrot (*née* Ollhoff), Lu Yin, John Hill Price, Ian McNish, Becky Zhong, and Alexander Susko, for stimulating conversation, project assistance, and constant encouragement.

I also thank my committee members, Drs. Rex Bernardo, Candice Hirsch, and Brian Steffenson, for advice and support related not only to my project, but also to career and life

goals. I am grateful to many others for productive collaborations and conversations, including Drs. Aaron Lorenz and Eric Watkins.

Just as diverse as my thesis projects were the sources of funding to make them possible. I received support from the Applied Plant Sciences Graduate Program, the Minnesota Department of Agriculture, Rahr Malting Company, the U.S. Wheat and Barley Scab Initiative, the Brewers Association, and the United States Department of Agriculture - National Institute of Food and Agriculture. Resources from the Minnesota Supercomputing Institute were used to complete all projects.

I will be eternally grateful for the many friendships I have cultivated during my graduate career, all of which made this time the most enjoyable yet of my life. Many thanks to my family, who tolerated long distances and intermittent reunions so that I could pursue this degree. I am thankful that I can count on your sustaining love and support.

Jeffrey L. Neyhart

St. Paul, Minnesota

July 2019

Chapter 1: Evaluating methods of updating training data in long-term genomewide selection

The following chapter was published in a peer-reviewed journal and can be located using the reference:

Jeffrey L. Neyhart, Tyler Tiede, Aaron J. Lorenz, and Kevin P. Smith. 2017. Evaluating Methods of Updating Training Data in Long-Term Genomewide Selection. *G3: Genes|Genomes|Genetics*, 7:1499-1510. doi: 10.1534/g3.117.040550.

Copyright © 2017 Neyhart et al.

Author contributions:

JLN conducted the simulations, collected/analyzed data, and wrote the manuscript and TT, AJL, and KPS provided edits. All authors conceived of the study and read and approved the final manuscript.

INTRODUCTION

The improvement of populations in plant breeding through recurrent selection may benefit tremendously from genomewide selection. Of particular worth are the high accuracies and shortened breeding cycles of genomewide selection, which allow for greater genetic gains per unit time (Bernardo and Yu, 2007; Heffner et al., 2009; Lorenz et al., 2011). While genomewide selection has already been employed in established breeding programs for major cultivated species (e.g. Asoro et al., 2013; Beyene et al., 2015; Sallam et al., 2015), this tool also has broad appeal across other species. For instance, breeding programs for tree or perennial crops with long generation times could find utility in making selections before the plants are mature enough to phenotype. Additionally, orphan, undomesticated, or unimproved crops may benefit from rapid breeding progress. Indeed, researchers have already investigated the use of genomewide selection in species such as apple (*Malus x domestica*; Kumar et al., 2012), Eucalyptus (Resende et al., 2012), oil palm (*Elaeis guineensis* Jacq.; Cros et al., 2015), and intermediate wheatgrass (*Thinopyrum intermedium* (Host) Barkworth & D.R. Dewey; Zhang et al., 2016). The population improvement necessary in newly established breeding programs, regardless of species, may be expedited through genomewide selection.

Of course, the aforementioned advantages of genomewide selection depend on maintaining sufficient genetic gain. This requires accurate predictions of the genotypic value of selection candidates based on markers located throughout the genome (Meuwissen et al., 2001). Accurate predictions depend on reliable phenotypic measurements and sufficient marker data on a training population. Genomewide marker coverage that captures genomic relationships between individuals and ensures linkage disequilibrium (LD) between markers and quantitative trait loci (QTL) will lead to higher prediction accuracy, especially when predictions are applied to selection candidates more distantly related to the training population (Habier et al., 2007; Lorenz et al., 2011). The predicted genotypic values under these conditions will more

closely reflect the true genotypic values, and selection can then act to increase the frequency of favorable QTL alleles in a population and shift the mean of a population in a desirable direction.

Characteristics of long-term recurrent selection create impediments to maintaining effective genomewide selection. Over generations, recombination between markers and QTL will cause LD to decay, while selection and drift will potentially act to generate new LD or tighten the LD between closely-linked loci (Hill and Robertson, 1968; Lorenz et al., 2011). Shifts in the pattern of QTL-marker LD, if not captured, will result in decreased prediction accuracy. This suggests that training populations must be updated during recurrent selection to maintain prediction accuracy, a notion that is indeed supported by studies using simulations and empirical data. Studies exploring simulations of recurrent selection in a clonally-propagated crop (Eucalyptus) and an inbreeding small grain (barley [*Hordeum vulgare* L.]) both revealed that the accuracy of genomewide selection was improved by updating the training population with data from previous breeding cycles (Jannink, 2010; Denis and Bouvet, 2013). Similarly, using empirical data from an advanced-cycle rye (*Secale cereal* L.) breeding program, Auinger et al. (2016) found that aggregating training population data over multiple cycles enhanced prediction accuracy. These investigations all demonstrated the benefit of including previous-cycle data into a training population, however they did not test different methods of selecting that data.

Though updating the training population may be required, there are practical considerations in how a breeder selects individuals to fulfill this need. Consider a breeding program employing genomewide recurrent selection in barley. Each year, the breeder must allocate phenotyping resources between testing potential cultivars and population improvement. Though genomewide selection offers to reduce the overall phenotyping costs of the latter (e.g. through early-generation selection), promising breeding lines will undoubtedly be included in field trials. Under genomewide selection, it seems a breeder must also contend with the composition of their training population, placing emphasis on methods to build or maintain this population

that both maximize prediction accuracy and minimize costs.

Given the resource limitations of practical breeding and the importance of the training population, it is fitting that much research has been devoted to the composition and design of such populations. Using data from a North American barley breeding program, Lorenz et al. (2012) reported reduced prediction accuracy when the training population and selection candidates belonged to separate subpopulations. Multiple studies have found that a training population that is more closely related to the selection candidates leads to more accurate predictions (Asoro et al., 2011; Lorenz and Smith, 2015). Other researchers have suggested more explicit criteria to determine the optimal training population for a set of selection candidates. Rincent et al. (2012) described training population design based on minimizing the mean prediction error variance (PEV) or maximizing the expected reliability of predictions (i.e. generalized coefficient of determination [CD]). When applied to empirical datasets, several investigations supported using the expected reliability criterion to optimally construct training populations (Rincent et al., 2012; Akdemir et al., 2015; Isidro et al., 2015; Rutkoski et al., 2015; Bustos-Korts et al., 2016a). These studies generally explored the construction of training populations from a single set of calibration individuals, therefore, the usefulness of this criterion over multiple breeding cycles to maintain prediction accuracy is unknown.

The objective of this study was to investigate various methods of updating a training population and their impact on genomewide recurrent selection. Using simulations, we envisioned a breeding program implementing genomewide recurrent selection for an inbreeding, small grain species (i.e. barley). Six different training population update methods were compared, along with two scenarios of training population composition. We tracked important variables in breeding, including prediction accuracy, response to selection, and genetic variance. Additionally, we attempted to explain some of our observations using other parameters, including persistence of LD phase and genomic relationship.

METHODS AND MATERIALS

A barley breeding program employing genomewide selection can realistically complete a breeding cycle in a single year (Fig. 1.1). Following this breeding timeline, our experiment simulates a breeding population undergoing 15 cycles of recurrent genomewide selection.

To incorporate the observed LD structure in barley breeding populations into our simulations, we used empirical marker data from two North American barley breeding programs: the University of Minnesota (UMN) and North Dakota State University (NDSU). Marker genotypes from 768 six-row spring inbred lines at 3,072 bi-allelic SNP loci were obtained from the Triticeae Toolbox (T3) database (Close et al., 2009; Blake et al., 2016). The genetic map position of markers was based on the consensus linkage map created by Muñoz-Amatriaín et al. (2011). Markers with more than 10% missing data and lines with more than 10% missing data were excluded. Markers were also filtered for redundancy, defined as those located at identical genetic map positions and with identical allele calls. A 0.01 cM interval was forced between markers with non-identical allele calls and shared map positions (i.e. due to low genetic map resolution). We set all heterozygous genotype calls to missing and imputed missing genotypes using the mode genotype across all samples. This left a set of 764 breeding lines and 1,590 homozygous markers spanning 1,137 cM.

Genetic Model to Simulate QTL

Each iteration of the simulation was initiated by randomly selecting $L = 100$ SNP loci to become causal QTL, regardless of genetic position or minor allele frequency. Genotypic values for QTL were drawn from a geometric series, as suggested by Lande and Thompson (1990). At the k th QTL, the value of the favorable homozygote was a^k , the value of the heterozygote was 0, and the value of the unfavorable homozygote was a^k , where $a = (1L)/(1 + L)$. The value of the first allele of a QTL was randomly assigned to be favorable or unfavorable. Dominance and

epistasis were assumed absent and higher values of the trait were considered favorable. The genotypic value of a given individual was calculated as the sum of the effects of QTL alleles carried by that individual.

Phenotypic values were simulated by adding nongenetic effects to the genotypic values according to the model $y_{ij} = g_i + e_j + \epsilon_{ij}$, where y_{ij} was the phenotypic value of the i th individual in the j th environment, g_i was the genotypic value of the i th individual, e_j was the effect of the j th environment, and ϵ_{ij} was the residual effect of the i th individual in the j th environment. Environmental effects were assumed to be samples of a normally-distributed random variable with mean 0 and standard deviation $\sqrt{\sigma_E^2}$, where σ_E^2 was eight times the variance among genotypic values (i.e. σ_G^2) (Bernardo, 2014). Residual effects were assumed to be samples of a normally-distributed random variable with mean 0 and standard deviation $\sqrt{\sigma_R^2}$, where σ_R^2 was scaled to achieve a target entry-mean heritability of $h^2 = 0.5$ in the base population. Phenotyping was assumed to take place in three environments with one replication, therefore within-environment variance and genotype-by-environment variance were confounded into σ_R^2 . The variance of environmental effects and the variance of residual effects remained unchanged over cycles of selection, allowing the heritability to vary. The mean phenotypic value of each individual over the three environments was used in genomewide prediction.

Base Population and Cycle 1 of Genomewide Selection

The base population (i.e. cycle 0 training population) consisted of genotypic and simulated phenotypic data on the 764 breeding lines. Based on these simulated phenotypes, the top fifty UMN lines and the top fifty NDSU lines were intermated between breeding programs to generate the cycle 1 population. Specifically, fifty crosses were simulated, using each parent once, and twenty F_3 -derived lines were generated per cross. Gametes were generated following Mendelian laws of segregation, with recombination events simulated according to the genetic map positions of all loci (Muñoz-Amatriaín et al., 2011) and assuming no cross-over interfer-

ence or mutation. Population development resulted in a pool of 1,000 F₃ selection candidates.

The marker data for the training population and selection candidates consisted of genotypes at all loci except the 100 QTL. This essentially simulated “genotyping” with complete accuracy. Monomorphic markers and those with a minor allele frequency less than 0.03 were removed prior to genomewide prediction. Marker effects were predicted using ridge-regression best linear unbiased prediction (RR-BLUP) according to the model

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}_{\text{TP}}\mathbf{u} + \mathbf{e}, \quad (1.1)$$

where \mathbf{y} was an $N \times 1$ vector of the phenotypic means of N training population lines, $\mathbf{1}$ was a $N \times 1$ vector of ones, μ was the grand mean, \mathbf{Z}_{TP} was a $N \times m$ incidence matrix of training population genotypes for m markers, \mathbf{u} was a $m \times 1$ vector of marker effects, and \mathbf{e} was a $N \times 1$ vector of residuals. Elements of \mathbf{Z}_{TP} were 1 if homozygous for the first allele, -1 if homozygous for the second allele, and 0 if heterozygous. Genotypic values of the F₃ selection candidates were predicted using the equation $\hat{\mathbf{g}} = \mathbf{Z}_{\text{SC}}\hat{\mathbf{u}}$, where $\hat{\mathbf{g}}$ was a $1,000 \times 1$ vector of predicted genotypic values, \mathbf{Z}_{SC} was a $1,000 \times m$ incidence matrix of selection candidate genotypes, and $\hat{\mathbf{u}}$ was a $m \times 1$ vector of predicted marker effects. Elements of \mathbf{Z}_{SC} were the same as those in \mathbf{Z}_{TP} .

Cycles 2 Through 15 of Genomewide Selection

Subsequent cycles of the simulation consisted of three steps: 1) crossing and population development, 2) prediction and selection, and 3) training population updating. These are outlined in the diagram presented in Fig. 1.2. Parents selected in the previous cycle were randomly intermated to form a pool of selection candidates. Again, fifty crosses were simulated and 1,000 F₃-derived selection candidates were generated. Prior to predictions, we removed monomorphic markers and those with a minor allele frequency less than 0.03 in both the pool of selection

candidates and in the training population. Since markers could become monomorphic due to selection or drift, the number of markers used for prediction decreased over breeding cycles. We predicted marker effects by Eq. 1.1, using phenotypic and genotypic data on the training population. These marker effects were then used to predict genotypic values of the 1,000 selection candidates, and those with the top 100 predicted genotypic values were designated as parents for the next cycle. A subset of all selection candidates were then designated as new additions to the training population according to one of the updating methods described below. We simulated phenotypes for these additions and merged the phenotypic and genotypic data to the pool of training population data.

Methods of updating the training population

Seven different methods of updating the training population were explored in the simulations. For each method, 150 selection candidates from each cycle were selected and added to the training population. These methods are termed “Top,” “Bottom,” “Random,” “PEVmean,” “CDmean,” “Tails,” and “No Change” and are described below. For “Top,” “Bottom,” and “Tails,” selection candidates were ranked based on predicted genotypic value. The 150 selection candidates with the highest (“Top”) or lowest (“Bottom”) values were added to the training population. For the “Tails” method, the 75 selection candidates with the highest values and the 75 selection candidates with the lowest values were added to the training population. For “Random,” a random sample of selection candidates were added to the training population, and for “No Change,” the training population was not updated over breeding cycles.

Two methods involved optimization algorithms previously described by other researchers, specifically “PEVmean” and “CDmean” (Rincent et al., 2012). Using only the genotypic data on all individuals, these algorithms aim to create a training population by optimally sampling individuals to be phenotyped in order to predict the value of individuals that would be unphenotyped. Our intention is similar, except that the individuals we sampled

to be phenotyped are one cycle removed from the individuals that would be unphenotyped. For PEV_{mean} , selection candidates were chosen to minimize the mean prediction error variance (PEV) of the genotypic values. As described in Rincent et al. (2012), the general PEV can be computed using a matrix of contrasts, C , between the “unphenotyped” individuals and the mean of the whole population (“phenotyped” and “unphenotyped” individuals). In solving Henderson’s (1984) equations, the PEV of any contrast can be computed as

$$PEV(\mathbf{C}) = \text{diag} \left[\frac{\mathbf{C}' \left(\mathbf{Z}'\mathbf{M}\mathbf{Z} + \frac{\sigma_e^2}{\sigma_a^2} \mathbf{A}^{-1} \right)^{-1} \mathbf{C}}{\mathbf{C}'\mathbf{C}} \right] \times \sigma_e^2, \quad (1.2)$$

where \mathbf{Z} is an incidence matrix, \mathbf{M} is an orthogonal projector (Rincent et al., 2012), and \mathbf{A} is the genomic relationship matrix (described below). For the variance of the residuals (σ_e^2), we used the restricted maximum likelihood estimate of σ_e^2 from the RR-BLUP linear model in Eq. 1.1. The additive genetic variance (σ_g^2) was calculated by multiplying the number of markers, N_m , by the restricted maximum likelihood estimate of the variance of marker effects (Bernardo, 2014). The PEV_{mean} was then calculated as $PEV_{mean} = \text{mean}[\text{diag}(PEV(\mathbf{C}))]$.

Similarly, for “CDmean,” candidates were chosen to maximize the reliability of the predictions, measured as the mean generalized coefficient of determination (CD). This can also be expressed as the expected reliability of the contrasts (Laloe, 1993; Rincent et al., 2012), computed as

$$CD(\mathbf{C}) = \text{diag} \left[\frac{\mathbf{C}' \left(\mathbf{A} - \frac{\sigma_e^2}{\sigma_a^2} \left(\mathbf{Z}'\mathbf{M}\mathbf{Z} + \frac{\sigma_e^2}{\sigma_a^2} \mathbf{A}^{-1} \right) \right)^{-1} \mathbf{C}}{\mathbf{C}'\mathbf{C}} \right]. \quad (1.3)$$

The values of σ_e^2 and σ_a^2 were the same as described for Eq. 1.2. The CDmean was then calculated as $CD_{mean} = \text{mean}[\text{diag}(CD(\mathbf{C}))]$.

We implemented an exchange algorithm similar to that described by Rincent et al.

(2012), with one modification in the designation of individuals to predict and individuals to sample for phenotyping. The situation outlined by Rincent et al. (2012) assumes that the genotypic data for the individuals to sample and for the individuals to predict is available concurrently. In our simulation, this is not the case, since phenotyping of the selections in one cycle (cycle n) will occur before genotypic data on selection candidates of the next cycle (cycle $n + 1$) becomes available (Fig. 1.1). We therefore chose the 100 parents of the cycle $n + 1$ selection candidates to be a proxy for the unphenotyped individuals, while the entire 1,000 selection candidates (including the parents) constituted the population of individuals to be sampled by the algorithm. To maintain a reasonable computation time, the exchange algorithms were iterated 500 times. Preliminary data showed that a reasonable optimum for either criterion was reached after 500 iterations (data not shown). The PEVmean or CDmean algorithms were used to select individuals from the selection candidates to be included in the training population for the next cycle.

We also considered two scenarios of using the updated training population data. The first scenario represented a situation where a breeder may want to use all available information, and in this case, the training population grew by 150 lines in each cycle. This was termed the “Cumulative” scenario, and over cycles the size of the training population ranged from 764 to 2,864 individuals. In the next scenario, we attempted to control for the effect of training population size by using a “sliding window” of 764 lines along breeding cycles. Specifically, in each cycle the 150 new training population additions from the latest breeding cycle took the place of the 150 training population additions from the earliest breeding cycle. Since the 764 base population lines all constituted cycle 0, these lines were discarded randomly until no base population lines remained in the training population. Afterwards, lines from earlier cycles were discarded as lines from later cycles were added. This was termed the “Window” scenario.

Variables tracked over breeding cycles

To better interpret the observations in the simulations, we tracked a number of additional variables, including persistence of LD phase, mean realized additive genomic relationship, prediction accuracy, genetic variance, mean genotypic value, inbreeding coefficient, and the frequency of QTL and marker alleles.

The genetic variance in each cycle was calculated as the variance among the genotypic values of the selection candidates. Prediction accuracy was measured by computing the correlation between the predicted genotypic values of the selection candidates and their true genotypic values.

We measured the LD between QTL and markers as such: for each and every polymorphic QTL in a given population (i.e. the training population or the selection candidates), we computed the correlation between that QTL and each and every polymorphic marker in the genome. We calculated persistence of LD phase by first measuring QTL-marker LD in the training population and in the selection candidates. QTL or markers that were not polymorphic in either of these populations were excluded. We then computed the correlation between the measures of QTL-marker LD in the training population and in the selection candidates. This metric, also known as the “correlation of r ,” evaluates whether patterns of QTL-marker LD are similar between two populations. High correlations of r indicate that QTL-marker LD phases are consistent, and presumably the predicted marker effects in one population would accurately represent the marker effects in the second population (de Roos et al., 2008; Toosi et al., 2010).

Additive relationships between lines in the simulation were measured with respect to the base population. Before initiating the simulations, a matrix \mathbf{P} was calculated as $2(p_i - 0.5)$, where p_i is the frequency of the second allele at locus i in the base population. Additionally, a normalization constant c was calculated as $2 \sum p_i(1 - p_i)$. Both calculations are described in VanRaden (2008). To compute additive relationships at any one cycle in the simulation, the

genotype matrices (including QTL) of the training population and selection candidates were combined into a matrix \mathbf{M} . The matrix \mathbf{P} was subtracted from \mathbf{M} to obtain matrix \mathbf{W} . We then calculated the relationship matrix as $\mathbf{A} = \frac{\mathbf{W}\mathbf{W}'}{c}$. This ensured that the relationship matrix was scaled to reflect the allele frequencies in the base population (VanRaden, 2008). We calculated the mean additive relationship as the mean value of the training population-selection candidate combinations. Inbreeding coefficients for each individual were also calculated from this matrix as the diagonal elements minus one.

All simulations were performed in R (version 3.3.1, R Core Team, 2018) using the packages *hypred* (version 0.5, Technow et al., 2014) and *rrBLUP* (version 4.4, Endelman, 2011). Each simulation experiment was repeated 250 times. The methods of updating the training population (i.e. “Top,” “Bottom,” “Random,” “CDmean,” “PEVmean,” “Tails,” and “No Change”) each constituted an independent experiment. With the two updating scenarios (i.e. “Window” and “Cumulative”), there were 14 different simulations.

Data Availability

Simulation scripts, starting marker genotypes, and summarized data are provided in the R package *GSSimTPUpdate*, available from the GitHub repository <https://github.com/UMN-BarleyOatSilphium/GSSimTPUpdate>. Included is a vignette on how to obtain the marker data from the T3 database.

RESULTS

Long-term prediction accuracy

Prediction accuracy (Fig. 1.3, Table S1) consistently decreased over cycles of selection for all methods of updating the training population and in both updating scenarios. Within and between scenarios, we observed differences among the update methods in the decay rate of prediction accuracy. A prominent observation was the precipitous decline in accuracy when not updating the training population (i.e. “No Change”). Early in breeding cycles, prediction accuracy for this method was similar to the remaining methods, but by cycle five had decayed beyond the remaining methods. As expected, identical trends were observed for “No Change” in both updating scenarios.

Among methods of actively updating the training population (i.e. excluding “No Change”), differences in prediction accuracy were observed in early cycles, but became increasingly similar in later cycles. The “Top” and “Tails” methods resulted in a non-significant, but noticeable accuracy advantage early on that persisted for several cycles (Fig. 1.3, Table S1). On the other hand, the “Bottom” method displayed a noticeable disadvantage that persisted for a similar length of time. The “Random,” “PEVmean,” and “CDmean” methods were highly comparable and yielded accuracies intermediate of the “Top” and “Bottom” methods. By cycle ten, the differences between active methods of updating were negligible. These patterns were observed in both the “Cumulative” and “Window” scenarios.

One noticeable difference between the trends in the “Cumulative” and “Window” scenarios was in the rate of prediction accuracy decay. Among the active methods of updating, the rate of prediction accuracy decay was slightly greater in the “Cumulative” scenario (Fig. 1.3A) compared to the “Window” scenario (Fig. 1.3B). By the fifteenth breeding cycle, the difference in these decay rates amounted to a difference in prediction accuracy of roughly 0.02 – 0.04.

Genetic variance and response to selection

Genetic variance among the selection candidates (Fig. 1.4A and 1.4B) similarly decreased across cycles for all training population update methods. For this variable, however, the rank among methods remained more consistent. That is, compared to the remaining update methods, the genetic variance in the “Top” and “Tails” methods was consistently less and the genetic variance in the “Bottom” method was consistently greater. The “Tails” method resulted in slightly higher genetic variance compared to the “Top” method, however this difference was never significant (95% confidence interval). Genetic variance across the “CDmean,” “PEVmean,” and “Random” methods was very similar within and between scenarios. Not updating the training population resulted in genetic variance similar to “CDmean,” “PEVmean,” and “Random” in early breeding cycles. After seven cycles, however, the loss of genetic variance was abated compared to remaining methods. By the end of the breeding timeline, the genetic variance for “No Change” was noticeably and significantly (95% confidence interval) higher than the remaining methods.

Overall, the mean genotypic value of the selection candidates (Fig. 1.4C and 1.4D) displayed a similar, but opposite pattern compared to the genetic variance. Updating the training population by the “Top” or “Tails” methods yielded an advantage in genotypic value, a trend that became more apparent in later breeding cycles. Conversely, the genotypic values under the “Bottom” method ranked lowest among the active updating methods. This disadvantage was often slight and non-significant, especially in the “Cumulative” scenario (Fig. 1.4C). As in the observations of genetic variance, the “CDmean,” “PEVmean,” and “Random” methods responded similarly. Most noticeable was the rapid plateau in genotypic value under the “No Change” method, particularly around the eighth breeding cycle. By the end of the breeding timeline, the “No Change” method appeared to have reached a limit, and although the trajectory of the remaining methods suggested further increases, their trends implied a limit as well (Fig.

1.4C and 1.4D). Curiously, the “Top” method was generally superior to the “Tails” method in the “Cumulative” scenario, however the opposite was true in the “Window” scenario. In both scenarios, the “Tails” method exhibited a trend suggesting that this method would eventually yield selection candidates with an average genotypic value superior to that of the “Top” method. The trends among the remaining training population update methods were similar in both updating scenarios.

Drivers of prediction accuracy

Average relationship between training population individuals and selection candidate individuals, as measured by marker information, varied among the update methods (Fig. 1.5A and 1.5B). As expected, the average relationship did not change in either updating scenario when the training population remained unaltered. Across both scenarios, the relationship generally remained highest under the “Top” method, lowest under the “Bottom” method, and intermediate under the “CDmean,” “PEVmean,” “Random,” and “Tails” methods. In the “Cumulative” scenario (Fig. 1.5A), actively updating the training population resulted in a linear increase in average relationship for all methods. Additionally, the different update methods, particularly “Top” and “Bottom,” displayed slight divergence, especially in later breeding cycles. The “Window” scenario (Fig. 1.5B) presented a more sigmoidal trend, eventually resulting in slight convergence in average relationship among active update methods. Interestingly, after cycle 12, the average relationship between the training population and the selection candidates in the “Tails” method remained greater than that in the “Top” method.

Generally, we observed a curvilinear increasing trend in the level of inbreeding (Fig. 1.5C and 1.5D). The “No Change” method performed similarly in the different updating scenarios, but differed markedly from the active updating methods. This method resulted in a more rapid increase in inbreeding, beginning after the fourth breeding cycle. By the end of the breeding timeline, the trend had not yet plateaued and suggested that inbreeding would

continue to increase. Considering the active updating methods, there were slight differences in inbreeding trends between the two updating scenarios. In the “Cumulative” scenario (Fig. 1.5C), these methods performed similarly, showing no significant differences. Inbreeding was slightly greater for these methods in this scenario than in the “Window” scenario (Fig. 1.5D). In this case, differences between the updating methods were more apparent. The “Top” method displayed noticeably lower levels of inbreeding, particularly after the eighth breeding cycle. Remaining methods performed similarly between each other.

We noticed consistent trends among methods of updating the training population in the rate of fixation of QTL (Fig. 1.5E and 1.5F). In both updating scenarios, the “Top” method maintain a higher number of fixed QTL across breeding cycles, followed by the “CDmean,” “PEVmean,” “Tails,” and “Random” methods, which performed similarly, followed by the “Bottom” and “No Change” methods, which also performed similarly. Additionally, we observed that roughly 10% of the QTL became fixed in cycle 1 of the breeding timeline, while by cycle 15 around 70% of the QTL were fixed. There were two slight, noteworthy differences in these trends between the updating scenarios. First, active updating methods generally displayed a higher proportion of fixed QTL in the “Window” scenario (Fig. 1.5E) than in the “Cumulative” scenario (Fig. 1.5F). Second, the degree of separation between the “Top” method and the “CDmean,” “PEVmean,” and “Random” methods appeared greater in the “Cumulative” scenario.

There were marked differences in the persistence of LD phase between the methods of updating the training population within and between the updating scenarios (Fig. 1.5G and 1.5H). Under the “Cumulative” scenario (Fig. 1.5G), persistence of phase for all update methods declined quickly in initial cycles, but reached equilibrium around the tenth cycle. The “Top” and “Tails” methods maintained the highest degree of persistence across breeding cycles, but the “Tails” method trended closer to the other active update methods by cycle twelve. Furthermore, the initial decay was much lower under the “Top” and “Tails” methods, and the

equilibrium point was higher than other methods. Persistence of phase under the “Bottom” method was initially much less than the other active update methods, and although it soon became similar to these methods, it still remained less. The remaining active update methods were quite similar in this scenario.

In comparison, actively updating the training population under the “Window” scenario (Fig. 1.5D) yielded increasing persistence of phase over the course of the breeding timeline. Each of these methods saw a small drop in persistence of phase initially, but after the fifth cycle values began to increase. Interestingly, none of these methods appeared to reach an equilibrium point. The disparity between update methods, especially “Top” and “Bottom,” was highly apparent under this scenario. Conversely, “CDmean,” “PEVmean,” and “Random” resulted in very similar levels of persistence of phase. Finally, the persistence of phase under the “Tails” method was initially intermediate of the “Top” method and the “CDmean,” “PEVmean,” and “Random” methods, however it eventually became more similar to the latter.

Expectedly, the “No Change” method resulted in identical trends in both updating scenarios. In the same way as prediction accuracy, we observed a precipitous, exponential decay in persistence of phase. The trend appeared to reach an equilibrium point at around the same breeding cycle as the active updating methods in the “Cumulative” scenario. However, this equilibrium point was much lower than the others.

DISCUSSION

Updating the training population can be simple and effective

We observed similar patterns in prediction accuracy (Fig. 1.3), mean genotypic value (Fig. 1.4C and 1.4D), and genetic variance (Fig. 1.4A and 1.4B) among active methods of updating the training population (i.e. excluding “No Change”). The high similarity between these methods suggests that simply including more recent data in the training population provides a marked advantage in improving the breeding population in the long-term. This is encouraging in a practical sense, as any phenotypic information generated on breeding lines, regardless of how they may have been selected, would probably be helpful in preventing severe long-term loss in prediction accuracy.

Although we only tested six active methods of updating the training population, we might expect that any method should outperform doing nothing. Over breeding cycles, including recent genotypic and phenotypic information in the training population helps to capture new LD generated by selection and drift (Hill and Robertson, 1968). Older training population lines will of course not provide any information on this new LD, however we may presume most or all selection candidates will share a proportion of this new LD as long as the parents of these lines are not unrelated. Therefore, even the selection candidates most distantly related to those chosen as parents will provide informative training data for the next cycle. In the long-term, we might expect a decrease in the relative importance of how selection candidates are chosen to add to the training population. Over continued cycles of selection in a closed population, parents will become increasingly related (Daetwyler et al., 2007), thus the pool of selection candidates will share a greater proportion of the new, informative LD.

Though it appears updating the training population is favorable regardless of method, it is worth pointing out differences in the methods we tested. The “Top” method achieved high prediction accuracy and high mean genotypic value across breeding cycles. These results are

not entirely surprising, since the candidates selected to update the training population were mostly those selected as parents for the next cycle (100 of 150). These additions to the training population will be highly related to the selection candidates in the next cycle, and will therefore provide the training population with the most useful information shared through genomic relationships and QTL-marker LD (Lorenz and Smith, 2015). Indeed, this is readily apparent in measures of relatedness between the training population and the selection candidates (Fig. 1.5A and 1.5B) and in measures of persistence of LD phase (Fig. 1.5C and 1.5D).

With this in mind, it is not surprising that the “Bottom” method delivers the lowest prediction accuracy (Fig. 1.3A and 1.3B) and lowest mean genotypic value (Fig. 1.4C and 1.4D), as zero lines added to the training population overlap with the selected parents. This lack of overlap would suggest that QTL-marker LD information in the training additions and that observed in the selection candidates will be in high disagreement. Indeed, we observe that this method produces training populations with the lowest average relationship to the selection candidates (Fig. 1.5A and 1.5B) and the lowest persistence of LD phase (Fig. 1.5G and 1.5H).

The “Tails” method, as a combination of the “Top” and “Bottom” method, offers some curious results. Though the prediction accuracy achieved from this method is, for the most part, not significantly different than that of the “Top” method, it is often higher, leading to low genetic variance (Fig. 1.4A and 1.4B) and high average genotypic value (Fig. 1.4C and 1.4D). This is in spite of the observation that under the “Tails” method, the average relationship between the training population and selection candidates (Fig. 1.5A and 1.5B) and persistence of LD phase (Fig. 1.5G and 1.5H) are roughly equal or lower than in the “Top” method. A possible explanation for this observation could be that this method produces training populations that satisfy different conditions for accurate genomewide predictions. First, 75 of the 150 training population additions overlap with the 100 selected parents. Just as in the “Top” method, these additions will be highly related to the selection candidates of the next cycle and contribute useful QTL-marker LD information. The other 75 additions will presumably be more unrelated

to these selection candidates, leading to the intermediate average relationship (Fig. 1.5A and 1.5B) and often lower persistence of LD phase (Fig. 1.5G and 1.5H). However, these training population additions may provide information for more reliable predictions. In a study where the training population was a subset of a larger population, Yu et al. (2016) found that individuals in the validation population (i.e. selection candidates) with the highest and lowest predicted genotypic values had the greatest upper bound for the reliability of those predictions (Karaman et al., 2016). It may be the case in our simulations that the training population additions in the “Tails” method had more reliably-predicted genotypic values. This reliability may have led to better identification of individuals that, when added to the training population, could provide information that more clearly differentiated the effects of QTL alleles, leading to more accurate predictions of marker effects. Thus, the “Tails” method may have taken advantage of both high relatedness and greater genotypic diversity in the training population.

The criterion-based updating methods (“CDmean” and “PEVmean”) performed very similarly to the “Random” method in prediction accuracy (Fig. 1.3A and 1.3B). This observation is generally in agreement with previous research (Akdemir et al., 2015; Isidro et al., 2015; Bustos-Korts et al., 2016a) and may be related to the size of the training population used in our simulations. In several examples in these studies, the prediction accuracy of a randomly selected training population was similar to that of a training population selected by the CD-mean or PEVmean criteria, particularly at larger sizes of the training population. While these investigations examined training populations ranging from 25 to 300 individuals, our simulations looked at much larger training populations, ranging from 764 to 2,864 individuals. It may be, then, that as the size of the training population becomes sufficiently large, the performance of the CDmean and PEVmean criteria becomes more similar to a random sampling. This, of course, does not suggest that these criteria have no use in selecting training populations. If these criteria are in fact superior in smaller training populations, they may be advantageous when performing genomewide selection on a trait that is expensive or low-throughput to phe-

notype.

It is worth addressing the continued loss in prediction accuracy in all updating methods and in both updating scenarios. This occurs even as two known components of prediction accuracy, persistence of LD phase and genomic relationship (de Roos et al., 2008; Toosi et al., 2010; Lorenz et al., 2011; Lorenz and Smith, 2015; Sallam et al., 2015) stabilize or increase. The primary reason for these observations is undoubtedly the reduction in heritability as genetic variance declines over cycles (Fig. 1.4A and 1.4B). Since residual variance remains constant, the phenotypic data measured on lines becomes increasingly uncorrelated with the true genotypic value (Bernardo and Yu, 2007; Bernardo, 2010). Thus, the data included in the training population will not capture the effects of QTL alleles, decreasing the accuracy of predicted marker effects. A second potential contributor is the fixation of marker loci over cycles. Since monomorphic markers are removed prior to model training, fewer markers will be used in later cycles. Indeed, by cycle 7, on average 55% of the original markers are used, and by cycle 15 this drops to 30% (data not shown). Though previous studies have stated the benefit of greater marker density (Combs and Bernardo, 2013), many others have noted diminishing returns (Lorenzana and Bernardo, 2009; Heffner et al., 2011; Lorenz et al., 2012). Reasonably high marker densities were maintained in our simulations, so this is likely not a strong driver of the decay in prediction accuracy.

The performance of the “Top” method suggests a simple procedure to optimize genomewide selection in an applied breeding program. Our results indicate that a breeder may prevent severe loss of prediction accuracy in recurrent selection by updating the training population to include information on lines that would be selected anyway. Ultimately, this method should be more cost effective than the others. A breeder would likely desire to evaluate selected parents in field trials, perhaps for variety development or to gather phenotypic data to accompany predicted genotypic values. The “Top” method provides an advantage here, as the number of additional lines to phenotype for updating the training population is minimal.

The breeder can use this information for dual purposes, using phenotypic data to build a more accurate training dataset while making informed decisions on potential variety selections.

Although the “Tails” method led to slightly greater prediction accuracy than the “Top” method, there are at least three reasons why it may not be the most practical method. First, the difference in prediction accuracy between these methods was generally not significant (Table S1). Second, the overlap between training population additions and candidates that would be prioritized for phenotyping by the breeder (i.e. parents and superior lines) is lower, and therefore, third, because of this lack of overlap, the breeder would expend costly resources on phenotyping lines that may not provide any utility outside of model training for genome-wide selection.

Encouragingly, empirical data in a barley breeding program supports the “Top” method in enhancing prediction accuracy. Over a few cycles of recurrent genome-wide selection for yield and deoxynivalenol content (a mycotoxin produced by the fungal pathogen *Fusarium graminearum* Schwabe.), Tiede (2017, in prep.) found that updating the training population improved prediction accuracy. Specifically, including data only on lines selected for favorable predicted genotypic values in previous cycles enhanced the prediction accuracy in subsequent cycles. This method was superior to a random selection of lines and was often superior to a selection based on criteria optimization.

Not updating the training population is unfavorable

It is quite apparent from our simulations that in the long-term, not updating the training population is highly unfavorable. Prediction accuracy decreases rapidly in this case (Fig. 1.3A and 1.3B), and as a consequence, response to selection also collapses, leading to the observed plateau in genotypic value (Fig. 1.4C and 1.4D). Here selection is acting on non-genetic noise, preventing the mean genetic value in the population from changing.

The genetic composition of the breeding populations underscores the negative conse-

quences of leaving the training population unaltered. Although genetic variance appears to be preserved in the long-term (Fig. 1.4A and 1.4B), considering the decrease in accuracy and the plateau in genotypic value, this may be due to a larger number of QTL that remain segregating. We do indeed observe this (Fig. 1.5E and 1.5F), but given the similarity in the number of fixed QTL under the “No Change” method and that under the remaining methods, we may also surmise that a greater proportion of QTL are becoming fixed for unfavorable alleles. We also observe alarming levels of inbreeding among the selection candidates when not updating the training population (Fig. 1.5C and 1.5D). This result is not surprising, since previous theory and simulations into genomewide selection show that more accurate predictions better capture the Mendelian sampling term (i.e. within-family variance), preventing high rates of inbreeding (Daetwyler et al., 2007; Jannink, 2010). Although higher inbreeding does not reduce genetic variance, it invariably will reduce the number of usable, polymorphic markers. Collectively, this suggests that continued genomewide selection without updating the training population will impose a lower selection limit on population improvement.

The results of our simulations indicate that severe consequences of not updating the training population were delayed until later cycles. Although prediction accuracy declines very rapidly (Fig. 1.3), mean genotypic value and genetic variance track closely with the other updating methods (Fig. 1.4). It is not until the fifth cycle or later that the impact of an unaltered training population is readily apparent. This can be encouraging in practical breeding scenarios. For instance, in a new breeding program, the stock of germplasm with phenotypic data may be low, and it may be several cycles before enough individual are tested to add to the training population. One may also consider a crop where the time between making a cross and gathering phenotypic data on the progeny is long. Several cycles of selection could be performed before data is available to update the training population. Our results suggest that the same training population could be used for a small number of cycles without serious detriment. A smaller and more recent training population may provide long-term advantages

We observed non-significant, but noticeable differences in prediction accuracy, mean genotypic value, and genetic variance between the “Cumulative” and “Window” updating scenarios. In the short-term, prediction accuracy was slightly greater under the “Cumulative” scenario for most of the active updating methods, particular the “Top” method (Fig. 1.3A). However, in the long-term, prediction accuracy was higher when the training population consisted of only more recent data (i.e. the “Window” scenario). Although the trends in genotypic value suggest that the “Cumulative” scenario is slightly advantageous in the short-term, the trend under the “Window” scenario suggested that additional gains may be greater (Fig. 1.4D). Indeed, given the slightly higher prediction accuracy observed at the end of the breeding timeline for this scenario, we would expect response to selection to be greater in the long-term (Bernardo, 2010).

In addition to the explanations provided earlier in the discussion, other factors may be responsible for these observations. Most notable are the differences between updating scenarios in genomic relationship (Fig. 1.5A and 1.5B) and persistence of LD phase (Fig. 1.5G and 1.5H). Retaining older training data results in lower average relationship between the training population and the selection candidates (Fig. 1.5A). This is not unexpected, since selection candidates in earlier cycles will be increasingly unrelated to those in later cycles. Maintaining a training population with more recent data results in higher average relationship and a higher rate of increasing relationship (Fig. 1.5B). This result corroborates previous research demonstrating higher prediction accuracy when retaining individuals in the training population that are more closely related to the selection candidates (Lorenz and Smith, 2015).

Perhaps most drastic are the differences in persistence of LD phase between updating scenarios. A training population with older data (i.e. “Cumulative”) results in decayed persistence of LD phase (Fig. 1.5G). Over cycles, recombination breaks down LD and training population additions capture new LD. Older training data does not reflect this new LD, decreasing the persistence of phase. The observed stabilization in Fig. 1.5C could be due to new

training data capturing as much LD as what is broken down by recombination. Evidence for this may be seen under the “Window” scenario (Fig. 1.5H), where persistence of LD phase increases when actively updating the training population. A training population of only recent data captures the new LD generated by recombination in the previous cycle, but without the uninformative LD present in older training data. In addition, it may be possible that recent training additions capture more of the informative new LD than what is lost through recombination, leading to the observed increase in persistence of phase.

Simulation considerations

It is important to address the limitations of our simulations, including assumptions that could be violated in a real-life breeding program. First, random mating may be unrealistic, and we might expect a breeder to impose a more sophisticated procedure for parent selection. For instance, mating pairs may be prioritized for complementation of favorable values of multiple traits. Additionally, an individual may be used as a parent over multiple breeding cycles, especially if observed phenotypic values agreed with the predicted genotypic values. More sophisticated methods of parental selection, such as those based on virtual bi-parental populations (Bernardo, 2014; Lian et al., 2015; Mohammadi et al., 2015), may be used. These non-random mating schemes could affect genetic variance or contribute to different patterns of LD, both of which would impact the accuracy of genomewide prediction. However, incorporating such nuances into our simulation would likely rest on additional assumptions and would be intractable to model. Random mating provides a simple approach, and given the recurrent selection scheme, it is a reasonable assumption. Our simulation also made the assumption that the breeding population was closed. This is obviously inaccurate in a practical program, as the exchange and incorporation of new germplasm is common. Realistically, we might expect prediction accuracy to decrease when adding germplasm from different breeding programs or subpopulations to the pool of selection candidates (Lorenz et al., 2012). In recurrent selec-

tion, however, the objective is to improve a population rapidly, so a closed population may be desirable (Bernardo, 2010).

Other assumptions may not reflect biological reality. First, our simulation forced QTL to be bi-allelic, but, as noted by Jannink (2010) and suggested in Buckler et al. (2009), many QTL may have multi-allelic genetic architecture. Second, we assumed the processes of mutation and crossover interference were absent, which is, of course, unrealistic.

Conclusions

In our simulation experiment of recurrent genomewide selection, we confirmed the need to update the training population over breeding cycles. Clearly, the LD between QTL and markers in the base population is decaying, likely as a result of recombination. When new data is not added to the training population, the change in LD is not captured, and prediction accuracy collapses. Among the tested methods of updating the training population, adding the lines predicted to have the greatest genotypic value (i.e. the “Top” method) is the most attractive. The desirability of this method stems not only from the resulting prediction accuracy and response to selection, but also from its simplicity and practicality. A breeder will undoubtedly desire to confirm the predictions of genotypic value with empirical phenotypic data, especially for the most promising lines or those selected to become parents. Updating the training population becomes simple, then, as this new data can be combined with previous training data. This method also facilitates updating the training model every cycle, likely the best option to capture the changes in LD as a result of recombination, selection, and drift. Nevertheless, our experiment leaves room for additional research, including fine-tuning the updating scenarios to choose the most informative training population from a pool of data. Additionally, optimizing other streams in the breeding program deserves research, including methods of selecting markers and parents. Long-term genomewide selection may benefit greatly from such investigations.

FIGURES

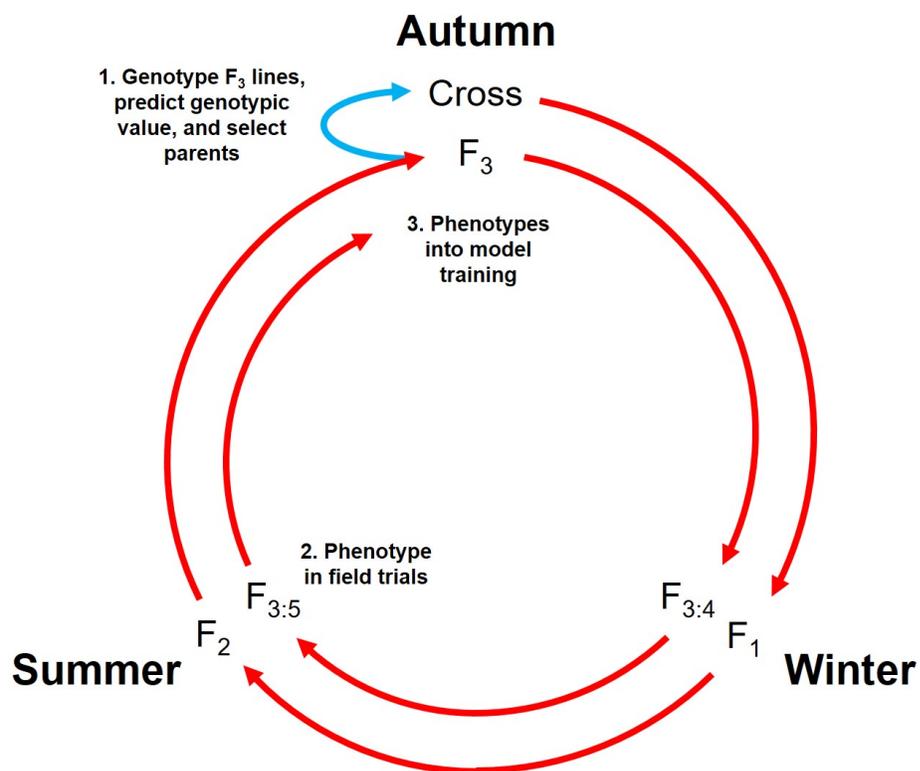


Figure 1.1: Realistically, a cycle of genomewide recurrent selection in barley may only be one year in length. Crosses are made in the autumn (year n) and progeny undergo single-seed descent through the following winter and summer. 1) At the F_3 generation during the next autumn (year $n + 1$), lines are genotyped and predicted genotypic value (PGVs) are determined using training data from the previous cycle. These predictions determine the lines to use as parents in the next cycle of crosses (blue arrow). 2) Predictions are also used to select lines to phenotype in the following summer (year $n + 2$). 3) This phenotypic information is then incorporated into the training data for the next cycle of predictions and crosses during the subsequent autumn.

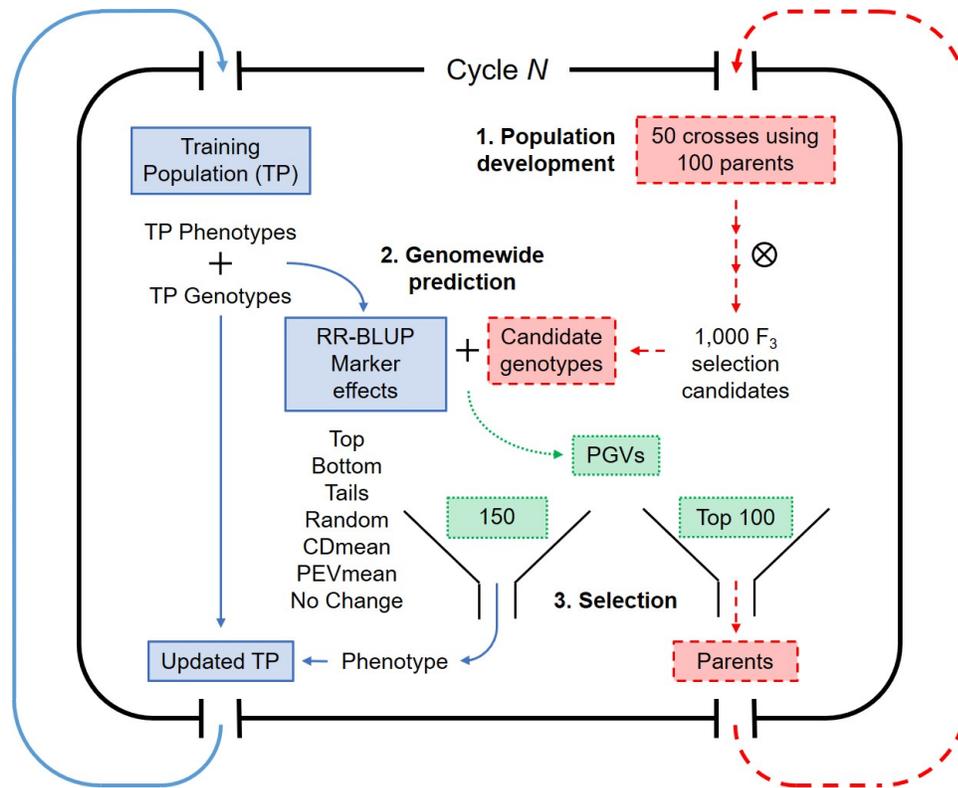


Figure 1.2: A single breeding cycle in our simulations may be broken down into two main streams. Blue (solid) indicates steps involving the training population, and red (dashed) indicates steps involving crossing and population development. Green (dotted) indicates the intermediate step of selection. 1) Fifty crosses are made using 100 randomly intermated parents from the previous cycle. Population development follows and 1,000 selection candidates are genotyped at the F₃ stage. Concurrently, marker effects are estimated using genotypic and phenotypic data from the training population (TP). 2) The predicted genotypic values of the selection candidates (PGVs) are used in decision-making. 3) The 100 selection candidates with the highest predicted genotypic values are selected as parents for the next cycle. Additionally, 150 selection candidates are selected based on the six different update methods. These candidates are “phenotyped”, and phenotypic and genotypic data are added to the pool of training data.

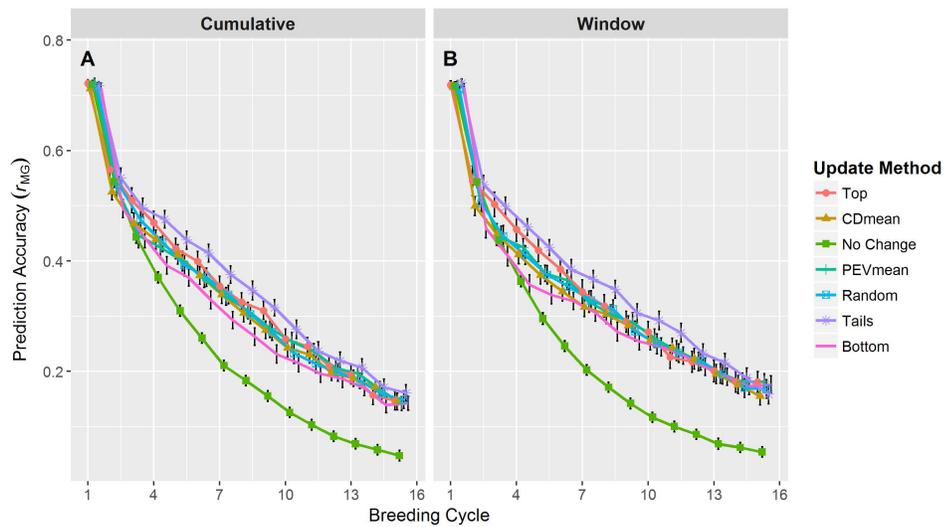


Figure 1.3: Prediction accuracy over breeding cycles of the simulation. Accuracy was measured as the correlation between the predicted and true genotypic values of the selection candidates. Point shapes delineate the different methods of updating the training population. Plots are separated into the Cumulative (**A**) and Window (**B**) updating scenarios. Average values are shown with 95% confidence intervals. To help reduce plot clutter, points for each update method are given a small, consistent jitter along the x-axis. Because the plotting jitter may accentuate small differences between updating methods, this data is also provided in Table S1.

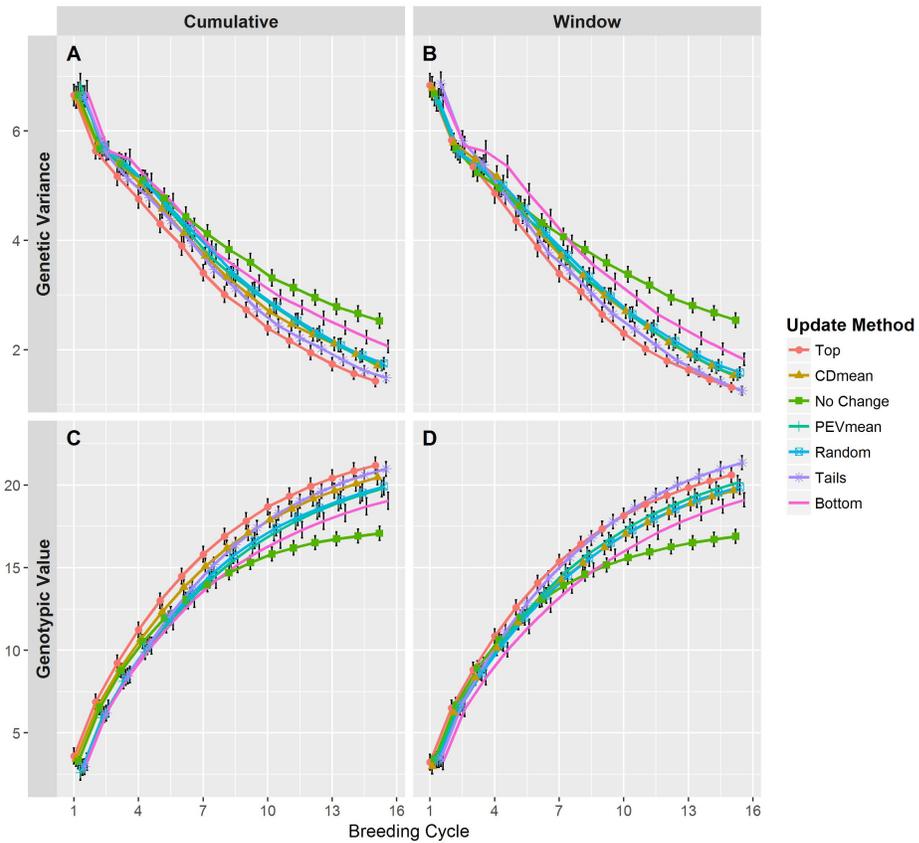


Figure 1.4: Genetic variance (**A** and **B**) and genotypic values (**C** and **D**) among the selection candidates over breeding cycles of the simulation. Point shapes delineate the different methods of updating the training population. Plots are separated into the Cumulative (**A** and **C**) and Window (**B** and **D**) updating scenarios. Average values are shown with 95% confidence intervals. To help reduce plot clutter, points for each update method are given a small, consistent jitter along the x-axis.

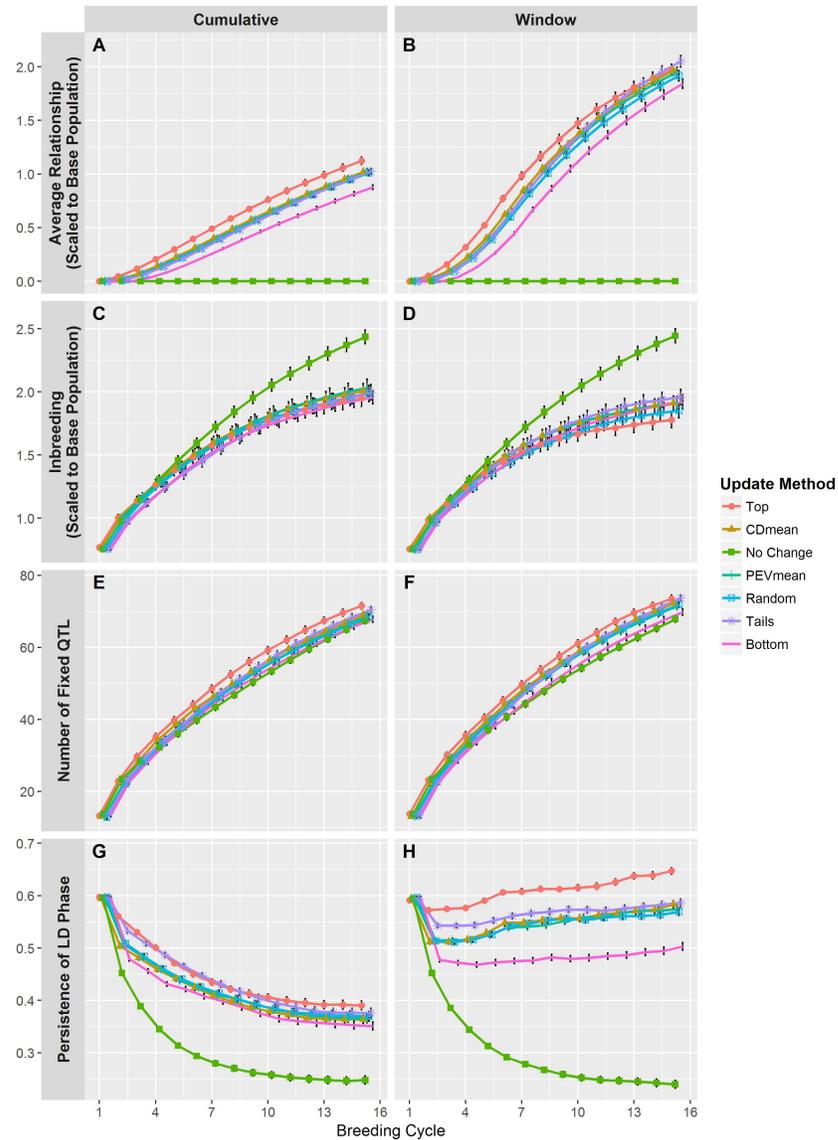


Figure 1.5: Other variables tracked over the course of the simulations. The average genomic relationship (A and B) was calculated between the training population and the selection candidates using marker genotypes. Relationships were scaled to reflect the allele frequencies in the base population. The level of inbreeding (C and D) was measured on the selection candidates and was derived from the relationship matrix described above. The number of QTL fixed for an allele (E and F) was measured in the selection candidates. Persistence of LD phase (G and H) was measured as the correlation of r between the training population and the selection candidates. Point shapes delineate the different methods of updating the training population. Plots are separated into the Cumulative (A, C, E, and G) and Window (B, D, F, and H) updating scenarios. Average values are shown with 95% confidence intervals. To help reduce plot clutter, points for each update method are given a small, consistent jitter along the x-axis.

Chapter 2: Strategies for subdividing environments in inter-generational genomewide selection

INTRODUCTION

Plant breeders routinely test germplasm in a target population of environments to identify broadly or locally superior genotypes. In such testing, breeders often encounter genotype-environment interactions (GEIs), or changes in relative genotypic performance across environments (Bernardo, 2010). In a plant breeding program, GEIs can be ignored as noise when selecting for genotypes that are broadly superior; alternatively, a breeder may select genotypes that are locally or regionally best by exploiting GEI.

Identifying superior genotypes across multiple environments may be accomplished by appropriate statistical models of GEI. Multiplicative models such as additive main effects and multiplicative interactions (AMMI) (Gauch and Zobel, 1988) or genotype main effects and genotype-by-environment interactions (GGE) (Yan et al., 2000) partition terms into genotypic and environment scores based on singular value decomposition. These models permit the visualization of GEI through biplots and the identification of mega-environments or clusters (Gauch and Zobel, 1997; Yan et al., 2000). Other models use factorial regression to identify the sensitivity of genotypes to environmental covariables. A popular version of these models is one proposed and modified by Yates and Cochran (1938), Finlay and Wilkinson (1963), and Eberhart and Russell (1966), which regresses the phenotypic observation of a genotype on the main effect of the environments. While this model is useful, environmental covariables independent of phenotypes (i.e. temperature, rainfall, soil conditions, etc.) are more desirable and would permit the prediction of observed genotypes in unobserved environments (reviewed by van Eeuwijk et al., 1996).

Just as environmental covariables can be used to make predictions for new environments, genetic covariables can be used to predict new genotypes. An application of this concept is genomewide selection (GWS), where a training population (TP) of genotyped and phenotyped individuals is used to predict the genotypic value of a population of new individuals (i.e. selec-

tion candidates) that have only been genotyped (Meuwissen et al., 2001). Although numerous studies point to the utility of GWS in plant breeding [e.g. Lorenzana and Bernardo (2009); Lorenz et al. (2012); Asoro et al. (2013); Rutkoski et al. (2015); Sallam et al. (2015)], standard models use phenotypic means to predict selection candidates, effectively ignoring GEI. Recently, several studies have proposed expanded GWS models to account for GEIs by borrowing information across environments (Burgueño et al., 2011, Burgueño et al. (2012); Lado et al., 2016; Malosetti et al., 2016; Saint Pierre et al., 2016; Cuevas et al., 2017). This method requires phenotypic observations in all environments, restricting the prediction space to observed environments.

An arguably important goal for GWS in plant breeding is the accurate prediction of new genotypes in new environments (Bustos-Korts et al., 2016b; Malosetti et al., 2016), though in practice there are constraints on the accuracy of these predictions. First, the population of new genotypes will likely be a generation or more removed from the TP (Lorenz and Smith, 2015; Sallam et al., 2015) and this decreased relatedness will negatively impact prediction accuracy (Habier et al., 2007). Second, predictions in new environments will require information on the relationship between these and already observed environments using explicit covariables (i.e. temperature, rainfall, soil fertility, geographic coordinates, etc.). Previous studies have examined the use of environmental covariables (ECs) in GWS (Heslot et al., 2014; Jarquín et al., 2014; Malosetti et al., 2016; Saint Pierre et al., 2016), noting improvements in prediction accuracy; however, the environments of most interest to a breeder are undoubtedly those in future years, the conditions of which are obviously difficult to predict (Allard and Bradshaw, 1964).

Instead of using the relationship between environments directly in prediction, this information could alternatively be used to assign mega-environments or clusters of environments. This assignment should ideally reduce GEI variance within clusters, converting it to genetic variance and leading to a greater response to selection (Atlin et al., 2000). For instance, clus-

ters may be identified statistically through AMMI or GGE models (Gauch and Zobel, 1997; Yan et al., 2000), followed by predictions within clusters (Lado et al., 2016); however, since this approach relies on phenotypic data, predictions of new environments are excluded. Instead, clusters could be defined using explicit information from ECs (Atlin et al., 2000; Piepho and Möhring, 2005; Bustos-Korts et al., 2016b) or through crop simulations based on these covariables (Löffler et al., 2005), allowing unobserved environments to be included.

In this study, we aimed to compare measures of calculating the distance between environments to construct sets of training data that maximize genomewide prediction accuracy. Our objectives were to assess the effect on prediction accuracy of (i) adding increasingly distant environments to a training set (TS), and (ii) using a TS of clustered environments. To meet these objectives, we relied on a multi-environment barley (*Hordeum vulgare* L.) dataset in which phenotype data on three quantitative traits were collected (Neyhart et al., 2019). An important advantage of this dataset is that a distinct TP and pool of selection candidates, or validation population (VP), were phenotyped together. The VP is one generation of recombination removed from the TP, which is typical of many breeding programs and permits parent-offspring validation of predictions (Lorenz and Smith, 2015). This is a unique advantage when assessing multi-environment genomewide predictions, where many studies have hitherto used a single population or generation.

METHODS AND MATERIALS

Phenotype and genotype data

We assembled a panel of 233 two-row spring barley breeding lines, including a TP of size 183 and a VP of size 50. Collection of phenotypic and genomewide marker data is described elsewhere (Neyhart et al., 2019); however, relevant details are included below. The panel was phenotyped (partially or wholly) for grain yield (kg ha^{-1}), heading date (days from planting), and plant height (cm) in 44 location-year environments between 2015 and 2017. Given our interest in studying the use of explicit covariables to group environments, we excluded data from five environments in which trials were conducted under irrigated conditions. We further restricted our dataset to environment-trait combinations in which data on both the TP and VP were collected and the heritability was greater than 0.1. The total number of retained environments was 23 for grain yield, 26 for heading date, and 27 for plant height (29 unique environments total) (Fig. 2.1).

Genomewide marker data was generated using genotyping-by-sequencing (Elshire et al., 2011). Using an in-house pipeline (Neyhart et al., 2019), we initially discovered 235,216 single nucleotide polymorphisms (SNPs), which were subsequently filtered on a minimum minor allele frequency of 0.05, maximum missing data proportion of 80%, and a maximum adjacent marker linkage disequilibrium (r^2) of 0.95. Additionally, lines with more than 80% missing marker data were removed. The final marker dataset included 6,361 SNPs, 175 TP lines, and 48 VP lines. Marker genotypes were imputed using the expectation maximization algorithm implemented in the rrBLUP R package (Endelman, 2011).

We used a two-step procedure for analyzing phenotypic data (Piepho et al., 2012b). In the first step, adjusted genotype means were calculated for each trait and environment, described in Neyhart et al. (2019). In the second step, phenotypic data were analyzed across all

environments using the following random effects models:

$$y_{ijk} = \mu + g_i + l_j + a_k + (la)_{jk} + (gl)_{ij} + (ga)_{ik} + (gla)_{ijk} + \epsilon_{ijk}, \quad (2.1)$$

where y_{ijk} is the mean of the i th genotype in the j th location and k th year (calculated from the first step), g_i is the effect of the i th genotype, l_j the effect of the j th location, a_k the effect of the k th year, $(la)_{jk}$, $(gl)_{ij}$, $(ga)_{ik}$, and $(gla)_{ijk}$ are the two- and three-way random interactions between genotype, location, and year, and ϵ_{ijk} is the associated error. To analyze random location-year environments, we also fitted the model:

$$y_{ij} = \mu + g_i + e_j + (ge)_{ij} + \epsilon_{ij}, \quad (2.2)$$

where y_{ij} is the mean of the i th genotype in the j th environment, e_j is the random effect of the j th environment, and $(ge)_{ij}$ is the random effect of genotype-environment interaction. We assumed all random effects were normally distributed with mean zero and proper variance. Residuals were assumed distributed such that $\epsilon_{ij} \sim N(0, \sigma_e^2 \mathbf{R})$, where σ_e^2 is the residual variance and \mathbf{R} is a diagonal matrix with elements equal to the inverse of the variances of the adjusted means obtained in the first step. All models were fitted on a per-trait basis, and the significance of variance components was tested using a likelihood ratio test. Broad-sense heritability across all environments was computed on an ad hoc basis (Holland et al., 2003).

Characterizing genotype-environment interaction

We analyzed GEI using variance components, correlations between environments, and AMMI models. These analyses impacted downstream clustering of environments (see below), so to reflect realistic breeding conditions, we restricted the dataset to only include lines in the TP. [Results did not change appreciably when using the entire dataset (Fig. S2.8, Table S2.4).] First, we calculated the proportion of GEI variance due to genetic variance heterogeneity (V) as

$V = var(\sigma_{gj})$, where σ_{gj} was the estimated genetic standard deviation in the j th environment (Li et al., 2018). The contribution of lack of genetic correlation (L) to GEI variance (σ_{ge}^2) was calculated as $L = \sigma_{ge}^2 - V$ (Cooper and DeLacy, 1994).

Second, we determined the pairwise genetic correlation between environments using genotype-environment means (y_{ij}). Simple genetic correlations were calculated using the Pearson correlation coefficient of genotypes shared between a given pair of environments. Relatedly, we calculated the phenotypic distance ($PD_{jj'}$) between environments as

$$PD_{jj'} = (n_{jj'})^{-1} \sum_{i=1}^{n_{jj'}} \left(\frac{y_{ij} - \mu_j}{s_j} - \frac{y_{ij'} - \mu_{j'}}{s_{j'}} \right), \quad (2.3)$$

where j and j' index two different environments, $n_{jj'}$ is the number of genotypes common to environments j and j' , y_{i*} is the mean of the i th genotype in an environment, μ_* is the mean of an environment, and s_* is the observed standard deviation among genotype means in an environment (Ouyang et al., 1995). This measurement is related to the genetic correlation between environments (Bernardo, 2010), where a distance near the minimum of 0 indicates little crossover GEI (high correlation), while a distance near the maximum of 4 implies abundant crossover GEI (low correlation). Similarly, the distance between locations was calculated using genotype-location means.

Third, we fitted AMMI models to assess the environmental component of GEI. To create a fully balanced two-way table of genotype-environment observations, we fitted a mixed model to predict missing phenotypic values [50 (1.1%) for heading date and plant height, 65 (1.6%) for grain yield]. As in Lado et al. (2016), the Kronecker product between the realized additive relationship matrix (VanRaden, 2008) and the estimated environmental correlation matrix (described above) was used to model the covariance of genotype-environment combina-

tions. Fitted values from this model were used as responses in the fixed-effect AMMI model:

$$y_{ij} = \mu + g_i + e_j + \sum_{m=1}^M \lambda_m \gamma_{im} \delta_{jm} + \epsilon_{ij}, \quad (2.4)$$

where λ_m is the singular value of the m th principal component, γ_{im} is the eigenvector value of the i th genotype and m th component, and δ_{jm} is the eigenvector value of the j th environment and m th component (Gauch, 2013). Interaction principal component axes (IPCA) were calculated for genotypes and environments as $\lambda_m^{0.5} \gamma_{im}$ and $\lambda_m^{0.5} \delta_{jm}$, respectively. Significant principal components were determined using a parametric bootstrapping procedure (Forkman and Piepho, 2014). We then used biplots to visually analyze spatial patterns in the environments.

Environmental covariables

We queried soil and weather data from publicly available databases to serve as ECs. For each trial, weather data were obtained from the National Oceanic and Atmospheric Administration (NOAA) Global Historical Climatology Network. Daily minimum and maximum temperature and precipitation readings were obtained in each of the ten years preceding a trial year (e.g. 2005 - 2014 for a 2015 trial) in order to characterize the intermediate-term climate history of a location. The NOAA weather station closest to the trial location was queried; however, if a station was missing more than 20% of the data during the 10-year timeframe, the next-closest weather station was used. The temporal range of weather observations within a given year was constrained to the average planting and harvest dates of that trial location. We summarized daily weather observations into ECs defined by O'Donnell and Ignizio (2012), either on an annual basis or in 30-day intervals beginning with the average planting date for a location (Table S2.3).

Soil characteristic data were obtained from the United States Department of Agriculture

- Natural Resources Conservation Service (Soil Survey Staff, National Resource Conservation Service, 2018) for trials in the United States and from the Agriculture and Agri-Food Canada - National Soil DataBase (Soil Landscapes of Canada Working Group, 2010) for trials in Canada. Measures of percent silt, sand, and clay, as well as soil pH and organic matter were obtained for each soil horizon within the most abundant soil type in each trial location. We calculated the mean of each soil variable in the topsoil or subsoil, weighted by the proportion of each soil horizon in each layer. The topsoil layer was defined as 0 - 20 cm below the surface and >20 cm below the surface, respectively (Anderson et al., 2016). Soil data was combined with the ten-year weather data to create two-way matrices of environments and covariables. In total, 61 ECs were created, which are listed and defined in Table S2.3.

Environmental distance measures and clustering

Geographic distance, phenotypes, and ECs were used to determine environmental relationships and clusters; we refer to these different sources of data as “distance measures” (Table 2.1). Distance measures were used as response variables in model-based clustering, implemented in the `mclust` package in R (Scrucca et al., 2016), to determine the appropriate number of clusters and their component environments (Dawson et al., 2013). Models were restricted to exclude clusters with a single environment. For the great circle distance measure, clustering was performed on the latitude and longitude coordinates of locations, while for phenotypic distance measurements (between environments or locations), we performed clustering on the first two principle coordinates of $PD_{jj'}$ values calculated in Eq. 2.3. The fitted values from AMMI1 models (1 interaction principal component) were used to create clusters by identifying environments with shared winning genotypes (Gauch and Zobel, 1997; Gauch, 2013). We explored three methods of selecting ECs to create clusters: i) simply including all ECs; ii) identifying ECs that were correlated with the environmental effect (e_j) from Eq. (Li et al., 2018); and iii) identifying ECs [scaled to mean zero and squared length unit variance (van Eeuwijk and

Elgersma, 1993)] that were correlated with the environmental IPCA score ($\lambda_m^{0.5}\delta_{jm}$) from Eq. 2.4. For the latter two methods, ECs were first ranked in descending order of the absolute estimated correlation coefficient; then, forward stepwise regression was used to add ECs, in this order, until the most appropriate model was determined. Here, ECs are chosen that reflect the environmental component of linear reaction norm models or the repeatable environmental portion of GEI (Bernardo, 2010).

We analyzed phenotypic data after clustering using a random effects model modified from Eq. 2.1 (Atlin et al., 2000):

$$y_{ijkm} = \mu + g_i + c_m + a_k + l_{j(m)} + (la)_{ij(m)} + (gc)_{im} + (ga)_{ik} + (gl)_{ij(m)} + (gac)_{ikm} + (gla)_{ijk(m)} + \epsilon_{ijkm}, \quad (2.5)$$

where c_m was the effect of the m th cluster, $(gc)_{im}$ the interaction of the i th genotype and m th cluster, $(gac)_{ikm}$ the interaction of the i th genotype, k th year, and m th cluster, and other terms as defined in Eq. 2.1. Subscripts in parentheses denote nesting (e.g. $l_{j(m)}$ is the effect of the j th location nested in the m th cluster). Again, we assumed that all random effects were normally distributed with zero mean and proper variance. To assess the effectiveness of clustering, we calculated the heritability of selection without clustering (H_1) and within clusters (H_2) using the following equations (Yan, 2016):

$$H_1 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{gc}^2}{n_c} + \frac{\sigma_{gl(c)}^2}{n_l} + \frac{\sigma_{ga}^2}{n_a} + \frac{\sigma_{gac}^2}{n_c n_a} + \frac{\sigma_{gla(c)}^2}{n_l n_a} + \frac{\sigma_\epsilon^2}{n_a n_l n_r}}, \quad (2.6)$$

$$H_2 = \frac{\sigma_g^2 + \sigma_{gc}^2}{\sigma_g^2 + \sigma_{gc}^2 + n_c \left(\frac{\sigma_{gl(c)}^2}{n_l} + \frac{\sigma_{gla(c)}^2}{n_l n_a} + \frac{\sigma_\epsilon^2}{n_a n_l n_r} \right) + \frac{\sigma_{ga}^2}{n_a} + \frac{\sigma_{gac}^2}{n_c n_a}}, \quad (2.7)$$

where n_c , n_l , n_a , and n_r are the number of clusters, locations, year, and replications, respectively. Generally, if $H_2 > H_1$, there is evidence supporting the assignment of environments into the respective clusters, as the response to selection will be higher within clusters than

across all environments

The distance measures were also used to rank environments relative to each other. The phenotypic distance values ($PD_{jj'}$) were used as is, and the great circle distance between environments was determined by latitude and longitude coordinates. The environmental main effect and IPCA1 scores were used to calculate a Euclidean distance between environments for AMMI. Finally, environmental kinship matrices were calculated using the different ECs subsets and the “GE-KE” model equation outlined by Malosetti et al. (2016).

Genomewide predictions

We assessed the usefulness of environmental distance measures by comparing prediction accuracies when i) using data from all available environments and ii) using a designated subset of training data environments. Prediction accuracy was assessed via cross-validation (CV) and parent-offspring validation (POV). We tested two different validation schemes, described in Jarquin et al. (2017): predictions of observed genotypes in unobserved environments (CV0 and POV0); and predictions of unobserved genotypes in unobserved environments (CV00 and POV00). Each instance of CV (i.e. for each validation environment, distance measure, and TS described below) relied on 25 independent realizations of five-fold training-test partitions. For each partition, predictions of genotypes in the k th fold were generated using data from the remaining four folds, and predictive ability was measured as the correlation between the observed phenotypic values and the predicted genotypic values of the entire TP after all folds were predicted. In the POV schemes, the entire TP was used to predict the genotypic values of the VP, and predictive ability was measured as above. In all cases, prediction accuracy was calculated by dividing the predictive ability by the square root of the estimated heritability within the validation environment for which predictions were generated (Dekkers, 2007).

Training sets for genomewide predictions were informed by the ranking or clustering of environments based on distance measures (Fig. 2.2). First, for each validation environment and

distance measure, we constructed training sets by incrementally adding TP data from environments, one-at-a-time, in order of increasing distance to that validation environment (Fig. 2.2A). At each step, the TS was used to assess predictions according to the CV and POV schemes described above. In addition to the distance measures, for each validation environment, we generated 25 random environment rankings as a comparison. Estimates of prediction accuracy were averaged over validation environments. Second, we assessed prediction accuracy within assigned clustered of environments. For each validation environment within a cluster defined by a distance measure, we formed training sets composed only of the remaining environments in that cluster (Fig. 2.2B). Prediction schemes described above were again tested, and accuracy results were analyzed using a mixed model that accounted for the fixed effects of distance measure, prediction scheme (CV or POV), and their interaction, and the random effects of cluster nested within distance measure, validation environment nested within cluster, and N_{TE} nested within cluster. To test whether the set of environments assigned to a cluster were informative, we sampled 25 sets of randomly selected environments (in number equivalent to the size of a cluster) to form a TS. This was repeated for each validation environment and distance measure. We measured the advantage of informative clusters by subtracting the prediction accuracy of random sets of training environments from that of the informative set of training environments; results were analyzed using the same mixed model described above, with the exception that the random effect of N_{TE} nested within cluster was removed from the model.

For all predictions, we used the following main-effect genomic best linear unbiased prediction model:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad (2.8)$$

where \mathbf{y} is a vector of genotype means in a set of environments (i.e. means from a model accounting for environment) (Lado et al., 2016), $\mathbf{1}$ is a matrix of 1s, μ is the grand mean, \mathbf{Z} is an incidence matrix relating \mathbf{y} to the vector of random genotypic values, \mathbf{u} , and \mathbf{e} is a vector

of residuals. We assumed genotypic values were distributed such that $\mathbf{u} \sim N(0, \sigma_g^2 \mathbf{G})$, where \mathbf{G} was the realized additive relationship matrix.

The size of our dataset allowed us to compare the results of using all available data versus a subset that might reasonably be used for time-structured predictions. We considered two dataset scenarios: a complete set that was used to predict each environment singly (termed “leave-one-environment-out”, LOEO), and a temporally structured set that included only observations from 2015 and 2016 (termed “time-forward”, TF). In the latter scenario, data were used to make predictions of 2017 environments. Relevant data for each scenario were used in the analyses described above, and the same genomewide prediction schemes (i.e. ranked environments and clustering) were tested. These scenarios permit a comparison of outcomes from using an ideal dataset versus one that would theoretically be available to a breeder looking to make predictions of future environments.

Data availability

The genotypic and phenotypic data used in this study are publicly available from the Triticeae Toolbox (<https://triticeaetoolbox.org/barley/>). Code to replicate the analyses described above and for generating figures is available at the GitHub repository https://github.com/UMN-BarleyOatSilphium/S2MET_Predictions. All analyses were performed in R (v. 3.5.1; R Core Team, 2018).

RESULTS

Analysis of phenotypic data and genotype-environment interactions

Our dataset was highly balanced, with more than 98% of possible genotype-environment combinations observed per trait. The mean (and range) of broad-sense heritability estimates within environments was 0.54 (0.18, 0.86) for grain yield, 0.85 (0.56, 0.98) for heading date, and 0.53 (0.11, 0.88) for plant height. The broad-sense heritability of each trait across all environments was 0.66 for grain yield, 0.96 for heading date, and 0.91 for plant height. Most variance components estimated in Eq. 2.1 were significant ($P < 0.05$; likelihood ratio test; Table 2.4, Table S2.4), with the exceptions of variance due to year (never significant), location (not significant for heading date), and genotype-year interaction (rarely significant). Location variance explained most of the phenotypic variance for grain yield (67%) and plant height (56%), and little of that for heading date (5.4%), which was mostly explained by location-year (i.e. environmental) variation (61%). Genetic variance account for 2.5% of the phenotypic variance of grain yield, 20% for heading date, and 5.1% for plant height. Most genetic interaction variance originated from genotype-location-year variation, and genotype-location variance was always greater than genotype-year variance (Table 2.4, Table S2.4). After decomposing the genotype-location-year variation into two different sources, we found a lack of genetic correlations was the primary source of GEI, accounting for 82% of this variance for heading date, 90% for plant height, and 93% for grain yield.

The mean (and range) of pairwise phenotypic correlations between environments was 0.26 (-0.17, 0.62) for grain yield, 0.71 (0.37, 0.89) for heading date, and 0.33 (-0.024, 0.62) for plant height. A visual assessment of genetic correlations, in the form of heatmaps (Fig. S2.9), revealed different patterns among traits. The correlations for heading date were typically high, and no discernable clusters were apparent in this initial visualization. In contrast, correlations between environments for grain yield were typically low, but environments tended to group

loosely together based on location. For plant height, year appeared to determine groupings of highly correlated environments (Fig. S2.9).

Nearly all principal components were identified as significant in our fitted AMMI models. We report only the results for the first interaction principal component (i.e. AMMI1), as it was used for further analysis and environmental clustering (Gauch, 2013). This component explained 51% of GEI variance for grain yield, 22% for heading date, and 54% for plant height (Fig. 2.3). Visualization of the AMMI model via biplots confirmed the relative contributions of genotypic and environmental variance to the total phenotypic variation (Fig. 2.3). For grain yield and plant height, there was little variation in the genotypic main effect and IPCA1 score relative to environments; for heading date, there was more variation in main effects but little variation in the IPCA1 score. In contrast, environments were arranged in a larger space, highlighting the relatively larger environmental variance, but also indicating that environmental IPCA1 scores accounted for more of the genotype-environment interaction effect (as opposed to the genotypic IPCA1 score).

Environmental characterization and clustering

Twenty-one ECs were determined to be associated with the environmental mean. This ranged from 2 (heading date) to 6 (plant height) under the leave-one-environment-out (LOEO) scenario, and from 3 (grain yield) to 13 (plant height) under the time-forward (TF) scenario. Similarly, 21 ECs were associated with the environmental IPCA score, and this ranged from 1 (plant height) to 2 (grain yield / heading date) under LOEO and from 2 (plant height) to 14 (heading date) under TF (Table S2.5). Some ECs were common to a pair or all three traits within a scenario and distance measure. For instance, latitude and maximum precipitation in a growing interval were associated with the environmental mean of at least 2 traits, while precipitation in the second and fourth growing intervals was associated with the IPCA score for 2 traits. Generally, associated ECs were unique to a scenario-distance measure-trait combination.

Phenotype-based distance measures (i.e. PD, LocPD, and AMMI) resulted in relatively fewer clusters for each trait, while distance measures defined by ECs (i.e. All-EC, Mean-EC, and IPCA-EC) generally led to a higher number of clusters (Fig. S2.9). Fewer clusters were assigned when grouping environment by shared winning genotypes (i.e. AMMI) (Gauch and Zobel, 1997; Gauch, 2013), with two clusters designated for all traits under the leave-one-environment-out scenario (and grain yield under the time-forward scenario), and only a single cluster designated for heading date and plant height under the time-forward scenario. Distance measures that emphasized attributes common to location (i.e. All-EC, Mean-EC, IPCA-EC, LocPD, and GCD) appeared to have greater overlap (defined as the proportion environment pairs designated to the same cluster) with each other, and clusters assigned by PD overlapped strongly with those assigned by AMMI (Fig. S2.9). Interestingly, clusters assigned by the AMMI measure overlapped more with those assigned by the IPCA-EC measure than any of the other EC or naïve distance measures (i.e. GCD, All-EC, Mean-EC).

Genomewide predictions

Overall, prediction accuracies under CV or POV were moderate to high when predicting a single environment using all available data (Fig. S2.12). Generally, CV prediction accuracies were higher than POV accuracies, and schemes that relied on information from previously observed genotypes (CV0 and POV0) usually yielded more accurate predictions than schemes targeting totally unobserved genotypes (CV00 and POV00). Additionally, predictions under the LOEO scenario were often numerically, but not significantly ($P > 0.05$, *t*-test), more accurate than corresponding prediction in the TF scenario. Differences in accuracy between prediction schemes were highly trait-dependent. The differences between CV accuracies and POV accuracies were marginal when considering grain yield, but differences were significant when predicting heading date. Notably, the POV predictions tended to be more accurate than CV predictions for plant height (Fig. S2.12).

When incrementally adding environments to a TS in order of increasing distance to a validation environment, the patterns of prediction accuracy (r_{MG}) were dependent on the trait, prediction scheme, and scenario. Under the LOEO scenario (Fig. 2.4), increasing the number of TS environments (N_{TE}) generally increased r_{MG} . This was true up until $N_{TE} = 5$, after which r_{MG} reached a plateau that roughly equaled the accuracy when using data from all environments. Notably, we observed a decline in r_{MG} with $N_{TE} > 5$ under CV schemes and the AMMI or PD distance measures. For instance, with the PD distance measure, increasing N_{TE} beyond 5 led to a loss in r_{MG} of up to 0.16 (29%) for grain yield and 0.10 (20%) for plant height. Under POV schemes, the AMMI and PD distance measures led to slightly higher r_{MG} , but difference among distance measures were marginal. When training sets were built by adding environments randomly, r_{MG} increased constantly, eventually reaching the accuracy when using all environments (Fig. 2.4). Aside from predictions under the POV00 scheme, the trend in r_{MG} for all distance measures was greater than that for random. Under the TF scenario (Fig. 2.5), increasing N_{TE} almost always led to a continuous gain in r_{MG} , reaching a peak only when data from all environment were used. (The exceptions were CV00 for grain yield and plant height, where the patterns more closely resembled those under LOEO.) We did not observe much difference between distance measures, though LocPD appeared to yield a slightly higher r_{MG} than others. Increasing N_{TE} by ordering environments randomly always led to continued gains in r_{MG} .

Summaries of r_{MG} when clustering environments are displayed in Fig. 2.6 and 2.7. Under the LOEO scenario (Fig. 2.6), the utility of clustering with each distance measure was dependent on the trait and prediction scheme. Aside from some notable exceptions, clustering never led to noticeably greater r_{MG} over using all available data without clustering (Fig. 2.6A). Predictions of grain yield and plant height assessed using CV schemes, distance measures that relied on environment-specific phenotypes (i.e. AMMI and PD) yielded a small (< 0.1) increase in r_{MG} over using all data; however, this gain disappeared under POV schemes. These distance

measure also appeared to create the most informative clusters, measured as the difference in r_{MG} compared with a random TS with an equal number of environments (Fig. 2.6B). Indeed, the AMMI and PD measures led to an advantage in r_{MG} of up to 0.3 for grain yield and almost 0.2 for plant height. For heading date, clustering almost universally led to lower r_{MG} than using all data, and this pattern did not change under the CV or POV prediction schemes (Fig. 6A). Distance measures for this trait that relied on geographic proximity or ECs (GCD, All-EC, Mean-EC, and IPCA-EC) led to slightly higher r_{MG} than phenotype-based measures. Clusters formed from these measures also tended to be more informative than random sets of training environments (Fig. 2.6B), though the advantage in r_{MG} was miniscule (< 0.1). Under the TF scenario (Fig. 2.7), the average r_{MG} when clustering was never significantly different than that when using all available data (Fig. 2.7A), and there was often a loss in r_{MG} when clustering. Unsurprisingly, distance measures in this scenario never created clusters that were appreciably more informative than random sets of environments (Fig. 2.7B), particularly under the POV scheme.

DISCUSSION

We compared the impact on prediction accuracy of different measures of grouping environments for creating environment-dependent genomewide prediction training sets. We attempted to recreate conditions likely to be encountered by a breeder by i) using a separate TP and VP from related, but distinct, generations, and ii) including measurements of environmental relatedness that assume the target environments are unobserved. Additionally, we highlighted two scenarios that compare an ideal dataset with one restricted to time-forward predictions. This study therefore takes a novel approach to answering questions of multi-environment genomewide prediction, one that may be more relatable to a practical breeding program.

Characterization of genotype-environment interactions and environmental relationships

Understanding the magnitude and cause of GEI can inform decisions about how to manage it. We found that variance attributable to genotype-location-year (σ_{gla}^2) interaction contributed meaningfully to the phenotypic variance of all traits (Table 2.4), and patterns observed within the TP were generally consistent with those observed within the VP (Fig. S2.8). Though σ_{gla}^2 was significant for all traits, it contributed differently to each. For instance, the ratio between the more readily selectable genetic variance (σ_g^2) and σ_{gla}^2 was highest for heading date and lowest for grain yield. Additionally, the fitted AMMI models revealed that the first interaction component explained a majority of GEI variation for grain yield and plant height, and little of that for heading date (Fig. 2.3). These results are consistent with the established understanding that heading date is relatively simple trait and grain yield a relatively complex trait, with plant height occupying some intermediate level of complexity (Hockett and Nilan, 1985).

By itself, the presence of GEI does not indicate how it may be managed. Such interactions may be the result of genetic variance heterogeneity across environments or a lack of genetic correlation between environments (Cooper and DeLacy, 1994). Lack of genetic cor-

relation is arguably more severe, since it implies cross-over interactions that change the rank of genotypes between environments (Bernardo, 2010). For all traits, the majority of variation attributable to GEI was due to this phenomenon. This has two implications for GWS: first, it indicates that variance homogeneity can reasonably be assumed, allowing for simpler prediction models; and second, it suggests that accuracy should be higher when making predictions between highly correlated environments.

The tested measures of calculating environmental distance aimed to take advantage of different sources of data on the environments. Measures that relied on phenotypic data (AMMI, PD, and LocPD) are arguably more informative for a breeder, but do not necessarily permit predictions of unobserved environments; meanwhile, measures that used explicit information, such as geographic location or ECs (GCD, All-EC, Mean-EC, and IPCA-EC), allow new environments to be classified, but may sacrifice more useful phenotypic information. We attempted to resolve that disparity by using measures incorporating ECs that were highly correlated with the environmental mean or interaction scores (Mean-EC and IPCA-EC, respectively). Both analyses are similar to factorial regression models used to identify the explicit variables that explain environmental and GEI effects (reviewed by van Eeuwijk et al., 1996), yet such models have generally not been used to cluster environments.

There are several breeding advantages and limitations to our ad hoc approach of clustering environments using ECs. First, we opted to use both soil conditions and a ten-year moving average of weather observations, which place greater emphasis on the characteristics of location, while allowing some variation due to year. Given the larger contribution of location and genotype-location interaction to total phenotypic variance of all traits (Table 2.4, Fig. S2.8), this emphasis may be justified, but we obviously lose information on year-specific causal factors. If years are considering uncorrelated, random realizations, it may be reasonable to ignore this information; realistically, we might expect a correlation between years in the short-term (e.g. increasing volatility as a result of climate change), and our approach recognizes this while

assigning more weight to location-specific information. Second, we identified relevant ECs and clustered environments independently for each trait. While it is reasonable to assume that traits will respond differently to environmental stimuli, breeders select for many traits simultaneously, and the deployment of cultivars to certain growing regions is generally not based on a single trait. We note, however, that the relevant ECs for each of our focal traits exhibit some overlap (Table S2.5), so it may be possible to select consensus ECs to form clusters for several traits.

Adding data from increasingly distant environments may reduce accuracy

To determine the impact on prediction accuracy of increasingly larger and more environmentally-heterogeneous training sets, we ranked training environments based on their relatedness (i.e. distance) to a validation environment, incrementally built training sets, and tested predictions. Similar lines of inquiry have been explored for constructing training sets. Heslot et al. (2013) used a backwards elimination approach to remove unproductive environments, while Lorenz and Smith (2015) instead ranked genotypes in a training population by their relationship to a VP and sequentially built training sets. Our analysis is more akin to the latter, where an additional data source describing the relationship among environments (i.e. the distance measures) was used. In general, we found that adding environments to a TS initially increases accuracy, but after a certain point, additional environments were not useful and occasionally reduced r_{MG} . We identified similarities in the trends of r_{MG} under nearly all circumstances, and we describe this and the impact of trait, distance measure, and prediction scheme below.

The initial gain in r_{MG} with additional environments in the TS can likely be ascribed to better estimates of genotypic values with additional data. Here, the gain in precision outweighs any noise that may be introduced by error or GEIs. Eventually, however, additional environments did not provide useful information, and an equilibrium formed between preci-

sion in estimating genotypic values and noise introduced by GEI. Interestingly, the number of environments in which this equilibrium was reached – typically five – was consistent across traits and prediction scenarios (LOEO vs. TF) This is encouraging for practical purposes, as it suggests that a small, reasonable number of environments is sufficient for accurate predictions.

Among the distance measures tested, AMMI and PD rankings of environments often led to the greatest advantage in r_{MG} of few, highly related environments over using all available data. This result was very apparent for the CV prediction schemes, where including more distant environments led to a precipitous decline in accuracy. Under the AMMI measure, environments with similarly repeatable GEI will be considered more related, and the PD measure defines higher relatedness by higher correlations between environments (Gauch and Zobel, 1997; Bernardo, 2010). Furthermore, the data used to calculate these distance measures was pulled exclusively from the TP, the population in which CV was performed. This might introduce an upward bias in predictions, so we compared the impact of distance measures between both CV and POV schemes. In POV, the superiority of the AMMI and PD measures was not as apparent, and differences between measures in the patterns of r_{MG} were marginal. It is possible that the VP experienced different GEI than the TP, which might limit the usefulness of distance measures calculated using observations on the TP; however, we observed little difference in variance components when analyzing the TP and VP separately (Fig. S2.8), suggesting that population-specific GEIs do not explain this observation. We suspect lower relatedness between the TP and VP is responsible, and we elaborate on this below.

Clustering environments marginally improves prediction accuracy

The ideal environment clustering method is one that reduces GEI sufficiently to allow simpler modeling and predictions of the genotypic mean (Bernardo, 2010). In our study, the same data employed to rank environments based on distance was used to group environments into ostensibly homogenous clusters within which genomewide predictions were assessed. While

there never was a significant gain in r_{MG} from using a subset of TS environments assigned to a particular cluster as opposed to simply using all available data, some interesting trends were present. As with sequentially adding TS environments, results were dependent on the trait, distance measure, and prediction scheme.

Clustering using phenotype-based distance measures (i.e. AMMI, PD, and LocPD) often yielded numerically higher prediction accuracies compared to other measures. This result is consistent with expectations, since these measures would inherently attempt to minimize GEI (in the case of AMMI) or select environments that were highly correlated (in the case of PD or LocPD) (Bernardo, 2010). As GEI variance in our experiment was due mostly to lack of correlation between environments (Table 2.4), we might expect both measures to yield similar clusters. Indeed, the proportion of overlap between clusters assigned by AMMI and PD was moderate to high (Fig. S2.9). These results suggest that these measures are effective in reducing GEI, leading to more precise estimates of genotypic values. This is in line with previous research indicating that training sets composed of environments more related or correlated to a validation environment allow borrowing of information and improvement of prediction accuracy (Burgueño et al., 2012; Lado et al., 2016; Jarquin et al., 2017). Such studies have typically addressed a single trait, and we note in our investigation that traits may be unevenly impacted.

The target trait for genomewide predictions appeared to influence the advantage in r_{MG} of using cluster-informed training sets compared with simply using all available data. Predictions of grain yield and plant height were more accurate with clustering based on AMMI and PD, while r_{MG} of heading date was often detrimentally impacted. Clustering based on the PD and AMMI measures always led to greater heritability efficiency, defined as the ratio between the heritability with clustering and that without (Fig. S2.10). For grain yield and plant height, the advantage of these distance measures was clear, as other measures led to low heritability efficiencies. Though heading date heritability was improved when clustering using AMMI or PD, the margin over not clustering was slim, and across all distance measures the heritability

efficiency was close to unity. This difference among traits may be due to their inherent complexity and preponderance to be influenced by GEIs. Grain yield and plant height are typically more complex traits that are subject to GEIs, and may therefore benefit most from attempts to reduce GEI by clustering. In contrast, heading date is highly heritable and less influenced by environment (Hockett and Nilan, 1985); therefore, clustering may simply reduce the amount of otherwise informative data for estimating genotypic values, leading to the observed reduction in accuracy.

The AMMI distance measure, though it often led to higher prediction accuracy, does not allow for assigning new locations to clusters, a plausible goal for a centralized breeding effort. An interesting alternative explored in our study was the use of ECs that are correlated with the environmental interaction PCA score from a fitted AMMI model (i.e. IPCA-EC). Theoretically, this measure would identify ECs that explained the repeatable portion of GEI (van Eeuwijk and Elgersma, 1993; Bernardo, 2010; Gauch, 2013) and then cluster similar environments based on those ECs. The resulting clusters seemed to confirm this expectation, as they overlapped with AMMI clusters (Fig. S2.9) and yielded the highest cluster heritability efficiency (Fig. S2.10) than any other EC-based distance measure. This may help explain why, particularly for grain yield and plant height, IPCA-EC clustering led to prediction accuracies similar to those of AMMI and superior to those of other EC-based clustering measures (Fig. 6A and 7A). Nevertheless, such clusters were often not any more informative than random groups of environments (Fig. 6B and 7B) and occasionally were not informative at all; such was the case for plant height in the LOEO scenario (Fig. 6A), where the IPCA-EC measure resulted in a single cluster (Fig. S2.9).

Simply using all data may be preferable when predicting a new generation

Our investigation addressed the realistic breeding scenario of predicting genotypes that are progeny of a distinct TP. It would be valuable to a breeder to determine, at the early generation

stage, what lines would be best in specific regions. This could potentially allow for larger populations to be developed, with selection relying on region-specific predictions instead of phenotypes (Lorenz et al., 2011). Though much research has focused on comparing different GWS models that attempt to exploit GEI by modeling interactions or reduce GEI by grouping environments (Burgueño et al., 2012; Dawson et al., 2013; Heslot et al., 2014; Lado et al., 2016; Saint Pierre et al., 2016; Jarquin et al., 2017), little work has explored predictions between two discrete and related populations.

In our study, predictions of an unobserved VP (i.e. POV) were consistently less accurate than predictions within a single TP (i.e. CV). Additionally, we frequently observed that any measure of grouping environments to reduce GEI that favorably impacted prediction accuracy under CV was ineffective under POV. We suspect this is the result of lower relatedness between the TP and VP than within the TP (Fig. S2.12). Research in GWS has clearly demonstrated that accurate predictions depend on high relatedness between a TP and VP (Habier et al., 2007; Lorenz et al., 2012; Wientjes et al., 2013; Lehermeier et al., 2014; Lorenz and Smith, 2015). Such research has typically focused on predictions of the genotypic value within a single environment or by otherwise ignoring GEI. Modeling to exploit or reduce GEI in GWS arguable adds complexity, and it may be unreasonable to expect that a TP that struggles to predict the overall value of a distinct VP will also permit accurate predictions when accounting for environmental context. This may help explain why adding environments to a TS, regardless of distance measure, generally increased accuracy under POV (Fig. 2.4 and 2.5), and clustering did not lead to improvements in accuracy over using all available data (Fig. 2.6 and 2.7).

Our results suggest that measures to rank and cluster environments were equivalent to simply using data from all available environments when predicting a new, unobserved generation. This has two implications for breeding. First, if phenotypic data is available across a broad swath of environments, as might be expected in a large breeding program or, in our case, a coordination among programs, GEIs may be simply be ignored and all data used for predic-

tions. Second, the few environments needed to achieve useful prediction accuracies (i.e. about 5) likely reflects the number of environments in which a breeding program of small or moderate size will have data. This indicates that accurate predictions of a new generation may be universally achievable. Simplicity is desirable when implementing GWS, and our results may be encouraging for a new breeding program or one that is considering using this tool.

This does not necessarily conflict with the body of research that advocates accounting for GEI to increase prediction accuracy (Jarquín et al., 2014; Cuevas et al., 2016; Lado et al., 2016; Saint Pierre et al., 2016; Jarquin et al., 2017). These studies examined a single population or generation, and the advantage of modeling GEI in GWS was generally realized when data from incompletely observed genotypes and environments can be borrowed to predict unobserved combinations (i.e. cross-validation type 2). With this information from prior research and the results of our investigation, we suggest strategic deployment of more complex GWS models in different stages of a breeding program. First, a new generation of unobserved selection candidates may be selected simply based on predictions of their overall genotypic value. Then, this population can be phenotyped using an appropriately designed partial replication experiment, and GWS models that borrow information from correlated environments can be used to predict unobserved genotype-environment combinations. This could permit a less stringent selection intensity in the first step in exchange for potential identification of superior, regionally-adapted genotypes in the second. More research is needed to test and validate this concept.

FIGURES

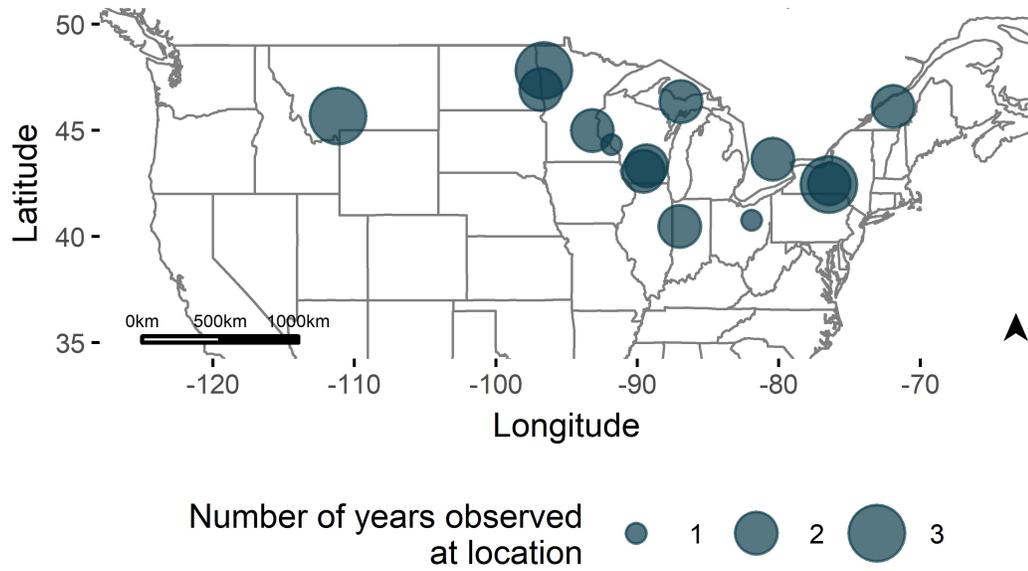


Figure 2.1: Phenotypic data was collected in 29 location-year environments across North America. Point size indicates the number of trial years at a specific location.

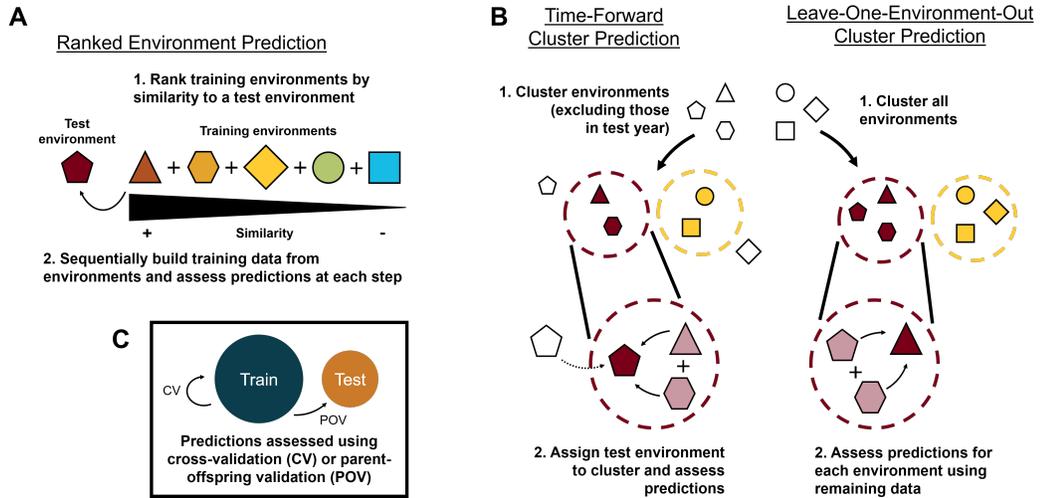


Figure 2.2: Distance measures were used to create genomewide prediction training sets. (A) Possible training environments were ranked in order of decreasing relatedness to a validation environment, then sequentially added to a training set. (B) Environments were clustered according to a distance measure, then predictions were assessed for environments in an unobserved year (time-forward) or individually (leave-one-environment-out). (C) Prediction accuracy was assessed using cross-validation (CV) within the training population (Test) and parent-offspring validation (POV) between the training and test populations (Test).

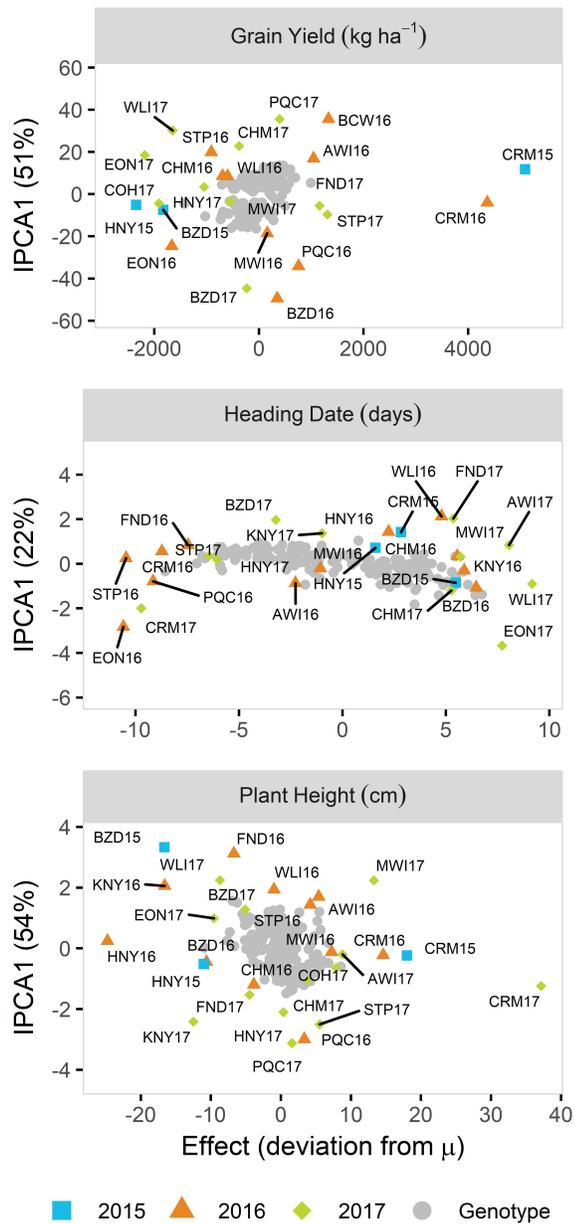


Figure 2.3: Biplots from additive main effects and multiplicate interaction (AMMI1) models illustrate the spatial pattern of the first interaction principal component axis (IPCA1) score against the genotypic (circles) or environmental (different shape per year) effect. The percentage of genotype-environment interaction variance explained by the IPCA1 score is noted in parentheses.

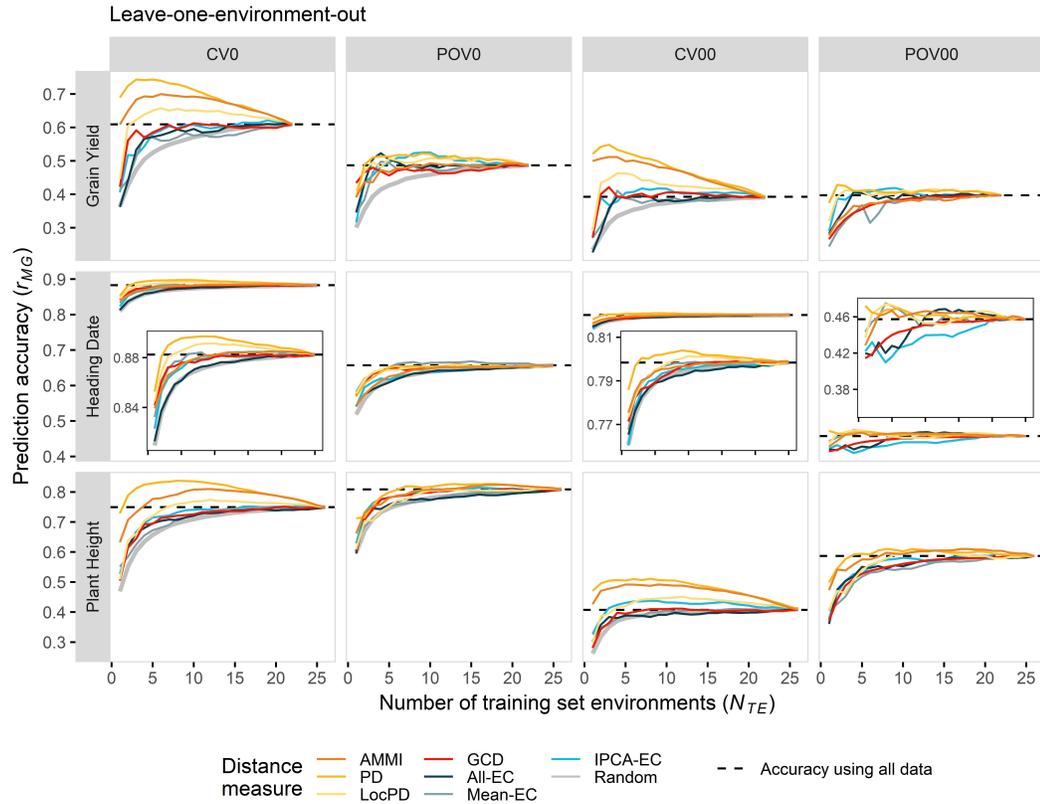


Figure 2.4: Prediction accuracy (r_{MG}) generally improved with increasing number of environments in a training set (N_{TE}) under the leave-one-environment-out scenario, but either reach a plateau or declined. Environments were added to a training set in order of increasing distance, defined by a distance measure to a validation environment. r_{MG} was assessed using cross-validation (CV0 or CV00) and parent-offspring validation (POV0 or POV00). Dashed lines denote r_{MG} when using all available data in the training set. Plot-in-plot windows for heading date highlight results on a more easily visualized scale.

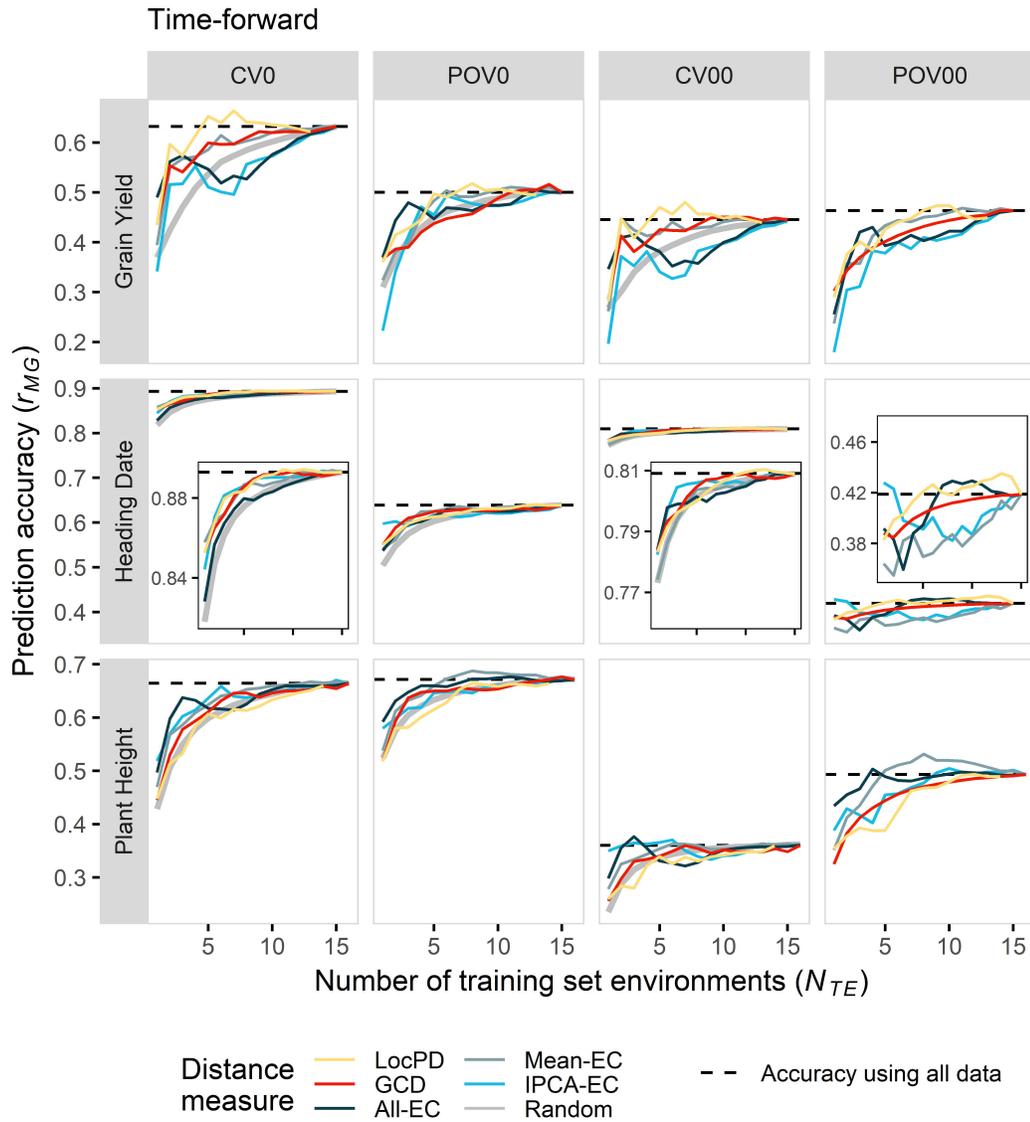


Figure 2.5: Prediction accuracy (r_{MG}) generally improved with increasing number of environments in a training set (N_{TE}) under the time-forward scenario. Environments were added to a training set in order of increasing distance, defined by a distance measure to a validation environment. r_{MG} was assessed using cross-validation (CV0 or CV00) and parent-offspring validation (POV0 or POV00). Dashed lines denote r_{MG} when using all available data in the training set. Plot-in-plot windows for heading date highlight results on a more easily visualized scale.

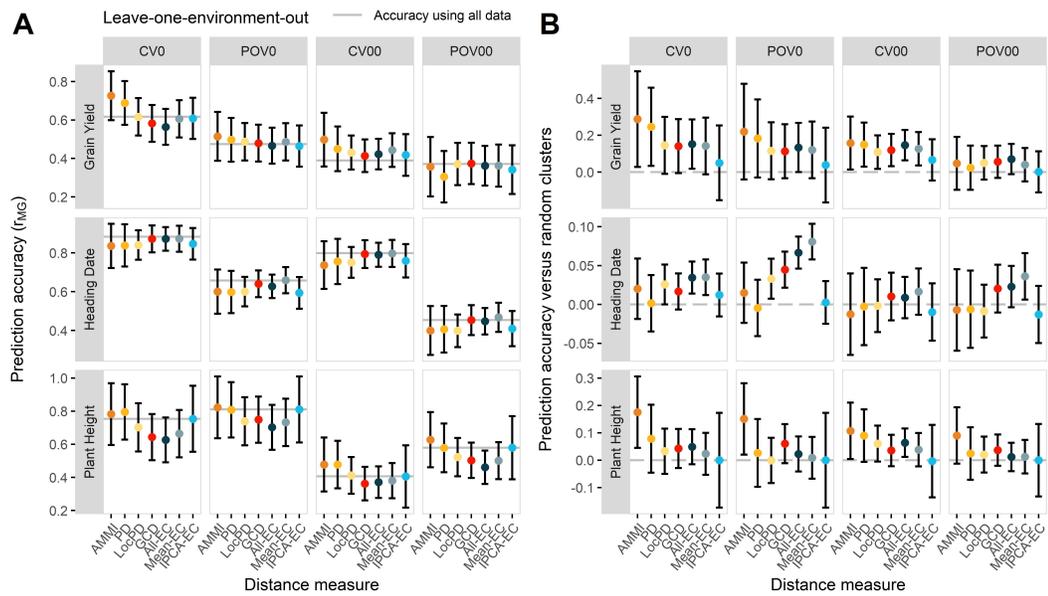


Figure 2.6: Effect plot of prediction accuracy (r_{MG}) when using data assigned to clusters based on a distance measure under the leave-one-environment-out scenario. r_{MG} was assessed using cross-validation (CV0 or CV00) and parent-offspring validation (POV0 or POV00). Error bars denote a 95% confidence interval and the grey line denote r_{MG} when using all available data in the training set.

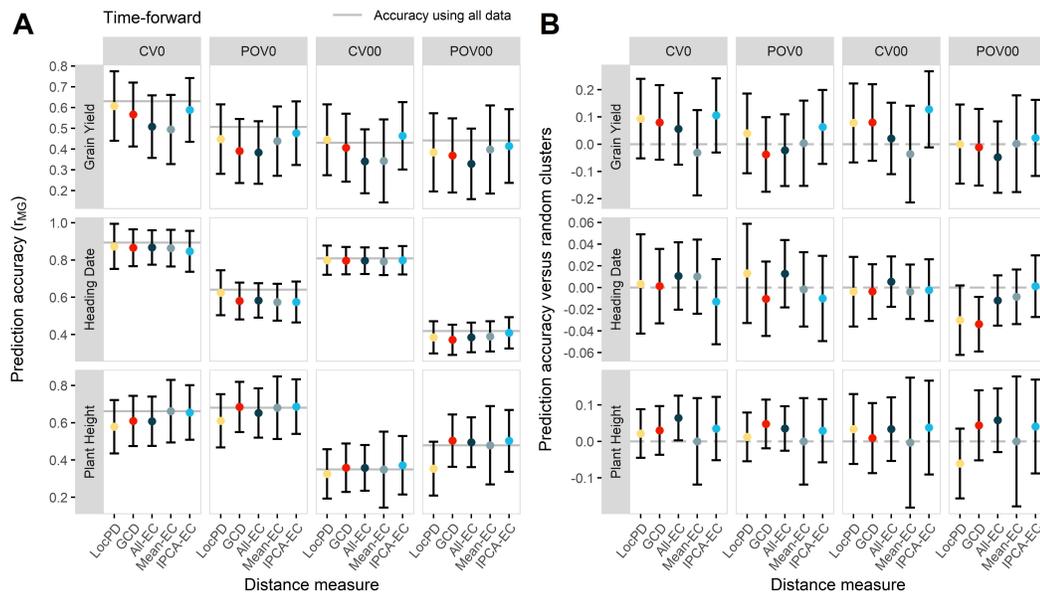


Figure 2.7: Effect plot of prediction accuracy (r_{MG}) when using data assigned to clusters based on a distance measure under the time-forward scenario. r_{MG} was assessed using cross-validation (CV0 or CV00) and parent-offspring validation (POV0 or POV00). Error bars denote a 95% confidence interval and the grey line denotes r_{MG} when using all available data in the training set.

TABLES

Table 2.1: Names, abbreviations, and descriptions of the distance measures used to characterize environments, along with an indication of trait-specificity and ability to make predictions of new environments.

Distance measure	Abbreviation	Description	Trait specific?	Enables predictions of new environments?
Phenotypic distance	PD	Distance based on phenotypic correlation of shared genotypes	Yes	No
Location phenotypic distance	LocPD	Similar to PD, but using phenotype average for locations	Yes	Yes, if location has been observed
AMMI mega-environments	AMMI	Mega-environments determined by shared winning genotypes	Yes	No
Great circle distance	GCD	Physical proximity of locations	No	Yes
All environmental covariables	All-EC	All available environmental covariables	No	Yes
Interaction-correlated environmental covariables	IPCA-EC	Covariables that are correlated with the environmental IPCA score	Yes	Yes
Mean-correlated environmental covariables	Mean-EC	Covariables that are correlated with the environmental mean	Yes	Yes

SUPPLEMENTAL FIGURES

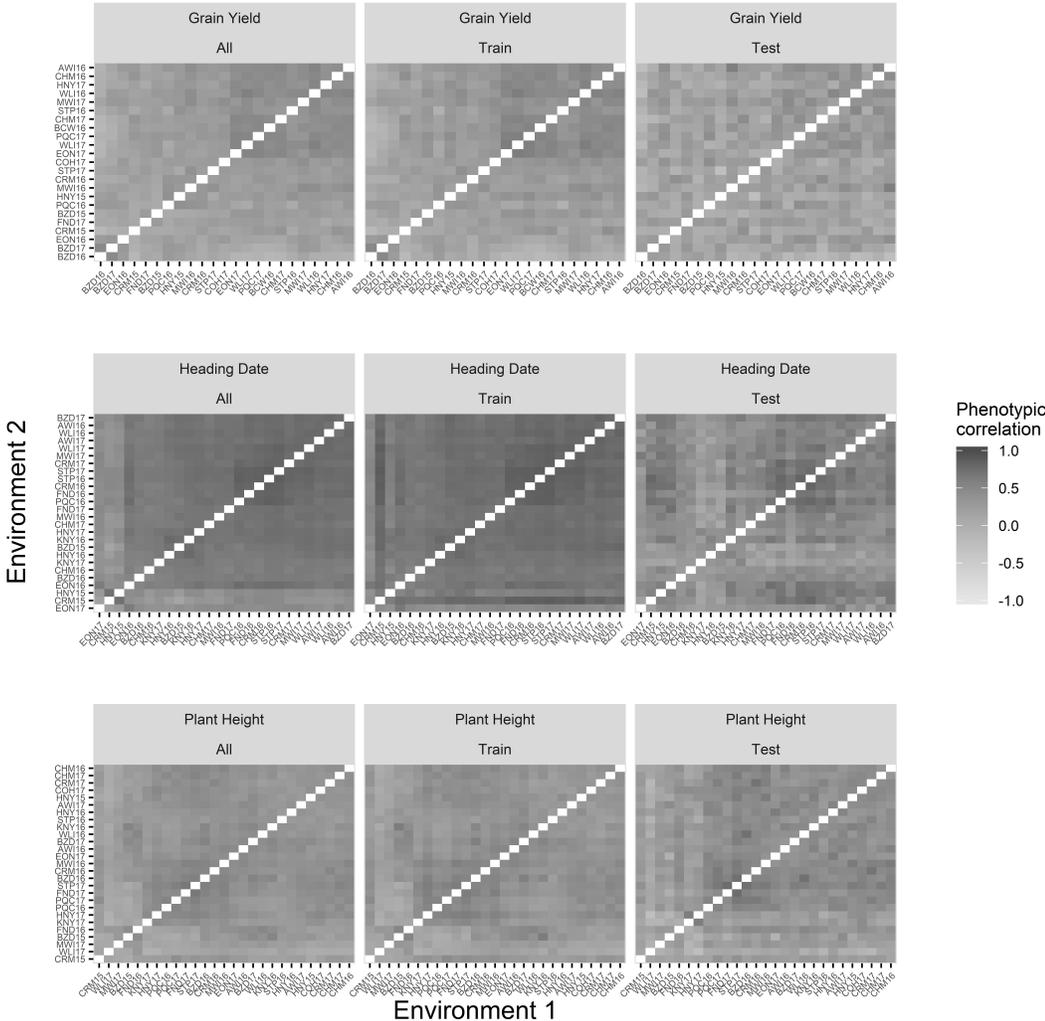


Figure 2.8: Heatmap of phenotypic correlations between environments, measured using all 233 barley lines (All), the 183-line training population (Train), or the 50-line test population (Test).

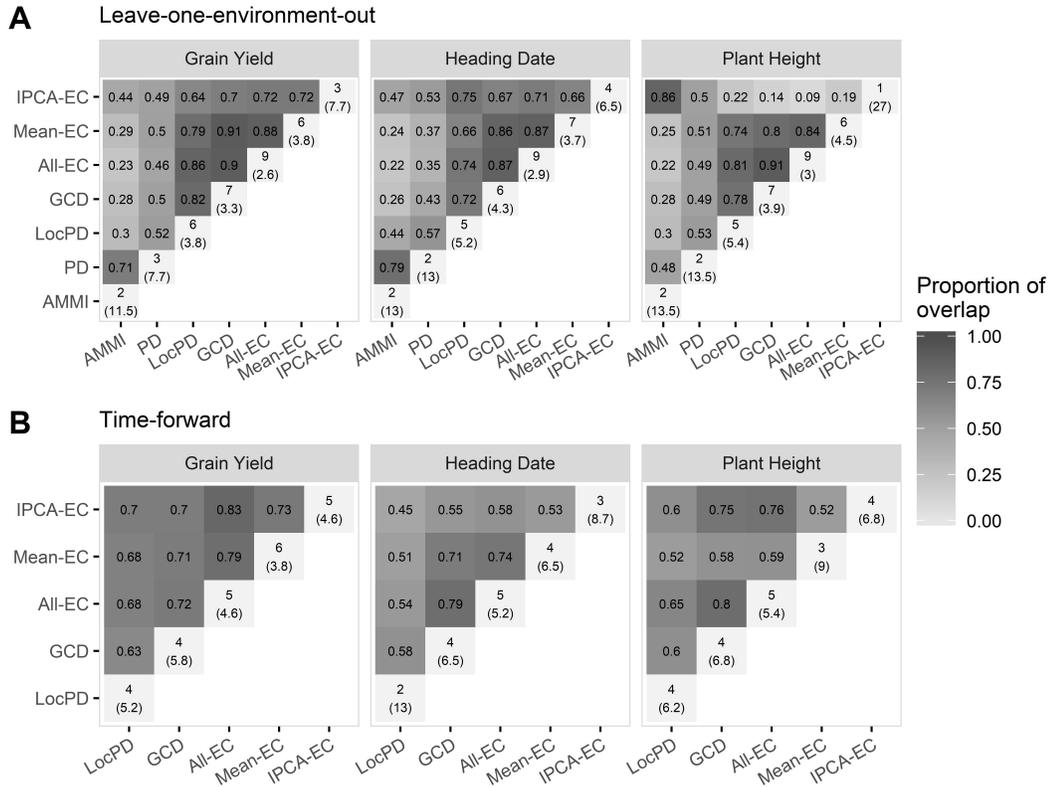


Figure 2.9: Overlap of clusters created by different distance measures. Off-diagonal elements measure the proportion of overlap between clusters assigned by two distance measures. Diagonal elements measure the number of clusters (and average number of environments per cluster in parentheses). Clusters were compared under the (A) leave-one-environment out and (B) time-forward scenarios.

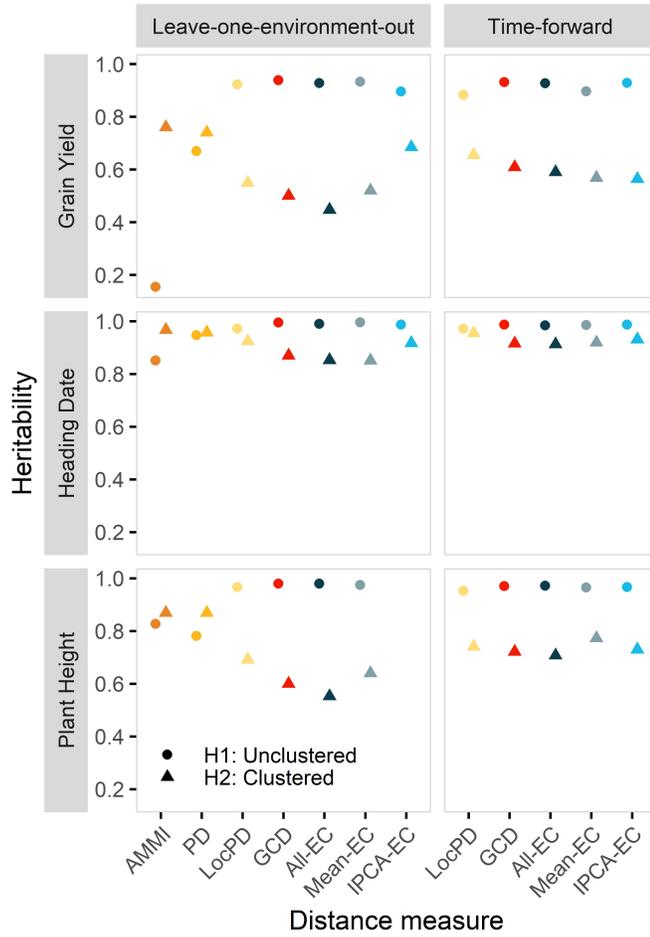


Figure 2.10: Comparison of heritability estimates when selection is performed across unclustered (circle) or clustered (triangle) environments. Estimates were made for three traits under the leave-one-environment-out or time-forward scenarios.

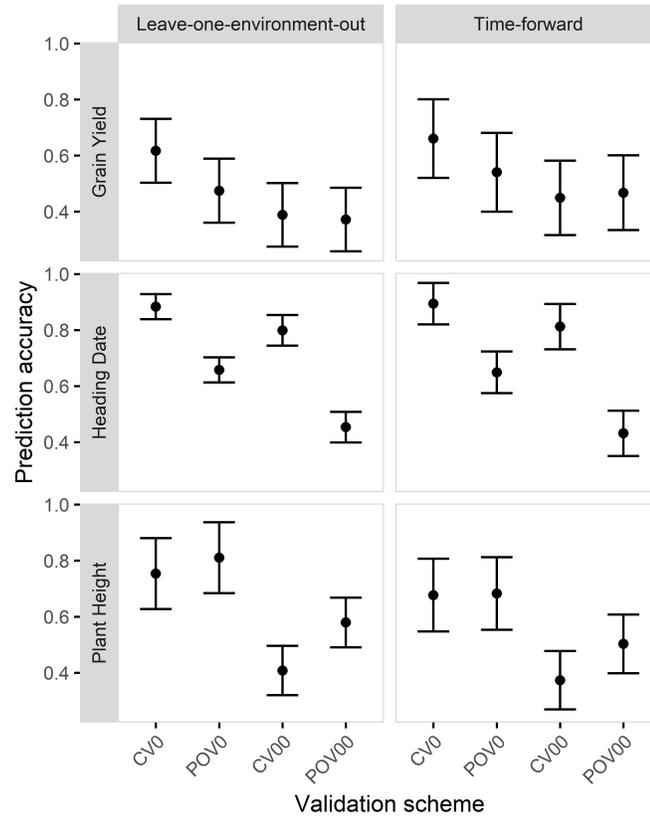


Figure 2.11: Effect plot of prediction accuracy when using all available data to make predictions. Under the leave-one-environment-out scenario, phenotypic data for all environments – except for the validation environment – were used in model training. Under the time-forward scenario, phenotypic data from all environments in 2015 and 2016 were used to predict each 2017 environment. Accuracy was measured using cross-validation [CV; predictions within the training population (TP)] or parent-offspring validation [POV; TP used to predict the validation population (VP)]. Numbers correspond to the validation schemes: 0 – predictions of observed genotypes in unobserved environments; 00 – predictions of unobserved genotypes in observed environments (see Jarquin et al., 2017 for more detailed definitions). (Note: POCV refers to predictions of the VP using the smaller TP subsets under CV as a control for TP size).

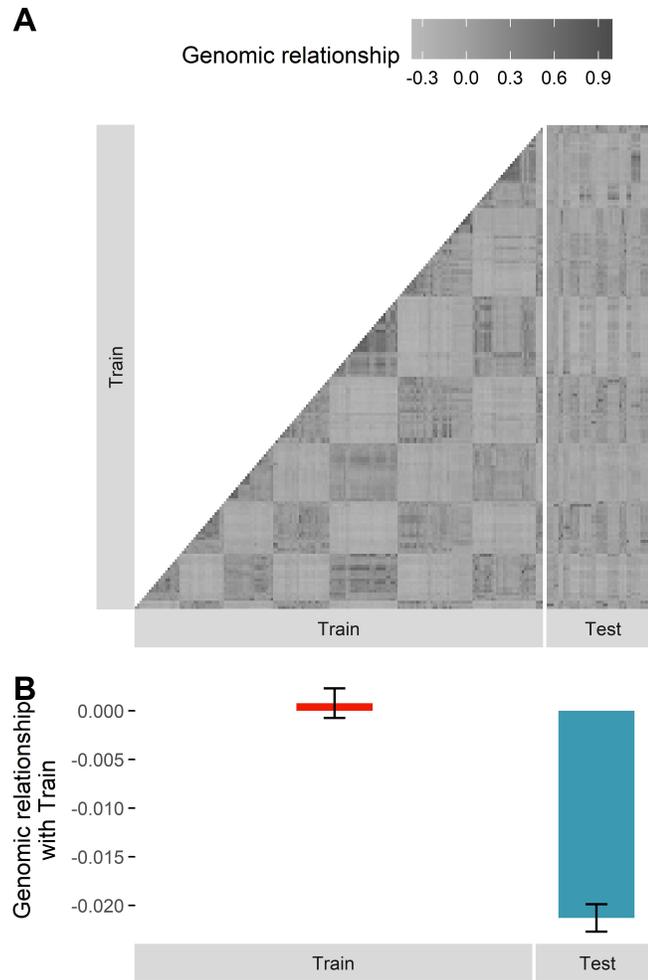


Figure 2.12: Comparison of realized marker-based genomic relatedness within the training population (Train) and between the training and test populations (Test). **(A)** heatmap of genomic relatedness between individual lines. **(B)** Average genomic relatedness within the training population and between the training and test populations. Error bars denote one standard deviation.

SUPPLEMENTAL TABLES

Table 2.2: Estimates of heritability (H) for each trait in each environment. Missing estimates (“NA”) are due to absent phenotypic data.

Environment	H		
	Grain Yield	Heading Date	Plant Height
AWI16	0.37	0.80	0.26
AWI17	NA	0.77	0.66
BCW16	0.60	NA	NA
BZD15	0.25	0.95	0.88
BZD16	0.18	0.77	0.46
BZD17	0.43	0.90	0.37
CHM16	0.73	0.69	0.67
CHM17	0.34	0.74	0.60
COH17	0.65	NA	0.75
CRM15	0.22	0.89	0.48
CRM16	0.64	0.87	0.11
CRM17	NA	0.92	0.71
EON16	0.66	0.90	NA
EON17	0.85	0.80	0.50
FND16	NA	0.87	0.60
FND17	0.20	0.76	0.79
HNY15	0.31	0.88	0.49
HNY16	NA	0.88	0.47
HNY17	0.75	0.92	0.69
KNY16	NA	0.92	0.51
KNY17	NA	0.64	0.59
MWI16	0.67	0.91	0.62
MWI17	0.68	0.94	0.31
PQC16	0.60	0.89	0.23
PQC17	0.84	NA	0.47
STP16	0.86	0.94	0.31
STP17	0.61	0.98	0.72
WLI16	0.80	0.90	0.75
WLI17	0.71	0.89	0.29

Table 2.3: Names and descriptions of environmental covariates [see O'Donnell and Ignizio (2012) for formulae].

Environmental covariable	Description
Int_i_Prp	Total precipitation (mm) in interval (Int) i
$Int_i_AvgTemp$	Average temperature (C) in interval i
$Int_i_MaxTemp$	Maximum temperature in interval i
$Int_i_MinTemp$	Minimum temperature in interval i
$Int_i_RangeTemp$	Average daily temperature range (maximum minus minimum) in interval i
$Int_i_SeasTemp$	Seasonality (standard deviation) of daily temperature ranges in interval i
$AnnoPrp$	Annual total precipitation
$MaxPrp$	Total precipitation of wettest interval
$MinPrp$	Total precipitation of driest interval
$AnnoAvgTemp$	Average annual temperature
$AnnoMaxTemp$	Average maximum temperature
$AnnoMinTemp$	Average minimum temperature
$AvgRangeTemp$	Average annual temperature range (maximum minus minimum)
$AnnoRangeTemp$	Peak annual temperature range (annual maximum minus annual minimum)
$AnnoSeasTemp$	Standard deviation of average interval temperature
$IsoTemp$	Isothermality ($AvgRangeTemp / AnnoRangeTemp$)
$MaxTemp$	Maximum temperature of warmest interval
$MinTemp$	Minimum temperature of coldest interval
$ClaySS$	Clay content (% , by weight) in the subsoil (SS, > 20 cm below surface)
$ClayTS$	Clay content in the topsoil (TS, 0 - 20 cm below surface)
$SandSS$	Sand content (% , by weight) in the subsoil (> 20 cm below surface)
$SandTS$	Sand content in the topsoil (0 - 20 cm below surface)
$SiltSS$	Silt content (% , by weight) in the subsoil (> 20 cm below surface)
$SiltTS$	Silt content in the topsoil (0 - 20 cm below surface)
$OrgMatSS$	Organic matter content (% , by weight) in the subsoil (> 20 cm below surface)
$OrgMatTS$	Organic matter content content in the topsoil (0 - 20 cm below surface)
$SoilpHSS$	pH in the subsoil (> 20 cm below surface)
$SoilpHTS$	pH in the topsoil (0 - 20 cm below surface)
$Elevation$	Meters above sea level
$Latitude$	Latitude
$Longitude$	Longitude

Table 2.4: Estimates of variance components and proportion of total variance for each source of variation in an analysis of phenotypic data on the entire 233-line population (ALL), the 183-line training population (TP), or the 50-line validation population (VP). The significance of variance components was determined using a likelihood ratio test.

Source	Variance (percent of total)		
	ALL	TP	VP
Grain Yield			
Genotype	101000*** (2.35%)	106000*** (2.45%)	54500*** (1.36%)
Location	2870000*** (66.5%)	2920000** (67.2%)	2870000*** (71.6%)
Year	0.041 (0%)	0 (0%)	90.7 (0.002%)
Genotype x Location	41200*** (0.955%)	43900*** (1.01%)	40100*** (1%)
Genotype x Year	6000* (0.139%)	5520 (0.127%)	4790 (0.119%)
Location x Year	682000*** (15.8%)	709000*** (16.3%)	645000*** (16.1%)
Genotype x Location x Year	256000*** (5.94%)	260000*** (5.98%)	210000*** (5.25%)
Genetic Heterogeneity [†]	22900 (8.93%)	20000 (7.7%)	19300 (9.2%)
Lack Of Correlation [†]	233000 (91.1%)	240000 (92.3%)	191000 (90.8%)
Residual	355000 (8.24%)	303000 (6.96%)	182000 (4.54%)
Heading Date			
Genotype	11*** (17%)	13.1*** (20.3%)	3.98*** (6.1%)
Location	6.38 (9.84%)	3.49 (5.41%)	14.7 (22.5%)
Year	1.73 (2.67%)	1.94 (3%)	0.639 (0.98%)
Genotype x Location	0.586*** (0.904%)	0.857*** (1.33%)	0.645** (0.989%)
Genotype x Year	0.217* (0.334%)	0.029 (0.045%)	0 (0%)
Location x Year	37.7*** (58.2%)	39.1*** (60.6%)	40*** (61.3%)
Genotype x Location x Year	7.18*** (11.1%)	6.05*** (9.37%)	5.32*** (8.15%)
Genetic Heterogeneity	1.06 (14.7%)	1.27 (21%)	1.18 (22.2%)
Lack Of Correlation	6.13 (85.3%)	4.78 (79%)	4.13 (77.8%)
Residual	0.001 (0.001%)	0 (0.001%)	0 (0%)
Plant Height			
Genotype	9.56*** (5.21%)	9.38*** (5.08%)	10.1*** (5.63%)
Location	97.9*** (53.4%)	103*** (55.6%)	80.9** (45.1%)
Year	11.1 (6.05%)	11.7 (6.31%)	11 (6.15%)
Genotype x Location	1.54*** (0.841%)	1.59*** (0.86%)	1.24 (0.69%)
Genotype x Year	0.075 (0.041%)	0.121 (0.066%)	0.031 (0.017%)
Location x Year	44.1*** (24.1%)	40.9*** (22.1%)	57.6*** (32.1%)
Genotype x Location x Year	19*** (10.4%)	18.5*** (10%)	18.4*** (10.3%)
Genetic Heterogeneity	2.03 (10.7%)	2 (10.8%)	2.41 (13.1%)
Lack Of Correlation	17 (89.3%)	16.5 (89.2%)	16 (86.9%)
Residual	0.02 (0.011%)	0.022 (0.012%)	0.018 (0.01%)

*, **, *** Significant at the 0.05, 0.01, and 0.001 probability levels, respectively.

[†]No test of significance was performed on these components of variation.

Table 2.5: Environmental covariables (ECs, Table S2) that were associated with the first interaction principal component axis (IPCA-EC) or environmental mean (Mean-EC), as determined using multiple linear regression using data under the leave-one-environment-out and time-forward scenarios.

Set	Group	Environmental covariables		
		Grain Yield	Heading Date	Plant Height
Leave-one-environment-out	EC_IPCA	Int4_SeasTemp	OrgMatTS	Int1_MaxTemp
	EC_Mean	Int2_Prcp	MaxMaxTemp	
		Latitude	Int2_MinTemp	SoilpHTS
		Int2_RangeTemp	Latitude	Latitude
		MaxPrcp		SoilpHSS
		MinPrcp		Int2_SeasTemp
				Int1_SeasTemp
		MaxPrcp		
Time-forward	EC_IPCA	Int2_Prcp	SiltTS	Int4_MaxTemp
		Int1_SeasTemp	Int1_Prcp	OrgMatTS
		Int4_MinTemp	AnnoMaxTemp	
		Int2_MaxTemp	MaxPrcp	
		SiltTS	Latitude	
		AnnoMaxTemp	ClaySS	
		Latitude	Int2_Prcp	
		ClaySS	Int3_AvgTemp	
		Int3_AvgTemp	SoilpHTS	
		Int4_Prcp	MinPrcp	
			Int4_Prcp	
			Int2_RangeTemp	
	EC_Mean		Int4_SeasTemp	
			Int3_RangeTemp	
		Int1_Prcp	Int2_AvgTemp	OrgMatTS
		Int3_RangeTemp	Int1_SeasTemp	Int3_MinTemp
		SoilpHSS	OrgMatTS	AnnoAvgTemp
			Int1_MaxTemp	AnnoMinTemp
				AvgRangeTemp
				SoilpHTS
				MaxPrcp
				MinPrcp
				Int2_RangeTemp
		Int3_AvgTemp		
		Int4_AvgTemp		
		Int2_Prcp		
		Latitude		

Chapter 3: Validating genomewide predictions of genetic variance in a contemporary breeding program

The following chapter was published in a peer-reviewed journal and can be located using the reference:

Jeffrey L. Neyhart and Kevin P. Smith. 2019. Validating Genomewide Predictions of Genetic Variance in a Contemporary Breeding Program. *Crop Science*. 59(3).

Author contributions:

JLN designed experiments, collected/analyzed data, and wrote the manuscript. All authors conceived of the study and read and approved the final manuscript.

INTRODUCTION

Improving quantitative traits in plants relies on maintaining the response to selection, a function of trait heritability, selection intensity, and genetic variance (Falconer and Mackay, 1996). Much breeding research has focused on better selection of individuals in a given population, chiefly by increasing the heritability or permitting more intense selection, under the implicit condition of fixed variance (Bernardo, 2010). This emphasis has persisted with the advent of genomewide selection, or the use of phenotyped individuals with genomewide marker data to predict the genotypic value of unphenotyped individuals (Meuwissen et al., 2001). While simulation and empirical studies have confirmed the usefulness of genomewide selection in plant breeding (Lorenz et al., 2012; Asoro et al., 2013; Combs and Bernardo, 2013; Krchov et al., 2015; Sallam et al., 2015; Dunckel et al., 2016), such research has primarily addressed improved selection of progeny from established populations derived from pre-chosen parents.

Beyond improvements to the selection of individuals within a population, the development of superior populations, by choosing favorable crosses between parents, would aid in increasing the response to selection. Ideally, a cross should yield a population with favorable mean and large genetic variance (Bernardo, 2010). The benefit of such a combination is formalized by the usefulness criterion (Schnell and Utz, 1975), or the mean value of the selected fraction of individuals in a population. The related superior progeny mean (μ_{sp}) (Zhong and Jannink, 2007) of a population can be derived from the expected population mean (μ) and genetic variance (V_G), assuming normally distributed trait values: $\mu_{sp} = \mu \pm i\sqrt{V_G}$, where i is the standardized selection intensity. (The superior progeny mean is the same as a usefulness criterion with a heritability of 1.) Simply, among populations with similar mean, those with larger genetic variance will yield more favorable superior progeny.

The superior progeny mean may be used to prioritize crosses in a breeding program if the expected mean and genetic variance is known or can be predicted. Such predictions have

long been sought by plant breeders, albeit with uneven ease and success. Classical quantitative genetics theory predicts the mean of a population of recombinant inbreds as the mean of the two parents (Bernardo, 2010). The genetic variance of a cross, however, has historically been difficult to predict. Earlier efforts used comparisons between parents of their phenotypic distance (Utz et al., 2001), coefficient of coancestry (Souza and Sorrells, 1991), or genetic distance (Bohn et al., 1999), all with limited or inconsistent success (reviewed in Bernardo (2010) and Mohammadi et al. (2015)]. The phenotypic distance method, while permitting different predictions of variance between traits, does not account for the segregation of the underlying quantitative trait loci (QTL) or markers linked to such loci. The coefficient of coancestry and genetic distance methods measure whole-genome distance, including potentially neutral loci, and do not offer trait-specific predictions. The ability to predict the effects of genomewide markers, coupled with affordable and powerful computing resources, has led to a new method of predicting genetic variance that is based on simulated bi-parental populations (Bernardo, 2014; Mohammadi et al., 2015). Initial validation research suggests that simulated populations, or methods that account for the segregation of loci in bi-parental families, are more predictive of genetic variance than the aforementioned methods based on phenotypic or genetic distance (Tiede et al., 2015). Though these results are encouraging, further validation work is needed, especially in the context of applied breeding.

Validating the use of simulated populations to predict the genetic variance of future crosses in a breeding program would add to the current theoretical evidence supporting its routine application. In barley (*Hordeum vulgare* L.), initial predictions suggest that both grain yield and deoxynivalenol (a mycotoxin produced by the fungal pathogen *Fusarium graminearum* Schwabe.) content could be improved by selecting crosses with favorable superior progeny mean, calculated using predictions of the mean and genetic variance of a cross (Mohammadi et al., 2015). Similar evidence was found for quality traits in wheat (*Triticum aestivum* L.), where low genetic variance may be desirable to restrict the trait value within indus-

try specifications (Lado et al., 2017). Finally, results from simulations suggest that crosses selected on the predicted superior progeny mean may lead to greater genetic gain compared to those selected simply on the predicted mean (Lehermeier et al., 2017).

The objective of this research was to validate predictions of the mean, genetic variance, and superior progeny mean in populations within a contemporary barley breeding program. We focused on three important quantitative traits in barley: Fusarium head blight (FHB) severity, heading date (a proxy for flowering time), and plant height. Given the known differences in genetic architecture and heritability among these traits (Hockett and Nilan, 1985), we hypothesized that predictions would be more accurate for the simple and highly heritable heading date and less accurate for the more complex FHB severity.

METHODS AND MATERIALS

Training population

A training population (TP) of 175 two-row barley lines was used for genomewide prediction. The assembly and genotyping of this population is described in Neyhart et al. (2019) and Chapter 2, and we describe relevant details below. The TP was phenotyped in Crookston, MN and St. Paul, MN in 2014 and 2015 (four location-year environments total). In each environment, data was collected in preliminary yield trials, FHB disease trials, or both. We measured FHB severity (percent diseased kernels), heading date (days from planting), and plant height (cm) in the disease trials, and we measured all traits, except FHB severity, in the yield trials. Trials in 2014 were planted in a type II modified augmented design (Lin and Poushinsky, 1985), with a primary check placed in the center of each block and three secondary checks randomly placed together in a subset of blocks. Trials in 2015 were planted in a similar incomplete block design, with the exceptions that the primary check was randomly placed in each block and secondary checks were randomly placed throughout the entire field. Entries were planted as single-row plots of size 0.46 m². Disease trials were conducted under spray-inoculated (St. Paul) and grain spawn-inoculated (Crookston) conditions, and mist irrigation was used in both locations to encourage disease development (described in Mesfin et al., 2003, and Massman et al. (2011)).

Due to differences in trial design, a two-step procedure was used to process and analyze phenotypic data (Piepho et al., 2012a). In the first step, obvious outliers were removed and observations were adjusted for field spatial variability using the *mvngGrAd* package in R (Technow, 2015). Spatially-adjusted phenotypic values were retained if the within-trial variance was reduced after adjustment. Subsequently, we fitted the following mixed model to estimate adjusted genotype means on a per-trial and per-trait basis:

$$y_{iklm} = \mu + G_i + r_k + c_l + b_m + \epsilon_{iklm}, \quad (3.1)$$

where y_{ijkl} was the observed phenotypic value of the i th genotype in the k th row, l th column, and m th block, μ was the grand mean in the trial, G_i was the fixed effect of the i th genotype, r_k was the effect of the k th row, c_l was the effect of the l th column, b_m was the effect of the m th block, and ϵ_{iklm} the associated random error. The effects r_k , c_l , and b_m were all considered random. When analyzing FHB severity, we included the observed heading date as a random effect, as plants may avoid higher inoculum load (and therefore may be less severely diseased), depending on the flowering time (Tiede et al., 2015). Backwards elimination was used to remove non-significant random effects using a p-value threshold of 0.05.

In the second step, the phenotypic best linear unbiased estimates (BLUEs) for each genotype in each trial calculated using Eq. 3.1 were used as the response in the following mixed model:

$$y_{ij} = \mu + g_i + t_j + (gt)_{ij} + \epsilon_{ij}, \quad (3.2)$$

where y_{ij} was the mean of i th genotype in the j th environment, g_i was the fixed effect of the i th genotype, t_j the fixed effect of the j th environment, $(gt)_{ij}$ the random interaction effect of the i th genotype and j th environment, and ϵ_{ij} the associated error. To account for differences in precision, the error was modelled as $\epsilon \sim N(0, \mathbf{R}V_\epsilon)$, where \mathbf{R} was a diagonal matrix containing the inverse of the variances of the genotype means estimated in the first step (Mohring and Piepho, 2009). All models were fitted using the *lme4* package in R (Bates et al., 2015). To avoid “double-shrinkage” of values (Piepho et al., 2012b), BLUEs for each genotype across all environments were calculated using Eq. 3.2 and subsequently used as phenotypes in predictions.

Marker data was generated using genotyping-by-sequencing (GBS; Elshire et al., 2011). Using a custom pipeline (Neyhart et al., 2019) we discovered 235,216 single nucleotide polymorphism (SNP) markers, pared to 6,361 markers after filtering on a minimum minor allele frequency of 0.05 and maximum missing data proportion of 80%. Additionally, lines were

filtered on a maximum missing marker data proportion of 80% missing marker data. Marker genotypes were imputed using the expectation maximization algorithm implemented in the *rrBLUP* package in R (Endelman, 2011). While the physical genomic positions of these markers were determined during read-alignment to the barley reference genome (Mascher et al., 2017), genetic positions are needed to model the segregation of markers in simulated bi-parental populations. To infer genetic map positions, we first obtained physical and genetic map information on 3,072 array-based SNPs developed during the Barley Coordinated Agricultural Project (BCAP) (Close et al., 2009). This marker set and genetic map (Muñoz-Amatriaín et al., 2011) have been used in studies of genomewide selection, association mapping, and population genetics of North American barley germplasm (e.g. Lorenz et al., 2012; Pauli et al., 2014; Poets et al., 2016). Since the TP was selected from the larger collection of two-row BCAP germplasm, we suspected that the genetic map would closely approximate true recombination frequencies. Next, local recombination rates, in cM Mb⁻¹, were calculated using the genetic and physical positions of the 3,072 SNPs, and rates were smoothed using a loess operation (Thomas Kono, pers. comm.). Finally, the genetic map positions of GBS markers were determined by linear interpolation, using the physical map position as a predictor and the smoothed local recombination rate as the estimated slope parameter.

Cross selection, population development, and phenotyping

Simulated bi-parental populations were used to inform the selection of parent combinations among a pool of 813 candidate lines. These lines were derived from crosses made between individuals in the TP (described in Neyhart et al., 2019) and were genotyped for the same 6,361 GBS markers. We used the PopVar package in R (Mohammadi et al., 2015) to predict the mean, variance, and superior progeny mean (denoted $\hat{\mu}$, \hat{V}_G , and $\hat{\mu}_{sp}$, respectively) for each of $C(813,2) = 330,078$ possible non-reciprocal parent combinations. First, phenotypic and genotypic data on the TP were first used to predict genomewide marker effects using the

following ridge-regression best linear unbiased prediction (RR-BLUP) model:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad (3.3)$$

where \mathbf{y} was a 175×1 vector of genotype BLUEs for the TP, $\mathbf{1}$ was a 175×1 vector of 1s, μ was the grand mean, \mathbf{Z} was a $175 \times 6,361$ matrix of marker genotypes coded as 1, 0, and -1 (homozygous for the reference allele, heterozygous, homozygous for the alternate allele, respectively), \mathbf{u} was a $6,361 \times 1$ vector of marker effects, and \mathbf{e} was a 175×1 vector of residuals. Next, the provided genetic map was used to simulated marker genotypes for individuals in a recombinant inbred line (RIL) family. The predicted genotypic values (PGVs) of these individuals were calculated using the simulated marker genotypes and the predicted marker effects. Markers were assumed to act additively, while dominance (not present due to complete inbreeding) and epistasis were assumed absent. Finally, the mean, variance, and mean of the bottom 10% of PGVs (lower trait values are favorable) were calculated as predictions of $\hat{\mu}$, $V_{\hat{g}}$, and $\hat{\mu}_{sp}$, respectively.

For each parent combination, we simulated families of 150 individuals 25 times, and predictions of each parameter were averaged over replicates. The number of replications was informed from previously published uses of the *PopVar* package (Mohammadi et al., 2015; Tiede et al., 2015). We also found little difference between predicted values averaged over 25 replications versus those averaged over 100 replications (data not shown). Fewer replications reduces the computational burden of the software, which can be immense. On a laptop computer equipped with 16GB of RAM and a single processor, we generated predictions of three traits for 25 crosses in ~360 seconds (0.1 hours). Predictions for all 330,078 crosses were completed after 32 hours on a supercomputing cluster equipped with 62 GB of RAM and parallelized across 24 processors.

A total of 27 crosses were selected using the predictions of all possible parent combi-

nations. Eighteen crosses were selected based on predictions of FHB severity, heading date, and plant height; these crosses were of interest to the breeder and represented a range of genetic variances for each trait. An additional 9 crosses were selected based on predictions of traits not validated in this study. Selected parent combinations were mated, and progeny were self-pollinated via single-seed descent to at least the $F_{3.5}$ stage and then given line designations. The final number of lines in each family ranged from 28 to 160, with a median of 90. Eleven of the families were originally intended for yield trials and were smaller (28-74 lines), while remaining families were larger (83-160 lines). No intentional selection was imposed during population development. In total, 2,661 lines were developed across the 27 families, which we will subsequently refer to as “validation families.”

The validation families were phenotyped in Crookston, MN and St. Paul, MN in 2017 and 2018. Ten of the families were planted exclusively in FHB disease trials, 12 in dryland trials, and five in both. The 12 families evaluated in dryland trials were phenotyped in 2017 and remaining families were phenotyped in both 2017 and 2018. All trials were planted in incomplete block designs as described above. Disease trials were again inoculated via grain-spawn (Crookston) or direct spray (St. Paul) and mist-irrigation was used to promote disease development. In the disease trials, heading date, plant height, and FHB severity were measured, and the same traits, except FHB severity, were measured in the dryland trials. When spray-inoculating disease and scoring severity in St. Paul, we created “inoculation sets” composed of lines at a similar developmental stage. This was done because of the large range in observed heading date (about 30 days) and the known association between heading date and disease severity (Mesfin et al., 2003). These sets were independently inoculated and scored. Though similar sets were not created in Crookston for inoculation, we grouped lines by maturity (i.e. early and late) when rating disease.

Statistical analysis

Phenotypic analysis followed the same two-step procedure described above, with noted changes. The analysis of per-trial and per-trait genotype means was modeled using Eq. 3.1; however, when analyzing FHB severity, we modified Eq. 3.1 such that:

$$y_{iklmn} = \mu + G_i + r_k + c_l + b_m + IS_n + (bIS)_{mn} + HD_i + \epsilon_{iklm}, \quad (3.4)$$

The genotypic BLUEs from this first step were used as phenotypes in the second step. Line means estimated in two different trials in the same environment were averaged. To estimate the empirical family mean and superior progeny mean, we fitted the same model as in Eq. 3.2 on a per-family basis, where terms were the same as in Eq. 3.2, except that μ was the family mean. The superior progeny mean was obtained as the mean of the lowest 10% of genotype BLUEs calculated from this model. We obtained empirical estimates of the family genetic variance by modifying this model such that g_i was treated as a random effect with $g_i \sim N(0, V_G)$. This modification was also used to calculate broad-sense heritability on a family-mean basis. We calculated 95% confidence intervals for the estimates of family μ , V_G , and μ_{sp} using bootstrapping. For the family mean and variance, 1,000 samples were first simulated from the per-family model fit, then parameters were re-estimated from the simulated data. For the superior progeny mean, the genotype BLUEs were sampled 1,000 times with replacement, and μ_{sp} was estimated using each sample.

The accuracy of predicting the mean, genetic variance, and superior progeny mean in each family was measured by correlating the observation of each parameter with their predictions. (We will refer to such correlations as predictive ability, abbreviated as r_{MP}) One family was removed from this analysis due to unreliable marker genotype data for one parent; therefore, estimates of predictive ability were based on 14 families for FHB severity and 26 families for heading date and plant height. The significance of correlation coefficients was determined

by bootstrapping with 1,000 replicates, and a 95% confidence interval was created using the 25th lowest value (lower bound) and the 975th lowest value (upper bound).

Stronger bias in the predictions of $V_{\hat{g}}$ relative to $\hat{\mu}$ will alter what crosses are predicted to have the most desirable $\hat{\mu}_{sp}$. The statistical bias of the predicted variance was calculated as $(\bar{V}_{\hat{g}} - \bar{V}_G)/\bar{V}_G$, where \bar{V}_G was the mean, across all families, of the estimated genetic variance, and $\bar{V}_{\hat{g}}$ was the mean of the predicted genetic variance. If the bias is negative, the predicted values are on a lower scale than the observed values, and if the bias is positive, predicted values are on a higher scale than the observed values. Five families for FHB severity and four families for plant height were removed when calculating bias due to extremely low estimates of V_G . These low estimates of V_G were likely floating point errors, therefore values less than 1×10^{-7} were considered 0 for the purpose of calculating bias and families with such low estimates were removed.

Data availability

Data used in this study are publicly available from the Triticeae Toolbox (<https://triticeaetoolbox.org/barley>). Scripts to perform the analyses and generate the figures described herein are available from the GitHub repository: <https://github.com/UMN-BarleyOatSilphium/PopVarVal>. A tutorial for accessing relevant data from the Triticeae Toolbox is also available from this repository. All analyses were performed in R (v. 3.5.1; R Core Team (2018)).

RESULTS AND DISCUSSION

Training population and cross predictions

Trait heritability in the training population (TP) was moderate to high, with estimates of 0.45 for FHB severity, 0.96 for heading date, and 0.52 for plant height. Genotype means and ranges (in parentheses) among lines in the TP were 16% (5.1, 39) for FHB severity, 50 days (44, 57) for heading date, and 74 cm (60, 87) for plant height. Genetic variance was significant for all traits ($P < 6 \times 10^{-7}$; likelihood ratio test), as was genotype-environment interaction variance ($P < 0.05$), particularly for FHB severity ($P < 5 \times 10^{-18}$). The relative contribution of genotype-environment interaction and genetic variance to total phenotypic variance (respectively, in parentheses) was high for FHB severity (83%/18%), low for heading date (7.7%/89%), and intermediate for plant height (64%/36%). These results are not surprising given prior heritability estimates for these traits (Hockett and Nilan, 1985; Massman et al., 2011), and they provide context when interpreting downstream results.

The relationship between predicted $\hat{\mu}$ and $V_{\hat{g}}$, and their relative contribution to variation in $\hat{\mu}_{sp}$, will dictate the importance of $V_{\hat{g}}$ when selecting crosses. A distinctly triangular pattern was observed when comparing $\hat{\mu}$ and $V_{\hat{g}}$ (Fig. 3.1), where potential crosses with extreme $\hat{\mu}$ were accompanied by low $V_{\hat{g}}$, but crosses with an intermediate $\hat{\mu}$ were associated with higher $V_{\hat{g}}$. Among crosses with a common parent, this relationship was approximately linear, as exemplified in Fig. 3.1. This pattern, based on the idea that lines with similarly extreme phenotypes will likely share alleles at most QTL influencing a trait, was predicted in theory (Zhong and Jannink, 2007) and observed in subsequent studies (Bernardo, 2014; Mohammadi et al., 2015; Lado et al., 2017). The variation in $\hat{\mu}_{sp}$ explained by $\hat{\mu}$ (estimated using linear regression) ranged from 95% for FHB severity to 97% for plant height. Variation in $V_{\hat{g}}$ explained 6.1% (plant height) to 8.2% (heading date) of that in $\hat{\mu}_{sp}$. Combined additively, these parameters explained more than 99% of the variance in $\hat{\mu}_{sp}$ for all traits. These proportions did

not change appreciably when analyzing the predictions of only the selected crosses (data not shown). While these results are consistent with previous findings (Mohammadi et al., 2015), the relative importance of $\hat{\mu}$ was greater among the predictions in our study. This is likely due to the absence of selection prior to making predictions, which would have reduced the variance in $\hat{\mu}$ and thereby its contribution to variation in $\hat{\mu}_{sp}$ (Zhong and Jannink, 2007).

Relationship between estimated family mean and genetic variance

Though relationships between predicted family means and genetic variances were consistent across traits, different patterns emerged in the empirical estimates of these parameters. The average estimated family mean (μ) was 25% (21, 31) for FHB severity, 53 days (48, 56) for heading date, and 82 cm (73, 91) for plant height (Table 3.1). Estimates of genetic variance (V_G) were on-average 5 (%)² (0, 15) for FHB severity, 2.3 days² (4.3, 8.6) for heading date, and 6.4 cm² (0, 18.4) for plant height. V_G was significantly greater than zero in 7 of 15 families (47%) for FHB severity and 23 of 27 families (85%) for both heading date and plant height. We did observe a family with an outlying estimate of heading date V_G (8.6 days²), and this family also had the highest estimated of plant height V_G (18.4 cm²). Beyond this coincidence, there was little distinguishing this family from the others, and we have no explanation for why it might be an outlier. Per-family distributions of genotypic values (Fig. 3.2) suggested that μ was less variable than V_G for FHB severity, but that the opposite was true for heading date. For plant height, these parameters appeared to be equally variable. Indeed, the ratio of the variance in V_G to the variance in μ was large for FHB severity (3.7), smaller for heading date (0.55) and roughly even for plant height (0.98). These results are not consistent with the ratio of variance in predicted $V_{\hat{g}}$ to that in $\hat{\mu}$ among the selected crosses, which was generally much lower. We suspect that this inconsistency is likely due to differences in trait heritability and shrinkage of predicted marker effects, causing a strong negative bias in $V_{\hat{g}}$ (discussed below). Nevertheless, these results have implications for the contributions of μ and V_G to superior progeny mean

(μ_{sp}). In our case, V_G will have a greater relative impact on variability in μ_{sp} for FHB severity and less so for heading date.

The estimates of trait heritability in the validation families were similar to those in the TP. Heritability estimates across all individuals in the validation families were 0.11 for FHB severity, 0.78 for heading date, and 0.74 for plant height. Family-wise estimates of heritability mirrored these overall estimates, but some variability was present among families (Table 3.1). For FHB severity, the mean h^2 (and range) was 0.10 (0, 0.28), while that for heading date was 0.49 (0.16, 0.84) and that for plant height was 0.41 (0, 0.76). Overall, genetic and genotype-by-environment variances were significant for all traits ($P < 0.005$), and the relative contribution of these variance components was consistent with the observations in the TP. While heritability estimates for some traits were quite low (e.g. FHB severity), the consistency of these estimates across populations (training and validation) and years suggests that we may draw conclusions about these traits at-large.

Relatedness and heritability likely drove predictive ability

Estimates of predictive ability are summarized in Table 3.2) and visualized in Fig. 3.3. The predictive ability for family mean was moderate to high for all traits ($r_{MP} = 0.46-0.62$). On a per-trait basis, these measurements were consistent with the estimates of heritability in the training population and the validation families, where lower heritability corresponded to lower predictive ability, as expected from genomewide selection theory (Daetwyler et al., 2008). For all traits, the predictive ability for genetic variance was always lower than that for μ , ranging from 0.01 (FHB severity; not significant) to 0.48 (plant height). This is not unexpected, since any error associated with the predicted marker effect will more strongly influence $V_{\hat{g}}$ than $\hat{\mu}$ (Zhong and Jannink, 2007; Lado et al., 2017). These estimates did not change appreciably when removing families with very low estimates of V_G (data not shown). Interestingly, the predictive abilities of μ_{sp} for FHB severity ($r_{MP} = 0.69$) and plant height ($r_{MP} = 0.62$) were greater than

those for μ . We might expect that the predictive ability for μ_{sp} would be intermediate of that for μ and V_G (i.e. the pattern observed for heading date), given the relative contributions of the mean and variance to the superior progeny mean, along with the unequal impact of prediction error described above. It is worth noting here the small sample size used to estimate predictive ability ($n = 14$ or 26). This will naturally lead to higher standard error estimates and larger confidence intervals for these coefficients (Table 3.2). Thus, this discrepancy could simply be due to sampling. Overall, these results are consistent with the understood impact of heritability on prediction accuracy.

In addition to heritability, other well-known drivers of prediction accuracy may be influencing our results. The size of our training population, at 175 lines, is arguably small for genomewide predictions. While studies in barley have consistently reported a pattern of diminishing returns in prediction accuracy with increasing training population size (Lorenz et al., 2012; Lorenz and Smith, 2015; Sallam et al., 2015), simulations have indicated that training population size may be particularly important when predicting V_G (Lehermeier et al., 2017). To our advantage, however, is the close relatedness of the training population to the validation families, the former being grandparents of the latter. Close relatedness is imperative for prediction accuracy (Habier et al., 2007), and genomewide selection research in barley has confirmed this (Lorenz et al., 2012; Lorenz and Smith, 2015). The structured pedigree relationships between the training and validation populations would have permitted alternative methods of predicting genetic variance, such as those that rely on relationship-based distance; however, previous research in a similar barley breeding program demonstrated such methods are inferior to the marker-based method used in this study (Tiede et al., 2015). Nevertheless, we note that the pedigree structure in our populations may be more reflective of realistic breeding program conditions, lending weight to the practicality of our findings.

Other factors influencing predictive ability may be unique to the circumstances of predicting V_G . First, the additive genetic variance in a bi-parental family is expected to increase

with successive generations of inbreeding (Bernardo, 2010). Our predictions assumed families of completely inbred recombinants, while the validation families in our study were, as per typical breeding procedures, inbred to at least the F_4 . It is possible that incomplete inbreeding led to deviations from the final additive genetic variance in families, reducing the predictive ability. Further, V_G could be upwardly biased due to dominance variance contributed by heterozygous loci, and such bias may be uneven across validation families. Second, error in the genetic map may have introduced additional bias in the predictions of covariance between loci, which would be upwardly biased if the true recombination frequency was greater than that implied by the genetic map, and vice versa. The impact of this error would likely depend on the genetic architecture. Predictions for simpler traits controlled by fewer and potentially larger-effect QTL, such as heading date and plant height, may be less severely affected, since the additive variance at each individual locus will account for most of the total genetic variance (Zhong and Jannink, 2007). In contrast, the covariance between loci may be more important for complex traits, such as FHB severity, that are influenced by many loci of small effect distributed more diffusely throughout the genome (Lado et al., 2017).

Implications of bias

An important consideration for selecting crosses based on the predictions of μ_{sp} is the bias in predicting V_G (Lian et al., 2015; Tiede et al., 2015). We found that predictions of $V_{\hat{g}}$ for all traits were associated with a strong negative bias, estimated at -95% for FHB severity, -83% for heading date, and -96% for plant height (Table 3.2). Among traits, these percentages are consistent with expectations given the estimates of heritability, where more heritable traits should suffer less bias. Within a given trait, the bias in $V_{\hat{g}}$ was relatively consistent across families (Fig. 3.4). For FHB severity, 8 of 9 families exhibited bias estimates between -90% to -98%, and for plant height, 19 of 22 families ranged in bias from -90% to -99%. For heading date, this range was larger, with 21 of 26 families exhibiting bias of -75% to -97%. Beyond trait

heritability, the estimates of bias we observed are likely due to our small training population size. Simulations have shown that larger training populations can help yield more unbiased predictions of genetic variance, particularly when the trait heritability is high (Lehermeier et al., 2017).

In addition to accuracy, the bias in the predictions of V_G is important for effectively discriminating among crosses. Negative estimates of bias have been reported in previous validation studies (Lian et al., 2015; Tiede et al., 2015), and this has been attributed to the shrinkage of marker effects towards zero when heritability is < 1 (Lian et al., 2015). The apparent association of heritability with bias was also reported by Tiede et al. (2015), albeit for a single trait, but it suggests that more precise training data should help mitigate against this problem. Though Lian et al. (2015) recommended against the use of genomewide markers and simulated populations to predict V_G out of concern for bias, it is worth noting that as long as it is consistent (and prediction accuracy is sufficient), bias will not change the ranking of potential crosses based on predicted genetic variance or superior progeny mean (Lehermeier et al., 2017). Given the relative consistency in bias across families in our analysis, we do not think this prediction method should be abandoned. Instead, as with any other implementation of genomewide selection, breeders should carefully consider whether this method will be effective given their trait of interest and the reliability of phenotypic data.

Confounding effects cloud the predictive ability of FHB severity

The separation of training and validation populations by environment may introduce unaccountable genotype-environment interactions and therefore reduce the heritability and predictive ability of a trait (Bernardo, 2010; Lian et al., 2015). We highlight further analysis of FHB severity, which may also be influenced by location-specific effects due to management or inoculation method. We modified Eq. 3.2 and fitted a random effects model to the training population data that included the main effects of genotype, year, and location, and their two-way and three-

way interactions. The variance components of genotype, location, and genotype-year-location interaction were all significant ($P < 0.005$; likelihood ratio test); therefore, we assessed the predictive ability of μ , V_G and μ_{sp} using training and validation data separated by location.

This analysis revealed sharp differences in heritability and predictive ability. While genotype-environment (i.e. genotype-year) interaction variance was significant within the TP in both locations ($P < 1 \times 10^{-5}$; likelihood ratio test), genetic variance was only significant in Crookston, a result apparent in the estimates of heritability ($h^2 = 0.46$, Crookston; $h^2 = 0.19$, St. Paul). This trend was also observed when analyzing data from the validation families ($h^2 = 0.27$, Crookston; $h^2 = 0.042$, St. Paul). As expected, predictions using data from Crookston were more accurate than those using data from St. Paul (Fig. 3.5). This difference was quite profound; the predictive ability of μ , V_G and μ_{sp} was 0.55, 0.67, and 0.35, respectively, when using Crookston data, but dropped to 0.08, -0.02, and 0.09, respectively, when using St. Paul data. This drastic difference is undoubtedly the result of more precise phenotypic data in Crookston, yet the cause of this precision is unknown.

It is possible that the inoculation method in Crookston confounded disease severity with heading date, increasing the observed heritability and predictive ability. In this location, grain-spawn inoculation is used, where infected grain is spread along the field rows, the pathogen sporulates from this medium, and infection is promoted. The inoculum is first applied when the earliest genotypes in the trial reach the heading stage, potentially allowing genotypes with delayed heading to escape the heaviest inoculum load. Our data suggest this may be the case, as we observed a stronger negative correlation between FHB severity and heading date in Crookston (-0.60) than in St. Paul (-0.23) for the training population; this trend was also observed in the validation families (-0.51, Crookston; 0.016, St. Paul). Alternatively, the lower heritability in St. Paul may be due to differences in environmental conditions following inoculation events, resulting in inconsistent disease development (Mesfin et al., 2003). These results serve as a reminder that the context and complexity of a trait should be considered prior to employing

genomewide selection.

Conclusions

We have shown that, under typical breeding program conditions, the mean, genetic variance, and superior progeny mean of potential crosses can be accurately predicted using genomewide markers and simulated populations. The favorable results observed for three relevant quantitative traits in barley indicate that this prediction method may be generalized across traits of varying complexity, though we note that heritability, a usual suspect influencing prediction accuracy, is again crucial for accurate predictions of these parameters in potential crosses. Though previous validation efforts have indicated mixed support for this method (Lian et al., 2015; Tiede et al., 2015), they relied on historical breeding data. Recently, Adeyemo and Bernardo (2019) used a diverse maize training population to predict V_G in subsequently developed populations, however they were unable to validate predictions for three focal traits in a set of 8 families. Our study adds to this work by validating predictions using a larger sample size under chronologically realistic breeding program conditions, where predictions are generated using a dedicated training population and families are subsequently developed.

The application of this method in a breeding program may depend on the objectives of the breeder or the age of the breeding program. In a young breeding program, populations may require improvement via recurrent selection, and choosing parent combinations simply based on $\hat{\mu}$ may be sufficient given the large contribution of variation in $\hat{\mu}$ to that in $\hat{\mu}_{sp}$. Alternatively, in a mature breeding program where parents are identified for cultivar development, selection on $\hat{\mu}_{sp}$ may be advantageous since the variation in $\hat{\mu}$ among potential crosses may be narrow. However, when variation in $\hat{\mu}$ is low, strong bias in the predictions of $V_{\hat{g}}$ will more severely limit the variation in $\hat{\mu}_{sp}$. At this point, care must be taken to ensure precise phenotypic data and an adequate training population size to help mitigate this risk. Finally, we emphasize that this method is computationally inexpensive and relies on data that is likely already available

to a breeder employing genomewide selection. Predictions of genetic variance and superior progeny mean, like standard genomewide predictions, can serve as an additional tool for breeders to make decisions.

FIGURES

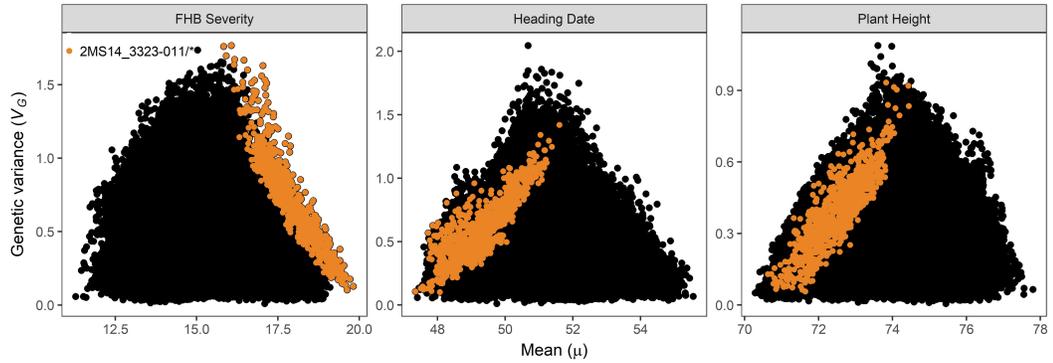


Figure 3.1: Across all 330,078 possible bi-parental crosses and the three focal traits, the relationship between predicted family mean and genetic variance formed a triangular pattern. This pattern is composed of many linear relationships within crosses sharing a common parent. Crosses with an example parent in common are highlighted.

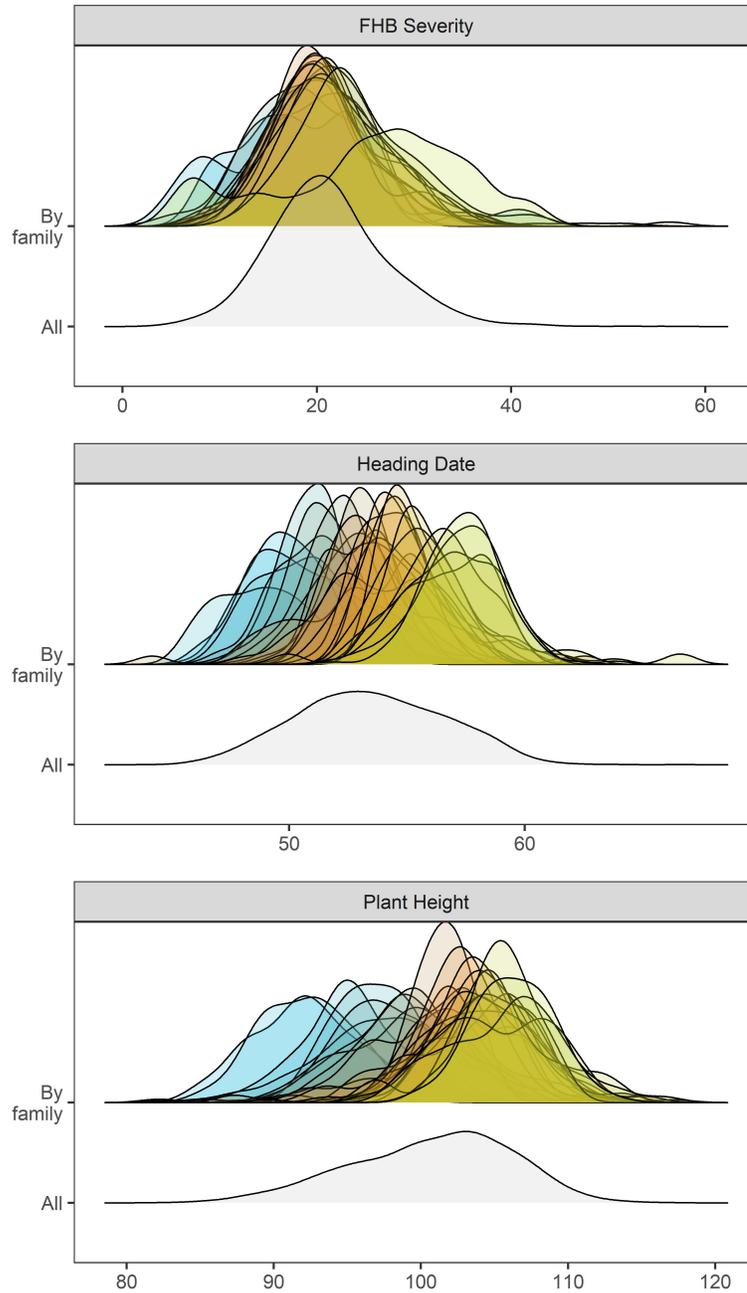


Figure 3.2: The distributions of genotype means in the validation families (shaded from blue to green to yellow in order of ascending mean) varied in their patterns across traits. For Fusarium head blight (FHB) severity, ($N_f = 14$ families), families appeared to differ more in variances than in mean, but this pattern was reversed for heading date ($N_f = 26$ families). Family mean and variance appeared to be equally variable for plant height ($N_f = 26$ families). The distribution across all individuals in the validation families (shaded in grey) may mask the patterns in the family-specific distributions.

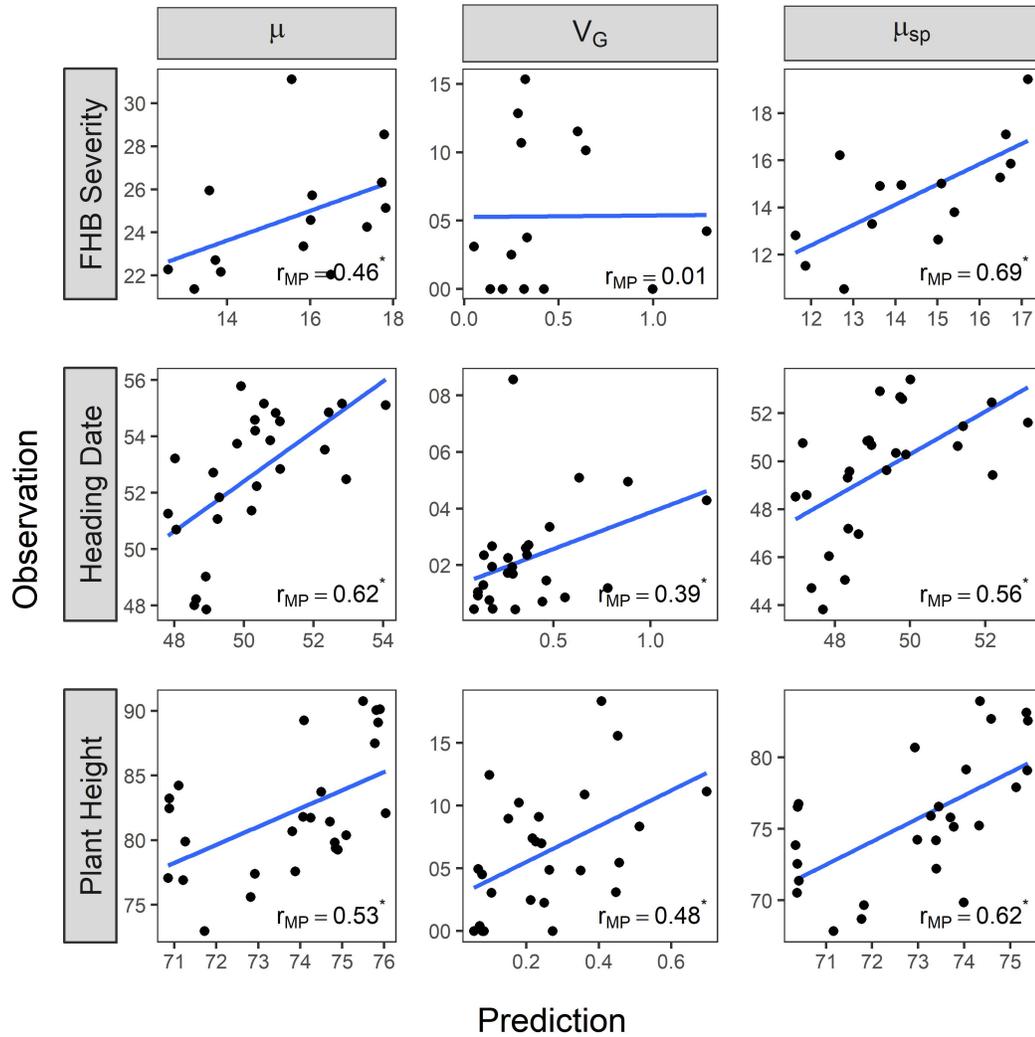


Figure 3.3: The estimates of predictive ability (r_{MP}) for family mean, genetic variance, and superior progeny mean were variable across three traits: Fusarium head blight (FHB) severity ($N_f = 14$ families), heading date ($N_f = 26$ families), and plant height ($N_f = 26$ families). Each point depicts the value of a parameter for a single validation family, and the line depicts a fitted linear regression line. Asterisks indicate that estimates of r_{MP} are significant at the 0.05 significant level (1,000 bootstrapping samples).

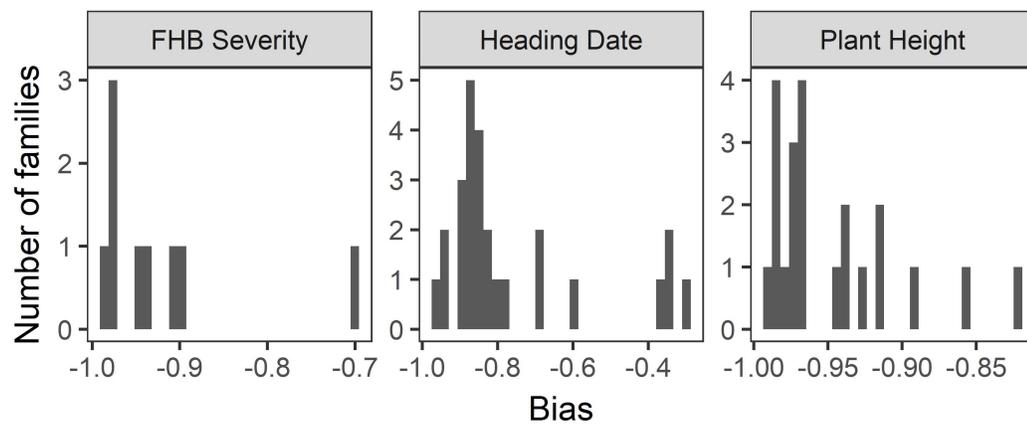


Figure 3.4: The bias in predicting genetic variance was relatively consistent across the validation families for a given trait. Families were removed if the observed estimate of genetic variance was very low ($< 1e-7$) and likely a floating point error: 5 families were removed for Fusarium head blight (FHB) severity and 4 families were removed for plant height.

TABLES

Table 3.1: Number of phenotyped families, and the mean (and range) of the observed family means (μ), genetic variances (V_G), superior progeny means (μ_{sp}), and heritabilities (h^2) among the validation families.

Trait	N_f	μ	V_G (units ²)	μ_{sp}	h^2
FHB [†] Severity	14	25 (21, 31)	4.9 (4.9e-14, 15)	15 (11, 19)	0.1 (1.5e-15, 0.28)
Heading Date	26	52 (48, 56)	2.3 (0.43, 8.6)	49 (44, 53)	0.49 (0.12, 0.84)
Plant Height	26	82 (73, 91)	6.4 (0, 18)	75 (68, 84)	0.41 (0, 0.76)

[†]FHB Fusarium head blight

Table 3.2: Estimates of predictive ability (r_{MP}) (and 95% confidence interval) for the family mean, genetic variance, and superior progeny mean, along with the average prediction bias of genetic variance.

Trait	Predictive ability (r_{MP})			Mean $V_{\hat{g}}$ Bias (%)
	$r(\mu, \hat{\mu})$	$r(V_G, V_{\hat{g}})$	$r(\mu_{sp}, \hat{\mu}_{sp})$	
FHB [†] Severity	0.46 (0.07, 0.85)	0.01 (-0.36, 0.56)	0.69 (0.28, 0.89)	-95
Heading Date	0.62 (0.45, 0.76)	0.39 (0.03, 0.77)	0.56 (0.38, 0.73)	-83
Plant Height	0.53 (0.26, 0.74)	0.48 (0.18, 0.7)	0.62 (0.39, 0.8)	-96

[†]FHB Fusarium head blight

SUPPLEMENTAL FIGURES

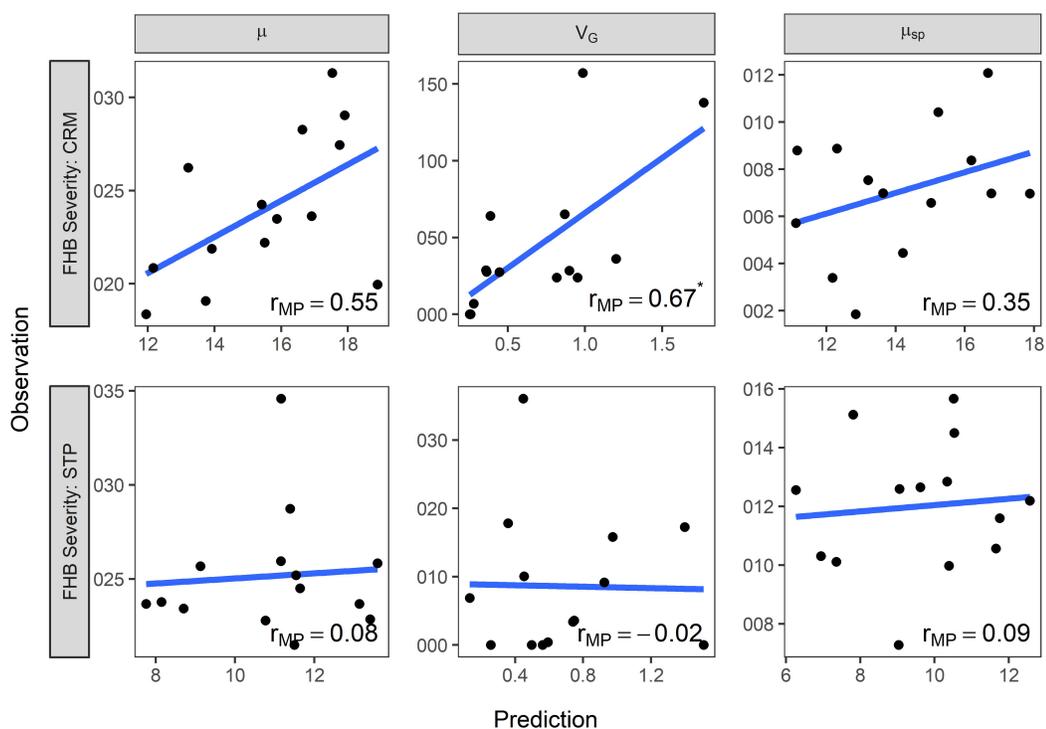


Figure 3.5: Predictions and observations of the family mean (μ), genetic variance (V_G), and superior progeny mean (μ_{sp}) for Fusarium head blight (FHB) severity among 14 families using data from Crookston, MN (CRM) or St. Paul, MN (STP). The predictive ability (r_{MP}) measures the correlation between the predictions and the observations, and is denoted by the regression line. *Significant at the 0.05 probability level (bootstrapping, 1,000 samples).

SUPPLEMENTAL TABLES

Table 3.3: Validation family number, parents, number of individuals (N_{ind}), number of phenotyping environments (N_e), and heritability (h^2) estimates for Fusarium head blight (FHB) severity, heading date, and plant height. Very low estimates of heritability (i.e. $< 1e-7$) were considered floating point errors and given a value of 0. Blanks indicate that the family was not phenotyped for that trait.

Family	Parent 1	Parent 2	N_{ind}	N_e	h^2		
					FHB Severity	Heading Date	Plant Height
4130	2MS14-3301-009	2MS14-3312-009	160	4	0	0.72	0.360
4161	2MS14-3307-008	2MS14-3307-024	160	4	0.1	0.56	0.340
4138	2MS14-3301-024	2MS14-3323-011	155	2		0.31	0.540
4135	2MS14-3301-021	2MS14-3335-014	154	4	0	0.75	0.660
4174	2MS14-3340-025	2MS14-3340-030	154	4	0.09	0.58	0.057
4166	2MS14-3323-001	2MS14-3323-011	147	2		0.29	0.450
4133	2MS14-3301-017	2MS14-3323-023	144	2		0.37	0.520
4145	2MS14-3304-030	2MS14-3307-024	131	4	0.061	0.67	0.570
4129	2MS14-3301-008	2MS14-3335-018	118	4	0	0.75	0.760
4142	2MS14-3304-013	2MS14-3307-006	118	4	0.22	0.66	0.420
4159	2MS14-3306-018	2MS14-3336-003	118	4	0.24	0.80	0.630
4175	2MS14-3323-011	2MS14-3341-006	100	4	0.25	0.84	0.740
4134	2MS14-3301-017	2MS14-3339-018	98	2		0.34	0.290
4143	2MS14-3304-013	2MS14-3307-024	90	4	0.28	0.68	0.620
4132	2MS14-3301-017	2MS14-3321-017	86	2		0.24	0.430
4131	2MS14-3301-015	2MS14-3323-011	83	2		0.45	0.300
4146	2MS14-3305-009	2MS14-3305-012	74	2		0.12	0.000
4150	2MS14-3305-009	2MS14-3305-023	71	2		0.23	0.000
4158	2MS14-3306-010	2MS14-3345-025	71	4	0	0.44	0.190
4151	2MS14-3305-009	2MS14-3305-027	63	2		0.12	0.550
4152	2MS14-3305-027	2MS14-3345-022	63	4	0.22	0.34	0.000
4179	2MS14-3306-022	2MS14-3346-008	63	4	0.053	0.63	0.660
4140	2MS14-3303-007	2MS14-3303-013	58	4	0	0.31	0.000
4139	2MS14-3302-026	2MS14-3305-009	57	2		0.44	0.320
4155	2MS14-3302-026	2MS14-3306-005	53	2		0.54	0.700
4154	2MS14-3306-002	2MS14-3346-002	44	4	0	0.79	0.480
4148	2MS14-3302-026	2MS14-3305-012	28	2		0.22	0.520

Epilogue: Future Research Directions

Research projects rarely answer their intended questions without generating more, and the aforementioned studies are no exception. Below, I describe several additional lines of inquiry that extend from my work.

Chapter 1

1. Recent work using empirical data offers support for the “Tails” method (Brandariz and Bernardo, 2018; Tiede and Smith, 2018), where a selection of lines that are predicted to be poorest is used to update the training population. A plant breeder might hesitate to implement this method in practice, as it would require a sacrifice of valuable field space for phenotyping lines that will never become cultivars. It would be beneficial to understand the theoretical reason why this method leads to more accurate predictions; this may lead to modifications of this method that may be more practical to a breeder.

Chapter 2

1. A notable body of research addressing genotype-environment interactions (GEIs) in the context of genomewide selection has focused on testing models that may or may not account for GEIs (e.g. Jarquín et al., 2014; Pérez-Rodríguez et al., 2015; Saint Pierre et al., 2016; Jarquin et al., 2017; Morais Júnior et al., 2018). Generally, the predictive utility of these models is tested by cross-validation, which may lead to inflated prediction accuracies if the training population is composed of a common population of related genotypes. Parent-offspring validation, and arguably more useful assessment of predictions, have not been tested in this model-building framework. This would be a natural extension of work in Chapter 2.

2. Only three traits - heading date, grain yield, and plant height - were tested in this study, as there were the most frequently measured (36-40 environments); however, another four traits were measured in at least nine environments, and many of these traits are related to malting quality. The analysis in Chapter 2 could be extended to these traits as well.
3. The training population was pulled from a larger collection of Barley Coordinated Agricultural Project germplasm. As such, both this population and the larger collection were genotyped with a 3,072-marker SNP array (Close et al., 2009). It would be possible, then, to use the extensive phenotypic data on the training population to make “backwards” predictions of the remaining two-row germplasm. Then, samples from this germplasm based on predictions could be phenotyped and those predictions validated. As a research topic, it would be useful to know the accuracy of these predictions, and while this topic has been addressed previously (see Yu et al., 2016, and Crossa et al. (2016)), the target germplasm has generally been restricted to unadapted landraces or breeding material. Practically, it is possible that there are other BCAP genotypes that are better-suited for the Minnesota breeding program, and such predictions may identify those genotypes that were overlooked.
4. The highly balanced nature of the dataset could allow for other multi-environment analysis, such as association mapping for environmentally-sensitive quantitative trait loci (QTL). One drawback, of course, would be the small population size ($n = 233$).

Chapter 3

1. The choice of traits for validating predictions of genetic variance was informed by logistical constraints (i.e. labor, field space, etc.). Higher-priority traits such as grain yield, malting quality, and deoxynivalenol content (a mycotoxin produced by *Fusarium graminearum*) would benefit from accurate predictions of genetic variance and the superior progeny mean. These traits can be resource-intensive (grain yield requires much

field space and malting quality data is expensive to obtain); it may be ideal to validate predictions on few families (~8-10) and use alternative measures of accuracy beyond the correlation coefficient.

2. This study looked only at the traditional bi-parental mating scheme used in the University of Minnesota barley breeding program. The program has typically eschewed more complicated crosses, such as multi-parent populations, given the longer time period necessary to develop them. Predictions of genetic variance could theoretically be extended to multi-parent populations, and it may be useful to determine 1) whether the accuracy of these predictions is similar to those in bi-parental populations, and 2) if informed selection of multi-parent crosses can justify the longer time requirements for their development.
3. The populations and data collected in this study could also be used to compare or validated other cross selection methods, such as the expected maximum haploid value (Müller et al., 2018), optimum haploid value (Daetwyler et al., 2015), predicted cross value (Han et al., 2017), and optimal cross selection (Gorjanc et al., 2018).

Bibliography

- Adeyemo, E. and R. Bernardo. 2019. Predicting Genetic Variance from Genomewide Marker Effects Estimated from a Diverse Panel of Maize Inbreds. *Crop Science* 59:1–8. doi:10.2135/cropsci2018.08.0525.
- Akdemir, D., J. I. Sanchez, and J.-L. Jannink. 2015. Optimization of genomic selection training populations with a genetic algorithm. *Genetics Selection Evolution* 47(1):38. doi:10.1186/s12711-015-0116-6. URL <http://www.gsejournal.org/content/47/1/38>.
- Allard, R. W. and a. D. Bradshaw. 1964. Implications of Genotype-Environmental Interactions in Applied Plant Breeding. *Crop Science* 4(5):503. doi:10.2135/cropsci1964.0011183X000400050021x.
- Anderson, J. E., T. J. Y. Kono, R. M. Stupar, M. B. Kantar, and P. L. Morrell. 2016. Environmental Association Analyses Identify Candidates for Abiotic Stress Tolerance in Glycine soja, the Wild Progenitor of Cultivated Soybeans. *G3: Genes|Genomes|Genetics* 6(4):835–843. doi:10.1534/g3.116.026914. URL <http://www.g3journal.org/content/6/4/835.abstract>.
- Asoro, F. G., M. A. Newell, W. D. Beavis, M. P. Scott, and J.-L. Jannink. 2011. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *The Plant Genome* 4(2):132–144. doi:10.3835/plantgenome2011.02.0007.
- Asoro, F. G., M. A. Newell, W. D. Beavis, M. P. Scott, N. A. Tinker, and J. L. Jannink. 2013. Genomic, marker-assisted, and pedigree-BLUP selection methods for beta-glucan concentration in elite oat. *Crop Science* 53(5):1894–1906. doi:10.2135/cropsci2012.09.0526.
- Atlin, G. N., R. J. Baker, K. B. McRae, and X. Lu. 2000. Selection response in subdivided target regions. *Crop Science* 40:7–13. doi:10.2135/cropsci2000.4017.
- Auinger, H. J., M. Schönleben, C. Lehermeier, M. Schmidt, V. Korzun, H. H. Geiger, H. P. Piepho, A. Gordillo, P. Wilde, E. Bauer, and C. C. Schön. 2016. Model training across multiple breeding cycles significantly improves genomic prediction accuracy in rye (*Secale cereale* L.). *Theoretical and Applied Genetics* 129:2043–2053. doi:10.1007/s00122-016-2756-5.
- Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software* 67(1). doi:10.18637/jss.v067.i01.
- Bernardo, R. 2010. *Breeding for Quantitative Traits in Plants*. Stemma Press, Woodbury, Minnesota, 2 edition.

- Bernardo, R. 2014. Genomewide Selection of Parental Inbreds: Classes of Loci and Virtual Biparental Populations. *Crop Science* 54(6):2586. doi:10.2135/cropsci2014.01.0088. URL <https://www.crops.org/publications/cs/abstracts/54/1/68><https://www.crops.org/publications/cs/abstracts/54/6/2586>.
- Bernardo, R. and J. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. *Crop Science* 47(3):1082–1090. doi:10.2135/cropsci2006.11.0690.
- Beyene, Y., K. Semagn, S. Mugo, A. Tarekegne, R. Babu, B. Meisel, P. Sehabiague, D. Makumbi, C. Magorokosho, S. Oikeh, J. Gakunga, M. Vargas, M. Olsen, B. M. Prasanna, M. Banziger, and J. Crossa. 2015. Genetic Gains in Grain Yield Through Genomic Selection in Eight Bi-parental Maize Populations under Drought Stress. *Crop Science* 55:154–163. doi:10.2135/cropsci2014.07.0460. URL <https://dl.sciencesocieties.org/publications/cs/abstracts/55/1/154>.
- Blake, V. C., C. Birkett, D. E. Matthews, D. L. Hane, P. Bradbury, and J.-L. Jannink. 2016. The Triticeae Toolbox: Combining Phenotype and Genotype Data to Advance Small-Grains Breeding. *The Plant Genome* 9(2):1–10. doi:10.3835/plantgenome2014.12.0099.
- Bohn, M., H. F. Utz, and A. E. Melchinger. 1999. Genetic Similarities among Winter Wheat Cultivars Determined on the Basis of RFLPs, AFLPs, and SSRs and Their Use for Predicting Progeny Variance. *Crop Science* 39(1):228. doi:10.2135/cropsci1999.0011183X003900010035x. URL <https://www.crops.org/publications/cs/abstracts/39/1/CS0390010228>.
- Brandariz, S. P. and R. Bernardo. 2018. Maintaining the accuracy of genomewide predictions when selection has occurred in the training population. *Crop Science* xx(xx):1–6. doi:10.2527/jas.2016-0672.
- Brewers Association. 2018. State Craft Beer Sales & Production Statistics. URL <https://www.brewersassociation.org/statistics/by-state/>.
- Buckler, E. S., J. B. Holland, P. J. Bradbury, C. B. Acharya, P. J. Brown, C. Browne, E. Ersoz, S. Flint-Garcia, A. Garcia, J. C. Glaubitz, M. M. Goodman, C. Harjes, K. Guill, D. E. Kroon, S. Larsson, N. K. Lepak, H. Li, S. E. Mitchell, G. Pressoir, J. A. Peiffer, M. O. Rosas, T. R. Rocheford, M. C. Romay, S. Romero, S. Salvo, H. Sanchez Villeda, H. S. da Silva, Q. Sun, F. Tian, N. Upadyayula, D. Ware, H. Yates, J. Yu, Z. Zhang, S. Kresovich, and M. D. McMullen. 2009. The genetic architecture of maize flowering time. *Science* 325:714–718. doi:10.1126/science.1174276. URL <http://www.ncbi.nlm.nih.gov/pubmed/19661422>{%}5Cn<http://www.sciencemag.org/cgi/doi/10.1126/science.1174276>.
- Burgueño, J., G. de los Campos, K. Weigel, and J. Crossa. 2012. Genomic prediction of breeding values when modeling genotype x environment interaction using pedigree and dense molecular markers. *Crop Science* 52(2):707–719. doi:10.2135/cropsci2011.06.0299.

- Burgueño, J., J. Crossa, J. M. Cotes, F. S. Vicente, and B. Das. 2011. Prediction Assessment of Linear Mixed Models for Multienvironment Trials. *Crop Science* 51(3):944. doi:10.2135/cropsci2010.07.0403. URL <https://www.crops.org/publications/cs/abstracts/51/3/944>.
- Bustos-Korts, D., M. Malosetti, S. Chapman, B. Biddulph, and F. V. Eeuwijk. 2016a. Improvement of predictive ability by uniform coverage of the target genetic space. *G3: Genes|Genomes|Genetics* 6(September):3733–3747. doi:10.1534/g3.116.035410.
- Bustos-Korts, D., M. Malosetti, S. Chapman, and F. A. van Eeuwijk. 2016b. Modelling of Genotype by Environment Interaction and Prediction of Complex Traits across Multiple Environments as a Synthesis of Crop Growth Modelling, Genetics and Statistics. In X. Yin and P. C. Struik, editors, *Crop Systems Biology*, pages 55–82. Springer, Cham, Switzerland. ISBN 978-3-319-20561-8. doi:10.1007/978-3-319-20562-5. URL <http://link.springer.com/10.1007/978-3-319-20562-5>.
- Close, T. J., P. R. Bhat, S. Lonardi, Y. Wu, N. Rostoks, L. Ramsay, A. Druka, N. Stein, J. T. Svensson, S. Wanamaker, S. Bozdog, M. L. Roose, M. J. Moscou, S. Chao, R. K. Varshney, P. Szűcs, K. Sato, P. M. Hayes, D. E. Matthews, A. Kleinhofs, G. J. Muehlbauer, J. DeYoung, D. F. Marshall, K. Madishetty, R. D. Fenton, P. Condamine, A. Graner, and R. Waugh. 2009. Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics* 10(582). doi:10.1186/1471-2164-10-582. URL <http://www.biomedcentral.com/1471-2164/10/582>.
- Combs, E. and R. Bernardo. 2013. Accuracy of Genomewide Selection for Different Traits with Constant Population Size, Heritability, and Number of Markers. *The Plant Genome* 6(1). doi:10.3835/plantgenome2012.11.0030. URL <https://www.crops.org/publications/tpg/abstracts/6/1/plantgenome2012.11.0030>.
- Cooper, M. and I. H. DeLacy. 1994. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *Theoretical and Applied Genetics* 88(5):561–572. doi:10.1007/BF01240919.
- Cros, D., M. Denis, L. Sánchez, B. Cochard, A. Flori, T. Durand-Gasselín, B. Nouy, A. Omoré, V. Pomiès, V. Riou, E. Suryana, and J.-M. Bouvet. 2015. Genomic selection prediction accuracy in a perennial crop: case study of oil palm (*Elaeis guineensis* Jacq.). *Theoretical and Applied Genetics* 128(3):397–410. doi:10.1007/s00122-014-2439-z. URL <http://link.springer.com/10.1007/s00122-014-2439-z>.
- Crossa, J., D. Jarquín, J. Franco, P. Pérez-rodríguez, J. Burgueño, C. Saint-pierre, P. Vikram, C. Sansaloni, C. Petrolí, D. Akdemir, C. Sneller, M. Reynolds, M. Tattaris, T. Payne, C. Guzman, R. J. Peña, P. Wenzl, and S. Singh. 2016. Genomic Prediction of Gene Bank Wheat Landraces. *G3: Genes|Genomes|Genetics* 6:1819–1834. doi:10.1534/g3.116.029637.

- Cuevas, J., J. Crossa, O. Montesinos-Lopez, J. Burgueno, P. Perez-Rodriguez, and G. de Los Campos. 2017. Bayesian Genomic Prediction with Genotype \times Environment Interaction Kernel Models. *G3: Genes|Genomes|Genetics* 7(January):g3.116.035584. doi:10.1534/g3.116.035584. URL <http://www.ncbi.nlm.nih.gov/pubmed/27793970>.
- Cuevas, J., J. Crossa, V. Soberanis, S. Pérez-Elizalde, P. Pérez-Rodríguez, G. de los Campos, O. A. Montesinos-López, and J. Burgueño. 2016. Genomic Prediction of Genotype \times Environment Interaction Kernel Regression Models. *The Plant Genome* August(november):1–20. doi:10.3835/plantgenome2016.03.0024.
- Daetwyler, H. D., M. J. Hayden, G. C. Spangenberg, and B. J. Hayes. 2015. Selection on optimal haploid value increases genetic gain and preserves more genetic diversity relative to genomic selection. *Genetics* 200(4):1341–1348. doi:10.1534/genetics.115.178038.
- Daetwyler, H. D., B. Villanueva, P. Bijma, and J. A. Woolliams. 2007. Inbreeding in genome-wide selection. *Journal of Animal Breeding and Genetics* 124(6):369–376. doi:10.1111/j.1439-0388.2007.00693.x.
- Daetwyler, H. D., B. Villanueva, and J. A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE* 3(10). doi:10.1371/journal.pone.0003395.
- Dawson, J. C., J. B. Endelman, N. Heslot, J. Crossa, J. Poland, S. Dreisigacker, Y. Manès, M. E. Sorrells, and J.-L. Jannink. 2013. The use of unbalanced historical data for genomic selection in an international wheat breeding program. *Field crops research* 154:12–22. doi:10.1016/j.fcr.2013.07.020. URL <http://linkinghub.elsevier.com/retrieve/pii/S0378429013002645>.
- Dekkers, J. 2007. Prediction of response to marker assisted and genomic selection using selection index theory. *Journal of Animal Breeding and Genetics* 124(6):331–341. doi:10.1111/j.1439-0388.2007.00701.x.
- Denis, M. and J. M. Bouvet. 2013. Efficiency of genomic selection with models including dominance effect in the context of Eucalyptus breeding. *Tree Genetics and Genomes* 9(1):37–51. doi:10.1007/s11295-012-0528-1.
- Dunckel, S., J. Crossa, S. Wu, D. Bonnett, and J. Poland. 2016. Genomic Selection for Increased Yield in Synthetic-Derived Wheat. *Crop Science* 0(0):0. doi:10.2135/cropsci2016.04.0209. URL <https://dl.sciencesocieties.org/publications/cs/abstracts/0/0/cropsci2016.04.0209>.
- Eberhart, S. A. and W. A. Russell. 1966. Stability Parameters for Comparing Varieties. *Crop Science* 6(3).
- van Eeuwijk, F. A., J.-B. Denis, and M. S. Kang. 1996. Incorporating Additional Information

- on Genotypes and Environments in Models for Two-Way Genotype By Environment Tables. In M. S. Kang and H. G. Gauch, editors, *Genotype-by-Environment Interaction*, pages 15–50. CRC Press, Boca Raton, FL.
- van Eeuwijk, F. A. and A. Elgersma. 1993. Incorporating environmental information in an analysis of genotype by environment interaction for seed yield in perennial ryegrass. *Heredity* 70(5):447–457. doi:10.1038/hdy.1993.66.
- Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* 6(5):e19379. doi:10.1371/journal.pone.0019379. URL <http://dx.plos.org/10.1371/journal.pone.0019379>.
- Endelman, J. B. 2011. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome* 4(3):250–255. doi:10.3835/plantgenome2011.08.0024. URL <https://www.crops.org/publications/tpg/abstracts/4/3/250>.
- Falconer, D. and T. F. Mackay. 1996. *Introduction to Quantitative Genetics*. Pearson Prentice Hall, Harlow, Essex, UK, fourth edition.
- Finlay, K. and G. Wilkinson. 1963. The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research* 14(6):742. doi:10.1071/AR9630742.
- Forkman, J. and H. P. Piepho. 2014. Parametric bootstrap methods for testing multiplicative terms in GGE and AMMI models. *Biometrics* 70:639–647. doi:10.1111/biom.12162.
- Gauch, H. G. 2013. A Simple Protocol for AMMI Analysis of Yield Trials. *Crop Science* 53:1860–1869. doi:10.2135/cropsci2013.04.0241.
- Gauch, H. G. and R. Zobel. 1997. Identifying mega environments and targeting genotypes. *Crop Science* 37(April):311–326. doi:10.2135/cropsci1997.0011183X003700020002x.
- Gauch, H. G. and R. W. Zobel. 1988. Predictive and postdictive success of statistical analyses of yield trials. *Theoretical and Applied Genetics* 76(1):1–10. doi:10.1007/BF00288824.
- Gorjanc, G., R. C. Gaynor, and J. M. Hickey. 2018. Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. *Theoretical and Applied Genetics* 131(9):1953–1966. doi:10.1007/s00122-018-3125-3. URL <https://doi.org/10.1007/s00122-018-3125-3>.
- Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177(4):2389–2397. doi:10.1534/genetics.107.081190.
- Han, Y., J. N. Cameron, L. Wang, and W. D. Beavis. 2017. The Predicted Cross Value for Ge-

- netic Introgression of Multiple Alleles. *Genetics* 205(4):1409–1423. doi:10.1534/genetics.116.197095. URL <http://www.genetics.org/lookup/doi/10.1534/genetics.116.197095>.
- Heffner, E. L., J.-L. Jannink, and M. E. Sorrells. 2011. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *The Plant Genome* 4(1):65–75. doi:10.3835/plantgenome2010.12.0029.
- Heffner, E. L., M. E. Sorrells, and J.-L. Jannink. 2009. Genomic selection for crop improvement. *Crop Science* 49(1):1–12. doi:10.2135/cropsci2008.08.0512. URL <https://www.crops.org/publications/cs/abstracts/49/1/1>.
- Henderson, C. R. 1984. *Applications of Linear Models in Animal Breeding*. University of Guelph, Guelph, 3 edition. ISBN 0889550301.
- Heslot, N., D. Akdemir, M. E. Sorrells, and J.-L. L. Jannink. 2014. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theoretical and Applied Genetics* 127(2):1–18. doi:10.1007/s00122-013-2231-5.
- Heslot, N., J.-L. Jannink, and M. E. Sorrells. 2013. Using genomic prediction to characterize environments and optimize prediction accuracy in applied breeding data. *Crop Science* 53(June):921–933. doi:10.2135/cropsci2012.07.0420.
- Hill, W. G. and A. Robertson. 1968. Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* 38(6):226–31. doi:10.1007/BF01245622. URL <http://www.ncbi.nlm.nih.gov/pubmed/24442307>.
- Hockett, E. A. and R. A. Nilan. 1985. Genetics. In D. Rasmusson, editor, *Barley*, pages 187–230. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, WI.
- Holland, J. B., W. E. Nyquist, and C. T. Cervantes-Martinez. 2003. Estimating and Interpreting Heritability for Plant Breeding. In J. Janick, editor, *Plant Breed. Rev.*, volume 22, pages 9–112. John Wiley & Sons, Inc. ISBN 9780471215417. doi:10.1002/9780470650202.ch2. URL <http://onlinelibrary.wiley.com/doi/10.1002/9780470650202.ch2/summary>.
- Isidro, J., J.-L. Jannink, D. Akdemir, J. Poland, N. Heslot, and M. E. Sorrells. 2015. Training set optimization under population structure in genomic selection. *Theoretical and Applied Genetics* 128(1):145–158. doi:10.1007/s00122-014-2418-4. URL <http://link.springer.com/10.1007/s00122-014-2418-4>.
- Iwata, H. and J.-L. Jannink. 2011. Accuracy of genomic selection prediction in barley breeding programs: A simulation study based on the real single nucleotide polymorphism data of barley breeding lines. *Crop Science* 51(5):1915–1927. doi:10.2135/cropsci2010.12.0732.

- Jannink, J.-L. 2010. Dynamics of long-term genomic selection. *Genetics Selection Evolution* 42(35). doi:10.1186/1297-9686-42-35.
- Jarquín, D., K. Kocak, L. Posadas, K. Hyma, J. Jedlicka, G. Graef, and A. Lorenz. 2014. Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC genomics* 15(1):740. doi:10.1186/1471-2164-15-740.
- Jarquín, D., C. Lemes da Silva, R. C. Gaynor, J. Poland, A. Fritz, R. Howard, S. Battenfield, J. Crossa, D. Jarquín, C. Lemes da Silva, R. C. Gaynor, J. Poland, A. Fritz, R. Howard, S. Battenfield, and J. Crossa. 2017. Increasing Genomic-Enabled Prediction Accuracy by Modeling Genotype x Environment Interactions in Kansas Wheat. *The Plant Genome* 10(2):1–15. doi:10.3835/plantgenome2016.12.0130. URL <https://dl.sciencesocieties.org/publications/tpg/abstracts/0/0/plantgenome2016.12.0130>.
- Karaman, E., H. Cheng, M. Z. Firat, D. J. Garrick, and R. L. Fernando. 2016. An upper bound for accuracy of prediction using GBLUP. *PLoS ONE* 11(8):1–18. doi:10.1371/journal.pone.0161054.
- Krchov, L.-M., G. A. Gordillo, and R. Bernardo. 2015. Multienvironment Validation of the Effectiveness of Phenotypic and Genomewide Selection within Biparental Maize Populations. *Crop Science* 55(3):1068. doi:10.2135/cropsci2014.09.0608. URL <https://dl.sciencesocieties.org/publications/cs/abstracts/55/3/1068>.
- Kumar, S., D. Chagné, M. C. a. M. Bink, R. K. Volz, C. Whitworth, and C. Carlisle. 2012. Genomic selection for fruit quality traits in apple (*Malus×domestica* Borkh.). *PLoS ONE* 7(5):1–10. doi:10.1371/journal.pone.0036674.
- Lado, B., P. G. Barrios, M. Quinke, P. Silva, and L. Gutiérrez. 2016. Modeling Genotype x Environment Interaction for Genomic Selection with Unbalanced Data from a Wheat Breeding Program. *Crop Science* 56(5):1–15. doi:10.2135/cropsci2015.04.0207.
- Lado, B., S. Battenfield, C. Guzman, M. Quinke, R. P. Singh, S. Dreisigacker, R. J. Peña, A. Fritz, P. Silva, J. Poland, and L. Gutiérrez. 2017. Strategies To Select Crosses Using Genomic Prediction in Two Wheat Breeding Programs. *The Plant Genome* 10(2):1–12. doi:10.3835/plantgenome2016.12.0128.
- Laloe, D. 1993. Precision and information in linear models of genetic evaluation. *Genetics Selection Evolution* 25(6):557–576. URL <http://www.biomedcentral.com/content/pdf/1297-9686-25-6-557.pdf>.
- Lande, R. and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124(3):743–756.
- Lehermeier, C., N. Krämer, E. Bauer, C. Bauland, C. Camisan, L. Campo, P. Flament,

- A. E. Melchinger, M. Menz, N. Meyer, L. Moreau, J. Moreno-González, M. Ouzunova, H. Pausch, N. Ranc, W. Schipprack, M. Schönleben, H. Walter, A. Charcosset, and C.-C. Schön. 2014. Usefulness of Multiparental Populations of Maize (*Zea mays* L.) for Genome-Based Prediction. *Genetics* 198(1):3–16. doi:10.1534/genetics.114.161943. URL <http://www.genetics.org/content/198/1/3.abstract>.
- Lehermeier, C., S. Teyssèdre, and C.-C. Schön. 2017. Genetic Gain Increases by Applying the Usefulness Criterion with Improved Variance Prediction in Selection of Crosses. *Genetics* 207:1651–1661. doi:10.1534/genetics.117.300403. URL <http://www.genetics.org/lookup/doi/10.1534/genetics.117.300403>.
- Li, X., T. Guo, Q. Mu, X. Li, and J. Yu. 2018. Genomic and environmental determinants and their interplay underlying phenotypic plasticity. *Proceedings of the National Academy of Sciences* 115(26):6679–6684. doi:10.1073/pnas.1718326115.
- Lian, L., A. Jacobson, S. Zhong, and R. Bernardo. 2015. Prediction of genetic variance in biparental maize populations: Genomewide marker effects versus mean genetic variance in prior populations. *Crop Science* 55(3):1181–1188. doi:10.2135/cropsci2014.10.0729.
- Lin, C.-S. and G. Poushinsky. 1985. A Modified Augmented Design (Type 2) for Rectangular Plots. *Canadian Journal of Plant Science* 65:743–749.
- Löffler, C. M., J. Wei, T. Fast, J. Gogerty, S. Langton, M. Bergman, B. Merrill, and M. Cooper. 2005. Classification of maize environments using crop simulation and geographic information systems. *Crop Science* 45(5):1708–1716. doi:10.2135/cropsci2004.0370.
- Lorenz, A. J., S. Chao, F. G. Asoro, E. L. Heffner, T. Hayashi, H. Iwata, K. P. Smith, M. E. Sorrells, and J.-L. Jannink. 2011. Genomic Selection in Plant Breeding: Knowledge and Prospects. In D. L. Sparks, editor, *Advances in Agronomy*, volume 110, pages 77–123. Elsevier Inc., 1 edition. ISBN 9780123855312. doi:10.1016/B978-0-12-385531-2.00002-5. URL <http://books.google.com/books?hl=en&lr=&id=C0sknXCy5nsC&oi=fnd&pg=PA77&dq=Genomic+Selection+in+Plant+Breeding:+Knowledge+and+Prospects&ots=zmj7tn4XJf&sig=j-PvIN3uHbTjBDUwNBSIzUojNow%5Cnhttp://linkinghub.elsevier.com/retrieve/pii/B9780123855312000025>.
- Lorenz, A. J., K. Smith, and J.-L. Jannink. 2012. Potential and Optimization of Genomic Selection for Fusarium Head Blight Resistance in Six-Row Barley. *Crop Science* 52(4):1609–1621. doi:10.2135/cropsci2011.09.0503. URL <https://www.crops.org/publications/cs/abstracts/52/4/1609>.
- Lorenz, A. J. and K. P. Smith. 2015. Adding Genetically Distant Individuals to Training Populations Reduces Genomic Prediction Accuracy in Barley. *Crop Science* 55:2657–2667. doi:10.2135/cropsci2014.12.0827.

- Lorenzana, R. E. and R. Bernardo. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theoretical and Applied Genetics* 120(1):151–161. doi:10.1007/s00122-009-1166-3.
- Malosetti, M., D. Bustos-Korts, M. P. Boer, and F. van Eeuwijk. 2016. Predicting Responses in Multiple Environments: Issues in Relation to Genotype x Environment Interactions. *Crop Science* 56:2210–2222. doi:10.2135/cropsci2015.05.0311.
- Mascher, M., H. Gundlach, A. Himmelbach, S. Beier, S. O. Twardziok, T. Wicker, V. Radchuk, C. Dockter, P. E. Hedley, J. Russell, M. Bayer, L. Ramsay, H. Liu, G. Haberer, Q. Zhang, Q. Zhang, R. A. Barrero, L. Li, S. Taudien, M. Groth, M. Felder, A. Hastie, H. Sta, J. Vr, S. Chan, R. Ounit, S. Wanamaker, D. Bolser, C. Colmsee, T. Schmutzer, L. Aliyeva, S. Grasso, J. Tanskanen, A. Chailyan, D. Sampath, D. Heavens, L. Clissold, S. Cao, B. Chapman, F. Dai, Y. Han, H. Li, X. Li, C. Lin, J. K. McCooke, C. Tan, P. Wang, S. Wang, S. Yin, G. Zhou, J. A. Poland, M. I. Bellgard, L. Borisjuk, A. Houben, J. Dole, S. Ayling, S. Lonardi, P. Kersey, P. Langridge, G. J. Muehlbauer, M. D. Clark, M. Caccamo, A. H. Schulman, K. F. X. Mayer, M. Platzer, T. J. Close, U. Scholz, M. Hansson, G. Zhang, I. Braumann, M. Spannagl, C. Li, R. Waugh, and N. Stein. 2017. A chromosome conformation capture ordered sequence of the barley genome. *Nature* 544(7651):1–43. doi:10.1038/nature22043.
- Massman, J., B. Cooper, R. Horsley, S. Neate, R. Dill-Macky, S. Chao, Y. Dong, P. Schwarz, G. J. Muehlbauer, and K. P. Smith. 2011. Genome-wide association mapping of Fusarium head blight resistance in contemporary barley breeding germplasm. *Molecular breeding* 27(4):439–454. doi:10.1007/s11032-010-9442-0. URL <http://link.springer.com/10.1007/s11032-010-9442-0>.
- Mesfin, A., K. P. Smith, R. Dill-Macky, C. K. Evans, R. Waugh, C. D. Gustus, and G. J. Muehlbauer. 2003. Quantitative Trait Loci for Fusarium Head Blight Resistance in Barley Detected in a Two-Rowed by Six-Rowed Population. *Crop Science* 43(1):307. doi:10.2135/cropsci2003.0307.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4):1819–1829. doi:11290733.
- Mohammadi, M., T. Tiede, and K. P. Smith. 2015. PopVar: A Genome-Wide Procedure for Predicting Genetic Variance and Correlated Response in Biparental Breeding Populations. *Crop Science* 55(5):2068–2077. doi:10.2135/cropsci2015.01.0030. URL <https://dl.sciencesocieties.org/publications/cs/abstracts/55/5/2068>.
- Mohring, J. and H. P. Piepho. 2009. Comparison of weighting in two-stage analysis of plant breeding trials. *Crop Science* 49(6):1977–1988. doi:10.2135/cropsci2009.02.0083.
- Morais Júnior, O. P., J. B. Duarte, F. Breseghello, A. S. Coelho, and A. M. Magalhães Júnior. 2018. Single-step reaction norm models for genomic prediction in multienvironment recur-

- rent selection trials. *Crop Science* 58(2):592–607. doi:10.2135/cropsci2017.06.0366.
- Müller, D., P. Schopp, and A. E. Melchinger. 2018. Selection on Expected Maximum Haploid Breeding Values Can Increase Genetic Gain in Recurrent Genomic Selection. *G3: Genes|Genomes|Genetics* 8(April):g3.200091.2018. doi:10.1534/g3.118.200091. URL <http://g3journal.org/lookup/doi/10.1534/g3.118.200091>.
- Muñoz-Amatriaín, M., M. J. Moscou, P. R. Bhat, J. T. Svensson, J. Bartoš, P. Suchánková, H. Šimková, T. R. Endo, R. D. Fenton, S. Lonardi, A. M. Castillo, S. Chao, L. Cistué, A. Cuesta-Marcos, K. L. Forrest, M. J. Hayden, P. M. Hayes, R. D. Horsley, K. Makoto, D. Moody, K. Sato, M. P. Vallés, B. B. Wulff, G. J. Muehlbauer, J. Doležel, and T. J. Close. 2011. An Improved Consensus Linkage Map of Barley Based on Flow-Sorted Chromosomes and Single Nucleotide Polymorphism Markers. *The Plant Genome Journal* 4(3):238–249. doi:10.3835/plantgenome2011.08.0023. URL <https://www.crops.org/publications/tpg/abstracts/4/3/238>.
- Neyhart, J. L., D. Sweeney, M. Sorrells, C. Kapp, K. D. Kephart, J. Sherman, E. J. Stockinger, S. Fisk, P. Hayes, S. Daba, M. Mohammadi, N. Hughes, L. Lukens, P. G. Barrios, L. Gutiérrez, and K. P. Smith. 2019. Registration of the S2MET Barley Mapping Population for Multi-Environment Genomewide Selection. *Journal of Plant Registrations* 13(2):270–280. doi:10.3198/jpr2018.06.0037crmp. URL <https://dl.sciencesocieties.org/publications/jpr/abstracts/13/2/270>.
- O'Donnell, M. S. and D. A. Ignizio. 2012. Bioclimatic Predictors for Supporting Ecological Applications in the Conterminous United States. U.S Geological Survey Data Series 691 page 10.
- Ouyang, Z., R. P. Mowers, A. Jensen, S. Wang, and S. Zheng. 1995. Cluster Analysis for Genotype × Environment Interaction with Unbalanced Data. *Crop Science* 35(5):1300. doi:10.2135/cropsci1995.0011183X003500050008x. URL <https://dl.sciencesocieties.org/publications/cs/abstracts/35/5/CS0350051300>.
- Pauli, D., G. Muehlbauer, K. Smith, and B. Cooper. 2014. Association Mapping of Agronomic QTLs in US Spring Barley Breeding Germplasm. *The Plant Genome* 7(3). doi:10.3835/plantgenome2013.11.0037. URL <https://dl.sciencesocieties.org/files/publications/tpg/first-look/tpg13-11-0037.pdf>.
- Pérez-Rodríguez, P., J. Crossa, K. Bondalapati, G. De Meyer, F. Pita, and G. De Los Campos. 2015. A pedigree-based reaction norm model for prediction of cotton yield in multi-environment trials. *Crop Science* 55(3):1143–1151. doi:10.2135/cropsci2014.08.0577.
- Piepho, H. P. and J. Möhring. 2005. Best Linear Unbiased Prediction of Cultivar Effects for Subdivided Target Regions. *Crop Science* 45(3):1151. doi:10.2135/cropsci2004.0398. URL <https://www.crops.org/publications/cs/abstracts/45/3/1151>.

- Piepho, H. P., J. Möhring, T. Schulz-Streeck, and J. O. Ogutu. 2012a. A stage-wise approach for the analysis of multi-environment trials. *Biometrical Journal* 54(6):844–860. doi:10.1002/bimj.201100219.
- Piepho, H. P., J. O. Ogutu, T. Schulz-Streeck, B. Estaghirou, a. Gordillo, and F. Technow. 2012b. Efficient computation of ridge-regression best linear unbiased prediction in genomic selection in plant breeding. *Crop Science* 52(3):1093–1104. doi:10.2135/cropsci2011.11.0592.
- Poets, A. M., M. Mohammadi, K. Seth, H. Wang, T. J. Y. Kono, Z. Fang, G. J. Muehlbauer, K. P. Smith, and P. L. Morrell. 2016. The effects of both recent and long-term selection and genetic drift are readily evident in North American barley breeding populations. *G3: Genes|Genomes|Genetics* 6(March):1–54. doi:10.1534/g3.115.022806.
- R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.r-project.org/>.
- Resende, M. F. R., P. Munoz, M. D. V. Resende, D. J. Garrick, R. L. Fernando, J. M. Davis, E. J. Jokela, T. a. Martin, G. F. Peter, and M. Kirst. 2012. Accuracy of Genomic Selection Methods in a Standard Data Set of Loblolly Pine (*Pinus taeda* L.). *Genetics* 190(4):1503–1510. doi:10.1534/genetics.111.137026.
- Rincent, R., D. Laloe, S. Nicolas, T. Altmann, D. Brunel, P. Revilla, V. M. Rodriguez, J. Moreno-Gonzalez, a. Melchinger, E. Bauer, C.-C. Schoen, N. Meyer, C. Giauffret, C. Bauland, P. Jamin, J. Laborde, H. Monod, P. Flament, a. Charcosset, and L. Moreau. 2012. Maximizing the Reliability of Genomic Selection by Optimizing the Calibration Set of Reference Individuals: Comparison of Methods in Two Diverse Groups of Maize Inbreds (*Zea mays* L.). *Genetics* 192(2):715–728. doi:10.1534/genetics.112.141473. URL <http://www.genetics.org/cgi/doi/10.1534/genetics.112.141473>.
- de Roos, A. P. W., B. J. Hayes, R. J. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* 179(3):1503–1512. doi:10.1534/genetics.107.084301.
- Rutkoski, J., R. P. Singh, J. Huerta-Espino, S. Bhavani, J. Poland, J. L. Jannink, and M. E. Sorrells. 2015. Efficient Use of Historical Data for Genomic Selection: A Case Study of Stem Rust Resistance in Wheat. *The Plant Genome* 8(1). doi:10.3835/plantgenome2014.09.0046. URL <https://dl.sciencesocieties.org/publications/tpg/abstracts/8/1/plantgenome2014.09.0046>.
- Saint Pierre, C., J. Burgueño, J. Crossa, G. Fuentes Dávila, P. Figueroa López, E. Solís Moya, J. Ireta Moreno, V. M. Hernández Muela, V. M. Zamora Villa, P. Vikram, K. Mathews, C. Sansaloni, D. Sehgal, D. Jarquin, P. Wenzl, and S. Singh. 2016. Genomic prediction models for grain yield of spring bread wheat in diverse agro-ecological zones. *Scientific*

- Reports 6:27312. doi:10.1038/srep27312. URL <http://www.nature.com/articles/srep27312>.
- Sallam, A. H., J. B. Endelman, J.-L. Jannink, and K. P. Smith. 2015. Assessing Genomic Selection Prediction Accuracy in a Dynamic Barley Breeding Population. *The Plant Genome* 8(1). doi:10.3835/plantgenome2014.05.0020. URL <https://dl.sciencesocieties.org/publications/tpg/abstracts/8/1/plantgenome2014.05.0020>.
- Schnell, F. and H. Utz. 1975. F1-leistung und elternwahl euphyder züchtung von selbstbefruchtern. In Bericht über die Arbeitstagung der Vereinigung Österreichischer Pflanzenzüchter, pages 243–248. BAL Gumpenstein, Gumpenstein, Austria.
- Scrucca, L., M. Fop, T. B. Murphy, and A. E. Raftery. 2016. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *The R journal* 8(1):289–317. doi:10.1177/2167702614534210. URL <http://www.ncbi.nlm.nih.gov/pubmed/27818791> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5096736>.
- Soil Landscapes of Canada Working Group. 2010. Soil Landscapes of Canada version 3.2. URL <http://sis.agr.gc.ca/cansis/nsdb/slc/v3.2/index.html>.
- Soil Survey Staff, National Resource Conservation Service, U. S. D. o. A. 2018. Web Soil Survey. URL <https://websoilsurvey.sc.egov.usda.gov/>.
- Souza, E. and M. E. Sorrells. 1991. Prediction of progeny variation in oat from parental genetic relationships. *Theoretical and Applied Genetics* 82(2):233–241. doi:10.1007/BF00226219.
- Technow, F. 2015. R package mvngGrAd: moving grid adjustment in plant breeding field trials.
- Technow, F., T. a. Schrag, W. Schipprack, E. Bauer, H. Simianer, and A. E. Melchinger. 2014. Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. *Genetics* 197(4):1343–1355. doi:10.1534/genetics.114.165860.
- Tiede, T., L. Kumar, M. Mohammadi, and K. P. Smith. 2015. Predicting genetic variance in bi-parental breeding populations is more accurate when explicitly modeling the segregation of informative genomewide markers. *Molecular Breeding* 35(10):199. doi:10.1007/s11032-015-0390-6.
- Tiede, T. and K. P. Smith. 2018. Evaluation and retrospective optimization of genomic selection for yield and disease resistance in spring barley. *Molecular Breeding* 38(5). doi:10.1007/s11032-018-0820-3.
- Toosi, A., R. L. Fernando, and J. C. M. Dekkers. 2010. Genomic selection in admixed and crossbred populations. *Journal of Animal Science* 88(1):32–46. doi:10.2527/jas.2009-1975.

- Utz, H. F., M. Bohn, and A. E. Melchinger. 2001. Predicting progeny means and variances of winter wheat crosses from phenotypic values of their parents. *Crop Science* 41(5):1470–1478. doi:10.2135/cropsci2001.4151470x.
- VanRaden, P. 2008. Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science* 91(11):4414–4423. doi:10.3168/jds.2007-0980. URL <http://linkinghub.elsevier.com/retrieve/pii/S0022030208709901>.
- Wientjes, Y. C. J., R. F. Veerkamp, and M. P. L. Calus. 2013. The Effect of Linkage Disequilibrium and Family Relationships on the Reliability of Genomic Prediction. *Genetics* 193(2):621–631. doi:10.1534/genetics.112.146290. URL <http://www.genetics.org/cgi/doi/10.1534/genetics.112.146290>.
- Yan, W. 2016. Analysis and Handling of $G \times E$ in a Practical Breeding Program. *Crop Science* 56:2106–2118. doi:10.2135/cropsci2015.06.0336. URL <https://dl.sciencesocieties.org/publications/cs/abstracts/0/0/cropsci2015.06.0336>.
- Yan, W., L. Hunt, Q. Sheng, and Z. Szlavnic. 2000. Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot. *Crop Science* 40(3):597. doi:10.2135/cropsci2000.403597x.
- Yates, F. and W. G. Cochran. 1938. The analysis of groups of experiments. *The Journal of Agricultural Science* 28(04):556. doi:10.1017/S0021859600050978. URL http://www.journals.cambridge.org/abstract_{_}S0021859600050978.
- Yu, X., X. Li, T. Guo, C. Zhu, Y. Wu, S. E. Mitchell, K. L. Roozeboom, D. Wang, M. L. Wang, G. A. Pederson, T. T. Tesso, P. S. Schnable, R. Bernardo, and J. Yu. 2016. Genomic prediction contributing to a promising global strategy to turbocharge gene banks. *Nature Plants* 2(10):16150. doi:10.1038/nplants.2016.150. URL <http://www.nature.com/articles/nplants2016150>.
- Zhang, X., A. Sallam, L. Gao, T. Kantarski, J. Poland, L. R. DeHaan, D. L. Wyse, and J. A. Anderson. 2016. Establishment and Optimization of Genomic Selection to Accelerate the Domestication and Improvement of Intermediate Wheatgrass. *The Plant Genome* 9(0). doi:10.3835/plantgenome2015.07.0059. URL <https://dl.sciencesocieties.org/publications/tpg/abstracts/0/0/plantgenome2015.07.0059>.
- Zhong, S. and J.-L. Jannink. 2007. Using quantitative trait loci results to discriminate among crosses on the basis of their progeny mean and variance. *Genetics* 177(1):567–576. doi:10.1534/genetics.107.075358.