Genomic Predictability of Interconnected Biparental Maize Populations

Christian Riedelsheimer,* Jeffrey B. Endelman,[†] Michael Stange,* Mark E. Sorrells,[†] Jean-Luc Jannink,^{†,*}
and Albrecht E. Melchinger^{*,1}

*Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany, and

[†]Department of Plant Breeding and Genetics, and [‡]U.S. Department of Agriculture–Agricultural Research Service, R. W. Holley Center for Agriculture and Health, Cornell University, Ithaca, New York 14853

ABSTRACT Intense structuring of plant breeding populations challenges the design of the training set (TS) in genomic selection (GS). An important open question is how the TS should be constructed from multiple related or unrelated small biparental families to predict progeny from individual crosses. Here, we used a set of five interconnected maize (*Zea mays* L.) populations of doubled-haploid (DH) lines derived from four parents to systematically investigate how the composition of the TS affects the prediction accuracy for lines from individual crosses. A total of 635 DH lines genotyped with 16,741 polymorphic SNPs were evaluated for five traits including Gibberella ear rot severity and three kernel yield component traits. The populations showed a genomic similarity pattern, which reflects the crossing scheme with a clear separation of full sibs, half sibs, and unrelated groups. Prediction accuracies within full-sib families of DH lines followed closely theoretical expectations, accounting for the influence of sample size and heritability of the trait. Prediction accuracies declined by 42% if full-sib DH lines were replaced by half-sib DH lines, but statistically significantly better results could be achieved if half-sib DH lines were available from both instead of only one parent of the validation population. Once both parents of the validation population were represented in the TS, including more crosses with a constant TS size did not increase accuracies. Unrelated crosses showing opposite linkage phases with the validation population resulted in negative or reduced prediction accuracies, if used alone or in combination with related families, respectively. We suggest identifying and excluding such crosses from the TS. Moreover, the observed variability among populations and traits suggests that these uncertainties must be taken into account in models optimizing the allocation of resources in GS.

GENOMIC prediction or selection, initially proposed and rapidly implemented in animal breeding (Meuwissen *et al.* 2001), is increasingly applied in plant breeding (Bernardo and Yu 2009; Lorenz *et al.* 2011; Morrell *et al.* 2011). However, with more investigations in plant breeding applications, new challenges emerge, mainly as the result of the greater possibilities of genetic manipulation and reproduction modes in plants compared to animals. Genomic predictions within diverse populations (Crossa *et al.* 2010; Riedelsheimer *et al.* 2012a,b; Windhausen *et al.* 2012) largely overlap with the scenarios in animal breeding. Pre-

dicting crossbred performance in animals also has some similarities with the prediction of maize hybrids from fully homozygous inbred lines drawn from genetically distant heterotic pools (Technow *et al.* 2012). Apart from using mixtures of close and distant relatives, little overlap with the field of animal breeding however exists in the case of multiple related or unrelated biparental families of inbred lines with a size ≤ 200 . The latter situation is investigated in this study, because it covers the most relevant population structure encountered in practical plant breeding programs of line and hybrid cultivars.

In maize breeding, the doubled-haploid (DH) technology reduced dramatically the time necessary to obtain fully homozygous inbred lines (Prigge and Melchinger 2012), enabling the generation of a huge number of inbred lines every year. These DH lines typically originate from distinct crosses of related parents from one heterotic pool. Theoretical and experimental results showed that the mean of DH

Copyright © 2013 by the Genetics Society of America

doi: 10.1534/genetics.113.150227

Manuscript received February 8, 2013; accepted for publication March 25, 2013

Supporting information is available online at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.150227/-/DC1>.

¹Corresponding author: Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany. E-mail: melchinger@uni-hohenheim.de

lines derived from an F_1 cross can be reliably predicted from the mean of the two parent lines and this applies to both *per se* and testcross performance (Melchinger 1987; Melchinger *et al.* 2005). Traditional pedigree-based approaches are, however, unable to predict the performance of individual DH lines within a full-sib family, because all of them have the same *expected* relationship. With the availability of high-density SNP data, it is now possible to calculate the *realized* relationship that is distributed around the expected value due to Mendelian sampling (Hayes *et al.* 2009a; Hill and Weir 2011). Since differences in the mean performance of populations are in most cases known *a priori* based on information about the parents, we did not follow the approach of previous studies (Albrecht *et al.* 2011; Rutkoski *et al.* 2012; Schulz-Streeck *et al.* 2012), which analyzed genomic predictions of DH lines across multiple families. Instead, our focus was on genomic prediction of DH lines from individual biparental populations.

For implementing genomic prediction in plant breeding programs, it is crucial to know how the design of the training set (TS) affects the achievable accuracy for predicting a single biparental population. Given a set of related populations, two questions arise: (i) If the TS composes only a fraction of the DH lines from a biparental cross, what is the accuracy of genomic prediction of the remaining full-sib DH lines? (ii) How does the composition of the TS in terms of related or unrelated families affect the accuracy of genomic prediction of DH lines from a cross not included in the TS?

The first question has been investigated from a theoretical point of view (Daetwyler *et al.* 2008, 2010; Goddard and Hayes 2009; Meuwissen 2009) and with low-density marker panels also empirically in maize (Lorenzana and Bernardo 2009). It is still unclear, however, how well theoretical expectations fit observed prediction accuracies in biparental maize populations, especially concerning their robustness across several traits and populations. Here we compare theoretical with empirical results for five important traits that so far have not been investigated in the field of genomic prediction: Giberella ear rot (GER) severity caused by the fungus *Fusarium graminearum*, the content of one of its major mycotoxins, deoxynivalenol (DON), and three kernel yield component traits: ear length, kernel rows, and kernels per row.

The second question is important because knowing the predictability of progeny of crosses based on existing related material is assumed to be one of the key elements for a better allocation of resources in plant breeding. This is especially true considering the high costs of phenotyping compared to automatable genotyping platforms. If multiple related or unrelated populations are combined into one TS, analytical solutions of the expected accuracy are complicated by the intense substructuring in the TS. However, it is evident from previous studies that besides the heritability of the trait and the TS size, relatedness between the TS and the validation population (VP) and the extent to which linkage disequilibrium (LD) with quantitative trait loci (QTL) is exploited across populations are the main factors (Goddard

Figure 1 Crossing scheme to obtain the five populations.

et al. 2006; Habier *et al.* 2007, 2010; Gianola *et al.* 2009; Clark *et al.* 2012). Nevertheless, a detailed understanding how the TS composition influences the predictability of individual crosses is still lacking.

From an applied point of view, several specific questions arise concerning the construction of the TS to predict progeny of a cross with related material: Is it favorable to have more crosses included in the TS? Does it matter if related material is available only for one parent of the VP? Does a mixture of related and unrelated crosses in the TS improve the prediction accuracy or can it even have negative effects? The objective of this study was to answer these questions empirically by changing the composition of the TS in a precisely defined manner to gain insights into the factors influencing the genomic prediction accuracy of individual crosses. Thus, the specific objectives of this study were to assess and compare the following prediction problems, while keeping the size of the TS constant: (i) prediction using full-sib DH lines, (ii) prediction using one to four half-sib families of DH lines that have together either only one or both parents in common with the parents of the VP, and (iii) prediction using one unrelated population of DH lines alone or in combination with one to three half-sib families of DH lines.

Materials and Methods

Populations

The genetic material consisted of five DH populations described in detail by Martin *et al.* (2012). Four European flint inbred lines from the maize breeding program of the University of Hohenheim were used as parents to develop the populations (Figure 1). Two of the four lines were resistant to GER (P1 = UH007, P2 = UH006) and two were susceptible to GER (P3 = UH009, P4 = D152). The five populations could be clearly separated in a principal component analysis on SNP data with no overlap in the first three dimensions (Supporting Information, Figure S1).

Field trials, trait recording, and phenotypic analysis

Field trials of all lines were conducted at two locations in Southwest Germany, using adjacent 10 × 20 α-lattice

designs with two replications. Phenotyping of GER and DON was performed in 2008 and 2009 as described previously (Martin *et al.* 2011). Briefly, artificial silk channel inoculation was performed with an aggressive isolate of *F. graminearum* on six (2008) or eight (2009) primary ears 5–6 days after silk emergence. At physiological maturity, inoculated ears were dehusked manually and GER was visually recorded as the area covered by mycelium. After drying to an approximate moisture content of 14%, 100 g of ground grain was taken to record spectral data, using near-infrared spectroscopy (NIRS). DON concentration was determined by applying the NIRS calibration model developed by Bolduan *et al.* (2009). Primary ears from five noninoculated plants of each plot were used in 2009 to record yield component traits ear length, kernel rows, and kernels per row.

Standard lattice analyses of phenotypic traits were performed using the software PLABSTAT (Utz 2005) as described previously (Martin *et al.* 2011), including an arcsine square-root transformation of GER severity and a natural logarithm transformation of DON content. Heritability (h^2) was estimated on a line-mean basis with 95% confidence intervals calculated according to Knapp and Bridges (1987). Both phenotypic and genotypic data can be found in [File S1](#) and were available for the following numbers of lines: 131 ($\pi_{1,2} = P1 \times P2$), 204 ($\pi_{1,3} = P1 \times P3$), 161 ($\pi_{1,4} = P1 \times P4$), 96 ($\pi_{2,3} = P2 \times P3$), and 43 ($\pi_{2,4} = P2 \times P4$).

Genotyping and genomic similarity measures

All inbred lines were genotyped using the Illumina MaizeSNP50 array containing 56,110 SNPs (Ganal *et al.* 2011). A total of 16,741 SNPs had a call rate of at least 0.95 and an across-populations minor allele frequency of at least 0.05 and hence were used for all further analysis. Imputation of the 0.3% missing alleles was performed using the expectation-maximization (EM) algorithm (Poland *et al.* 2012) implemented in the R-package rrBLUP (Endelman 2011). Identity-by-state (IBS) genetic similarities among all n lines were calculated as

$$s_{\text{IBS}} = \frac{1}{2} \frac{(\mathbf{J} + \mathbf{Z}\mathbf{Z}^T)}{m},$$

where m is the number of SNPs, $\mathbf{J}$ is a matrix of 1's, and $\mathbf{Z}$ is a matrix of size $n \times m$ with elements z_{ij} reflecting the SNP genotype of the i th line at the j th marker locus, where z_{ij} is -1 , 0 , or $+1$ for the aa , Aa , and AA genotypes (Astle and Balding 2009).

As a second measure for genetic similarity, we calculated the Pearson correlation coefficients among rows of $\mathbf{Z}$. This genomic correlation measure, denoted s_{GC} , gives values ranging from -1 to $+1$ in the case of different or the same alleles for all SNPs, respectively, and captures the correlation pattern of the individual DH lines based on their SNP alleles.

To investigate the persistence of linkage phases across the DH populations, we considered consecutive marker pairs with ≈ 10 -Mb distance. Following Technow *et al.* (2012),

Table 1 Prediction schemes with different compositions of the training set (TS) to predict the validation population (VP)

Abbreviation	No. families included in the TS	Relationship	No. p	No. UR
1A	1	FS	2	0
1B	1	HS	1	0
1C	1	UR	0	1
2A	2	HS	2	0
2B	2	HS	1	0
2C	2	HS + UR	1	1/2
3A	3	HS	2	0
3B	3	HS + UR	2	1/3
3C	3	HS + UR	1	1/3
4A	4	HS	2	0
4B	4	HS + UR	2	1/4

FS, full sibs; HS, half sibs; UR, unrelated; No. p, number of parents of the VP that are also a parent in at least one of the families in the TS; No. UR, fraction of genotypes in the TS that belong to an UR family.

linkage phase similarity (s_{LP}) between two populations was then calculated as the proportion of marker pairs with the same linkage phase (reflected by the sign of the r statistic) in both populations.

Genomic prediction model

Genomic prediction was performed using ridge-regression best linear unbiased prediction (RR-BLUP) (Piepho 2009) implemented in the R package rrBLUP (Endelman 2011). RR-BLUP assumes normally distributed genetic effects and has been found to be robust toward trait architecture in elite maize with a high level of LD (Riedelsheimer *et al.* 2012b). The model was fitted using the equivalent and computationally efficient genomic (G-)BLUP formulation (Hayes *et al.* 2009a) with a realized relationship matrix computed according to “method I” of VanRaden (2008).

Genomic prediction cross-validation schemes

Prediction accuracies were always calculated as the Pearson correlation between observed and predicted trait values divided by the square root of the heritability of the trait in the VP.

To answer the questions stated in the objectives of this article, we varied the composition of the TS in terms of (i) relatedness to the VP ranging from full sibs (FS) and half sibs (HS) to unrelated (UR), assuming that the parents are unrelated, (ii) number of families (1–4) from which genotypes for the TS are sampled, and (iii) number of parents of the VP that are also a parent in at least one of the families in the TS. A summary of all 11 TS compositions (1A–4B) is given in Table 1. It is important to note that the VP always consisted of the DH lines from one cross with variable size.

The size of the TS ranged from 12 to a maximum of 168 in steps of 12. If the TS consisted of multiple families, sampling was done with equal proportion from each family. To obtain stable mean estimates, 100 repetitions of the sampling procedure for constructing the TS were performed. If the TS could be constructed in multiple combinations of families, the average over all combinations is reported.

Table 2 Summary of population characteristics

Characteristic	Trait	Population				
		$\pi_{1,2}$	$\pi_{1,3}$	$\pi_{1,4}$	$\pi_{2,3}$	$\pi_{2,4}$
$\bar{x}$	GER	39.21	53.06	59.40	42.97	56.58
	DON	3.42	4.84	4.87	3.70	4.64
	EL	13.25	13.37	13.94	14.08	14.34
	KR	12.60	14.67	12.92	12.84	11.86
	KPR	17.61	19.05	19.22	18.68	19.69
$\sigma_g^2 \pm \text{SE}$	GER	101.8 $\pm$ 15.31	74.0 $\pm$ 10.11	101.9 $\pm$ 13.66	131.0 $\pm$ 22.54	115.1 $\pm$ 28.79
	DON	0.399 $\pm$ 0.060	0.253 $\pm$ 0.038	0.470 $\pm$ 0.062	0.556 $\pm$ 0.100	0.438 $\pm$ 0.123
	EL	1.186 $\pm$ 0.188	1.253 $\pm$ 0.158	1.369 $\pm$ 0.251	1.220 $\pm$ 0.244	1.193 $\pm$ 0.296
	KR	1.265 $\pm$ 0.222	2.056 $\pm$ 0.213	1.789 $\pm$ 0.212	1.570 $\pm$ 0.250	0.818 $\pm$ 0.218
	KPR	4.533 $\pm$ 0.946	4.393 $\pm$ 0.635	8.326 $\pm$ 1.183	5.436 $\pm$ 1.251	3.292 $\pm$ 1.249
h^2	GER	0.77	0.70	0.79	0.82	0.83
	DON	0.77	0.64	0.80	0.80	0.74
	EL	0.75	0.76	0.61	0.73	0.84
	KR	0.69	0.91	0.89	0.89	0.79
	KPR	0.60	0.68	0.76	0.65	0.59
C.I. _{h^2}	GER	0.70–0.82	0.62–0.75	0.73–0.83	0.75–0.87	0.71–0.89
	DON	0.70–0.82	0.55–0.71	0.75–0.84	0.71–0.85	0.57–0.84
	EL	0.66–0.82	0.69–0.82	0.48–0.71	0.60–0.82	0.71–0.92
	KR	0.57–0.78	0.88–0.93	0.86–0.92	0.84–0.93	0.63–0.88
	KPR	0.44–0.71	0.58–0.75	0.68–0.82	0.48–0.77	0.26–0.77

Subscripts of the populations (π) refer to the involved parents (see Figure 1). $\bar{x}$, Average phenotypic value; σ_g^2 , genotypic variance; h^2 , heritability; C.I. _{h^2} , 95% confidence interval of h^2 ; SE, standard error. GER, Giberella ear rot severity; DON, deoxynivalenol content; EL, ear length; KR, kernel rows; KPR, kernels per row.

Comparing actual with expected within-population prediction accuracy

We used the formula suggested by Daetwyler *et al.* (2010) to compare the observed with the expected within-population prediction accuracy. It was suggested to approximate for highly polygenic quantitative traits, $\hat{r}_{\text{gg}} = \sqrt{N_T h^2 / (N_T h^2 + M_e)}$, where N_T is the number of genotypes in the training set and M_e is the effective number of loci. Following Meuwissen and Goddard (2010), M_e was estimated as $M_e = 2N_e L$, where L is the genome size in morgans. $L = 16.34$ M was adopted from a previous linkage-mapping study, using population $\pi_{1,2}$ (Martin *et al.* 2011). The effective population size (N_e) was calculated using the harmonic mean approximation for two generations (Hartl and Clark 1997, p. 291), resulting in $N_e = 3.94, 3.96, 3.95, 3.92$, and 3.82 for populations $\pi_{1,2}, \pi_{1,3}, \pi_{1,4}, \pi_{2,3}$, and $\pi_{3,4}$, respectively.

Model-based comparison of TS compositions

A model-based approach was used to test for statistical differences among different TS compositions. The total TS size was fixed to 84, which was the maximum possible size for which balanced comparisons of TS compositions were possible. The observed prediction accuracies were modeled with the linear mixed model

$$y_{ijk} = \mu + p_i + t_j + c_k + e,$$

where y_{ijk} is the obtained prediction accuracy averaged over all repetitions and possible combinations for constructing the TS; p_i is the i th VP with $i = 1, \dots, 5$; t_j is the j th trait with $j = 1, \dots, 5$; c_k is the k th TS composition with $k =$

$1, \dots, 11$; and the residual error $e \sim N(0, \sigma_e^2)$. We assumed c_k to be a fixed effect and populations and traits to be random draws from normal distributions; e.g., $p_i \sim N(0, \sigma_p^2)$, $t_j \sim N(0, \sigma_t^2)$. Diagnostic plots were used to detect possible violations of the model assumptions. Variance components were estimated using restricted maximum likelihood (REML) and tested for being >0 , using an exact restricted likelihood-ratio test as implemented in the R package RLRsim (Scheipl *et al.* 2007). All pairwise comparisons between compositions of TSs were tested using a two-sided t -test with a P -value adjustment using the Tukey method.

Results

Population characteristics and phenotypic traits

IBS similarity (s_{IBS}) values among parents ranged from 0.324 to 0.522 (19.8% of possible range), and genomic correlation (s_{GC}) values showed a broader range between -0.055 and -0.503 (27.9% of possible range, Table S1). Average IBS similarities of DH lines within populations correlated almost perfectly ($r = 0.999$) with the IBS similarities of the corresponding parents.

The populations showed an IBS similarity pattern, which reflects the crossing scheme with a clear separation of FS, HS, and UR relationship groups (Figure 2). IBS similarities among FS, HS, and UR relatedness groups showed expected differences in mean values (Figure S2). Since actual IBS sharing between two genotypes is the outcome of a stochastic process due to Mendelian sampling, expected variation in IBS values was observed with substantial overlap between UR and HS as well as HS and FS families. The average

s_{IBS} value for HS was close to the average between UR and FS (Figure S2). Across all pairs of lines, genomic correlations were strongly correlated with IBS similarities ($r = 0.92$) with clearly separable densities for FS, HS, and UR families.

Linkage phase similarities (s_{LP}) between populations were highly correlated ($r = 0.91$) with average s_{GC} between populations (Table S2). Average s_{IBS} values of pairs of populations were also highly correlated with average s_{LP} ($r = 0.93$) and s_{GC} values ($r = 0.89$).

Heritabilities were high for all traits and stable over most populations but showed in general large confidence intervals especially for yield component traits (Table 2). The average difference between the highest and lowest heritability for each trait was 0.18 with a coefficient of variation ranging from 6.7 to 11.1%.

Genotypic variances (σ_g^2) of the traits showed considerable variation among populations and correlations with average s_{IBS} ranged from -0.17 for kernels per row to -0.61 for DON content (Table 2).

Prediction using full sibs from the same cross (1A)

A nonlinear increase in prediction accuracy with increasing size of the TS was observed for all traits within all populations (Figure 3). The higher the initial accuracy obtained with the smallest TS size of 12, the faster was the gain in additional prediction accuracy, which leveled off with increasing TS size. At a TS size of 84, the difference among traits between highest and lowest prediction accuracy ranged from 0.20 for $\pi_{2,3}$ to 0.39 for $\pi_{1,4}$. At this TS size, Spearman's rank correlation between accuracies and heritabilities was found to be 0.45 ($P = 0.044$). Within the evaluated range in TS size, only the prediction accuracies for kernel rows in $\pi_{1,3}$ reached the level of phenotypic selection accuracy (square root of heritability). For all except the smallest population ($\pi_{2,4}$), a good agreement between observed and theoretically expected within-population prediction accuracy (Figure 4) was found. However, some traits were systematically over- or underestimated.

Comparison of TS compositions

Averaged over traits, all TS compositions without FS (2B–4B) showed initial accuracies and slopes markedly lower than with FS (Figure 5). General trends were observed when comparing scenarios 1A (1 FS) with 1B (1 HS) and 1C (1 UR). Prediction accuracies obtained with one HS family were for all traits, all populations, and all sample sizes lower than prediction with full sibs (Figure 5, Figure S4, Figure S5, Figure S6, Figure S7, and Figure S8). If the TS comprised one UR population, the lowest prediction accuracies for all TS sizes were observed in 12 of 20 possible trait $\times$ population combinations and in 7 combinations, this yielded even negative accuracies with a decreasing trend with increasing TS size (Figure S4B, Figure S4E, Figure S5C, Figure S6E, Figure S7C, Figure S7D, and Figure S8E). Most pronounced was this observation in population $\pi_{2,4}$ for traits GER (Figure S4E) and kernels per row (Figure S8E).

Figure 2 Heatmap showing IBS similarities (red, above diagonal) and genomic correlations (blue, below diagonal) among all genotypes. The numbers in the blocks refer to average IBS similarities within and between populations. Average genomic correlations in comparisons with linkage phase similarities can be found in Table S2. FS, full sibs; HS, half sibs; UR, unrelated.

A good agreement was found with the assumptions underlying the model-based statistical testing of the influence of the TS composition on prediction accuracy (Figure S3). Quantile–quantile (Q-Q) plots showed one outlier for factor VP. Setting all factors of the model as random, REML-based estimated variance components revealed that factors VP and trait explained together 32.4%, whereas factor TS composition explained 41.7% (Table 3).

Least-squares estimates of prediction accuracies obtained with the different TS compositions with a fixed maximum possible TS size of 84 showed large 95% confidence intervals across traits and populations (Table 4) with the highest estimate obtained for full sibs. Pairwise comparisons between TS compositions were significant in several specific cases (Table 4 and Table S3). Using prediction with FS (1A) as a baseline, prediction using one HS family (1B) declined significantly from 0.59 to 0.25 and dropped significantly to almost zero (0.05) with one UR family. Having both parents of the VP represented in the TS resulted in significantly better prediction accuracies than having only one common parent between the VP and the crosses involved in the TS. With two populations (comparison 2A – 2B), the increase was 36.3% and with three populations (comparison 3B – 3C) it was 38.7%. Once both parents were represented in the TS, no significant differences were found if the TS consisted of more than two families. If an UR family was part of a TS involving otherwise one, two, or three HS families, no significant change in accuracy was observed. However, replacing a HS family with an UR family led to a significant reduction in

Figure 3 (A–D) Prediction accuracies for within-population prediction using full sibs (1A). Shown are the average values for all traits obtained from 100 repetitions. The dashed lines indicate phenotypic accuracies calculated as the square root of the heritability.

prediction accuracy, if both parents of the VP were no longer represented in the TS (comparison 2A – 2C).

Discussion

The objective of this study was to systematically investigate the influence of the composition of the TS on the genomic prediction accuracies in an interconnected biparental population scenario. Independent of the TS composition, we focused on predicting DH lines from a single cross and not across different crosses. Reasoning for this lies in the well-established finding that the mean of DH lines derived from an F₁ cross can be reliably predicted from the mean of the two parent lines (Melchinger 1987; Melchinger *et al.* 2005). In addition, it can be shown that population means estimated this way can already accurately predict DH lines across populations. Let y_{ij} be the performance of DH line j from cross i . Then, regressing estimated population means (c_i) on y_{ij} yields the model

$$y_{ij} = \mu + c_i + d_{ij},$$

where μ is the overall mean and d_{ij} is the deviation from y_{ij} to c_i . The total variation among all DH lines (σ_y^2) can be decomposed into variation between crosses (σ_c^2) and among DH lines within crosses (σ_d^2). Assuming that the parents of the crosses are unrelated, quantitative genetic theory yields $\sigma_y^2 = 2\sigma_A^2$ and $\sigma_c^2 = \sigma_d^2 = \sigma_A^2$, where σ_A^2 is the additive genetic variance of an outbred reference population. Assuming a purely additive model and that the parent means are perfectly known, the correlation between c_i and y_{ij} can be derived as

$$r(y_{ij}, c_i) = \frac{\sigma_A^2}{\sqrt{2\sigma_A^2 \times \sigma_A^2}} = \frac{1}{\sqrt{2}} = 0.71.$$

Thus, if population means estimated from the parents are used to predict DH lines from multiple crosses, a prediction accuracy of 0.71 can already be expected. This correlation is

Figure 4 (A and B) Observed vs. expected within-population prediction accuracy using full sibs (1A) averaged over (A) all populations except the smallest one ($\pi_{2,4}$) and (B) traits. The expected prediction accuracy was calculated using the formula suggested by Daetwyler *et al.* (2010), assuming an average genome size $L = 16.34$ M and effective number of loci $M_e = 2N_eL$. Effective population size (N_e) was calculated using the harmonic mean approximation for two generations.

used in scenarios where prediction is evaluated across populations but not if DH lines are predicted within a single cross, which was the focus of this study. To disentangle the factors influencing prediction accuracies of progeny from a genotyped but not a phenotyped cross, TSs were designed in 11 defined ways. Inferences about the influence of the TS design were drawn from a model-based approach while keeping the total TS size constant.

Prediction using full sibs (1A)

For most trait $\times$ population combinations, full-sib prediction did not reach phenotypic accuracy measured as the square root of heritability (Figure 3). However, correlated response to selection also depends on selection cycle length and the intensity of selection (Falconer and Mackay 1996). The latter can be increased markedly by shifting resources from phenotyping toward production and genotyping of many more DH lines than under traditional phenotypic selection.

For full sibs, prediction accuracy was strongly dependent on TS size and moderately dependent on trait heritability. This was expected from theoretical simulations (Meuwissen 2009), as well as from empirical work in wheat (Heffner *et al.* 2011a,b), barley (Lorenz *et al.* 2012), and oats (Asoro *et al.* 2011). Within an individual biparental population, genomic prediction information can come from (i) variation in relatedness (*i.e.*, Mendelian sampling) captured by the SNPs and (ii) cosegregation of loci linked to QTL (linkage information). Artifacts originating from population structure can be assumed to be absent, which is empirically supported by (i) the homogeneous data clouds of the populations in the principal components analysis (Figure S1) and (ii) the unstructured distribution of IBS values (Figure 2 and Figure S2). Hence, within populations, it should be possible to model the expected prediction accuracy ($\hat{r}_{gg}$) by population and trait parameters.

In dairy cattle, deterministic predictions could be successfully used as a guide for the required TS size (Hayes *et al.* 2009b). Our results support this also for within-population prediction using full sibs (Figure 4). One possible reason for

the systematic over- and underestimations of traits lies in deviations from the polygenic genetic architecture, which is assumed in the applied deterministic formula. Indeed, genome partitioning of genetic variance indicates a less polygenic architecture for certain trait $\times$ population combinations (Figure S9). On the other side, the robustness of RR-BLUP toward the underlying number of QTL in a high LD scenario (Riedelsheimer *et al.* 2012b) suggests that differences in the number of underlying QTL play a less important role here. A more likely explanation lies in the imprecise heritability estimates reflected by their large confidence intervals (Table 4). This was especially true for the yield component traits that were evaluated in only 1 year and showed the strongest deviation from the expected values. This result emphasizes the need for high-quality phenotypic data for the TS for genomic prediction.

Nevertheless, for highly heritable traits, the high correlations between actual and observed values suggest that within biparental populations, prediction accuracies can be predicted well using information of (i) training set size, (ii) heritability, and (iii) effective number of loci. Given that good heritability estimates are available, it seems possible to use deterministic formulas to optimize resource allocation between TSs and VPs for prediction using full sibs.

Notably, the minimum training set sizes required to achieve reasonable prediction accuracies within biparental maize DH populations are by at least two orders of magnitude smaller than those required in animal breeding (Goddard and Hayes 2009). This results mostly because of the smaller genome size (in morgans) and the much smaller N_e of biparental populations compared to, *e.g.*, cattle for which $N_e \approx 100$ (Sørensen *et al.* 2005). Thus, results from simulation or empirical studies in animal breeding should therefore be examined carefully before drawing inferences for situations in plant breeding.

Influence of TS composition

Prediction within populations using full sibs is appealing because of (i) the high accuracies achievable with a moderate TS size and (ii) the good agreement of theoretical

Figure 5 (A–E) Prediction accuracies for individual validation populations (VP) depending on the composition of the training set (TS) and the total TS size. Shown are the mean values over all traits, repetitions, and possible combinations to construct the TS. The dashed lines indicates the TS size at which a model-based statistical comparison of TS compositions was performed. See Table 1 for details about the different TS compositions. Results for individual traits can be found in Figure S4, Figure S5, Figure S6, Figure S7, and Figure S8.

expectations with empirical results. However, in an ongoing breeding program, phenotypic and genotypic data are available for multiple related and/or unrelated populations often before the actual cross is being made. With the current body of theoretical knowledge in the field of genomic prediction, it is not possible to answer the question of how the composition of the TS affects the genomic prediction of other crosses, which is needed to help breeders in their decision-making process. We therefore followed an empirical model-based approach with statistical testing for significant differences between TS compositions, while keeping the TS size, and hence total resource capacity, constant. TS size had to be fixed to a rather small number because 84 was the maximum possible size for which balanced comparisons among all TS compositions were possible. However, relative differences between TS compositions for which larger TS sizes were possible remained relatively constant beyond this value because prediction accuracies began to level off at this TS size (Figure 5).

The strong decline in prediction accuracy from FS (1A) to HS (1B) and to UR (1C) was expected from theory (Habier *et al.* 2007) as well as from recent empirical studies emphasizing that the level of relatedness between genotypes in the TS and the VP has a strong impact on prediction accuracies

(Habier *et al.* 2010; Clark *et al.* 2012). Having multiple related HS families, one important question is whether resources should be spent to increase the number of crosses or the number of DH lines per cross. Here, we found no statistical difference between different numbers of involved crosses once both parents of the VP are parents of at least one cross involved in the TS. In addition, increasing the number of HS families with one parent in common with the VP did not compensate for the lack of representation of the second parent of the VP. In a genomic prediction scenario with HS families, our results suggest that an optimal TS should (i) be of sufficient total size and (ii) have both parents of the VP represented. The number of families used to generate such a TS, however, does not seem to have a strong impact (Table 4). Nevertheless, real or simulated data sets with many more HS families should be analyzed to investigate this question in greater detail.

One interesting finding of our study was the negative accuracies obtained with UR families for certain traits, especially in case of $\pi_{2,4}$ as the VP. We note here that the effect of an UR family could not be evaluated for population $\pi_{1,2}$ because its UR family $\pi_{3,4}$ is not present in our study. It is also important that the term “unrelated” was used to distinguish this type of cross from half sibs and full sibs. UR

Table 3 Estimated variance components and proportion of total variance for all factors in the model-based analysis of the influence of TS composition on prediction accuracy

Factor	Variance component $\times 10^2$	d.f.	<i>P</i>	Portion of total variance (%)
VP	0.197	4	7.1×10^{-7}	4.8
Trait	1.125	4	$<2.2 \times 10^{-16}$	27.6
TS composition	1.700	10	$<2.2 \times 10^{-16}$	41.7
Error	1.053	186		25.8

TS size was fixed to 84, which was the maximum possible size for which estimates of all TS compositions were possible. d.f., degrees of freedom.

families still showed a substantial IBS similarity with the VP (Table S2) and were therefore not unrelated in a population genetic sense. However, since prediction accuracies became increasingly negative with increasing TS size, the UR family must have provided a negative prediction signal coming from opposite linkage phases with important QTL in the VP. Genome partitioning of genetic variance following Riedelsheimer *et al.* (2012b) suggests that important QTL for the measured traits are indeed especially present in population $\pi_{2,4}$ (Figure S9). This also indicates that causal variants might not segregate in all DH populations. In addition, linkage phase similarity measured as s_{LP} between population $\pi_{2,4}$ and its UR family was among the lowest ones of all combinations and genomic correlation measured as s_{GC} was negative (-0.233 , Table S2). If such an UR family is combined with one HS family to a mixed TS, the negative predictive signal coming from the UR family could even counteract the positive signal coming from the HS family (decreasing curve of TS composition 2C compared to 1B in Figure S4E). Therefore it is recommended to exclude such families from the TS. This finding is also in agreement with a simulation study showing that increasing accuracies by combining different populations into a single TS requires persistence of marker–QTL LD across populations (de Roos *et al.* 2009).

The strong correlation ($0.89 < r < 0.93$) among s_{IBS} , s_{GC} , and s_{LP} suggests that any of these measures can be used to estimate genetic similarity between populations to identify populations with negative predictive signal. However, we found the s_{GC} measure appealing because of its simplicity, applicability to individual genotypes, and defined range of values that fall between -1 and $+1$. This measure might also be suitable in identifying outlier genotypes in the TS or the VP. In the case of monogenic or oligogenic genetic architecture with known positions of major QTL, a similarity measure that takes linkage phases in the vicinity of these QTL into account might be even more accurate than a simple genome-wide measure. Future simulation studies with large sets of related populations and known QTL positions and effects could shed more light on this issue.

Replacing a HS family with an UR family led to significant reductions in prediction accuracy only if one parent of the VP was no longer represented in the TS (comparison 2A – 2C); otherwise the difference was not significant. With a constant size of the TS, a drop in accuracy might be expected if the TS is diluted with unrelated genotypes, especially if they show reverse linkage phases. However, since all line crosses

in our study originated from only four parents, the DH lines from every UR family were necessarily HS to both HS families included in the TS. For a single locus with additive effect a , let

$$\begin{aligned} A \times B : 2a_1 &= \mu_{AA} - \mu_{BB} \\ A \times C : 2a_2 &= \mu_{AA} - \mu_{CC} \\ B \times D : 2a_3 &= \mu_{BB} - \mu_{DD} \\ C \times D : 2a_4 &= \mu_{CC} - \mu_{DD}. \end{aligned}$$

Then, it follows $\mu_{AA} - \mu_{BB} = 2a_2 + 2a_4 - 2a_3 = 2a_1$. Thus, with the assumption of no epistasis, additive effects in one population can be derived from a linear combination of the additive effects in the HS families of each parent and the UR family. In a general setting with more than four parents, the inclusion of an UR family might, in contrast to our results, reduce prediction accuracy even if both parents are represented. Careful investigation of the linkage phase similarities (or genomic correlations) between UR families and the VP should therefore be carried out before such families are included in the TS.

Influence of population and trait on TS composition

Systematic population and trait effects accompany the described contrasts among the TS compositions. Variability among traits was substantially higher than variability among VPs (Table 3) and both VP and trait effects accounted for 34.9%, which was of similar magnitude compared to the variation introduced due to different TS compositions (41.7%). This was not unexpected as previous studies also found large variability in prediction accuracy attributable to traits (corrected for their heritability) and populations (Ornella *et al.* 2012).

Although differences between TS compositions could be quantified and tested, the population and trait variability hampered a direct quantification of interchangeabilities, *e.g.*, how many HS from one or more families are needed to obtain the same prediction accuracy as with a certain number of FS. For example, whereas 48 HS from two families were sufficient to replace 24 FS in the TS for predicting GER severity in population $\pi_{1,4}$ (Figure S4C), 72 HS were needed for population $\pi_{1,3}$ (Figure S4B), and 168 HS were barely enough in population $\pi_{1,2}$ (Figure S4A). In addition, these numbers were in most cases not transferable to other traits. This variability among traits and populations has important implications for resource optimizations. If such calculations are based on only one cross or trait, results can be misleading.

Table 4 Least-squares estimates of accuracies obtained with different TS compositions at a fixed TS size of 84

TS composition	Significant difference ^a	Estimate	SE	d.f.	95% C.I.
1A (1 FS)	a	0.59	0.06	7.36	0.459–0.732
1B (1 HS)	bc	0.25	0.06	10.34	0.114–0.387
1C (1 UR)	d	0.05	0.06	8.32	–0.083–0.184
2A (2 HS, no. p = 2)	b	0.34	0.06	6.82	0.204–0.457
2B (2 HS, no. p = 1)	c	0.21	0.06	6.82	0.083–0.346
2C (1 HS + 1 UR)	c	0.21	0.06	7.39	0.078–0.342
3A (3 HS)	b	0.33	0.06	6.82	0.195–0.458
3B (2 HS, no. p = 2 + 1 UR)	b	0.31	0.06	7.39	0.178–0.442
3C (2 HS, no. p = 1 + 1 UR)	c	0.19	0.06	7.39	0.058–0.323
4A (4 HS)	b	0.31	0.07	18.80	0.162–0.463
4B (3 HS + 1 UR)	b	0.30	0.06	7.39	0.171–0.435

SE, standard error of estimate; d.f., degrees of freedom; C.I., confidence interval; no. p, number of parents of the VP that are also a parent in at least one of the families in the TS.

^a Estimates with the same letter are not significantly different ($P < 0.05$).

Therefore, any optimization calculation should be based on a large set of crosses and results for one trait should be extrapolated to another trait only with utmost caution. Further research should focus on elucidating in more detail the reasons underlying the variability of the prediction accuracy among traits and populations. Possible factors to be investigated with very large sets of simulated or real populations and traits include differences in genetic architecture with different linkage phases between TSs and VPs in regions harboring important QTL for the trait of interest.

Conclusions

From the analysis of our set of five interconnected biparental DH populations, we draw the following conclusions:

- Within-population prediction is feasible for highly heritable traits with moderate training set sizes and agrees on average well with theoretical expectations.
- Prediction accuracies decline strongly if FS are replaced by HS, but statistically significantly better results can be achieved if HS families are available from both instead of only one parent of the VP.
- Once both parents are represented, it is not favorable to increase the number of crosses.
- Unrelated families with opposite linkage phase similarities with the VP can have negative prediction accuracies or reduced prediction accuracies if combined with related material. Such crosses should be identified using linkage phase or genomic correlation measures and excluded from the TS.
- The observed variability among populations and traits requires resource optimization that incorporates knowledge about the variability of parameters.

Acknowledgments

This research was funded by the German Federal Ministry of Education and Research within the Agro-ClustEr “Synbreed—Synergistic plant and animal breeding” (grant 0315528D) and by the Deutsche Forschungsgemeinschaft research grant ME 2260/6-1. Funding to J.L.J. and M.E.S.

in support of this research came from a Bill and Melinda Gates Foundation grant “Genomic selection: the next frontier for rapid genetic gains in maize and wheat.”

Literature Cited

- Albrecht, T., V. Wimmer, H.-J. Auinger, M. Erbe, C. Knaak *et al.*, 2011 Genome-based prediction of testcross values in maize. *Theor. Appl. Genet.* 123: 339–350.
- Asoro, F. G., M. A. Newell, W. D. Beavis, M. P. Scott, and J.-L. Jannink, 2011 Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *Plant Genome* 4: 132–144.
- Astle, W., and D. J. Balding, 2009 Population structure and cryptic relatedness in genetic association studies. *Stat. Sci.* 24: 451–471.
- Bernardo, R., and J. Yu, 2009 Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* 47: 1082–1090.
- Bolduan, C., J.-M. Montes, B. S. Dhillon, V. Mirdita, and A. E. Melchinger, 2009 Determination of mycotoxin concentration by ELISA and near-infrared spectroscopy in fusarium-inoculated maize. *Cereal Res. Commun.* 37: 521–529.
- Clark, S. A., J. M. Hickey, H. D. Daetwyler, and J. H. 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.
- Crossa, J., G. D. L. Campos, P. Pérez, D. Gianola, J. Burgueño *et al.*, 2010 Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186: 713–724.
- 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*: e3395.
- Daetwyler, H. D., R. Pong-Wong, B. Villanueva, and J. A. Woolliams, 2010 The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185: 1021–1031.
- de Roos, A. P. W., B. J. Hayes, and M. E. Goddard, 2009 Reliability of genomic predictions across multiple populations. *Genetics* 183: 1545–1553.
- Endelman, J. B., 2011 Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome* 4: 250–255.
- Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics*, Ed. 4. Longmans Green, Harlow, Essex, UK.
- Ganal, M. W., G. Durstewitz, A. Polley, A. Bérard, E. S. Buckler *et al.*, 2011 A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS ONE* 6: e28334.

- Gianola, D., G. de los Campos, W. G. Hill, E. Manfredi, and R. Fernando, 2009 Additive genetic variability and the Bayesian alphabet. *Genetics* 183: 347–363.
- Goddard, M. E., and B. J. Hayes, 2009 Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10: 381–391.
- Goddard, M. E., B. Hayes, H. McPartlan, and A. J. Chamberlain, 2006 Can the same genetic markers be used in multiple breeds? Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brazil, August 13–18, 2006. CD-ROM communication no. 22-16.
- Habier, D., R. L. Fernando, and J. C. M. Dekkers, 2007 The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177: 2389–2397.
- Habier, D., J. Tetens, F.-R. 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.
- Hartl, D. L., and A. G. Clark, 1997 *Principles of Population Genetics*. Sinauer Associates, Sunderland, MA.
- Hayes, B. J., P. M. Visscher, and M. E. Goddard, 2009a Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res.* 91: 47–60.
- Hayes, B. J., H. D. Daetwyler, P. J. Bowman, G. Moser, B. Tier *et al.*, 2009b Accuracy of genomic selection: comparing theory and results. Proceedings of the Association for the Advancement of Animal Breeding and Genetics, Barossa Valley, South Australia, pp. 34–37.
- Heffner, E. L., J.-L. Jannink, H. Iwata, E. Souza, and M. E. Sorrells, 2011a Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop Sci.* 51: 2597–2606.
- Heffner, E. L., J.-L. Jannink, and M. E. Sorrells, 2011b Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome* 4: 65–75.
- Hill, W. G., and B. S. Weir, 2011 Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genet. Res.* 93: 47–64.
- Knapp, S. J., and W. C. Bridges, 1987 Confidence interval estimators for heritability for several mating and experimental designs. *Theor. Appl. Genet.* 73: 759–763.
- Lorenz, A. J., S. M. Chao, F. G. Asoro, E. L. Heffner, T. Hayashi *et al.*, 2011 Genomic selection in plant breeding: knowledge and prospects, pp. 77–123 in *Advances in Agronomy*, Vol. 110, edited by D. L. Sparks. Elsevier–Academic Press, San Diego.
- Lorenz, A. J., K. P. Smith, and J.-L. Jannink, 2012 Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. *Crop Sci.* 52: 1609–1621.
- Lorenzana, R. E., and R. Bernardo, 2009 Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor. Appl. Genet.* 120: 151–161.
- Martin, M., T. Miedaner, B. S. Dhillon, U. Ufermann, B. Kessel *et al.*, 2011 Colocalization of QTL for Gibberella ear rot resistance and low mycotoxin contamination in early European maize. *Crop Sci.* 51: 1935–1945.
- Martin, M., B. S. Dhillon, T. Miedaner, and A. E. Melchinger, 2012 Inheritance of resistance to Gibberella ear rot and deoxynivalenol contamination in five flint maize crosses. *Plant Breed.* 131: 28–32.
- Melchinger, A. E., 1987 Expectation of means and variances of testcrosses produced from F2 and backcross individuals and their selfed progenies. *Heredity* 59: 105–115.
- Melchinger, A. E., C. F. H. Longin, H. F. Utz, and J.-C. Reif, 2005 Hybrid maize breeding with doubled haploid lines: quantitative genetic and selection theory for optimum allocation of resources. Proceedings of the 41st Illinois Corn Breeders' School, University of Illinois at Urbana–Champaign, Urbana, IL, pp. 8–21.
- Meuwissen, T. H. E., 2009 Accuracy of breeding values of “unrelated” individuals predicted by dense SNP genotyping. *Genet. Sel. Evol.* 41: 35.
- Meuwissen, T. H. E., and M. Goddard, 2010 Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics* 185: 623–631.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.
- Morrell, P. L., E. S. Buckler, and J. Ross-Ibarra, 2011 Crop genomics: advances and applications. *Nat. Rev. Genet.* 13: 85–96.
- Ornella, L., S. Singh, P. Perez, J. Burgueño, R. Singh *et al.*, 2012 Genomic prediction of genetic values for resistance to wheat rusts. *Plant Genome* 5: 136–148.
- Piepho, H. P., 2009 Ridge regression and extensions for genome-wide selection in maize. *Crop Sci.* 49: 1165–1176.
- Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S. Wu *et al.*, 2012 Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* 5: 103–113.
- Prigge, V., and A. E. Melchinger, 2012 Production of haploids and doubled haploids in maize, pp. 161–172 in *Plant Cell Culture Protocols, Methods in Molecular Biology*, Vol. 877, edited by V. M. Loyola-Vargas and N. Ochoa-Alejo. Humana Press, New York.
- Riedelsheimer, C., A. Czedik-Eysenberg, C. Grieder, J. Lisec, F. Technow *et al.*, 2012a Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat. Genet.* 44: 217–220.
- Riedelsheimer, C., F. Technow, and A. E. Melchinger, 2012b Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines. *BMC Genomics* 13: 452.
- Rutkoski, J., J. Benson, Y. Jia, G. Brown-Guedira, J.-L. Jannink *et al.*, 2012 Evaluation of genomic prediction methods for Fusarium head blight resistance in wheat. *Plant Genome* 5: 51–61.
- Scheipl, F., S. Greven, and H. Küchenhoff, 2007 Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Comput. Stat. Data Anal.* 52: 3283–3299.
- Schulz-Streeck, T., J. O. Ogutu, Z. Karaman, C. Knaak, and H. P. Piepho, 2012 Genomic selection using multiple populations. *Crop Sci.* 52: 2453–2461.
- Sørensen, A. C., M. K. Sørensen, and P. Berg, 2005 Inbreeding in Danish dairy cattle breeds. *J. Dairy Sci.* 88: 1865–1872.
- Technow, F., C. Riedelsheimer, T. A. Schrag, and A. E. Melchinger, 2012 Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. *Theor. Appl. Genet.* 125: 1181–1194.
- Utz, H. F., 2005 *PLABSTAT—A Computer Program for Statistical Analysis of Plant Breeding Experiments*. 3A. Universität Hohenheim, Stuttgart.
- VanRaden, P. M., 2008 Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91: 4414–4423.
- Windhausen, V., G. N. Atlin, J. Crossa, J. M. Hickey, P. Grudloyma *et al.*, 2012 Effectiveness of genomic prediction of maize hybrid performance in different populations and environments. *G3: Genes, Genomes, Genetics* 2: 1427–1436.

Communicating editor: D. J. de Koning

GENETICS

Supporting Information

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.150227/-/DC1>

Genomic Predictability of Interconnected Biparental Maize Populations

**Christian Riedelsheimer, Jeffrey B. Endelman, Michael Stange, Mark E. Sorrells, Jean-Luc Jannink,
and Albrecht E. Melchinger**

Figure S1 Principal component analysis of the 5 populations using SNP data. Shown are results for the first two (A) or second and third (B) principal components.

Figure S2 Smoothed scatter plot of genomic correlation against IBS similarity of all pairs of genotypes. Contour lines for the three relationship groups were obtained from a two-dimensional kernel density estimation. Histograms of IBS similarity and genomic correlations are shown on the top and right, respectively.

Figure S3 Diagnostic plots for model-based analysis of the influence of TS composition on prediction accuracy.

Figure S4 Prediction accuracies for trait GER severity for individual validation populations (VP) depending on the composition of the training set (TS) and the total TS size. Shown are the mean values over 100 repetitions and all possible combinations to construct the TS. The dashed line indicates the TS size at which a model-based statistical comparison of TS compositions was performed. See Table 1 for details about the different TS compositions. Results for the other traits can be found in Figures 4 and S5-S8.

Figure S5 Prediction accuracies for trait DON content for individual validation populations (VP) depending on the composition of the training set (TS) and the total TS size. Shown are the mean values over 100 repetitions and all possible combinations to construct the TS. The dashed line indicates the TS size at which a model-based statistical comparison of TS compositions was performed. See Table 1 for details about the different TS compositions. Results for the other traits can be found in Figures 4, S4, and S6-S8.

Figure S6 Prediction accuracies for trait ear length for individual validation populations (VP) depending on the composition of the training set (TS) and the total TS size. Shown are the mean values over 100 repetitions and all possible combinations to construct the TS. The dashed line indicates the TS size at which a model-based statistical comparison of TS compositions was performed. See Table 1 for details about the different TS compositions. Results for the other traits can be found in Figures 4, S4-S5, and S7-S8.

Figure S7 Prediction accuracies for trait kernel rows for individual validation populations (VP) depending on the composition of the training set (TS) and the total TS size. Shown are the mean values over 100 repetitions and all possible combinations to construct the TS. The dashed line indicates the TS size at which a model-based statistical comparison of TS compositions was performed. See Table 1 for details about the different TS compositions. Results for the other traits can be found in Figures 4, S4-S6, and S8.

Figure S8 Prediction accuracies for trait kernels per row for individual validation populations (VP) depending on the composition of the training set (TS) and the total TS size. Shown are the mean values over 100 repetitions and all possible combinations to construct the TS. The dashed line indicates the TS size at which a model-based statistical comparison of TS compositions was performed. See Table 1 for details about the different TS compositions. Results for the other traits can be found in Figures 4, and S4-S7.

Figure S9 Genome partitioning of genetic variance. Genetic variance explained by individual chromosomes was calculated following Riedelsheimer *et al.* (2012b). The numbers in the points correspond to the chromosome numbers.

File S1

Genotypes and phenotypes

File S1 is available for download at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.150227/-/DC1>.

Table S1 Genetic similarities measured as identity-by-state (s_{IBS} , above diagonale) and genomic correlations (s_{GC} , below diagonale) among parents of DH populations. Calculations were based on the 17,803 SNPs which were polymorphic across the four parents. Both measures were correlated with $r = 0.71$.

	P1	P2	P3	P4
P1	1	0.522	0.579	0.414
P2	-0.055	1	0.494	0.399
P3	-0.251	-0.083	1	0.324
P4	-0.379	-0.228	-0.503	1

Table S2 Linkage phase similarities of 10 Mb distant marker pairs (s_{LP} , above diagonale) and average genomic correlations (s_{GC} , below diagonale) among the five populations. Both measures correlated with $r = 0.91$.

	$\pi_{1,2}$	$\pi_{1,3}$	$\pi_{1,4}$	$\pi_{2,3}$	$\pi_{2,4}$
$\pi_{1,2}$	1	0.983	0.911	0.968	0.966
$\pi_{1,3}$	0.026	1	0.852	0.947	0.769
$\pi_{1,4}$	-0.066	-0.098	1	0.811	0.917
$\pi_{2,3}$	0.177	0.147	-0.251	1	0.949
$\pi_{2,4}$	0.134	-0.233	0.092	0.095	1

Table S3 Pairwise comparisons of training set (TS) compositions relevant for answering questions (a) – (e). TS size was fixed to 84 which was the maximum possible size for which comparisons of all TS compositions were possible. The linear model with prediction accuracy as response variable included the factors population, trait and its interaction as random effects and TS composition as fixed effects (see main text). *P*-values from two-side *t*-tests were adjusted using the Tukey method. SE, standard error of estimate; DF, degrees of freedom calculated according to Kenward and Roger (1997).

Comparison	Estimate	SE	DF	t-ratio	<i>P</i> -value
<i>(a) How does prediction accuracy decline when moving from FS to HS and UR?</i>					
1A - 1B	0.341	0.04	184.68	8.19	0.000***
1B - 1C	0.200	0.04	185.36	4.60	0.000***
1A - 1C	0.541	0.04	185.23	14.92	0.000***
<i>(b) Is it favorable to have both parents p_i and p_j of the VP represented in the TS?</i>					
2A - 2B	0.121	0.03	186.01	4.18	0.002**
3B - 3C	0.120	0.03	186.01	3.69	0.013*
<i>(c) Is it favorable to have more crosses included in the TS?</i>					
2A - 3A	0.010	0.03	186.01	0.34	1.000
3A - 4A	0.013	0.05	179.06	0.24	1.000
2A - 4A	0.023	0.05	179.06	0.42	1.000
<i>(d) Does adding one UR cross change the accuracy if total TS size is kept constant?</i>					
1B - 2C	0.040	0.04	185.77	0.98	0.996
2A - 3B	0.026	0.03	185.97	0.83	0.999
2B - 3C	0.024	0.03	185.97	0.77	1.000
3A - 4B	0.023	0.03	185.97	0.75	1.000
<i>(e) When and to what extent is prediction accuracy reduced if a related cross is replaced by an UR cross?</i>					
2A - 2C	0.126	0.03	185.97	4.03	0.004**
2B - 2C	0.004	0.03	185.97	0.14	1.000
3A - 3B	0.016	0.03	185.97	0.51	1.000
4A - 4B	0.010	0.06	172.15	0.18	1.000

*** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$