Theor Appl Genet (2008) 117:251–260
DOI 10.1007/s00122-008-0770-y

ORIGINAL PAPER

Hybrid maize breeding with doubled haploids. IV. Number versus size of crosses and importance of parental selection in two-stage selection for testcross performance

Thilo Wegenast · C. Friedrich H. Longin ·
H. Friedrich Utz · Albrecht E. Melchinger ·
Hans Peter Maurer · Jochen C. Reif

Received: 10 October 2007 / Accepted: 8 April 2008 / Published online: 26 April 2008

Abstract Parental selection influences the gain from selection and the optimum allocation of test resources in breeding programs. We compared two hybrid maize (*Zea mays* L.) breeding schemes with evaluation of testcross progenies: (a) doubled haploid (DH) lines in both stages (DHTC) and (b) S_1 families in the first stage and DH lines within S_1 families in the second stage (S_1TC-DHTC). Our objectives were to (1) determine the optimum allocation regarding the number of crosses, S_1 families, DH lines, and test locations, (2) investigate the impact of parental selection on the optimum allocation and selection gain (ΔG), and (3) compare the maximum ΔG achievable with each breeding scheme. Selection gain was calculated by numerical integration. Different assumptions were made regarding the budget, variance components, correlation between the mean phenotypic performance of the parents and the mean genotypic value of the testcross performance of their progenies (ρ_P), and the composition of the finally selected test candidates. In comparison with randomly chosen crosses, maximum ΔG was largely increased with parental selection in both breeding schemes. With an increasing correlation ρ_P, this superiority increased strongly, while the optimum number of crosses decreased in favor of an increased number of test candidates within crosses. Thus, concentration on few crosses among the best parental lines might be a promising approach for short-term success in advanced cycle breeding. Breeding scheme S_1TC-DHTC led to a larger ΔG but had a longer cycle length than DHTC. However, with further improvements in the DH technique and the realization of more than two generations per year, early testing of S_1 families prior to production of DH lines would become very attractive in hybrid maize breeding.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-008-0770-y) contains supplementary material, which is available to authorized users.

T. Wegenast and C. F. H. Longin contributed equally to this work.

Communicated by H. C. Becker.

T. Wegenast · C. F. H. Longin · H. F. Utz ·
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

H. P. Maurer · J. C. Reif
State Plant Breeding Institute, University of Hohenheim,
70593 Stuttgart, Germany

Introduction

In hybrid maize breeding, new genetic variation is generally generated by crossing two or more elite inbred lines within each heterotic group (Bernardo 2002). From the S_0 plants of these crosses, doubled haploid (DH) lines may be developed by in vivo haploid induction (Schmidt 2004; Seitz 2005). Owing to the finite financial resources in a breeding program, a plant breeder must find a compromise between (1) the number of crosses, (2) the number of test candidates within each cross, and (3) the intensity of their testing as determined by the number of test locations, years, and replications. Selection among crosses enables breeders to discard inferior crosses in early stages and to concentrate their resources on selection within the most promising crosses (cf., Schnell 1982).

Selection gain (ΔG) is the commonly used criterion to investigate the optimum allocation of test resources for maximizing progress from selection (Baker 1984; Wricke and Weber 1986; Hühn 1996, 2005a, b, 2006; Bernardo 2003). For one-stage selection among and within crosses and resources for testing 2,000 candidates, ΔG was maximum for 50–100 crosses with 20–40 lines within each cross (Baker 1984; Hühn 1996). With increasing budget, costs of line development, heritability, and number of selected crosses, the number of crosses increased at the expense of a reduced number of test candidates within crosses (Hühn 1996; 2005a, b, 2006). The above studies assumed selection among randomly chosen crosses and did not consider parental selection.

The mean performance of a cross can be predicted by the average performance of its parental lines (cf., Choo et al. 1979; Wricke and Weber 1986). The accuracy of this prediction depends on the genetic correlation between the mean performance of the parents and the mean performance of their progeny. Assuming absence of epistasis, this correlation equals to the square root of the heritability of the parental performance. In advanced cycle breeding, the performance of parental lines is well-known from earlier breeding cycles (Bernardo 2003). Thus, breeders can estimate the potential of a cross before testing its progeny extensively.

For wheat (*Triticum aestivum* L.) breeding, parental selection applying best cross combinations of selected parental lines led to a largely increased ΔG compared with crossing randomly chosen parents (Utz 1982). The optimum allocation comprised a small number of crosses and a large number of test candidates within crosses. While Utz (1982) focused on selfing generations in pedigree selection, breeding schemes involving DH lines were not considered and basic assumptions, such as underlying variance components and the budget, differed largely from those in maize breeding.

We calculated the maximum ΔG by numerical integration to optimize the allocation of test resources in hybrid maize breeding with DH lines. Two-stage selection schemes were considered with evaluation of testcross progenies: (1) DH lines in both stages and (2) S_1 families in the first and DH lines within S_1 families in the second selection stage. Different assumptions were made regarding the budget, variance components (VC), correlation between the mean phenotypic performance of the parents and the mean genotypic value of the testcross performance of their progenies (ρ_P), and composition of finally selected test candidates (N_f). Our objectives were to (1) determine the optimum allocation regarding the number of crosses, S_1 families, DH lines, and test locations, (2) investigate the impact of parental selection on the optimum allocation and ΔG, and (3) compare the maximum ΔG achievable with each of the two breeding schemes.

Materials and methods

Breeding schemes

We extended two two-stage breeding schemes on optimum allocation of test resources for a single cross (Longin et al. 2007b) to selection among and within several crosses (Supplementary Figure S1). In breeding scheme DHTC, DH lines were produced by in vivo haploid induction from S_0 plants before the first selection stage. In breeding scheme S_1TC-DHTC, early testing for testcross performance of S_1 families was made and remnant seed was used for a simultaneous in vivo haploid induction of these S_1 families. Chromosome doubling was only performed with haploid kernels produced in selected S_1 families.

In both stages of each breeding scheme, selection was first made among and then within crosses. The target variable throughout this treatise is the genotypic value of the testcross performance for grain yield with a given tester. At a given selection stage j, selection among N_j test candidates was based on the phenotypic mean of testcross performance at this stage with the tester, evaluated at L_j test locations. Without restrictions on L_j in stage j, ΔG is maximum for one replication per test location (Sprague and Federer 1951; Utz 1969; Melchinger et al. 2005). Thus, we set the number of replications to one for all calculations.

After two stages of selection, the best $N_f = 10$ DH lines were selected. To assess the optimum composition of finally selected test candidates (N_f), we investigated all possible integer combinations of test candidates for $N_f = 10$, i.e., the number of finally selected crosses × DH lines within crosses in DHTC and the number of finally selected crosses × S_1 families within crosses × DH lines within S_1 families in S_1TC-DHTC. An overview of the notation used in this treatise is given in Table 1.

Calculation of selection gain

We calculated ΔG on a per-cycle basis (Longin et al. 2007a, b), using the well-known formula of Cochran (1951). We assumed that (1) parental selection and selection among crosses were independent from selection within crosses, and (2) selection among S_1 families was independent from selection within S_1 families. Calculation of ΔG among and within S_1 families was based on the formulas for among-family and strict within-family selection of Hill et al. (1996). These formulas were extended to selection among and within crosses (Supplementary Table S2). Parental crosses for a new breeding cycle were selected before the start of the breeding program. This parental selection was based on the parental mean $\overline{P}$ of all possible pairwise parental combinations $P_m \times P_n$, because general varietal ability (Wright 1974; Gallais 1979) of a cross $P_m \times P_n$ can

Table 1 Notation used in this treatise

j	Selection stage
α_P	Selected fraction in the parental selection
ρ_G	Genotypic correlation between the mean GCA of the parental lines and the mean GCA of their progenies
ρ_P	Phenotypic correlation between the mean performance of the parental lines and the mean genotypic value of the testcross performance of their progenies
σ^2_C, $\sigma^2_{\mathrm{DH}/C}$, $\sigma^2_{F/C}$, $\sigma^2_{\mathrm{DH}/F}$	Genotypic variances among test candidates, for details see Table 2
DH	Doubled haploid
ΔG	Selection gain in two-stage selection
ΔG*	Value of ΔG at the corresponding optimum allocation of L_j*, N_j*
L_j, N_j,	Number of test locations and test candidates in stage j in performance trials
L_j*, N_j*	Optimum number of test locations and test candidates in stage j in performance trials, maximizing ΔG in the set of admissible allocations
N_f	Composition of finally selected test candidates, for details see Table 4
$N_{j_C}, N_{j_{\mathrm{DH}/C}}, N_{j_{F/C}}, N_{j_{\mathrm{DH}/F}}$	Number of crosses, DH lines within crosses, S$_1$ families within crosses, and DH lines within S$_1$ families in stage j in performance trials
VC	Variance components, for details see Table 2

be predicted by the mean of the general combining abilities (GCA) of its parents P_m and P_n, This enables a reduction of the number of crosses, but reduces the genotypic variance among testcross means of crosses ($\sigma_C{}^2$) to (Cochran 1951)

$$\sigma^{2\prime}_C = \sigma^2_C\left[1 - \rho^2_P i_{\alpha_P}(i_{\alpha_P} - k_P)\right], \tag{1}$$

where $\rho_P{}^2$ is the squared correlation coefficient between the mean phenotypic testcross performance of the parental lines and the mean genotypic value of testcross progeny from their cross, and i_{α_P} and k_P are the selection intensity and truncation point of the normal distribution for the selection based on parental means. We assumed $\rho^2_P = 0.25$, 0.5, and 0.75, which covers the range expected for a quantitative trait like grain yield. In addition, we investigated $\rho^2_P = 0$ to consider also the case of randomly chosen parents.

In both stages of DHTC, selection among crosses was based on the phenotypic testcross mean of all DH lines from the corresponding cross. Selection among DH lines within the selected crosses was based on the phenotypic testcross mean of the particular DH line evaluated at L_j test locations. In the second stage of DHTC, an optimum index of the phenotypic means of the test candidates evaluated in both stages was used as the selection criterion. The independence of selection among and within crosses requires separate optimum indices for selection among and within crosses. Selection gain was calculated according to Utz (1982) as

$$\varDelta G = \sigma\left(\frac{\rho_P o_P}{\alpha_P} + \frac{\sum_{j=1}^{2}\rho_{j_C} o_{j_C} J_{j_C}}{\alpha_{1_C}\alpha_{2_C}} + \frac{\sum_{j=1}^{2}\rho_{j_{\mathrm{DH}/C}} o_{j_{\mathrm{DH}/C}} J_{j_{\mathrm{DH}/C}}}{\alpha_{1_{\mathrm{DH}/C}}\alpha_{2_{\mathrm{DH}/C}}}\right), \tag{2}$$

where σ is the standard deviation of the target variable, o_P the ordinate of the univariate normal distribution at the truncation point of the parental selection, α_j the selected fraction in stage j (i.e., the ratio of selected by tested candidates), ρ_j the correlation between the phenotypic mean of testcross performance in stage j and the target variable, o$_j$ the ordinate of the univariate normal distribution at the truncation point of selection stage j, and J_j the convergent improper integral of the standardized bivariate normal distribution in selection stage j. The indices P, C and DH/C refer to the selection among parental means, crosses and DH lines within crosses.

In both stages of S$_1$TC-DHTC, selection among crosses was based on the phenotypic testcross means of the S$_1$ families of the corresponding cross. In the first stage, selection among S$_1$ families within selected crosses was based on the phenotypic testcross mean of the respective S$_1$ families evaluated at L_1 test locations. In the second stage, selection among S$_1$ families within selected crosses was based on the phenotypic testcross mean of all DH lines of the corresponding S$_1$ family. Selection among DH lines within selected S$_1$ families was based on the phenotypic testcross mean of the respective DH lines evaluated at L$_2$ test locations. Thereby, selection among crosses and S$_1$ families within crosses in the second stage was based on two optimum indices combining the respective phenotypic testcross means of both stages. Selection gain was calculated as

$$\varDelta G = \sigma\left(\frac{\rho_P o_P}{\alpha_P} + \frac{\sum_{j=1}^{2}\rho_{j_C} o_{j_C} J_{j_C}}{\alpha_{1_C}\alpha_{2_C}} + \frac{\sum_{j=1}^{2}\rho_{j_{F/C}} o_{j_{F/C}} J_{j_{F/C}}}{\alpha_{1_{F/C}}\alpha_{2_{F/C}}} + \frac{\rho_{x_{2_{\mathrm{DH}/F}}} o_{2_{\mathrm{DH}/F}}}{\alpha_{2_{\mathrm{DH}/F}}}\right). \tag{3}$$

The indices F/C and DH/F refer to the selection among S$_1$ families within crosses and DH lines within S$_1$ families.

For calculation of the selection gain in parental selection, σ_C^2 is used, whereas in selection among crosses, the genetic variance needs to be modified to $\sigma_C^{2\prime}$ (Eqs. 1–3), Supplementary Table S2).

Optimum allocation of resources

The allocation of test resources refers to $(L_1, N_{1_C}, N_{1_{\mathrm{DH}/C}}, L_2, N_{2_C}, N_{2_{\mathrm{DH}/C}})$ for DHTC and to $(L_1, N_{1_C}, N_{1_{F/C}}, L_2, N_{2_C}, N_{2_{F/C}}, N_{2_{\mathrm{DH}/F}})$ for $\mathrm{S_1}$TC-DHTC . The allocation of test resources was considered optimum if it maximized ΔG in the set of all integer allocation combinations feasible for a given scenario, i.e., budget, variance components (VC), and ρ_P. The optimum allocation as well as the corresponding ΔG are denoted by an asterisk, e.g., L_1*, ΔG*.

Economic frame and quantitative-genetic parameters

A fixed total budget for the production of test candidates and evaluation of their testcross progenies in two selection stages was defined in terms of testcross plot equivalents. Equal plot sizes in both selection stages were assumed. In DHTC, the budget equals $N_{1_C} N_{1_{\mathrm{DH}/C}}[K_{\mathrm{DH}} + L_1(1 + K_T)] + N_{2_C} N_{2_{\mathrm{DH}/C}} L_2(1 + K_T)$, where K_{DH} refers to the production costs of one DH line and K_T to the production costs of testcross seed for one plot. In $\mathrm{S_1}$TC-DHTC, the budget equals $N_{1_C} N_{1_{F/C}}[K_F + L_1(1 + K_T)] + N_{2_C} N_{2_{F/C}} N_{2_{\mathrm{DH}/F}}[K_{\mathrm{DH}} + L_2(1 + K_T)]$, where K_F refers to the production costs of one $\mathrm{S_1}$ family. All costs are based on actual costs in the maize breeding program of the University of Hohenheim. We assumed K_{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 each heterotic group of one heterotic pattern.

Three different ratios of VC (Table 2) were chosen based on estimates for grain yield in recent official variety performance tests in Germany including early- and late-maturing germplasm (VC1), DH populations in maize programs of Central Europe breeding companies (VC2), and official maize variety performance tests of early-maturing germplasm in Southwest Germany (VC3, Longin et al. 2007b). In DHTC, the total genotypic variance among testcross progenies of DH lines from different crosses was $\sigma^2 = \sigma_C^2 + \sigma_{\mathrm{DH}/C}^2$, where σ^2 is the genotypic variance of the target variable in Eq. (2) and $\sigma_{\mathrm{DH}/C}^2$ the genotypic variance among testcross progenies of the DH lines within crosses. In the absence of epistasis and linkage disequilibrium, $\sigma_{\mathrm{DH}/C}^2 = \sigma_C^2$ according to quantitative genetic expectations (Melchinger 1988). In $\mathrm{S_1}$TC-DHTC, the total genotypic variance among testcross progenies of DH lines within $\mathrm{S_1}$ families from different crosses was $\sigma^2 = \sigma_C^2 + \sigma_{F/C}^2 + \sigma_{\mathrm{DH}/F}^2$, where σ^2 is the genotypic variance of the target variable in Eq. (3), $\sigma_{F/C}^2$ the genotypic variance among testcross progenies of the $\mathrm{S_1}$ families within crosses, and $\sigma_{\mathrm{DH}/F}^2$ the genotypic variance among testcross progenies of DH lines within $\mathrm{S_1}$ families. In the absence of epistasis and linkage disequilibrium, $\sigma_{\mathrm{DH}/F}^2 = \sigma_{F/C}^2 = 1/2\ \sigma_C^2$ according to quantitative genetic expectations (Melchinger 1988). In both stages of each breeding scheme, we assumed that the ratio of the interaction variances was equal to the ratio of the corresponding genotypic variances. However, the plot error variance was assumed to be constant for testcrosses of all test candidates.

Results

For parameters only marginally affected by varying budgets and variance component (VC) ratios, representative results were shown for intermediate values of the budget (20,000 testcross plot equivalents) and variance components (VC2). In both breeding schemes, maximum selection gain (ΔG*) was largely increased when parental selection was considered ($\rho_P > 0$, Table 3). $\mathrm{S_1}$TC-DHTC was superior over DHTC in ΔG* but the difference in ΔG*

Table 2 Three ratios of variance components (*VC*) were considered in both breeding schemes (DHTC, $\mathrm{S_1}$TC-DHTC), where the total genotypic variance σ^2 is subdivided into the variance among (1) crosses (*C*), (2) DH lines within crosses (DH/*C*), (3) $\mathrm{S_1}$ families within crosses (*F*/*C*), or (4) DH lines within $\mathrm{S_1}$ families (DH/*F*)

Breeding scheme	Variance among	VC1					VC2					VC3				
		σ_G^2	$\sigma_{G\times l}^2$	$\sigma_{G\times y}^2$	$\sigma_{G\times l\times y}^2$	σ_e^2	σ_G^2	$\sigma_{G\times l}^2$	$\sigma_{G\times y}^2$	$\sigma_{G\times l\times y}^2$	σ_e^2	σ_G^2	$\sigma_{G\times l}^2$	$\sigma_{G\times y}^2$	$\sigma_{G\times l\times y}^2$	σ_e^2
DHTC	*C*	1/2	1/8	1/8	1/4	1	1/2	1/4	1/4	1/2	2	1/2	1/2	1/2	1	4
	DH/*C*	1/2	1/8	1/8	1/4	1	1/2	1/4	1/4	1/2	2	1/2	1/2	1/2	1	4
$\mathrm{S_1}$TC-DHTC	*C*	1/2	1/8	1/8	1/4	1	1/2	1/4	1/4	1/2	2	1/2	1/2	1/2	1	4
	F/*C*	1/4	1/16	1/16	1/8	1	1/4	1/8	1/8	1/4	2	1/4	1/4	1/4	1/2	4
	DH/*F*	1/4	1/16	1/16	1/8	1	1/4	1/8	1/8	1/4	2	1/4	1/4	1/4	1/2	4

σ_G^2 refers to the genotypic variance among testcross progenies of the candidates with a given tester, $\sigma_{G\times l}^2$ to the variance of the genotype × location interactions, $\sigma_{G\times y}^2$ to the variance of the genotype × year interactions, $\sigma_{G\times l\times y}^2$ to the variance of the genotype × location × year interactions, and σ_e^2 to the plot error variance. The index *G* refers to the respective test candidates, i.e., *C*, DH/*C*, *F*/*C*, DH/*F*

Table 3 Optimum allocation of test resources maximizing ΔG (ΔG^*) in two-stage selection with evaluation of testcross progenies of (1) DH lines in both stages (breeding scheme DHTC) and (2) S_1 families in the first stage and DH lines within S_1 families in the second stage (breeding scheme S_1TC-DHTC) and its dependence on ρ_P

Restrictions	ρ_P	Optimum allocation				
		N_1^{*a}	N_2^{*b}	L_1*	L_2*	ΔG*
Breeding scheme DHTC						
–	0	5,822 = 41 × 142	208 = 4 × 52	2	23	2.413
–	0.50	4,044 = 4 × 1,011	184 = 2 × 92	3	28	3.191
–	0.71	3,926 = 2 × 1,963	214 = 2 × 107	3	26	3.486
–	0.87	4,025 = 1 × 4,025	174 = 1 × 174	3	30	3.783
$N_{1_C} \leq 100, L_j \leq 15$	0	6,210 = 46 × 135	255 = 5 × 51	2	15	2.405
$N_{1_C} \leq 100, L_j \leq 15$	0.50	4,456 = 4 × 1,114	248 = 2 × 124	3	15	3.145
$N_{1_C} \leq 100, L_j \leq 15$	0.71	4,292 = 2 × 2,146	286 = 2 × 143	3	15	3.472
$N_{1_C} \leq 100, L_j \leq 15$	0.87	4,456 = 1 × 4,456	248 = 1 × 248	3	15	3.762
Breeding scheme S_1TC-DHTC						
–	0	1,310 = 655 × 2	597 = 3 × 1 × 199	8	14	3.138
–	0.50	517 = 11 × 47	638 = 2 × 1 × 319	18	15	3.616
–	0.71	432 = 4 × 108	614 = 2 × 1 × 307	21	16	3.880
–	0.87	404 = 2 × 202	666 = 2 × 1 × 333	22	15	4.111
$N_{1_C} \leq 100, L_j \leq 15$	0	900 = 100 × 9	639 = 3 × 1 × 213	11	14	3.074
$N_{1_C} \leq 100, L_j \leq 15$	0.50	616 = 11 × 56	642 = 2 × 1 × 321	15	15	3.615
$N_{1_C} \leq 100, L_j \leq 15$	0.71	576 = 4 × 144	680 = 2 × 1 × 340	15	15	3.874
$N_{1_C} \leq 100, L_j \leq 15$	0.87	558 = 2 × 279	698 = 2 × 1 × 349	15	15	4.101

Assumptions: a budget of 20,000 testcross plot equivalents, variance components VC2 and a final selection of the $N_f = 10$ best DH lines within the best cross. For explanation of abbreviations, see Table 1

[a] *DHTC* Number of crosses × DH lines within crosses at first stage, *S_1TC-DHTC* Number of crosses × S_1 families within crosses at first stage

[b] *DHTC* Number of crosses × DH lines within crosses at second stage, *S_1TC-DHTC* Number of crosses × S_1 families within crosses × DH lines within S_1 families at second stage

decreased from 30% with randomly chosen parents to 8% with the most effective parental selection ($\rho_P = 0.87$). For $\rho_P = 0.71$, the optimum allocation in S_1TC-DHTC was $L_1^* = 21$ and $L_2^* = 16$ test locations in stage one and two, respectively, $N^*_{1_{F/C}} = 108 S_1$ families in each of the $N^*_{1_C} = 4$ crosses in the first stage as well as $N^*_{2_{\mathrm{DH}/F}} = 307$ DH lines within each $N^*_{2_{F/C}} = 1 S_1$ family within $N^*_{2_C} = 2$ crosses in the second stage. In DHTC, N_1*, L_2* were larger and N_2*, L_1* were smaller in comparison with S_1TC-DHTC. In both breeding schemes, an increase in ρ_P decreased the optimum number of crosses in favor of an increased number of test candidates within crosses in both stages. Restricting N_{1_C} to 100 and L_j to 15 led to an increase in the number of test candidates within crosses but decreased ΔG* only slightly ($\leq$2%).

The impact of varying budgets, VC ratios, and compositions for $N_f = 10$ on ΔG* and the optimum allocation of test resources was hardly affected by ρ_P (data not shown). Thus, results were presented only for $\rho_P = 0.71$ (Table 4). Increasing the budget from 10,000 to 40,000 testcross plot equivalents in S_1TC-DHTC resulted in more than tripled values of N_1* and N_2* as well as an increase in ΔG^* by about 9%. For DHTC, an increased budget led to a higher increase in N_1* and to smaller increases in N_2* and ΔG* in comparison with S_1TC-DHTC. A fourfold increase in the non-genetic variance from VC1 to VC3 resulted in S_1TC-DHTC in a 50% increase in $N^*_{1_{F/C}}$ and a doubled number of $N^*_{2_{F/C}}$, an increase in L_j* up to the maximum possible number of 15 in both stages, large reductions in $N^*_{1_C}$ and $N^*_{2_{\mathrm{DH}/F}}$, as well as a reduction in ΔG* of approximately 20%. In DHTC, increased non-genetic variance had a larger impact on N_1* but a smaller impact on N_2* and ΔG*. In S_1TC-DHTC, ΔG* was highest for the final selection of 10 DH lines within the best S_1 family within the best cross ($N_f = 1 \times 1 \times 10$). The final selection of the 10 best crosses each with its best S_1 family and its best DH line ($N_f = 10 \times 1 \times 1$) led to a more than fivefold increase in $N^*_{1_C}$ and $N^*_{2_C}$, a decrease in $N^*_{1_{F/C}}$ and $N^*_{2_{\mathrm{DH}/F}}$ by more than 60%, a decrease in L_1* to 8, and a decrease in ΔG* by almost 17% compared to $N_f = 1 \times 1 \times 10$. In DHTC, ΔG* was highest for $N_f = 1 \times 10$, but the composition of N_f had a smaller impact on ΔG* and the optimum allocation of test resources than in S_1TC-DHTC.

Table 4 Optimum allocation of test resources maximizing ΔG (ΔG*) in two-stage selection with evaluation of testcross progenies of (1) DH lines in both stages (breeding scheme DHTC) and (2) S_1 families in the first stage and DH lines within S_1 families in the second stage (breeding scheme S_1TC-DHTC) and its dependence on the budget, VC, and composition of N_f, assuming $\rho_P = 0.71$ and $L_j \leq 15$

Assumptions			Optimum allocation				
Budget	VC	N_f^{a}	$N_1^{*\mathrm{b}}$	$N_2^{*\mathrm{c}}$	L_1*	L_2*	ΔG*
Breeding scheme DHTC							
10,000	VC2	1 × 10	2,038 = 2 × 1,019	168 = 2 × 84	3	15	3.331
20,000	VC2	1 × 10	4,292 = 2 × 2,146	286 = 2 × 143	3	15	3.472
40,000	VC2	1 × 10	8,938 = 2 × 4,469	490 = 2 × 245	3	15	3.601
20,000	VC1	1 × 10	5,998 = 2 × 2,999	290 = 2 × 145	2	15	3.734
20,000	VC3	1 × 10	2,830 = 2 × 1,415	248 = 2 × 124	5	15	3.172
20,000	VC2	2 × 5	4,400 = 4 × 1,116	246 = 3 × 82	3	15	3.377
20,000	VC2	5 × 2	4,455 = 9 × 495	248 = 8 × 31	3	15	3.236
20,000	VC2	10 × 1	4,484 = 19 × 236	240 = 16 × 15	3	15	3.124
Breeding scheme S_1TC-DHTC							
10,000	VC2	1 × 1 × 10	296 = 4 × 74	382 = 2 × 1 × 191	15	13	3.700
20,000	VC2	1 × 1 × 10	576 = 4 × 144	680 = 2 × 1 × 340	15	15	3.874
40,000	VC2	1 × 1 × 10	1,158 = 6 × 193	1,356 = 3 × 1 × 452	15	15	4.020
20,000	VC1	1 × 1 × 10	645 = 5 × 129	884 = 2 × 1 × 442	14	11	4.252
20,000	VC3	1 × 1 × 10	573 = 3 × 191	684 = 2 × 2 × 171	15	15	3.459
20,000	VC2	1 × 2 × 5	633 = 3 × 211	666 = 2 × 3 × 111	15	15	3.752
20,000	VC2	1 × 10 × 1	940 = 2 × 470	630 = 1 × 18 × 35	10	14	3.551
20,000	VC2	2 × 1 × 5	581 = 7 × 83	676 = 4 × 1 × 169	15	15	3.688
20,000	VC2	10 × 1 × 1	1,144 = 22 × 52	644 = 14 × 2 × 23	8	15	3.230

For explanation of abbreviations, see Tables 1 and 2

[a] *DHTC* Number of finally selected crosses × DH lines within selected crosses, *S_1TC-DHTC* Number of finally selected crosses × S_1 families within selected crosses × DH lines within selected S_1 families

[b] *DHTC* Number of crosses × DH lines within crosses at first stage, *S_1TC-DHTC* Number of crosses × S_1 families within crosses at first stage

[c] *DHTC* Number of crosses × DH lines within crosses at second stage, *S_1TC-DHTC* Number of crosses × S_1 families within crosses × DH lines within S_1 families at second stage

In both breeding schemes, for values of $\rho_P < 0.71$, response curves of ΔG increased strongly with increasing N_{1_C} up to a maximum and decreased thereafter slightly (Fig. 1). In the vicinity of the maximum, the response curves of ΔG were flat for varying N_{1_C}. For values of $\rho_P \geq 0.71$, deviations from $N_{1_C}^*$ led to a clear reduction in ΔG. Even for non-optimal N_{1_C}, the use of parental selection ($\rho_P > 0$) was by far superior to random crosses ($\rho_P = 0$).

Discussion

We extended formulas of Hill et al. (1996) for calculation of ΔG to selection among and within crosses (Supplementary Material S2). Thereby, ΔG was calculated assuming an infinite sample size, although populations of medium size are commonly used in plant breeding. This simplifies the calculations considerably and results in similar optimum allocation of test resources and only marginally inflated ΔG compared to finite sample sizes (Longin et al. 2006a).

Impact of the parental selection

The selection among parental lines before starting a new breeding cycle ($\rho_P > 0$) was identified as the most important factor for maximizing ΔG (Table 3, Fig. 1), which is in accordance with previous studies (Utz 1982; Bernardo 2003). The effectiveness of parental selection is influenced by three parameters: (1) the correlation between the mean phenotypic performance of the parental lines and the mean genotypic value of the testcross performance of their progenies (ρ_P), (2) the selected fraction of the parental crosses (α_P), and (3) the reduction in the genetic variance among means of testcross progenies of the crosses (σ_C^2) to $\sigma_C^{2}{}'$ (Eqs. 1–3).

The correlation ρ_P is the product of (1) the genotypic correlation (ρ_G) between the mean GCA of the parental

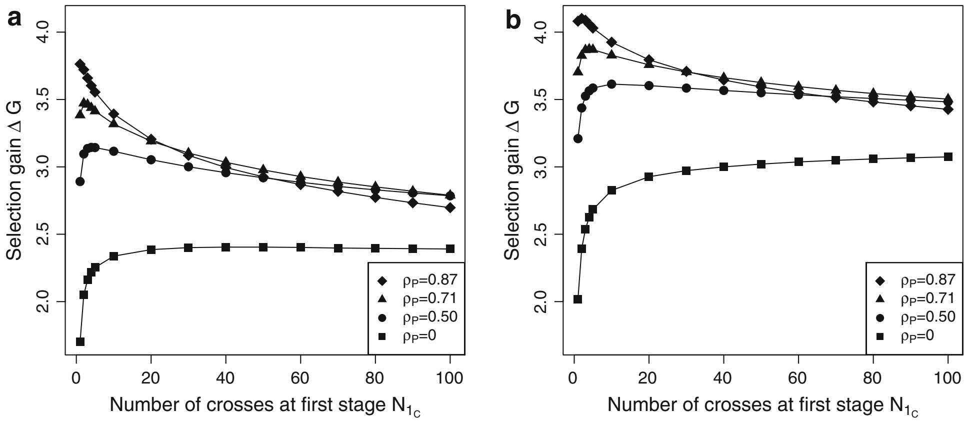

Fig. 1 Selection gain (ΔG) in breeding scheme DHTC (**a**) and $\mathrm{S_1}$TC-DHTC (**b**) as a function of N_{1_C} for varying ρ_P, assuming a budget of 20,000 testcross plot equivalents, VC2, $L_j \leq 15$, and $N_f = 1$ cross × 10 DH lines within that cross for DHTC and 1 cross × 1 $\mathrm{S_1}$ family within that cross × 10 DH lines within that $\mathrm{S_1}$ family for $\mathrm{S_1}$TC-DHTC. For explanation of abbreviations, see Table 1

lines and the mean GCA of their progenies and (2) the square root of the heritability h^2 of the parental GCA. The correlation ρ_G depends solely on the ratio of additive to additive × additive variance (σ_A^2:σ_{AA}^2) of the considered trait in populations in linkage equilibrium, when higher order epistasis is neglected (Supplementary Material S3). For the assumption of negligible σ_A^2 and large σ_{AA}^2, ρ_G reaches its minimum value of 0.71. With increasing importance of σ_A^2 in comparison with σ_{AA}^2, ρ_G rapidly surpasses 0.9 for $\sigma_A^2{:}\sigma_{AA}^2 \geq 1$. The heritability of the parental GCA depends on performance trials in previous breeding cycles and did not depend on the allocation of the actual breeding cycle. Elite inbred lines are usually tested in numerous locations leading to a high h^2. In a study with elite lines tested in 19 environments, h^2 of 0.64 was found (Schön et al. 2004). Based on this study, we assumed a h^2 of 0.64. Even for the pessimistic assumption of $\rho_P = 0.5$, ΔG^* was more than 18% larger than with $\rho_P = 0$, indicating the large importance of parental selection to maximize progress from selection.

For calculating the selected fraction α_P, we assumed 1,000 feasible cross combinations, thus, $\alpha_P = N_{1_C}/1000$. Therefore, at least 46 parental lines are needed for pairwise crosses disregarding reciprocals. This is within the range of 40–120 elite inbreds available in each heterotic group (Bernardo 1996). An increase in the number of feasible cross combinations leads to a smaller α_P and, thus, to a higher ΔG^*. However, the optimum allocation of test resources was not affected (data not shown).

We assumed equal genetic variances within all crosses. This corresponds to the assumption that the individuals are a random sample of a population in Hardy–Weinberg equilibrium. For this situation, the assumption of equal genetic variances was confirmed with simulation studies, where differences in the variance within crosses were much smaller than differences in their mean performance (Zhong and Jannink 2007). However, in applied breeding programs, the pedigree structure leads to complex relationships among elite inbreds within the heterotic groups. This may lead to different genetic variances within different crosses and might reduce ΔG and the optimum number of crosses (N_{jC}^*) in comparison with our results. However, the complex pedigree structure in maize breeding programs with varying coefficients of coancestry among elite inbreds (Bernardo 1996) makes exact investigations cumbersome and requires simplifying assumptions. In addition, the implementation of pedigree information to parental selection requires further research on index selection and optimum weights for combining the performance of the parental lines and their coefficient of coancestry.

The flip side of a large ρ_P and a small α_P is a reduction in σ_C^2. However, the advantage of an increased selection gain in parental selection seems to outweigh the effect of the reduction in σ_C^2 by far. The GCA of inbreds is available from BLUP analysis of single crosses (Bernardo 1996) so this GCA information should be readily available in breeding programs. Consequently, the prediction of the mean performance of a cross by the mean performance of its parental lines allows an early rejection of inferior crosses without exhausting the budget. With unknown or imprecisely evaluated parental lines, an additional pretest of these parents might be an approach to reduce the number of crosses (Choo et al. 1979; Baker 1984; Wricke and Weber 1986).

Optimum allocation of test resources

With increasing ρ_P, the optimum number of crosses decreased strongly in favor of increased numbers of test candidates within crosses in both breeding schemes (Table 3, Fig. 1). This result is in accordance with that of Utz (1982) and can be explained by different amounts of genetic variance available for selection among crosses, $\mathrm{S_1}$ families and DH lines. While σ_C^2 is reduced by parental selection, the variances within crosses ($\sigma_{\mathrm{DH}/C}^2$, $\sigma_{F/C}^2$, and $\sigma_{\mathrm{DH}/F}^2$) are not affected by parental selection and, consequently favors selection within crosses. In contrast to these findings, a large optimum number of crosses (among selected parents) at the expense of a small optimum number of test candidates within crosses was reported in the literature (Bernardo 2003). The difference is due to different crossing procedures for the parental lines. We assumed that a breeder uses the best cross combinations based on the parental mean, whereas Bernardo (2003) selected the best 25 or 10% of the parental lines and crossed them randomly.

In maize breeding programs, generally 10–50 crosses per cycle are produced (Bernardo 2003) and the progenies of several crosses are finally selected. The very low $N_{1_C}^*$ in our study is attributable to the focus on line development within one heterotic group. Taking into account that several heterotic groups within one breeding program are commonly used, our results confirm the number of crosses generally used in applied maize breeding. The consideration of multiple traits, effective population size for long-term selection (Gordillo and Geiger 2007), and the risk of focusing on an inferior cross might increase $N_{j_C}^*$. However, this needs further research on index and long-term selection.

Response curves of ΔG as a function of N_{1_C} depended strongly on ρ_P (Fig. 1). With $\rho_P < 0.50$, response curves were flat in the vicinity of the maximum. This is in accordance with findings on ΔG as a function of the number of DH lines within one cross (Longin et al. 2006b). In contrast, for $\rho_P \geq 0.50$, deviations from the optimum $N_{1_C}^*$ led to strong reductions in ΔG. For instance, in S_1TC-DHTC assuming $\rho_P = 0.87$, an increase in N_{1_C} from two to five led to a decrease in ΔG of almost 2%. This difference may be due to the increasing importance of a larger number of test candidates within crosses with increasing ρ_P, as described in the previous section.

For $\rho_P > 0$, the number of DH lines within S_1 families varied between 20 and 470 in all scenarios of S_1TC-DHTC. In DHTC, the number of DH lines within crosses was up to ten times larger (Tables 3, 4). The high number of DH lines per cross in DHTC is cumbersome with the actual DH technique. With the current rates of haploid induction (10–15%) and chromosome doubling (20–30%, Röber et al. 2005), the production of more than 1,000 DH lines per cross in DHTC is difficult, requiring multiple production and induction of the same cross. In contrast, the production of more than 300 DH lines per S_1 family is possible by using a bulk of S_2 plants within each S_1 family for haploid induction.

In all scenarios, ΔG^* was highest for the final selection of 10 DH lines within the best cross (Table 4). With higher numbers of finally selected crosses, $N_{j_C}^*$ increased at the expense of a reduced number of test candidates within crosses. However, ΔG^* was reduced by more than 4%, even though the total number of finally selected DH lines N_f has not been changed (Table 4). Modifying our selection approach to (1) an evaluation of varying numbers of test candidates within crosses according to the performance level of the cross in the first stage and (2) selecting the best test candidates across all crosses and S_1 families tested in the second stage might increase $N_{1_C}^*$ and ΔG^*. However, to our knowledge, no quantitative genetic formulas are available for such breeding strategies, requiring further research. Additionally, the effective population size decreases more strongly by final selection of only one cross with 10 DH lines, than by the final selection of several crosses with a smaller number of DH lines per cross. Thus, for long-term selection, a final selection of more than one cross might be preferable.

Relative efficiency of breeding schemes

The considerable superiority of S_1TC-DHTC over DHTC reported in a previous study (Longin et al. 2007b) was confirmed in this study also for the extension of those schemes to selection among and within crosses (Tables 3, 4). The differences in ΔG^* among the breeding schemes are mainly due to the possibility to concentrate a larger part of the budget on DH lines of the best S_1 families and the new arising genetic variance in the second selection stage of S_1TC-DHTC (Longin et al. 2007b). The relative superiority of S_1TC-DHTC over DHTC decreased with increasing ρ_P (Table 3). The stronger impact of an increasing ρ_P on DHTC than on S_1TC-DHTC may be due to a larger increase in selection intensity in the second stage of the selection within crosses in DHTC. In DHTC, the released capacity due to parental selection can mostly be used for increasing the selection intensity within crosses. In contrast, in S_1TC-DHTC, the large number of test locations in both stages limits the increase in the number of test candidates within crosses. Nevertheless, even for $\rho_P = 0.87$, ΔG^* was more than 9% higher in S_1TC-DHTC than in DHTC.

Breeding scheme S_1TC-DHTC has a longer cycle length than DHTC (Supplementary Figure S1). The length of S_1TC-DHTC could be shortened by using individual S_1 plants as males for production of testcross seed and in parallel as females in crosses with the inducer (Longin et al. 2007b). Chromosome doubling must then be performed simultaneously with early testing. However, even with maximum rates of haploid induction, the high optimum number of DH lines per S_1 plant determined in this study cannot be realized in practice. S_1TC-DHTC could also be shortened, if per se trials were feasible already 1 year after haploid induction. This might be realized with further improvement of haploid induction, identification of haploid kernels, and chromosome doubling, as well as the use of winter nurseries, where more than two generations per year can be realized. In contrast, a shortening of DHTC is hardly feasible, because performance trials must be made in the summer season. Consequently, the larger ΔG^* in breeding schemes with early testing prior to DH production could be combined with a short cycle length, representing a very promising breeding scheme for the development of maize hybrids.

In conclusion, the possibility to reduce the number of crosses by selection among parental lines is of utmost importance for the optimum allocation of test resources and maximizing ΔG. For advanced cycle breeding, where ρ_P is expected to be high, the concentration on few crosses among the best parental lines might be a promising approach for short-term success in hybrid maize breeding. However, for long-term success, the effective population size should also be considered. With the current limitations in the DH technique, S_1TC-DHTC has a higher ΔG at the expense of a longer cycle length than DHTC. With improvements in the DH technique and the realization of more than two generations per year, early testing prior to production of DH lines would become very attractive in hybrid maize breeding. For S_1TC-DHTC with parental selection, assuming a budget of 20,000 fields for testing one heterotic group, the allocation of test resources is close to its optimum, if (1) the number of crosses does not exceed five in the first stage ($N_{1_C} \leq 5$), (2) the number of test locations is similar in both selection stages ($L_1 \approx L_2$), and (3) >50% of the budget are spent in the second stage.

Acknowledgments This research was supported by funds from DFG, Grant No 1070/1, International Research Training Group "Sustainable Resource Use in North China" to T. Wegenast. The authors thank Dr. W. Schipprack, Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, Stuttgart, Germany for his valuable suggestions. In addition, 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 reviewer.

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

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

Choo TM, Christie BR, Reinbergs E (1979) Doubled haploids for estimating genetic variances and a scheme for population improvement in self-pollinating crops. Theor Appl Genet 54:267–271

Cochran WG (1951) Improvement by means of selection. In: Proc second Berkeley symp math stat prob, pp 449–470

Gallais A (1979) The concept of varietal ability in plant breeding. Euphytica 28:811–823

Gordillo GA, Geiger HH (2007) Optimizing of DH-line based recurrent selection procedures in maize under a restricted annual loss of genetic variance. Euphytica. doi:10.1007/s10681-007-9616-y

Hill WG, Caballero A, Dempfle L (1996) Prediction of response to selection within families. Genet Sel Evol 28:379–383

Hühn M (1996) Optimum number of crosses and progeny per cross in breeding self-fertilizing crops. I. General approach and first numerical results. Euphytica 91:365–374

Hühn M (2005a) Optimum number of crosses and progeny per cross in breeding self-fertilizing crops. II. Numerical results based on expected selection responses (special case). Cereal Res Commun 33:493–500

Hühn M (2005b) Optimum number of crosses and progeny per cross in breeding self-fertilizing crops. II. Numerical results based on expected selection responses (general case). Cereal Res Commun 33:501–508

Hühn M (2006) Optimum number of crosses and progeny per cross in breeding self-fertilizing crops. III. Economic constraints. Cereal Res Commun 34:903–910

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, Melchinger AE, Reif JC (2007a) Hybrid maize breeding with doubled haploids: II. Optimum type and number of testers in two-stage selection for general combining ability. Theor Appl Genet 114:393–402

Longin CFH, Utz HF, Reif JC, Wegenast T, Schiprack W, Melchinger AE (2007b) 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

Melchinger AE (1988) Means, variances, and covariances between relatives in hybrid populations with disequilibrium in the parent population. In: Weir BS, Eisen EJ, Goodman MM, Namkoong G (eds) Proc second int conf quantit genet. Sinnauer, Sunderland, pp 400–415

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, USA, pp 8–21

Röber F, Gordillo GA, Geiger HH (2005) In vivo haploid induction in maize—performance of new inducers and significance of doubled haploid lines in hybrid breeding. Maydica 50:275–283

Schmidt W (2004) Hybridmaiszüchtung bei der KWS SAAT AG (in German). In: Bericht über die 54. Tagung der Vereinigung der Pflanzenzüchter und Saatgutkaufleute Österreichs 2003, Gumpenstein, Austria, pp 1–6

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

Schön CC, Utz HF, Groh S, Truberg B, Openshaw S, Melchinger AE (2004) Quantitative trait locus mapping based on a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. Genetics 167:485–498

Seitz G (2005) The use of doubled haploids in corn breeding. In: Proceedings of the forty first annual Illinois corn breeders' School 2005, Urbana-Champaign, USA, pp 1–7

Sprague GF, Federer WT (1951) A comparison of variance components in corn yield trials: II. Error, year × variety, location × variety and variety components. Agron J 42:535–541

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

Utz HF (1982) Der Selektionserfolg in spaltenden Generationen mit experimentellen Untersuchungen und Modellrechnungen (in German). Professorial dissertation, Universität Hohenheim, Stuttgart

Wricke G, Weber WE (1986) Quantitative genetics and selection in plant breeding. Walter de Gruyter, Berlin, pp 41–65, 172–179, 214–218

Wright AJ (1974) A genetic theory of general varietal ability for diploid crops. Theor Appl Genet 45:163–169

Zhong S, Jannink J-L (2007) Using QTL results to discriminate among crosses based on their progeny mean and variance. Genetics 177:567–576