

10-19-2015

Adding Genetically Distant Individuals to Training Populations Reduces Genomic Prediction Accuracy in Barley


Aaron Lorenz

University of Nebraska-Lincoln, lore0149@unl.edu

Kevin P. Smith

University of Minnesota

Follow this and additional works at: <http://digitalcommons.unl.edu/agronomyfacpub>

 Part of the [Agricultural Science Commons](#), [Agriculture Commons](#), [Agronomy and Crop Sciences Commons](#), [Botany Commons](#), [Horticulture Commons](#), [Other Plant Sciences Commons](#), and the [Plant Biology Commons](#)

Lorenz, Aaron and Smith, Kevin P., "Adding Genetically Distant Individuals to Training Populations Reduces Genomic Prediction Accuracy in Barley" (2015). *Agronomy & Horticulture -- Faculty Publications*. 834.
<http://digitalcommons.unl.edu/agronomyfacpub/834>

This Article is brought to you for free and open access by the Agronomy and Horticulture Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Agronomy & Horticulture -- Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Adding Genetically Distant Individuals to Training Populations Reduces Genomic Prediction Accuracy in Barley

Aaron J. Lorenz* and Kevin P. Smith

ABSTRACT

One of the most important factors affecting genomic prediction accuracy appears to be training population (TP) composition. The objective of this study was to evaluate the effect of genomic relationship on genomic prediction accuracy and determine if adding increasingly unrelated individuals to a TP can reduce prediction accuracy. To accomplish this, a population of barley (*Hordeum vulgare* L.) lines from the University of Minnesota (lines denoted as MN) and North Dakota State University (lines denoted as ND) breeding programs were used for model training. Predictions were validated using two independent sets of progenies derived from MN \times MN crosses and ND \times ND crosses. Predictive ability sharply decreased with decreasing relationship between the TP and validation population (VP). More importantly, it was observed that adding increasingly unrelated individuals to the TP can actually reduce predictive ability compared with smaller TPs consisting of highly related individuals only. Reported results are possibly conditional on the relatively low marker density (342 single nucleotide polymorphisms [SNPs]) used. Nevertheless, these findings suggest plant breeding programs desiring to use genomic selection could benefit from focusing on good phenotyping of smaller TPs closely related to the selection candidates rather than developing large and diverse TPs.

A.J. Lorenz, Dep. of Agronomy and Horticulture, Univ. of Nebraska, Lincoln, NE 68583; and K.P. Smith, Dep. of Agronomy and Plant Genetics, Univ. of Minnesota, St. Paul, MN 55108. Received 15 Dec. 2014. Accepted 30 Apr. 2015. *Corresponding author (lore0149@umn.edu).

Abbreviations: BLUE, best linear unbiased estimate; CAP, Barley Coordinated Agricultural Project; DON, deoxynivalenol; FHB, Fusarium head blight; G-BLUP, genomic best linear unbiased prediction; HT, plant height; IBS, identity-by-state; LD, linkage disequilibrium; QTL, quantitative trait loci; RCBD, randomized complete-block design; RR-BLUP, ridge regression best linear unbiased prediction; SNP, single nucleotide polymorphism; TP, training population; VP, validation population.

MANY FACTORS affect genomic prediction accuracy, including model, marker density, TP composition, trait complexity, and precision of phenotyping (Combs and Bernardo, 2013; Lorenz et al., 2012; Lorenz, 2013; Heffner et al., 2011). One of the most important factors under control of the breeder appears to be TP, or calibration set, composition (Lorenz et al., 2012; Riedelsheimer et al., 2013; Rincent et al., 2012; Wientjes et al., 2013). Intelligent sampling of a TP from a larger population of individuals could enhance the effectiveness and efficiency of genomic selection for plant breeding (Jannink et al., 2010). If a plant breeder desires information on trait performance for a population too large to fully phenotype with available resources, genomic predictions could be used in place of phenotypes if the expense of genotyping is substantially lower than that of phenotyping, a scenario that is increasingly becoming reality for many

Published in Crop Sci. 55:2657–2667 (2015).

doi: 10.2135/cropsci2014.12.0827

© Crop Science Society of America | 5585 Guilford Rd., Madison, WI 53711 USA

All rights reserved.

traits. The goal then is to identify the most informative subset of individuals for model training. Rincent et al. (2012) showed that an exchange algorithm combined with an objective function, consisting of the generalized coefficient of determination, chose TPs that were more informative than randomly sampled TPs. This method was applied to a diversity panel of maize (*Zea mays* L.) lines rather than a population with genetic structure more typical of a breeding program.

Another objective of TP sampling could be to choose the most informative subset of records from an extant TP database containing genotype and phenotype information. An active genomic selection program could continually build a database as additional generations of individuals are phenotyped and genotyped. As each additional set of progenies from a new cycle of selection is genotyped, a model is trained and genomic predictions are calculated. Should all records be used for model training or only a subset of those records? As selection proceeds, relationships between selection candidates and the TP comprised of individuals from early cycles of selection decrease, potentially making TP individuals from distant generations less informative. Data could potentially be shared between public breeding programs, but is it beneficial to combine germplasm from different programs into a single TP to increase TP size, or could it actually be detrimental?

A number of studies have shown the importance of genetic relationships between training individuals and selection candidates on genomic prediction accuracy (Clark et al., 2012; Lorenz et al., 2012; Wientjes et al., 2013; Pszczola et al., 2012; Habier et al., 2010). Prediction accuracies have been found to be low when TPs and VPs are composed of germplasm from different breeding programs (Lorenz et al., 2012) or even between different full-sib families (Riedelsheimer et al., 2013). It has also been shown that measures of the genetic relationship between TPs and selection candidates are the best predictors of genomic prediction accuracy (Clark et al., 2012; Wientjes et al., 2013). There are at least three possible reasons for these observations. First, more closely related individuals share a common ancestor fewer generations back in time, and, therefore, fewer opportunities existed for recombination between markers and quantitative trait loci (QTL), preserving QTL–marker linkage phases. This is especially true for genomic prediction models that rely on relationships and long-range linkage disequilibrium (LD) between markers and QTL, such as genomic best linear unbiased prediction (G-BLUP) and ridge regression BLUP (RR-BLUP; Habier et al., 2013). Second, training and selection candidate populations with a closer genetic relationship are more likely to share polymorphic loci generating genetic variation. In other words, it is possible that genetic variation within distantly related populations is controlled by different sets of polymorphic loci caused by drift and mutation operating separately

through time. Finally, QTL \times genetic-background interactions could exist (Lorenz and Cohen, 2012; Mohammadi et al., 2015). More closely related individuals share a larger fraction of their genetic background than distantly related individuals and are, therefore, more likely to share these interaction deviations if they exist.

The objective of this study was to evaluate the effect of genomic relationship on genomic prediction accuracy and determine if adding increasingly unrelated individuals to a TP can reduce prediction accuracy. To address these objectives, we used phenotypic and genotypic data from the University of Minnesota genomic selection program for resistance to Fusarium head blight (FHB). This dataset holds many advantages for studying genomic-selection-related questions than many other datasets used in similar publications. Most importantly, the set of selection candidates, or VP, is a generation of selection and sexual recombination removed from the TP. Because the biggest advantage of genomic selection lies in its potential to expedite cycles of selection through circumventing phenotyping, progenies being predicted will be at least one cycle of selection removed from the TP. Second, the VP consists of many groups of biparental families derived from a series of crosses between selected lines in the breeding program, typical of most plant breeding programs.

MATERIALS AND METHODS

Germplasm

Detailed information on the germplasm composing the TP can be found in Massman et al. (2011) and Lorenz et al. (2012). Briefly, the whole TP consisted of 768 six-row barley F_4 lines, of which 384 lines were taken from the University of Minnesota breeding program and 384 lines were taken from the North Dakota State University breeding program. These lines were submitted to the Barley Coordinated Agricultural Project (CAP) Years 1 through 4 (i.e., 2006–2009; <http://barleycap.org>). Ninety-six lines were submitted per year for genotyping and phenotyping.

The VP consisted of 300 $F_{3.5}$ progenies derived from a set of crosses between advanced breeding lines selected from the TP. Fourteen parents were crossed in different combinations to create three cross types: MN \times MN, MN \times ND, and ND \times ND. Ten crosses were made per cross type, and 24 progenies were derived per cross, resulting in 240 progenies per cross type. After derivation of the $F_{3.5}$ lines, 100 progenies were randomly selected per cross type from the original 240.

Phenotyping and Genotyping

For the TP, plant height (HT) was evaluated at four locations in a randomized complete-block design (RCBD) with two replications in each CAP year. Ninety-six lines from each breeding program were evaluated in separate years so that the 96 lines evaluated in 2006 were completely different than the 96 evaluated in 2007 and so forth. The FHB ratings and deoxynivalenol (DON) concentrations were collected in single-row disease nurseries conducted at four locations between Minnesota and North Dakota each year. A RCBD with two replications was

used. Because of the large number of trials and high degree of unbalance between years, common checks were used to adjust for trial effects and calculate best linear unbiased estimates (BLUEs) of each line with a linear model including fixed effects for year, complete block nested within year, and line. Residuals were assumed to be random effects independent and identically distributed. Three to nine checks were in common between trials across years. More information on TP phenotyping can be found in Massman et al. (2011) and Lorenz et al. (2012).

Phenotypic data on the VP was collected in separate trials for agronomic and disease traits. Protocols and methods for trait measurement were the same as those used for the TP. For agronomic traits, lines were planted in an augmented design with six incomplete blocks and three check varieties per incomplete block. Each check was replicated two times in each incomplete block. Data on agronomic traits was collected at three locations in 2011 (St. Paul, MN; Crookston, MN; and Nesson Valley, ND) and four locations in 2012 (St. Paul, MN; Crookston, MN; Nesson Valley, ND; and Fargo, ND). All lines were evaluated in all seven location–year combinations. Disease traits were evaluated in single-row plots planted in disease nurseries across the four locations in both 2011 and 2012. All lines were evaluated in all eight location–year combinations. The same design as that used for agronomic trait evaluation was used in the disease nurseries, with the exception that each trial was replicated two times at each location. Best linear unbiased estimates adjusted for block effects were calculated for each line by fitting a linear model including fixed effects for environment (year–location combination), complete block nested within environment, incomplete block nested within complete block, and line. Residuals were assumed to be random effects independent and identically distributed.

Three thousand and seventy-two SNPs were scored on the TP using two Illumina GoldenGate oligonucleotide pool assays (Close et al., 2009). More detail on genotyping the TP can be found in Massman et al. (2011) and Lorenz et al. (2012). From this set of 3072 SNPs, a subset of 384 SNPs was scored on the VP. The SNPs were selected on the basis of polymorphism level across the 14 selected parents and uniform distribution across the genetic map. After filtering out 19 failed SNPs and 23 SNPs with low minor-allele frequency and excessive heterozygosity, 342 SNPs remained.

Genomic Prediction Model Training

An RR-BLUP genomic prediction model was trained:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

Where \mathbf{y} is a vector of BLUEs of the reference lines; $\mathbf{1}\mu$ is an intercept vector; \mathbf{Z} is an $n \times p$ incidence matrix containing the allelic states of the p marker loci ($z = \{-1, 0, 1\}$), where -1 represents the minor allele; \mathbf{u} is the $p \times 1$ vector of marker effects; and \mathbf{e} is a $n \times 1$ vector of residuals. Under RR-BLUP, $\mathbf{u} \sim \text{MVN}(0, \mathbf{I}\sigma_u^2)$, where σ_u^2 is the variance of the common distribution of marker effects and is estimated using restricted maximum likelihood. The RR-BLUP model was implemented in the R package rrBLUP version 4.2 (R Development Core Team, 2012; Endelman, 2011). Predictions of the individuals

comprising the VP were calculated as $\hat{\mathbf{g}} = \mathbf{Z}_v\hat{\mathbf{u}}$, where $\hat{\mathbf{g}}$ is a vector of genomic predictions, \mathbf{Z}_v is the marker incidence matrix of the VP, and $\hat{\mathbf{u}}$ is the vector of predicted marker effects output from the RR-BLUP model.

Matrices containing realized relationships between all individuals were calculated four different ways: Method 1 of Van Raden (2008) (\mathbf{G}), method 2 of Van Raden (2008) (\mathbf{G}_2), identity-by-state (IBS) similarities (\mathbf{S}_{IBS}), and genomic correlations (\mathbf{S}_{GC}). The latter two were calculated as in Riedelsheimer et al. (2013). The formulas can be found in the provided references. Briefly, \mathbf{G} is the centered and scaled genomic relationship matrix. The \mathbf{G}_2 method is similar to \mathbf{G} except that markers are weighted by the reciprocal of their variances (i.e., markers with low minor-allele frequency are weighted more). The matrix \mathbf{S}_{IBS} contains the proportion of marker alleles shared between individuals, and \mathbf{S}_{GC} contains the Pearson correlation coefficient of allelic states between individuals.

A sliding-window approach was used to study the effect of genetic relationship between TP and selection candidates on prediction accuracy. First, three VPs were defined: MN \times MN progenies ($n = 100$), MN \times ND progenies ($n = 100$), and ND \times ND progenies ($n = 100$). For a given VP, the TP was sorted, in descending order, according to the mean G_{ij} between an individual in the TP and the whole VP. A window of size $N = 200$ individuals was used and slid down the gradient of relationship in increments of 10 individuals (i.e., the first TP was individuals 1–200, the second 11–210, etc.). For each sampled TP, a model was trained and predictions were correlated to observed values to calculate predictive ability. The sliding window was incremented by 10 individuals down the gradient of relationship until individual 760. Predictive ability, defined as the correlation between observed value and predicted value, was plotted against mean G_{ij} between the TP and VP. Quadratic functions were fit to the points with only significant terms ($P < 0.05$) retained.

To study the effect of adding increasingly unrelated individuals to a TP, a similar approach was taken where TP individuals were sorted according their average relationship to the VP. A TP was started by selecting the 10 lines with highest mean G_{ij} with the VP. The TP was increased in size by increments of 10 by adding the next 10 lines with the highest mean G_{ij} with the VP. The TP was increased to $N = 760$ following this procedure. For each TP, a model was trained and predictions were correlated against observations.

Related to adding individuals to the TP based on their mean G_{ij} with the VP, two additional methods were attempted. First, a TP was developed for every single individual in the VP. That is, individuals in the TP were sorted based on their relatedness to the single validation individual being predicted. The TP of size N was selected and used to predict the genetic value of that single validation individual. This was repeated for all VP individuals at all TP sizes ranging from 10 to 700. A similar algorithm was used on a family basis, where TPs were selected for each family rather than each individual.

A standard error for each correlation coefficient was estimated using the bootstrap procedure with 1000 bootstrap replicates (Efron and Gong, 1983).

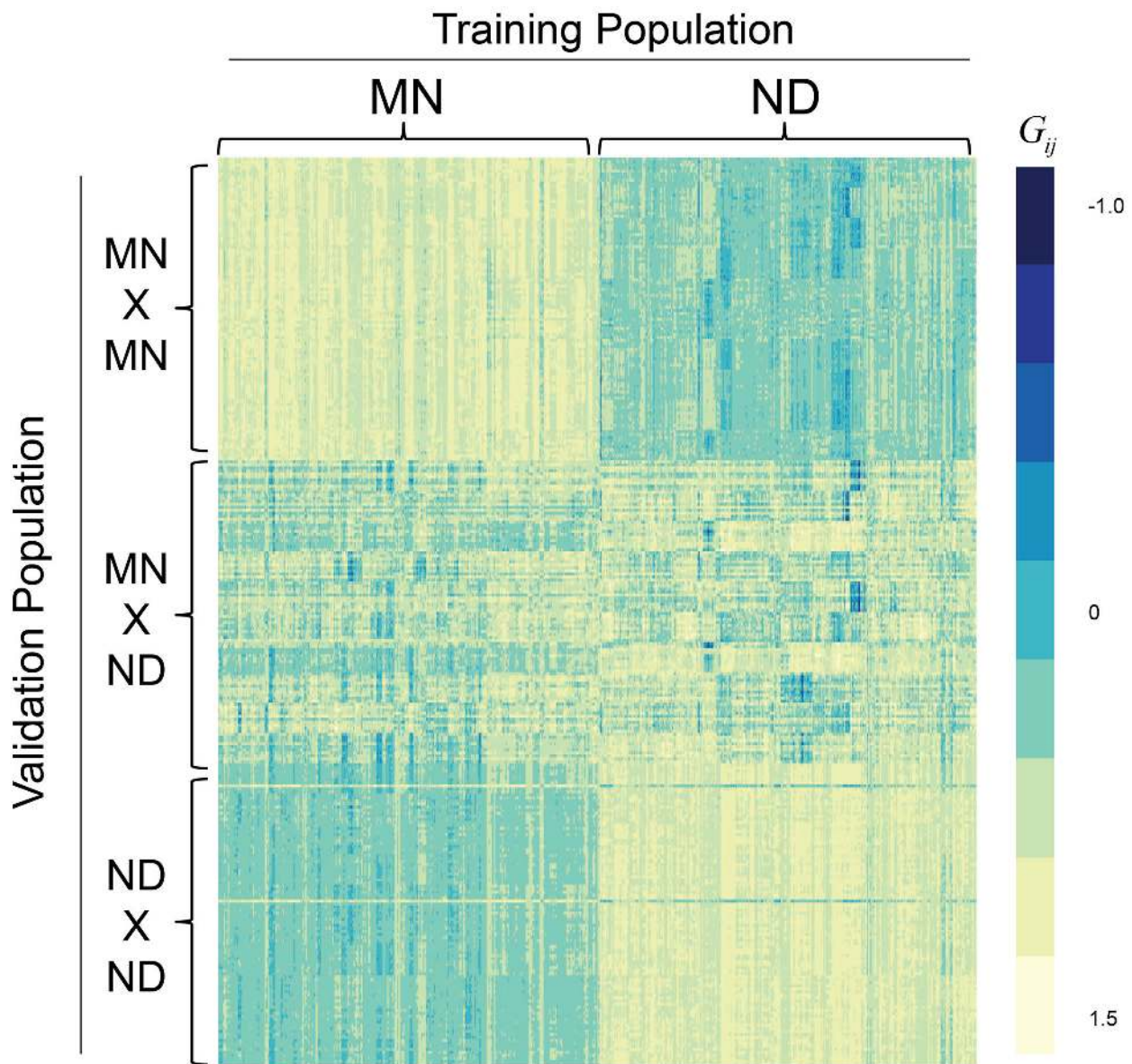


Figure 1. Heat map representing realized genomic relationship (G ; Van Raden Method 1) among training population barley lines and validation lines derived from MN \times MN, MN \times ND, and ND \times ND crosses.

RESULTS

Variation in the realized genomic relationships between lines within breeding programs and between breeding programs is displayed in Fig. 1. As expected, the average relationship is higher between TP lines and progenies derived from the same breeding program. There are, however, a number of instances where relationships were higher between programs than within programs (Fig. 2) resulting from the fact that the MN and ND breeding programs have exchanged germplasm and are much less diverged than germplasm of other barley breeding programs (Hamblin et al., 2010). The average relationship between the MN and ND TPs and the MN \times ND progenies was centered at zero and the variance in relationships was at least as great as the between and within-program comparisons (Fig. 1, 2).

Using the sliding-window approach, a clear positive relationship between predictive ability and mean G_{ij} between TP and VP was observed for two of the three traits and both the MN \times MN and ND \times ND VPs (Fig. 3). The relationship was less linear for FHB, especially in the ND \times ND VP. When the 200 least-related individuals were used to train a model, predictive abilities were negative. Predictive abilities approached 0.50 when the most closely related individuals composed the TP. These predictive abilities are expected to be lower than the prediction accuracy (i.e., the correlation between the prediction and the true breeding value) because of the random environmental deviations included in the validation phenotype, which was not adjusted for. Overall, predictive ability of the MN \times ND progenies was poor and no relationship

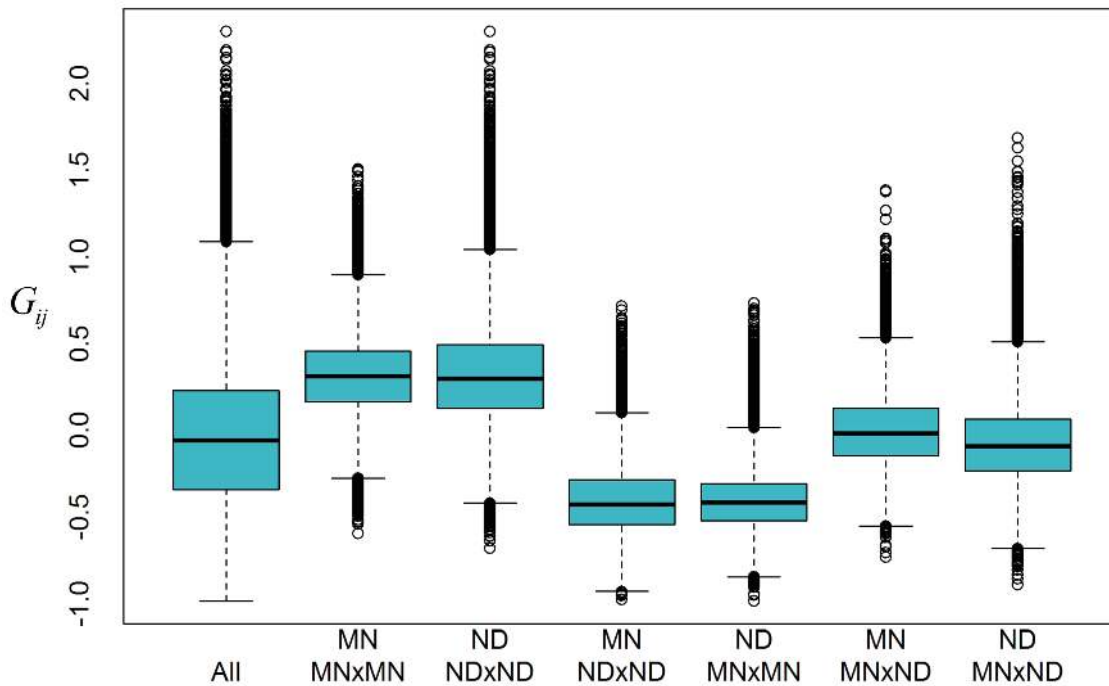


Figure 2. Box plots of realized genomic relationships (G_{ij}) among all possible pairs of individuals (All) and between pairs of individuals between the group species by the top lines of the x-axis and bottom line of the x-axis. For example, the second box from the left represents all G_{ij} values between the MN training population and MN \times MN validation population.

between predictive ability and G_{ij} was observed, except for a very weak linear relationship for DON (Fig. 3C).

To address the question of whether adding increasingly unrelated lines to the TP can actually reduce predictive ability, the TP was built up by adding individuals according to their mean genomic relationship to the VP (see Materials and Methods section). This analysis was not performed for the MN \times ND VP because the relationship between predictive ability and mean G_{ij} was very weak or nonexistent (Fig. 3). For both the MN \times MN and ND \times ND VPs, prediction accuracy increased with increasing TP size to a point and then, in most instances, began to decline as increasingly unrelated individuals were added to the TP (Fig. 4). For DON in the MN \times MN VP, prediction accuracy was maximized at 0.49 when the most closely related 250 lines composed the TP. As less-related lines were added to the TP, with some of those lines being from the ND program, predictive ability began to gradually decline. Predictive ability dropped to 0.42 when 760 lines were used for model training. For FHB in the ND \times ND VP, predictive ability reached 0.45 at a TP size of 410 then gradually declined to 0.37 when $N = 760$. Most traits didn't exhibit a prediction accuracy decline until mean G_{ij} of the added set was <0 and a large fraction of the newly added lines was from the other breeding program. The trend was less pronounced for HT in both VPs, but the predictive ability still trended downward with mean G_{ij} of the added set. This pattern can also be seen by directly comparing the optimal predictive ability for mean G_{ij} and that when the whole TP is used (Fig. 5). The smaller

Tps selected by mean G_{ij} are always more predictive than the whole TP, sometimes by more than 45% (ND \times ND DON; Fig. 5). Interestingly, predictive ability of DON in the ND \times ND VP rapidly peaks at $N = 310$, quickly drops, then gradually starts to increase again around $N = 525$. This trend seems to suggest that adding MN training individuals quickly affects predictive ability, but then the effect of relatedness is overcome by increased N . It appears that predictive ability of DON is the most affected by adding unrelated individuals in the MN \times MN VP also (Fig. 4).

Other forms of a relationship matrix may be used in addition to the realized genomic relationship matrix calculated using Method 1 of Van Raden (2008), such as IBS similarities and genomic correlations as calculate by Riedelsheimer et al. (2013) and Method 2 of Van Raden (2008). These relationship matrices (\mathbf{S}_{IBS} , \mathbf{S}_{GC} , and \mathbf{G}_2) are all highly correlated with \mathbf{G} , with r^2 values ranging from 0.71 to 0.97. The same analysis displayed in Fig. 4 was repeated, but this time the \mathbf{S}_{IBS} , \mathbf{S}_{GC} , and \mathbf{G}_2 relationship matrices were used in place of \mathbf{G} . Table 1 displays the maximum predictive ability and the population size used to achieve that predictive ability when TPs were built on the basis of these forms of the relationship matrix. While differences are nonsignificant and quite small in most instances, it appears that \mathbf{G} tends to achieve higher predictive abilities with smaller TPs presumably because it better identifies the most closely related, and therefore informative, individuals. This is most apparent for FHB in both VPs in which considerably smaller TPs selected by

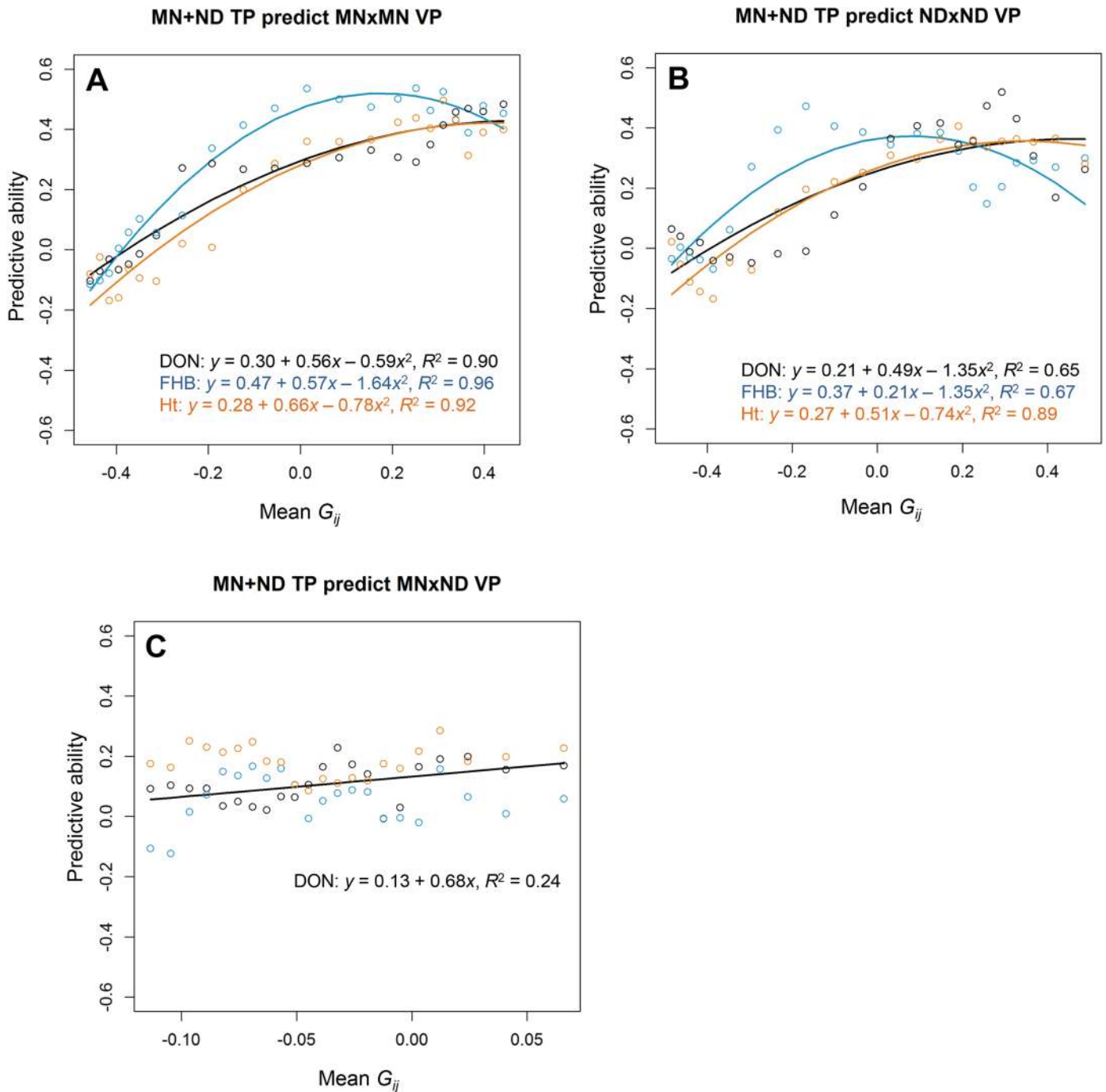
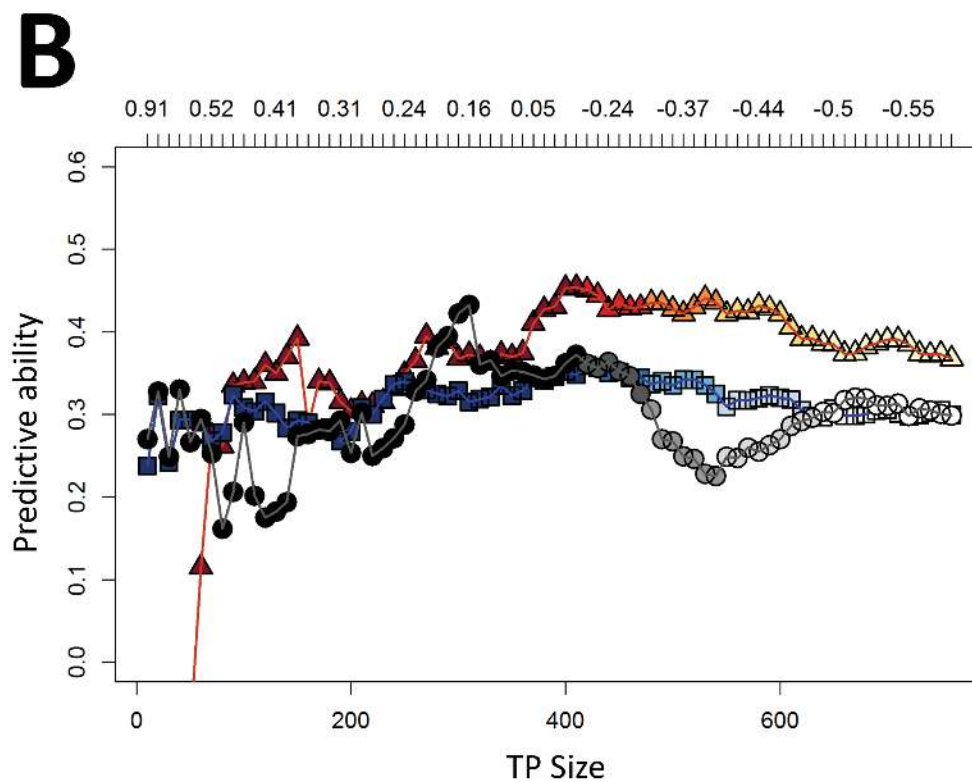
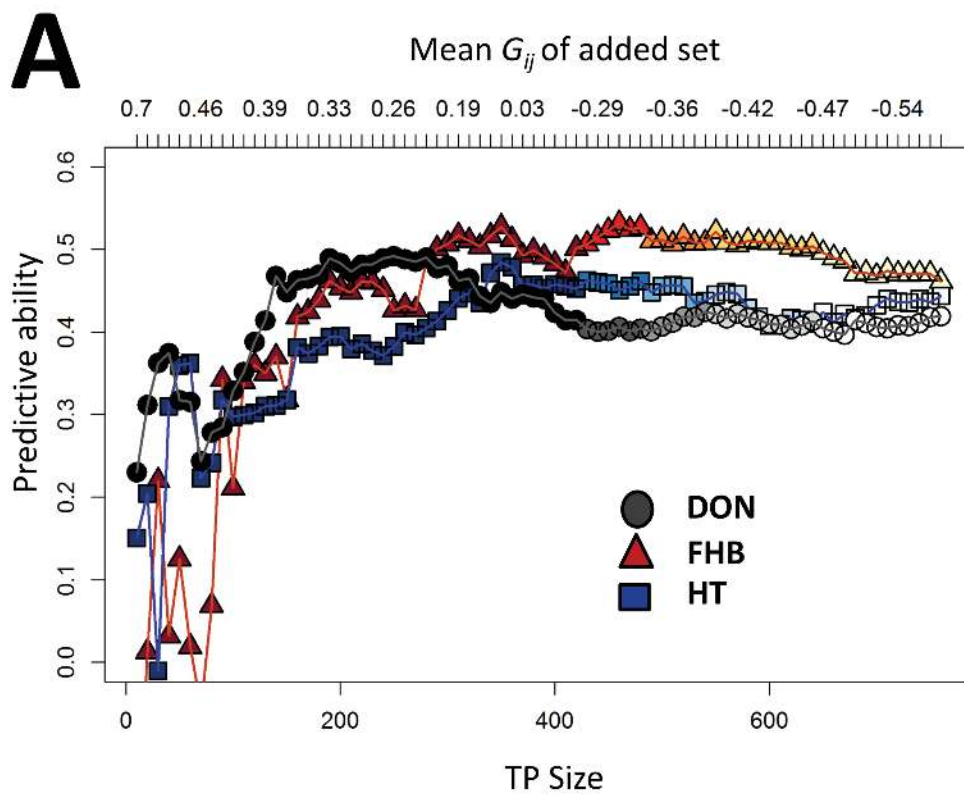


Figure 3. Plots of predictive ability vs. mean genomic relationship between the training population (TP) and validation population (VP) for deoxynivalenol (DON) concentration (black points), Fusarium head blight ratings (blue points), and plant height (orange points). Individuals in the TP were ordered based on their realized genomic relationship to the VP. A sliding window of 200 individuals, incremented by 10, was used to create the gradient in average relationship (Mean G_{ij}).

G provided higher prediction accuracies than larger TPs selected using the other relationship matrices.

An issue with using **G** to quantify relationships is that they are relative to the current population. Sets of individuals with similar ancestry may have different values of G_{ij} within different populations depending on the overall relatedness among individuals in the same population. It is not possible, therefore, to extrapolate a G_{ij} threshold to apply generally for inclusion or exclusion of individuals in

a TP. Identity-by-state similarities could be more generally applied because they are simply the shared fraction of polymorphisms. To evaluate an IBS cutoff, the average S_{ij} between newly added TP lines and the VP at which predictive ability begins to decline was tabulated. Critical S_{ij} values ranged from 0.74 for DON in the MN \times MN VP to 0.62 for FHB in the ND \times ND VP, indicating TPs can be confidently built up when adding individuals with S_{ij} values to the VP that are greater than 0.70. As S_{ij} values approach 0.60,



% MN (A) or ND (B)	100-	89-	79-	69-	59-
	90%	80%	70%	60%	50%

Figure 4. Plots of predictive ability vs. training population (TP) size when sets of 10 lines are added according to their genomic relationship (G_{ij}) to the validation population (VP). The mean G_{ij} of the newly added set is displayed along the top axis. (A) MN \times MN VP; (B) ND \times ND VP. The point shades of color represent the proportion of the whole TP that is either MN (A) or ND (B).

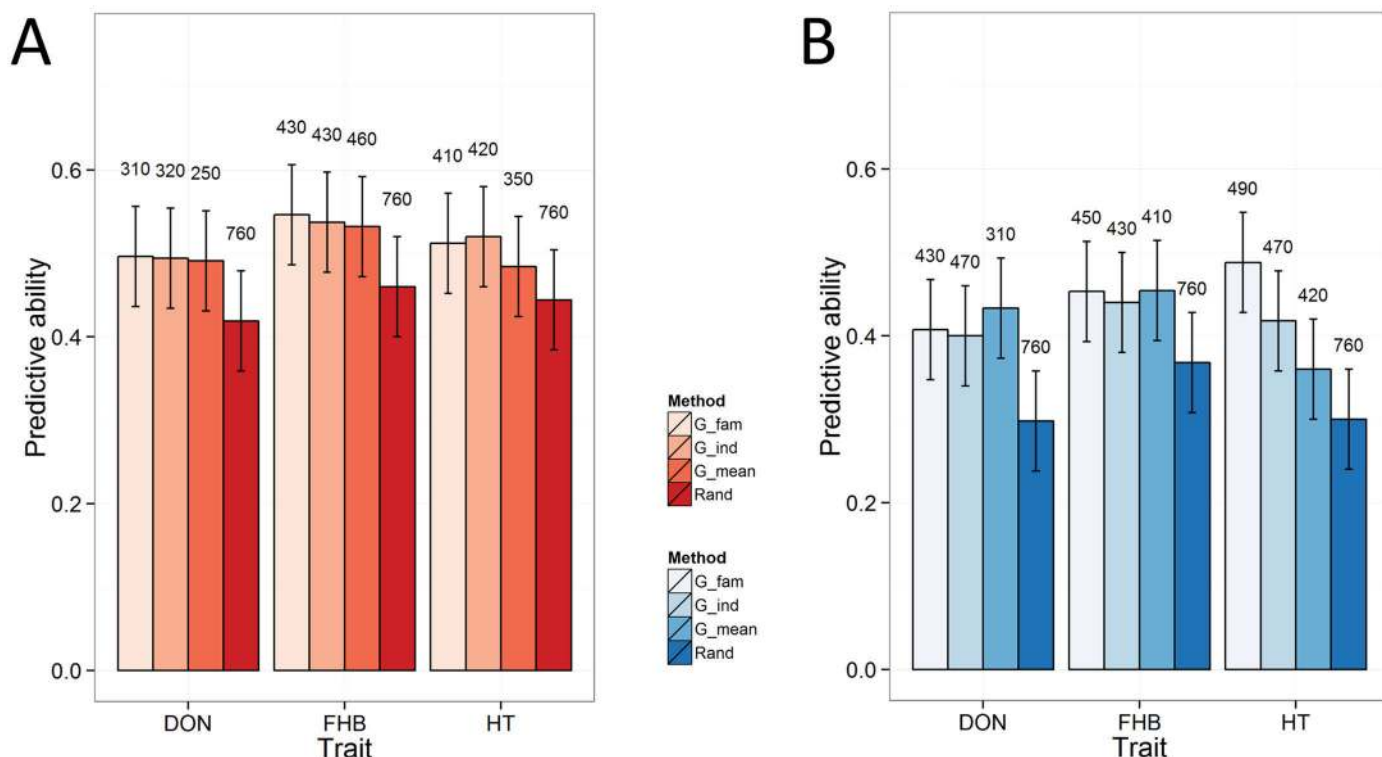


Figure 5. Predictive ability for four training population (TP) selection schemes: G_fam, TPs selected for specific families based on their average genomic relationship (G_{ij}) to that family; G_ind, TPs selected for specific individuals based on their genomic relationship to that individual; G_mean, TPs selected based on their average relationship to the entire validation population; Rand, training population randomly selected. The numbers above each bar indicate the size of the best performing training population. (A) MN × MN validation population; (B) ND × ND validation population. DON, deoxynivalenol; FHB, Fusarium head blight; HT, plant height.

Table 1. Maximum predictive abilities realized when training populations are built according to their relationship with the validation population measured using either Van Raden Method 1 genomic relationship matrix (G), Van Raden Method 2 genomic relationship matrix (G_2), identity-by-state similarity (S_{IBS}), or genomic correlation (S_{GC}). Size of the training population (N) at which maximum predictive ability is displayed.

Trait†	Method	Validation population					
		MN × MN			ND × ND		
		Predictive ability	N	SE	Predictive ability	N	SE
DON	G	0.492	250	0.071	0.433	310	0.067
	G_2	0.510	290	0.073	0.422	310	0.066
	S_{IBS}	0.480	300	0.074	0.396	320	0.067
	S_{GC}	0.450	420	0.073	0.398	330	0.072
FHB	G	0.532	460	0.075	0.454	410	0.077
	G_2	0.537	560	0.077	0.454	410	0.108
	S_{IBS}	0.527	510	0.076	0.452	470	0.077
	S_{GC}	0.470	380	0.085	0.440	470	0.089
HT	G	0.484	350	0.065	0.359	420	0.065
	G_2	0.480	360	0.065	0.353	410	0.067
	S_{IBS}	0.484	390	0.061	0.358	430	0.063
	S_{GC}	0.471	550	0.065	0.350	420	0.067

† DON, deoxynivalenol; FHB, Fusarium head blight; HT, plant height.

caution should be taken to prevent avoidable decreases in prediction accuracy. These values could depend on the population diversity, marker number, and marker set selection.

Given the observation that the average relationship between the TP and VP is an important factor in

predictive ability, this analysis was extended to include TP formation based on individual and family relatedness. An individual specific TP is one in which TP individuals are selected based on relationship to a VP individual. A model is trained and a prediction is made for that VP individual

and this is repeated for every individual in the VP. For family-specific TPs, a very similar algorithm is used, but TPs are selected for each family rather than for each individual. It was observed that individual- and family-specific TPs performed better than randomly selected TPs, but no substantial improvement over TP selected on mean G_{ij} was observed except in the case of HT in the ND \times ND VP (Fig. 5). In this case selecting a TP unique to each individual family was the most predictive, with a predictive ability greater than two standard errors over the mean G_{ij} predictive ability. It is not known why family-based TP selection worked particularly well for this one case.

DISCUSSION

Similar to findings by other researchers (Riedelsheimer et al., 2013; Lorenz et al., 2012; Wientjes et al., 2013; Clark et al., 2012; Lehermeier et al., 2014), we found that prediction accuracy is maximized when the TP and VP are closely related, and prediction accuracy is abysmal when TP and VP are relatively distantly related. Specifically, using a sliding-window approach, we observed that the most closely related set of 200 TP individuals provided much better predictions than the least-related set, which provided zero predictive ability. Potential underlying causes of this common observation are listed in the introductory section. It was not our objective to determine the causes underlying this trend, and, moreover, the relatively low marker density used in this study precluded a detailed analysis on marker linkage phases such as those performed by Technow et al. (2013) and Riedelsheimer et al. (2013). Our results on this topic clearly show that even within a structured breeding program consisting of related germplasm of low diversity, which characterizes the MN and ND barley germplasm used in this study (Hamblin et al., 2010), close attention should be given to relationships between TP and selection candidates.

A more novel aspect of this study is the investigation into the effect on predictive ability when increasingly genetically distant individuals are added to a standing TP. By sorting individuals based on relatedness to the VP and building a TP based on \mathbf{G} , it was shown that adding unrelated individuals to the TP does indeed hold potential for reducing predictive ability in the case of these barley populations. While absolute differences between whole TPs and TPs selected based on relatedness were mostly not significant because standard errors on correlations are typically large, we did find that using the whole TP was consistently less predictive than using a subset of the TP selected based on genomic relatedness. Clearly the addition of these individuals is adding error to marker-effect estimates through differing marker-QTL linkage phases, QTL \times genetic-background interactions, or differing sets of loci controlling variation for the traits between genetic groups. Little evidence has been found for this in prior

studies looking at real data. Technow et al. (2013) found that while prediction accuracy for Northern corn leaf blight in maize using an unrelated TP was much poorer than that achieved using a related TP, these authors observed no detrimental effects to combining related and unrelated TPs. In fact, prediction accuracy was slightly increased relative to the related TP alone. Using a series of biparental crosses, Riedelsheimer et al. (2013) found that adding an unrelated population to a TP with a half-sib relationship to the VP only slightly (and nonsignificantly) reduced prediction accuracy. Conversely, using simulations, Habier et al. (2013) showed that the addition of unrelated individuals to a TP can reduce the accuracy provided by additive genetic relationships. Increased accuracy contributed from unrelated individuals comes through the historical LD source of accuracy, which can be exploited if TPs are large and marker densities are high (Habier et al., 2013; Hickey et al., 2014).

It is recognized that the number of markers used in this study was relatively small and that the effect of relatedness on accuracy could be diminished if higher marker densities were used. This would apply if opposite linkage phases between unrelated individuals were the primary cause of reduced prediction accuracy and not QTL \times genetic-background interactions. The RR-BLUP model works best in situations of extensive relatedness resulting in long stretches of identical-by-descent DNA in populations and thus good preservation of LD phases, a common situation in plant breeding programs. Since marker effects are assumed to be sampled from a common distribution in RR-BLUP, the estimated marker effects are spread across many markers, that is, a large-effect QTL will have its effect spread across many markers rather than be captured by only the most proximal marker. The degree to which this shrinkage occurs depends on number of markers scored relative to the population size or, in other words, the severity of large p -small n problem. Therefore lack of shared linkage phase between more distant marker loci could still be an issue because effects are distributed across greater distances in RR-BLUP regardless of the marker density. The RR-BLUP model, and its equivalent G-BLUP, is the most commonly used model in genomic prediction for plant breeding because of its simplicity and good performance relative to more complex, computationally intensive models (Heslot et al., 2012). If greater genotyping densities were applied, like those found to be effective by Hickey et al. (2014), it could be advantageous to explicitly model the additive genetic relationships as well as short-range-marker-QTL LD to capture information coming from unrelated individuals as suggested by Habier et al. (2013).

Given variation in breeding program structure, levels of genetic diversity, and trait genetic architectures within and between crop species, it would be highly speculative to extrapolate these specific results to genomic prediction

for plant breeding in general. Nevertheless, it is useful to know that there is potential for this effect and if any general guidelines exist to help decide which individuals to include in a training set and which individuals to exclude. For selecting individuals to include in a TP, we used the realized genomic relationship (**G**) calculated using Method 1 of Van Raden (2008). While differences were small, it was found that this method tended to select the combination of smallest and best TP. This formulation of **G**, however, expresses the relationships among individuals relative to the current population. The mean relationship is zero, with negative coefficients indicating pairs of individuals are less related than the expected relationship of two randomly sampled individuals and positive coefficients indicating higher genetic relatedness. To develop some general guidelines for inclusion of individuals in a TP based on genetic distance requires that genetic relationship coefficients be fairly constant across populations rather than relative to the current population. Identity-by-state similarities were chosen for this. It was found that when individuals with <0.62 to $0.74 S_{ij}$ were added to the TP, predictive ability began to trend downward with some variation across traits. If the results reported herein hold up, they suggest that genomic selection programs in plant breeding should focus on developing training sets consisting of a few hundred (e.g., 200–500) individuals closely related to the selection candidates rather than large and diverse training sets. This approach would not only circumvent genetical reasons underlying reduced prediction accuracy, but would also minimize confounding effects of phenology on model training resulting from the inclusion of lines with wide variation in morphology and flowering time. In maize breeding, where breeding families are typically large (e.g., 50–200), families are routinely developed using doubled-haploid technology and seed quantity is enough to allow yield trials during early generations of selection, the focus has been on biparental family-specific TPs. In small grains like barley, however, often only a few individuals per family survive the first few generations of inbreeding and selection based on flowering time, disease resistance, and overall plant health and morphology. Therefore, the number of individuals per family making it to yield trials is generally very small, preventing the use of family-specific TPs. Our results suggest that these multifamily TPs should be closely related to the set of selection candidates. In this study, selecting based on relatedness to the whole VP was at least as good as family-specific and individual-specific TPs, but this result may be influenced by the fact that each VP used in this study was fairly narrow, only resulting from 10 related, within-breeding program crosses. If a great amount of variation exists among families within the selection candidate set, then it's likely a family-specific TP could be beneficial.

Acknowledgments

We thank Vikas Vikram for making the data available and Richard Horsley for providing the North Dakota phenotypic data. Thanks are also due to Ed Scheifelbein, Karen Beaubien, and Martin Hochhalter for providing technical assistance and Shioaman Chao for genotyping. This research was supported by USDA–NIFA Grant No. 2009–85606–05701, “Barley Coordinated Agricultural Project” and the US Wheat and Barley Scab Initiative USDA–ARS Agreement No. 59–0206–9–072. This is a cooperative project with the US Wheat and Barley Scab Initiative. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the USDA.

References

- Clark, S.A., J.M. Hickey, H.D. Daetwyler, and J.H.J. van der Werf. 2012. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genet. Sel. Evol.* 44:4. doi:10.1186/1297-9686-44-4
- 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. Szuvecs, 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
- Combs, E., and R. Bernardo. 2013. Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. *Plant Gen.* 6. doi:10.3835/plantgenome2012.11.0030
- Efron, B., and G. Gong. 1983. A leisurely look at the bootstrap, the jackknife, and cross-validation. *Am. Stat.* 37:36–48.
- Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Gen.* 4:250–255. doi:10.3835/plantgenome2011.08.0024
- Habier, D., R.L. Fernando, and D.J. Garrick. 2013. Genomic BLUP decoded: A look into the black box of genomic prediction. *Genetics* 194:597–607. doi:10.1534/genetics.113.152207
- Habier, D., J. Tetens, F. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet. Sel. Evol.* 42:5. doi:10.1186/1297-9686-42-5
- Hamblin, M.T., T.J. Close, P.R. Bhat, S.M. Chao, J.G. Kling, K.J. Abraham, T. Blake, W.S. Brooks, B. Cooper, C.A. Griffey, P.M. Hayes, D.J. Hole, R.D. Horsley, D.E. Obert, K.P. Smith, S.E. Ullrich, G.J. Muehlbauer, and J.L. Jannink. 2010. Population structure and linkage disequilibrium in US barley germplasm: Implications for association mapping. *Crop Sci.* 50:556–566. doi:10.2135/cropsci2009.04.0198
- Heffner, E.L., J. Jannink, and M.E. Sorrells. 2011. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Gen.* 4:65–75. doi:10.3835/plantgenome.2010.12.0029
- Heslot, N., H. Yang, M.E. Sorrells, and J. Jannink. 2012. Genomic selection in plant breeding: A comparison of models. *Crop Sci.* 52:146–160. doi:10.2135/cropsci2011.06.0297
- Hickey, J.M., S. Dreisigacker, J. Crossa, S. Hearne, R. Babu, B.M. Prasanna, M. Grondona, A. Zambelli, V.S. Windhausen and K.

- Mathews. 2014. Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Sci.* 54:1476–1488. doi:10.2135/cropsci2013.03.0195
- Jannink, J., A.J. Lorenz, and H. Iwata. 2010. Genomic selection in plant breeding: From theory to practice. *Brief. Funct. Genomics Proteomics* 9:166–177. doi:10.1093/bfpg/elq001
- Lehermeier, C., N. Krämer, E. Bauer, C. Bauland, C. Camisan, L. Campo, P. Flament, A.E. Melchinger, M. Menz, and N. Meyer. 2014. Usefulness of multiparental populations of maize (*Zea mays* L.) for genome-based prediction. *Genetics* 198:3–16. doi:10.1534/genetics.114.161943
- Lorenz, A.J. 2013. Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: A simulation experiment. *G3: Genes, Genomes, Genet.* 3:481–491.
- Lorenz, K., and B.A. Cohen. 2012. Small- and large-effect quantitative trait locus interactions underlie variation in yeast sporulation efficiency. *Genetics* 192:1123–1132. doi:10.1534/genetics.112.143107
- Lorenz, A.J., K.P. Smith, and J. Jannink. 2012. Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. *Crop Sci.* 52:1609–1621. doi:10.2135/cropsci2011.09.0503
- 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. *Mol. Breed.* 27:439–454. doi:10.1007/s11032-010-9442-0
- Mohammadi, M., T. Blake, A.D. Budde, S. Chao, P.M. Hayes, R.D. Horsley, D.E. Obert, S.E. Ullrich, and K. Smith. 2015. A genome-wide association study of malting quality across eight U.S. barley breeding programs. *Theor. Appl. Genet.* 128:705–721. doi:10.1007/s00122-015-2465-5
- Pszczola, M., T. Strabel, H.A. Mulder, and M.P.L. Calus. 2012. Reliability of direct genomic values for animals with different relationships within and to the reference population. *J. Dairy Sci.* 95:389–400. doi:10.3168/jds.2011-4338
- R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Riedelsheimer, C., J.B. Endelman, M. Stange, M.E. Sorrells, J. Jannink, and A.E. Melchinger. 2013. Genomic predictability of interconnected biparental maize populations. *Genetics* 194:493–503. doi:10.1534/genetics.113.150227
- 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* 192:715–728. doi:10.1534/genetics.112.141473
- Technow, F., A. Bürger, and A.E. Melchinger. 2013. Genomic prediction of northern corn leaf blight resistance in maize with combined or separated training sets for heterotic groups. *G3: Genes, Genomes, Genet.* 3:197–203.
- Van Raden, P. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423. doi:10.3168/jds.2007-0980
- 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* 193:621–631. doi:10.1534/genetics.112.146290