Theor Appl Genet (2011) 123:1–10
DOI 10.1007/s00122-011-1561-4

ORIGINAL PAPER

Best linear unbiased prediction and optimum allocation of test resources in maize breeding with doubled haploids

Xuefei Mi · Thilo Wegenast · H. Friedrich Utz · Baldev S. Dhillon · Albrecht E. Melchinger

Received: 12 May 2010 / Accepted: 21 February 2011 / Published online: 9 March 2011

Abstract With best linear unbiased prediction (BLUP), information from genetically related candidates is combined to obtain more precise estimates of genotypic values of test candidates and thereby increase progress from selection. We developed and applied theory and Monte Carlo simulations implementing BLUP in 2 two-stage maize breeding schemes and various selection strategies. Our objectives were to (1) derive analytical solutions of the mixed model equations under two breeding schemes, (2) determine the optimum allocation of test resources with BLUP under different assumptions regarding the variance component ratios for grain yield in maize, (3) compare the progress from selection using BLUP and conventional phenotypic selection based on mean performance solely of the candidates, and (4) analyze the potential of BLUP for further improving the progress from selection. The breeding schemes involved selection for testcross performance either of DH lines at both stages (DHTC) or of S_1 families at the first stage and DH lines at the second stage (S_1TC-DHTC). Our analytical solutions allowed much faster calculations of the optimum allocations and superseded matrix inversions to solve the mixed model equations. Compared to conventional phenotypic selection, the progress from selection was slightly higher with BLUP for both optimization criteria, namely the selection gain and the probability to select the best genotypes. The optimum allocation of test resources in S_1TC-DHTC involved ≥ 10 test locations at both stages, a low number of crosses (≤ 6) each with 100–300 S_1 families at the first stage, and 500–1,000 DH lines at the second stage. In breeding scheme DHTC, the optimum number of test candidates at the first stage was 5–10 times larger, whereas the number of test locations at the first stage and the number of test candidates at the second stage were strongly reduced compared to S_1TC-DHTC.

Introduction

In commercial maize (*Zea mays* L.) breeding, a very large number of new inbreds and hybrids is generated in each breeding cycle. Selection of the best ones is a challenge and, therefore, should be based on the most reliable criterion for the true performance of genotypes. Furthermore, resources (such as human resources, finances, infrastructure) are limited and, thus, their optimum allocation is of crucial importance for the efficiency and competitiveness of breeding programs. For optimal progress from selection, a plant breeder has to strike a balance among the number of (1) parental crosses, (2) test candidates within each cross (e.g., testcrosses of S_1 families and doubled haploid (DH) lines in the present study), and (3) test environments as well as field replications.

Estimation of the genotypic value is commonly based on the phenotypic value of the tested candidate itself, whereas information on related genotypes is largely neglected. The best linear unbiased prediction (BLUP) method was originally developed to estimate breeding values in animal breeding (cf., Robinson 1991). With BLUP, the prediction

---

Communicated by H. Becker.

X. Mi and T. Wegenast contributed equally to this work.

X. Mi · T. Wegenast · H. F. Utz · B. S. Dhillon ·
A. E. Melchinger (✉)
Institute of Plant Breeding, Seed Science,
and Population Genetics, University of Hohenheim,
70593 Stuttgart, Germany
e-mail: melchinger@uni-hohenheim.de

of breeding values is based on (1) known genetic relationships among genotypes and (2) the available phenotypic data of related genotypes (Bernardo 1996). It also enables to cope with unbalanced data sets. Bernardo (1996) and Piepho et al. (2008) advocated application of the BLUP approach in plant breeding, because it has a good prediction accuracy and promises to enhance the efficiency of selection. Nevertheless, BLUP is rarely used in plant breeding (Piepho et al. 2008).

Line development by the DH technology is increasingly replacing conventional recurrent selfing in maize breeding. It enables extremely rapid development of a large number of lines and testing of their experimental hybrids very early in a breeding cycle. DH lines may be developed from S_0 plants in the source germplasm, as is general practice (Smith et al. 2008), or from S_1 families selected on the basis of their testcross performance (Longin et al. 2007; Bernardo 2009). Longin et al. (2007) and Wegenast et al. (2008, 2010) compared these two breeding schemes considering two-stage selection for testcross performance of (1) DH lines at both stages (DHTC) and (2) S_1 families at the first stage and DH lines at the second stage (S_1TC-DHTC) for optimization of allocation of test resources and progress from selection. Further, Wegenast et al. (2010) evaluated the efficiency of selection assuming equal or variable numbers of DH candidates per cross at one selection stage. The same assumption was also made for S_1 families. In these studies, classical phenotypic selection based solely on simple phenotypic testcross means was used, i.e., without giving weight to information from related genotypes. For variable numbers of candidates, heritabilities differ for each candidate making prediction of the selection gain hardly feasible, thus, simulation studies are used to optimize these scenarios. However, BLUP has so far not been considered for optimizing the selection process.

In BLUP, pedigree information is reflected in the mixed model equations (MME) by the genetic relationship matrices $\boldsymbol{A}$ and $\boldsymbol{R}$, which is calculated by using the coefficient of coancestry among the test candidates. The matrix inversion to solve the MME is generally computation intensive, and, thus, hardly feasible even when breeding programs are of moderate size. To calculate the optimum allocation of resources in a breeding program, the MME need to be solved for every combination of allocation of resources under consideration. This is only feasible if the time-consuming matrix inversions are avoided by using general mathematical solutions of the MME, but no such formulas are available for DHTC and S_1TC-DHTC. Further, to the best of our knowledge, no formula or simulation program is published in the literature to compute the optimum allocation of test resources and estimates of selection gain in hybrid maize breeding.

The objectives of our study were to (1) derive analytical solutions of the mixed model equations in the breeding schemes DHTC and S_1TC-DHTC, (2) determine the optimum allocation of test resources with BLUP under different assumptions regarding variance component ratios for grain yield in maize, (3) compare the progress from selection with BLUP and classical phenotypic selection based on mean performance solely of the candidate, and (4) analyze the potential of BLUP for further improving the progress from selection with both breeding schemes.

Materials and methods

Breeding schemes

The two breeding schemes, DHTC and S_1TC−DHTC considered in this treatise, were described in detail by Wegenast et al. (2010). Briefly, to start a new breeding cycle, parents for generating new base populations were selected on the basis of their performance in previous breeding cycles. For each combination of different variables, an initial number of DH lines or S_1 families were produced from randomly taken S_0 plants in the parental crosses, and the numbers of DH lines or S_1 families were limited by the total resources. In S_1TC-DHTC, selection at the first stage was performed among S_1 families; subsequently, selected S_1 families were used to develop DH lines for evaluation in the second stage of selection. In DHTC, two-stage selection among DH lines was conducted. In both breeding schemes, the best 10 DH lines were selected based on testcross performance. The breeding schemes used in the aforementioned and the present study are identical, except for the selection criterion. Wegenast et al. (2010) considered conventional phenotypic selection based on only mean performance of the candidates themselves, whereas in the present study, selection was based on BLUPs, which were estimated using the data on phenotypic performance of the testcrosses of the candidate itself and its relatives, employing the BLUP approach and MME (see below).

Selection strategies

For each breeding scheme, we compared two selection strategies (Wegenast et al. 2010). In all cases, selection among DH lines and/or S_1 families was based on their testcross performance. In strategy 1, sequential selection was performed among parental crosses and then in two stages among DH lines in DHTC or among parental crosses, S_1 families at the first stage, and DH lines at the second stage in S_1TC-DHTC. The numbers of DH lines (DHTC) or S_1 families within each parental cross and DH lines within each S_1 family (S_1TC-DHTC) were assumed to be uniform. In strategy 2, selection among parental crosses was not considered. Selection was performed among DH lines at both stages in DHTC and among S_1 families at the first stage and among DH lines at the second stage in S_1TC-DHTC. Strategy 2 had three variants: in strategy 2a, the numbers of DH lines or S_1 families were constant as in strategy 1; in strategies 2b and 2c, these numbers varied depending on the parental performance. In DHTC, the number of DH lines within each parental cross in the first stage depended on the rank of the performance of the parental cross. In S_1TC-DHTC, the number of S_1 families and DH lines within S_1 families within parental crosses depended on the testcross performance of the cross. With lower rank, the number of candidate DH lines or S_1 families decreased moderately in strategy 2b and strongly in strategy 2c.

Economic frame and quantitative-genetic parameters

A fixed total budget for the production and evaluation of the test candidates was defined in terms of testcross plot equivalents. The budget and the production costs were calculated as described by Wegenast et al. (2010). The assumed budget was 20,000 testcross plot equivalents in a heterotic pool.

The values of variance components (σ^2_G, $\sigma^2_{G\times y}$, $\sigma^2_{G\times l}$, $\sigma^2_{G\times l\times y}$, σ^2_e) for grain yield were based on the evaluation of DH lines in maize breeding programs in Central Europe (Wegenast et al. 2008), where σ^2_G is the genotypic variance among testcrosses of the candidate lines with a given tester, $\sigma^2_{G\times y}$ the variance of the genotype × year interactions, $\sigma^2_{G\times l}$ the variance of the genotype × location interactions, $\sigma^2_{G\times l\times y}$ the variance of the genotype × location × year interactions, and σ^2_e the variance of the residual error. The index G in the variance component ratios refers to the respective test candidates, i.e., parental crosses (C), DH lines within crosses (D), S_1 families within crosses (F), or DH lines within S_1 families (D/F). Three scenarios of relative ratios of variance components were considered. In scenario VC1, we assumed a variance component ratio $\sigma^2_G{:}\sigma^2_{G\times y}{:}\sigma^2_{G\times l}{:}\sigma^2_{G\times l\times y}{:}\sigma^2_e =$ 0.5:0.125:0.125:0.25:1, respectively, for $C = G$ and $D = G$, and 0.25:0.0625:0.0625:0.125:1, respectively, for $F = G$ and $D/F = G$ (Wegenast et al. 2008). In other scenarios, the contribution of σ^2_G was kept constant, but all other variances were doubled (VC2) and quadrupled (VC3).

Mixed model equations

Analysis of the simulated phenotypic data was based on a mixed linear model (Piepho et al. 2008). Following Bauer et al. (2006), we assumed all effects as random, except the population mean, which was treated as a fixed effect and set equal to 0 as is common practice (Bernardo 2003). All non-genetic effects were combined together into the random error ϵ, following Bernardo (2002, Chap. 10). The target variable was the genotypic value of the testcross performance for grain yield. In the absence of epistasis, the average effect of gene substitution in testcrosses behaves like the additive effect for per se performance (Melchinger 1987; Melchinger et al. 1998) and consequently, the covariance between testcross progeny due to the common tester is 0. The variance components for grain yield were assumed to be known, with values as defined in the previous section.

In breeding scheme DHTC, the linear model used was

$$\mathbf{y} = \mathbf{Z}_C\mathbf{u}_C + \mathbf{Z}_D\mathbf{u}_D + \epsilon \tag{1}$$

where $\mathbf{y}$ is the vector of phenotypic observations, ϵ is the vector of the random error with variance σ^2_ϵ, $\mathbf{u}$ is the vector of random effects, $\mathbf{Z}$ is the design matrix associated to the random effects, the index C refers to crosses, and D to DH lines within crosses. The MME to calculate the BLUPs were

$$\begin{bmatrix} \mathbf{Z}'_C\mathbf{R}^{-1}\mathbf{Z}_C + \sigma_C^{-2}\mathbf{A}_C^{-1} & \mathbf{Z}'_C\mathbf{R}^{-1}\mathbf{Z}_D \\ \mathbf{Z}'_D\mathbf{R}^{-1}\mathbf{Z}_C & \mathbf{Z}'_D\mathbf{R}^{-1}\mathbf{Z}_D + \sigma_D^{-2}\mathbf{A}_D^{-1} \end{bmatrix} \times \begin{bmatrix} \hat{\mathbf{c}} \\ \hat{\mathbf{d}} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}'_C\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'_D\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}, \tag{2}$$

where $\mathbf{A}$ is the genetic relationship matrix associated with the respective random effects, $\mathbf{R}$ is the design matrix associated with the random error, and $\hat{\mathbf{c}}$ and $\hat{\mathbf{d}}$ are the vectors of predicted BLUPs of the crosses and DH lines.

In breeding scheme S_1TC-DHTC, the linear model used at the first stage was

$$\mathbf{y} = \mathbf{Z}_C\mathbf{u}_C + \mathbf{Z}_F\mathbf{u}_F + \epsilon \tag{3}$$

where the index F refers to S_1 families within crosses. The MME to calculate the BLUPs were

$$\begin{bmatrix} \mathbf{Z}'_C\mathbf{R}^{-1}\mathbf{Z}_C + \sigma_C^{-2}\mathbf{A}_C^{-1} & \mathbf{Z}'_C\mathbf{R}^{-1}\mathbf{Z}_F \\ \mathbf{Z}'_F\mathbf{R}^{-1}\mathbf{Z}_C & \mathbf{Z}'_F\mathbf{R}^{-1}\mathbf{Z}_F + \sigma_F^{-2}\mathbf{A}_F^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{c}} \\ \hat{\mathbf{f}} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}'_C\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'_F\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}, \tag{4}$$

where $\hat{\mathbf{f}}$ is the vector of the predicted BLUPs of the S_1 families. The linear model used at second stage of breeding scheme S_1TC-DHTC was

$$\mathbf{y} = \mathbf{Z}_C\mathbf{u}_C + \mathbf{Z}_F\mathbf{u}_F + \mathbf{Z}_{D/F}\mathbf{u}_{D/F} + \epsilon \tag{5}$$

where the index D/F refers to DH lines within S_1 families. The MME to calculate the BLUPs were

$$\begin{bmatrix} \mathbf{Z}'_C\mathbf{R}^{-1}\mathbf{Z}_C + \sigma_C^{-2}\mathbf{A}_C^{-1} & \mathbf{Z}'_C\mathbf{R}^{-1}\mathbf{Z}_F & \mathbf{Z}'_C\mathbf{R}^{-1}\mathbf{Z}_{D/F} \\ \mathbf{Z}'_F\mathbf{R}^{-1}\mathbf{Z}_C & \mathbf{Z}'_F\mathbf{R}^{-1}\mathbf{Z}_F + \sigma_F^{-2}\mathbf{A}_F^{-1} & \mathbf{Z}'_F\mathbf{R}^{-1}\mathbf{Z}_{D/F} \\ \mathbf{Z}'_{D/F}\mathbf{R}^{-1}\mathbf{Z}_C & \mathbf{Z}'_{D/F}\mathbf{R}^{-1}\mathbf{Z}_F & \mathbf{Z}'_C\mathbf{R}^{-1}\mathbf{Z}_{D/F} + \sigma_{D/F}^{-2}\mathbf{A}_{D/F}^{-1} \end{bmatrix}$$

$$\times \begin{bmatrix} \widehat{\mathbf{c}} \\ \widehat{\mathbf{f}} \\ \widehat{\mathbf{d}} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}'_C\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'_F\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'_{D/F}\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}. \tag{6}$$

The variance–covariance matrices **A** are identity matrices. The matrix **R** is not an identity but a full rank matrix with several blocks, each block representing a group of DH lines or S_1 families within the same cross and/or S_1 family and DH lines within S_1 family. As we assumed all progenies within a cross to have a constant covariance with each other, the blocks have identical values in the non-diagonal elements. The lines used for producing the parental crosses were assumed to be unrelated and each line was used in only one cross. Further, the inbreeding coefficient of DH lines was considered to be equal to 1.

Simulation model

We applied Monte Carlo simulations to investigate the selection strategies (1, 2a, 2b, and 2c) in both breeding schemes DHTC and S_1TC-DHTC. Gaussian normal distributions of the genotypic and phenotypic values were assumed for grain yield. Parental cross selection was based on the genotypic values of the crosses, which were taken from a distribution $N(0, \sigma^2{}_C)$. The phenotypic values for testcrosses of (1) DH lines in DHTC-2 and (2) S_1 families and DH lines in S_1TC-DHTC-2 were modeled as described by Wegenast et al. (2010). For faster random number generation, the simulation programs were written in C and implemented in the statistical software R (R Development Core Team 2006), using a method equivalent to the R package mvtnorm. Test locations common at both stages (L_c) were assumed such that $L_c = min(L_1, L_2)$. Without restriction on L_j in stage j, one replication at each test location was considered, as ΔG has been reported to be maximum for one replication per test location for both stages of selection (cf., Bernardo 2002; Melchinger et al. 2005).

Optimum allocation of test resources and optimization criteria

The optimum allocation of test resources for the two optimization criteria, namely selection gain (ΔG) and the probability to identify superior genotypes [$P(q)$], as well as their standard deviations ($\mathrm{SD}_{\Delta G}, \mathrm{SD}_{P(q)}$) were estimated following Wegenast et al. (2010). Thus, $P(q)$ corresponds to the probability that the selected DH lines belong to the fraction of the q% DH lines with the highest genotypic value of testcrosses in the base population. The value of q considered was 0.1%. The allocation of test resources refers to the pairs (N_j, L_j) for both stages, where N_j is the number of test candidates at stage j. It was considered optimum if it maximized the corresponding optimization criterion ΔG or $P(q)$. An asterisk was used to indicate the optimum allocation as well as the corresponding optimization criteria (e.g., $L_1^*, \Delta G^*$). Since the optimization criteria were estimated with a precision of 0.01, the optimum allocation (N_J^*, L_J^*) was determined such that $\mathrm{SD}_{\Delta G}$ or $\mathrm{SD}_{P(q)}$ was minimum among all allocations within a 0.01 drop-off of ΔG or $P(q)$.

Results

Solutions of the mixed model equations

General analytical solutions of the MME were derived for both breeding schemes DHTC and S_1TC-DHTC. The formulas hold true for any variance component ratio and allocation of test resources (numbers of crosses, DH lines, S_1 families, test locations, and replications), provided the parental crosses and parental lines are unrelated and the population mean is equal to zero. In DHTC, the BLUP of the kth cross ($\widehat{c}_k$) and the mth DH line within the kth cross ($\widehat{d}_{km}$) at both selection stages are as follows:

$$\widehat{c}_k = \bar{y}_k \frac{\sigma_C^2}{\sigma_C^2 + \sigma_{\epsilon_C}^2 + \sigma_D^2/n_k + \sigma_{\epsilon_D}^2/n_k} \tag{7}$$

$$\widehat{d}_{km} = \left(y_{km} - \bar{y}_k \frac{\sigma_C^2 + \sigma_{\epsilon_C}^2}{\sigma_C^2 + \sigma_{\epsilon_C}^2 + \sigma_D^2/n_k + \sigma_{\epsilon_D}^2/n_k} \right) \frac{\sigma_D^2}{\sigma_D^2 + \sigma_{\epsilon_D}^2}, \tag{8}$$

where $\bar{y}_k = n_k^{-1} \sum_{m=1}^{n_k} y_{km}, n_k$ is the number of DH lines within the kth cross, y_{km} is the phenotypic value of the mth DH line within the kth cross, $\sigma_{\epsilon_C}^2 = \frac{\sigma_{C\times l}^2}{L} + \frac{\sigma_{C\times y}^2}{Y} + \frac{\sigma_{C\times l\times y}^2}{LY}$, and $\sigma_{\epsilon_D}^2 = \frac{\sigma_{D\times l}^2}{L} + \frac{\sigma_{D\times y}^2}{Y} + \frac{\sigma_{D\times l\times y}^2}{LY} + \frac{\sigma_e^2}{LYR}$, where L is the number of test locations, Y is the number of test years, and R is the number of replications ($Y = 1$ and $R = 1$ in the present study).

In breeding scheme S_1TC-DHTC, the BLUPs of the crosses and S_1 families at the first stage are as follows

$$\widehat{c}_k = \bar{y}_k \frac{\sigma_C^2}{\sigma_C^2 + \sigma_{\epsilon_C}^2 + \sigma_F^2/n_k + \sigma_{\epsilon_F}^2/n_k} \tag{9}$$

$$\widehat{f}_{kl} = \left(y_{kl} - \bar{y}_k \frac{\sigma_C^2 + \sigma_{\epsilon_C}^2}{\sigma_C^2 + \sigma_{\epsilon_C}^2 + \sigma_F^2/n_k + \sigma_{\epsilon_F}^2/n_k} \right) \frac{\sigma_F^2}{\sigma_F^2 + \sigma_{\epsilon_F}^2}, \tag{10}$$

where $\overline{y}_k = n_k^{-1} \sum_{l=1}^{n_k} y_{kl}, n_k$ is the number of S_1 families within the kth cross, y_{kl} is the phenotypic value of the lth S_1 family within the kth cross, and $\sigma^2_{\epsilon_F} = \frac{\sigma^2_{F\times l}}{L} + \frac{\sigma^2_{F\times y}}{Y} + \frac{\sigma^2_{F\times l\times y}}{LY} + \frac{\sigma^2_e}{LYR}$.

In the second stage of the breeding scheme S_1TC-DHTC, the BLUPs of the crosses, S_1 families, and DH lines are

$$\widehat{c}_k = b\sum_1^{n_k} \frac{\overline{y}_{kl}}{a_l} \tag{11}$$

$$\widehat{f}_{kl} = \left(\overline{y}_{kl} - \widehat{c}_k\left(1 + \frac{\sigma^2_{\epsilon_C}}{\sigma^2_C}\right)\right)\frac{\sigma^2_F}{\sigma^2_F + \sigma^2_{\epsilon_F} + \sigma^2_{D/F}/n_{kl} + \sigma^2_{\epsilon_{D/F}}/n_{kl}}, \tag{12}$$

$$\widehat{d}_{klm} = \left(y_{klm} - \widehat{c}_k\left(1 + \frac{\sigma^2_{\epsilon_C}}{\sigma^2_C}\right) - \widehat{f}_{kl}\left(1 + \frac{\sigma^2_{\epsilon_F}}{\sigma^2_F}\right)\right)\frac{\sigma^2_{D/F}}{\sigma^2_{D/F} + \sigma^2_{\epsilon_{D/F}}}, \tag{13}$$

where $b = \sigma^2_C\left(1 + (\sigma^2_C + \sigma^2_{\epsilon_C})\sum_1^{n_k}\frac{1}{a_l}\right)^{-1}$, $a_l = \sigma^2_F + \sigma^2_{\epsilon_F} + \frac{\sigma^2_{D/F}}{n_{kl}} + \frac{\sigma^2_{\epsilon_{D/F}}}{n_{kl}}$, $\overline{y}_{kl} = n_{kl}^{-1}\sum_{l=1}^{n_{kl}} y_{klm}, n_k$ is the number of S_1 families within the kth cross, n_{kl} is the number of DH lines within the lth S_1 family of the kth cross, y_{klm} is the phenotypic value of the mth DH line within the lth S_1 family of the kth cross, $\sigma^2_{\epsilon_F} = \frac{\sigma^2_{F\times l}}{L} + \frac{\sigma^2_{F\times y}}{Y} + \frac{\sigma^2_{F\times l\times y}}{LY}$, and $\sigma^2_{\epsilon_{D/F}} = \frac{\sigma^2_{D/F\times l}}{L} + \frac{\sigma^2_{D/F\times y}}{Y} + \frac{\sigma^2_{D/F\times l\times y}}{LY} + \frac{\sigma^2_e}{LYR}$.

Simulation results

In DHTC, selection strategy 2a generally resulted in a slightly larger ΔG^* and $P(q)^*$ compared with the other strategies (Table 1). The standard deviation of the optimization criteria generally increased with decreasing relative contribution of the genetic variance from VC1 to VC3. For ΔG^* in DHTC–2a and VC1 (with highest ΔG^* and $P(q)^*$ compared to VC2 and VC3), the optimum allocation was 1 (L_1^*) and 14 (L_2^*) test locations, four crosses ($N^*_{1_C}$) and 2,290 DH lines ($N^*_{1_D}$) of each cross in the first stage, and 404 DH lines (N^*_{2D}) in the second stage. For $P(q)^*$ in DHTC–2a and VC1, N_J^* was smaller, and $L_1^*$was larger than that for ΔG^*.

In S_1TC-DHTC, selection strategy 2 generally resulted in a larger ΔG^* than strategy 1 (Table 2). The standard deviation of the optimization criteria increased with decreasing relative contribution of the genetic variance from VC1 to VC3. The optimum number of initial crosses $N^*_{1_C}$ under VC1 and VC3 was generally larger for $P(q)^*$ than for ΔG^*. For $P(q)^*$, the optimum number of test locations L_J^* generally increased from VC1 to VC3. The optimum allocation for ΔG^* in S_1TC-DHTC–2c in combination with VC1 (with highest ΔG^* and $P(q)^*$ compared to VC2 and VC3) was 10 (L_1^*) and 11 (L_2^*) test locations, five crosses, and 196 S_1 families of each cross in the first stage, and 812 DH lines (N_2^*) in the second stage. In S_1TC-DHTC–2c and VC1, N_1^* was reduced and L_J^* increased for $P(q)^*$ compared to ΔG^*, whereas N_2^* was similar.

Breeding scheme S_1TC-DHTC (Table 2), compared with DHTC (Table 1), was superior for both optimization criteria, ΔG^* and $P(q)^*$ under all assumptions. The superiority was relatively more pronounced for $P(q)^*$ than for ΔG^*, and for VC2 than for VC1 and VC3. Further, for both breeding schemes and all selection strategies, the estimates of both optimization criteria were generally reduced as the contributions of the genetic variance relative to the masking variances decreased from VC1 to VC3.

Discussion

Effect of BLUP on the optimization criteria

In an earlier paper, we studied the breeding schemes with the same assumptions concerning the variance component ratios, test resources, and production costs as in the present one, but employed conventional phenotypic selection based on mean performance only of the candidates rather than BLUP (Wegenast et al. 2010). Selection based on BLUP (Tables 1, 2) instead of mean performance solely of the test candidates slightly increased the progress from selection in both breeding schemes. This enhancement of ΔG^* and $P(q)^*$ was due to the theoretical model underlying the selection based on BLUP, i.e., the use of the information of related genotypes in addition to that of the test candidate as a selection index in BLUP. The impact of BLUP might have been reduced by the large test capacity of the breeding program which lead to very reliable values of the mean performance of the test candidates leaving a little scope of their improvement with the use of the additional information of related genotypes. Therefore, in breeding programs with smaller budgets, the relative advantage of BLUP compared to simple means is expected to be larger.

Our study focused on investigating the general influence of BLUP on the progress from selection and the optimum allocation of test resources. In order to keep the formulas and the computing time for the simulations at an acceptable level, we made a few simplifying assumptions. These were unrelated parental crosses, a constant coancestory coefficient Θ among the candidates within crosses expected on the basis of quantitative genetic theory (neglecting the

Table 1 Two-stage breeding scheme DHTC (evaluation of testcross progenies of DH lines within crosses at both stages) with selection based on BLUPs: optimum allocation of test resources to maximize the optimization criteria (OC), namely selection gain ($\Delta\widehat{G}^*$) and the probability of selecting superior genotypes [$\widehat{P}(q = 0.1\%)^*$], for three different ratios of variance components (VC) $\sigma^2_G{:}\sigma^2_{G\times y}{:}\sigma^2_{G\times l}{:}\sigma^2_{G\times l\times y}{:}\sigma^2_e$ assuming equal ratios for crosses and DH lines

Selection strategy[a]	VC[b]	Optimum allocation					
		N_1^*	N_2^*	L_1^*	L_2^*	OC	SD_{OC}
Optimization criterion $\Delta\widehat{G}^*$							
1	1	5,502 (2 × 2,751)[c]	372 (2 × 186)[c]	2	15	3.62	0.29
2a	1	9,160 (4 × 2,290)[c]	404	1	14	3.64	0.31
2b	1	8,676 (3 × N_{1_D})[d]	490	1	13	3.63	0.32
2c	1	5,866 (2 × N_{1_D})[d]	334	2	14	3.64	0.28
1	2	3,956 (2 × 1,917)[c]	392 (2 × 196)[c]	3	15	3.33	0.33
2a	2	4,590 (3 × 1,530)[c]	232	3	14	3.36	0.33
2b	2	3,850 (2 × N_{1_D})[d]	388	3	15	3.34	0.29
2c	2	3,548 (2 × N_{1_D})[d]	221	4	15	3.36	0.29
1	3	3,566 (2 × 1,783)[c]	232 (2 × 116)[c]	4	14	3.06	0.34
2a	3	2,524 (2 × 1,262)[c]	191	6	15	3.08	0.30
2b	3	3,234 (2 × N_{1_D})[d]	338	4	14	3.06	0.30
2c	3	3,704 (2 × N_{1_D})[d]	188	4	14	3.07	0.32
Optimization criterion $\widehat{P}(q = 0.1\%)^*$							
1	1	5,684 (4 × 1,421)[c]	342 (2 × 171)[c]	2	15	0.80	0.23
2a	1	5,760 (4 × 1,440)[c]	329	2	15	0.83	0.20
2b	1	9,075 (5 × N_{1_D})[d]	386	1	15	0.82	0.21
2c	1	8,180 (4 × N_{1_D})[d]	473	1	15	0.83	0.19
1	2	5,622 (3 × 1,874)[c]	352 (2 × 176)[c]	2	15	0.65	0.29
2a	2	5,620 (3 × 1,810)[c]	384	2	15	0.67	0.26
2b	2	4,595 (5 × N_{1_D})[d]	231	3	14	0.65	0.26
2c	2	4,572 (3 × N_{1_D})[d]	254	3	13	0.65	0.29
1	3	3,210 (2 × 1,605)[c]	346 (2 × 173)[c]	4	14	0.50	0.30
2a	3	3,570 (2 × 1,785)[c]	231	4	14	0.50	0.28
2b	3	3,798 (2 × N_{1_D})[d]	400	3	15	0.50	0.28
2c	3	3,436 (2 × N_{1_D})[d]	255	4	15	0.50	0.28

N_J^* optimum number of test candidates in stage j, L_J^* optimum number of test locations in stage j, SD standard deviation

[a] 1: Sequential selection among and within parental crosses; selection among all DH lines with 2a = fixed numbers of DH lines, 2b = moderately, or 2c = strongly variable numbers of DH lines within crosses

[b] VC1 = 0.5:0.125:0.125:0.25:1; VC2 = 0.5:0.25:0.25:0.5:2; VC3 = 0.5:0.5:0.5:1:4

[c] Number of crosses × DH lines within crosses

[d] Number of crosses × variable number of N_{1_D} DH lines within crosses

variance of Θ due to segregation), and known instead of estimated variance components. The use of empirical BLUP using variance components estimated with restricted maximum likelihood instead of true BLUP (Piepho et al. 2008) would probably slightly reduce the advantage of BLUP. A direct and more reliable estimate of the actual coancestory coefficient Θ for each pair of candidates based on molecular markers, might have enhanced the impact of BLUP. Therefore, incorporation of marker information (Bauer et al. 2006; Lorenzana and Bernardo 2009; Schrag et al. 2010) or the estimation of variance component ratios for each trial separately could make the selection based on BLUP more efficient and also influence the optimum allocation of test resources. Then, the relationship matrix ***A*** would be non-diagonal.

In breeding scheme S_1TC-DHTC, the use of BLUP led to a very high probability $[P(q)^*]$ for the selection of the best $q = 0.1\%$ genotypes under the assumption of VC1, i.e., when relative contribution of genetic variance is high. Response to selection increased with the application of BLUP. The positive effect of BLUP was also found in studies on the prediction of the performance of maize

Table 2 Two-stage breeding scheme S_1TC-DHTC (evaluation of testcrosses of S_1 families at the first stage and of DH lines within S_1 families at second stage) with selection based on BLUPs: Optimum allocation of test resources to maximize the optimization criteria (OC), namely selection gain ($\Delta\widehat{G}^*$) and the probability of selecting superior genotypes [$\widehat{P}(0.1\%)^*$], for three different ratios of variance components (VC) σ_G^2:$\sigma_{G\times y}^2$:$\sigma_{G\times l}^2$:$\sigma_{G\times l\times y}^2$:$\sigma_e^2$ for crosses, S_1 families, and DH lines assuming equal ratios for S_1 families and DH lines

Selection strategy[a]	VC[b]	Optimum allocation					
		N_1^*	N_2^*	L_1^*	L_2^*	OC	SD_{OC}
Optimization criterion $\Delta\widehat{G}^*$							
1	1	800 (4 × 200)[c]	760 (2 × 1 × 380)[d]	13	11	3.96	0.46
2a	1	864 (4 × 216)[c]	716 (2 × 1 × 358)[d]	11	13	3.98	0.43
2b	1	516 (4 × 129)[c]	956 (2 × N_{2_F} × N_{2_D})[e]	14	12	3.99	0.45
2c	1	980 (5 × 196)[c]	812 (2 × N_{2_F} × N_{2_D})[e]	10	11	4.00	0.44
1	2	736 (4 × 184)[c]	712 (2 × 2 × 178)[d]	11	15	3.58	0.51
2a	2	940 (5 × 188)[c]	568 (2 × 2 × 142)[d]	11	15	3.65	0.47
2b	2	788 (4 × 197)[c]	600 (2 × N_{2_F} × N_{2_D})[e]	14	13	3.64	0.49
2c	2	1,080 (6 × 180)[c]	360 (2 × N_{2_F} × N_{2_D})[e]	11	15	3.64	0.49
1	3	604 (2 × 302)[c]	698 (2 × 1 × 349)[d]	14	14	3.23	0.54
2a	3	537 (3 × 179)[c]	716 (2 × 2 × 179)[d]	15	15	3.27	0.50
2b	3	600 (2 × 300)[c]	754 (2 × N_{2_F} × N_{2_D})[e]	15	13	3.28	0.49
2c	3	424 (2 × 212)[c]	952 (2 × N_{2_F} × N_{2_D})[e]	15	13	3.28	0.52
Optimization criterion $\widehat{P}(0.1\%)^*$							
1	1	944 (8 × 118)[c]	764 (2 × 1 × 382)[d]	11	11	0.92	0.17
2a	1	770 (7 × 110)[c]	918 (2 × 1 × 367)[d]	10	12	0.93	0.14
2b	1	655 (5 × 131)[c]	800 (2 × N_{2_F} × N_{2_D})[e]	14	12	0.93	0.16
2c	1	660 (6 × 120)[c]	794 (2 × N_{2_F} × N_{2_D})[e]	13	13	0.93	0.15
1	2	955 (5 × 191)[c]	680 (3 × 1 × 200)[d]	9	15	0.77	0.29
2a	2	760 (4 × 190)[c]	744 (2 × 3 × 100)[d]	11	14	0.81	0.25
2b	2	975 (4 × 197)[c]	600 (2 × N_{2_F} × N_{2_D})[e]	14	13	0.79	0.27
2c	2	1,080 (6 × 180)[c]	398 (2 × N_{2_F} × N_{2_D})[e]	12	15	0.80	0.24
1	3	666 (3 × 222)[c]	636 (2 × 3 × 171)[d]	14	15	0.58	0.35
2a	3	716 (4 × 179)[c]	544 (2 × 3 × 187)[d]	15	15	0.59	0.33
2b	3	564 (2 × 282)[c]	794 (2 × N_{2_F} × N_{2_D})[e]	15	13	0.58	0.36
2c	3	720 (4 × 180)[c]	586 (2 × N_{2_F} × N_{2_D})[e]	14	15	0.60	0.33

N_j^* = optimum number of test candidates in stage j, L_j^* optimum number of test locations in stage j, SD standard deviation

[a] 1 = Sequential selection among and within parental crosses; sequential selection among and within parental crosses at the first stage and selection among all DH lines at the second stage with 2a = fixed numbers of S_1 families and DH lines, 2b = moderately, or 2c = strongly variable numbers of S_1 families and DH lines

[b] VC1 = 0.5:0.125:0.125:0.25:1 (crosses) and 0.25:0.0625:0.0625:0.125:1 (S_1 families and DH lines), VC2 = 0.5:0.25:0.25:0.5:2 (crosses) and 0.25:0.125:0.125:0.25:2 (S_1 families and DH lines), VC3 = 0.5:0.5:0.5:1:4 (crosses) and 0.25:0.25:0.25:0.5:4 (S_1 families and DH lines)

[c] Number of crosses × S_1 families within crosses

[d] Number of crosses × S_1 families within crosses × DH lines within S_1 families

[e] Number of crosses × variable number of N_{2_F} S_1 families within crosses and N_{2_D} DH lines within S_1 families

single crosses (Bernardo 1996). In addition, the use of BLUP compared to conventional phenotypic selection led to a closer relatedness of the selected DH lines, expressed as the average coefficient of coancestry, in selection strategies 2b and 2c with variable numbers of S_1 families and DH lines in both breeding schemes (data not shown). In the selection based on BLUP, candidates having a larger number of relatives with desirable performance have a bonus to be selected, because more information is available leading to more accurate BLUPs. Thus, it is likely that DH lines within crosses or within S_1 families with a larger number of related DH lines may get selected rather than those within a smaller cross or S_1 family.

Advantage of formulas to estimate genotypic values and optimally allocate test resources

Since we assumed special structures of the covariance matrices $\boldsymbol{A_C}, \boldsymbol{A_D}, \boldsymbol{A_F}, \boldsymbol{A_{D/F}}$, and $\boldsymbol{R}$ (e.g., constant covariances within crosses and S_1 families within crosses), the development of mathematical formulas to estimate the genotypic values of the test candidates rendered matrix inversion of the MME unnecessary. These formulas facilitated and accelerated the simulations considerably. Using the statistical software R (R Development Core Team 2006) in breeding scheme S_1TC-DHTC, the mathematical formulas saved approximately two-thirds of the simulation time compared to the matrix inversions (data not shown), whereas in breeding scheme DHTC, BLUP-based simulations could not be undertaken as this required more computing power than available. The simplifying assumption of unrelated crosses was, however, necessary for the development of the mathematical formulas. We combined all non-genetic effects (i.e., the genotype × environment interactions as well as the residual error) into the random error ϵ. Thus, the error effects of related candidates may be correlated and consequently, $\boldsymbol{R}$ was a block diagonal matrix. The model used in this study is therefore an extension to that used in the literature on BLUPs in plant breeding, where $\boldsymbol{R}$ was assumed to be diagonal (Bernardo 1996).

We developed our formulas by calculating a general inverse of the MME analytically using methods described in the literature (cf., Searle 1996). The calculations were made with the statistical software Mathematica 7 (Wolfram Research Inc 2008), which is widely used for symbolic matrix calculations. Finally, the general formulas were triple-checked numerically for correctness using the software programs Mathematica 7, SAS Proc IML (SAS Institute 2004), and R (R Development Core Team 2006) for various allocations of resources. In addition, our formulas are in agreement with others given in the literature (Searle 1996; Holland et al. 2003).

With our formulas, the BLUP of each cross $(\widehat{c}_k)$, S_1 family within a cross $(\widehat{f}_{kl})$, and DH line within a cross $(\widehat{d}_{km})$ or S_1 family $(\widehat{d}_{klm})$ can be calculated directly from the phenotypic data (Eqs. 7–13). Thus, BLUPs of individual candidates can be computed without calculating the BLUPs of all others. The formulas solely depend on the phenotypic values of the candidates, the variance component ratios, the allocation of test resources, as well as the number of test candidates n, and the genetic relationship of the candidates, which are all fixed for a given breeding scheme. Thus, the influence of (1) the resource allocation, i.e., numbers of test candidates, locations and replications and (2) the underlying variance component ratios on the BLUPs can be calculated directly from the formulas without matrix inversion. In breeding scheme DHTC, the BLUP of a cross is the phenotypic mean of the cross multiplied with an adjusted heritability. The BLUP of a DH line within a cross is the phenotypic value minus the effect of a cross multiplied with the heritability. The calculation of the BLUPs at the first stage of S_1TC-DHTC is analogous to DHTC. At the second stage of breeding scheme S_1TC-DHTC, the BLUP of a cross can be considered as the harmonic mean of the families within that cross.

Optimum allocation of test resources

In both breeding schemes, the optimum number of test locations increased with decreasing relative contribution of genotypic variance from VC1 to VC3 (Tables 1, 2). In breeding scheme DHTC, the optimum number of test candidates in the first (N_1^*), and also in most cases in the second stage (N_2^*), decreased to counterbalance the additional budget for the test locations. In breeding scheme S_1TC-DHTC, the same trend was generally observed for the test locations.

When comparing the optimum allocation of test resources for selection based on BLUPs in the present study with that based on mean performance solely of the candidates (Wegenast et al. 2010) in breeding scheme DHTC, the optimum number of test candidates N_1^* at the first stage was strongly reduced with BLUP. This leads to a shift of resources from the first to the second stage of selection, and, thus, a more intensive evaluation of the promising candidates in the second stage. For S_1TC-DHTC, no clear trend was discernible from the comparison of the optimum allocation of resources for selection based on BLUP versus mean performance solely of the candidate.

In applied breeding programs, elite inbreds within the heterotic groups have complex pedigree structure and relationships, leading to varying coefficients of coancestry among and within crosses (Bernardo 1996). This complex structure requires simplifying assumptions. However, when crosses are related, a larger number of crosses with a lower number of candidates within crosses might be optimum for selection based on BLUP, because data of related crosses could also be used to estimate the BLUPs of the candidates.

Comparing the optimum allocation for ΔG^* with $P(q)^*$ in S_1TC-DHTC, the optimum number of crosses at the first stage ($N_{1_C}^*$) was generally larger for $P(q)^*$ (Table 2). However, the values of N_J^* and L_J^* did not show any consistent trend between the two optimization criteria. Similarly in breeding scheme DHTC, the optimization criterion had no consistent effect on the optimum allocation.

Comparison of selection strategies

In both breeding schemes, the values of both optimization criteria decreased from VC1 to VC3 (Tables 1, 2). This decrease was due to the progressive relative increase in non-genetic variances from VC1 to VC3 and hence, reduced heritability. The relative decrease for ΔG^* was smaller in DHTC than in S_1TC-DHTC, whereas for $P(q)$ the trend was the opposite. This trend underpins the results of an earlier study (Wegenast et al. 2010) that breeding scheme S_1TC-DHTC is more robust for the criterion $P(q)$.

With BLUP, differences among selection strategies narrowed down compared to that for mean performance solely of the test candidates (Wegenast et al. 2010). This may be due to the fact that with BLUP, the values of the individuals are adjusted with the performance of related individuals which leads to shrinkage as compared with conventional phenotypic selection.

In breeding scheme DHTC, all three variants of selection strategy 2 involving selection among DH lines at both stages resulted in equal or higher ΔG^* and $P(q)^*$ than selection strategy 1 involving sequential selection among and within crosses (Table 1). Further, strategy 2 resulted generally in equal or lower SD_{OC} for both criteria, particularly $P(q)^*$, compared with strategy 1. In breeding scheme S_1TC-DHTC, all three variants of selection strategy 2 generally resulted in larger estimates of ΔG^*, whereas for $P(q)^*$, both selection strategies 1 and 2 showed comparable results.

In both breeding schemes, selection strategies using variable rather than fixed numbers of S_1 families and DH lines did not consistently lead to higher ΔG^* and $P(q)^*$ (Tables 1, 2). Thus, the advantage of strategies 2b and 2c in breeding scheme S_1TC-DHTC reported earlier in Wegenast et al. (2010) using selection based solely on phenotypic values of the candidate itself was not supported by the present study. This may be due to the effect of cross and/or family size, i.e., the number of DH lines or S_1 families within crosses, and the number of DH lines within S_1 families. The identification of a superior DH line belonging to a very small-sized cross or S_1 family compared with larger-sized ones is less likely.

Conclusions

The formulas for the calculation of the BLUPs, developed in this study greatly facilitate the computation of ΔG^* and $P(q)^*$ estimates as well as the optimum allocation of test resources compared with the approach using the MME. These formulas not only accelerated the computations but also facilitated investigations of the effect of various factors on the BLUP estimates more explicitly. Selection based on BLUPs rather than phenotypic values had a positive, though small, effect on progress from selection in both breeding schemes. The progress is higher for breeding scheme S_1TC-DHTC than DHTC, whereas the differences among the selection strategies are small for both optimization criteria and breeding schemes. Further research is warranted to study the effect of BLUP on the progress from selection by the additional integration of marker information to determine directly the relationship matrices $\boldsymbol{A}$ and $\boldsymbol{R}$.

Acknowledgments This research was supported by funds from DFG, Grant 1070, International Research Training Group "Sustainable Resource Use in North China" to T. Wegenast. Financial support for X. Mi was provided by Grant 0315072B "GABI Gain" within the framework "Genome analysis of the plant biological system" supported by the German Federal Ministry of Education, Research, and Technology. We greatly appreciate the helpful comments and suggestions of two anonymous reviewers.

References

Bauer AM, Reetz TC, Léon J (2006) Estimation of breeding values of inbred lines using best linear unbiased prediction (BLUP) and genetic similarities. Crop Sci 46:2685–2691

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

Bernardo R (2002) Breeding for quantitative traits in plants. Stemma Press, Woodbury, Minnesota

Bernardo R (2003) Parental selection, number of breeding populations, and size of each population in inbred development. Theor Appl Genet 107:1252–1256

Bernardo R (2009) Should maize doubled haploids be induced among F_1 or F_2 plants? Theor Appl Genet 119:255–262

Holland JB, Nyquist WE, Cervantes-Martinez CT (2003) Estimating and interpreting heritability for plant breeding: an update. Plant Breed Rev 22:9–112

Longin CFH, Utz HF, Reif JC, Wegenast T, Schipprack W, Melchinger AE (2007) Hybrid maize breeding with doubled haploids: III. Efficiency of early testing prior to doubled haploid production in two-stage selection for testcross performance. Theor Appl Genet 115:519–527

Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor Appl Genet 120:151–161

Melchinger AE (1987) Expectation of means and variances of testcrosses produced from from F_2 and backcross individuals and their selfed progenies. Heredity 59:105–115

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, Longin CFH, Utz HF, Reif JC (2005) Hybrid maize breeding with doubled haploid lines: quantitative genetic and selection theory for optimum allocation of resources. In: Proceedings of the forty-first annual Illinois Corn Breeders' School 2005, Urbana-Champaign, Illinois, USA, pp 8–21

Piepho HP, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209–228

R Development Core Team (2006) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN: 3-900051-07-0, URL http://www.R-project.org

Robinson GK (1991) That BLUP is a good thing: the estimation of random effects. Stat Sci 6:15–51

SAS Institute (2004) SAS Version 9.1. SAS Institute, Cary, NC

Schrag TA, Möhring J, Melchinger AE, Kusterer B, Dhillon BS, Piepho HP, Frisch M (2010) Prediction of hybrid performance in maize using molecular markers and joint analyses of hybrids and parental inbreds. Theor Appl Genet 120:451–461

Searle SR (1996) The matrix handling of BLUE and BLUP in the mixed linear model. Linear Algebra Appl 264:291–311

Smith JSC, Hussain T, Jones ES, Graham G, Podlich D, Wall S, Williams M (2008) Use of doubled haploids in maize breeding: implications for intellectual property protection and genetic diversity in hybrid crops. Mol Breed 22:51–59

Wegenast T, Longin CFH, Utz HF, Melchinger AE, Maurer HP, Reif JC (2008) Hybrid maize breeding with doubled haploids. IV. Number versus size of crosses and importance of parental selection in two-stage selection for testcross performance. Theor Appl Genet 117:251–260

Wegenast T, Utz HF, Longin CFH, Maurer HP, Dhillon BS, Melchinger AE (2010) Hybrid maize breeding with doubled haploids: V. Selection strategies for testcross performance with variable sizes of crosses and S_1 families. Theor Appl Genet 120:699–708

Wolfram Research, Inc. (2008) Mathematica, Version 7.0. Champaign, IL