



Genomic prediction with multiple biparental families

Pedro C. Brauner¹ · Dominik Müller¹ · Willem S. Molenaar¹ · Albrecht E. Melchinger¹

Received: 3 May 2019 / Accepted: 18 September 2019 / Published online: 8 October 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Key message For genomic prediction within biparental families using multiple biparental families, combined training sets comprising full-sibs from the same family and half-sib families are recommended to reach high and robust prediction accuracy, whereas inclusion of unrelated families is risky and can have negative effects.

Abstract In recycling breeding, where elite inbreds are recombined to generate new source material, genomic and phenotypic information from lines of numerous biparental families (BPFs) is commonly available for genomic prediction (GP). For each BPF with a large number of candidates in the prediction set (PS), the training set (TS) can be composed of lines from the same full-sib family or multiple related and unrelated families to increase the TS size. GP was applied to BPFs generated in silico and from two published experiments to evaluate the prediction accuracy (ρ) of different TS compositions. We compared ρ for individual pairs of BPFs using as TS either full-sib, half-sib, or unrelated BPFs. While full-sibs yielded highly positive ρ and half-sibs also mostly positive ρ values, unrelated families had often negative ρ , and including these families in a combined TS reduced ρ . By simulations, we demonstrated that optimized TS compositions exist, yielding 5–10% higher ρ than the TS including all available BPFs. However, identification of poorly predictive families and finding the optimal TS composition with various quantitative-genetic parameters estimated from available data was not successful. Therefore, we suggest omitting unrelated families and combining in the TS full-sib and few half-sib families produced by specific mating designs, with a medium number (~50) of genotypes per family. This helps in balancing high ρ in GP with a sufficient effective population size of the entire breeding program for securing high short- and long-term selection progress.

Introduction

Genomic prediction (GP; see Supplementary Table 1 for a list of abbreviations) was pioneered by Meuwissen et al. (2001) for estimating breeding values in cattle breeding. It has proven to be a powerful tool for improving the efficiency of selection in animal and plant breeding (García-Ruiz et al. 2016; Crossa et al. 2017). GP consists of training a statistical

model by using phenotypic and genomic information from genotypes of a so-called training set (TS) to predict the breeding value of candidates with only genomic information, the prediction set (PS). Prediction accuracy (ρ), defined as the correlation between the genomic estimated breeding values (GEBVs) and true breeding values (TBVs), represents a key parameter when comparing the expected response from phenotypic and genomic selection (Isidro et al. 2015).

According to theoretical results, ρ depends on the trait heritability (h^2), the TS size (N_{TS}), and the effective number of chromosome segments (Daetwyler et al. 2008; 2010). Additionally, information from close relatives in the TS has a strong positive influence on the accuracy of GEBVs (Habier et al. 2007; Clark et al. 2012). In animal breeding of major breeds such as Holstein–Friesian cattle (Olson et al. 2012), TS sizes are usually large ($N_{TS} > 1000$), leading to high ρ and a substantial increase in the selection gain and efficiency of breeding programs employing GEBVs (Meuwissen et al. 2016; Hickey et al. 2017).

In plant breeding for line and hybrid cultivars, TS sizes are generally small because production and phenotyping of

Communicated by Jiankang Wang.

Pedro C. Brauner and Dominik Müller have contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00122-019-03445-7>) contains supplementary material, which is available to authorized users.

✉ Albrecht E. Melchinger
melchinger@uni-hohenheim.de

¹ Institute of Plant Breeding, Seed Sciences and Population Genetics, University of Hohenheim, Fruwirthstraße 21, 70599 Stuttgart, Germany

the candidates are usually laborious and expensive (Riedelsheimer and Melchinger 2013; Akdemir and Isidro-Sánchez 2019). Moreover, different from the large panmictic populations handled in animal breeding that include thousands of genotypes in the TS, the germplasm is composed of numerous biparental families (BPFs) with small size, because recycling the best lines from the previous cycle(s) represents the predominant method for development of new lines (Mikel and Dudley 2006). Despite these less favorable conditions, GP promises a quantum leap in selection gain due to higher selection intensity and reduced breeding cycle length, provided ρ is sufficiently large compared with h^2 (Bernardo and Yu 2007; Heffner et al. 2009).

With multiple BPFs, there are several options for composing the TS to be used in GP for selecting genotypes within each BPF. In the simplest case, the genotypes in the TS and PS originate from the same BPF, i.e., they are full-sibs. This is expected to yield highest ρ for a given family size (Lehermeier et al. 2014; Crossa et al. 2017), because Mendelian sampling can be fully accounted for through co-segregation between quantitative trait loci (QTL) and genome-wide markers (Habier et al. 2007; Schopp et al. 2017b). Numerous experimental studies in various crops (e.g., Riedelsheimer et al. 2013; Lehermeier et al. 2014; Liu et al. 2015; Würschum et al. 2017) and simulations (Schopp et al. 2017b) found high ρ for GP within BPFs, albeit with considerable variation among BPFs. However, even in this most favorable case, the size of the TS must still be sufficiently large ($N_{TS} \geq 50$) to reach an adequate ρ (Marulanda et al. 2015; Schopp et al. 2017b), which poses severe restrictions for GP of multiple BPFs under a limited budget.

A second option is to include information from a different BPF for composing the TS, adopting the idea from animal breeding to borrow information from major breeds for GP of minor breeds (Iheshiulor et al. 2016). For a given family F targeted for GP, candidates of which represent the PS, the TS may comprise related or unrelated BPFs having one or no parent in common with F . As expected under weaker pedigree relationships (Schopp et al. 2017b), experimental studies demonstrated a substantial decrease in ρ from full-sib to half-sib and unrelated families for given N_{TS} including negative estimates in the latter case (Riedelsheimer et al. 2013). Thus, N_{TS} must be large when constructing the TS with a half-sib or an unrelated family, but ρ increases at a much smaller rate beyond $N_{TS} > 100$ and approaches much lower asymptotic values than for full-sib families (Lehermeier et al. 2014). Moreover, simulations revealed a large variation in ρ among BPFs used as TS, which increased from full-sib to half-sib and unrelated families (Schopp et al. 2017b).

A third alternative is merging several BPFs in a combined TS in order to increase N_{TS} beyond what is possible if only a single BPF is used. In an extensive study with testcrosses of maize, Lehermeier et al. (2014) obtained for several traits

the same ρ when the TS was composed of 10 half-sib families instead of the full-sib family, because N_{TS} was fivefold increased. However, Riedelsheimer et al. (2013) reported negative ρ values between pairs of unrelated BPFs and Schopp et al. (2017b) found in simulations a large variation in ρ when individual half-sib or unrelated families were used as TS. This raises the question whether uncritical pooling of information from all available BPFs into a joint TS might lead to lower ρ than using a selected subset of all available families.

To avoid possible negative effects of incorporating in the TS any BPF having poor ρ with the prediction set family F , identifying of such families by simple criteria would be beneficial. Schopp et al. (2017b) found that linkage phase similarity (LPS) has a relatively small effect on ρ compared to the proportion of shared segregating markers between TS and PS (θ). Further criteria that come into consideration are the genomic relationship between the genotypes in the TS and PS or forecasts of ρ described by Wientjes et al. (2015). The most promising BPFs identified by these criteria could be merged in the TS for approaching the optimum TS composition. Whether any of these criteria are useful for reaching this goal has not yet been investigated to the best of our knowledge.

Our main objective was to find for a given PS family F and a given set A of all families, including half-sib and unrelated BPFs as well as full-sibs of F , the best possible subset O^F of A that yields highest ρ . In particular, we investigated four questions: (1) Whether the BPFs are ranked in descending order according to their ρ with F , and how does sequentially adding the BPFs, from highest to lowest rank, to the TS affect ρ ? (2) How big is the difference in ρ for set O^F and set A ? (3) Since true values of ρ are needed for finding set O^F , which is not available in practice, can we find a subset E^F approaching ρ for set O^F based on empirical criteria between F and each BPF in A , such as estimates of ρ , the average simple matching coefficient (SM), the linkage phase similarity or the forecasted prediction accuracy ρ^W ? (4) What are the implications of our findings regarding the design of breeding programs applying GP? To answer these questions, we (a) used sets of BPFs simulated from real genomic data of elite germplasm as well as genetic resources and (b) re-analyzed data sets of two experiments from the literature.

Materials and methods

Genotypic data

Genotypic data obtained with the Illumina MaizeSNP50 BeadChip for two populations of maize were taken from two publications (Melchinger et al. 2017; Schrag et al. 2018) and served as starting point for our simulations, following

the study of Schopp et al. (2017b). As detailed by these authors, the ancestral population *Elite* consisted of 149 elite flint lines from the maize breeding program of the University of Hohenheim and displayed mainly long-range linkage disequilibrium (LD). The ancestral population *Landrace* consisted of 59 doubled haploid (DH) lines derived from the German maize landrace “Gelber Badischer” and displayed mainly short-range LD. The original set of 29,729 polymorphic SNPs was reduced to 10,769 SNPs after removing loci with a minor allele frequency less than 0.05 in either germplasm.

Simulation of families and traits

In each simulation run, a set of BPFs was simulated by selecting at random a number of parents (n_p) that were drawn from the respective ancestral population (either *Elite* or *Landrace*) and crossed in silico in a half-diallel mating design, yielding $n_p(n_p - 1)/2$ crosses (Supplementary Fig. 1). One of these crosses was randomly chosen as PS family F . From the $2(n_p - 2)$ crosses sharing one common parent with F , set H composed of n_H half-sib families was developed. Likewise, from the remaining crosses having no parent in common with F , set U composed of n_U unrelated families was developed. From each cross, n_g DH lines were generated in silico using the R-package Meiosis (Müller and Broman 2017), which simulated meioses in the F_1 generation and subsequently duplicated the genome of the resulting gametes. For family F , we had two disjoint subfamilies of DH lines: (1) subfamily F^{TS} comprising n_g DH lines that were included in the TS either alone or in combination with other families from H and/or U ; (2) subfamily F^{PS} comprising $n_g^{PS} = 1000$ DH lines that served as independent PS for precise estimation of ρ within F . Set $A = F^{TS} \cup H \cup U$ with $n_A = 1 + n_H + n_U$ families, each having n_g genotyped and phenotyped DH lines, represented the maximum TS available for GP of set F^{PS} and its composition is specified for each scenario described below.

Next, we randomly sampled in each simulation run two disjoint sets of n_{SNP} and n_{QTL} loci from the 10,769 SNPs, which served as markers and QTL, respectively. For each QTL j , an allele substitution effect α_j was simulated as $\alpha_j = (2\beta - 1)\gamma$. Here, γ was sampled from a Gamma distribution $\Gamma(1.66, 0.4)$ (Meuwissen et al. 2001) and β was drawn from a Bernoulli distribution with probability 0.5. We assumed an additive-genetic model underlying the allele substitution effects, which is expected to hold true in the absence of epistasis for per se performance of pure-breeding lines as well as for testcross performance of lines in hybrid breeding, as described by Melchinger et al. (1998).

The true breeding value (TBV) for each DH line was computed by summation of the allele substitution effects across the whole genome: $\sum_j q_j \alpha_j$, where $q_j \in \{0, 2\}$ is the genotypic score

at QTL j observed for the respective DH line and α_j the allele substitution effect. Phenotypes were simulated by adding a noise variable $e \sim N(0, \sigma_e^2)$ to the TBV. For achieving a desired heritability (h^2) across all BPFs, the variance σ_e^2 was calculated as numerical solution of the following equation:

$$h^2 = \frac{1}{n_A} \sum_{v \in A} \{h_v^2\} = \frac{1}{n_A} \sum_{v \in A} \left\{ \frac{\sigma_{g_v}^2}{\sigma_{g_v}^2 + \sigma_e^2} \right\}, \quad (1)$$

where $\sigma_{g_v}^2$ and h_v^2 refer to the variance of TBVs and heritability in family v , respectively.

This procedure was exercised in each of 200 simulation runs, from which we calculated the mean and standard deviation (SD) of ρ . The parameter values underlying our simulations are described in Supplementary Table 2. Default values were ancestral population *Elite*, $n_p = 8$, $n_A = 28$, $n_H = 12$, $n_U = 15$, $n_g = 50$, $n_{SNP} = 5000$, $n_{QTL} = 1000$ and $h^2 = 0.6$, which were subsequently modified depending on the scenarios described below.

Genomic prediction

We used the standard genomic best linear unbiased prediction model (Habier et al. 2007; VanRaden 2008):

$$y_{TS} = \mu + Z\alpha + \varepsilon \quad (2)$$

where y_{TS} is the vector of phenotypes of all DH lines in the TS, μ is the general intercept, α is the vector of breeding values, Z is an incidence matrix associating phenotypes with breeding values and ε is a vector of residuals. Standard assumptions were $\alpha \sim N(0, \mathbf{G}\sigma_\alpha^2)$ and $\varepsilon \sim N(0, \mathbf{I}\sigma_\varepsilon^2)$, where σ_α^2 is the additive-genetic variance and σ_ε^2 the residual error variance.

The genomic relationship matrix \mathbf{G} based on the n_{SNP} SNP markers was calculated by extending the approach of Schopp et al. (2017b) to multiple BPFs. Let $x_{v,i,l} \in \{0, 2\}$ be the genotypic score for the genotype at locus l of DH line i from family v . The genomic relationship $G_{i,j}$ between two DH lines i and j from families v and w , respectively, was then calculated as:

$$G_{i,j} = \frac{\sum_l (x_{v,i,l} - p_{v,l})(x_{w,j,l} - p_{w,l})}{\sqrt{2 \sum_l p_{v,l}(1 - p_{v,l})} \sqrt{2 \sum_l p_{w,l}(1 - p_{w,l})}} \quad (3)$$

where $p_{v,l}, p_{w,l} \in \{0, 0.5, 1\}$ refer to the allele frequencies at locus l in family v and w , respectively. For $v = w$, the formula simplifies to Method 1 of VanRaden (2008). Estimates of σ_α^2 and σ_ε^2 in the TS as well as GEBVs of the PS were calculated using the R-package rrBLUP (Endelman 2011) using the *mixed.solve* function. We computed for each PS family F and TS composed of families in set T the Pearson correlation between the TBV and GEBV for the $n_g^{PS} = 1000$ DH lines from F^{PS} and regarded these values as true prediction

accuracy $\rho(T, F^{PS})$ for GP of F^{PS} using TS T , given the large sample size of the PS.

Potential for optimizing the training set composition

Investigating the potential for optimizing the TS composition involved three steps. First, we calculated $\rho(A, F^{PS})$ as benchmark, because set A is often uncritically used as TS. Second, we searched among all possible subsets $T \subset A$ for the optimum TS composition, denoted as set O^F , which maximizes $\rho(T, F^{PS})$. In principle, this task can be solved by complete enumeration, but since computation time grows exponentially with n_A , this approach becomes excessive for large n_A . To approach O^F in such a situation, we resorted to an algorithm for numerical optimization called binary particle swarm optimization (BPSO). We implemented a modification of this algorithm described by Khanesar et al. (2007) in a publicly available R-package BPSO (Müller 2019). For $n_A = 10$, we compared $\rho(O^F, F^{PS})$ for O^F identified with this package and complete enumeration and found close agreement between both methods (data not shown).

Empirical optimization of the training set composition

Since the TBVs were known in our simulation study, we were able to calculate ρ for F^{PS} . In real data sets, however, TBVs are unknown and only phenotypic data of the lines in F^{TS} are available. As an empirical method for approaching O^F by a set E^F , we used the phenotypic values of DH lines available from F^{TS} to obtain a subset $T \subset A$ yielding the highest estimate of ρ . The procedure for identifying E^F was almost identical to the one described above for identifying O^F with two exceptions. First, we temporarily restricted our search to all possible subsets of $T \subset H \cup U$, using either complete enumeration or the BPSO algorithm. At this stage, subfamily F^{TS} was excluded from the search in A , because including full-sibs in the TS raises ρ to a high level and, thus, would reduce the differences in $\rho(T, F^{TS})$ between alternative TS compositions $T \subset A$. Second, we used phenotypic values of the lines in F^{TS} as PS. Strictly, this allows only estimation of the predictive ability r_a (i.e., the correlation of phenotypic values with the GEBVs in the DH lines of F^{TS}), which, divided by $\sqrt{h^2}$ of F^{TS} , yields an estimate of the prediction accuracy (Dekkers 2007). However, this step is unnecessary because it does not alter the ranking of different TS compositions. Set E^F was obtained by combining F^{TS} with the best TS composition identified in the previous step, because full-sibs contribute most information to GP (Schopp et al. 2017b). Finally, we computed $\rho(E^F, F^{PS})$.

Training set composition scenarios

We analyzed the influence of different TS compositions on ρ in three scenarios. In Scenario 1 detailed in Supplementary Fig. 2, we investigated the possibility that predictively poor families can reduce $\rho(T, F^{PS})$ when they are included in a combined TS $T \subset A$. We simulated a half-diallel with $n_p = 12$ parents and selected at random one of the BPFs produced from these crosses as family F , kept the $n_H = 20$ half-sib families of F as set H and sampled from the remaining families $n_U = 20$ for set U , but used default values for all other factors. We ordered the families H_i within set $H = \{H_i | i = 1, \dots, 20\}$ according to decreasing values of $\rho(H_i, F^{PS})$. Likewise, we ordered the families U_i within set $U = \{U_i | i = 1, \dots, 20\}$ according to decreasing values of $\rho(U_i, F^{PS})$. For both H and U , starting with family $i = 1$ we built combined TS of increasing size N_{TS} by sequentially including individual families in the order of their rank i , denoted as $\cup H_i$ and $\cup U_i$. The same procedure was applied using other criteria for ranking the BPFs: (1) the genome-wide linkage phase similarity (LPS) calculated after de Roos et al. (2008), (2) the forecasted pairwise prediction accuracy ρ^W obtained from deterministic equations described by Wientjes et al. (2015), (3) the average simple matching coefficient (SM) (Sneath and Sokal 1973) between the parents of BPF H_i or U_i with family F , and (4) the proportion of polymorphic markers θ shared between F and H_i or U_i (Schopp et al. 2017b).

In Scenario 2 detailed in Supplementary Fig. 3, we investigated different TS compositions. We compared $\rho(T, F^{PS})$ for TS compositions $T = F^{TS}, H, U, F^{TS} \cup H, F^{TS} \cup U, H \cup U, A, O^F$ and E^F . Besides the default values, we also investigated under *ceteris paribus* conditions how changes in $h^2, n_{SNP}, n_{QTL}, n_H$ or n_U affect ρ (Supplementary Table 2).

In Scenario 3 detailed in Supplementary Fig. 4, we investigated how different choices of n_H influence ρ for set $F^{TS} \cup H$, because preliminary results showed that this TS composition yields high and robust estimates of ρ . We used $n_H = 0, 2, 4, 12, 20$, and varied $n_g = 25, 50, 100$ as well as $h^2 = 0.3, 0.6, 0.9$. For all other factors, we used default values (Supplementary Table 2).

Experimental data

We used two data sets of multiple BPFs of maize available from the literature. The first was the set of $n_H = 10$ half-sib families from Lehermeier et al. (2014), in which 10 dent founder lines (B73, D06, D09, EC169, F252, F618, Mo17, UH250, UH304, and W117) were crossed to the central dent line F353. Population size per family ranged between 53 and 104 DH lines, which were evaluated as testcrosses using a tester from the opposite heterotic pool. We chose three traits to cover different genetic architecture and h^2 : dry matter

yield, plant height, and female flowering. A total of 34,116 polymorphic SNPs were available across all BPFs for GP.

The second data set was composed of 10 BPFs from tropical maize germplasm published by Zhang et al. (2015) with different pedigree relationships, including both half-sib and unrelated families. Population sizes ranged between 172 and 191 $F_{2,3}$ lines, which were evaluated tropical environments as testcrosses using a tester from the opposite heterotic pool. We used three traits: grain yield, male flowering and plant height recorded under well-watered conditions. The lines for each BPF were genotyped using genotyping by sequencing. After quality check, 3866 SNPs were available across all families.

In each data set, we performed a cross-validation scheme, where $n_g = 50$ genotypes were sampled from each BPF to construct F^{TS} , H and U . The remaining genotypes from the PS family not sampled for F^{TS} were used as prediction set F^{PS} . Prediction accuracy ρ was calculated as the predictive ability r_a divided by $\sqrt{h^2}$ of the trait estimated from all lines in F (Dekkers 2007).

Referring to Scenario 1, we investigated also with experimental data if predictively poor families can reduce the ρ for F^{PS} when they are sequentially included in a combined TS. Using each family once as PS, we computed ρ of F^{PS} with every other BPF and ranked these estimates accordingly. As described above, we built combined TS of increasing size N_{TS} by sequentially including individual BPFs in the order of their rank. This procedure was applied to the half-sib panel from Lehermeier et al. (2014) as well as the group of unrelated families from Zhang et al. (2015). In the latter study, using the numbering of BPFs employed by these authors,

only BPFs F9 and F10 served as PS family, because they had no parent in common with all other BPFs. Referring to Scenario 2, we investigated the optimization of the TS composition for BPFs F1 to F8 from Zhang et al. (2015). Using each BPF once as PS, we analyzed ρ using different TS compositions. For both scenarios, we calculated the mean and SD of ρ across 50 repetitions of the cross-validation scheme and every family treated once as PS.

Data availability statement

Simulations and data analysis were carried out within the R environment (R core 2019). All functions used in our simulations, including the data of both ancestral populations, can be found within the R-package “HOT”, available at: <https://gitlab.com/HOT>.

Results

Genomic prediction with pairs and sequential union of families in the training set (Scenario 1)

All subsequent simulation results refer to means over the 200 simulation runs. Using full-sib family F^{TS} as TS yielded $\rho(F^{TS}, F^{PS}) = 0.71$, whereas the pairwise prediction accuracies $\rho(H_i, F^{PS})$ for the half-sib families $H_i \in H$ and $\rho(U_i, F^{PS})$ for the unrelated families $U_i \in U$ averaged 0.37 and 0.20, respectively (Fig. 1). Estimates of $\rho(H_i, F^{PS})$ varied from 0.68 to -0.06 and those of $\rho(U_i, F^{PS})$ from 0.54 to -0.19 . Combining the two half-sib families ($i = 1, 2$)

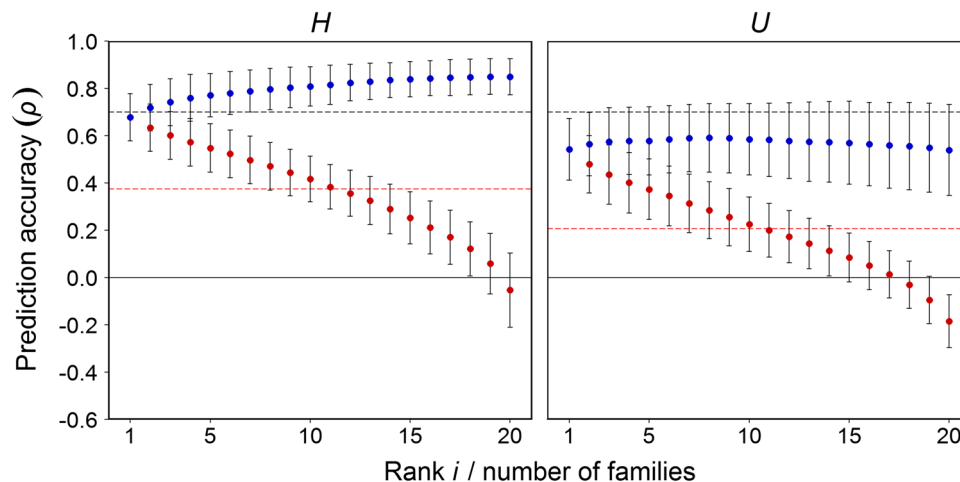


Fig. 1 Prediction accuracy ($\rho \pm SD$) for genomic prediction of family F^{PS} with single half-sib families ($H_i \in H$) and unrelated families ($U_i \in U$) shown in red and composite training sets (TS) shown in blue. Families were ranked in descending order according to the magnitude of ρ with F^{PS} , with ranks i shown on the x -axis, and sequentially added to the composite TS. Black horizontal dashed line: mean

prediction accuracy obtained for the full-sib family F^{TS} ; red horizontal dashed line: mean pairwise prediction accuracies for all $H_i \in H$ and $U_i \in U$. Results were averaged across 200 simulation runs using default values: ancestral population *Elite*, $n_g = 50$, $n_{SNP} = 5000$, $n_{QTL} = 1000$, $h^2 = 0.6$

with the highest pairwise ρ in the TS with $N_{TS}=100$ yielded $\rho(\cup H_i, F^{PS})=0.72$. Adding further half-sib families with rank $i=3$ to 20 sequentially to the TS yielded a steady increase in $\rho(H_i, F^{PS})$ up to 0.84, after which the beneficial effect on ρ flattened off.

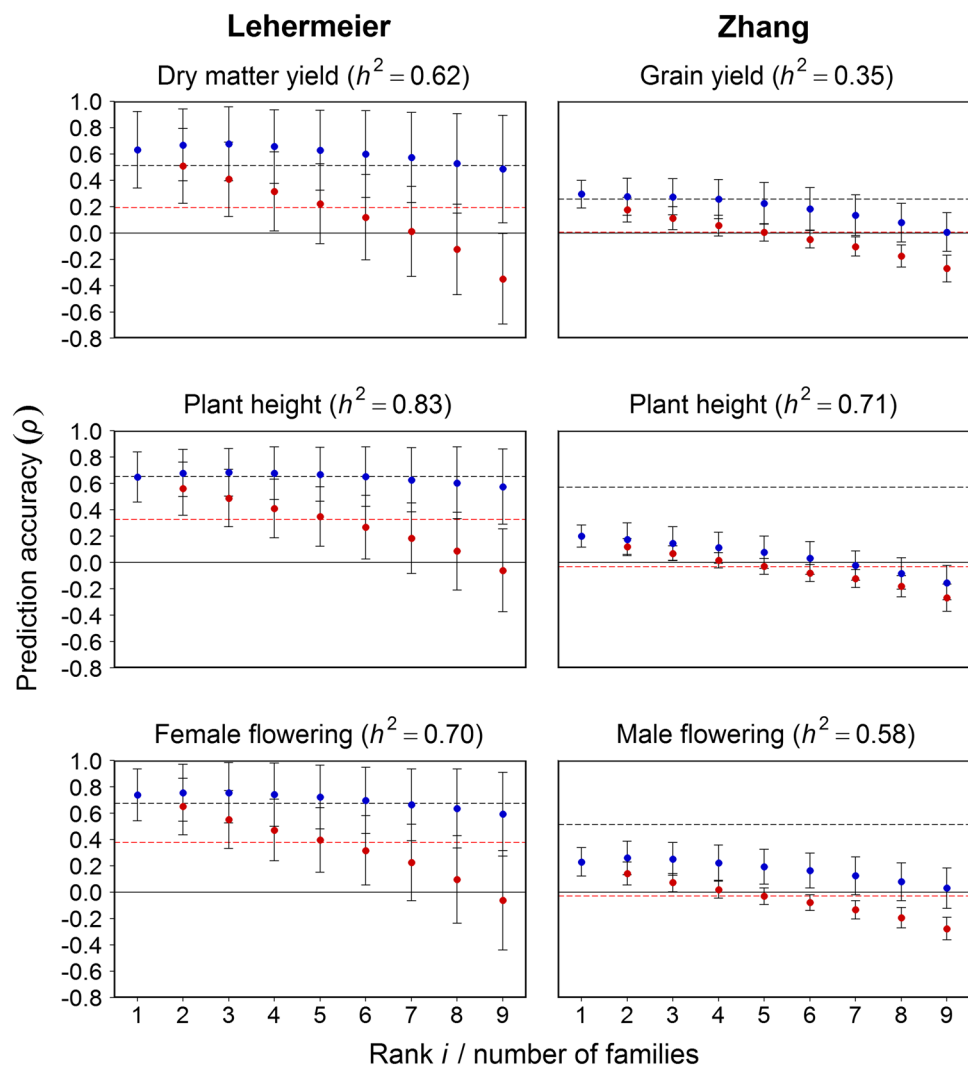
Combining unrelated families from set U in the order of their rank resulted in a concave curve (Fig. 1). It reached a maximum of $\rho(\cup U_i, F^{PS})=0.59$ with eight families in the TS, but decreased afterward to 0.53, even slightly below $\rho(U_1, F^{PS})=0.54$ for the best unrelated family. While SD of ρ for the combined TS decreased as the number of sequentially added families from set H increased, SD increased when merging families with lower rank from set U .

For the panel of 10 half-sib families from Lehermeier et al. (2014), ρ for F^{TS} averaged 0.52 for dry matter yield, 0.65 for plant height and 0.67 for female flowering, whereas $\rho(H_i, F^{PS})$ averaged 0.19 for dry matter yield, 0.33 for plant height and 0.38 for female flowering (Fig. 2). Estimates of $\rho(H_i, F^{PS})$ ranged from 0.62 to -0.34 for dry matter yield,

0.65 to -0.10 for plant height, and 0.72 to -0.05 for female flowering for rank $i=1$ to 9. Sequentially adding families to the TS improved $\rho(\cup H_i, F^{PS})$ for dry matter yield (0.68), plant height (0.69) and female flowering (0.75) up to rank 3, yielding values higher than predictions with F^{TS} . However, including further families from rank 4 to 9 to the TS resulted in $\rho(\cup H_i, F^{PS})$ lower than for $\rho(H_1, F^{PS})$. In contrast to the simulations, the SD for ρ of the combined TS became larger with an increasing number of half-sib families.

For the set of 10 unrelated families taken from Zhang et al. (2015), ρ for F^{TS} as TS averaged 0.26 for grain yield, 0.57 for plant height and 0.51 for male flowering (Fig. 2). Estimates of $\rho(U_i, F^{PS})$ varied from 0.29 to -0.27 for grain yield, from 0.20 to -0.27 for plant height, and from 0.23 to -0.28 for male flowering for rank $i=1$ and 9, respectively, with a mean close to zero for all traits. In contrast to the simulations, $\rho(\cup U_i, F^{PS})$ immediately decreased after including the second ranked family. Only for male flowering did $\rho(\cup U_i, F^{PS})$ increase

Fig. 2 Prediction accuracy ($\rho \pm$ SD) for genomic prediction of family F^{PS} with single half-sib families ($H_i \in H$) and unrelated families ($U_i \in U$) shown in red and composite training sets (TS) shown in blue for experimental data taken from Lehermeier et al. (2014) and Zhang et al. (2015) for test-crosses of maize lines. Families were ranked in descending order according to the magnitude of ρ with F^{PS} , with ranks i shown on the x -axis, and sequentially added to the composite TS. Results were averaged over 10 families for Lehermeier et al. (2014) and across two families (populations F9 and F10) unrelated with all other populations for Zhang et al. (2015). The black horizontal dashed line and the red horizontal dashed line have the same meaning as in Fig. 1

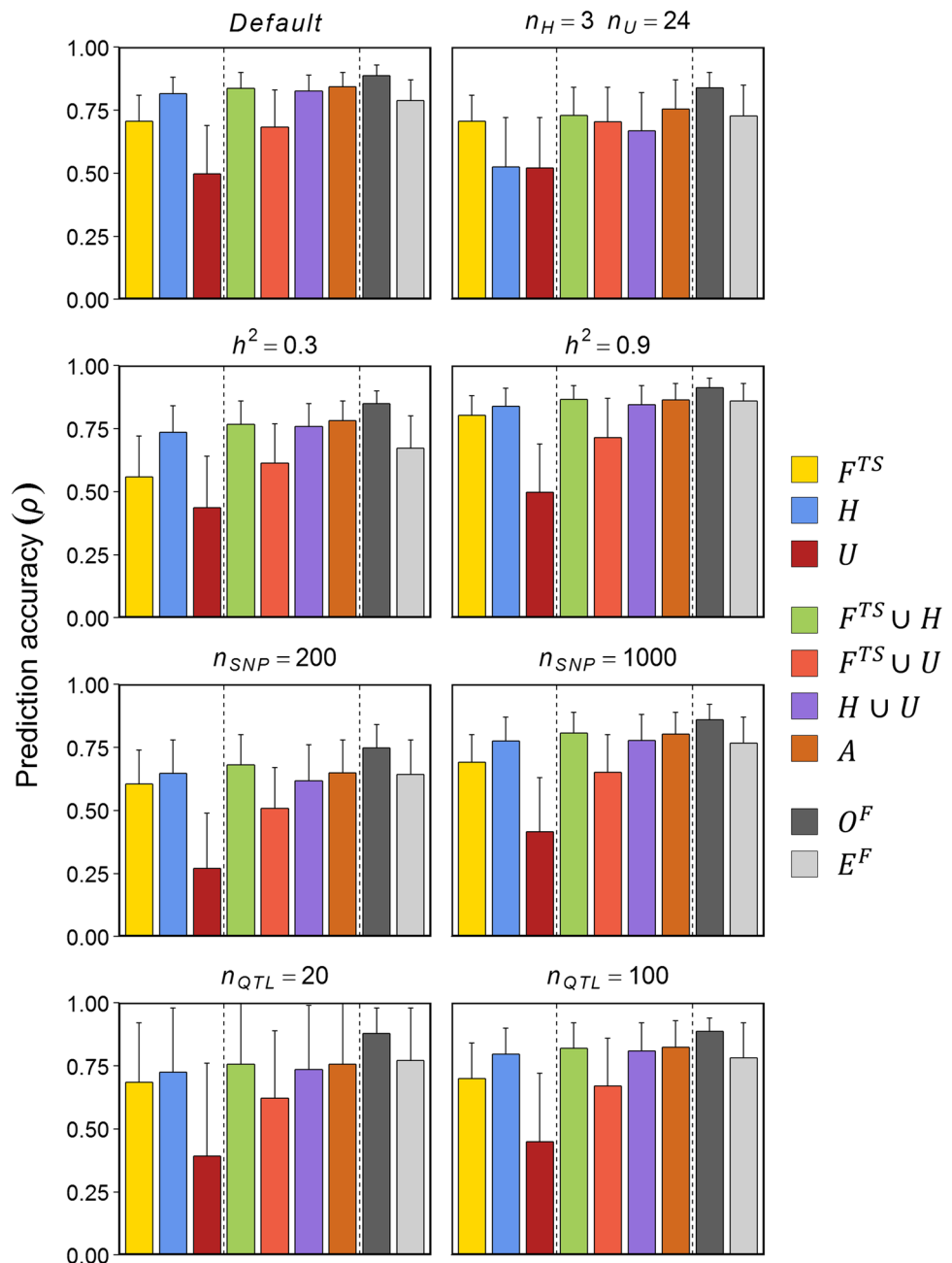


initially when combining the families with rank $i = 1, 2$. However, sequentially adding further families into the TS up to rank $i = 4$ hardly altered the prediction accuracy for grain yield and plant height. Adding further families had actually a negative effect so that when all nine unrelated families were combined, $\rho(UU_i, F^{PS})$ was either zero for grain yield and male flowering or even negative for plant height. For all traits, the SD of the combined TS increased with a larger number of families.

Genomic prediction with different compositions of the training set (Scenario 2)

Simulations with default values averaged $\rho = 0.71, 0.81$ and 0.50 for F^{TS}, H and U , respectively (Fig. 3). Using all families in sets $F^{TS} \cup H$ or $H \cup U$ resulted in a higher ρ of ~ 0.83 , whereas $F^{TS} \cup U$ yielded a lower ρ value than F^{TS} alone. Combining all available families $A = F^{TS} \cup H \cup U$ yielded $\rho = 0.84$. In comparison with $\rho(A, F^{PS})$, ρ was 0.05 higher for O^F and 0.05 lower for E^F . SD was highest for set U and $F^{TS} \cup U$ and lowest for set O^F , with all other TS compositions showing intermediate SD values.

Fig. 3 Prediction accuracy (ρ +SD) in the prediction set family F^{PS} by genomic prediction with training sets (TS) composed of the full-sib family (F^{TS}), half-sib families ($H_i \in H$) and unrelated families ($U_i \in U$) or complete and partial combinations of them as well as the optimum TS composition (O^F) and empirically optimal TS composition (E^F). Results were averaged across 200 simulations using default values (ancestral population *Elite*, $n_p = 8, n_A = 28, n_H = 12, n_U = 15, n_g = 50, n_{SNP} = 5000, n_{QTL} = 1000$ and $h^2 = 0.6$) or modifications in one of these parameters mentioned above each graph



For the set of BPFs from Zhang et al. (2015), the mean of ρ across all traits was 0.42, 0.34 and 0.07 when the TS comprised F^{TS} or all families in H or U , respectively (Table 1). Constructing the TS with $F^{TS} \cup H$ yielded highest ρ for plant height (0.52) and male flowering (0.46) and the second highest for grain yield (0.49). Combinations $F^{TS} \cup U$ and $H \cup U$ both had ρ lower than F^{TS} and H alone for plant height and male flowering, whereas for grain yield the values were higher. Combining all families (set A) yielded the highest ρ for grain yield (0.51) and third highest for plant height (0.46) and male flowering (0.41). Set E^F had higher ρ than A for plant height and male flowering, but for grain yield, the reverse was true.

Factors influencing the prediction accuracy for different compositions of the training set

Decreasing the number of half-sib families from default $n_H = 12$ to 3 and increasing the number of unrelated families from default $n_U = 15$ to 24 reduced ρ for all TS compositions including H by at least 19%, with the largest reduction from 0.81 to 0.50 for set H (Fig. 3). However, it hardly changed ρ for set U and combinations with U , whereas the difference between set A and O^F increased from 0.05 to 0.08. Increasing h^2 , n_{QTL} or n_{SNP} improved ρ uniformly for all TS compositions but hardly changed their relative differences. One exception was $F^{TS} \cup H$, which was after O^F the best performing TS composition under low n_{SNP} . Another

Table 1 Prediction accuracy (ρ) in the prediction set (PS) family F^{PS} for genomic prediction from composite training sets (TS) using different families with $n_g = 50$ as well as different number of half-sib and unrelated families (n_H and n_U , respectively)

$F^{PS\dagger}$	n_H	n_U	Prediction accuracy (ρ)							
			F^{TS}	H	U	$F^{TS} \cup H$	$F^{TS} \cup U$	$H \cup U$	A	E^F
Grain yield										
F1	4	5	0.42	0.51	0.20	0.57	0.44	0.54	0.58	0.49
F3	4	5	0.12	0.37	0.22	0.35	0.21	0.40	0.37	0.25
F2	3	6	0.60	0.26	0.00	0.66	0.60	0.30	0.68	0.60
F4	3	6	0.15	0.18	0.20	0.24	0.28	0.25	0.29	0.26
F5	2	7	0.51	0.39	0.10	0.54	0.52	0.41	0.57	0.54
F8	2	7	0.55	0.34	0.04	0.55	0.57	0.33	0.60	0.58
F6	1	8	0.39	0.31	0.15	0.47	0.39	0.33	0.44	0.39
F7	1	8	0.37	0.27	0.12	0.41	0.40	0.26	0.43	0.41
		Mean	0.40	0.34	0.14	0.49	0.44	0.35	0.51	0.45
Plant height										
F1	4	5	0.55	0.60	0.03	0.65	0.45	0.52	0.58	0.60
F3	4	5	0.53	0.16	-0.04	0.48	0.41	0.12	0.39	0.47
F2	3	6	0.53	0.37	0.22	0.61	0.49	0.41	0.57	0.59
F4	3	6	0.42	0.27	0.01	0.50	0.31	0.17	0.37	0.41
F5	2	7	0.52	0.42	0.24	0.60	0.52	0.47	0.59	0.58
F8	2	7	0.56	0.44	-0.08	0.61	0.36	0.33	0.45	0.51
F6	1	8	0.36	0.33	-0.06	0.47	0.29	0.23	0.37	0.42
F7	1	8	0.10	0.10	0.11	0.16	0.28	0.13	0.31	0.29
		Mean	0.46	0.35	0.05	0.52	0.40	0.30	0.46	0.49
Male flowering										
F1	4	5	0.21	0.32	-0.03	0.34	0.19	0.29	0.32	0.26
F3	4	5	0.53	0.40	-0.33	0.57	0.32	0.15	0.37	0.53
F2	3	6	0.45	0.19	0.09	0.45	0.39	0.24	0.43	0.43
F4	3	6	0.37	0.38	-0.14	0.47	0.29	0.24	0.40	0.40
F5	2	7	0.21	0.19	0.19	0.31	0.29	0.28	0.34	0.28
F8	2	7	0.50	0.46	0.02	0.58	0.41	0.39	0.48	0.51
F6	1	8	0.51	0.24	0.25	0.52	0.51	0.14	0.50	0.51
F7	1	8	0.45	0.36	0.16	0.51	0.43	0.27	0.48	0.49
		Mean	0.39	0.32	0.03	0.46	0.34	0.25	0.41	0.42

Results refer to experimental data of testcrosses of maize lines from Zhang et al. (2015) recorded for grain yield (GY), plant height (PH) and male flowering (MF). The largest ρ value within each row is shown in bold

[†]Numbering of prediction set family F^{PS} is identical with the number used by Zhang et al. (2015)

exception was E^F , which outperformed all other TS compositions except O^F for low n_{QTL} and showed similar ρ as A for high h^2 . SD of ρ substantially increased with lower n_{QTL} but increased only slightly with higher n_U (and smaller n_H at the same time), lower h^2 and smaller n_{SNP} . When sampling the parents from ancestral population *Landrace*, the observed ρ were slightly lower than for ancestral population *Elite*, with the strongest reductions observed for U and combinations with U (Supplementary Fig. 5), but the relative differences in ρ were hardly affected.

When n_H increased from 0 to 20, $\rho(F^{TS} \cup H, F^{PS})$ increased from 0.43 to 0.72 for $n_g = 25$ and $h^2 = 0.3$, but only from 0.88 to 0.93 for $n_g = 100$ and $h^2 = 0.9$ (Fig. 4). Thus, the gain in ρ obtained from adding half-sib families

was smaller under large n_g and high h^2 . At intermediate values of n_g and h^2 , the differences between ρ for $n_H = 0, 2$ and 4 were at most 0.05.

Ranking families for their predictive value using other criteria

Calculating SM, θ , LPS and ρ^W between each BPF and F yielded for both half-sib and unrelated families substantial differences between the top and lowest ranking family for each of the four criteria (Fig. 5). The largest range was found for θ in set H (from 0.77 to 0.23) and LPS in set U (from 0.82 to 0.19) and the smallest for SM in set U . The top ranking families identified with θ and LPS had $\rho(H_1, F^{PS})$ and $\rho(U_1, F^{PS})$

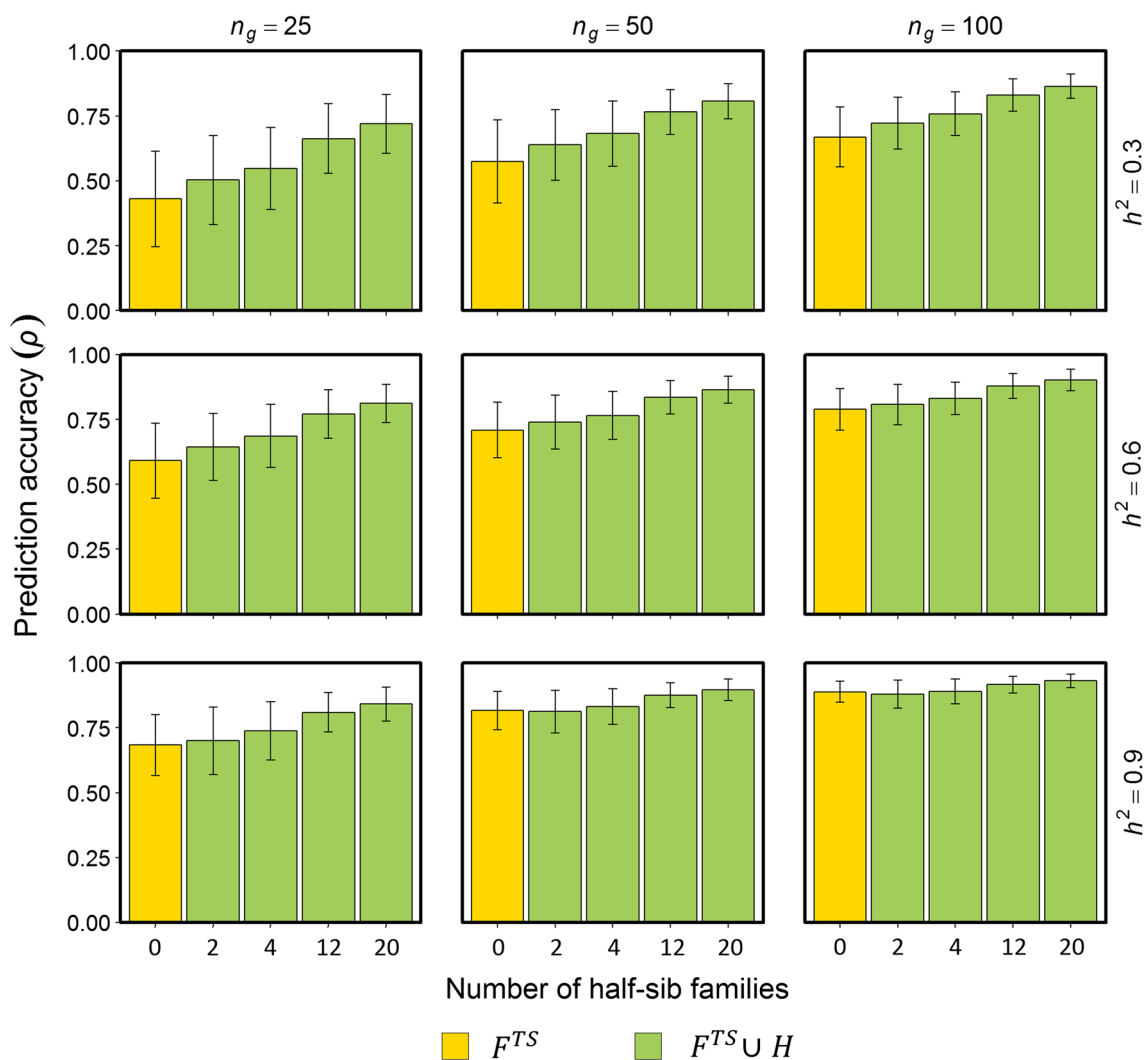
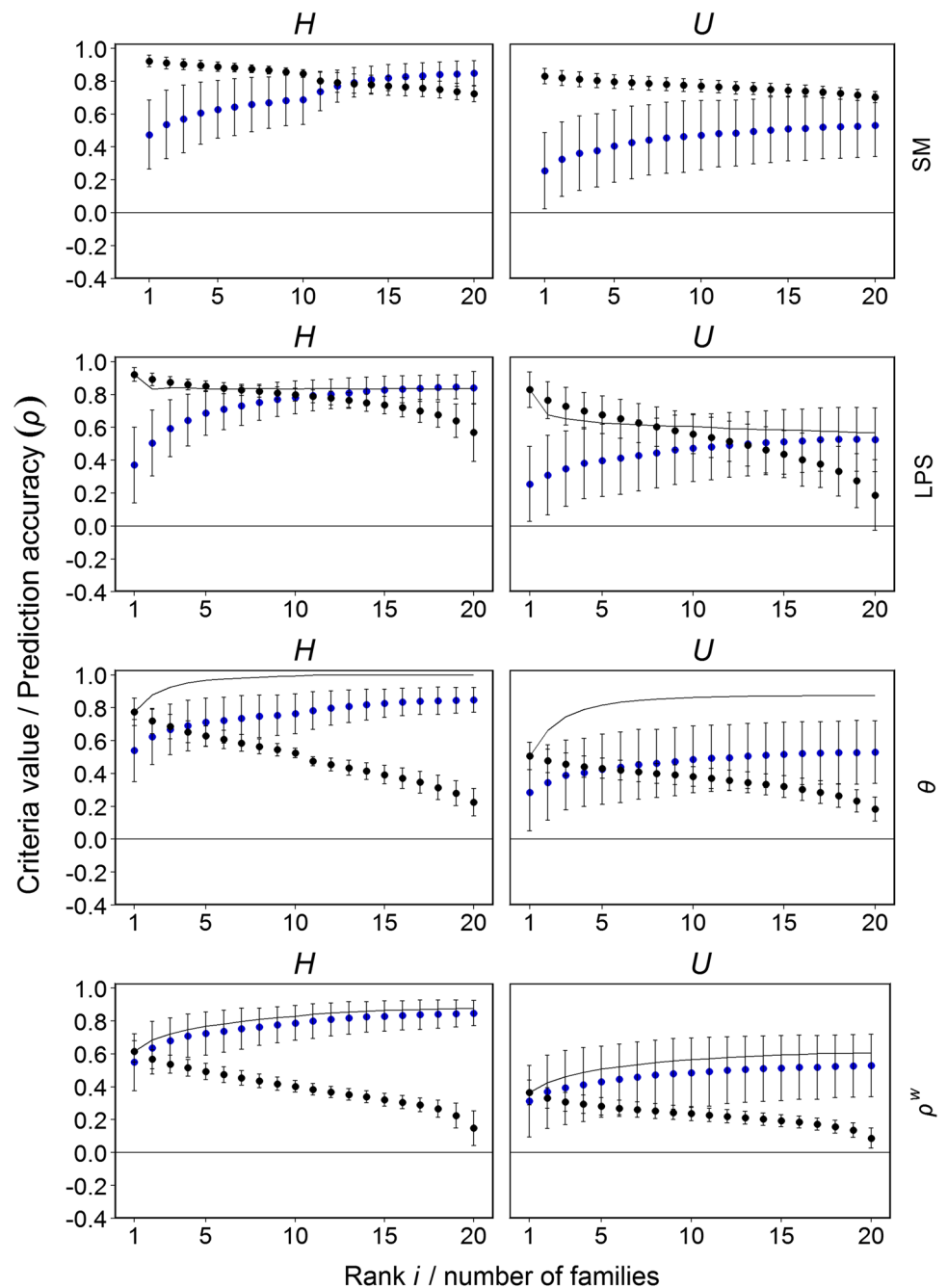


Fig. 4 Prediction accuracy ($\rho \pm SD$) in the prediction set family F^{PS} by genomic prediction with the full-sib family (F^{TS}) and combinations of F^{TS} with half-sib families from H in the training set $F^{TS} \cup H$. Set H had $n_H = 0, 2, 4, 12, 20$ families, with both parents of F being

equally represented. Results were averaged across 200 simulation runs using default values (ancestral population *Elite*, $n_{SNP} = 5000$, $n_{QTL} = 1000$) as well as three different sizes of n_g and h^2 mentioned above and to the side of graphs, respectively

Fig. 5 Prediction accuracy ($\rho \pm \text{SD}$) for composite training sets (TS) shown in blue and estimates of simple matching coefficient (SM), linkage phase similarity (LPS), θ and $\rho^W(x \pm \text{SD})$ shown in black between single half-sib families ($H_i \in H$) or unrelated families ($U_i \in U$) with family F . Estimates of LPS, θ and ρ^W for the composite TS are shown in the black line. Families were ranked in descending order according to the magnitude of the criteria (SM, LPS, θ , ρ^W) with F , with ranks i shown on the x -axis, and sequentially added to the composite TS. Results were averaged across 200 simulation runs using default values (ancestral population *Elite*, $n_g = 50$, $n_{\text{SNP}} = 5000$, $n_{\text{QTL}} = 1000$, $h^2 = 0.6$)



values that were on average 34 and 51% lower, respectively, than when the ranking was based on the actual pairwise ρ values shown in Fig. 1. Sequentially combining families in the TS ranked by these criteria yielded monotonic increasing concave curves for $\rho(\cup H_i, F^{\text{PS}})$ and $\rho(\cup U_i, F^{\text{PS}})$ with biggest increases observed for SM and LPS.

Discussion

Previous studies evaluated methods for constructing an optimum TS prior to recording phenotypic data in field trials using two optimization criteria based on genomic data

alone: the coefficient of determination (Laloë 1993) or the prediction error variance (Rincent et al. 2012). These criteria can be used to select a subset of lines within a population that improves ρ compared to a random sample of genotypes (Akdemir et al. 2015; Bustos-Korts et al. 2016; Akdemir and Isidro-Sanchez 2019). In contrast, we evaluated how both genomic and preexisting phenotypic information from genotypes of multiple BPFs can be used to identify for each BPF the best TS composition in GP. This reflects the situation often faced in breeding programs, where numerous BPFs with different degree of relatedness to the PS are produced and at least partly tested in field trials (Albrecht et al. 2011; Lian et al. 2014; Zhang et al. 2015).

Comparison of simulation and experimental results

We used both simulated and experimental data to analyze GP with multiple BPFs under Scenario 1 and 2. Simulations can be rapidly conducted and, due to their flexibility, allow new research questions to be readily addressed. Thus, important factors influencing the prediction accuracy can be investigated and new insights be gained for how to improve the development of genetic materials to best exploit the benefits of GP. Addressing such questions with experimental data is not feasible in practice due to excessive expenditures for large experiments that would be required. However, the conclusions drawn from simulations strongly depend on how well the underlying models reflect reality. For this reason, we complemented our simulations with a re-analysis of experimental data from two studies in the literature (Lehermeier et al. 2014; Zhang et al. 2015) to validate the simulations.

In general, prediction accuracies of the TS compositions investigated in Scenario 1 were higher in simulations than in experiments (Figs. 1, 2). Most strikingly, the decline in $\rho(UU_i, F^{PS})$, when higher ranking unrelated families were sequentially added to the TS, occurred already for $i \geq 3$ with experimental data and was much more severe than in the simulations. These discrepancies are partly attributable to differences in the parental inbreds used for producing the BPFs. The lines of the ancestral populations used for generating the BPFs in silico were genetically not very diverse because they originated from the same breeding program or a single landrace. In contrast, the dent founder lines used for producing the half-sib families in the study of Lehermeier et al. (2014) were genetically very distant, because they originated from different maize breeding programs for temperate maize in Europe and the US corn belt. Thus, values for θ and LPS were lower, which contributes to a reduction in ρ . Similar arguments apply to the study of Zhang et al. (2015), in which BPFs were derived from crosses among tropical elite lines from CIMMYT and evaluated in tropical environments.

The SNP data for our simulations were taken from a public breeding program to reflect allele frequencies and LD decay encountered in practice, following Daetwyler et al. (2013). However, SNPs with marker allele frequency < 0.05 were removed during the quality check of our data so that rare QTL alleles were not accounted for, which may be important in reality. Furthermore, simulations of the trait architecture assumed a purely additive-genetic model and borrowed distributions of QTL effects from the literature (Meuwissen et al. 2001). Ignoring epistasis is expected to reduce prediction accuracy (Jiang and Reif 2015; Martini et al. 2017) and may, together with the above-mentioned reasons, explain the differences between simulation and experimental results (Figs. 1, 2).

We considered parents from ancestral population *Elite* as unrelated although latent genetic relationships cannot be excluded. Hence, the prediction accuracies of the simulated half-sib and unrelated families might be biased upward compared to the experimental results because GP benefits from closer relationships between the TS and PS (Clark et al. 2012; Schopp et al. 2017b). However, this bias was most likely small because prediction accuracies obtained for ancestral population *Landrace*, which included only unrelated lines (Brauner et al. 2018), were only slightly lower than for ancestral population *Elite* (Fig. 1 and Supplementary Fig. 6).

Factors influencing the prediction accuracy under different training set compositions

Factors influencing ρ were evaluated under *ceteris paribus* conditions because their main effects are of primary interest and interactions among them are presumably small. In general, modulating the factor levels did not change the ranking of ρ for different TS compositions. Reducing the number of half-sib families from $n_H = 12$ to 3 in favor of an increase in the number of unrelated families resulted for the simulations in a more than 10% reduction of ρ for all TS compositions except U , O^F , and E^F (Fig. 3). This finding was confirmed by the experimental data (Table 1), where the families with $n_H = 4$ or 3 had generally higher ρ values than those with $n_H = 2$ or 1 for plant height and male flowering. This suggests that in the design of breeding programs, clear preference should be given to half-sib families over unrelated families, which might be totally excluded for compiling TS, given the potential negative effects on ρ .

Increasing h^2 had the most favorable effect on ρ for F^{TS} and E^F , which benefited more from higher h^2 values than set O^F (Fig. 3). As candidates are phenotyped with higher precision, predictive abilities $r_a(T, F^{TS})$ become more precise for every $T \subset A$. Therefore, identification of the best TS composition E^F gets more reliable.

Increasing n_{SNP} was mainly beneficial, when the TS included DH lines from set U either alone or in combination with other sets (Fig. 3). This is in harmony with results for GP in synthetics, where ρ is primarily driven by linkage between QTL and markers, which requires a high marker density and large N_{TS} (Schopp et al. 2017a). However, an increase in marker density beyond one marker per cM hardly pays off for increasing ρ . Reducing n_{QTL} from 1000 to 100 or 20 slightly reduced the differences in ρ among TS compositions except for O^F and E^F , which outperformed the other TS compositions. Thus, the genetic architecture of the trait is apparently of secondary importance for the strategy to find the best TS composition.

Causal analysis of prediction accuracy under different training set compositions

In addition to h^2 and N_{TS} , ρ between BPFs is associated with the LPS between the PS and TS (Fig. 5) (Riedelsheimer et al. 2013; Lehermeier et al. 2014). Moreover, Schopp et al. (2017b) found the parameter θ , which reflects the proportion of polymorphic QTL in the PS that are also polymorphic in the TS, to be of equal or even greater importance than LPS. This was our rationale for exploring these measures as well as SM and ρ^W for identifying families having a negative effect on ρ when merging multiple BPFs in the TS. However, none of these criteria could be effectively used to assemble an optimum TS, as was evident from the lower ρ for the BPF in rank one identified with the criteria (Fig. 5) compared to the pairwise ρ (Fig. 1).

Obviously, θ limits the proportion of the genetic variance in the PS, which can be explained by the QTL in the TS, and is more variable among half-sib families than among unrelated families (Fig. 5). In contrast, LPS is more variable among unrelated BPFs, because sampling the two parents of a BPF from an ancestral population generates extensive new “sample” LD that can differ substantially between BPFs (Schopp et al. 2017a). Different linkage phases among loci (both QTL and markers) in two BPFs can cause different substitution effects of short chromosome segments, which reduce the “local” prediction accuracy for the segment and ultimately reduce ρ (Schopp et al. 2017b). Regarding the effect of LPS on ρ for a combined TS, we speculate that chromosome segment substitution effects estimated from the latter correspond closely to the average of these effects across all BPFs included in the TS. This can explain the observed reduction in ρ of set $U \cup U_i$, when higher rank families U_i were included in the TS, despite an increase in N_{TS} (Figs. 1, 2). Moreover, we hypothesize that the superior ρ of set O^F is due to a specific combination of BPFs, which optimally balances N_{TS} , θ and LPS for the chromosome segments relevant for the trait and family F , because set O^F

often comprised approximately the same number of half-sib and unrelated families (Supplementary Fig. 7).

To investigate the effect of N_{TS} on the prediction accuracy independent of θ , we determined ρ for individual families H_i or U_i with rank $i = 1, 10, 20$ under increasing values of n_g (Supplementary Fig. 8). Besides drastic differences in the level of ρ among these families, increasing n_g markedly improved $\rho(H_1, F^{PS})$ and $\rho(U_1, F^{PS})$, yet the improvement leveled off for $n_g \geq 200$. Conversely, $\rho(H_{20}, F^{PS})$ and $\rho(U_{20}, F^{PS})$ hardly changed for larger n_g . This agrees with the strong increase in both ρ and θ for $i < 4$ during sequential addition of half-sib families (Fig. 5). However, sequentially including BPFs by order of their value of θ with family F soon approaches a plateau for $i \geq 4$ close to 1.00 for set H and 0.87 for set U . Hence, combining more than four half-sib or unrelated families hardly increases θ and ρ .

Strategies for improving the training set composition

The optimum TS composition O^F had ρ values about 5–10% higher than set A and other combinations of F^{TS} , H and U (Fig. 3). In reality, TBVs required for determining set O^F are unknown and finding set E^F with the goal to approach O^F was not successful. Even with $h^2 = 1.0$, ρ of E^F is considerably lower than for O^F because the predictive ability $r_a(T, F^{TS})$ of each subset $T \subset A$ employed for identifying E^F is associated with considerable uncertainty due to the limited size n_g of subfamily F^{TS} used for calculating the correlation r_a (data not shown). Thus, the large standard error of the r_a values used for identifying the best TS composition is likely the reason for the gap in ρ between set O^F and E^F . Increasing n_g for F^{TS} is an obvious solution for reducing this gap, but this remedy has limitations regarding the optimum allocation of resources discussed below. Moreover, for large values of n_g , the prediction accuracy achieved with subfamily F^{TS} is already very high so that including half-sib families, not to mention unrelated families, is of little value.

Other methods envisioned for determining the optimal training set composition by sequential inclusion of BPFs in the TS may employ various criteria such as SM, LPS, θ or forecasts of ρ with ρ^W . In previous studies, these criteria were associated with ρ for pairs of BPFs (Lehermeier et al. 2014; Schopp et al. 2017b). However, for multiple BPFs we were unable to approach O^F empirically by using these criteria (data not shown). One reason why genomic information is insufficient for finding O^F is that ρ between pairs of BPFs varies considerably among traits (Schopp et al. 2017b), which is not accounted for by any of these criteria. For a given family F , set O^F may vary between traits due to differences in θ and LPS for the underlying QTL. Use of parental information from previous breeding cycles might

help to improve the forecasting ability of these criteria if the genomic information for haplotypes, and their substitution effects are integrated in order to restrict these measures to specific chromosome segments influencing the trait of interest. However, this warrants further research beyond the scope of this study.

Although pooling all available BPFs did not have a detrimental effect on ρ in the simulations irrespective of the ancestral population (Fig. 3 and Supplementary Fig. 5), this result did not hold true for the experimental data. With the data of Zhang et al. (2015), combining F^{TS} with unrelated families U in the TS had mainly negative effects on ρ , although for grain yield there was a slight advantage (0.02) of set A over $F^{TS} \cup H$ (Table 1). Moreover, inclusion of unrelated families generally increased SD of ρ , reflecting that GP becomes more risky to apply. Thus, we recommend omitting unrelated families from the TS in agreement with the concerns expressed by Riedelsheimer et al. (2013).

Design of breeding programs using GP with multiple biparental families

So far, we focused on the composition of the TS for maximizing the prediction accuracy of a single PS family. In practice, breeders work with multiple BPFs and want to apply GP simultaneously to each of them. Consequently, an important question is how to design A , the set of all BPFs produced in a breeding cycle, for making best use of GP. Adopting the above recommendation to exclude unrelated families from GP, we limited ourselves to consider only full-sib and half-sib families in the TS for every family F in A . Moreover, we assumed a fixed budget for producing, genotyping and phenotyping altogether N_{Tot} DH lines in the entire breeding program, ignoring the problem of balancing the resources spent for the TS and PS addressed by Riedelsheimer and Melchinger (2013) for a single BPF. As main criterion for comparison, we used ρ for each family, because this determines the selection gain achieved by GP. Long-term consequences for the genetic variation in the breeding program, which can erode by uncritical implementation of GP (Jannink 2010), might be addressed by considering the effective population size (Falconer and Mackay 1996) of all BPFs in set A as further criterion.

The first decision to be made by the breeder concerns the number of all families n_A versus the size n_g of each family under the side condition that $N_{Tot} = n_A \times n_g$. The second decision concerns the mating design employed to produce the BPFs for set A , which determines n_H , the number of half-sib families available for inclusion in the TS of each BPF. One extreme case is that the BPFs in A are produced by a half-diallel mating design from n_p parents so that $n_A = n_p(n_p - 1)/2$ and $n_H = 2(n_p - 2)$. Another extreme case is that set A consists only of unrelated crosses between

n_p parents so that $n_A = n_p/2$, $n_H = 0$ and merely the n_g full-sibs from each family serve as TS. In between these two extremes is the round-robin design, where n_p parents are used to produce $n_A = n_H n_p/2$ BPFs, each having n_H half-sib families to be determined by the breeder. These formulas show that for given n_A , the number of parents n_p used for producing the BPFs in A , which is directly proportional to the effective population size, is much smaller for the half-diallel mating design than for the unrelated crosses, and in between for the Round-Robin design.

With unrelated BPFs and $n_g = 50$ or 100 full-sibs for GP of each family, a sufficient level of ρ is reached unless the heritability is extremely low (Fig. 3). However, the improvement in ρ observed for traits with medium and high h^2 , when n_g is doubled from 50 to 100, hardly justifies halving n_A and n_p . Thus, composing A only of unrelated crosses is appealing for warranting a long-term selection progress in GP. Conversely, producing A with a half-diallel mating design has a high risk of narrowing the genetic diversity without substantial improvement in the prediction accuracy, as is evident from the comparison of ρ values for $n_H = 12$ or 20 with $n_H = 0$, unless h^2 is low and n_g is small (Fig. 4). However, $n_H = 12$ or 20 correspond to half-diallels produced from only $n_p = 8$ and 12 parents requiring $n_A = 28$ and 66 BPFs, respectively, which limits n_g so that a positive net effect on ρ is rather questionable. Using a round-robin design with $n_H = 2$ or 4 and $n_g = 50$ seems to be a good compromise. It slightly increases ρ and reduces its SD compared to unrelated BPFs especially under those conditions, where full-sibs alone yield only medium ρ values. Moreover, if seed production or phenotyping of testcrosses of some BPFs fail, so that they cannot be included in the TS, their half-sib families can still serve as backup. However, compared to composing A from unrelated crosses, the number of parents n_p in a Round-Robin design is inversely related to n_H .

In this study, we investigated the design of A for a single breeding cycle. In practice, GP is applied in subsequent breeding cycles, which are closely interconnected because the top lines selected in one cycle generally serve as parents for producing the BPFs of the next cycle. Thus, further research is warranted to investigate how information from previous breeding cycles can be integrated in the design of subsequent cycles.

Conclusions

When implementing GP in recycling breeding with elite lines, breeders are faced with the conflict of working with a sufficiently large TS for predicting candidates in each of multiple BPFs and yet utilizing large number of BPFs to warrant a sufficient effective population size. A popular solution to this problem is combining all available BPFs in the TS, including

unrelated families. We confirmed with simulations and experimental data that big differences exist in the predictive value of individual BPFs as TS, depending on their relationship to the PS family. Whereas the mean prediction accuracy was highest for full-sibs followed by half-sib and unrelated families, the variation among ρ values showed an opposite trend. Thus, ρ of unrelated BPFs was often negative and including these families in a combined TS is usually detrimental. Consequently, merging all available BPFs in the TS for the sake of increasing N_{TS} is not the best option, because other TS compositions exist that yield higher ρ values. However, identifying the optimal set of BPFs to be combined in a TS, based on existing genomic and phenotypic data, is still an unsolved problem. To be on the safe side for achieving a positive selection response with GP, we recommend to include only full-sib and half-sib families in the TS for each BPF. For producing the entire set A of BPFs subject to GP in a breeding cycle, we propose a mating scheme such as the Round-Robin design, which yields two to four half-sib families for each BPF. A medium number of DH lines (e.g., $n_g \sim 50$) from each BPF is genotyped and phenotyped for training the prediction model. Finally, GP of the numerous candidates in the PS of each BPF is based on a specific TS, which comprises n_g DH lines from (a) the BPF itself (i.e., full-sibs) and (b) few ($n_H = 2$ to 4) half-sib families of this BPF. This represents a compromise between achieving high ρ values and securing a sufficiently large effective population size of the entire breeding program, thus, warranting a balance between short- and long-term selection progress with genomic selection.

Acknowledgements We acknowledged the researchers and institutions that developed the two data sets evaluated in this study and made them publicly available. We are indebted to Dr. Eva Bauer (Technical University of Munich) and Dr. Xuecai Zhang (CIMMYT) for assistance and sharing of additional information from the experimental data. PB was funded by the German Ministry of Education and Research (BMBF) within the scope of the funding initiative MAZE “Plant Breeding Research for the Bioeconomy” (Funding ID: 031B0195).

Author contribution statements AEM and DM designed the study. DM wrote the core simulation code with package HOT. DM and PB performed simulations and data analysis. PB created all tables and figures. AEM, WM, and PB wrote the manuscript based on an initial draft by DM. All authors read and approved the manuscript.

Compliance with ethical statements

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standards The experiments reported in this study comply with the current laws of Germany.

References

- Akdemir D, Isidro-Sánchez J (2019) Design of training populations for selective phenotyping in genomic prediction. *Sci Rep* 9:1446. <https://doi.org/10.1038/s41598-018-38081-6>
- Akdemir D, Sanchez JI, Jannink J-L (2015) Optimization of genomic selection training populations with a genetic algorithm. *Genet Sel Evol* 47:38. <https://doi.org/10.1186/s12711-015-0116-6>
- Albrecht T, Wimmer V, Auinger H-J et al (2011) Genome-based prediction of testcross values in maize. *Theor Appl Genet* 123:339–350. <https://doi.org/10.1007/s00122-011-1587-7>
- Bernardo R, Yu J (2007) Prospects for genomewide selection for quantitative traits in maize. *Crop Sci* 47:1082. <https://doi.org/10.2135/cropsci2006.11.0690>
- Brauner PC, Müller D, Schopp P et al (2018) Genomic prediction within and among doubled-haploid libraries from maize landraces. *Genetics* 210:1185–1196. <https://doi.org/10.1534/genetics.118.301286>
- Bustos-Korts D, Malosetti M, Chapman S et al (2016) Improvement of predictive ability by uniform coverage of the target genetic space. *G3 (Bethesda)* 6:3733–3747. <https://doi.org/10.1534/g3.116.035410>
- Clark SA, Hickey JM, Daetwyler HD, van der Werf JHJ (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. <https://doi.org/10.1186/1297-9686-44-4>
- Core Team R (2019) A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Crossa J, Pérez-Rodríguez P, Cuevas J et al (2017) Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci* 22:961–975. <https://doi.org/10.1016/j.tplants.2017.08.011>
- Daetwyler HD, Villanueva B, Woolliams JA (2008) Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE* 3:e3395. <https://doi.org/10.1371/journal.pone.0003395>
- Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185:1021–1031. <https://doi.org/10.1534/genetics.110.116855>
- Daetwyler HD, Calus MPL, Pong-Wong R et al (2013) Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* 193:347–365. <https://doi.org/10.1534/genetics.112.147983>
- de Roos APW, Hayes BJ, Spelman RJ, Goddard ME (2008) Linkage disequilibrium and persistence of phase in holstein–friesian, jersey and angus cattle. *Genetics* 179:1503–1512. <https://doi.org/10.1534/genetics.107.084301>
- Dekkers JCM (2007) Marker-assisted selection for commercial crossbred performance. *J Anim Sci* 85:2104. <https://doi.org/10.2527/jas.2006-683>
- Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome J* 4:250. <https://doi.org/10.3835/plantgenome2011.08.0024>
- Falconer D, Mackay T (1996) Introduction to quantitative genetics, 4th edn. Longmans Green, London
- García-Ruiz A, Cole JB, VanRaden PM et al (2016) Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc Natl Acad Sci* 113:E3995–E4004. <https://doi.org/10.1073/pnas.1519061113>
- Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389–2397. <https://doi.org/10.1534/genetics.107.081190>

- Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection for crop improvement. *Crop Sci* 49:1. <https://doi.org/10.2135/cropsci2008.08.0512>
- Hickey JM, Chiurugwi T, Mackay I, Powell W (2017) Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat Genet* 49:1297–1303. <https://doi.org/10.1038/ng.3920>
- Iheshiulor OOM, Woolliams JA, Yu X et al (2016) Within- and across-breed genomic prediction using whole-genome sequence and single nucleotide polymorphism panels. *Genet Sel Evol* 48:15. <https://doi.org/10.1186/s12711-016-0193-1>
- Isidro J, Jannink J-L, Akdemir D et al (2015) Training set optimization under population structure in genomic selection. *Theor Appl Genet* 128:145–158. <https://doi.org/10.1007/s00122-014-2418-4>
- Jannink J-L (2010) Dynamics of long-term genomic selection. *Genet Sel Evol* 42:35. <https://doi.org/10.1186/1297-9686-42-35>
- Jiang Y, Reif JC (2015) Modeling epistasis in genomic selection. *Genetics* 201:759–768. <https://doi.org/10.1534/genetics.115.177907>
- Khanesar MA, Teshnehlab M, Shoorehdeli MA (2007) A novel binary particle swarm optimization. In: 2007 Mediterranean conference on control and automation. IEEE, pp 1–6
- Laloë D (1993) Precision and information in linear models of genetic evaluation. *Genet Sel Evol* 25:557. <https://doi.org/10.1186/1297-9686-25-6-557>
- Lehermeier C, Krämer N, Bauer E et al (2014) Usefulness of multiparental populations of maize (*Zea mays* L.) for genome-based prediction. *Genetics* 198:3–16. <https://doi.org/10.1534/genetics.114.161943>
- Lian L, Jacobson A, Zhong S, Bernardo R (2014) Genomewide prediction accuracy within 969 maize biparental populations. *Crop Sci* 54:1514. <https://doi.org/10.2135/cropsci2013.12.0856>
- Liu L, Du Y, Huo D et al (2015) Genetic architecture of maize kernel row number and whole genome prediction. *Theor Appl Genet* 128:2243–2254. <https://doi.org/10.1007/s00122-015-2581-2>
- Martini JWR, Gao N, Cardoso DF et al (2017) Genomic prediction with epistasis models: on the marker-coding-dependent performance of the extended GBLUP and properties of the categorical epistasis model (CE). *BMC Bioinform* 18:3. <https://doi.org/10.1186/s12859-016-1439-1>
- Marulanda JJ, Melchinger AE, Würschum T (2015) Genomic selection in biparental populations: assessment of parameters for optimum estimation set design. *Plant Breed* 134:623–630. <https://doi.org/10.1111/pbr.12317>
- Melchinger AE, Utz HF, Schön CC (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics* 149:383–403
- Melchinger AE, Schopp P, Müller D et al (2017) Safeguarding our genetic resources with libraries of doubled-haploid lines. *Genetics* 206:1611–1619. <https://doi.org/10.1534/genetics.115.186205>
- Meuwissen THE, Hayes B, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829
- Meuwissen THE, Hayes B, Goddard M (2016) Genomic selection: a paradigm shift in animal breeding. *Anim Front* 6:6–14. <https://doi.org/10.2527/af.2016-0002>
- Mikel MA, Dudley JW (2006) Evolution of north American dent corn from public to proprietary germplasm. *Crop Sci* 46:1193. <https://doi.org/10.2135/cropsci2005.10-0371>
- Müller D (2019) BPSO: binary particle swarm optimization. R package version 1.0.0. <https://github.com/DominikMueller64/BPSO>
- Müller D, Broman KW (2017) Meiosis: simulation of meiosis in plant breeding research. R package. Version 1.0.0. <https://github.com/DominikMueller64/Meiosis>
- Olson KM, VanRaden PM, Tooker ME (2012) Multibreed genomic evaluations using purebred Holsteins, Jerseys, and Brown Swiss. *J Dairy Sci* 95:5378–5383. <https://doi.org/10.3168/jds.2011-5006>
- Riedelsheimer C, Melchinger AE (2013) Optimizing the allocation of resources for genomic selection in one breeding cycle. *Theor Appl Genet* 126:2835–2848. <https://doi.org/10.1007/s00122-013-2175-9>
- Riedelsheimer C, Endelman JB, Stange M et al (2013) Genomic predictability of interconnected biparental maize populations. *Genetics* 194:493–503. <https://doi.org/10.1534/genetics.113.150227>
- Rincent R, Laloë D, Nicolas S et al (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. <https://doi.org/10.1534/genetics.112.141473>
- Schopp P, Müller D, Technow F, Melchinger AE (2017a) Accuracy of genomic prediction in synthetic populations depending on the number of parents, relatedness, and ancestral linkage disequilibrium. *Genetics* 205:441–454. <https://doi.org/10.1534/genetics.116.193243>
- Schopp P, Müller D, Wientjes YCJ, Melchinger AE (2017b) Genomic prediction within and across biparental families: means and variances of prediction accuracy and usefulness of deterministic equations. *G3 (Bethesda)* 7:3571–3586. <https://doi.org/10.1534/g3.117.300076>
- Schrag TA, Westhues M, Schipprack W et al (2018) Beyond genomic prediction: combining different types of omics data can improve prediction of hybrid performance in maize. *Genetics* 208:1373–1385. <https://doi.org/10.1534/genetics.117.300374>
- Sneath PH, Sokal RR (1973) Numerical taxonomy: the principles and practice of numerical classification. Freeman, San Francisco
- VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Wientjes Y, Veerkamp RF, Bijma P et al (2015) Empirical and deterministic accuracies of across-population genomic prediction. *Genet Sel Evol* 47:5. <https://doi.org/10.1186/s12711-014-0086-0>
- Würschum T, Maurer HP, Weissmann S et al (2017) Accuracy of within- and among-family genomic prediction in triticale. *Plant Breed* 136:230–236. <https://doi.org/10.1111/pbr.12465>
- Zhang X, Pérez-Rodríguez P, Semagn K et al (2015) Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. *Heredity (Edinb)* 114:291–299. <https://doi.org/10.1038/hdy.2014.99>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.