Aus dem Institut für

Pflanzenzüchtung, Saatgutforschung und Populationsgenetik

der Universität Hohenheim

Fachgebiet Angewandte Genetik und Pflanzenzüchtung

Prof. Dr. A.E. Melchinger

Optimum schemes for hybrid maize breeding with doubled haploids

Dissertation

zur Erlangung des Grades eines Doktors

der Agrarwissenschaften

vorgelegt der

Fakultät Agrarwissenschaften

von

M.Sc. agr.

Thilo Wegenast

aus Sulz am Neckar

Münster 2011

Die vorliegende Arbeit wurde am 18.04.2011 von der Fakultät Agrarwissenschaften der Universität Hohenheim als "Dissertation zur Erlangung des Grades des Doktors der Agrarwissenschaften" angenommen.

Tag der mündlichen Prüfung: 09.06. 2011

1. Prodekan: Prof. Dr. A. Fangmeier

Berichterstatter 1. Prüfer: Prof. Dr. A.E. Melchinger

Mitberichterstatter 2. Prüfer: Prof. Dr. H.-P. Piepho

3. Prüfer: Prof. Dr. J. Bennewitz

CONTENTS

---

[1] 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

[2] Wegenast T, Dhillon BS, Utz HF, Longin CFH, Maurer HP, Melchinger AE (2009) Comparison of selection strategies and optimization criteria in hybrid maize breeding with doubled haploids. Maydica 54:343-352

[3] Wegenast T, Utz HF, Longin CFH, Maurer HP, Dhillon BS, Melchinger AE (2010) Hybrid maize breeding with doubled haploids: V. Selection strategies for testcross performance with variable sizes of crosses and S_1 families. Theor Appl Genet 120:699-708
Erratum to: Wegenast T, Utz HF, Longin CFH, Maurer HP, Dhillon BS, Melchinger AE (2010) Hybrid maize breeding with doubled haploids: V. Selection strategies for testcross performance with variable sizes of crosses and S_1 families. Theor Appl Genet 121:1391-1393

[4] Mi X, Wegenast T, Utz HF, Dhillon BS, Melchinger AE (2011) Best linear unbiased prediction (BLUP) and optimum allocation of test resources in maize breeding with doubled haploids. Theor Appl Genet 123:1-10

ABBREVIATIONS

BLUE	Best linear unbiased estimator
BLUP	Best linear unbiased prediction
DH	Doubled Haploid
GCA	General combining ability
MME	Mixed Model Equation
$P(q)$	Probability of selecting superior genotypes
SD	Standard deviation
ΔG	Selection gain
Θ	Coancestry coefficient
ρ_G'	Genotypic correlation between the mean GCA of the parental lines and the GCA of a single random progeny line
ρ_G	Genotypic correlation between the mean GCA of the parental lines and the mean GCA of their progeny lines
ρ_P	Correlation between the mean phenotypic performance of the parental lines and the mean genotypic performance of the testcross progeny
σ^2_A	Additive variance
σ^2_{AA}	Additive x additive variance
σ^2_G	Genetic variance

1 GENERAL INTRODUCTION

In hybrid maize breeding, inbred lines were traditionally produced by recurrent selfing of heterozygous plants for five to six generations (Hallauer 1990; Eder and Chalyk 2002). With the use of the doubled haploid (DH) technique, collections of homozygous plants can be obtained directly from breeding material, which is a segregating population of gametes, without the time consuming process of inbreeding (Lee and Tracy 2009). The application of the DH technique has several advantages: (i) DH development is very rapid and considerably shortens the length of a breeding cycle. (ii) The total genetic variances among DH for *per se* and testcross performances are expressed in the first DH generation and, thus, the selection process is accelerated and its efficiency enhanced. (iii) DH lines are completely homozygous and, therefore, meet better the criteria for variety protection and also require fewer resources for selfing and maintenance breeding. (iv) Logistics of seed transfer between main season and off-season sites is simplified during DH development. (vii) Inbred line development and evaluation can be separated into two distinct activities: the DH line development, which can be globally centralized, and their evaluation, which has to continue as a decentralized activity to meet adaptation needs of agro-ecologically diverse regions (Schmidt 2004; Röber et al. 2005; Geiger and Gordillo 2009; Lee and Tracy 2009).

The successful use of DH lines is mainly attributable to the substantial progress made in the *in vivo* haploid development technique during the past 10-15 years (Geiger 2009). Coe (1959) laid the foundation of the *in vivo* production of haploid plants by the identification of genotypes (known as inducers), which induce maternal haploids when used as a pollen parent. The availability of efficient inducer lines with haploid induction rates above 8% (Röber et al. 2005) combined with a reliable system for distinguishing kernels with a haploid embryo from those with diploid embryo greatly facilitated the adoption and routine use of DH lines in hybrid maize breeding. As the integration of the DH technique has an impact on many steps in a breeding program, new breeding schemes and selection strategies, which are more cost-effective than those prevalent, need to be worked out and the optimum allocation of the available resources have to be investigated. Thus, breeding and selection schemes need to be studied for the optimum allocation of resources to maximize progress from selection. The focus of the present study was to conduct investigations on the optimum allocation of resources to maximize progress from two recently suggested breeding schemes. A breeding program having more than one breeding population, as is practically the case, was considered and new statistical tools were employed to explore these breeding schemes for the maximization of progress from selection.

Target criteria to measure the progress from selection

Two target criteria are defined in the literature to quantify the progress from selection: (1) the selection gain ΔG (Falconer and Mackay 1996, pp 184-194) and (2) the probability $P(q)$ of identifying superior genotypes (Keuls and Sieben 1955). The selection gain reflects the superiority of the population generated by intermating the selected genotypes in comparison with the genotypic mean of the base population, whereas $P(q)$ quantifies the probability to develop superior varieties without reference to the mean of the selected group (Wricke and Weber 1986, pp 172-194). With ΔG, the selection progress of breeding programs employing recurrent selection is quantified and optimized (Gordillo and Geiger 2008), whereas the use of $P(q)$, focuses on quantifying the success in the development of competitive varieties (Johnson 1989).

For a given breeding population and a fixed number of selected genotypes, both ΔG and $P(q)$, are functions of the heritability and the selection intensity and are positively correlated with each other (Bernardo 2002, pp 204-206; Johnson 1989). For a given breeding population, heritability increases with larger numbers of test locations, years and replications used to conduct performance trials. Selection intensity increases with larger numbers of initial test candidates. With limited financial resources, a plant breeder has to find a compromise between (1) the number of breeding populations, *i.e.*, parental crosses, to be used to develop new lines, (2) the number of candidates within each parental cross, and (3) the intensity of their testing as determined by the number of locations, years, and replications. Thus, the optimum allocation of test resources is of crucial importance for having a high efficiency of breeding programs. The use of $P(q)$, focuses on quantifying the success of the development of competitive varieties (Johnson 1989). Since ΔG and $P(q)$ quantify the progress in selection with different focuses, their deployment may lead to diverse optimum allocations of test resources or even choice of different breeding schemes or selection strategies for these two situations.

Selection among parental lines of crosses

In hybrid maize breeding, new lines are usually derived from crosses of the elite inbreds within heterotic pools, which is known as advanced cycle breeding (Bernardo 1996, 2002). Since a large number of elite inbreds is available within each heterotic pool, the number of possible parental crosses that can be obtained is extremely large. Breeders have two options: develop a large number of parental crosses and a small number of DH lines within each cross or develop a small number of parental crosses and a large number of DH lines. In applied breeding as well as in theoretical studies, both approaches are used, but there is no consensus on the relative superiority of these (*cf.*, Wricke and Weber 1986, pp 172-179; Bernardo 2003). A large number of parental crosses were considered superior by Hühn (1996) and Bernardo (2003), whereas others favored a medium (Baker 1984) or a small number (Utz 1982) of crosses with larger numbers of lines derived from them.

The mean phenotypic performance of the DH lines derived from a cross between two inbred parents can be predicted by the average performance of these parental inbreds (*cf.*, Choo *et al.* 1979; Wricke and Weber 1986, pp 41-65). The efficiency of this prediction depends on the genetic correlation between mean performance of the parental inbreds and the mean performance of their progeny lines, as well as the square root of the heritability of the parental performance. The efficiency of this prediction is expected to have strong influence on the optimum allocation of test resources as well as the progress from selection (Baker 1984). An efficient prediction of the performance of the breeding population from its parental mean, would allow allocation of resources to the most promising parental crosses. Despite its crucial importance for success of a breeding program, no study is available in the literature on maize breeding with DH considering selection among parental inbreds before crossing them for initiating a new breeding cycle.

Alternative breeding schemes

A key question in hybrid maize breeding with DH is the generation of the crosses that should be used for haploid induction. Actually, the general practice in applied breeding is to induce haploids in S_0 plants (Seitz 2005; Longin et al. 2007; Bernardo 2009). With S_0 we refer to the F_1 of a biparental cross. With early haploid induction and the use of off-season nurseries, a breeding cycle can be shortened down to four years (Gordillo and Geiger 2008) and the evaluation of potential hybrid cultivars is possible very early after beginning a new breeding cycle. However, in this approach a large amount of the budget is spent to produce DH lines from untested parental crosses and parental plants. Further, with this approach, DH lines are produced after only one meiosis and, thus, the chances of recombination are low. The limited recombination may decrease the genetic variance among the DH lines, leading to a reduced selection gain. On the other hand, positive epistatic combinations of alleles may be conserved with this approach (Bernardo 2009).

Alternatively to the production of DH lines directly from the S_0, an early test on testcross performance can be made in the S_1 or S_2 generation before DH production. This leads to a longer breeding cycle but allows focusing the production of DH lines from the most promising parental crosses and plants/families within crosses. Early testing has been proposed by Jenkins (1935, cited by Bernardo 2002, p 183). The advantage of early testing is that test resources can be concentrated on the most promising material. A prerequisite for the success of early testing is an adequately high genetic correlation of testcross of an S_1 plant with the inbred(s) derived from that S_1 plant. A theoretical value of this correlation, in absence of epistasis, larger than 0.7 (Bernardo 1991) supports early testing. In conventional inbred line development using recurrent selfing both, early and late testing, are possible and widely used (Bauman 1981). However, the use of DH technique needs a clear decision whether the first testcross is made before or after production of DH lines. Longin *et al.* (2007) compared breeding schemes with production of DH lines in the S_0 and S_1 generation with a focus on selection within one given cross population. However, the potential of early testing applied to a hybrid maize breeding programs working on many breeding populations, as is the general case, and its influence on the optimum allocation of resources for selection among and within parental crosses is lacking in the literature.

Alternative selection strategies

Under both breeding schemes, using early testing or producing DH directly in the S_0 generation, selection of superior genotypes might be practiced according to the following two strategies: (1) Selection can be performed first among parental crosses for their mean performance and then, in a second step, for the best candidates within the selected parental crosses. (2) Selection can be performed immediately among all DH lines without considering the mean performance of a parental cross. The selection among crosses enables breeders to discard inferior ones and to concentrate their resources on the most promising crosses (*cf.*, Schnell 1982). However, discarding entire crosses implies the risk of rejecting single superior candidates within crosses of lower mean performance. Therefore, selection among all DH lines disregarding their pedigree relationships (*cf.*, Lush 1947; Falconer and Mackay 1996, pp 229-239) might be a promising approach. A comparison of these selection strategies considering the optimum allocation of test resources is not available in the literature.

In most studies on the optimum allocation of resources, a constant size of crosses or families within crosses at one selection stage is assumed (*cf.*, Baker 1984; Bernardo 2003). However, it is expected to be a superior strategy to vary the size of crosses depending on their performance in earlier breeding cycles or stages, *i.e.*, produce a larger number of DH lines from superior crosses at the expense of a reduced number of DH lines from inferior crosses. In consequence, a higher proportion of the budget could be allocated to the most promising crosses without totally neglecting the chance of testing less promising ones. In few studies on animal breeding, the influence of unequal family sizes on ΔG was investigated; however, only specific situations with very small family sizes were considered (Burrows 1984; Toro and Nieto 1984; Phocas and Colleau 1995). In addition, there is no study calculating the optimum allocation of test resources to parental crosses in the literature on plant breeding.

Best linear unbiased prediction (BLUP) of genotypic values

Genotypic values are typically estimated using phenotypic values of the candidate without considering information on related genotypes. With the best linear unbiased prediction (BLUP), unbalanced data and pedigree information using the genetic covariance among relatives can be used to predict the genotypic value of a candidate. The BLUP procedure was originally developed in animal breeding to estimate breeding values (Robinson 1991). Predictions are based on (1) known genetic relationships among genotypes and (2) available phenotypic data of related genotypes (Bernardo 1996). Thus, genotypes developed from different but related breeding populations and tested in different environments can be compared. The BLUP method has high prediction accuracy and is expected to enhance the efficiency of selection. Thus, its use in plant breeding is being advocated (Bernardo 1996; Piepho 2009) but so far is seldom used. Thus, studies on the effect of BLUP on the progress from selection and the optimum allocation of test resources are of interest.

The basic statistical theory of BLUP is expressed in the mixed model equation (MME), integrating fixed and random effects (Bernardo 2002, Chapter 10). Random effects have a covariance structure, whereas fixed effects have not. The covariance structure of the pedigree information from related candidates is reflected in the genetic relationship matrix **A**, which is based on the coefficients of coancestry among the test candidates. The matrix inversions to solve the MME are generally very computation intensive. For assessing the optimum allocation of test resources, an immense number of allocation combinations need to be calculated, which is extremely time consuming because large matrices must be handled in applied breeding programs. With the availability of general formulas to solve the MME, rapid simulation studies could be performed avoiding matrix inversions. However, no such formula or optimization study is published in the literature to compute the progress from selection or the optimum allocation of test resources in hybrid maize breeding.

Objectives

The aim of the research for this thesis was to investigate the optimum use of the DH technique in hybrid maize breeding. We compared breeding schemes and selection strategies for the optimum allocation of test resources to maximize progress from selection. The specific objectives were to:

(1) compare different target criteria, ΔG and $P(q)$ as well as their respective standard deviations, for the optimum allocation of test resources by using numerical integration and Monte Carlo simulations;

(2) investigate the effect of parental selection, varying variance components and budgets on the optimum allocation of test resources for different target criteria;

(3) assess the optimum filial generation for DH production by comparing two breeding schemes based on DH production in generation S_0 or S_1;

(4) compare selection strategies - sequential selection among and within parental crosses vs. selection among all test candidates ignoring the cross and family structure – for optimum allocation of test resources and progress from selection;

(5) examine the effect of producing a larger number of candidates within promising crosses and S_1 families on the optimum allocation of test resources and progress from selection; and

(6) determine the effect of selection based on criteria developed by integrating information from genetically related candidates by employing BLUP, on the optimum allocation of test resources and progress from selection.

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

Bauman LF (1981) Review of methods used by breeders to develop superior inbreds. Proc Corn Sorghum Industr Res Conf 36:199-208

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, USA

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 255-262

Burrows PM (1984) Inbreeding under selection from related families. Biometrics 40:895-906

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

Coe EH (1949) A line of maize with high haploid frequency. Am Nat 93:3821-382

Eder J, Chalyk S (2002) In vivo haploid induction in maize. Theor Appl Genet 104:703-708

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman

Scientific and Technical Ltd, Essex, England

Geiger HH (2009) Doubled haploids. In: Bennetzen JL, Hake S (eds) Maize handbook – Volume II: Genetics and genomics. Springer Science and Business, New York, pp 641-657

Gordillo GA and Geiger HH (2008) Optimization of DH-line based recurrent selection procedures in maize under a restricted annual loss of genetic variance. Euphytica 161:141-154

Hallauer AR (1990) Methods used in developing maize inbreds. Maydica 35:1-16

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

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

Lee EA, Tracy WF (2009) Modern maize breeding. In: Bennetzen JL, Hake S (eds) Maize handbook – Volume II: Genetics and genomics, Springer Science and Business, New York, pp 142-160

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

Phocas F, Colleau JJ (1995) Approximating selection differentials and variances for correlated selection indices. Genet Sel Evol 27:551-565

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

Röber FK, 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

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

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

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

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

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

Wricke G, Weber WE (1986) Quantitative genetics and selection in plant breeding. Walter de Gruyter, Berlin

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, and Jochen C. Reif

T. Wegenast, C. F. H. Longin, H. F. Utz, and A. E. Melchinger, Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany

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

Theor Appl Genet (2008) 117:251-260

The original publication is available at www.springerlink.com

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

Abstract Parental selection influences the optimum allocation of test resources and the gain from selection in breeding programs. In our study, 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. The 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 before the production of DH lines might become very attractive in hybrid maize breeding.

Comparison of selection strategies and optimization criteria in hybrid maize breeding with doubled haploids

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

T. Wegenast, B.S. Dhillon, H.F. Utz, and A.E. Melchinger, Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany

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

Maydica 54 (2009): 343-352

The original publication is available at www.maydica.org

Abstract Doubled haploids (DH) are increasingly replacing conventionally developed inbred lines in hybrid maize (*Zea mays* L.) breeding. We considered two breeding schemes with selection based on the evaluation of testcross progenies of DH lines in two stages (DHTC) and of S_1 families in the first stage and of DH lines within S_1 families in the second stage (S_1TC-DHTC). For both breeding schemes, we considered different selection strategies with or without optimum index at second stage and uniform or variable sizes of crosses and S_1 families. Our objectives were to (1) determine the optimum number of test environments as well as the optimum number and sizes of parental crosses and S_1 families for various variance component ratios and (2) identify the best selection strategy for each breeding scheme with respect to the selection gain (ΔG) and the probability to select superior genotypes ($P(q)$) as well as to minimize their standard deviation. Breeding scheme S_1TC-DHTC had larger ΔG and $P(q)$ but a higher standard deviation of ΔG than DHTC. For breeding scheme S_1TC-DHTC, strategies with selection among all DH lines led to larger progress than the strategy with sequential selection among and within crosses and S_1 families. The results for the optimization criteria in strategies with variable numbers of S_1 families and DH lines within crosses and S_1 families were larger than with uniform numbers. With decreasing contributions of the genetic variance, the number of test locations should be increased at the expense of the test candidates in both breeding schemes. Our study showed that a lower standard deviation of the expected values for both optimization criteria was feasible without a significant loss in ΔG or $P(q)$ with an increased number of test locations.

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, and Albrecht E. Melchinger

T. Wegenast, H.F. Utz, B.S. Dhillon, and A.E. Melchinger, Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany

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

Theor Appl Genet (2010) 120:699-708; 121:1391-1393

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 from S_1 families. In both 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 in order 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. The 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 selection strategies and breeding schemes showed similar relative efficiency for both optimization criteria ΔG and $P(0.1\%)$. In comparison to 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.

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

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

X. Mi, T. Wegenast, H.F. Utz, B.S. Dhillon, and A.E. Melchinger, Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany

Theor Appl Genet (2011) 123:1-10

The original publication is available at www.springerlink.com

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

Abstract With best linear unbiased prediction (BLUP), information from genetically related candidates is combined to obtain more precise estimates of genotypic values of test candidates. Thereby, progress from selection is increased. We developed and applied theory and Monte Carlo simulations implementing BLUP in two two-stage maize breeding schemes and various selection strategies. Our objectives were to (1) derive analytical solutions of the mixed model equations in two breeding schemes, (2) determine the optimum allocation of test resources with BLUP under different assumptions regarding the variance component ratios for grain yield in maize, (3) compare the progress from selection using BLUP and conventional phenotypic selection based on mean performance solely of the candidates, and (4) analyze the potential of BLUP for further improving the progress from selection. The breeding schemes involved selection for testcross performance either of DH lines at both stages (DHTC) or of S_1 families at the first stage and DH lines at the second stage (S_1TC-DHTC). Our analytical solutions

allowed much faster calculations of the optimum allocations and superseded matrix inversions to solve the mixed model equations. Compared to conventional phenotypic selection, the progress from selection was slightly higher with BLUP for both optimization criteria, namely the selection gain and the probability to select the best genotypes. The optimum allocation of test resources in S_1TC-DHTC involved 10 test locations at both stages, a low number of crosses (6) each with 100-300 S_1 families at the first stage, and 500-1000 DH lines at the second stage. In breeding scheme DHTC, the optimum number of test candidates at the first stage was 5-10 times larger, whereas the numbers of test locations at the first stage and the number of test candidates at the second stage were strongly reduced compared to breeding scheme S_1TC-DHTC.

6 GENERAL DISCUSSION

The use of doubled haploids (DH) is being implemented as a standard procedure in many commercial maize breeding programs. It has opened many avenues for developing and deploying new breeding schemes with higher selection efficiency. Additionally, rapid progress in computation power has enabled the evaluation of progress through new breeding schemes and selection strategies by the application of new statistical methods. The synergistic combination of new statistical methods as well as the DH technology can be effectively employed for an optimization of breeding schemes. We focused on the optimum allocation of test resources for DH development in maize breeding. Thereby, we considered two breeding schemes involving two-stage selection for testcross performance: A breeding scheme with evaluation of DH lines from the very beginning of the selection process (DHTC) and another scheme that employed early testing based on testcross performance of S_1 families and then producing and testing DH lines derived from selected S_1 families and testing the testcrosses of DH lines (S_1TC-DHTC, Wegenast *et al.* 2008, Suppl. Mat.). For both breeding schemes, we evaluated different selection strategies and optimization criteria. We considered grain yield as the target criterion, controlled by many loci with small effects resulting in a Gaussian normal distribution of genotypic and phenotypic values (*cf.*, Dekkers and Hospital 2002)

Comparison of numerical integration and Monte Carlo simulations for modeling breeding schemes

Optimization and comparisons of selection strategies in plant breeding are usually made by using model calculations based on quantitative genetic theory (Bouchez and Gallais

2000; Bordes *et al.* 2006; Longin *et al.* 2007a, b; Gordillo and Geiger 2008). These calculations represent non-linear optimizations and require the numerical computation of multivariate integrals for specific probability distributions. Alternatively to this numerical computation, multivariate integrals can be determined by Monte Carlo simulations (Jannink and Abadie 1999; Wang *et al.* 2003; Bauer *et al.* 2006; Longin *et al.* 2006a; Wang and Pfeiffer 2007).

For numerical integration, formulas for finite population sizes are only available for one-stage selection (Robson *et al.* 1967; Hill 1976, 1977). With the simplifying assumption of an infinite population size, exact formulas for selection gain (ΔG) in multi-stage selection have been developed for self pollinating crops (Utz 1967). As breeding populations are usually small, the expected gain in numerical integration results in overestimating the realized gains (Hill and Caballero 1992). The optimum allocation of test resources for ΔG was found to be similar assuming finite or infinite population sizes (Utz 1969; Longin 2007), and our results are in accordance to that (Wegenast *et al.* 2008, 2010). However, no numerical formulas are available to calculate the probability of identifying superior genotypes ($P(q)$). In addition, the standard deviation (SD) of both ΔG and ($P(q)$) cannot be calculated and no information is available on the relationship among the selected candidates, when numerical integration is used.

Monte Carlo simulations are a flexible tool for modeling selection processes (Jannink and Abadie 1999), and multiple optimization criteria such as ΔG or ($P(q)$; moreover, the *SD* of these criteria can be calculated at the same time. In addition, extra information can be integrated in the simulation, *e.g.*, the genetic relationship of the selected candidates, which is necessary to calculate the coancestry coefficient among the selected candidates. The use of Monte Carlo simulations is limited by the computing time, because the simulation has to be run at least 10,000 times for each allocation of resources to achieve an adequate accuracy. However, no quantitative genetic formulas were available to model variable family sizes or selection based on BLUPs. Thus, to model scenarios which are close to applied breeding programs or which explore new methods of selection, Monte Carlo simulations are necessary (Wegenast *et al.* 2010; Mi *et al.* 2011).

Comparison of optimization criteria

The selection gain ΔG is a widely used criterion to measure the progress from selection. Another criterion, the probability to select superior genotypes, has been employed in few studies only. By using both these criteria and their SD, we focused on progress from selection on a per cycle basis.

The estimates of ΔG reflect the superiority of the population generated by intermating the selected genotypes over the genotypic mean of the base population, whereas $P(q)$ quantifies the probability to develop superior varieties without reference to the mean of the genotypes in the base population and the selected group (Wricke and Weber 1986, pp 172-194). ΔG reflects gain through recurrent selection and it is an indicator of long-term breeding success, whereas $P(q)$ focuses the identification of superior candidates on a short-term basis. Besides the estimation of genetic progress expected by using ΔG or $P(q)$), which are here considered as main optimization criteria, it is important to know the accuracy of the estimates, which is reflected by the SD of ΔG or $P(q)$.

Exact formulas are available for ΔG assuming infinite sample sizes but not for $P(q)$. Thus, in one of the studies, where we used quantitative genetic formulas, we determined ΔG (Wegenast *et al.* 2008). Further, Monte Carlo simulations were employed to estimate at the same time both optimization criteria and their SD. To examine long-term success of a breeding program, the genetic relationship among the selected DH lines was considered by calculating their coancestry coefficient Θ. The influence of a breeding scheme or selection strategy on long-term breeding success was estimated by assessing the genetic relationship of the selected candidates and thereof, the coancestry coefficient among these candidates.

The optimum allocation of test resources generally differed for *ΔG* with *$P(q)$*, but there was no consistent trend (Wegenast *et al.* 2010; Mi *et al.* 2011). The results are in contrast to those reported by Longin *et al.* (2007), who investigated selection among DH lines within one cross, and observed that optimum allocation was similar for both optimization criteria. Further, Yonezawa and Yamagata (1978), who considered one-stage selection in segregating populations, had reported that the probability to identify superior genotypes was maximized rather by increasing the number of crosses than the number of progenies within crosses. The different results in the present and two earlier studies are most probably mainly due to the more complex situation considered by us. The larger number of variables used in the present study opened various possibilities for adjustment of the optimum allocation of resources under different situations.

The relative ranking of the breeding schemes or selection strategies considered did not differ much for the two optimization criteria, but the differences between the breeding schemes were relatively more pronounced for *$P(q)$* than for *ΔG* (Wegenast *et al.* 2010; Mi *et al.* 2011). With an increasing number of test locations at the expense of a slightly lower number of test candidates, the SD of both optimization criteria decreased nearby the optimum, without any significant loss in *ΔG* or *$P(q)$* (Wegenast *et al.* 2009). The coancestry coefficient among the selected DH lines was larger for breeding scheme S_1TC-DHTC than DHTC, caused by the additional selfing of the S_0 generation before DH induction (Wegenast *et al.* 2010). This higher coancestry coefficient reflects a reduced genetic variance σ^2_G among the selected DH lines and might lead to a reduced progress from long-term selection.

Factors influencing the optimization of selection progress

Progress from selection and the optimum allocation of resources are influenced by three global parameters: (1) the mode of selection (stage of selection, *i.e.,* among parental crosses, S_0 plants and/or DH lines, number vis-à-vis the size of crosses, selection criterion), (2) the budget available in a breeding program, and (3) the genetic architecture (*i.e.*, the ratio of variance components) of the breeding population(s) and trait(s) under consideration. The present study focused on optimizing the mode of selection for three different variance component ratios and levels of financial resources.

The genetic base of the breeding population can be broadened and the magnitude of genetic variance relative to other sources can be enhanced by incorporating and/or introgressing diverse germplasm. But given a breeding population, the genetic architecture (*i.e.*, variance component ratio) is trait and target environment specific. Sustaining or broadening the genotypic variance is a strategic, but the basic aim of a breeder is to ensure breeding success. In the present study, with a decreasing contribution of the genotypic variance, the optimum number of test locations increased at the expense of a smaller number of test candidates (Wegenast *et al.* 2008, 2009, 2010; Mi *et al.* 2011). This decreased genotypic variance led to lower values of ΔG and $P(q)$ and generally larger $SD_{\Delta G}$ and $SD_{P(q)}$. The study further showed that an enhanced budget led an to increased number of test candidates at both selection stages and enabled more intense selection. However, an increment in the budget has a non-linear, though, positive effect on selection intensity and heritability and consequently on progress from selection. A quadrupling of the budget from 10,000 to 40,000 testcross plots increased ΔG and $P(q)$ by about 8% and 20%, respectively.

The study also showed that an increase in the progress from selection by improving the mode of selection can be achieved by (1) application of parental selection and concentration of test resources on few crosses among the best elite parent lines, (2) the choice of the breeding scheme, (3) selection among all DH lines rather than performing selection first among and then within parental crosses and S_1 families, and (4) selection decision based on BLUPs or other criteria (Wegenast *et al.* 2008, 2009, 2010; Mi *et al.* 2011).

Selection of parental crosses

Selection of parental crosses at the time of initiation of a breeding cycle was found to be the most important factor maximizing the progress from selection and influencing the optimum allocation of test resources, which is in accordance with Utz (1982). Bernardo (2003) found the parental selection to be more important than the optimum allocation of test resources. In the present study, effective parental selection decreased the optimum number of initial parental crosses and sharply increased progress from selection (Wegenast *et al*. 2008). The parental selection was based on the parental mean of all possible pairwise combinations of parental inbreds, because the general varietal ability of a progeny can be predicted by the mean of the general combining abilities (GCA) of its parents (Wegenast *et al*. 2008). Thus, we measured the reliability of the parental selection by using the correlation between the mean phenotypic performance of the parental lines and the mean genotypic performance of the testcross progeny (ρ_P). The correlation ρ_P is the product of (1) the genotypic correlation ρ_G between the mean GCA of the parental lines and the mean GCA of their progeny lines and (2) the square root of the heritability of the parental cross. We developed a formula to calculate ρ_G, showing that it is a function of the ratio of additive x additive to additive variance (σ^2_{AA}:σ^2_A) if the trait in the considered population is in linkage equilibrium and higher order epistasis is neglected (Wegenast *et al.* 2008, Suppl. Mat.). The correlation ρ_G reaches its minimum value of 0.71 for large values of σ^2_{AA} and negligible σ^2_A, and it rapidly surpasses 0.9 for

$\sigma^2_{AA}:\sigma^2_A<1$ (Fig. 1). The correlation between a *single* random DH line and the parental mean (ρ_G') also depends solely on σ^2_A and σ^2_{AA}:

$$\rho_G' = \sqrt{0.5}\left(1 - \frac{\sigma^2_{AA}}{4\sigma^2_A + 2\sigma^2_{AA}}\right). \tag{1}$$

However, this correlation is lower than 0.71 (Fig. 1), indicating that single DH lines might differ strongly from their parental mean.

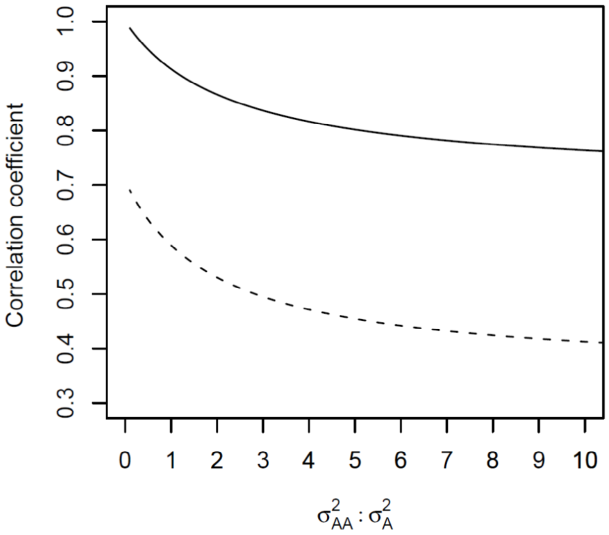

Fig. 1 Correlation coefficient between the mean GCA of two parental inbreds and (1) the mean GCA of their progeny lines (ρ_G, solid line), and (2) the GCA of a randomly chosen progeny line from the offspring of these parents (ρ_G', dashed line). The correlation is presented as a function of the ratio of additive x additive to additive variance $\boldsymbol{\sigma^2_{AA}:\sigma^2_A}$.

With more reliable parental selection (*i.e.*, increasing ρ_P), the optimum number of parental crosses at the first stage is decreased (Wegenast et al. 2008). In addition, the loss in ΔG, if the allocation is non-optimal, is clearly larger if ρ_P is high. This indicates that optimum allocation is much more critical if parental selection is considered in a breeding scheme. In the case of randomly chosen parents, the optimum curves are very flat, but ΔG is by far lower.

Breeding schemes

In breeding scheme DHTC, DH lines were produced from S_0 plants taken at random *i.e.*, without any selection. The testcrosses of these DH lines were evaluated at the first stage and then, the promising testcrosses were evaluated at the second stage, and the superior

DH lines were selected. In breeding scheme S_1TC-DHTC, S_1 families were produced from S_0 plants taken at random, an early test for testcross performance of the S_1 families was made at the first stage, and remnant seed of superior S_1 families was used for simultaneous haploid induction. Chromosome doubling was then performed using only haploid kernels produced in selected S_1 families after the first selection stage. Then testcrosses of DH lines were produced and their performance evaluated at the second selection stage.

Gordillo and Geiger (2008a, b) compared selection gains through recurrent selection for population improvement in which selection in each cycle was conducted in one, two, or three stages, and observed one-stage selection to be superior. The superiority of one-stage over multi-stage selection was basically due to shorter length of a cycle and it led to larger selection gain per year. As the accuracy of selection is as important as the length of a selection cycle (Longin *et al.* 2006), we focused on two-stage breeding schemes to balance the needs of a short cycle length and the accuracy of selection. In addition, the length of a selection cycle in applied breeding program depends on the number of generations that can be grown in one year by using off-season nurseries and on the number of generations that are needed to produce a sufficient amount of seed for testing. Thus, we focused on progress from selection on a per cycle basis. In breeding scheme DHTC, the optimum allocation of test resources was (1) a very large number of DH lines within the few best parental crosses, and testing their testcrosses at a small number of locations at the first stage and (2) testing testcrosses of a small number of the most promising DH lines at a large number of locations at the second stage. Thus, high selection intensity in the first stage was combined with an intensive evaluation of the candidates at the second stage.

The additional early testing for testcross performance in generation S_1 before production of DH lines in breeding scheme S_1TC-DHTC lengthens the selection cycle. However, this approach allows limiting the production of DH lines to the most promising S_1 families on the basis of early testing. Selection progress in breeding scheme S_1TC-DHTC was significantly larger than for DHTC for both optimization criteria ΔG and $P(q)$. The advantage of S_1TC-DHTC can be the result of the early testing and, additionally, the generation of new genetic variability during a second meiosis. This led to allocation of a larger part of the budget to the testing of DH lines of the selected S_1 families in combination with high selection intensity and an intensive evaluation at both selection stages. The optimum allocation of test resources in breeding scheme S_1TC-DHTC with parental selection involved similar numbers of test locations at both stages, a large number of S_1 families within the most promising crosses at the first stage, and a large number of DH lines within the best S_1 families at the second stage. There is allocation of more than 50% of the budget on the evaluation of the DH lines at the second selection stage.

Further, in S_1TC-DHTC the optimum number of initial parental crosses is larger than that for DHTC, which may enhance the probability to focus on promising parental crosses. This may be another factor responsible for the higher selection progress through S_1TC-DHTC than DHTC. In addition, the optimum number of total DH lines to be produced is extremely high compared with S_1TC-DHTC, resulting in lower costs of production of DH lines in S_1TC-DHTC. Thus, S_1TC-DHTC may be preferred when well-developed labs and protocols for DH production are not available or breeding populations are not well adapted (such as breeding populations having a large proportion of exotic germplasm) and are expected to yield a low proportion of DH lines with good performance. However, in the case of breeding programs with routine DH development and well adapted breeding populations, DHTC with the shorter cycle length and a larger annual progress from selection may be preferable. However, the production of DH lines from S_0 plants has a low number of recombinations and, thus, large blocks of parental chromosomes are transmitted to the progeny (Smith *et al.* 2008). This might cause negative effects on long-term selection (Bernardo 2009). Thus, it might be an option to introduce a selfing generation of S_0 plants before haploid induction in breeding scheme DHTC.

The considered breeding schemes were based on the assumption that two selfing generations of the DH lines are required to produce the necessary amount of seed for field evaluations. Selection of DH lines *per se* prior to testcross evaluation is considered but is neglected for the selection process on yield as the correlation between line *per se* and testcross evaluations is low for grain yield (*cf.,* Mihaljevic *et al.* 2005).

Selection strategies

Model studies in the literature are usually based on balanced datasets assuming equal numbers of candidates within each parental cross (*cf.*, Baker 1984; Bernardo 2003; Longin 2007). In addition, selection is usually performed first among parental crosses and then among the candidates produced within selected crosses. This selection among parental crosses enables breeders to concentrate their test resources on the most promising parental crosses in the subsequent steps in the selection cycle. However, this strategy also entails the risk of discarding individual superior candidates within rejected parental crosses. In applied breeding, however, the number of DH lines within parental crosses usually differs due to various reasons other than the superiority of crosses, such as the success of DH induction. No formula was available in the literature to work out the selection progress for unequal numbers of DH lines within parental crosses or S_1 families. Therefore, we developed Monte Carlo simulations to model this selection strategy. These models for variable cross and family sizes met the demand of (1) modeling a strong and a moderate variation in the number of DH lines produced in parental crosses and/or S_1 families depending on their expected performance or due to other factors and (2) a full utilization of budget. In this process, we developed two geometric series that depended on the total number of DH lines available, the rank of a given parental cross or S_1 family based on its relative performance, and a factor defining moderate or strong variation in the number of DH candidates within parental crosses and/or S_1 families (Wegenast *et al.* 2010).

The efficiency of both breeding schemes, DHTC and S_1TC-DHTC, increased when selection among all DH lines instead of sequential selections among parental crosses, S_1 families, and DH lines (Wegenast *et al.* 2010; Mi *et al.* 2011). The variable instead of fixed numbers of DH lines within crosses and S_1 families had small additional positive effect. The higher efficiency of the selection among all DH lines may be due to the selection of superior DH lines out of parental crosses or S_1 families with inferior mean performance. Thus, on the average, the final number of DH lines selected originated from more than one cross. This led also to a lower coancestry coefficient Θ among selected DH lines, and consequently, to a larger long-term progress from selection.

The present study revealed that for sequential selection among and within crosses and S_1 families, ΔG was largest when the best cross was selected and a large number of a DH lines was derived and evaluated from that, instead of evaluating a smaller number of DH lines derived from several crosses (Wegenast *et al.* 2008). However, the selection of all DH lines of one cross led to a high Θ among the selected DH lines and, thus, a smaller effective population size. Consequently, the genetic variance available for subsequent selection cycles is diminished with negative effects on success for long-term breeding. From a simulation study, Sánchez Rodríguez (1999) concluded that selection with unequal contributions of parents to the next generation, is often more efficient than selection limited by equal contributions, as progress from selection and genetic diversity of the selected fraction have a positive relationship. In the present study, a larger progress due to selection was achieved in combination with a smaller coancestry coefficient through the selection of DH lines disregarding the cross and family structure. Thus, this selection strategy takes care of both short-term and long-term gains in a balanced manner.

BLUP

BLUP is a general method for predicting random effects, whereas BLUE, the best linear unbiased estimator, is used for the estimation of fixed effects; random effects have a covariance structure whereas fixed effects have not (Lynch and Walsh 1998, Chap. 26; Bernardo 2002, Chap. 10). In contrast to animal breeding, the estimation of genotypic effects in plant breeding has up to now mostly been based on models with fixed genetic effects (Piepho *et al.* 2008). In BLUP, the variance components to calculate the mixed model equations (MME) are assumed to be known. In practice, these variance components need to be estimated by another statistical method (*i.e.,* analysis of variance or restricted maximum likelihood). Then, the term BLUP needs to be replaced by empirical BLUP, because the variance components are replaced by their estimates (Piepho *et al.* 2008). In this thesis, we considered given values of variance components, as our focus was rather on determining the influence of BLUP on the optimum allocation of test resources and not on the estimation of variance components.

The optimum allocation of test resources based on BLUP in the two maize breeding schemes considered reduced the number of test candidates in the first stage of the breeding scheme DHTC. Thus, a shift of the resources from the first to the second selection stage occurred. Consequently, promising candidates could be more extensively tested at the second selection stage. In breeding scheme S_1TC-DHTC, no clear trend was observed for the optimum allocation of test resources based on BLUPs compared with those based on the mean performance of the candidate, *i.e.,* BLUEs. On the whole, selection based on BLUPs had a small positive effect on the progress from selection in both breeding schemes compared to that based on BLUEs.

The formulas developed in the present study, though based on some simplifying assumptions, greatly facilitated the conduct of Monte Carlo simulations. With these formulas, BLUPs can be calculated separately for each candidate directly from the phenotypic data without matrix inversions. Based on these formulas, further studies on BLUPs can be conducted, where relationship matrices based on the marker data and/or variance component estimates based on field data are available.

General trends

For both breeding schemes and all selection strategies, it was advantageous to concentrate on few parental crosses among the best parent lines selected using reliable information from the previous breeding cycle. In breeding scheme S_1TC-DHTC, where early testing of S_1 families is carried out, it was optimal to conduct testing at 9-15 locations in both stages, in combination with ≥500 S_1 families at the first and ≥650 DH lines at the second stage. In breeding scheme DHTC, testing of ≥2500 DH lines at ≤6 test locations in the first and ≤450 DH lines at ≥12 locations in the second stage were optimal.

Response curves revealed that the optimum number of parental crosses was the most sensitive allocation factor influencing the progress from selection. As long as the number of parental crosses was near to the optimum, response curves were flat. A shift in the number of initial parental crosses, however, led to a significant decrease in the progress from selection. Response curves for S_1TC-DHTC were flatter than for DHTC, thus, the risk of working with a non-optimum allocation is lower for S_1TC-DHTC.

Prospects of model calculations

For the estimation of the progress from parental selection, the available phenotypic data were assumed to be balanced, *i.e.,* the same amount of phenotypic data was available to estimate the phenotypic value of each parental inbred. In addition, all parents were assumed to be unrelated and only crossed once, leading to the development of unrelated parental crosses. In practice, phenotypic values of parental lines are obtained with different magnitudes of testing (number of testers as well as test years and locations and related factors), parental lines are usually related and have varying coefficients of coancestry (Bernardo 1996), and are used to generate different number of parental crosses. Thus, varying test reliability and genetic relationship need to be considered and more complex crossing designs implemented to get results reflecting the situation prevalent in applied breeding programs, because this might change the choice of the parental lines and alter the optimum allocation of test resources.

The present study focused on the assumption of a balanced field design, *i.e.*, all candidates were assumed to be tested at the same number of test environments and replications using an identical field design. However, the number of test candidates and test environments often times differ among years due to the dynamic nature of applied plant breeding programs, and even differ within a year with respect to replications and due to the partial or complete loss of data on some test candidates. Consequently, the field design may be included in the model calculations instead of taking the simple entry means as basic data. We concentrated on selection for a single trait, whereas in applied breeding, additional traits (such as maturity, silage yield and quality, and stress resistance) are of interest. Thus, it might be preferable to work with a higher number of parental crosses as calculated in the present study; though, the general recommendation to concentrate on rather few parental crosses is clearly justified by the present study.

The genetic relationship matrices used in the MME for calculating the genotypic values in BLUP were based on quantitative-genetic theory. With the use of marker information, these genetic relationships could be estimated more precisely and this will enable the researcher to consider various relationships among DH lines and to account for the same.

With the increasing availability of cheaper molecular markers amenable to high-throughput genotyping, efforts are being made to predict the genotypic value of a genotype using information from all available markers. In simulation studies, this genomic selection method was found useful in DH populations for traits which are controlled by many QTL, have low heritability, and when the selection experiment is based on a small population size (Mayor and Bernardo 2009, Heffner *et al.* 2010). With BLUP, the prediction of genotypic values based on marker data was reported to be reliable (Lorenzana and Bernardo 2009). With the integration of BLUP-based genomic selection in the two breeding schemes investigated, new and more efficient schemes could be designed for hybrid maize breeding. This would require further studies on hybrid maize breeding schemes based on DH and using marker technology as well as high computation power.

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

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

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

Bernardo R (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 255-262

Bordes J, Charmet G, Dumas de Vaulx R,Lapierre A, Pollacsek M, Beckert M, Gallais A (2007) Doubled-haploid versus single-seed descent and S_1-family variation for testcross performance in a maize population. Euphytica 154:41-51

Bouchez A, Gallais A (2000) Efficiency of the use of doubled-haploids in recurrent selection for combining ability. Crop Sci 40:23-29

Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. Nature Rev Genet 3:22-32

Gordillo GA, Geiger HH (2008a) Optimization of DH-line based recurrent selection procedures in maize under a restricted annual loss of genetic variance. Euphytica 161: 141-154

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

Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant breeding with genomic selection: Gain per unit time and cost. Crop Sci 50:1681-1690

Hill WG (1976) Order statistics of correlated variables and implications in genetic selection programmes. Biometrics 32:889-902

Hill WG (1977) Order statistics of correlated variables and implications in genetic selection programmes. II. Response to selection. Biometrics 33:703-712

Hill WG, Caballero A (1992) Artificial selection experiments. Annu Rev Ecol Syst 23:287-310

Holland JB (2009) Increasing yield. In: Bennetzen JL, Hake S (eds) Maize handbook – Volume II: Genetics and genomics. Springer Science and Business, New York, pp 469-482

Jannink J-L, Abadie T.E. (1999) Inbreeding method effects on genetic mean, variance, and structure of recurrent selection populations. Crop Sci 39:988-997

Longin CFH (2007) Optimum allocation of test resources and comparison of alternative breeding schemes for hybrid maize breeding with doubled haploids. Dissertation, University of Hohenheim, Germany.

Longin CFH, Utz HF, Reif JC, Schipprack W, Melchinger AE (2006) 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 number and type of testers in two-stage selection for general combining ability. Theor. Appl. Genet. 114:393-402

Longin CFH, Utz HF, Reif JC, Wegenast T, Schipprack 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

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

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinnauer Associates Inc., Sunderland, MA, USA.

Mayor PJ, Bernardo R (2009) Genomewide selection and marker-assisted recurrent selection in doubled haploid versus F_2 populations. Crop Sci 49:1719-1725

Mi X, Wegenast T, Utz HF, Dhillon BS, Melchinger AE (2011) Best linear unbiased prediction (BLUP) and optimum allocation of test resources in maize breeding with doubled haploids. Theor Appl Genet 123:1-10

Neuner S, Edel C, Emmerling R, Thaller G, Götz K-U (2009) Precision of genetic parameters and breeding values estimated in marker-assisted BLUP genetic evaluation. Genet Sel Evol 41:26

Piepho HP (2009) Ridge regression and extensions for genomewide selection in maize. Crop Sci 49:1165-1176

Piepho HP, Möhring J (2006) Selection in cultivar trials – is it ignorable? Crop Sci 46:192-201

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

Robson DS, Powers L, Urquhart DS (1967) The proportion of genetic deviates in the tails of a normal population. Theor Appl Genet (Züchter) 37:205-216

Sánchez Rodríguez L (1999) Balanced vs. slightly unbalanced selection. Heredity 84:685-691

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

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

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

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

Wang J, van Ginkel M, Podlich D, Ye G, Trethowan R, Pfeiffer W, DeLacy IH, Cooper M, Rajaram S (2003) comparison of two breeding strategies by computer simulation. Crop Sci 43:1764-1773

Wang J, Pfeiffer WH (2007) Simulation modeling in plant breeding: Principles and applications. Agric Sci China 6:908-921

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

Wegenast T, Dhillon BS, Utz HF, Longin CFH, Maurer HP, Melchinger AE (2009) Comparison of selection strategies and optimization criteria in hybrid maize breeding with doubled haploids. Maydica 54:343-352

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

Wricke G, Weber WE (1986) Quantitative genetics and selection in plant breeding. Walter de Gruyter, Berlin

Yonezawa K, Yamagata H (1978) On the number and size of cross combinations in a breeding programme of self-fertilizing crops. Euphytica 27:113-116

7. SUMMARY

In hybrid maize breeding, the doubled haploid technique is increasingly replacing conventional recurrent selfing for the development of new lines. In addition, novel statistical methods have become available as a result of enhanced computing facilities. This has opened up many avenues to develop more efficient breeding schemes and selection strategies for maximizing progress from selection.

The overall aim of the present study was to compare the selection progress by employing different breeding schemes and selection strategies. Two breeding schemes were considered, each involving selection in two stages: (i) developing DH lines from S_0 plants and evaluating their testcrosses in stage one and testcrosses of the promising DH lines in stage two (DHTC) and (ii) early testing for testcross performance of S_1 families before production of DH lines from superior S_1 families and then evaluating their testcrosses in the second stage (S_1TC-DHTC). For both breeding schemes, we examined different selection strategies, in which variance components and budgets varied, the cross and family structure was considered or ignored, and best linear unbiased prediction (BLUP) of testcross performance was employed. The specific objectives were to (1) maximize through optimum allocation of test resources the progress from selection, using the selection gain (ΔG) or the probability to select superior genotypes ($P(q)$) as well as their standard deviations as criteria, (2) investigate the effect of parental selection, varying variance components and budgets on the optimum allocation of test resources for maximizing the progress from selection, (3) assess the optimum filial generation (S_0 or S_1) for DH production, (4) compare various selection strategies - sequential selection considering or ignoring the cross and family structure - for maximizing progress from selection, (5) examine the effect of producing a larger number of candidates within promising crosses and S_1 families on the progress from selection, and (6) determine the effect of BLUP, where information from genetically related candidates is integrated in the selection criteria, on the progress from selection.

For both breeding schemes, the best strategy was to select among all S_1 families and/or DH lines ignoring the cross structure. Further, in breeding scheme S_1TC-DHTC, the progress from selection increased with variable sizes of crosses and S_1 families, *i.e.,* larger numbers of DH lines devoted to superior crosses and S_1 families. Parental cross selection strongly influenced the optimum allocation of test resources and, consequently, the selection gain ΔG in both breeding schemes. With an increasing correlation between the mean testcross performance of the parental lines and the mean testcross performance of their progenies, the superiority in progress from selection compared to randomly chosen parents increased markedly, whereas the optimum number of parental crosses decreased in favor of an increased number of test candidates within crosses.

With BLUP, information from genetically related test candidates resulted in more precise estimates of their genotypic values and the progress from selection slightly increased for both optimization criteria ΔG and $P(q)$, compared with conventional phenotypic selection. Analytical solutions to enable fast calculations of the optimum allocation of test resources were developed. This analytical approach superseded matrix inversions required for the solution of the mixed model equations. In breeding scheme S_1TC-DHTC, the optimum allocation of test resources involved (1) 10 or more test locations at both stages, (2) 10 or fewer parental crosses each with 100 to 300 S_1 families at the first stage, and (3) 500 or more DH lines within a low number of parental crosses and S_1 families at the second stage. In breeding scheme DHTC, the optimum number of test candidates at the first stage was 5 to 10 times larger, whereas the number of test locations at the first stage and the number of DH lines at the second stage was strongly reduced compared with S_1TC-DHTC.

The possibility to reduce the number of parental crosses by selection among parental lines is of utmost importance for the optimization of the allocation of test resources and

maximization of the progress from selection. Further, the optimum allocation of test resources is crucial to maximize the progress from selection under given economic and quantitative-genetic parameters. By using marker information and BLUP-based genomic selection, more efficient selection strategies could be developed for hybrid maize breeding.

8. ZUSAMMENFASSUNG

In der Hybridmaiszüchtung werden Elternlinien in zunehmendem Maße durch die Doppelhaploiden (DH) Technik hergestellt, anstatt wie bisher durch fortgesetzte Selbstbefruchtung. Zudem ermöglicht eine verbesserte Rechenleistung die Nutzung neuer statistischer Methoden. Dies eröffnet neue Möglichkeiten, den Selektionserfolg durch effizientere Züchtungsprogramme und Selektionsstrategien zu maximieren.

Übergeordnetes Ziel dieser Studie war, den Selektionserfolg für zweistufige Züchtungsschemata und unterschiedliche Selektionsstrategien zu ermitteln. Folgende Züchtungsschemata für Zweistufenselektion wurden zu Grunde gelegt: (i) Schema DHTC, bei dem DH Linien aus S_0 Pflanzen hergestellt werden. Die Testkreuzungen dieser DH Linien werden in der ersten Stufe geprüft und die besten daraus erneut in der zweiten Selektionsstufe. (ii) In Schema S_1TC-DHTC wird ein früher Test auf Testkreuzungsleistung von S_1 Familien vor der Herstellung von DH Linien aus überlegenen S_1 Familien durchgeführt, welche dann in der zweiten Selektionsstufe auf ihre Testkreuzungsleistung geprüft werden. Für beide Zuchtschemata wurden unterschiedliche Selektionsstrategien untersucht, wobei die Varianzkomponenten und Budgets variierten, die Struktur der Kreuzungen und Familien berücksichtigt oder ignoriert wurde und die Methode der Best Linear Unbiased Prediction (BLUP) für Testkreuzungsleistung angewendet wurde. Die spezifischen Ziele waren (1) den Selektionserfolg zu maximieren durch das Ermitteln der optimalen Allokation der Testressourcen bezüglich des Selektionsgewinns (ΔG) und der Wahrscheinlichkeit ($P(q)$), überlegene Genotypen zu identifizieren, (2) den Effekt der Elternwahl, unterschiedlicher Varianzkomponenten und Budgets auf die optimale Allokation der Ressourcen zu untersuchen, (3) die optimale Generation (S_0 oder S_1) für die Herstellung von DH Linien in zwei Zuchtschemata zu ermitteln, (4) verschiedene Selektionsstrategien zu vergleichen, bei denen die Kreuzungs- und Familienstruktur berücksichtigt oder vernachlässigt wurde, (5) den Effekt der Herstellung einer größeren Kandidatenzahl aus vielversprechenden Kreuzungen und S_1 Familien auf den Selektionserfolg zu ermitteln und (6) den Einfluss von BLUP auf den Selektionserfolg zu untersuchen.

Die beste Strategie für beide Zuchtschemata war, zwischen allen S_1 Familien und/oder DH Linien unter Vernachlässigung der Kreuzungsstruktur zu selektieren. Im Zuchtschema S_1TC-DHTC wurde der Selektionserfolg noch weiter erhöht bei variabler Kreuzungs- und Familiengröße, indem eine größere Anzahl DH Linien in überlegenen Elternkreuzungen und S_1 Familien hergestellt wird. Die Elternwahl hatte in beiden Zuchtschemata einen bedeutenden Einfluss auf die optimale Allokation von Testressourcen und den Selektionsgewinn. Mit einer zunehmenden Korrelation zwischen der mittleren Testkreuzungsleistung der Elternlinien und ihrer Nachkommen stieg diese Überlegenheit deutlich an, während die optimale Anzahl an Elternkreuzungen abnahm zugunsten einer größeren Anzahl an Prüfkandidaten innerhalb der Elternkreuzungen.

Mit BLUP wird Information genetisch verwandter Kandidaten genutzt, um zu präziseren Schätzungen der genotypischen Werte zu gelangen. Der Selektionserfolg für beide Optimierungskriterien ΔG und $P(q)$ war etwas größer mit der auf BLUP basierenden Selektion als mit konventioneller phänotypischer Selektion. Analytische Lösungen wurden entwickelt, um die optimale Allokation der Testressourcen schnell berechnen zu können. Dieser analytische Ansatz machte die Matrixinversionen zur Lösung der Mixed Model Gleichungen überflüssig. In Zuchtschema S_1TC-DHTC beinhaltete die optimale Allokation von Testressourcen (1) 10 oder mehr Prüforte in beiden Selektionsstufen, (2) 10 oder weniger Elternkreuzungen mit jeweils 100 bis 300 S_1 Familien in der ersten Stufe, sowie (3) mindestens 500 DH Linien aus einer geringen Anzahl von Elternkreuzungen und S_1 Familien in der zweiten Selektionsstufe. In Zuchtschema DHTC war die optimale Anzahl Prüfkandidaten in der ersten Selektionsstufe 5 bis 10 mal größer als in der zweiten Selektionsstufe, während die Anzahl Prüforte in der ersten und die Anzahl DH Linien in der zweiten Stufe deutlich reduziert war im Vergleich zu S_1TC-DHTC.

Die Möglichkeit, die Anzahl der Elternkreuzung durch Selektion zwischen Elternlinien zu reduzieren, ist von größter Bedeutung für die optimale Allokation der Testressourcen und die Maximierung des Selektionserfolgs. Die sorgfältige Wahl der optimalen Allokation ist entscheidend für den maximalen Selektionserfolg unter gegebenen ökonomischen Bedingungen und quantitativ-genetischen Parametern. Die Einführung neuer Selektionsmethoden wie z.B. BLUP verspricht eine weitere Zunahme des Selektionserfolgs. Mit Hilfe von Markerinformationen und auf BLUP basierender genomischer Selektion können zukünftig noch effizientere Selektionsstrategien für die Hybridmaizüchtung entwickelt werden.

9. ACKNOWLEDGEMENTS

I am very grateful to my academic supervisor Prof. Dr. A.E. Melchinger for his advice, suggestions, and support during this thesis work. Thanks to Prof. Dr. H.-P. Piepho and Prof. Dr. J. Bennewitz for serving my graduate commitee.

Sincere thanks to Dr. C.F.H. Longin for his valuable support and for his introduction to the IRTG. Special thanks to Prof. Dr. H.F. Utz to his help in program development as well as his support and the fruitful discussions on selection theory. Many thanks to Dr. X. Mi for his valuable support in theory and simulation development and our interesting discussions. Thanks to Prof. Dr. B.S. Dhillon for his great support in paper-writing. Thanks to Dr. H.P. Maurer and PD Dr. J.C. Reif for being coauthors of my publications.

I very much appreciate the support of Prof. Dr. S. Chen in China. I want to deeply thank Dr. Li Liang, Li Haochuan, all members of the Maize Centre at CAU and all Chinese members of the IRTG for their support in China and their introduction to the Chinese culture. Thanks to C. Biernath, R. Bisharat, Dr. C. Böber, S.-Y. Chen, Dr. D. Ebersberger, Dr. T. Feike, Dr. H. Knörzer, Dr. Y. Kühl, R. Mendoza-Huaitalla, A. Menegat, Dr. A. Roth, and S. Zia for the time we shared in China and Germany. The financial support from Deutsche Forschungsgemeinschaft is also gratefully acknowledged.

Many thanks to H. Beck, S.U. Bhosale, C. Bolduan, B. Devezi Savula, Dr. S. Fischer, N. Friedl, C. Grieder, J. Jesse, E. Kokai-Kota, Dr. B. Kranz, M. Martin, F.J. Mauch, Dr. M. Muraya, Dr. E. Orsini, PD Dr. H.K. Parzies, H. Pöschel, V. Prigge, Dr. I.Y. Rabbi, PD Dr. J.C. Reif, Dr. W. Schipprack, Dr. T.A. Schrag, PD Dr. B. Stich, A. Strigens, D. van Inghelandt, V. Weber, and all unmentioned members of the institute and LSA for their support and for creating a pleasant work environment.

Finally, I want to thank Christina for her support during my studies and the beginning of my thesis as well as the good time we shared. I want to deeply thank my family, my friends, as well as Stephanie with Philipp and Johanna for their great support, especially during the stressful part of the thesis.

10 CURRICULUM VITAE

Name	Thilo Wegenast
Birth	June 16, 1976 in Sulz am Neckar
School Education	1982-1986, Elementary School, Sulz-Mühlheim 1986-1995, High School, Albeck Gymnasium Sulz a.N.; Abitur, 06/1995
Civil Service	09/1995 – 09/1996, Jugendzentrum Freudenstadt
University Education	10/1996 – 11/2000, "Lehramt an Realschulen", Pädagogische Hochschule Heidelberg; Erstes Staatsexamen, 11/2000 10/2002-03/2005, Agricultural Sciences, University of Hohenheim, Stuttgart; Bachelor of Science, 03/2005 04/2005-10/2006, Plant Breeding, University of Hohenheim, Stuttgart; Master of Science 10/2006 03/2007-01/2010, Doctoral student, Institute of Plant Breeding, Seed Science, and Population Genetics as well as International research training group "Sustainable resource use in the North China Plain", University of Hohenheim,Stuttgart

Professional experience	Collaboration on parental farm "Waldhof", Gerhard Wegenast, Sulz-Renfrizhausen 02/2001-04/2002, "Referendariat", Staatliches Realschullehrerseminar Schwäbisch Gmünd and Adalbert-Stifter-Realschule Heidenheim 05/2002-06/2002, "Schlossgut" Walther Badmann, Giengen an der Brenz 08/2004-10/2004, "Ministerium für Ernährung und Ländlichen Raum Baden-Württemberg", Stuttgart 10/2006-02/2007, Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim, Stuttgart Since 02/2010, Dow AgroSciences GmbH Rastatt

EIDESSTATTLICHE ERKLÄRUNG

Hiermit erkläre ich an Eides statt, dass die vorliegende Arbeit von mir selbst verfasst und lediglich unter Zuhilfenahme der angegebenen Quellen und Hilfsmittel angefertigt wurde. Wörtlich oder inhaltlich übernommene Stellen wurden als solche gekennzeichnet.

Die vorliegende Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Institution oder Prüfungsbehörde vorgelegt.

Insbesondere erkläre ich, dass ich nicht früher oder gleichzeitig einen Antrag auf Eröffnung eines Promotionsverfahrens unter Vorlage der hier eingereichten Dissertation gestellt habe.

Münster, Februar 2011,