Theor Appl Genet (2010) 120:699–708
DOI 10.1007/s00122-009-1187-y

ORIGINAL PAPER

Hybrid maize breeding with doubled haploids: V. Selection strategies for testcross performance with variable sizes of crosses and S_1 families

Thilo Wegenast · H. Friedrich Utz · C. Friedrich H. Longin · Hans Peter Maurer · Baldev S. Dhillon · Albrecht E. Melchinger

Received: 20 February 2009 / Accepted: 9 October 2009 / Published online: 29 October 2009

Abstract In hybrid maize (*Zea mays* L.) breeding, doubled haploids (DH) are increasingly replacing inbreds developed by recurrent selfing. Doubled haploids may be developed directly from S_0 plants in the parental cross or via S_1 families. In both these breeding schemes, we examined 2 two-stage selecting strategies, i.e., considering or ignoring cross and family structure while selection among and within parental crosses and S_1 families. We examined the optimum allocation of resources to maximize the selection gain ΔG and the probability $P(q)$ of identifying the $q\%$ best genotypes. Our specific objectives were to (1) determine the optimum number and size of crosses and S_1 families, as well as the optimum number of test environments and (2) identify the superior selection strategy. Selection was based on the evaluation of testcross progenies of (1) DH lines in both stages (DHTC) and (2) S_1 families in the first stage and of DH lines within S_1 families in the second stage (S_1TC-DHTC) with uniform and variable sizes of crosses and S_1 families. We developed and employed simulation programs for selection with variable sizes of crosses and S_1 families within crosses. The breeding schemes and selection strategies showed similar relative efficiency for both optimization criteria ΔG and P (0.1%). As compared with DHTC, S_1TC-DHTC had larger ΔG and P (0.1%), but a higher standard deviation of ΔG. The superiority of S_1TC-DHTC was increased when the selection was done among all DH lines ignoring their cross and family structure and using variable sizes of crosses and S_1 families. In DHTC, the best selection strategy was to ignore cross structures and use uniform size of crosses.

Introduction

Optimum allocation of test resources is of crucial importance for the efficiency and competitiveness of breeding programs. With limited test resources, a plant breeder has to strike a balance among the number of crosses, test candidates within each cross, as well as test environments and replications within environments. Selection among crosses enables breeders to discard inferior crosses in early stages of line development and to assign the resources to the promising ones (cf., Schnell 1982). This selection is generally based on the mean performance of the crosses, but this entails the risk of discarding individual superior candidates within the rejected crosses. Therefore, some researchers favored selection among the total number of test candidates disregarding their cross structure (cf., Lush 1947).

In hybrid maize (*Zea mays* L.) breeding, doubled haploids (DH) are increasingly becoming popular, replacing conventionally developed inbred lines. Alternative

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1187-y) contains supplementary material, which is available to authorized users.

Communicated by H. Becker.

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

T. Wegenast
e-mail: wegenast@uni-hohenheim.de

C. F. H. Longin
Limagrain Verneuil Holding, BP 58 Route de Lavardac, 47600 Nérac, France

H. P. Maurer
State Plant Breeding Institute, University of Hohenheim, 70593 Stuttgart, Germany

breeding schemes for recurrent selection with DH, disregarding cross structures, were optimized by Gordillo and Geiger (2008). However, there is no such study on the identification of inbred lines for utilization in hybrid breeding as well as the comparison of selection strategies considering or ignoring cross structures. Further, Gordillo and Geiger (2008) used approximations to simplify the estimation of the selection gain (ΔG, Falconer and Mackay 1996).

A constant number of lines or families within crosses is generally considered in literature (cf., Baker 1984; Bernardo 2003). Under this assumption, Longin et al. (2007) and Wegenast et al. (2008) developed and applied theory for ΔG in breeding schemes for various situations in maize breeding. In practical maize breeding, the numbers of S_1 families and DH lines vary among crosses. Larger numbers of S_1 families or DH lines are produced in promising crosses than others, based on prior information of their parents or their mean performance in earlier testing stages. The same is true for DH lines within S_1 families. Thus, a higher proportion of the resources is allocated to the more promising crosses. However, no formula or simulation program is available in literature to compute the optimum allocation of test resources and estimates of ΔG under this situation, which is an apparent gap between theory and practical breeding.

Besides ΔG, progress from selection has been quantified by the probability $P(q)$ of identifying the $q\%$ superior genotypes (cf., Keuls and Sieben 1955). Longin et al. (2006a, b) found similar optimum allocations for both these criteria in a DH breeding scheme, but they considered only one cross.

We developed and employed simulation programs to assess the optimum allocation of test resources for DH line development and their evaluation in testcrosses to maximize ΔG or $P(q)$. Two breeding schemes were considered in which the selection was based on the performance of testcrosses of (1) DH lines derived fromseveral crosses at two stages (DHTC) or (2) S_1 families fromseveral crosses in the first and of DH lines within S_1 families, in the second stage (S_1TC-DHTC). In both breeding schemes, we investigated the effect of different selection strategies on the optimum allocation of test resources. Selection strategies used in earlier studies (Longin et al. 2007; Wegenast et al. 2008) were extended to include (a) selection among and within crosses and S_1 families disregarding cross structures and (b) variable sizes of crosses and S_1 families. Our specific objectives were to (1) determine the optimum number and sizes of crosses and S_1 families as well as the optimum number of test environments for maximizing ΔG and $P(q)$ and (2) identify the superior selection strategy for each breeding scheme with respect to ΔG and $P(q)$.

Materials and methods

Breeding schemes

We considered two breeding schemes, DHTC and S_1TC-DHTC (Wegenast et al. 2008, Supplementary Fig. S1). In DHTC, DH lines are produced from S_0 plants randomly taken from a parental cross. In S_1TC-DHTC, S_1 families are developed from similarly taken S_0 plants and evaluated in test crosses, then DH lines are produced in selected S_1 families. Further, parental crosses for a new breeding cycle are selected on the basis of the parental testcross mean in the preceding breeding cycle. For selection among parental crosses, a phenotypic correlation (ρ_P) of the mean performance of the parental lines, known from preceding breeding cycles with the mean genotypic value of the testcross performance of their progenies, was assumed to be 0.71 or 0.50 (Wegenast et al. 2008). The selection at both stages was based on the phenotypic mean of testcross performance of the candidates with a fixed given tester, evaluated at L_j test locations, where j refers to the selection stages 1 and 2. The target variable throughout this study is the genotypic value of the testcross performance for grain yield.

Selection strategies

For each breeding scheme, we evaluated two selection strategies (Table 1). In strategy 1, we selected first among parental crosses and then among DH lines within crosses in DHTC–1 and additionally among S_1 families in S_1 TC-DHTC–1. In the second strategy, selection at both stages was performed among all DH lines disregarding the cross structure in DHTC–2. In S_1TC-DHTC–2, selection at the first stage was performed first among crosses and then among S_1 families; at the second stage, selection was performed among all DH lines disregarding the cross and family structure. The second selection strategy had three variations as described in "Size of crosses and S_1 families".

Size of crosses and S_1 families

Three different procedures were used to determine the number of DH lines or S_1 families per cross and DH lines per S_1 family (hereafter referred to as size of crosses and S_1 families; Table 1). In selection strategies 1 and 2a, we assumed a uniform size of crosses and a uniform size of S_1 families in each stage. In selection strategies 2b and 2c, variable sizes of crosses and S_1 families were assumed. In DHTC–2b and 2c, the size of the crosses depended on their rank, calculated from their performance in the parental selection. In S_1TC-DHTC–2b and 2c, the size of the

Table 1 Selection strategies for the two-stage breeding schemes DHTC and S_1TC-DHTC showing the selection procedures in both stages as well as the calculation of the size of the crosses and S_1 families, i.e., the number of DH lines or S_1 families per cross and DH lines per S_1 family

Breeding scheme/ selection strategy	Selection among						Size of			
	Crosses		S_1 families		DH lines		Crosses		S_1 families	
	Stage 1	Stage 2	Stage 1	Stage 2	Stage 1	Stage 2	Stage 1	Stage 2	Stage 1	Stage 2
DHTC–1	Yes	Yes	NA	NA	Yes	Yes	$\frac{N_1}{N_{1_C}}$	$\frac{N_2}{N_{2_C}}$	NA	NA
DHTC–2	No	No	NA	NA	Yes	Yes			NA	NA
DHTC–2a							$\frac{N_1}{N_{1_C}}$	–/–		
DHTC–2b							$N_1\left(\frac{7}{25}\right)^{\sqrt{k}}$	–/–		
DHTC–2c							$N_1\left(\frac{1}{2}\right)^{k}$	–/–		
S_1TC-DHTC–1	Yes	Yes	Yes	Yes	NA	Yes	$\frac{N_1}{N_{1_C}}$	$\frac{N_2}{N_{2_C}}$	$\frac{N_1}{N_{1_C}N_{1_{F/C}}}$	$\frac{N_2}{N_{2_C}N_{2_{F/C}}}$
S_1TC-DHTC–2	Yes	No	Yes	No	NA	Yes				
S_1TC-DHTC–2a							$\frac{N_1}{N_{1_C}}$	$\frac{N_2}{N_{2_C}}$	$\frac{N_1}{N_{1_C}N_{1_{F/C}}}$	$\frac{N_2}{N_{2_C}N_{2_{F/C}}}$
S_1TC-DHTC–2b							$\frac{N_1}{N_{1_C}}$	$N_2\left(\frac{7}{25}\right)^{\sqrt{k}}$	$\frac{N_1}{N_{1_C}N_{1_{F/C}}}$	$N_{2_{\mathrm{DH}/F}}\left(\frac{7}{25}\right)^{\sqrt{k}}$
S_1TC-DHTC–2c							$\frac{N_1}{N_{1_C}}$	$N_2\left(\frac{1}{2}\right)^{k}$	$\frac{N_1}{N_{1_C}N_{1_{F/C}}}$	$N_{2_{\mathrm{DH}/F}}\left(\frac{1}{2}\right)^{k}$

S_1 families occur only in breeding scheme S_1TC-DHTC. DH lines at first stage occur only in breeding scheme DHTC

–/–, no fixed cross size in stage 2, due to non-hierarchical selection in stage 1

N_j = the number of test candidates in stage j, N_{j_C} = the number of crosses in stage j, $N_{j_{F/C}}$ = the number of S_1 families within a cross in stage j, $N_{j_{\mathrm{DH}/C}}$ = the number of DH lines within a cross in stage j, and k = the rank of a given cross or S_1 family

crosses and S_1 families within crosses in the second stage depended on their rank, calculated from their performance at the first selection stage. In both breeding schemes, with poorer rank, the size of crosses and S_1 families decreased moderately in strategy 2b and strongly in strategy 2c (an example is given in Supplementary Fig. S1). In some allocations, the available test candidates could not be fully allocated to the crosses or S_1 families within crosses. In these cases, the remaining small number of DH lines were assigned to the best cross and S_1 family within crosses.

Test locations common to both stages (L_c) were assumed such that $L_c = \min(L_1, L_2)$. Without restrictions on L_j in stage j, ΔG is maximum for one replication per test location for both stages of selection (cf., Bernardo 2002; Melchinger et al. 2005). Thus, we considered the number of replications equal to one. After two stages of selection, the best $N_f = 10$ DH lines were selected. In strategy 1, the best 10 DH lines with the highest testcross performance within the best cross (and S_1 family within that cross) were selected, based on an earlier study showing that this approach maximized ΔG (Wegenast et al. 2008).

Economic frame and quantitative-genetic parameters

A fixed total budget for the production and evaluation of the test candidates in two selection stages was defined in terms of testcross plot equivalents. An equal plot size at both selection stages was assumed. In DHTC, the budget equals $N_1[K_{\mathrm{DH}} + L_1(1 + K_T)] + N_2L_2(1 + K_T)$, where N_j refers to the total number of test candidates available in stage j, L_j to the total number of test locations at stage j, K_{DH} to the production costs of one DH line and K_T to the production costs of testcross seed for one plot. In S_1TC-DHTC, the budget equals $N_1[K_F + L_1(1 + K_T)] + N_2[K_{\mathrm{DH}} + L_2(1 + K_T)]$, where K_F refers to the production costs of one S_1 family. All costs are based on actual costs in the maize breeding program of the University of Hohenheim. We assumed that $K_{\mathrm{DH}} = 1/2$, $K_T = 1/25$, and $K_F = 1/12$ testcross plot equivalents. Three budgets were compared with a total of 10,000, 20,000 and 40,000 testcross plot equivalents available for line development 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) were obtained from the evaluation of DH populations for grain yield in maize programs of 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., crosses (C), DH lines within crosses (DH/C), S_1 families within crosses (F/C) or DH lines within S_1 families (DH/F). In scenario VC1, we assumed a variance component ratio

σ^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 for *C* and DH/*C*, respectively, and 0.25:0.0625:0.0625:0.125:1 for *F*/*C* and DH/*F*, respectively (Wegenast et al. 2008). In other scenarios, the contribution of σ^2_G was kept constant, but the non-genetic variances were doubled (VC2) and quadrupled (VC3).

Simulation model

The selection strategies were investigated by Monte Carlo simulations. Since grain yield is a quantitative trait, we assumed a Gaussian distribution of the genotypic and phenotypic values. Parental selection was based on the genotypic values of the crosses, assuming $N(0, \sigma^2_C)$.

In DHTC, the phenotypic value of a DH line was modeled by

$$Y = c + \mathrm{dh}/c + m_c + m_{\mathrm{dh}/c} + e \tag{1}$$

with

$$c \sim N(0, \sigma^{2\prime}_C),$$
$$\mathrm{dh}/c \sim N(0, \sigma^2_{\mathrm{DH}/C}),$$
$$m_c \sim N\left(0, \sigma^2_{C\times y} + \frac{\sigma^2_{C\times l}}{L_j} + \frac{\sigma^2_{C\times l\times y}}{L_j}\right), \text{ and}$$
$$m_{\mathrm{dh}/c} \sim N\left(0, \sigma^2_{\mathrm{DH}/C\times y} + \frac{\sigma^2_{\mathrm{DH}/C\times l}}{L_j} + \frac{\sigma^2_{\mathrm{DH}/C\times l\times y}}{L_j}\right),$$
$$e \sim N(0, \frac{\sigma^2_e}{L_j}),$$

where *c* and dh are the effects of the crosses and DH lines within crosses, respectively, m_c as well as $m_{\mathrm{dh}/c}$ the effects masking the former effects, and *e* the residual error. $\sigma^{2\prime}_C$ is the genotypic variance among crosses after parental selection (Wegenast et al. 2008).

In $\mathrm{S_1}$TC-DHTC, the phenotypic value of an $\mathrm{S_1}$ family in stage $j = 1$ was modeled by

$$Y = c + f/c + m_c + m_{f/c} + e \tag{2}$$

with *c*, m_c, and *e* as defined above and

$$f/c \sim N(0, \sigma^2_{F/C})$$

and

$$m_{f/c} \sim N\left(0, \sigma^2_{F/C\times y} + \frac{\sigma^2_{F/C\times l}}{L_j} + \frac{\sigma^2_{F/C\times l\times y}}{L_j}\right),$$

where *f*/*c* is the effect of the $\mathrm{S_1}$ families within crosses and $m_{f/c}$ their masking effect. The phenotypic value of a DH line in stage $j = 2$ in $\mathrm{S_1}$TC-DHTC was modeled by

$$Y = c + f/c + \mathrm{dh}/f + m_c + m_{f/c} + m_{\mathrm{dh}/f} + e \tag{3}$$

with *c*, *f*/*c*, m_c, $m_{f/c}$ and *e* as defined above and

$$\mathrm{dh}/f \sim N(0, \sigma^2_{\mathrm{DH}/F}) \text{ and}$$
$$m_{\mathrm{dh}/f} \sim N\left(0, \sigma^2_{\mathrm{DH}/F\times y} + \frac{\sigma^2_{\mathrm{DH}/F\times l}}{L_j} + \frac{\sigma^2_{\mathrm{DH}/F\times l\times y}}{L_j}\right),$$

where dh/*f* is the effect of the DH lines within $\mathrm{S_1}$ families and $m_{\mathrm{dh}/f}$ their masking effect. The covariance between the phenotypic values of both selection stages was determined as $\sigma^2_G + \frac{L_c\sigma^2_{G\times l}}{L_1L_2}$.

The number of simulation runs required to have an accuracy of 0.01 for ΔG was determined on the basis of the standard error of the arithmetic mean as $(3\mathrm{SD}/0.01)^2$ (Berry and Lindgren 1996). Thus, between 9,000 and 29,000 simulation runs were performed. The simulation programs were written in C and implemented in the statistical software R (R Development Core Team 2006).

Optimum allocation of test resources and optimization criteria

The optimum allocation of test resources for per-cycle selection gain $(\Delta\widehat{G})$ or the probability of identifying superior genotypes $(\widehat{P}(q))$ as well as their standard deviations ($\mathrm{SD}_{\Delta\widehat{G}}$ and $\mathrm{SD}_{\widehat{P}(q)}$) were estimated by extending the approach of Longin et al. (2006b). The value of *q*, the *q*% best genotypes, considered was 5, 1, and 0.1%. For example, $\widehat{P}(5\%)$ corresponds to the probability that the selected DH lines comprise the fraction of the 5% DH lines with the highest genotypic value of testcrosses in the unselected base population of all crosses considered before parental selection. Additionally, we calculated the average coefficient of coancestry $(\bar{\Theta})$ among the selected DH lines for each allocation (Longin et al. 2009). The allocation of test resources refers to tuples (N_j, L_j) for both stages *j*. It was considered optimum if it maximized the corresponding optimization criterion. The optimum allocation as well as the corresponding optimization criteria are denoted by an asterisk, e.g., $L^*_1, \Delta\widehat{G}^*$. This optimum allocation was obtained for each scenario by a grid search in the space of all admissible resource allocations. Since $\Delta\widehat{G}$ was only estimated with a precision of 0.01, the optimum allocation (N^*_j, L^*_j) was determined such that (1) the number of locations was minimum (Utz 1969; Longin et al. 2006b) and (2) the number of DH lines per cross or $\mathrm{S_1}$ family was minimum among all allocations within a 0.01 drop-off of $\Delta\widehat{G}$, to facilitate the conduct of field trials.

Results

The breeding schemes and selection strategies showed similar relative efficiency for both optimization criteria ΔG

and $P(0.1\%)$. S$_1$TC-DHTC was distinctly superior to DHTC, the superiority being more pronounced for $\widehat{P}(0.1\%)^*$ than for $\Delta\widehat{G}^*$ (Table 2). The optimization criteria, $\widehat{P}^*(5\%)$ and $\widehat{P}^*(1\%)$, showed only minor differences between the two breeding schemes and various selection strategies, because all values were close to unity (data not shown). Consequently, results were presented only for $\Delta\widehat{G}^*$ and $\widehat{P}(0.1\%)^*$. The highest values of both optimization criteria were observed for S$_1$TC-DHTC in strategy 2c, and for DHTC in strategy 2a. For $\Delta\widehat{G}^*$ in S$_1$TC-DHTC–2c, the optimum allocation was 12 (L_1^*) and 15 (L_2^*) test locations, four crosses ($N_{1_C}^*$) and 153 S$_1$ families ($N_{1_{F/C}}^*$) in each cross at the first stage, and 764 DH lines (N_2^*) at the second stage. In DHTC–2a, N_1^* was larger, while N_2^* and L_1^* were smaller than the corresponding numbers for S$_1$TC-DHTC–2c. For $\widehat{P}(0.1\%)^*$ in S$_1$TC-DHTC–2c, $N_{1_C}^*$ was about 20% larger and L_1^* was smaller in comparison with $\Delta\widehat{G}^*$. In DHTC–2a, the optimum allocation was similar for both optimization criteria. The standard deviation, $\mathrm{SD}_{\Delta\widehat{G}^*}$ was larger in S$_1$TC-DHTC compared with DHTC, but $\mathrm{SD}_{\widehat{P}(0.1\%)^*}$ was smaller. The average coancestry coefficient $\bar{\Theta}$ was 50–100% larger for both optimization criteria in S$_1$TC-DHTC than in DHTC, but differed only slightly between the optimization criteria.

For both optimization criteria, with an increasing number of crosses in the first stage (N_{1_C}), selection response increased up to 3–4 crosses and decreased thereafter in all selection strategies (Fig. 1). Deviations from the optimum N_{1_C} led to a smaller decrease in S$_1$TC-DHTC than in DHTC for $\Delta\widehat{G}$. For S$_1$TC-DHTC, response curves were almost flat in the vicinity of the maximum. Furthermore, differences for the optimization criteria among the selection strategies increased with increasing N_{1_C} in DHTC.

The effect of varying budgets, variance component ratios and ρ_P on the optimization criteria and the optimum allocation of test resources are presented for the best selection strategy in both breeding schemes, namely

Table 2 Optimum allocation of test resources maximizing the optimization criteria (OC), selection gain ($\Delta\widehat{G}^*$) and the probability of selecting superior genotypes $\widehat{P}(0.1\%)^*$, in two-stage selection with evaluation of testcross progenies of (1) DH lines at both stages (DHTC) and (2) S$_1$ families at first stage and DH lines of S$_1$ families at second stage (S$_1$TC-DHTC)

Breeding scheme/Selection strategy	Optimum allocation				OC	SD$_{OC}$	$\bar{\Theta}$
	N_1^{*a}	N_2^{*b}	L_1^*	L_2^*			
Optimization criterion $\boldsymbol{\Delta\widehat{G}^*}$							
DHTC–1	5,538 = 3 × 1,846	392 = 2 × 196	2	14	3.322	0.385	0.500
DHTC–2a	5,812 = 4 × 1,453	320	2	15	3.384	0.324	0.311
DHTC–2b	6,152	330	2	12	3.335	0.340	0.324
DHTC–2c	4,704	203	3	14	3.347	0.326	0.323
S$_1$TC-DHTC–1	648 = 4 × 162	694 = 2 × 1 × 347	13	15	3.916	0.516	0.750
S$_1$TC-DHTC–2a	834 = 3 × 278	752 = 2 × 1 × 376	9	15	3.916	0.485	0.663
S$_1$TC-DHTC–2b	708 = 4 × 177	735	11	15	3.920	0.496	0.623
S$_1$TC-DHTC–2c	612 = 4 × 153	764	12	15	3.941	0.483	0.619
Optimization criterion $\boldsymbol{\widehat{P}(0.1\%)^*}$							
DHTC–1	5, 655 = 3 × 1, 885	400 = 2 × 200	2	13	0.631	0.295	0.500
DHTC–2a	5, 644 = 4 × 1, 411	348	2	15	0.671	0.252	0.316
DHTC–2b	6,204	320	2	12	0.640	0.258	0.322
DHTC–2c	4,640	220	3	14	0.651	0.258	0.322
S$_1$TC-DHTC–1	760 = 4 × 190	600 = 2 × 1 × 300	13	15	0.906	0.230	0.750
S$_1$TC-DHTC–2a	1023 = 3 × 341	756 = 2 × 2 × 189	8	14	0.905	0.208	0.615
S$_1$TC-DHTC–2b	792 = 6 × 132	675	11	15	0.908	0.210	0.616
S$_1$TC-DHTC–2c	752 = 4 × 188	752	10	15	0.915	0.200	0.619

Assumptions: a budget of 20,000 testcross plot equivalents, variance component ratios VC2, and a correlation $\rho_P = 0.71$ between the mean performance of the parental lines and the mean genotypic value of the testcross performance of their progeny

N_j^* = optimum number of test candidates in stage j, L_j^* = optimum number of test locations in stage j, SD = the standard deviation, and $\bar{\Theta}$ = the average coefficient of coancestry among the selected DH lines

[a] DHTC–1 and 2a: number of crosses × DH lines within crosses, DHTC–2b and 2c: the number of DH lines within crosses depended on the rank of the cross; S$_1$TC-DHTC: number of crosses × S$_1$ families within crosses

[b] DHTC–1: number of crosses × DH lines within crosses, DHTC–2: number of DH lines; S$_1$TC-DHTC–1 and 2a: number of crosses × S$_1$ families within crosses × DH lines within S$_1$ families; S$_1$TC-DHTC–2b and 2c: the number of S$_1$ families within crosses and DH lines within S$_1$ families depended on the rank of the cross and the S$_1$ family

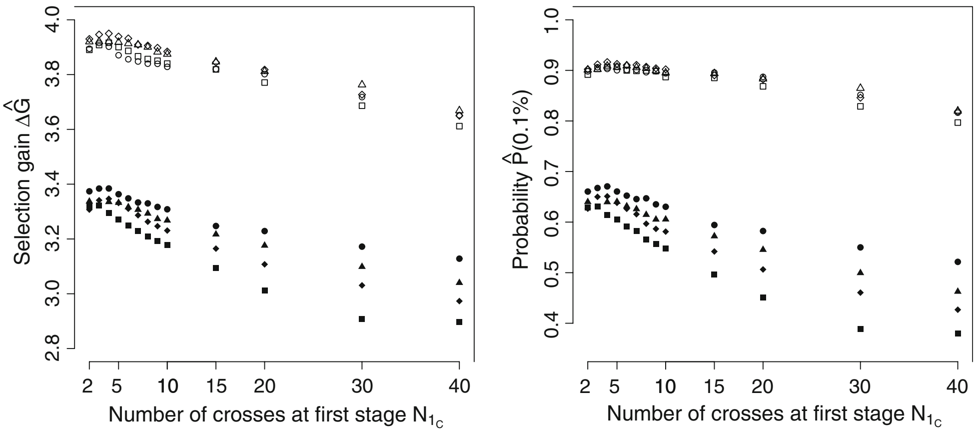

Fig. 1 Selection gain ($\Delta\widehat{G}$) and the probability of selecting superior genotypes ($\widehat{P}(0.1\%)$) as a function of the number of crosses in the first stage (N_{1_C}) for selection strategies 1 (*open square*), 2a (*open circle*), 2b (*open triangle*) and 2c (*open diamond*) in breeding scheme DHTC (*solid symbols*) and S_1TC-DHTC (*hollow symbols*)

DHTC–2a and S_1TC-DHTC–2c (Table 3). An increase in the budget (from 10,000 to 40,000 plot equivalents) resulted in higher values of N_1^* and N_2^*, and increased optimization criteria in both breeding schemes. An increase in the non-genetic variances (from VC1 to VC3) generally reduced N_1^* and N_2^* values, with the exception of N_2^* in S_1TC-DHTC-2c. Furthermore, the rise in non-genetic variances decreased both optimization criteria, but caused an increase in L_j^*. A reduction in ρ_P (from 0.71 to 0.50) resulted in increased $N_{1_C}^*$ values for both breeding schemes and both optimization criteria. On the other hand, the reduction in ρ_P did not result in considerable changes in N_2^* values, though they were increased for $\Delta\widehat{G}$ and decreased for $\widehat{P}(0.1\%)^*$. However, the reduction in ρ_P affected the changes in optimization criteria, by reducing them in case of DHTC–2a and by improving them in case of S_1TC-DHTC–2c.

Discussion

Our results on the optimum allocation of test resources and estimates of ΔG considering breeding schemes S_1TC-DHTC and DHTC were in conformity with the earlier studies of Longin et al.(2007) and Wegenast et al. (2008). In particular, S_1TC-DHTC was superior to DHTC for $\Delta\widehat{G}$. The optimum number of crosses at the first stage ($N_{1_C}^*$) increased with an increasing budget and ρ_P and decreasing non-genetic variances; however, in all situations of both breeding schemes, $N_{1_C}^* \leq 12$. A larger proportion of the test resources was allocated to the second selection stage in S_1TC-DHTC in comparison with DHTC. The production cost for an S_1 family K_F hardly influenced allocation of the budget to the selection stages and the estimates of both optimization criteria (data not shown). An increasing budget or ρ_P had a larger impact on $\Delta\widehat{G}^*$ in DHTC than in S_1TC-DHTC. In DHTC, deviations from $N_{1_C}^*$ led to a larger reduction in ΔG in comparison with S_1TC-DHTC.

Comparison of the breeding schemes

As in the case of ΔG, breeding scheme S_1TC-DHTC was superior to DHTC for $\widehat{P}(0.1\%)$, considering non-hierarchical selection, and variable cross and family sizes (Tables 2 and 3; Fig. 1). A comparison of the respective selection strategies in the breeding schemes showed S_1TC-DHTC to have more than 15% higher ΔG^* and 34% higher $\widehat{P}(0.1\%)^*$ (Table 2). However, the longer cycle length of S_1TC-DHTC than DHTC was not considered in the present study, because per-year estimation cannot be made for $P(q)$. The higher $SD_{\Delta\widehat{G}^*}$ in S_1TC-DHTC than in DHTC (Tables 2, 3) was mainly attributable to a larger SD at the first selection stage (data not shown). In consequence, the deployment of S_1TC-DHTC instead of DHTC offers the chance of having a larger mean ΔG^*, but a larger variation in these estimates.

The selfing of the S_0 generation in S_1TC-DHTC led to larger average coancestry coefficient $\bar{\Theta}$ among the selected candidates in comparison with DHTC (Tables 2, 3). A larger $\bar{\Theta}$ indicates a reduced genetic variance among the selected candidates, leading to a reduced ΔG in a long-term recurrent selection program. On the other hand, the S_1 development offers an additional generation for recombinations, thereby increasing the genetic variance, which results in larger response to long-term selection (Bernardo 2009) The present study focused on the identification of DH lines for the development of hybrids. Thus, it is recommended to employ S_1TC-DHTC. For recurrent selection, DH lines from additional crosses could be selected and intermated for the next breeding cycle to ensure a long-term breeding success.

Comparison of selection strategies

We compared two selection strategies: (1) selection first among and then within crosses in DHTC–1 and additionally within S_1 families in S_1TC-DHTC–1 and (2) selection

Table 3 Optimum allocation of test resources maximizing the optimization criteria (OC), selection gain ($\Delta\widehat{G}^*$) and the probability of selecting superior genotypes ($\widehat{P}(0.1\%)^*$), in two-stage selection with evaluation of testcross progenies of (1) DH lines at both stages (DHTC–2a) and (2) S_1 families at first stage and DH lines of S_1 families at second stage (S_1TC-DHTC–2c) and its dependence on the phenotypic correlation ρ_P (between the mean performance of the parents and the mean genotypic value of the testcross performance of their progenies), the variance component ratios (VC), and the budget in terms of testcross plot equivalents

Breeding scheme/Selection strategy	Assumptions			Optimum allocation				OC	SD_{OC}	$\bar{\Theta}$
	Budget	VC	ρ_P	N_1^{*a}	N_2^*	L_1^*	L_2^*			
Optimization criterion $\boldsymbol{\Delta\widehat{G}^*}$										
DHTC–2a	10,000	2	0.71	2,874 = 3 × 958	191	2	13	3.257	0.316	0.340
DHTC–2a	20,000	2	0.71	5,812 = 4 × 1,453	320	2	15	3.384	0.324	0.311
DHTC–2a	40,000	2	0.71	9,325 = 5 × 1,865	427	3	14	3.491	0.322	0.296
DHTC–2a	20,000	1	0.71	9,556 = 4 × 2,389	390	1	13	3.662	0.309	0.318
DHTC–2a	20,000	3	0.71	3,564 = 2 × 1,782	232	4	14	3.089	0.317	0.399
DHTC–2a	20,000	2	0.50	6,069 = 7 × 867	297	2	14	3.070	0.369	0.268
S_1TC-DHTC–2c	10,000	2	0.71	219 = 3 × 73	436	13	15	3.746	0.496	0.657
S_1TC-DHTC–2c	20,000	2	0.71	612 = 4 × 153	764	12	15	3.941	0.483	0.619
S_1TC-DHTC–2c	40,000	2	0.71	1,482 = 6 × 247	1,040	15	15	4.115	0.508	0.665
S_1TC-DHTC–2c	20,000	1	0.71	845 = 5 × 169	858	11	11	4.330	0.472	0.667
S_1TC-DHTC–2c	20,000	3	0.71	582 = 3 × 194	712	14	15	3.430	0.514	0.609
S_1TC-DHTC–2c	20,000	2	0.50	610 = 10 × 61	726	13	15	3.666	0.563	0.667
Optimization criterion $\boldsymbol{\widehat{P}(0.1\%)^*}$										
DHTC–2a	10,000	2	0.71	3,153 = 3 × 1,051	149	2	12	0.595	0.273	0.338
DHTC–2a	20,000	2	0.71	5,644 = 4 × 1,411	348	2	15	0.671	0.252	0.316
DHTC–2a	40,000	2	0.71	9,375 = 5 × 1,875	447	3	13	0.730	0.225	0.296
DHTC–2a	20,000	1	0.71	9,728 = 4 × 2,432	371	1	13	0.833	0.197	0.315
DHTC–2a	20,000	3	0.71	3,222 = 2 × 1,611	342	4	14	0.500	0.284	0.397
DHTC–2a	20,000	2	0.50	6,216 = 7 × 888	293	2	13	0.479	0.271	0.271
S_1TC-DHTC–2c	10,000	2	0.71	355 = 5 × 71	344	12	15	0.848	0.272	0.654
S_1TC-DHTC–2c	20,000	2	0.71	752 = 4 × 188	752	10	15	0.915	0.200	0.619
S_1TC-DHTC–2c	40,000	2	0.71	1,729 = 7 × 247	1,094	13	14	0.951	0.156	0.665
S_1TC-DHTC–2c	20,000	1	0.71	1,314 = 6 × 219	406	11	11	0.981	0.105	0.655
S_1TC-DHTC–2c	20,000	3	0.71	477 = 3 × 159	896	13	14	0.670	0.321	0.610
S_1TC-DHTC–2c	20,000	2	0.50	627 = 11 × 57	848	11	14	0.808	0.306	0.667

N_j^* = optimum number of test candidates in stage j, L_j^* = optimum number of test locations in stage j, SD = the standard deviation, and $\bar{\Theta}$ = the average coefficient of coancestry among the selected DH lines

[a] DHTC–2a: number of crosses × DH lines within crosses; S_1TC-DHTC–2c: number of crosses × S_1 families within crosses

among DH lines disregarding the cross and family structure in DHTC–2 and S_1TC-DHTC–2. Normally, not all candidates within the topmost cross are superior to the best candidates of other crosses (cf., Lush 1947), because the distributions of the test candidates of different crosses overlap. In the present study, in selection strategy 2, superior DH lines from crosses or S_1 families irrespective of parental performance were selected; thus, no superior genotype was rejected due to the mean performance of its cross or S_1 family, improving $\Delta\widehat{G}^*$ and $\widehat{P}(0.1\%)^*$ (Table 2; Fig. 1).

In S_1TC-DHTC, selection strategy 2c with variable size of crosses and S_1 families was superior and achieved higher $\Delta\widehat{G}^*$ and $\widehat{P}(0.1\%)^*$, and both criteria had SD_{OC} and $\bar{\Theta}$ equal or lower than the other selection strategies (Table 2; Fig. 1). In strategy 2c, at least half of the test candidates belonged to the best cross after the first selection stage and one quarter of the test candidates to the second best cross (Table 1; Supplementary Fig. S1). In strategy 2b, however, only 7/25 of the test candidates belonged to the best and about 1/6 to the second best cross after the first selection stage. Thus, it was worth to have variable sizes of crosses and S_1 families and devote a large budget to the better cross. Positive effects of variable sizes of families on ΔG were also observed in animal breeding (Toro and Nieto 1984; Toro et al. 1988; Toro and Pérez-Encisco 1990). In DHTC, the application of variable sizes of crosses did not improve the values for the optimization criteria. The dimension of the crosses was calculated before the first selection stage based solely on parental information. Therefore, a more effective use of the parental cross and family information, e.g., using best linear unbiased prediction (BLUP, Bernardo 1996) might further improve the use of variable cross and family sizes.

Considering the best selection strategy in each breeding scheme, the relative superiority of S$_1$TC-DHTC–2c over DHTC–2a increased with a decreasing budget and ρ_P for both optimization criteria (Table 3). On comparing S$_1$TC-DHTC–2c with DHTC–2a for varying variance component ratios, opposite trends were found for $\Delta\widehat{G}^*$ and $\widehat{P}(0.1\%)^*$: smaller non-genetic variances led to an increase in the relative superiority of S$_1$TC-DHTC–2c over DHTC–2a for $\Delta\widehat{G}^*$, but a decreasing relative superiority for $\widehat{P}(0.1\%)^*$. Most likely, the low impact of small non-genetic variances on $\widehat{P}(0.1\%)^*$ in S$_1$TC-DHTC was due to the fact that for $\widehat{P}(0.1\%)^*$, the values were already very high and, consequently, there was little scope for improvement.

The reason for the lower SD$_{OC}$ in strategy 2 than in strategy 1 in S$_1$TC-DHTC (Table 2) might be that progress from selection is less variable when selection in the second stage is performed among all DH lines and not only among the DH lines of selected crosses and S$_1$ families. In DHTC, the decrease in SD$_{\Delta\widehat{G}^*}$ in selection strategy 2 compared with strategy 1 was stronger than in S$_1$TC-DHTC. This may be attributable to the non-hierarchical selection among all DH lines in both selection stages.

The lower (11%) value of $\bar{\Theta}$ in S$_1$TC-DHTC–2, as compared with S$_1$TC-DHTC–1 (Table 2), is attributable to the final selection of all DH lines of the best S$_1$ family within the best cross in strategy 1. On the other hand, in S$_1$TC-DHTC–2, the finally selected DH lines originate mostly from more than one S$_1$ family and cross (data not shown). In animal breeding, variable family sizes also led to a decrease in the inbreeding coefficient (Toro and Nieto 1984.) If DH lines in S$_1$TC-DHTC–1 were finally selected out of more than one cross or S$_1$ family within crosses, $\bar{\Theta}$ could be reduced. However, this will reduce $\Delta\widehat{G}^*$ (Wegenast et al. 2008).

Comparison of the optimization criteria

The estimate of $\Delta\widehat{G}$ reflects the superiority of the population generated by intermating the selected genotypes in comparison with the genotypic mean of the base population, whereas $\widehat{P}(q)$ quantifies the chance to develop superior varieties without reference to the mean of the whole selected group (Wricke and Weber 1986). To have a realistic chance of success in identifying a superior genotype, $P(q)$ should be greater than 0.75 (Longin et al. 2006a, b), which can be achieved in both breeding schemes for $\widehat{P}(5\%)^*$ and $\widehat{P}(1\%)^*$ (data not shown). However, the probability of having the top 0.1% genotypes under evaluation, actually included in the selected fraction, i.e., $P(0.1\%)$, is distinctly higher with S$_1$TC-DHTC than with DHTC (Tables 2, 3). The optimization criteria did not affect the ranking of the breeding schemes or selection strategies. However, the differences among the breeding schemes, selection strategies, as well as the assumptions concerning the total budget and the variance component ratios, were relatively more pronounced for $\widehat{P}(0.1\%)^*$ than for $\Delta\widehat{G}^*$.

The relative decrease in the optimization criteria due to non-optimum allocation in both breeding schemes was larger for $\widehat{P}(0.1\%)$ than for $\Delta\widehat{G}$ (Fig. 1). The higher sensitivity of $\widehat{P}(0.1\%)$ to non-optimal allocation in comparison with $\Delta\widehat{G}$ might be due to the binomial character of $P(q)$: for the calculation of $\widehat{P}(q)$, selected genotypes that belong to the upper 0.1% quantile are recorded as 1 and those that are below the 0.1% quantile are recorded as 0. For the calculation of $\Delta\widehat{G}$, however, the continuously distributed phenotypic values of the genotypes are used.

The binomial nature of $\widehat{P}(q)$ with genotypes surpassing the defined threshhold or not also influences its SD. The values of $SD_{\widehat{P}(q)}$ assume their maximum value for $\widehat{P}(q) = 0.5$ and their minimum for $\widehat{P}(q) = 0$ and $\widehat{P}(q) = 1$ (Longin et al. 2006b). This explains the considerably lower $SD_{\widehat{P}(0.1\%)^*}$ in S$_1$TC-DHTC in comparison with DHTC.

Optimum allocation of test resources

An increase in the number of test candidates in the first stage (N_1^*) at the expense of a reduced optimum number of test candidates in the second stage (N_2^*) enhanced $\widehat{P}(0.1\%)^*$ in S$_1$TC-DHTC (Tables 2, 3). This is consistent with results of previous studies (Robson et al. 1967; Johnson 1989; Knapp 1998). In DHTC, however, N_j^* was hardly influenced by the choice of the optimization criteria. This might be due to the very high N_1^*; thus, a further increase in N_1^* would not have any additional positive effect on $\widehat{P}(0.1\%)^*$.

In S$_1$TC-DHTC, selection strategy 2 led to a decrease in L_1^* at the expense of an increase in N_2^* in comparison with strategy 1 (Table 2). The reason for the increase in N_2^* being that an increase in the total number of genotypes enhances the chance to select superior DH genotypes irrespective of the cross performance. In DHTC–2, N_2^* decreased in favor of an increased N_1^* or L_1^* in comparison with DHTC–1. However, the allocation of the test resources to the first and second selection stage was not affected by the selection strategy. Furthermore, in DHTC–2, selection among all DH lines disregarding the cross structure was also applied in the first stage; thus, a shift of the budget from the first to the second stage did not result in more effective selection.

Response curves of S_1TC-DHTC were flat in the vicinity of the maximum (Fig. 1): for all strategies, when the number of crosses in the first stage (N_{1_C}) did not exceed 15, and the decrease in the value of the optimization criteria was less than 4% in comparison with the maximum in S_1TC-DHTC. For these calculations, the allocation of all other factors except N_{1_C} was optimized. The reason for the lower sensitivity of S_1TC-DHTC to a non-optimal allocation might be that a loss in response to selection among parents is compensated by an increase in the response to selection among S_1 families in the first stage of selection without reducing the selection intensity in the second stage of selection.

In S_1TC-DHTC–2, a larger part of the resources was allocated to the second selection stage as compared with S_1TC-DHTC–1 (Table 2). The reason for this additional shifting of the test resources to the second stage in S_1TC-DHTC–2 might be that the topmost DH lines disregarding their cross and family structure, were selected and, it proved beneficial to test these lines more extensively.

Conclusions

Breeding scheme S_1TC-DHTC had larger $\Delta\widehat{G}^*$ and $\widehat{P}(0.1\%)^*$ than DHTC. The superiority of S_1TC-DHTC was further enhanced when selection was done among all DH lines disregarding their cross and family structure, as well as the use of variable instead of uniform sizes of crosses and S_1 families. With variable sizes of crosses and S_1 families, the allocation of a large fraction of the budget to the crosses on top after the first selection stage was the superior strategy. Although the ranking was not altered with the use of ΔG^* or $P(0.1\%)^*$, differences between breeding schemes were higher for $P(0.1\%)^*$. Thus, $P(q)$ offers a very sensitive tool to differentiate among breeding schemes and selection strategies. Further investigations are warranted to examine whether selection progress under various breeding schemes and selection strategies can be upgraded by giving optimal weights to parental information using BLUP approaches.

Acknowledgments This research was supported by funds from DFG, Grant No 1070, International Research Training Group "Sustainable Resource Use in North China" to T. Wegenast. The authors appreciate the editorial work of Dr. J. Muminović, whose suggestions considerably improved the style of the manuscript. We greatly appreciate the helpful comments and suggestions of the anonymous reviewers.

References

Baker RJ (1984) Quantitative genetic principles in plant breeding. In: Gustafson JP (ed) Gene manipulation in plant improvement. Plenum Press, New York, pp 147–176

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, pp 152–156

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

Berry DA, Lindgren BW (1996) Statistics. Duxberry Press, New York, p 371

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman Scientific and Technical Ltd, Essex, England, pp 184–194, 230–231

Gordillo GA, Geiger HH (2008) Alternative recurrent selection strategies using doubled haploid lines in hybrid maize breeding. Crop Sci 48:911–922

Johnson B (1989) The probability of selecting genetically superior S_2 lines from a maize population. Maydica 34:5–14

Keuls M, Sieben JW (1955) Two statistical problems in plant selection. Euphytica 4:34–44

Knapp SJ (1998) Marker-assisted selection as a strategy for increasing the probability of selecting superior genotypes. Crop Sci 38:1164–1174

Longin CFH, Maurer HP, Melchinger AE, Frisch M (2009) Optimum allocation of test resources and relative efficiency of alternative procedures of within-family selection in hybrid breeding. Plant Breed 128:213–216

Longin CFH, Utz HF, Melchinger AE, Reif JC (2006a) Hybrid maize breeding with doubled haploids: comparison between selection criteria. Acta Agron Hung 54:343–350

Longin CFH, Utz HF, Reif JC, Schipprack W, Melchinger AE (2006b) Hybrid maize breeding with doubled haploids: I. One-stage versus two-stage selection for testcross performance. Theor Appl Genet 112:903–912

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

Lush JL (1947) Family merit and individual merit as bases for selection. Part 1. Am Nat 81:241–261

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

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, http://www.R-project.org

Robson DS, Powers L, Urquhart NS (1967) The proportion of genetic deviates in the tails of a normal population. Genet Breed Res 37:205–216

Schnell FW (1982) A synoptic study of the methods and categories of plant breeding. Z Pflanzenzuecht 89:1–18

Toro MA, Nieto BM (1984) A simple method for increasing the response to artificial selection. Genet Res Camb 44:347–349

Toro MA, Nieto B, Salgado C (1988) A note on minimization of inbreeding in small-scale selection programmes. Livest Prod Sci 20:317-323

Toro M, Pérez-Encisco M (1990) Optimization of selection response under restricted inbreeding. Genet Sel Evol 22:93–107

Utz HF (1969) Mehrstufenselektion in der Pflanzenzüchtung. (In German) Arbeiten der Universität Hohenheim, vol 49. Verlag Eugen Ulmer, Stuttgart, Germany

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

Wricke G, Weber WE (1986) Quantitative genetics and selection in plant breeding. Walter de Gruyter, Berlin, New York, pp 172–194