Theoretical and Applied Genetics (2023) 136:176
https://doi.org/10.1007/s00122-023-04413-y

ORIGINAL ARTICLE

Genomic prediction in hybrid breeding: I. Optimizing the training set design

Albrecht E. Melchinger[1,2] · Rohan Fernando[3] · Christian Stricker[1] · Chris-Carolin Schön[1] · Hans-Jürgen Auinger[1]

Received: 8 February 2023 / Accepted: 23 June 2023 / Published online: 2 August 2023

Abstract

Key message **Training sets produced by maximizing the number of parent lines, each involved in one cross, had the highest prediction accuracy for H0 hybrids, but lowest for H1 and H2 hybrids.**

Abstract Genomic prediction holds great promise for hybrid breeding but optimum composition of the training set (TS) as determined by the number of parents (n_{TS}) and crosses per parent (c) has received little attention. Our objective was to examine prediction accuracy (r_a) of GCA for lines used as parents of the TS (I1 lines) or not (I0 lines), and H0, H1 and H2 hybrids, comprising crosses of type I0 × I0, I1 × I0 and I1 × I1, respectively, as function of n_{TS} and c. In the theory, we developed estimates for r_a of GBLUPs for hybrids: (i)$\hat{r}_a$ based on the expected prediction accuracy, and (ii) $\tilde{r}_a$ based on r_a of GBLUPs of GCA and SCA effects. In the simulation part, hybrid populations were generated using molecular data from two experimental maize data sets. Additive and dominance effects of QTL borrowed from literature were used to simulate six scenarios of traits differing in the proportion ($\tau_{SCA} = 1\%$, 6%, 22%) of SCA variance in σ_G^2 and heritability ($h^2 = 0.4$, 0.8). Values of $\tilde{r}_a$ and $\hat{r}_a$ closely agreed with r_a for hybrids. For given size $N_{TS} = n_{TS} \times c$ of TS, r_a of H0 hybrids and GCA of I0 lines was highest for $c = 1$. Conversely, for GCA of I1 lines and H1 and H2 hybrids, $c = 1$ yielded lowest r_a with concordant results across all scenarios for both data sets. In view of these opposite trends, the optimum choice of c for maximizing selection response across all types of hybrids depends on the size and resources of the breeding program.

Introduction

Genomic prediction has a huge potential for improving the efficiency of hybrid breeding as demonstrated by numerous studies with various allogamous crops such as maize, sunflower, rye, sugar beet, autogamous crops such as wheat, barley, triticale and partially allogamous crops such as oilseed rape (see Seye et al. (2020) for a recent review). The main reason is that the number of potential hybrids, which can be predicted using genomic data from their parents, is determined by multiplying the number of parents in each heterotic group. Consequently, the size of the prediction set (PS) increases in quadratic terms with the number of parents and allows to apply extremely high selection intensities (Westhues et al. 2017). Besides the optimum allocation of resources to be spent on phenotyping hybrids in the training set (TS) versus genotyping parent lines, the composition of the TS as determined by the number of parents and number of crosses per parent has a strong influence on the prediction accuracy (r_a) of the various types of hybrids in the PS (Technow et al. 2014; Seye et al. 2020). A parent of a hybrid in the PS is classified as an I1 line, if it has one or more hybrids in the TS; all other parents having no hybrid in the TS are classified as I0 lines. Thus, the PS is composed of H0, H1 and H2 hybrids, corresponding to crosses of type I0 × I0, I0 × I1 or I1 × I0, and I1 × I1 lines. Optimal integration of genomic selection in hybrid breeding requires detailed knowledge, how prediction accuracy of these different types of hybrids depends on the population structure of TS (Kadam et al. 2021). This applies not only to genomic but also to pedigree

Communicated by Antonio Augusto Franco Garcia.

✉ Albrecht E. Melchinger
albrechtmelchinger@gmail.com

[1] Plant Breeding, TUM School of Life Sciences, Technical University of Munich, 85354 Freising, Germany

[2] Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, 70599 Stuttgart, Germany

[3] Department of Animal Science, Iowa State University, Ames, IA 50011, USA

data (Bernardo 1996) and other types of "omics" data suggested in the literature for hybrid prediction including transcriptomic (Westhues et al. 2017; Zenke-Philippi et al. 2017; Seifert et al. 2018), and metabolomic data (Riedelsheimer et al. 2012; Westhues et al. 2017) or combinations of them (Schrag et al. 2018).

In principle, two approaches exist for organizing the TS for genomic prediction in hybrid breeding. In the "testcross" approach, a separate TS is generated for each parent population by crossing candidate lines with the same tester(s) from the opposite population and testcross performance is used as a proxy for the GCA of the lines. In the "factorial" approach, the TS consists of inter-population hybrids between candidates from each parent population produced according to an incomplete factorial mating design. Both approaches were compared with simulations (Seye et al. 2020) and experiments (Fristche-Neto et al. 2018; Lorenzi et al. 2022). Since these studies demonstrated the superiority of sparse factorial designs for hybrid prediction and reciprocal recurrent genomic selection, we focused our investigations on this approach, with a main focus on the prediction accuracy of H0, H1, and H2 hybrids, as well as the GCA of I0 and I1 lines.

Hybrid performance is the main criterion for cultivar development, while general combining ability (GCA) of the parent lines is the main criterion for selecting the most promising parents for generating the base material for the next breeding cycle (Hallauer et al. 2010). In general, testcross performance with a genetically narrow tester (i.e., inbred line or single cross) from the opposite population is used as a proxy for the GCA of the candidate lines (Albrecht et al. 2011; Lian et al. 2014; Auinger et al. 2021), but these predictions are confounded with specific combining ability (SCA) effects of the candidates with the tester. In particular, a link between the prediction accuracy of hybrids and the prediction accuracy of their SCA and GCA of their parents is missing. A deeper understanding of the relationship between these components could help integrating product development with recurrent improvement of the parent populations in a genomic-based comprehensive approach to hybrid breeding.

Investigations on the optimum composition of the TS for hybrid breeding have been based on the prediction error variance (Fristche-Neto et al. 2018) or the CDmean criterion (Kadam et al. 2021). The genomic relationship matrix required for these approaches was calculated treating the hybrid population itself as reference base. However, this ignores the fact that unlike in other breeding categories in plant breeding (e.g., line, clonal and population breeding), the two parents of a hybrid usually originate from two genetically diverse populations for optimum exploitation of heterosis (Melchinger and Gumber 1998). Using a model with GCA and SCA effects, Seye et al. (2020) investigated the optimum composition of the TS with simulations. They analyzed a hybrid population consisting of several families. Each family was composed of inter-population hybrids that were produced from parent lines derived from diallel crosses of four founder lines in each parent population. The authors based their comparisons on prediction accuracies calculated across all families of hybrids but it is unknown to what extent the variance among subgroup means affected the prediction accuracy of H0, H1 and H2 hybrids within families, which is of main interest to the breeder.

The magnitude of the prediction accuracy is of fundamental importance for the optimum design of the TS. Estimating the prediction accuracy for sets of H0, H1 and H2 hybrids by cross-validation is hardly feasible in hybrid breeding, because this would require a huge TS, exceeding by far the capacity of most breeding programs. Alternative to cross-validation, r_a can be estimated by an approximation of the expected prediction accuracy ($\hat{r}_a$) calculated from population parameters as described by Ould Estaghvirou et al. (2013). However, application of $\hat{r}_a$ to determine the prediction of H0, H1 and H2 hybrids under a GCA–SCA model has not been described in the literature hitherto.

Selection index formulas have been used in various studies of hybrid prediction with the GCA–SCA model assuming that the fixed effects in the mixed linear model are known (e.g. Seye et al. 2020). For an extension to the general case when fixed effects are unknown, we present formulas for calculating GBLUPs and r_a for hybrids, GCA and SCA effects using well-known results from mixed models (Henderson 1975). In the theory part, our goal was to derive formulas connecting r_a of hybrids to r_a of their GCA and SCA effects and provide estimates for r_a of hybrids, GCA and SCA effects. In the simulation part, our objective was to investigate r_a and two types of estimates ($\hat{r}_a$ and $\tilde{r}_a$) for different types of hybrids (H0, H1, H2) and GCA of I0 and I1 lines under various scenarios differing in the relative importance of SCA effects and heritabilities. In particular, we examined how these statistics are influenced by the number of hybrids in the TS $\left(N_{TS}\right)$ and its composition regarding the number of parent lines $\left(n_{TS}\right)$ sampled from each parent population versus the number of crosses per parent line (c) in the TS. Finally, we discuss the implications of our results for the optimized design of the TS in hybrid breeding programs.

Theory

We begin by formulating the statistical model and providing formulas for calculating GBLUPs (= best linear unbiased predictors based on genomic data) and their variances for hybrid performance and GCA and SCA effects under a general mixed model. Let F be the set of n_F female lines and

M the set of n_M male lines for which genomic data are available, and $H = F \times M$ the set of all $n_F \times n_M$ possible hybrid combinations in the factorial between the lines of F and M. The N_{TS} hybrids in the TS are a subset of H. The lines in sets $F_1 \subset F$ and $M_1 \subset M$, denoted as I1 lines, serve as female or male parents of at least one hybrid in the TS. (In our notation, we use capital and lower case letters for numbers referring to hybrids and parent lines, respectively). By contrast, the lines in $F_0 = F \backslash F_1$ and $M_0 = M \backslash M_1$ are referred to as I0 lines, because they are not used as parent of any hybrid in the TS.

We assume a fixed number N_{TS} of TS hybrids, which depends primarily on the budget of the breeder assigned for phenotyping of hybrids. The goal is to choose a subset $HT \subset H$ as TS such that the prediction accuracy for untested hybrids $H \backslash HT$ is maximized. However, in hybrid breeding we have to consider that the PS consists of different subsets of hybrids $\Phi_{s,t}(s,t = 0,1)$, with $\Phi_{0,0} = [F_0 \times M_0]$ comprising the H0 hybrids, $\Phi_{1,0} = [F_1 \times M_0]$ and $\Phi_{0,1} = [F_0 \times M_1]$ comprising the H1 hybrids, and $\Phi_{1,1} = [F_1 \times M_1] \backslash HT$ comprising the H2 hybrids. Owing to different relatedness to other members of the TS, the H0, H1, and H2 types of hybrids have different prediction accuracies (Technow et al. 2014).

The common model to subdivide the genotypic value $G_{i\times j}$ of a hybrid $i \times j$ from the cross of female line $i \in F$ with male line $j \in M$ is (Hallauer et al. 2010)

$$G_{i\times j} = \mu + g_{F,i} + g_{M,j} + s_{i\times j}, \tag{1}$$

where μ is the mean of the hybrids in H, $g_{F,i}$ and $g_{M,j}$ are the GCA effects of i and j, respectively, and $s_{i\times j}$ is the SCA effect of their hybrid combination. The mixed model linking the phenotypic data of the hybrids in the TS to the GCA of the female and male lines and the SCA of all possible hybrid combinations in H can be written as

$$\boldsymbol{y}_{TS} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}, \tag{2}$$

with $E(\boldsymbol{y}_{\mathrm{TS}}) = \boldsymbol{X}\beta$, $E(\boldsymbol{u}) = 0$, $E(\boldsymbol{e}) = 0$, $\mathrm{var}\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\boldsymbol{G} & 0\\ 0 & \boldsymbol{R}\end{bmatrix}$, $\mathrm{var}[\boldsymbol{y}_{TS}] = \boldsymbol{V} = \boldsymbol{R} + \boldsymbol{ZGZ}^{\boldsymbol{T}}$, where

$$\boldsymbol{Z} = \begin{bmatrix}\boldsymbol{Z}_F & \boldsymbol{Z}_M & \boldsymbol{Z}_H\end{bmatrix},\ \boldsymbol{u} = \begin{bmatrix}\boldsymbol{g}_F\\ \boldsymbol{g}_M\\ \boldsymbol{s}_H\end{bmatrix} \quad \text{and} \quad \mathrm{var}\begin{bmatrix}\boldsymbol{g}_F\\ \boldsymbol{g}_M\\ \boldsymbol{s}_H\end{bmatrix} = \boldsymbol{G} = \begin{bmatrix}\boldsymbol{G}_F & 0 & 0\\ 0 & \boldsymbol{G}_M & 0\\ 0 & 0 & \boldsymbol{G}_H\end{bmatrix} \tag{3}$$

Here, $\boldsymbol{y}_{TS} = (y_k)_{k\in \mathrm{TS}}$ is a vector of dimension N_{TS} of phenotypic observations of the hybrids in the TS, $\boldsymbol{\beta}$ is the vector of non-genetic fixed effects of the hybrids in the TS and $\boldsymbol{X}$ the design matrix linking these effects to the observations in $\boldsymbol{y}_{\mathrm{TS}}$, $\boldsymbol{u}$ is a vector of random effects with dimension $N = n_F + n_M + n_F \times n_M$, composed of the vectors $\boldsymbol{g}_F$ and $\boldsymbol{g}_M$ of GCA effects of all n_F female and n_M male lines in F and M, respectively, and the vector $\boldsymbol{s}_H$ of SCA effects of all $N_H = n_F \times n_M$ hybrid combinations in H, and $\boldsymbol{e}$ is the residual error. The vectors $\boldsymbol{g}_F$, $\boldsymbol{g}_M$ and $\boldsymbol{s}_H$ are assumed to be (i) pairwise uncorrelated because the parent lines are sampled independently from the female and male population, and (ii) arranged in the order of the numbering of the lines in set F and M, respectively, and the element s_k in vector $\boldsymbol{s}_H = (s_k)$ with $k = (i-1) \times n_M + j$ refers to $s_{i\times j}$ $\boldsymbol{Z}_F$, $\boldsymbol{Z}_M$ and $\boldsymbol{Z}_H$ are incidence matrices relating the phenotypic data in $\boldsymbol{y}_{TS}$ with the vectors $\boldsymbol{g}_F$, $\boldsymbol{g}_M$ and $\boldsymbol{s}_H$, which have columns of zeros, if the respective line is $\in F_0$ or $\in M_0$ or the hybrid combination is $\in H \backslash HT$, respectively. $\boldsymbol{G}_F = \sigma^2_{\mathrm{gca}F}\boldsymbol{K}_F$, $\boldsymbol{G}_M = \sigma^2_{\mathrm{gca}M}\boldsymbol{K}_M$ and $\boldsymbol{G}_H = \sigma^2_{\mathrm{sca}}\boldsymbol{K}_H$, where $\boldsymbol{K}_F$ and $\boldsymbol{K}_M$ are the kinship matrices among the n_F female lines and among the n_M male lines, respectively. The relationship matrix $\boldsymbol{K}_H$ for SCA effects is obtained as Kronecker product $\boldsymbol{K}_H = \boldsymbol{K}_F \otimes \boldsymbol{K}_M$, as originally shown for pedigree-based coancestries (Cockerham 1961). $\sigma^2_{\mathrm{gca}F}$ and $\sigma^2_{\mathrm{gca}M}$ are the GCA variances among unrelated homozygous lines from the female and male parent population, respectively, and σ^2_{sca} the SCA variance of unrelated single cross hybrids between lines of the two parent populations, from which the lines in F and M were sampled. $\boldsymbol{R}$ denotes the corresponding "error" matrix pertaining to the model in Eq. (2), where $\boldsymbol{R} = \sigma^2_e \boldsymbol{I}_{N_{\mathrm{TS}}}$ is assumed in most cases. The elements of $\boldsymbol{K}_F$ and $\boldsymbol{K}_M$ can be calculated from genomic data using established methods (VanRaden 2008).

Using results of Henderson (1975), the best linear unbiased predictor (BLUP) $\hat{\boldsymbol{u}}$ for vector $\boldsymbol{u}$ is obtained as:

$$\hat{\boldsymbol{u}} = \boldsymbol{GZ}^{\boldsymbol{T}}\boldsymbol{V}^{-1}\left(\boldsymbol{y}_{\mathrm{TS}} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right) = \boldsymbol{GZ}^{\boldsymbol{T}}\boldsymbol{P}\boldsymbol{y}_{\mathrm{TS}} = \boldsymbol{B}\boldsymbol{y}_{\mathrm{TS}} \tag{4}$$

with $\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y_{\mathrm{TS}}$, $\boldsymbol{P} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}$,and $\boldsymbol{B} = \boldsymbol{GZ}^{\boldsymbol{T}}\boldsymbol{P} = \begin{bmatrix}\boldsymbol{G}_F\boldsymbol{Z}_F^T\boldsymbol{P}\\ \boldsymbol{G}_M\boldsymbol{Z}_M^T\boldsymbol{P}\\ \boldsymbol{G}_H\boldsymbol{Z}_H^T\boldsymbol{P}\end{bmatrix} = \begin{bmatrix}\boldsymbol{B}_F\\ \boldsymbol{B}_M\\ \boldsymbol{B}_H\end{bmatrix}$, yielding for the vectors GCA and SCA effects the BLUPs

$$\hat{\boldsymbol{g}}_F = \boldsymbol{B}_F\boldsymbol{y}_{\mathrm{TS}},\ \hat{\boldsymbol{g}}_M = \boldsymbol{B}_M\boldsymbol{y}_{\mathrm{TS}},\ \hat{\boldsymbol{s}}_H = \boldsymbol{B}_H\boldsymbol{y}_{\mathrm{TS}}. \tag{5}$$

Thus, we get for the variance of the BLUPs

$$\mathrm{var}(\hat{\boldsymbol{u}}) = \boldsymbol{BVB}^T = \boldsymbol{L} \tag{6}$$

and

$$\mathrm{var}(\hat{\boldsymbol{g}}_F) = \boldsymbol{B}_F\boldsymbol{V}\boldsymbol{B}_F^T = \boldsymbol{L}_F,\ \mathrm{var}(\hat{\boldsymbol{g}}_M) = \boldsymbol{B}_M\boldsymbol{V}\boldsymbol{B}_M^T = \boldsymbol{L}_M \quad \text{and} \quad \mathrm{var}(\hat{\boldsymbol{s}}_{FH}) = \boldsymbol{B}_H\boldsymbol{V}\boldsymbol{B}_H^T = \boldsymbol{L}_H \tag{7}$$

Since all hybrids have a common mean μ, selection among them can be based on prediction of $h_{i\times j} = g_{F,i} + g_{M,j} + s_{i\times j}$ and does not require prediction of $G_{k\times l}$. This corresponds to predicting $\boldsymbol{h} = \boldsymbol{g}_F \otimes 1_{N_M} + 1_{N_F} \otimes \boldsymbol{g}_M + \boldsymbol{s}_H$, or

$$\boldsymbol{h} = \boldsymbol{W}\boldsymbol{u} \quad \text{with} \quad \boldsymbol{W} = \left[\boldsymbol{I}_{n_F} \otimes 1_{n_M} \;\; 1_{n_F} \otimes \boldsymbol{I}_{n_M} \;\; \boldsymbol{I}_{n_F \times n_M}\right], \tag{8}$$

where $\boldsymbol{I}_n$ refers to a unity diagonal matrix, 1_n to a unity vector of dimension n, and $\otimes$ denotes the Kronecker product. Because the BLUP of a linear function of random effects is equal to the linear function of the BLUPs of the random effects (Henderson 1984), the BLUP of $\boldsymbol{h}$ and its variance are obtained as

$$\begin{aligned}\hat{\boldsymbol{h}} &= \boldsymbol{W}\hat{\boldsymbol{u}} = \boldsymbol{W}\boldsymbol{B}\boldsymbol{y}_{TS} = \hat{\boldsymbol{g}}_F \otimes 1_{n_M} + 1_{n_F} \otimes \hat{\boldsymbol{g}}_M + \hat{\boldsymbol{s}}_H \\ &= \left(\boldsymbol{B}_F \otimes 1_{n_M} + 1_{n_F} \otimes \boldsymbol{B}_M + \boldsymbol{B}_H\right)\boldsymbol{y}_{\mathrm{TS}}\end{aligned} \tag{9}$$

and

$$\begin{aligned}\mathrm{var}\left(\hat{\boldsymbol{h}}\right) = \boldsymbol{W}\boldsymbol{B}\boldsymbol{V}\boldsymbol{B}^T\boldsymbol{W}^T = \boldsymbol{W}\boldsymbol{L}\boldsymbol{W}^T &= \boldsymbol{B}_F\boldsymbol{V}\boldsymbol{B}_F^T \otimes \boldsymbol{J}_{n_M} \\ &+ \boldsymbol{J}_{n_F} \otimes \boldsymbol{B}_M\boldsymbol{V}\boldsymbol{B}_M^T + \boldsymbol{B}_H\boldsymbol{V}\boldsymbol{B}_H^T + 2 \\ &\times \left[1_{n_F}^T \otimes \boldsymbol{B}_F\boldsymbol{V}\boldsymbol{B}_M^T \otimes 1_{n_M} + \boldsymbol{B}_F\boldsymbol{V}\boldsymbol{B}_H^T \otimes 1_{n_M}\right. \\ &\left. + 1_{n_F} \otimes \boldsymbol{B}_M\boldsymbol{V}\boldsymbol{B}_H^T\right],\end{aligned} \tag{10}$$

where $\boldsymbol{J}_n$ is a $n \times n$ matrix of ones.

For any subset $\Phi \subset \left\{1, \ldots, n_F + n_M + n_F \times n_M\right\}$ of elements in $\boldsymbol{u}$, we can express the prediction accuracy r_a as correlation $r\left(\hat{\boldsymbol{u}}_\Phi, \boldsymbol{u}_\Phi\right)$ between the true genetic values (TGV) $\boldsymbol{u}_\Phi$ and their GBLUPs $\hat{\boldsymbol{u}}_\Phi$:

$$r_a\left(\hat{\boldsymbol{u}}_\Phi\right) = r\left(\hat{\boldsymbol{u}}_\Phi, \boldsymbol{u}_\Phi\right) = \frac{\sum_{k\in\Phi}\left(\hat{u}_k - \bar{\hat{u}}_k\right)\left(u_k - \bar{u}_k\right)}{\sqrt{\left(\sum_{k\in\Phi k}\left(\hat{u}_k - \bar{\hat{u}}_k\right)^2\right)\left(\sum_{k\in\Phi}\left(u_k - \bar{u}_k\right)^2\right)}} = \frac{\hat{\boldsymbol{u}}^T\boldsymbol{S}_\Phi\boldsymbol{u}}{\sqrt{\left(\hat{\boldsymbol{u}}^T\boldsymbol{S}_\Phi\hat{\boldsymbol{u}}\right)\left(\boldsymbol{u}^T\boldsymbol{S}_\Phi\boldsymbol{u}\right)}}, \tag{11}$$

where a bar denotes the mean of $\hat{u}_k$ or u_k over $k \in \Phi$, $\boldsymbol{S}_\Phi = \boldsymbol{I}_\Phi - \frac{1}{|\Phi|}\boldsymbol{J}_\Phi$ is a centering matrix such that $\boldsymbol{I}_\Phi$ is a matrix of dimension $n_F + n_M + n_F \times n_M$ having values of 1 on the diagonal, if the corresponding index $n \in \Phi$, and zeros elsewhere, $\boldsymbol{J}_\Phi$ is a matrix of the same dimension having 1's, if both indices $k, l \in \Phi$, and zeros elsewhere, and $|\Phi|$ is the number of elements in Φ. For hybrid values, we have for $\Phi \subset \left\{1, \ldots, N_H = n_F \times n_M\right\}$

$$r_a\left(\hat{\boldsymbol{h}}_\Phi\right) = r\left(\hat{\boldsymbol{h}}_\Phi, \boldsymbol{h}_\Phi\right) = \frac{\hat{\boldsymbol{u}}^T\boldsymbol{W}^T\boldsymbol{S}_\Phi\boldsymbol{W}\boldsymbol{u}}{\sqrt{\left(\hat{\boldsymbol{u}}^T\boldsymbol{W}^T\boldsymbol{S}_\Phi\boldsymbol{W}\hat{\boldsymbol{u}}\right)\left(\boldsymbol{u}^T\boldsymbol{W}^T\boldsymbol{S}_\Phi\boldsymbol{W}\boldsymbol{u}\right)}} \tag{12}$$

In simulations with known values of $\boldsymbol{u}$, this formula can be used to calculate r_a for hybrid prediction. If the TGV are unknown, as is the case in practice, one could use cross-validation to determine the predictive ability R_a of GBLUPs replacing in Eqs. (11 and 12) the TGV by phenotypic data and get an estimate of r_a by the ratio $R_a/\sqrt{h^2}$, where h^2 is the heritability of the phenotypic data (cf. Dekkers 2007). However, cross-validation can be circumvented by an alternative approach suggested by Ould Estaghvirou et al. (2013). Accordingly, an approximation of $E\left[r\left(\hat{\boldsymbol{u}}_\Phi, \boldsymbol{u}_\Phi\right)\right]$ and consequently an estimate of $r_a\left(\hat{\boldsymbol{u}}_\Phi\right)$ can be obtained for any subset Φ as (see Appendix 1):

$$\hat{r}_a\left(\hat{\boldsymbol{u}}_\Phi\right) = \sqrt{\frac{tr\left(\boldsymbol{S}_\Phi\boldsymbol{L}\right)}{tr\left(\boldsymbol{S}_\Phi\boldsymbol{G}\right)}} = \sqrt{\frac{\sum_{i\in\Phi}\left(l_{ii} - l_{i.}\right)}{\sum_{i\in\Phi}\left(g_{ii} - g_{i.}\right)}} \tag{13}$$

and

$$\hat{r}_a\left(\hat{\boldsymbol{h}}_\Phi\right) = \sqrt{\frac{tr\left(\boldsymbol{S}_\Phi\boldsymbol{W}\boldsymbol{L}\boldsymbol{W}^T\right)}{tr\left(\boldsymbol{S}_\Phi\boldsymbol{W}\boldsymbol{G}\boldsymbol{W}^T\right)}} \tag{14}$$

The matrices $\boldsymbol{G}$ and $\boldsymbol{L}$ can be calculated (*cf.* Equations (3 and 6)) without phenotypic data of hybrids using genomic data of the parent lines, variance components σ^2_{gcaF}, σ^2_{gcaM}, σ^2_{sca}, σ^2_e estimated from previous breeding cycles and the incidence matrices $\boldsymbol{X}$ (usually $\boldsymbol{X} = 1$) and $\boldsymbol{Z}$ pertaining to the fixed and random effects in Eq. (2), respectively, and matrix $\boldsymbol{W}$ as defined in Eq. (8).

As shown in Appendix 2, the expected individual prediction accuracy (ρ_a) for a randomly chosen hybrid $i \times j$ can be approximated by ρ_a of its SCA and parental GCA effects as

$$\rho_a\left(\hat{h}_{i\times j}, h_{i\times j}\right) \approx \sqrt{\rho_a^2\left(\hat{g}_{F,i}, g_{F,i}\right)\tau_{\mathrm{gcaF}} + \rho_a^2\left(\hat{g}_{M,j}, g_{M,j}\right)\tau_{\mathrm{gcaM}} + \rho_a^2\left(\hat{s}_{i\times j}, s_{i\times j}\right)\tau_{\mathrm{sca}}}, \tag{15}$$

where $\tau_{gcaF} = \frac{\sigma^2_{\mathrm{gcaF}}}{\sigma^2_G}$, $\tau_{\mathrm{gcaM}} = \frac{\sigma^2_{\mathrm{gcaM}}}{\sigma^2_G}$ and $\tau_{\mathrm{sca}} = \frac{\sigma^2_{\mathrm{sca}}}{\sigma^2_G}$ is the proportion of the total genetic variance $\sigma^2_G = \sigma^2_{\mathrm{gcaF}} + \sigma^2_{\mathrm{gcaM}} + \sigma^2_{\mathrm{sca}}$

among unrelated hybrids attributable to the GCA and SCA variances, respectively, with $\tau_{\mathrm{gca}F} + \tau_{\mathrm{gca}M} + \tau_{\mathrm{sca}} = 1$. If set Φ is equal to $\Phi_{s,t}$ $(s, t = 0, 1)$, corresponding to the set of hybrids of type H0, H1 and H2, respectively, a similar approximation applies to the prediction accuracies r_a of hybrids and GCA and SCA effects (see Appendix 2, Eq. (28)). This relationship can be used to obtain a further estimate $\tilde{r}_a\left(\hat{\boldsymbol{h}}_{\Phi_{s,t}}\right)$ of $r_a\left(\hat{\boldsymbol{h}}_{\Phi_{s,t}}\right)$ (see Appendix 2, Eq. (30))

$$\tilde{r}_a\left(\hat{\boldsymbol{h}}_{\Phi_{s,t}}\right) = \sqrt{r_a^2\left(\hat{\boldsymbol{g}}_{F,F_s}\right)\tau_{\mathrm{gca}F} + r_a^2\left(\hat{\boldsymbol{g}}_{M,M_t}\right)\tau_{\mathrm{gca}M} + r_a^2\left(\hat{\boldsymbol{s}}_{\Phi_{s,t}}\right)\tau_{\mathrm{sca}}} \tag{16}$$

where $r_a\left(\hat{\boldsymbol{g}}_{F,F_s}\right)$, $\tau_{\mathrm{gca}F}$ etc. must be substituted by appropriate estimates. This relationship holds true for H0 and H1 hybrids irrespective of the structure of the TS. For H2 hybrids, the TS must have the structure of a balanced incomplete factorial design, where $|F_1| = |M_1|$ and each line $i \in F_1$ was crossed to the same number c of parents from M_1 and vice versa.

Genetic materials, markers and genomic relationships

We based our simulations on the SNP marker genotypes of maize inbreds from two experiments conducted by the maize breeding program of the University of Hohenheim. Since the lines were expected to be fully homozygous, heterozygous marker genotypes were treated as "missing." Markers with more than 5% missing values were removed; otherwise, "missing" values were imputed using BEAGLE version 5.0 (Browning and Browning 2016). To avoid clustering of markers in small genomic segments, we restricted the number of markers to a maximum of 10 per Mbp and retained from markers with linkage disequilibrium $r^2 \geq 0.999$ only one. The marker genotypes of hybrids were inferred from the genotypes of their parent lines.

Data set DS1 comprised SNP data of $n_F = 145$ dent lines (=females) and $n_M = 111$ flint lines (=males) genotyped with the 50 k Illumina SNP chip MaizeSNP50 (Ganal et al. 2011). The lines had been evaluated for GCA of grain yield and other important agronomic traits in testcross trials and selected to be used as parents of hybrids evaluated in factorials analyzed in various studies on hybrid prediction with different types of "omics" data (Technow et al. 2014; Westhues et al. 2017; Schrag et al. 2018). From the original set of 26,795 polymorphic markers in the 256 lines, 13,813 markers remained after quality check and pruning, denoted as set *SNP*1. From these, 12,058 were polymorphic within the 145 dent lines and 12,053 within the 111 flint lines. None of the SNPs was polymorphic between the two groups but monomorphic within each group. The polymorphic markers provided a largely uniform coverage of the entire maize genome. The genetic diversity of the lines and their linkage disequilibrium (LD) structure has been described in detail elsewhere (Technow et al. 2012, 2014). Our analyses were based on the $N_H = n_F \times n_M = 16,095$ hybrids that could be simulated from the crosses between the factorial of the female and male parent lines.

Data set DS2 included SNP data from an unpublished experiment with $n_F = 182$ dent inbreds and $n_M = 162$ flint inbreds developed by in vivo haploid induction (Chaikam et al. 2019) from a single cross of two elite dent founder parents and a single cross of two related (coancestry $f = 1/4$) elite flint founder parents, respectively. The lines were unselected except for seed set in the first generation (D_0) of doubled-haploid plants to secure efficient line multiplication. All lines and their four homozygous parents were genotyped with the 600 k Affymetrix® Axiom® Maize Array (Unterseer et al. 2014). The number of polymorphic markers in set *SNP*2 amounted to 107,135 SNPs, with 2428 remaining after pruning because many of them were completely linked, out of which 2157 SNPs segregated in the dent lines and 1311 in the flint lines, and 1040 in both populations. As expected for doubled-haploid lines from bi-parental crosses, there were numerous monomorphic regions in both crosses, especially for in the flint population as a consequence of the relatedness of the founder parents. All analyses were based on the $N_H = n_F \times n_M = 29,484$ inter-population hybrids that could be produced in silico among the female and male parent lines.

For each data set, genomic kinship matrices $\boldsymbol{K}_F$ and $\boldsymbol{K}_M$ of the female and male lines were calculated with Method 1 of VanRaden (2008) using the respective parent population to determine the frequency of the reference allele. The matrix $\boldsymbol{K}_H$ was subsequently obtained as the Kronecker product $\boldsymbol{K}_F \otimes \boldsymbol{K}_M$.

Traits

To simulate traits by our software module A (Figure S1), we followed the procedure described in previous papers (Technow et al. 2012; Esfandyari et al. 2015; Seye et al. 2020) with modifications. Briefly, a random set *QP* of 3000 SNPs from set *SNP*1 were chosen as possible QTL positions in DS1, with the restriction that the number of SNPs was approximately proportional to the genetic length of the chromosomes. In DS2 a random set *QP* of 500 SNPs from set *SNP*1 were chosen as possible QTL with the restriction that number of QTL was proportional the map length of the polymorphic regions (DS2). A random subset $Q \subset QP$ of n_Q QTL were assigned additive (a_l) and dominance effects ($d_l = a_l \times k_l$), where k_l is the degree of dominance, defined according to Lynch and Walsh (1998). The additive effects

a_l were drawn from a Gamma distribution with parameter scale = 1.66 and shape = 0.4 and assigned to the reference allele coded as 1 in the SNP data. The degree of dominance k_l was drawn from a normal distribution $N(\mu_k, \sigma_k^2)$. A subset $Q_d \subset Q$ of n_{Q_d} QTL displaying only dominance effects d_l was finally obtained by setting $a_l = 0$. The parameters μ_k, σ_k^2 were chosen based on the average degree of dominance summarized by Hallauer et al. (2010) from numerous experiments with maize for grain yield, maturity, resistance and quality traits. QTL studies of heterotic traits in elite hybrids with the NC design III (Garcia et al. 2008; Schön et al. 2010), triple testcross design (Frascaroli et al. 2007), or the F2 and immortalized F2 design (Stuber et al. 1992; Tang et al. 2010; Guo et al. 2014) served as reference point for determining the ratio $n_{Q_d} : n_Q$. Table S1 shows the values of n_Q, n_{Q_d}, μ_k, σ_k leading to three types of "target" traits with $\tau_{sca} = 1\%, 6\%, 22\%$, which in combination with $h^2 = 0.4, 0.8$ define the six scenarios analyzed in our study. The plausibility of our model assumptions for simulating different types of traits was confirmed by the close agreement of the τ_{sca} values in the simulated hybrid populations with experimental estimates for yield, maturity, and quality traits from the maize literature (Table S2).

Based on the simulated QTL genotypes for set Q, the genotypic value of every hybrid was determined by summing the corresponding additive and dominance effects, respectively, across all QTL. Subsequently, the genotypic values were scaled to unit variance and centered to zero mean. Phenotypic values of the hybrids were obtained by adding to the genotypic values a normally distributed noise variable with variance $\sigma_e^2 = 1/h^2 - 1$ to obtain the desired broad sense heritabilities. By averaging the genotypic values over all hybrid combinations of a given line, we obtained its "true" GCA. The "true" SCA of each hybrid combination was obtained by subtracting the GCA of both parents from the genotypic value of the hybrid. Variance components σ_{gcaF}^2, σ_{gcaM}^2, σ_{sca}^2, σ_G^2 and the ratio $\tau_{sca} = \sigma_{sca}^2 : \sigma_G^2$ were determined from the GCA and SCA values of the complete factorial (= set H) and used for calculating GBLUPs as described below. For each of the six scenarios, simulation of each type

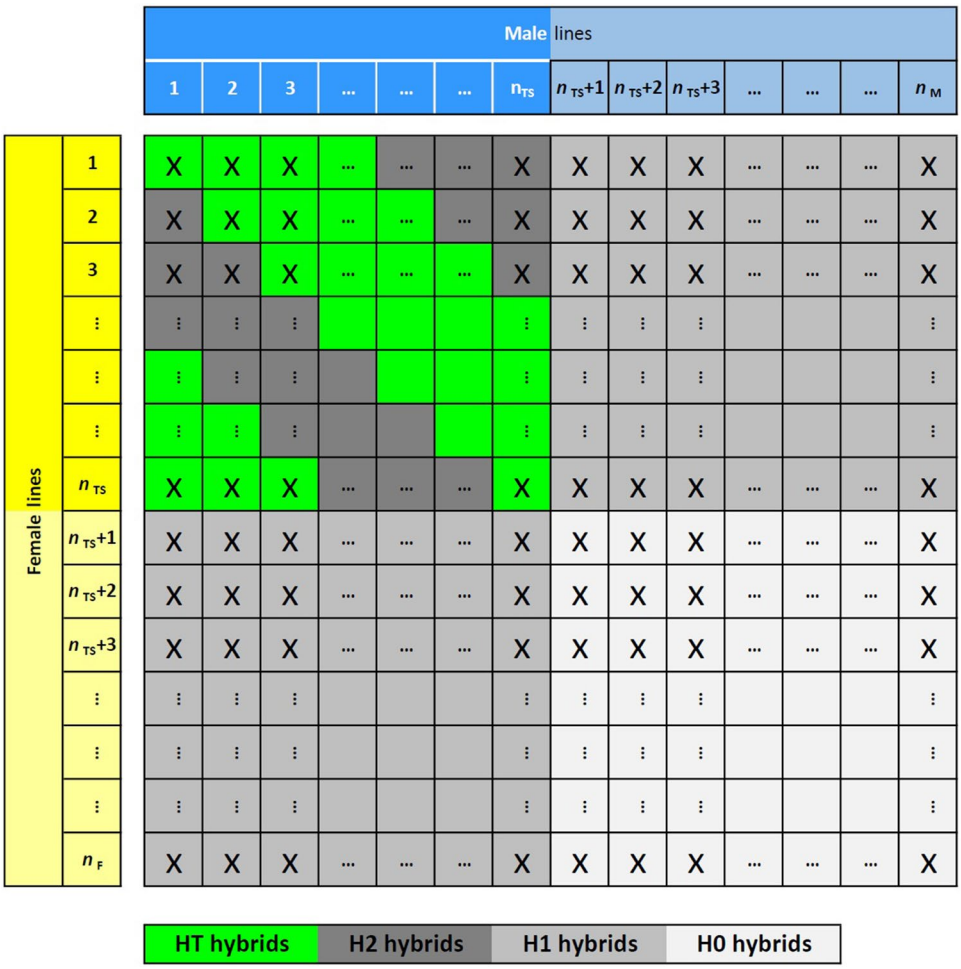

Fig. 1 Schematic representation of the training set (TS) of hybrids (HT hybrids, $N_{TS} = 28$ green) as determined by the number of lines ($n_{TS} = 7$) sampled from each parent population (females = yellow, males = blue) and crosses per parent line (here $c = 4$) used for genomic prediction of hybrid performance and GCA of the parent lines. I0 and I1 lines are shown with weak and strong color intensity, respectively, and H0, H1 and H2 hybrids by increasing intensity levels of gray (color figure online)

of target trait was replicated 50 times by sampling always anew QTL positions and effects and the noise variable.

Prediction of hybrids

For simulating the TS and PS for data set DS1 and DS2 in software module B (Figure S1), we first sampled randomly n_{TS} lines from each of the sets F and M to obtain the subsets F_1 and M_1 of I1 lines for producing the TS hybrids. The parental lines were crossed such that each male from M_1 was mated to c females from F_1 and each female from F_1 was mated to c male lines from M_1 to obtain a total of $N_{\mathrm{TS}} = n_{\mathrm{TS}} \times c$ hybrid combinations for the TS according to the scheme depicted in Fig. 1 for$c = 4$. We varied $n_{\mathrm{TS}} = 12, 24, 36, ..., 96$ and $n_{\mathrm{TS}} = 12, 24, 36, ..., 144$ for DS1 and DS2, respectively, and $c = 1, 2, 4$ to investigate the effect of the composition of the TS on the prediction accuracy for GCA of both I0 and I1 lines as well as the SCA and hybrid performance for H0, H1, and H2 hybrids marked by different colors in Fig. 1. Sampling of the subset of I1 lines in each parent population and production of crosses for generating the TS was repeated 20 times. Hence, for each scenario we had a total of 1000 simulation runs, corresponding to 50 replications for each type of target trait × 20 parent samplings for the TS.

In software module C (Figure S1), we performed GBLUP for genomic prediction of all hybrids in set H and GCA effects of all lines in set F and M on the basis of the GCA–SCA model in Eq. (1). The variance components $\sigma^2_{\mathrm{gca}F}$, $\sigma^2_{\mathrm{gca}M}$, σ^2_{sca} calculated from set H as described above were used in our calculations, which enabled us to finish 1000 simulation runs for each scenario within acceptable time. This procedure is only feasible in simulations but in practice, estimated values of the variance components obtained from phenotypic values and genomic relationships in the TS must be employed in GBLUP. For comparison, we therefore calculated GBLUPs with variance components estimated from the phenotypic data and genomic relationships of the hybrids in the TS for $N_{\mathrm{TS}} = 96$ in both data sets and $N_{\mathrm{TS}} = 144$ in DS2. The variance components were estimated as posterior mean obtained by a Gibbs sampler with 2500 burn-ins and a chain length of 10,000 (Sorensen and Gianola 2002), which warranted satisfactory convergence. The close agreement of r_a values obtained with both methods shown in Figs. 2 and 3 justified this procedure.

For final analysis, we calculated in each simulation run the prediction accuracy for genomic prediction of hybrids, GCA and SCA as Pearson correlation of their GBLUPs and their known genetic values across all genotypes in the respective set using Eqs. (11, 12). Furthermore, we calculated the approximation $\widetilde{r}_a$ of r_a with Eq. (16). Likewise, $\widehat{r}_a$ of the GBLUPs was calculated by inserting the required population parameters ($\boldsymbol{K}_F$, $\boldsymbol{K}_M$, $\boldsymbol{K}_H$,$\sigma^2_{\mathrm{gca}F}$, $\sigma^2_{\mathrm{gca}M}$, σ^2_{sca}, σ^2_e) in Eqs. (13, 14). Finally, we calculated the mean of each statistic across the 1000 simulation runs for each scenario as well as the corresponding 95% confidence interval, which was for the r_a, $\widehat{r}_a$ and $\widetilde{r}_a$ values of all types of hybrids smaller than 1% of the mean. All computations were performed with the Julia programming language (Bezanson et al. 2017).

Data availability statement

The marker data for sets DS1 and DS2, the positions of the markers and the Julia program are available at https://github.com/TUMplantbreeding/OptimTrainingSetDesign and can be downloaded from there.

Results

Figure 2A shows the curves of r_a for the three types (H0, H1, H2) of hybrids in the PS and all hybrids in set H of data set DS1 as a function of the number of parent lines n_{TS} and number of crosses $c = 1, 2, 4$ per parent in the TS, which together determine the size of the TS ($N_{\mathrm{TS}} = n_{\mathrm{TS}} \times c$). For given h^2 and type of hybrid, the shape of the curves was almost congruent irrespective of c and τ_{SCA}, with a concave curvature that flattened out for larger n_{TS} except for almost horizontal curves for H2 hybrids if $h^2 = 0.8$ and $c = 4$. The level of r_a increased substantially by doubling c from 1 to 2 but a further doubling to $c = 4$ yielded a much smaller increase for high h^2 as indicated by the distance between the curves. Increasing h^2 from 0.4 to 0.8 increased r_a by ~ 30% for all types of hybrids. The r_a curves differed little between $\tau_{\mathrm{SCA}} = 1\%$ and 6% but were at a substantially lower level for $\tau_{\mathrm{SCA}} = 22\%$. The r_a values for all hybrids in H followed closely the curves for H0 hybrids but with a steeper slope.

Using a fixed size $N_{\mathrm{TS}} = 96$ of the TS as a benchmark (circles in Fig. 2A), the value of c maximizing r_a depended on the type of hybrid but for each type, the ranking of r_a values for $c = 1, 2, 4$ was identical for the six scenarios. For H0 hybrids, r_a was maximum for $c = 1$, exceeding r_a values for $c = 2$ and $c = 4$ by ~ 1–3% and ~ 5–10%, respectively, with largest differences for high h^2. The r_a values for H1 and H2 hybrids were always smallest for $c = 1$ followed by $c = 2$ and $c = 4$, with differences being much larger for H2 than H1 hybrids. The curves of r_a for GCA of I0 and I1 lines had a striking similarity with the corresponding curves for H0 and H2 hybrids, albeit at a slightly higher level (Fig. 2B).

The curves of r_a for data set DS2 showed essentially the same picture as those for DS1, with some trends being amplified (Fig. 3). Again, the curves for H0 and H2 hybrids

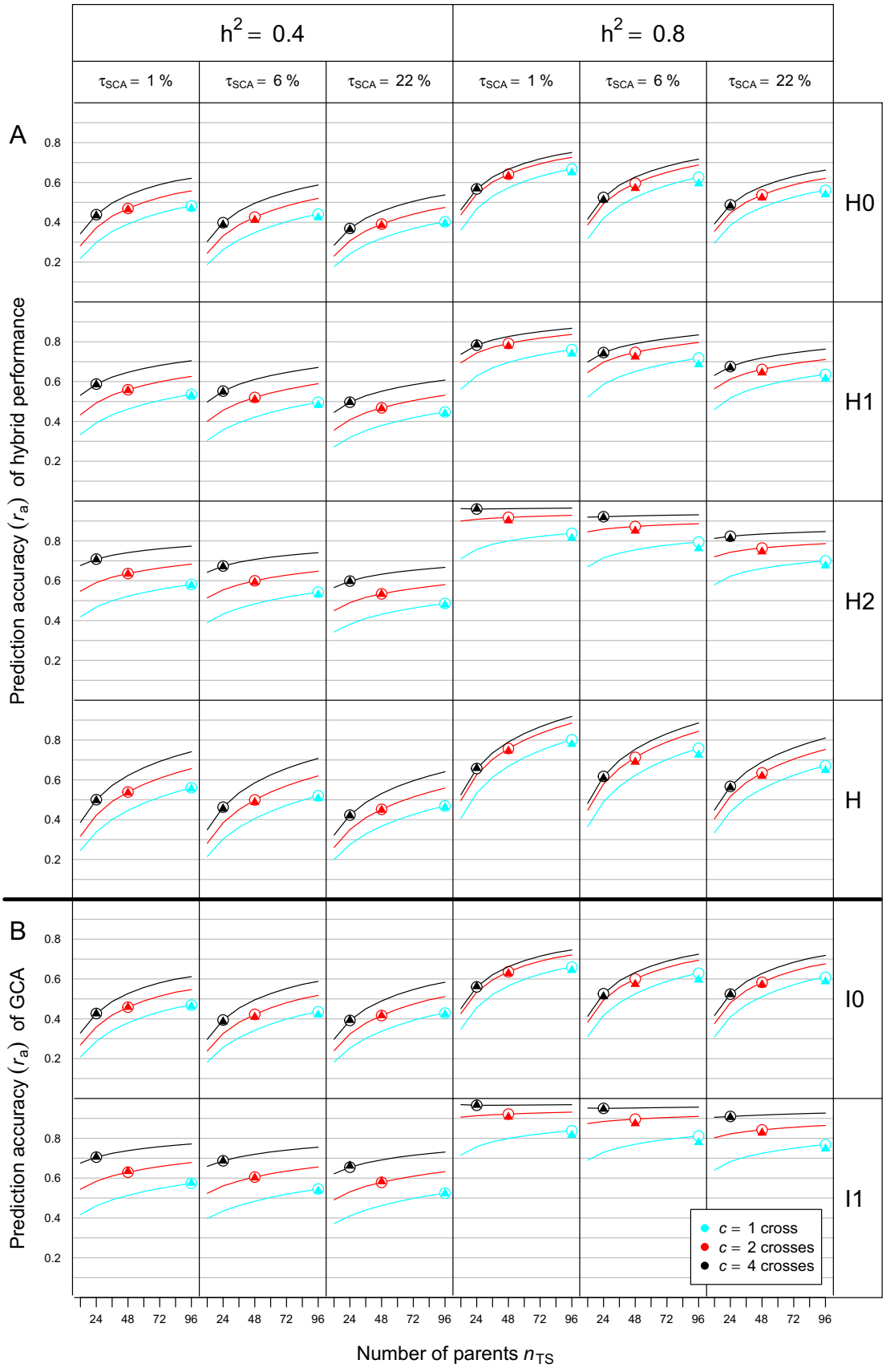
h² = 0.4
h² = 0.8
τSCA = 1 %
τSCA = 6 %
τSCA = 22 %
A
B
H0
H1
H2
H
I0
I1
Prediction accuracy (ra) of hybrid performance
Prediction accuracy (ra) of GCA
0.2
0.4
0.6
0.8
24
48
72
96
Number of parents nTS
c = 1 cross
c = 2 crosses
c = 4 crosses

Fig. 2 Prediction accuracy (r_a) for **A** H0, H1, and H2 type of hybrids and all hybrids in set *H* and **B** GCA of I0 and I1 lines as a function of the number n_{TS} of parent lines and number of crosses per parent ($c = 1, 2, 4$) used for producing the training set (TS). Results refer to means of 1000 simulation runs based on data set DS1 for different values of h^2 and τ_{SCA} (proportion of the SCA variance in σ_g^2 of hybrids). Circles and triangles refer to r_a values for $N_{TS} = n_{TS} \times c = 96$ obtained with GBLUPs calculated with "true" and estimated variance components, respectively

were almost identical with those for GCA of I0 and I1 lines, respectively. However, the curves had initially a steeper slope, especially for GCA of I0 lines, and approached a plateau at $N_{\mathrm{TS}} = 144$ for high h^2. For $N_{\mathrm{TS}} = 96$, the maximum r_a value for H0 hybrids showed in all six scenarios a larger difference between $c = 1$ and $c = 4$ than for DS1. The maximum r_a for $c = 1$ referring to $\tau_{\mathrm{SCA}} = 22\%$ and 1% ranged for low h^2 between 0.50 and 0.65 and for high h^2 between 0.70 and 0.85, respectively. By comparison, r_a values of H1 hybrids hardly differed between $c = 1, 2, 4$ for $N_{\mathrm{TS}} = 96$ and $N_{\mathrm{TS}} = 144$. For H2 hybrids and given values of N_{TS}, differences between r_a for different values of c were substantial for all scenarios, most notably for $h^2 = 0.4$. The prediction accuracy across all hybrids in set H had its maximum for $c = 1$ in all scenarios. For $N_{\mathrm{TS}} = 96$, the difference to r_a for $c = 2$ ranged between 1% ($h^2 = 0.4$, $\tau_{\mathrm{sca}} = 22\%$) and 14% ($h^2 = 0.8$, $\tau_{\mathrm{sca}} = 1\%$), but for $N_{TS} = 144$, these differences became smaller.

For the subset of simulations with $N_{\mathrm{TS}} = 96$ for DS1 and $N_{\mathrm{TS}} = 96$ and 144 for DS2, the r_a values for hybrid performance obtained with GBLUPs calculated with estimated variance components (triangles) were almost identical with the corresponding r_a values of GBLUPs calculated with the "true" variance components (circles and diamonds) (Figs. 2 and 3). The latter estimates showed for $h^2 = 0.8$ a small upward bias of less than 3% for $c = 1, 2$ in all types of hybrids. As expected the SD of r_a values was generally much larger for r_a values calculated with estimated variance components due to the estimation error associated with them (results not shown).

The r_a values for SCA were for all types of hybrids in both data sets (Figures S2 and S3) much smaller than those for GCA of I0 and I1 lines. For all scenarios and types of hybrids, r_a values were much lower for DS1 than DS2. For $h^2 = 0.4$, the prediction accuracy of SCA was lower than 0.28 for all cases. For $h^2 = 0.8$, the prediction accuracy was of moderate size for H2 hybrids in both data sets, but even in the most favorable case with $\tau_{\mathrm{SCA}} = 22\%$ and $c = 4$, r_a for the hybrids in *H* barely exceeded 0.4 for $N_{\mathrm{TS}} = 144$ in DS2. Estimates of r_a for SCA effects obtained from GBLUPs calculated with estimated variance components (triangles) were almost identical to those obtained with "true" variance components (circles and diamonds).

Calculating $\widetilde{r}_a$ by Eq. (16) as a function of the r_a and τ values of GCA and SCA effects provided an excellent approximation of r_a values for all scenarios and values of n_{TS} and c in both data sets (Figures S4 and S5). Similarly, the curves for $\widehat{r}_a$ of hybrids and GCA effects calculated according to Eqs. (13, 14) had identical shape as the curves for r_a in all scenarios (Figures S6 and S7). In both data sets, $\widehat{r}_a$ showed in comparison with r_a a slight upward and minor downward bias for low and high h^2, respectively.

Discussion

Our research pertains to the prediction of inter-population hybrids produced by crossing lines of two genetically distant populations. This setting is typical for established hybrid breeding programs in maize and other allogamous crops, because organizing the germplasm in genetically divergent parent populations warrants optimum exploitation of heterosis and reduces the proportion τ_{SCA} of the SCA variance in σ_g^2 of hybrids according to experimental results (Melchinger and Gumber 1998) and theoretical arguments (Reif et al. 2007). Our simulations show that smaller values of τ_{SCA} increase the prediction accuracy for all types of hybrids in the PS, irrespective of the data set and the size and design of the TS (Figs. 2 and 3). Thus, our results support the conclusion of Zhao et al. (2015) that genetically distant heterotic groups are advantageous for both conventional hybrid breeding and implementation of genomic prediction.

Choice of trait architecture, parent populations and genetic model

By using simulations, we were able to investigate various scenarios in hybrid breeding with a large number of replications and to determine the prediction accuracy directly by correlating predicted and "true" genotypic values, thereby bypassing the estimation of r_a by means of cross-validation. Following Fisher (1918), we assumed a large number of QTL underlying the genetic architecture of complex quantitative traits with small additive and dominance effects as practiced in previous studies with similar objectives (Technow et al. 2012; Seye et al. 2020). We ignored epistasis given its minor importance in experimental studies with maize (Melchinger et al. 1988; Lamkey et al. 1995) and the low importance of statistical epistasis at the level of populations even in the presence of significant physiological epistasis at

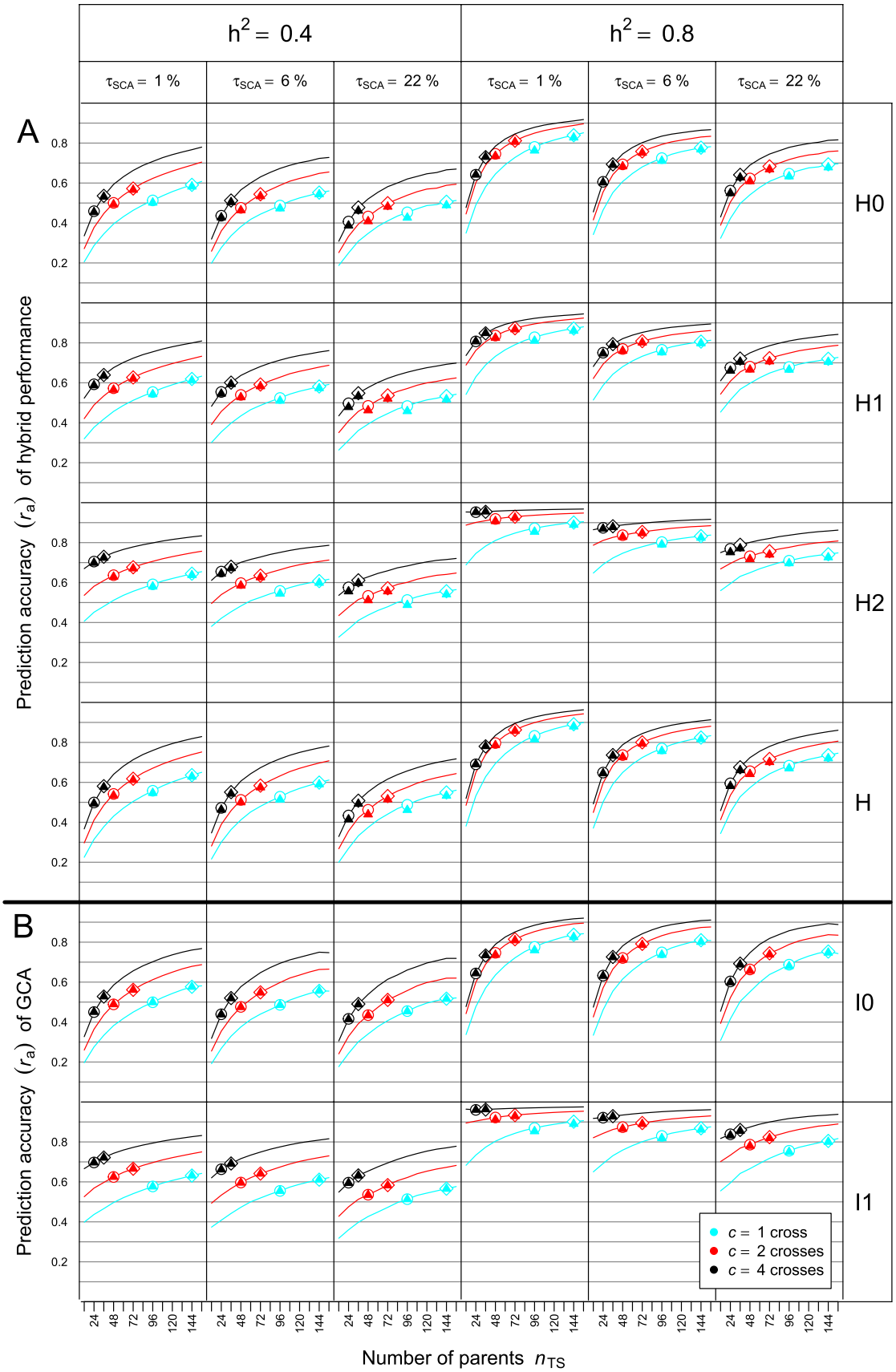
h² = 0.4
h² = 0.8
τSCA = 1 %
τSCA = 6 %
τSCA = 22 %
A
B
H0
H1
H2
H
I0
I1
Prediction accuracy (ra) of hybrid performance
Prediction accuracy (ra) of GCA
0.2
0.4
0.6
0.8
24
48
72
96
120
144
Number of parents nTS
c = 1 cross
c = 2 crosses
c = 4 crosses

Fig. 3 Prediction accuracy (r_a) for **A** H0, H1, and H2 type of hybrids and all hybrids in set H and **B** GCA of I0 and I1 lines as a function of the number n_{TS} of parent lines and number of crosses per parent ($c = 1, 2, 4$) used for producing the training set (TS). Results refer to means of 1000 simulation runs based on data set DS2 for different values of h^2 and τ_{SCA} (proportion of the SCA variance in σ_g^2 of hybrids). Circles and triangles refer to results for $N_{TS} = n_{TS} \times c = 96$ and diamonds and triangles refer to results for $N_{TS} = n_{TS} \times c = 144$ obtained with GBLUPs calculated with "true" and estimated variance components, respectively

the level of individual genotypes (Hill et al. 2008; Sackton and Hartl 2016).

The genotypic data underlying our simulations were taken from an active maize breeding program to warrant high practical relevance. Data set DS1 represents the situation when numerous preselected lines are available and a limited number of most promising hybrid combinations are to be produced and evaluated for final product development. The lines in DS1 had been selected based on their line perse and testcross performance but we expect similar outcomes if they were identified with genomic selection. Data set DS2 can be viewed as an application of the Hallauer (1967) proposal of full-sib selection for hybrid breeding, where single crosses instead of testcrosses are evaluated at each stage of the breeding cycle, in which case genomic prediction is the only way to exploit the effects of Mendelian sampling in the parent populations.

We limited our investigation to the classical GCA–SCA model of Sprague and Tatum (1942) for modeling hybrid performance. Kadam et al. (2021) used a model with only additive effects for investigating the optimum TS composition, whereas Fristche-Neto et al. (2018) used additionally a model with both additive and dominance effects. The GCA–SCA model has the advantage that it captures population-specific effects of SNPs, which is important if the linkage phase and strength of linkage disequilibrium (LD) between QTL and SNPs and/or the QTL effects differ between the parent populations. A systematic comparison of the prediction accuracy of the different models is missing in the literature and warrants further research with experimental data but we do not expect that choice of the model will strongly affect the optimum design of the TS.

We restricted our analyses to GBLUPs for four reasons. First, GBLUP allowed to calculate (i) an approximation $\widehat{r}_a$ of the expectation of the prediction accuracy r_a for hybrid performance (Eq. (14)) and GCA and SCA effects (Eq. (13)) based on population parameters, and (ii) an approximation $\widetilde{r}_a$ of r_a as function of r_a estimates for GCA and SCA effects (Eq. (16)). Second, BLUP is relatively simple to compute and under multivariate normality, it is the conditional mean, which has well-known optimality properties for selection based on predicted values (Fernando and Gianola 1986). Third, GBLUP proved to be competitive in comparison with other parametric and nonparametric methods for prediction targeting a single population (Heslot et al. 2012; Crossa et al. 2013) or a hybrid population (Kadam and Lorenz 2019). Fourth, GBLUP can be easily adopted to cope with genotype × environment interactions (Ferrão et al. 2020) and epistasis by using appropriate Gaussian kernels based on ordinary genomic relationship matrices (Jiang and Reif 2015).

Prediction accuracy of hybrids, GCA and SCA effects

By assuming absence of covariances between GBLUPs of GCA and SCA effects, we were able to derive $\widetilde{r}_a$ in Eq. (16) as an approximation of r_a for hybrid performance, which depends on the r_a of GCA and SCA effects and their contribution to σ_G^2. As confirmed by almost identical curves for $\widetilde{r}_a$ and r_a (Figures S4 and S5), Eq. (16) yielded for both data sets an excellent approximation of the prediction accuracy of hybrids, which provides a key for their interpretation in the light of r_a for GCA and SCA effects.

For all scenarios, we observed striking differences in the composition of the TS maximizing r_a (Figs. 2 and 3). As known from other applications of genomic prediction in breeding (Clark et al. 2012; Riedelsheimer et al. 2013; Auinger et al. 2021), the degree of relationship between the genotypes in the TS and PS has a strong influence on the prediction accuracy. In the case of I1 lines, the cross(es) of each line in the TS can be regarded as member(s) of the virtual family of half-sibs underlying the definition of its GCA. Thus, increasing c results for each I1 line in more hybrid relatives in the TS, which is expected to improve r_a of its own GCA and that of I0 lines related to it. However, there will be fewer of such I1 lines that can contribute to r_a of I0 lines. Therefore, one must be cautious in generalizing that $c = 1$ is always the best choice as suggested by our findings.

Prediction of SCA was not promising for all scenarios in both data sets, because r_a was generally too low for all sets of hybrids (Figure S2 and S5). Even in the most favorable case ($h^2 = 0.8$, $\tau_{\mathrm{SCA}} = 22\%$, H2 hybrids), a TS with $n_{\mathrm{TS}} = 144$ and $c = 4$ was needed in data set DS2 to achieve $r_a \sim 0.5$. However, this result may depend on the large genetic distance among the parent populations in our study and might differ, if no clearly defined heterotic groups are available as applies to autogamous crops at the beginning of hybrid breeding, where τ_{SCA} can exceed 25%.

Combining the statements of the previous sections provides an explanation why the curves for prediction accuracy of H0 and H2 hybrids were almost identical to those for GCA of I0 and I1 lines, respectively, and the curves of H1 hybrids are in between those for H0 and H2 hybrids. Firstly, $\widetilde{r}_a$ is in close agreement with r_a for all types of hybrids and all scenarios in both data sets (Figs. 2 and 3), indicating that Eq. (16) provides a solid basis for assessing the importance of GCA and SCA effects in hybrid prediction. Secondly, the contribution of SCA effects to $\widetilde{r}_a$ is close to zero, because τ_{SCA} and even more so r_a^2 for SCA effects are minor in comparison with the contribution of GCA effects. Thus, it follows that prediction accuracy of hybrids depends almost exclusively on r_a of GCA effects.

The steeper increase in r_a values for GCA of I0 lines and the higher level of the curves for data set DS2 than DS1 (Figs. 2 and 3) can be explained by differences in the structure of their parent populations. Data set DS1 included some related lines and displayed a rapid decay of LD between adjacent loci (cf. Technow et al. 2014). Consequently, pedigree relationships captured by markers were presumably an important driver of prediction accuracy. By comparison, because the DH lines of data set DS2 had undergone only one generation of genetic recombination, large haploblocks were present in each parent population as reflected by the large number of completely linked markers over long physical distances. Therefore, co-segregation of QTL and markers and high LD among them were most likely the main reasons for reaching very high levels of prediction accuracy even with moderate TS size (Habier et al. 2013; Schopp et al. 2017).

Importance of additional factors influencing prediction accuracy

For a given number of field plots available for phenotyping, the breeder has in addition to the choice of c the option to increase N_{TS} at the expense of evaluating the TS in fewer environments. However, in all scenarios doubling N_{TS} had generally a much smaller effect on increasing r_a than doubling h^2, which increased prediction accuracy for all sets of hybrids by ~ 30% (Figs. 2 and 3). High h^2, which of course would need more than doubling the number of test environments, was particularly important for reliable prediction of H0 hybrids. Furthermore, increasing N_{TS} was more rewarding under high than low h^2, indicating that low h^2 can only partly be compensated by larger N_{TS}.

Increasing n_{TS} beyond 60 resulted in a rapidly diminishing increase in r_a except for H0 hybrids and this held true for all values of c. For data set DS2, this can be explained by the large haplotypes in the parent populations similar to the results obtained with testcrosses of lines from bi-parental populations (Lehermeier et al. 2014; Lian et al. 2014). However, for data set DS1 this was unexpected and most likely attributable to its breeding history, because in genomic prediction with testcross data (Albrecht et al. 2014; Krchov and Bernardo 2015; Auinger et al. 2021), r_a approached a plateau at much larger TS sizes. Consequently, regarding the optimum allocation of resources assigned to the TS and PS, one reaches soon the point, where a further increase in n_{TS} and/or c hardly pays off in a higher prediction accuracy.

Larger contributions ($\tau_{\mathrm{SCA}} = 22\%vs.1\%$) of SCA to σ_G^2 of hybrids reduced r_a for all scenarios and types of hybrids by less than 10% (Figs. 2 and 3). Hence, the decision on the optimum design of the TS is largely independent of the degree of heterosis in trait expression. This conclusion can be extended to the complexity of the trait as confirmed by simulations with smaller numbers of QTL (data not shown).

Optimum design of the TS for genomic prediction of hybrids and GCA

Optimal implementation of GP in breeding programs requires a balanced compromise between the expenditures spent on the TS and PS (Riedelsheimer and Melchinger 2013) whereby the former influences mainly the prediction accuracy and the latter the selection intensity. Fortunately, the optimal design of the TS for prediction of hybrids and for prediction of GCA effects coincides due to almost perfect congruency of their curves for r_a shown in Figs. 2 and 3. A further important finding was that the heritability and genetic trait architecture had essentially no influence, because for a given N_{TS} and type of hybrids, the ranking of r_a values was identical, independent of h^2 and τ_{SCA}.

Nevertheless, the task of finding the optimal TS in hybrid breeding is complicated by the fact that one has to deal with three types (H0, H1, H2) of hybrids differing in their prediction accuracy. For $c = 1$, the prediction accuracy was highest for H0 hybrids but lowest for H1 and H2 hybrids. Moreover, depending on the design of the TS, the composition of the PS will also change. If for example in DS1 and $N_{\mathrm{TS}} = 96$, c is increased from 1 to 4 and consequently n_{TS} is reduced by ¼, the number of H2 hybrids will be reduced by ~ 1/16 whereas the number of H0 hybrids will grow from 4.6 to 65.4%. Hence, in most cases H2 and HT hybrids have by far the lowest proportion in H as the size of these sets depends on n_{TS}, which is generally small due the limited size of the TS as a result of the high costs of phenotyping. By comparison, H0 hybrids will represent by far the largest subset in H as the cost of producing and genotyping a large number of lines in each parent population is rather inexpensive with the

use of modern technologies such as doubled-haploid production and genotyping by sequencing, respectively. Thus, in most cases it will be most advantageous to have $c = 1$ for designing the TS as confirmed by the prediction accuracy calculated across the entire set H for $N_{TS} = 96$ in data set DS1 (Fig. 2) and $N_{TS} = 96$ and 144 in DS2 (Fig. 3). This conclusion is in line with the results of the experimental study of Lorenzi et al. (2022) demonstrating the potential of sparse factorial designs for genomic prediction in hybrid breeding.

An exact solution of the optimization problem would require calculating the selection gain under truncation selection in multiple populations with different prediction accuracies and costs for H0, H1 and H2 hybrids, which is beyond the scope of this study. Obtaining reliable estimates of r_a for each set of hybrids by cross-validation would require a large TS, exceeding by far the capacity of most breeding programs. As a practicable alternative, one might consider to determine $\hat{r}_a$ for each type of hybrid, which can be calculated from genomic data and estimates of the relevant variance components that could be borrowed from previous breeding cycles. Using $\hat{r}_a$ for optimizing the allocation of resources would also offer flexibility with regard to the choice of n_{TS}, c and h^2 in order to balance the sample size N_{TS} versus the number of test environments used in phenotyping the TS for a given total number of test plots.

Besides genetic and economic aspects for the optimum design of the TS when adopting the "factorial" approach, breeders might prefer to use $c = 1$ for practical reasons. First, fewer seeds are required from each parent, which may obviate the necessity of seed multiplication and thereby the loss of one generation, if seed multiplication is a problem, as applies often to the production of doubled haploids. Second, nicking in flowering of the female and male parents is more likely to be successful for a single pair than if two or more crosses are to be produced per parent in a partially balanced incomplete factorial design. If chemical agents or partitions are used for producing seed for testing purposes, only a corresponding spatial arrangement of the female and male genotypes is required.

In our simulations, the parents of the TS were randomly sampled from the set F and M of all females and males, respectively. In practice, however, the breeder has the option of a target sampling. Fristche-Neto et al. (2018) used the prediction error variance of the hybrids in the PS as selection criterion. They found a slightly higher prediction accuracy for selected over randomly chosen genotypes, when applying an additive model to factorial crosses, but this trend was reversed, when dominance effects were included. Kadam et al. (2021) also used an additive model in combination with the CDmean criterion of Rincent et al. (2012) applied to the hybrid population. With regard to applications of the GCA–SCA model, we suggest to perform the selection separately in each parent population, using search algorithms and criteria recently described for a single population (Isidro y Sánchez et al. 2022; Rio et al. 2022), but to replace the genetic variance in the formulas by the GCA variance of the respective parent population. The TS for hybrid prediction would then be produced by randomly crossing the lines selected from each parent population. In the presence of population structure, this method is expected to improve the prediction accuracy but further research is warranted to assess its merits.

Topics for further research

In established hybrid breeding programs, the majority of activities deals with recycling breeding, where numerous (un-)related bi-parental families are produced anew in each parent population and cycle. As demonstrated with experimental data (Lehermeier et al. 2014) and simulations (Brauner et al. 2019), a genomic model trained with ~ 10 half-sib families of ~ 100 DH lines can yield the same prediction accuracy for testcross prediction of DH lines within a bi-parental family as if trained with ~ 100 full sibs from the same family. Hence, one might argue that including numerous related families of inter-population hybrids in the TS might also be possible for hybrid prediction. Seye et al. (2020) used this approach and obtained for all types of hybrids very high prediction accuracies that were calculated by pooling all 36 hybrid families. However, it remains unknown to what extent these results apply to genomic prediction of hybrids within families, which is the only way to exploit the within family variance and therefore of main interest to the breeder. When including members of all families in the PS, the variation among family means explains one third of the total additive genetic variance among the hybrids so that predicting their performance by their family mean would yield already a prediction accuracy of $\sqrt{1/3} = 0.58$. A direct transfer of the results for genomic prediction across families for testcross performance to that for hybrid performance seems problematic because in the former case the gametic array of the tester is identical for all candidates, whereas in the latter case the gametes of the hybrids in the TS would mostly be sampled from different families. Thus, further research with well-designed experiments is warranted to examine the variation in the prediction accuracy of genomic prediction of hybrids between and across hybrid families similar to the study of Lehermeier et al. (2014) on testcross performance.

Our theoretical results were presented for homozygous lines but they also apply to heterozygous parents. Compared with homozygous lines, using heterozygous S_0 plants as parents reduces the GCA variances by ½ and the SCA variance by ¼ so that τ_{sca} is expected to be very small even for highly heterotic traits. We therefore conjecture that the optimum design of the TS for this scenario will not fundamentally

differ from that for homozygous parents. Confirmation of this hypothesis with simulations would be important for breeding of crops such as oil palm, where hybrids are produced from crosses between heterozygous parents and genomic selection holds great promise due to the drastic reduction of the generation interval (Cros et al. 2015; Kwong et al. 2017).

Further progress for genomic prediction is expected from machine learning methods (Yin et al. 2020) because they do not require assumptions about the distribution of marker effects or their statistical independence and are particularly suited to recognize and exploit patterns in the data regarding the action and interaction of alleles and similarities of genotypes. However, they need to be adapted to the special features of hybrid populations. Thus, these methods may further improve the already high prediction accuracy achieved by GBLUP, but again we do not expect that this will fundamentally change the optimum design of the TS in hybrid breeding unless the parent population have a strong population structure or other type of pattern.

Conclusions

Genomic prediction holds great promise to cope with the huge number of potential hybrids enabled by recent progress in the development of new lines and high-throughput genotyping. Our simulations clearly show that the prediction accuracy of hybrids depends largely on the GCA of their parent lines. Therefore, optimizing the design of the TS for fixed N_{TS} goes hand in hand with the prediction of hybrids in product development and the prediction of GCA of the parent lines in recurrent selection.

We found that including one cross per parent line ($c = 1$) in the TS yields highest prediction accuracy for H0 hybrids and GCA of I0 lines but lowest prediction accuracy for H2 hybrids and GCA of I1 lines, with H1 hybrids taking an intermediate position. The optimal design of the TS is therefore complicated by these opposite trends and depends heavily on the fraction of H0, H1 and H2 hybrids and I0 and I1 lines in the entire sets *H, F* and *M*. A solution for this problem is calculating the expected selection response across all sets of hybrids and lines, taking into consideration the costs of genotyping versus phenotyping and other relevant aspects. However, if the bulk of predicted genotypes are H0 hybrids due to the low costs of genotyping, $c = 1$ seems to be generally the best option for constructing the TS as confirmed by our results on the prediction accuracy across all types of hybrids in set *H* in both data sets. With regard to the entailed paradigm shift in hybrid breeding from the "testcross" to the "factorial" approach, we recommend to complement our simulations with investigations based on experimental data and examine also the potential influence of epistasis and genotype × environment interactions on the optimum design of the TS.

Appendix 1

Derivation of the parametric estimate $\hat{r}_a$ of prediction accuracy for hybrids and their GCA and SCA effects

The expectation of the prediction accuracy $r_a(\hat{\boldsymbol{u}}_\Phi)$ in Eq. (11) is

$$E\left[r\left(\hat{\boldsymbol{u}}_\Phi, \boldsymbol{u}_\Phi\right)\right] = E\left[\frac{\hat{\boldsymbol{u}}^T \boldsymbol{S}_\Phi \boldsymbol{u}}{\sqrt{\left(\hat{\boldsymbol{u}}^T \boldsymbol{S}_\Phi \hat{\boldsymbol{u}}\right)\left(\boldsymbol{u}^T \boldsymbol{S}_\Phi \boldsymbol{u}\right)}}\right] \tag{17}$$

Following Ould Estaghvirou et al. (2013), we approximate the expectation of the ratio and product of random variables by the ratio and product of their expectations

$$E\left[r\left(\hat{\boldsymbol{u}}_\Phi, \boldsymbol{u}_\Phi\right)\right] \approx \frac{E\left[\hat{\boldsymbol{u}}^T \boldsymbol{S}_\Phi \boldsymbol{u}\right]}{\sqrt{E\left[\hat{\boldsymbol{u}}^T \boldsymbol{S}_\Phi \hat{\boldsymbol{u}}\right] E\left[\hat{\boldsymbol{u}}^T \boldsymbol{S}_\Phi \hat{\boldsymbol{u}}\right]}} \tag{18}$$

Using results about (i) the expectation of bilinear forms (Searle 1971, p.65), and (ii) $\mathrm{cov}(\hat{\boldsymbol{u}}, \boldsymbol{u}) = \mathrm{var}(\hat{\boldsymbol{u}})$ (Henderson 1975, p.425), we have

$$E\left[\hat{\boldsymbol{u}}^T \boldsymbol{S}_\Phi \boldsymbol{u}\right] = tr\left(\boldsymbol{S}_\Phi \mathrm{cov}(\hat{\boldsymbol{u}}, \boldsymbol{u})\right) = \mathrm{tr}\left(\boldsymbol{S}_\Phi \mathrm{var}(\hat{\boldsymbol{u}})\right) = \mathrm{tr}\left(\boldsymbol{S}_\Phi \boldsymbol{L}\right) \tag{19}$$

$$E\left[\hat{\boldsymbol{u}}^T \boldsymbol{S}_\Phi \hat{\boldsymbol{u}}\right] = \mathrm{tr}\left(\boldsymbol{S}_\Phi \mathrm{cov}(\hat{\boldsymbol{u}}, \hat{\boldsymbol{u}})\right) = \mathrm{tr}\left(\boldsymbol{S}_\Phi \mathrm{var}(\hat{\boldsymbol{u}})\right) = \mathrm{tr}\left(\boldsymbol{S}_\Phi \boldsymbol{L}\right) \tag{20}$$

and

$$E\left[\boldsymbol{u}^T \boldsymbol{S}_\Phi \boldsymbol{u}\right] = \mathrm{tr}\left(\boldsymbol{S}_\Phi \mathrm{cov}(\boldsymbol{u}, \boldsymbol{u})\right) = \mathrm{tr}\left(\boldsymbol{S}_\Phi \boldsymbol{G}\right) \tag{21}$$

with $\boldsymbol{L} = \boldsymbol{B}\boldsymbol{V}\boldsymbol{B}^T$ (see Eq. (6)). Inserting these expressions in Eq. (18), we obtain $\hat{r}_a(\hat{\boldsymbol{u}}_\Phi)$ as an estimate of $r_a(\hat{\boldsymbol{u}}_\Phi)$

$$\hat{r}_a\left(\hat{\boldsymbol{u}}_\Phi\right) = \sqrt{\frac{\mathrm{tr}\left(\boldsymbol{S}_\Phi \boldsymbol{L}\right)}{\mathrm{tr}\left(\boldsymbol{S}_\Phi \boldsymbol{G}\right)}} = \sqrt{\left\{\sum_{i\in\Phi}\left(l_{ii} - l_{i.}\right)\right\} \Big/ \sum_{i\in\Phi}\left(g_{ii} - g_{i.}\right)} \tag{22}$$

where $\boldsymbol{G} = \left(g_{ij}\right)$, $\boldsymbol{L} = \left(l_{ij}\right)$, $g_{i.} = \frac{1}{|\Phi|}\sum_{j\in\Phi} g_{ij}$ and $l_{i.} = \frac{1}{|\Phi|}\sum_{j\in\Phi} l_{ij}$.

An estimate of r_a for a subset $\Phi \subset \left\{1, \ldots, N_F \times N_M\right\}$ of hybrids is obtained as

$$\hat{r}_a\left(\hat{\boldsymbol{h}}_\Phi\right) = \frac{E\left(\hat{\boldsymbol{h}}^T \boldsymbol{S}_\Phi \boldsymbol{h}\right)}{\sqrt{E\left(\hat{\boldsymbol{h}}^T \boldsymbol{S}_\Phi \hat{\boldsymbol{h}}\right) E\left(\boldsymbol{h}^T \boldsymbol{S}_\Phi \boldsymbol{h}\right)}} = \sqrt{\frac{\mathrm{tr}\left(\boldsymbol{S}_\Phi \boldsymbol{W}\boldsymbol{L}\boldsymbol{W}^T\right)}{\mathrm{tr}\left(\boldsymbol{S}_\Phi \boldsymbol{W}\boldsymbol{G}\boldsymbol{W}^T\right)}} \tag{23}$$

Note that matrices $\boldsymbol{G}$, $\boldsymbol{B}$ and $\boldsymbol{L}$ can be calculated from the genomic data of the genotypes in set F and set M (i.e., matrices $\boldsymbol{K}_F$ and $\boldsymbol{K}_M$), variance components $\sigma^2_{\mathrm{gca}F}$, $\sigma^2_{\mathrm{gca}M}$, σ^2_{sca}, σ^2_e, and the incidence matrices $\boldsymbol{X}$ (usually $\boldsymbol{X} = 1$) and $\boldsymbol{Z}$ pertaining to the fixed and random effects in Eq. (2), respectively.

Appendix 2

Linking the prediction accuracies of hybrids to those of their SCA and parental GCA effects and derivation of $\breve{r}_a$ for estimating the prediction accuracy of hybrids

Henderson (1975) proved for BLUP that $\mathrm{cov}(\boldsymbol{u}, \hat{\boldsymbol{u}}) = \mathrm{var}(\hat{\boldsymbol{u}})$. Thus, the expected (individual) prediction accuracy of a randomly chosen hybrid $i \times j$ can be obtained by extracting the corresponding diagonal elements from the matrices $\mathrm{var}\left(\hat{\boldsymbol{h}}\right)$ and $\mathrm{var}(\boldsymbol{h})$

$$\rho_a\left(\hat{h}_{i\times j}, h_{i\times j}\right) = \frac{\mathrm{cov}\left(\hat{h}_{i\times j}, h_{i\times j}\right)}{\sqrt{\mathrm{var}\left(\hat{h}_{i\times j}\right)\mathrm{var}\left(h_{i\times j}\right)}} = \sqrt{\frac{\mathrm{var}\left(\hat{h}_{i\times j}\right)}{\mathrm{var}\left(h_{i\times j}\right)}} \tag{24}$$

Ignoring covariances among BLUPs of GCA and SCA effects, we get from Eq. (10)

$$\mathrm{var}\left(\hat{\boldsymbol{h}}\right) \approx \mathrm{var}\left(\hat{\boldsymbol{g}}_F\right) \otimes \boldsymbol{J}_{N_M} + \boldsymbol{J}_{N_F} \otimes \mathrm{var}\left(\hat{\boldsymbol{g}}_M\right) + \mathrm{var}\left(\hat{\boldsymbol{s}}_H\right) \tag{25}$$

and using $\mathrm{var}(\boldsymbol{h}) = \boldsymbol{W}^T\boldsymbol{G}\boldsymbol{W} = \boldsymbol{G}_F \otimes \boldsymbol{J}_{n_M} + \boldsymbol{J}_{n_F} \otimes \boldsymbol{G}_M + \boldsymbol{G}_H$, we get the approximation

$$\rho_a\left(\hat{h}_{i\times j}, h_{i\times j}\right) \approx \sqrt{\frac{\mathrm{var}\left(\hat{g}_{F,i}\right) + \mathrm{var}\left(\hat{g}_{M,j}\right) + \mathrm{var}\left(\hat{s}_{i\times j}\right)}{\mathrm{var}\left(h_{i\times j}\right)}} = \sqrt{\rho_a^2\left(\hat{g}_{F,i}, g_{F,i}\right)\tau_{\mathrm{gca}F} + \rho_a^2\left(\hat{g}_{M,j}, g_{M,j}\right)\tau_{\mathrm{gca}M} + \rho_a^2\left(s_{i\times j}, \hat{s}_{i\times j}\right)\tau_{\mathrm{sca}}} \tag{26}$$

where $\tau_{\mathrm{gca}F} = \sigma^2_{\mathrm{gca}F}/\sigma^2_G$, $\tau_{\mathrm{gca}M} = \sigma^2_{\mathrm{gca}M}/\sigma^2_G$ and $\tau_{\mathrm{sca}} = \sigma^2_{\mathrm{sca}}/\sigma^2_G$.

Following Ould Estaghvirou et al. (2013), an estimate $\breve{r}_a(\Phi)$ of the prediction accuracy $r_a(\Phi)$ of the GBLUPs for the genotypes in a population Φ can be obtained from the (individual) prediction accuracies $\rho_a(k)$ of $k \in \Phi$ as

$$\breve{r}_a(\Phi) = \sqrt{\frac{1}{|\Phi|}\sum_{k\in\Phi}\rho_a^2(k)} \tag{27}$$

In our setting, $\Phi \subset H = \{(1,1), ..., (1, n_M), (2,1), ..., (2, n_M), (n_F, 1), ..., (n_F, n_M)\}$ describes a subset of hybrids. Regarding the prediction accuracy of GBLUPs for H0 hybrids, we have $\Phi_{0,0} = F_0 \times M_0$; for H1F and H1M hybrids, we have $\Phi_{1,0} = F_1 \times M_0$ and $\Phi_{0,1} = F_0 \times M_1$, respectively; and for H2 hybrids we have $\Phi_{1,1} = (F_1 \times M_1)\backslash HT$. Thus, we obtain for H0 and H1 hybrids.

$$\breve{r}_a\left(\hat{\boldsymbol{h}}_{\Phi_{s,t}}\right) = \sqrt{\frac{1}{|F_a| \cdot |M_b|}\sum_{(i,j)\in\Phi_{s,t}}\rho_a^2\left(\hat{h}_{i\times j}, h_{i\times j}\right)}$$

.

Together with Eq. (26), we get

$$\breve{r}_a\left(\hat{\boldsymbol{h}}_{\Phi_{s,t}}\right) \approx \sqrt{\frac{1}{|F_s|}\sum_{i\in F_s}\rho_a^2\left(\hat{g}_{F,i}, g_{F,i}\right)\tau_{\mathrm{gca}F} + \frac{1}{|M_s|}\sum_{i\in M_s}\rho_a^2\left(\hat{g}_{M,j}, g_{M,j}\right)\tau_{\mathrm{gca}M} + \frac{1}{|F_s| \cdot |M_t|}\sum_{(i,j)\in F_a\times M_b}\rho_a^2\left(\hat{s}_{i\times j}, s_{i\times j}\right)\tau_{\mathrm{sca}}}$$

From Eq. (27), we have $r_a\left(\hat{\boldsymbol{g}}_{F,F_s}\right) \approx \breve{r}_a\left(\hat{\boldsymbol{g}}_{F,F_s}\right) = \sqrt{\frac{1}{|F_s|}\sum_{i\in F_s}\rho_a^2\left(\hat{g}_{F,i}, g_{F,i}\right)}$,

$r_a\left(\hat{\boldsymbol{g}}_{M,M_t}\right) \approx \breve{r}_a\left(\hat{\boldsymbol{g}}_{M,M_t}\right) = \sqrt{\frac{1}{|M_t|}\sum_{j\in M_t}\rho_a^2\left(\hat{g}_{M,M_t}, g_{M,M_t}\right)}$

and $r_a\left(\hat{s}_{F_s\times M_t}\right) \approx \breve{r}_a\left(\hat{s}_{F_s\times M_t}\right) = \sqrt{\frac{1}{|F_s|\cdot|M_t|}\sum_{(i,j)\in F_s\times M_t}\rho_a^2\left(\hat{s}_{i\times j}, s_{i\times j}\right)}$.

Combining these results, we get as an approximation for $r_a\left(\hat{\boldsymbol{h}}_{F_s\times M_t}\right) \approx \breve{r}_a\left(\hat{\boldsymbol{h}}_{\Phi_{s,t}}\right)$

$$r_a\left(\hat{\boldsymbol{h}}_{F_s\times M_t}\right) \approx \sqrt{r_a^2\left(\hat{\boldsymbol{g}}_{F,F_s}\right)\tau_{\mathrm{gca}F} + r_a^2\left(\hat{\boldsymbol{g}}_{M,M_t}\right)\tau_{\mathrm{gca}M} + r_a^2\left(\hat{\boldsymbol{s}}_{F_s\times M_t}\right)\tau_{\mathrm{sca}}} \tag{28}$$

Substituting $r_a\left(\hat{\boldsymbol{g}}_{F,F_s}\right)$, $\tau_{\text{gca}F}$ and the other terms on the right hand side by estimates such as $\hat{r}_a\left(\hat{\boldsymbol{g}}_{F,F_s}\right)$ etc. yields the estimate $\tilde{r}_a\left(\hat{\boldsymbol{h}}_{F_s\times M_t}\right)$ for $r_a\left(\hat{\boldsymbol{h}}_{F_s\times M_t}\right)$:

$$\tilde{r}_a\left(\hat{\boldsymbol{h}}_{F_s\times M_t}\right)=\sqrt{\hat{r}_a^2\left(\hat{\boldsymbol{g}}_{F,F_s}\right)\tau_{\text{gca}F}+\hat{r}_a^2\left(\hat{\boldsymbol{g}}_{M,M_t}\right)\tau_{\text{gca}M}+\hat{r}_a^2\left(\hat{\boldsymbol{s}}_{F_s\times M_t}\right)\tau_{\text{sca}}} \tag{29}$$

which was used for preparing Figures S4 and S6.

If the TS is constructed according to a balanced incomplete factorial design with c crosses per parent line in F_1 and M_1 and $|F_1|=|M_1|$, using Eqs. (26 and 27) we have for H2 hybrids

$$\tilde{r}_a\left(\hat{\boldsymbol{h}}_{\Phi_{1,1}}\right)=\sqrt{\frac{1}{|F_1|\cdot(|M_1|-c)}\sum_{(i,j)\in\Phi_{1,1}}\rho_a^2\left(\hat{h}_{i\times j},h_{i\times j}\right)}\approx\sqrt{\frac{1}{|F_1|}\sum_{i\in F_1}\rho_a^2\left(\hat{g}_{F,i},g_{F,i}\right)\tau_{\text{gca}F}+\frac{1}{|M_1|}\sum_{j\in M_1}\rho_a^2\left(\hat{g}_{M,j},g_{M,j}\right)\tau_{\text{gca}M}+\frac{1}{|F_1|\cdot(|M_1|-c)}\sum_{(i,j)\in\Phi_{1,1}}\rho_a^2\left(\hat{s}_{i\times j},s_{i\times j}\right)\tau_{\text{sca}}}$$

Using again Eq. (28), we get

$$\tilde{r}_a\left(\hat{\boldsymbol{h}}_{\Phi_{1,1}}\right)\approx\sqrt{r_a^2\left(\hat{\boldsymbol{g}}_{F,F_1}\right)\tau_{\text{gca}F}+r_a^2\left(\hat{\boldsymbol{g}}_{M,M_1}\right)\tau_{\text{gca}M}+r_a^2\left(\hat{\boldsymbol{s}}_{\Phi_{1,1}}\right)\tau_{\text{sca}}} \tag{30}$$

as an approximation of $r_a\left(\hat{\boldsymbol{h}}_{\Phi_{1,1}}\right)$.

In summary, Eqs. (29 and 30) allow to estimate $\tilde{r}_a$ for the prediction accuracy r_a of H0 hybrids $\left(\Phi_{0,0}=F_0\times M_0\right)$, H1 hybrids $(\Phi_{1,0}=F_1\times M_0$ or $\Phi_{0,1}=F_0\times M_1)$ and H2 hybrids $\left(\Phi_{1,1}=\left(F_1\times M_1\right)\backslash HT\right)$ by substituting appropriate r_a estimates of their I0 and/or I1 parent lines and r_a estimates of the SCA effects of their hybrids.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00122-023-04413-y.

Acknowledgements The authors are indebted to Tobias Lanzl and Yan-Chen Lin for preparing data set DS1 and DS2, respectively, for this study and to Matthias Frisch for critical reading of the manuscript and valuable suggestions.

Author contributions AEM conceived and designed the study. AEM, RF and CS developed the theory. HJA, RF and CS developed the software. HJA executed all computations and prepared the graphs. AEM, HJA, RF, CS and CCS interpreted the results, wrote the manuscript and approved the final version.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was funded by intra-mural funds of the Technical University of Munich.

Declarations

Conflict of interest The authors declare that they have no conflict of interest. AEM is editor-in-chief and CCS is member of the editorial board of Theor. Appl. Genetics.

Ethical approval The authors declare that their work complies with the current laws of Germany.

References

Albrecht T, Wimmer V, Auinger H-J, Erbe M, Knaak C, Ouzunova M, Simianer H, Schön C-C (2011) Genome-based prediction of testcross values in maize. Theor Appl Genet 123:339–350

Albrecht T, Auinger H-J, Wimmer V, Ogutu JO, Knaak C, Ouzunova M, Piepho H-P, Schön C-C (2014) Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. Theor Appl Genet 127:1375–1386

Auinger H-J, Lehermeier C, Gianola D, Mayer M, Melchinger AE, da Silva S, Knaak C, Ouzunova M, Schön C-C (2021) Calibration and validation of predicted genomic breeding values in an advanced cycle maize population. Theor Appl Genet 134:3069–3081

Bernardo R (1996) Best linear unbiased prediction of maize single-cross performance. Crop Sci 36:50–56

Bezanson J, Edelman A, Karpinski S, Shah VB (2017) Julia: a fresh approach to numerical computing. SIAM Rev 59:65–98

Brauner PC, Schipprack W, Utz HF, Bauer E, Mayer M, Schön C-C, Melchinger AE (2019) Testcross performance of doubled haploid lines from European flint maize landraces is promising for broadening the genetic base of elite germplasm. Theor Appl Genet 132:1897–1908

Browning BL, Browning SR (2016) Genotype imputation with millions of reference samples. Am J Hum Genet 98:116–126

Chaikam V, Molenaar W, Melchinger AE, Boddupalli PM (2019) Doubled haploid technology for line development in maize: technical advances and prospects. Theor Appl Genet 132:3227–3243

Clark SA, Hickey JM, Daetwyler HD, van der Werf JH (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 44:1–9

Cockerham CC (1961) Implications of genetic variances in a hybrid breeding program 1. Crop Sci 1:47–52

Cros D, Denis M, Sánchez L, Cochard B, Flori A, Durand-Gasselin T, Nouy B, Omoré A, Pomiès V, Riou V (2015) Genomic selection

prediction accuracy in a perennial crop: case study of oil palm (Elaeis guineensis Jacq.). Theor Appl Genet 128:397–410
Crossa J, Beyene Y, Kassa S, Pérez P, Hickey JM, Chen C, De Los CG, Burgueño J, Windhausen VS, Buckler E (2013) Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3: Genes, Genom, Genet 3:1903–1926
Dekkers J (2007) Prediction of response to marker-assisted and genomic selection using selection index theory. J Anim Breed Genet 124:331–341
Esfandyari H, Sørensen AC, Bijma P (2015) A crossbred reference population can improve the response to genomic selection for crossbred performance. Genet Sel Evol 47:1–12
Fernando R, Gianola D (1986) Optimal properties of the conditional mean as a selection criterion. Theor Appl Genet 72:822–825
Ferrão LFV, Marinho CD, Munoz PR, Resende MF Jr (2020) Improvement of predictive ability in maize hybrids by including dominance effects and marker× environment models. Crop Sci 60:666–677
Fisher R (1918) The correlation between relatives on the supposition of mendelian inheritance. Earth Environ Sci Trans R Soc Edinb 52(2):399–433
Frascaroli E, Cane MA, Landi P, Pea G, Gianfranceschi L, Villa M, Morgante M, Pe ME (2007) Classical genetic and quantitative trait loci analyses of heterosis in a maize hybrid between two elite inbred lines. Genetics 176:625–644
Fristche-Neto R, Akdemir D, Jannink J-L (2018) Accuracy of genomic selection to predict maize single-crosses obtained through different mating designs. Theor Appl Genet 131:1153–1162
Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner E-M, Hansen M, Joets J (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
Garcia AAF, Wang S, Melchinger AE, Zeng Z-B (2008) Quantitative trait loci mapping and the genetic basis of heterosis in maize and rice. Genetics 180:1707–1724
Guo T, Yang N, Tong H, Pan Q, Yang X, Tang J, Wang J, Li J, Yan J (2014) Genetic basis of grain yield heterosis in an "immortalized F2" maize population. Theor Appl Genet 127:2149–2158
Habier D, Fernando RL, Garrick DJ (2013) Genomic BLUP decoded: a look into the black box of genomic prediction. Genetics 194:597–607
Hallauer AR (1967) Development of single-cross hybrids from two-eared maize populations 1. Crop Sci 7:192–195
Hallauer AR, Carena MJ, Miranda Filho Jd (2010) Quantitative genetics in maize breeding. Springer, Berlin
Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423–447
Henderson C (1984) Estimation of variances and covariances under multiple trait models. J Dairy Sci 67:1581–1589
Heslot N, Yang HP, Sorrells ME, Jannink JL (2012) Genomic selection in plant breeding: a comparison of models. Crop Sci 52:146–160
Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet 4:e1000008
Isidro y Sánchez J, Simon R, Deniz A (2022) Hands on training optimization in genomic selection. In: Elias AA, Goel S (eds) Genomic selection in plants a guide for breeders. CRC Press, Boca Raton, pp 23–50
Jiang Y, Reif JC (2015) Modeling epistasis in genomic selection. Genetics 201:759–768
Kadam DC, Lorenz AJ (2019) Evaluation of nonparametric models for genomic prediction of early-stage single crosses in maize. Crop Sci 59:1411–1423
Kadam DC, Rodriguez OR, Lorenz AJ (2021) Optimization of training sets for genomic prediction of early-stage single crosses in maize. Theor Appl Genet 134:687–699
Krchov LM, Bernardo R (2015) Relative efficiency of genomewide selection for testcross performance of doubled haploid lines in a maize breeding program. Crop Sci 55:2091–2099
Kwong QB, Ong AL, Teh CK, Chew FT, Tammi M, Mayes S, Kulaveerasingam H, Yeoh SH, Harikrishna JA, Appleton DR (2017) Genomic selection in commercial perennial crops: applicability and improvement in oil palm (Elaeis guineensis Jacq.). Sci Rep 7:1–9
Lamkey KR, Schnicker BJ, Melchinger AE (1995) Epistasis in an elite maize hybrid and choice of generation for inbred line development. Crop Sci 35:1272–1281
Lehermeier C, Krämer N, Bauer E, Bauland C, Camisan C, Campo L, Flament P, Melchinger AE, Menz M, Meyer N (2014) Usefulness of multiparental populations of maize (*Zea mays* L.) for genome-based prediction. Genetics 198:3–16
Lian L, Jacobson A, Zhong S, Bernardo R (2014) Genomewide prediction accuracy within 969 maize biparental populations. Crop Sci 54:1514–1522
Lorenzi A, Bauland C, Mary-Huard T, Pin S, Palaffre C, Guillaume C, Lehermeier C, Charcosset A, Moreau L (2022) Genomic prediction of hybrid performance: comparison of the efficiency of factorial and tester designs used as training sets in a multiparental connected reciprocal design for maize silage. Theor Appl Genet 135:3143–3160
Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer, Sunderland, MA
Melchinger AE, Gumber RK (1998) Overview of heterosis and heterotic groups in agronomic crops. Concepts Breed Heterosis Crop Plants 25:29–44
Melchinger A, Schmidt W, Geiger H (1988) Comparison of testcrosses produced from F2 and first backcross populations in maize. Crop Sci 28:743–749
Ould Estaghvirou SB, Ogutu JO, Schulz-Streeck T, Knaak C, Ouzunova M, Gordillo A, Piepho H-P (2013) Evaluation of approaches for estimating the accuracy of genomic prediction in plant breeding. BMC Genom 14:1–21
Reif J, Gumpert F, Fischer S, Melchinger A (2007) Impact of interpopulation divergence on additive and dominance variance in hybrid populations. Genetics 176:1931–1934
Riedelsheimer C, Melchinger AE (2013) Optimizing the allocation of resources for genomic selection in one breeding cycle. Theor Appl Genet 126:2835–2848
Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet 44:217–220
Riedelsheimer C, Endelman JB, Stange M, Sorrells ME, Jannink J-L, Melchinger AE (2013) Genomic predictability of interconnected biparental maize populations. Genetics 194:493–503
Rincent R, Laloë D, Nicolas S, Altmann T, Brunel D, Revilla P, Rodriguez VM, Moreno-Gonzalez J, Melchinger A, Bauer E (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
Rio S, Charcosset A, Mary-Huard T, Moreau L, Rincent R (2022) Building a calibration set for genomic prediction genomic predictions (GP), characteristics to be considered, and optimizationoptimizations approaches. In: Ahmadi N, Bartholomé J (eds) Genomic prediction of complex traits: methods and protocols. Springer, New York, pp 77–112
Sackton TB, Hartl DL (2016) Genotypic context and epistasis in individuals and populations. Cell 166:279–287

Schön CC, Dhillon BS, Utz HF, Melchinger AE (2010) High congruency of QTL positions for heterosis of grain yield in three crosses of maize. Theor Appl Genet 120:321–332

Schopp P, Müller D, Technow F, Melchinger AE (2017) Accuracy of genomic prediction in synthetic populations depending on the number of parents, relatedness, and ancestral linkage disequilibrium. Genetics 205:441–454

Schrag TA, Westhues M, Schipprack W, Seifert F, Thiemann A, Scholten S, Melchinger AE (2018) Beyond genomic prediction: combining different types of omics data can improve prediction of hybrid performance in maize. Genetics 208:1373–1385

Searle SR (1971) Linear models. John Wiley & Sons Inc, New York

Seifert F, Thiemann A, Schrag TA, Rybka D, Melchinger AE, Frisch M, Scholten S (2018) Small RNA-based prediction of hybrid performance in maize. BMC Genom 19:1–14

Seye A, Bauland C, Charcosset A, Moreau L (2020) Revisiting hybrid breeding designs using genomic predictions: simulations highlight the superiority of incomplete factorials between segregating families over topcross designs. Theor Appl Genet 133:1995–2010

Sorensen D, Gianola D (2002) Likelihood, Bayesian and MCMC methods in quantitative genetics. Springer, Berlin

Sprague GF, Tatum LA (1942) General vs. specific combining ability in single crosses of corn. J Am Soc Agrono

Stuber CW, Lincoln SE, Wolff D, Helentjaris T, Lander E (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823–839

Tang J, Yan J, Ma X, Teng W, Wu W, Dai J, Dhillon BS, Melchinger AE, Li J (2010) Dissection of the genetic basis of heterosis in an elite maize hybrid by QTL mapping in an immortalized F2 population. Theor Appl Genet 120:333–340

Technow F, Riedelsheimer C, Schrag TA, Melchinger AE (2012) Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. Theor Appl Genet 125:1181–1194

Technow F, Schrag TA, Schipprack W, Bauer E, Simianer H, Melchinger AE (2014) Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. Genetics 197:1343–1355

Unterseer S, Bauer E, Haberer G, Seidel M, Knaak C, Ouzunova M, Meitinger T, Strom TM, Fries R, Pausch H (2014) A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. BMC Genom 15:1–15

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423

Westhues M, Schrag TA, Heuer C, Thaller G, Utz HF, Schipprack W, Thiemann A, Seifert F, Ehret A, Schlereth A (2017) Omics-based hybrid prediction in maize. Theor Appl Genet 130:1927–1939

Yin L, Zhang H, Zhou X, Yuan X, Zhao S, Li X, Liu X (2020) KAML: improving genomic prediction accuracy of complex traits using machine learning determined parameters. Genome Biol 21:1–22

Zenke-Philippi C, Frisch M, Thiemann A, Seifert F, Schrag T, Melchinger AE, Scholten S, Herzog E (2017) Transcriptome-based prediction of hybrid performance with unbalanced data from a maize breeding programme. Plant Breed 136:331–337

Zhao Y, Li Z, Liu G, Jiang Y, Maurer HP, Würschum T, Mock H-P, Matros A, Ebmeyer E, Schachschneider R (2015) Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. Proc Natl Acad Sci 112:15624–15629