Maydica 54 (2009): 353-362

MARKER-BASED PREDICTION OF HYBRID PERFORMANCE IN MAIZE SINGLE-CROSSES INVOLVING DOUBLED HAPLOIDS

T.A. Schrag[1], M. Frisch[2], B.S. Dhillon[1], A.E. Melchinger[1,*]

[1] *Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, D-70593 Stuttgart, Germany*
[2] *Institute of Agronomy and Plant Breeding II, Justus-Liebig-University Giessen, D-35392 Giessen, Germany*

Received October 8, 2009

ABSTRACT - Successful prediction of hybrid performance (HP) is an attractive alternative to expensive field testing for identification of superior hybrids and can greatly accelerate hybrid breeding programs. In combination with the doubled-haploid technology, which has greatly facilitated the development of maize lines, HP prediction based on molecular markers has gained importance in recent years. Here, we review the methods for HP prediction of single-crosses between heterotic groups. These methods are based on line *per se* performance, general combining ability, best linear unbiased prediction, genetic distances, and genetic effects estimated from molecular markers. We consider marker-based HP prediction with a focus on the use of doubled-haploid lines and discuss its application for three stages in hybrid breeding programs: (1) after field evaluation of experimental hybrids, (2) after field evaluation of testcross performance, and (3) before field evaluation of testcross performance. Types of hybrids, whose performance is to be predicted, are those having two (Type 2), one (Type 1), or no (Type 0) parental lines that are already evaluated for testcross performance. Various studies in maize indicate that efficiency of HP prediction depends not only on the prediction approach and the actual experimental data, but also on the type of hybrids. Finally, we present an outlook on the future of HP prediction, with a focus on the integration of high-throughput genotyping platforms in hybrid breeding programs.

KEY WORDS: General combining ability; Hybrid performance prediction; Selection; Best linear unbiased prediction.

INTRODUCTION

Performance prediction of single crosses has the potential to greatly improve the efficiency of applied hybrid breeding programs in various crop plants. Maize (*Zea mays* L.) germplasm is commonly organized in genetically divergent heterotic groups, and commercial hybrid varieties are commonly inter-group single crosses (HALLAUER, 1990; SMITH *et al*., 1999). In each heterotic group, a large number of inbred lines are developed per cycle. With the introduction of the doubled haploid (DH) technology, the number of DH lines generated per cycle has greatly increased as compared with inbred lines developed through conventional recurrent selfing (SCHMIDT, 2004; GALLAIS and BORDES, 2007). Hence, the number of possible cross-combinations between lines from two heterotic groups becomes extremely large. But only a small proportion of these crosses are evaluated in field trials owing to limited resources. Thus, it is of great importance to the breeder, to identify the most promising hybrids for field evaluation. The performance of hybrids, which have not undergone any field evaluation, can be predicted by employing field data available from related crosses and molecular marker data of the parental lines. Here, we (1) review the methods for hybrid performance (HP) prediction of inter-group single-crosses, (2) consider three different application scenarios for marker-based HP prediction with a focus on the use of DH lines, and (3) present an outlook on the future of HP prediction.

METHODS FOR PREDICTION OF HP

In recent years, prediction of HP of single-crosses has been a major research issue in view of its expected effect on the acceleration of hybrid breeding programs. A multitude of approaches have been proposed that are described below.

Line per se performance

Predicting the grain yield of single crosses in maize based on the *per se* performance of their

---

* For correspondence (fax + 49 711 45922343; e.mail: melchinger@uni-hohenheim.de).

parental inbred lines has not been effective due to masking non-additive effects (SMITH, 1986; HALLAUER, 1990). Further, genotype-environment interactions also lower this correlation (BERNARDO, 1991). The difference between the performance of a hybrid and the mean of its parents accounted for up to 76% of the grain yield of maize hybrids (HALLAUER and MIRANDA FILHO, 1988). Thus, line *per se* performance alone does not sufficiently explain the variance of grain yield among maize hybrids (DUDLEY *et al*., 1992).

General combining ability

SPRAGUE and TATUM (1942) developed the concept of general (GCA) and specific combining ability (SCA), which has been very useful in hybrid breeding. Estimates of GCA of the parental lines provide an established and simple approach to predict HP (COCKERHAM, 1967; MELCHINGER *et al*., 1987). But prediction based on GCA alone ignores SCA, which constitutes an important component of HP (GARDNER and EBERHART, 1966). Thus, including both GCA and SCA in the model improved the prediction of HP (BERNARDO, 1994; CHARCOSSET *et al*., 1998).

Phenotypic BLUP method

Best linear unbiased prediction (BLUP) was developed by HENDERSON (1975) originally for the prediction of breeding values in animal breeding. Its application in plant breeding was demonstrated by HILL and ROSENBERGER (1985) for the analysis of unbalanced data of germplasm evaluation trials in multiple environments. BERNARDO (1994) adopted BLUP for the performance prediction of untested single crosses based on phenotypic data of related hybrids. The BLUP approach also uses information on genetic relationships among the parental inbreds, based on coancestry coefficients estimated from pedigree records or molecular marker data. BERNARDO demonstrated the successful application of the approach with hybrids in an 6 × 9 factorial mating design (BERNARDO, 1994), and on a much wider scale with 16 data sets on 61 to 636 hybrids (4099 in total) belonging to different heterotic patterns and evaluated in multiple environments (BERNARDO, 1996a,c). Further studies indicated the robustness of the BLUP approach when inbred relationships were erroneously specified (BERNARDO, 1996b). BLUP was also successfully used to predict HP of 24 crosses in soybean (PANTER and ALLEN, 1995) and 66 hybrids in rice (XU and VIRMANI, 2000). The results of these studies have been promising. However, the full potential of molecular markers is not utilized. The marker-based coancestry coefficients indicate overall expectations for the whole genome and ignore specific genomic regions that are relevant for trait expression (CHARCOSSET *et al*., 1991).

Trait- and marker BLUP method

The above discussed BLUP approach based on trait data, also named T-BLUP, was extended by BERNARDO (1998a) to make use of both trait and marker data (TM-BLUP) for prediction of HP. In the TM-BLUP approach, identity by descent of unobservable quantitative trait locus (QTL) alleles was inferred from molecular marker data and used for modelling the covariances associated with QTL. However, when applied to an empirical data set of 464 single crosses from one heterotic pattern, TM-BLUP resulted only in marginal improvement for predicting single-cross performance, compared with the T-BLUP approach (BERNARDO, 1998b). This result was corroborated by a simulation study, which analyzed the influence of various factors, such as heritability, number of QTL, and map distance between markers and flanking QTL (BERNARDO, 1999).

Molecular genetic distances

Predicting HP with estimates of genetic distances (GD) between the parental lines based on random DNA markers was not successful for inter-group hybrids, as reviewed by MELCHINGER (1999). These findings were in agreement with theoretical (CHARCOSSET *et al*., 1991; CHARCOSSET and ESSIOUX, 1994) and simulation results (BERNARDO, 1992), that attributed the low correlation between mid-parent heterosis and GD to (1) no or only loose linkage of heterotic QTL with the molecular markers employed to estimate GD and (2) different linkage phases between the QTL and marker alleles in the maternal and paternal gametic arrays, as generally expected with inter-group hybrids.

The potential of GD estimated from trait-related markers has often been discussed (e.g., MELCHINGER, 1993), but there are not many published studies, especially for inter-group hybrids. In maize, DUDLEY *et al.* (1991) reported low correlations of Modified Rogers Distance (MRD) with hybrid yield if random markers were used, but also for markers that were significantly associated with the trait. They evaluated single crosses developed through diallel matings of 14 inbred lines, which originated from several heterotic groups, so that the 91 single crosses were not only inter-group hybrids. In rice, Joshi *et al.* (2001) tested 28 hybrids (14 restorer × 2 CMS lines), and mostly recorded low correlations of GD with yield of hybrids, irrespective of the use of random (RAPD, ISSR) or trait-related (RFLP, STMS) molecular markers to measure GD.

In contrast, Zhang *et al.* (1994) in diallel crosses of eight parents in rice, found higher correlations of HP or mid-parent heterosis with specific heterozygosity (calculated only with significant marker loci) than with general heterozygosity (calculated with all marker loci). Similarly, in intra-group crosses from a diallel of 10 rice lines, Cho *et al.* (2004) reported closer correlations of GD with HP or mid-parent heterosis when markers were pre-selected for association with heterosis. Also, in a study based on six CMS × five restorer lines in oilseed rape, pre-selection among several markers (morphological, isozymes, soluble proteins, RAPD) increased the magnitude of correlations between GD and yield (Yu *et al.*, 2005).

Marker-based GD was assessed not only for prediction of HP or mid-parent heterosis, but also for prediction of SCA (Lee *et al.*, 1989; Melchinger *et al.*, 1990). Such marker-based SCA predictions were used in combination with experimentally estimated GCA effects by Charcosset *et al.* (1998) to predict HP of untested single crosses. In diallel matings among 21 diverse silage maize lines, they compared several marker-based approaches to account for SCA. For inter-group crosses, the study indicated no advantage of predicting SCA with GD if compared with the additive model in which GCA was considered but SCA was ignored. However, the authors obtained higher prediction efficiencies using BLUP and factorial regression models.

Marker effect estimates

Molecular markers can be tested for their association with hybrid yield, and the estimates of their allelic effects can be employed for prediction of HP. Dudley *et al.* (1991) demonstrated that hybrid genotypic values which were estimated by scores across 29 significant loci were useful to predict yield of single crosses in maize. Vuylsteke *et al.* (2000) investigated the associations of large numbers of AFLP markers with HP and SCA for grain yield across inter-group maize hybrids. The sum of marker effects across significantly associated markers (“total sum of selected markers“, TCSM) provided an estimate for the genotypic value of the hybrids. The predictions obtained with a linear regression of HP on TCSM were encouraging, but the authors did not compare the approach with established procedures for prediction of inter-group hybrids such as GCA-based methods. Schrag *et al.* (2006) applied the TCSM approach to four factorials of Dent x Flint hybrids and compared it with a GCA approach and two approaches based on a combination of GCA and SCA for grain yield and grain dry matter content. Further, Schrag *et al.* (2007) extended the TCSM approach to consider multiple testing of loci, robustness to missing marker observations, and suitability for multiple alleles, which they named “total effects of associated markers” (TEAM) approach. The suitability of TEAM for multiple alleles enabled its application to multi-allelic haplotype blocks, which account for linkage disequilibrium between markers. Further, combining this method with a BLUP analysis of phenotypic data enabled the use of highly unbalanced field data (Schrag *et al.*, 2009), which is prevalent in applied breeding programs for hybrid maize. Joint analyses of performance of hybrids and *per se* performance of their parental lines allowed marker-based prediction of mid-parent heterosis, which in combination with the *per se* performance of parents was used for prediction of untested hybrids. Marker-based TEAM approaches were outperformed by BLUP based on GCA and SCA for prediction of grain yield and grain dry matter content of hybrids provided that pedigree-based relationship measures and covariance between GCA and line *per se* performance were considered (Schrag *et al.*, 2010). Further, the study indicated the potential of marker-based approaches for prediction of hybrids especially between untested lines, as is the case for newly developed lines.

APPLICATION OF MARKER-BASED PERFORMANCE PREDICTION OF HYBRIDS INVOLVING DH LINES

In recent years, the use of the DH technology has greatly facilitated and accelerated the development of maize lines (Schmidt, 2004; Röber *et al.*, 2005; Seitz, 2005). In conventional inbred line development, several generations of recurrent selfing are necessary to achieve high degrees of homozygosity. In contrast, in DH line development, complete homozygosity is achieved in an extremely short time, just by haploidization and subsequent chromosome doubling. Therefore, unlike conventionally developed inbreds, DH lines are neither subjected to *per se* performance assessment nor to early testing for testcross performance. Thus, the main differences between DH technology and conventional recurrent selfing are (1) line development through DH technology is more rapid than the conventional inbreeding process, needs less resources, and the number of lines can be greatly enhanced and (2) unlike conventional inbreds, the DH lines represent a random sample of lines from the parental cross because these have not undergone any selection for *per se* or testcross performance. For these reasons, the prediction of HP with molecular markers is especially attractive for DH lines as an alternative to expensive field testing and is greatly expected to accelerate the identification of superior hybrids.

The stages in DH-based hybrid breeding are: development of DH lines, evaluation of the *per se* performance of DH lines, development and evaluation of testcrosses of DH lines, and development and evaluation of experimental hybrids. Marker-based HP prediction can be introduced at various stages: (1) after field evaluation of experimental hybrids to predict performance of missing hybrids, (2) after field evaluation of testcross performance, and (3) before field evaluation of testcross performance. Three scenarios at these stages, involving different types of hybrids, are considered for performance prediction. Hybrid types are those having two (Type 2), one (Type 1), or no (Type 0) parental lines already evaluated for their testcross performance.

Scenario 1:
Performance prediction of Type 2 hybrids after field evaluation of experimental hybrids

Promising lines, which are selected after several stages of evaluation for their *per se* performance and GCA, are used as parents to produce experimental hybrids. The experimental hybrids are then field evaluated to identify the best crosses among these candidates. In complete factorials, all possible combinations between two sets of parental lines from opposite heterotic groups are produced. However, in practice, some crosses may fail. Because the number of such crosses is rather small, it is expected that an appreciable number of crosses of each parental line are available in the factorial set (Fig. 1a) and, therefore, GCA of all these parental lines can be estimated. Such missing crosses, of which both parental lines have been tested in hybrid combinations, are of Type 2. To assess the utility of producing and evaluating these crosses, their HP can be predicted utilizing data on all the other experimental hybrids in the factorial. As more and more parental lines are being routinely genotyped (Eathington *et al.*, 2007), the available molecular marker data of the parental lines can also be employed for the prediction.

A leave-one-out-validation scheme has been employed to validate and compare different approaches for performance prediction of Type 2 hybrids. Bernardo (1996a) employed this scheme to assess BLUP-based prediction. He studied 16 heterotic patterns with 480 to 8100 potential hybrids, of which 61 to 636 hybrids were actually tested (4099 in to-

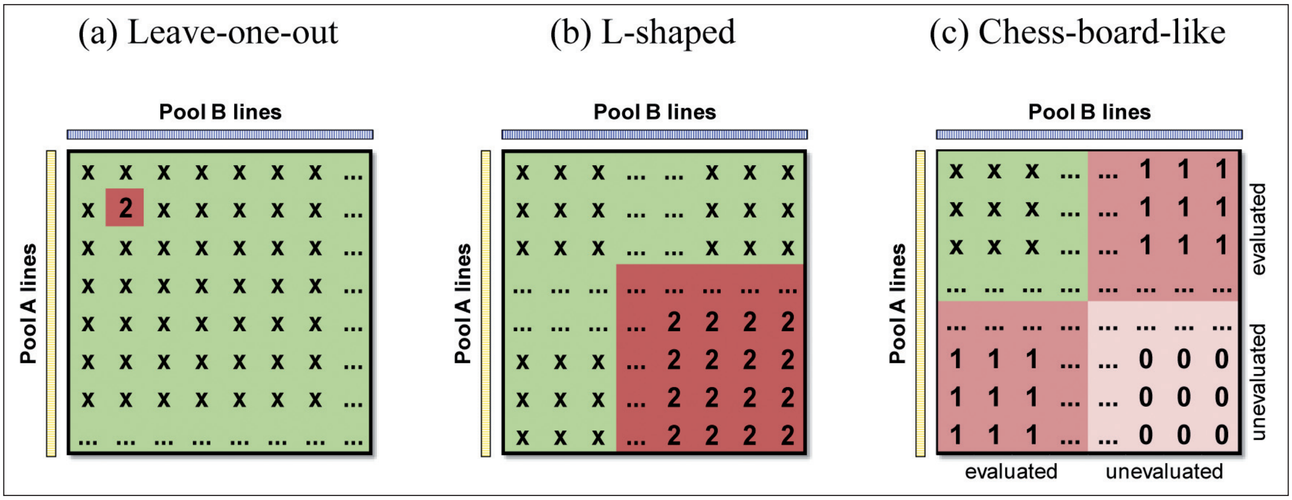

FIGURE 1 – Schemes for cross-validation within factorial mating designs, which comprise hybrids between lines from two heterotic groups A and B: (a) Leave-one-out, (b) L-shaped (rotated 90° clockwise in this diagram), and (c) chess-board-like cross validations comprising tested hybrids (“x”) and untested hybrids (“0”,”1”,”2”), where untested hybrids may have no (Type 0, “0”), one (Type 1, “1”), or two (Type 2, “2”) parental lines evaluated in testcrosses.

tal) and available for cross-validation. He iterated the prediction for each hybrid and determined the coefficient of correlation (*r*) between observed and predicted values for each heterotic pattern. For the sake of easy comparison among various studies (Table 1), we here report the square of the correlation coefficient (r^2). For grain yield, r^2 was 0.33 on average across the 16 heterotic patterns and ranged from 0.02 to 0.58. For grain dry matter content, r^2 ranged from 0.56 to 0.86 and was 0.79 on the average. In another study (BERNARDO, 1998b), the data set across 16 heterotic patterns was augmented to 4775 tested hybrids by including data of an additional year. In this study, r^2 ranged from 0.21 to 0.75 with an average of 0.41 for grain yield and from 0.75 to 0.88 and an average of 0.82 for grain dry matter content. VUYLSTEKE *et al.* (2000) evaluated the TCSM prediction approach also by employing the leave-one-out-validation scheme. They considered 53 hybrids among unrelated lines from a set of 78 diallel crosses. Performance prediction in different scenarios resulted in r^2 between 0.35 and 0.45. Comparison of the TCSM and GCA-based predictions was carried out by SCHRAG *et al.* (2006). Across four complete factorials, each with 44 to 98 hybrids (a total of 270 hybrids), the prediction efficiency r^2 for grain yield ranged from 0.46 to 0.73 for TCSM and 0.54 to 0.86 for GCA. For grain dry matter content, r^2 ranged from 0.59 to 0.87 for TCSM and 0.78 to 0.96 for GCA. MAENHOUT *et al.* (2007) used support vector machine regression based on AFLP and SSR markers for HP prediction in a leave-one-out-validation scheme. In a data set based on the evaluation of 2371 hybrids plus a 10-fold higher number of check hybrids, they obtained an r^2 of 0.34 for grain yield and 0.71 for moisture content.

The procedure of hybrid breeding in which performance of untested Type 2 crosses is predicted using marker data may involve five steps. These are: (1) compilation of field trial data of experimental hybrids from the current breeding cycle, and optionally from previous cycles as well as on *per se* performance of the parental lines; (2) compilation of routine marker data of all parental lines, and computation of coefficients of coancestry using pedigree or marker data if to be used in the prediction procedure; (3) estimation of marker effects from the marker and phenotypic data; (4) prediction of the performance of Type 2 hybrids with TCSM (VUYLSTEKE *et al.*, 2000; SCHRAG *et al.*, 2006), TEAM (SCHRAG *et al.*, 2007), or if line *per se* data are available, with the TEAM-LM approach (SCHRAG *et al.*, 2009); (5) seed production of the untested Type 2 hybrids identified to be superior on the basis of their predicted HP, and evaluation of these hybrids in field trials.

The aim in this scenario is the identification of superior experimental hybrids that were unintentionally missed in evaluations of factorial crosses in the current cycle. Through this approach, the risk of missing superior hybrids is reduced, but efforts and resources are required for the prediction analysis as well as for seed production and field evaluation of the hybrids that were selected as a result of the marker-based predictions. Further, the evaluation of these hybrids is delayed by one test season, which may be disadvantageous with regards to the selection gain per year.

Scenario 2: Performance prediction of Type 2 hybrids after testcross evaluation of lines

The first phase of the HP evaluation of candidate lines of each heterotic group involves their crossing with the best lines of the opposite group (treated as testers) and the evaluation of the resulting testcrosses because crossing all candidate lines of one heterotic group to all candidate lines from the opposite group would result in an unmanageable complete factorial. Such a procedure results in an incomplete factorial with an L-shape as shown in Fig.1b. In this incomplete factorial mating design, all untested hybrids are of Type 2, i.e., crosses between tested lines. SCHRAG *et al.* (2007) mimicked such a scenario and analyzed four Dent x Flint maize factorials of dimensions 11 x 4 to 14 x 7. In these factorials, five Dent and three Flint lines were randomly selected as opposite testers, so that 63 to 86% of the hybrids were regarded as tested hybrids. The prediction efficiency r^2 was 0.47 to 0.90 for GCA, 0.42 to 0.86 for the TEAM approach based on single markers (SM-TEAM), and 0.47 to 0.86 for a TEAM approach based on haplotype blocks (HB2-TEAM). CHARCOSSET *et al.* (1998) analyzed HP prediction in a half-diallel among 21 maize lines. They developed a validation procedure, which differed from the L-shape and used a restricted and balanced sampling procedure to ensure, that all untested hybrids were of Type 2 and to allow the variation of the percentage of tested hybrids. Considering the 182 crosses among unrelated lines only, a BLUP approach for marker-based prediction of SCA together with GCA provided the best results, the r^2 being 0.62 when the proportion of tested hybrids was approximately 30%, 0.70 for 50%, and 0.73 for 80%.

TABLE 1 - *Efficiency (r^2) of prediction of hybrid performance in various studies in maize*[a].

Author(s)	Validation approach	Type of hybrids[b]	Prediction approach	r^2 Grain yield	r^2 Grain dry matter/moisture	Remarks
Scenario 1						
Bernardo (1996a)	LOOV	Type 2	T-BLUP	0.02-0.58[c]	0.56-0.86	4099 hybrids
Bernardo (1998b)	LOOV	Type 2	T-BLUP	0.21-0.75	0.75-0.88	4775 hybrids
Vuylsteke *et al.* (2000)	LOOV	Type 2	TCSM	0.35-0.45	–	53 hybrids among unrelated lines
Schrag *et al.* (2006)	LOOV	Type 2	GCA	0.54-0.86	0.78-0.96	270 hybrids (4 experiments)
Schrag *et al.* (2006)	LOOV	Type 2	TCSM	0.46-0.73	0.59-0.87	270 hybrids (4 experiments)
Maenhout *et al.* (2007)	LOOV	Type 2	SVR	0.34	0.71	2371 hybrids
Scenario 2						
Charcosset *et al.* (1998)	Restricted sampling	Type 2	BLUP	0.73	–	182 hybrids, 80% predictors,
Charcosset *et al.* (1998)	Restricted sampling	Type 2	BLUP	0.70	–	182 hybrids, 50% predictors
Charcosset *et al.* (1998)	Restricted sampling	Type 2	BLUP	0.62	–	182 hybrids, 30% predictors
Schrag *et al.* (2007)	L-shape	Type 2	GCA	0.47-0.90	–	270 hybrids (4 exp.), 63 to 86% predictors
Schrag *et al.* (2007)	L-shape	Type 2	SM-TEAM	0.42-0.86	–	270 hybrids (4 exp.), 63 to 86% predictors
Schrag *et al.* (2007)	L-shape	Type 2	HB2-TEAM	0.47-0.86	–	270 hybrids (4 exp.), 63 to 86% predictors
Scenario 3						
Bernardo (1996c)	Modified LOOV	Type 1	T-BLUP	0.23	0.62	4099 hybrids
Maenhout *et al.* (2007)	Modified LOOV	Type 1	SVR	0.45	0.61	2371 hybrids
Schrag *et al.* (2008)	Chess-board	Type 1	TEAM-LM	0.51	0.72	400 hybrids
Bernardo (1996c)	Modified LOOV	Type 0	T-BLUP	0.13	0.43	4099 hybrids
Maenhout *et al.* (2007)	Modified LOOV	Type 0	SVR	0.20	0.41	2371 hybrids
Schrag *et al.* (2008)	Chess-board	Type 0	TEAM-LM	0.37	0.66	400 hybrids
Bernardo (1994, 1995)	Sampling	all types	T-BLUP	0.36-0.64	–	54 and 67 hybrids, 15 to 56% predictors

[a] Abbreviations: best linear unbiased prediction (BLUP), trait-data BLUP (T-BLUP), general combining ability (GCA), leave-one-out-validation (LOOV), support vector machine regression (SVR), total contribution of selected markers (TCSM), total effects of associated markers (TEAM), single-marker TEAM (SM-TEAM), haploblock TEAM (HB2-TEAM), line *per se* plus mid-parent heterosis TEAM (TEAM-LM).

[b] For prediction, hybrids with two (Type 2), one (Type 1), or no (Type 0) testcross-evaluated parental lines were considered.

[c] range of r^2.

The aim in this scenario is to predict the performance of untested crosses between DH lines that have already been evaluated for line *per se* performance and for GCA in testcrosses. Application of marker-based HP prediction of such untested experimental hybrids may involve the following four steps: (1) marker genotyping of DH lines that were selected on the basis of *per se* performance; (2) production and evaluation of testcrosses using a small number of lines or F1 testers from the opposite heterotic pool; (3) prediction of the performance of untested hybrids with TEAM-H and TEAM-LM (Schrag *et al.*, 2009) using available testcross data, and optionally also field data from the previous cycles; (4) seed production and evaluation of the experimental hybrids identified to be superior on the basis of predicted performance and testcross results. In this scenario, the use of line testers may be preferable over F1 testers as these have advantages for the estimation of marker effects and larger genetic variation, and result in testcrosses that are potential single-cross hybrid varieties.

Scenario 3: *Performance prediction of Type 1 and Type 0 hybrids before testcross evaluation of lines*

The DH lines, being completely homozygous, can be genotyped with molecular markers immediately after their production. Thus, marker-based performance prediction of hybrids involving new DH lines can be carried out before their field evaluations of *per se* or testcross performance. This can greatly enhance the efficiency particularly of hybrid breeding programs using DH technology. The HP of new DH lines can be predicted in combination with (1) already tested lines (Type 1 hybrids) or (2) other new DH lines, i.e., between two new untested lines (Type 0 hybrids) (Fig. 1c).

Extended forms of the leave-one-out-validation scheme, described earlier with regard to Type 2 hybrids, were used by BERNARDO (1996c) and MAENHOUT *et al.* (2007) to evaluate the efficiency of prediction of Type 1 and Type 0 hybrids. They discarded the data of all evaluated hybrids that involved the parents (one parent for Type 1 hybrid and both for Type 0 hybrid) of the hybrid to be predicted. Such a procedure was employed by BERNARDO (1996c) for BLUP-based prediction of HP in a large data set. For grain yield, the efficiency of prediction r^2 was on average 0.33 for Type 2, 0.23 for Type 1, and 0.13 for Type 0 hybrids. For grain dry matter content, r^2 was on average 0.79 for Type 2, 0.62 for Type 1, and 0.43 for Type 0 hybrids. Using a similar cross-validation scheme, performance of Type 1 and Type 0 hybrids was predicted by MAENHOUT *et al.* (2007) and of Type 2 hybrids by MAENHOUT *et al.* (2008). They used support vector machine regression based on AFLP and SSR markers, and reported the r^2 for the prediction of performance to be 0.34 for Type 2, 0.45 for Type 1, and 0.20 for Type 0 hybrids. For grain moisture, the r^2 was 0.71 for Type 2, 0.61 for Type 1, and 0.41 for Type 0 hybrids.

The studies presented above were all based on voluminous but highly unbalanced data, and the data structure represented the situation in practical breeding programs very well. However, evaluating the efficiency of prediction with cross-validation leaving out just one hybrid or at maximum all hybrids having a common parent with the hybrid in question (e.g., the entire column and row in the factorial data set), does not represent the situation in practical hybrid breeding programs, in which large numbers of both tested as well as new untested lines are evaluated together. Cross-validation with larger proportions of hybrids eliminated from the complete data set mimics this situation more closely. BERNARDO (1994, 1995) applied such a validation approach. Using a full BLUP model, he examined the prediction efficiencies for grain yield with varying proportions (15 to 56%) of predictor hybrids in small data sets of 54 and 67 hybrids, and reported r^2 to vary from 0.36 to 0.64. But he did not distinguish among Type 2, 1, and 0 hybrids.

Combining a sampling strategy for parental lines together with a large set of unbalanced data, SCHRAG *et al.* (2009) differentiated between the performance prediction of Type 1 and Type 0 hybrids. The data set was based on the evaluation of 400 of the 1504 possible factorial crosses of 47 Dent and 32 Flint lines. In every cross-validation round, 50% of the lines in each group were randomly chosen as tested lines, so that there were 25% hybrids of Type 2, 50% of Type 1, and 25% of Type 0. With the TEAM-LM approach for HP prediction, they observed r^2 for grain yield to be 0.51 for Type 1 and 0.37 for Type 0 hybrids, and for grain dry matter content to be 0.72 for Type 1 and 0.66 for Type 0 hybrids.

The marker-based HP prediction of crosses between unevaluated DH lines involves seven steps. These are: (1) genotyping of yet untested and unselected DH lines; (2) prediction of performance of Type 1 and 0 hybrids based on data of experimental hybrids and testcrosses from previous cycles e.g. with the TEAM-H approach (SCHRAG *et al.*, 2009). Additionally, marker-based prediction of *per se* performance of the DH lines may be undertaken; (3) selection of DH lines, using the predictions of HP and, if available, *per se* performance. In the absence of any other information on the new DH lines, this step of selection has to rely completely on marker effects estimated from data of the previous breeding cycles; (4) field evaluation of *per se* performance concurrently with production of testcrosses (line testers from the opposite heterotic pool may be chosen based on HP predictions); (5) selection of promising lines for seed increase on the basis of the *per se* performance and for field evaluation of the testcrosses; (6) HP prediction of experimental hybrids based on testcross performance of the lines under consideration and all data generated earlier; (7) seed production and evaluation of the experimental hybrids identified to be superior. The aim of using this procedure, which relies heavily on marker-based prediction, is to select DH lines based on their marker genotypes well in time to, first, reduce the expenditures on labour and field capacities devoted to seed increase and *per* se evaluation of DH lines, seed production and evaluation of testcrosses and experimental hybrids, and second, to accelerate the breeding process by identifying superior hybrids at a very early stage. The procedure may be modified by first field evaluating *per se* performance of DH lines and then genotyping selected DH lines.

CONCLUSIONS AND FUTURE PROSPECTS

High-throughput genotyping platforms have been established for large plant breeding programs during the last decade (Eathington *et al*., 2007). Marker data, generated by routine fingerprinting of new lines, has multiple uses such as quality control of breeding procedures and seed production, grouping of germplasm, mapping of traits, marker-assisted selection and marker-assisted backcrossing. Thus, molecular markers became an integral component of applied breeding programs. Multiplex marker platforms, which analyze thousands of single nucleotide polymorphism (SNP) markers simultaneously in a single step, are available for routine application in breeding of major crops (Hyten *et al*., 2008). In animal breeding, marker assays are in use, which provide more than 54,000 evenly spaced SNP probes spanning the bovine genome (Sellner *et al*., 2007). Platforms for whole-genome genotyping in human genetics already provide more than 1,000,000 SNPs on a single chip (Ziegler *et al*., 2008). Large numbers of known SNPs are required for development of such high-density chips. In maize, the genome sequencing of inbred B73 is completed (www.maizesequence.org) and will serve as a basis for re-sequencing efforts of several maize inbred lines in public and private research institutions, thereby, providing the foundation to identify huge numbers of SNPs. Owing to the ongoing advances in automated high-throughput marker technologies, the costs per marker data point are expected to considerably decline relative to the costs of phenotypic trait evaluation in field trials (Bernardo, 2008). In summary, marker-based HP prediction for selection of promising single crosses in combination with recent advances in DH and molecular marker technologies has the potential to further accelerate hybrid breeding programs and improve their efficiency.

Genomic selection (Meuwissen *et al*., 2001), which aims to overcome the problem of marker selection in fixed effects regression, has mostly been considered in animal breeding (Hayes *et al*., 2009), and very recently for inbred line development in crop plants (Bernardo and Yu, 2007; Heffner *et al*., 2009). Piepho (2009) considered it as a promising application for HP prediction. Despite of the availability of an extremely large number of SNP markers, an optimum marker density for cost-efficient prediction of HP in a given germplasm remains an issue to be investigated. In addition to molecular markers, transcriptome and metabolome data are becoming available and may also provide means for HP prediction, as was demonstrated for transcriptome analysis in maize (Frisch *et al*., 2010) and metabolome analysis in Arabidopsis (Gärtner *et al*., 2009). Joint analysis of the genomic, transcriptomic and metabolomic data for molecular-based prediction of HP needs to be studied for its potential to improve the efficiency of applied hybrid breeding programs.

REFERENCES

Bernardo R., 1991 Correlation between testcross performance of lines at early and late selfing generations. Theor. Appl. Genet. **82:** 17-21.

Bernardo R., 1992 Relationship between single-cross performance and molecular marker heterozygosity. Theor. Appl. Genet. **83:** 628-634.

Bernardo R., 1994 Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci. **34:** 20-25.

Bernardo R., 1995 Genetic models for predicting maize single-cross performance in unbalanced yield trial data. Crop Sci. **35:** 141-147.

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

Bernardo R., 1996b Best linear unbiased prediction of maize single-cross performance given erroneous inbred relationships. Crop Sci. **36:** 862-866.

Bernardo R., 1996c Best linear unbiased prediction of the performance of crosses between untested maize inbred. Crop Sci. **36:** 872-876.

Bernardo R., 1998a A model for marker-assisted selection among single crosses with multiple genetic markers. Theor. Appl. Genet. **97:** 473-478.

Bernardo R., 1998b Predicting the performance of untested single crosses: trait and marker data. pp. 117-127. *In:* K.R. Lamkey, J.E. Staub (Eds.), Concepts and breeding of heterosis in crop plants. ASA, CSSA, and SSSA, Madison, WI, USA.

Bernardo R., 1999 Marker-assisted best linear unbiased prediction of single-cross performance. Crop Sci. **39:** 1277-1282.

Bernardo R., 2008 Molecular markers and selection for complex traits in plants: learning from the last 20 Years. Crop Sci. **48:** 1649-1664.

Bernardo R., J. Yu, 2007 Prospects for genomewide selection for quantitative traits in maize. Crop Sci. **47:** 1082-1090.

Charcosset A., B. Bonnisseau, O. Touchebeuf, J. Burstin, P. Dubreuil, Y. Barriere, A. Gallais, J.B. Denis, 1998 Prediction of maize hybrid silage performance using marker data: comparison of several models for specific combining ability. Crop Sci. **38:** 38-44.

Charcosset A., L. Essioux, 1994 The effect of population-structure on the relationship between heterosis and heterozygosity at marker loci. Theor. Appl. Genet. **89:** 336-343.

Charcosset A., M. Lefort-Buson, A. Gallais, 1991 Relationship between heterosis and heterozygosity at marker loci: A theoretical computation. Theor. Appl. Genet. **81:** 571-575.

Cho Y.I., C.W. Park, S.W. Kwon, J.H. Chin, H.S. Ji, K.J. Park, S. McCouch, H.J. Koh, 2004 Key DNA markers for predicting heterosis in F1 hybrids of japonica rice. Breed. Sci. **54:** 389-397.

Cockerham C.C., 1967 Prediction of double crosses from single crosses. Der Züchter **37:** 160-169.

Dekkers J.C.M., 2007 Marker-assisted selection for commercial crossbred performance. J. Anim. Sci. **85:** 2104-2114.

Dudley J.W., M.A. Saghai Maroof, G.K. Rufener, 1991 Molecular markers and grouping of parents in maize breeding programs. Crop Sci. **31:** 718-723.

Dudley J.W., M. Saghai Maroof, K. Rufener, 1992 Molecular marker information and selection of parents in corn breeding programs. Crop Sci. **32:** 301-304.

Eathington S.R., T.M. Crosbie, M.D. Edwards, R.S. Reiter, J.K. Bull, 2007 Molecular markers in a commercial breeding program. Crop Sci. **47:** S154-S163.

Frisch M., A. Thiemann, J. Fu, T.A. Schrag, S. Scholten, A.E. Melchinger, 2010 Transcriptome-based distance measures for grouping of germplasm and prediction of hybrid performance in maize. Theor. Appl. Genet. (accepted).

Gallais A., J. Bordes, 2007 The use of doubled haploids in recurrent selection and hybrid development in maize. Crop Sci. **47:** S190-S201.

Gardner C.O., S.A. Eberhart, 1966 Analysis and interpretation of the variety cross diallel and related populations. Biometrics **22:** 439-452.

Gärtner T., M. Steinfath, S. Andorf, J. Lisec, R.C. Meyer, T. Altmann, L. Willmitzer, J. Selbig, 2009 Improved heterosis prediction by combining information on DNA- and metabolic markers. PLoS ONE **4:** e5220.

Hallauer A.R., 1990 Methods used in developing maize inbreds. Maydica **35:** 1-16.

Hallauer A.R., J.B. Miranda Filho, 1988 Quantitative genetics in maize breeding. Iowa State University Press, Ames, IA.

Hayes B.J., P.J. Bowman, A.J. Chamberlain, M.E. Goddard, 2009 Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. **92:** 433-443.

Heffner E.L., M.E. Sorrells, J.L. Jannink, 2009 Genomic selection for crop improvement. Crop Sci. **49:** 1-12.

Henderson C.R., 1975 Best linear unbiased estimation and prediction under a selection model. Biometrics **31:** 423-447.

Hill R.R., J.L. Rosenberger, 1985 Methods for combining data from germplasm evaluation trials. Crop Sci. **25:** 467-470.

Hyten D.L., Q. Song, I.Y. Choi, M.S. Yoon, J.E. Specht, L.K. Matukumalli, R.L. Nelson, R.C. Shoemaker, N.D. Young, P.B. Cregan, 2008 High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor. Appl. Genet. **116:** 945-952.

Joshi S.P., S.G. Bhave, K.V. Chowdari, G.S. Apte, B.L. Dhonukshe, K. Lalitha, P.K. Ranjekar, V.S. Gupta, 2001 Use of DNA markers in prediction of hybrid performance and heterosis for a three-line hybrid system in rice. Biochem. Genet. **39:** 179-200.

Lee M., E.B. Godshalk, K.R. Lamkey, W.L. Woodman, 1989 Association of restriction length polymorphism among maize inbreds with agronomic performance of their crosses. Crop Sci. **29:** 1067-1071.

Maenhout S., B. De Baets, G. Haesaert, E. Van Bockstaele, 2007 Support vector machine regression for the prediction of maize hybrid performance. Theor. Appl. Genet. **115:** 1003-1013.

Maenhout S., B. De Baets, G. Haesaert, E. Van Bockstaele, 2008 Marker-based screening of maize inbred lines using support vector machine regression. Euphytica **161:** 123-131.

Melchinger A.E., 1993 Use of RFLP markers for analysis of genetic relationships among breeding materials and prediction of hybrid performance. pp. 621-628. *In:* D.R. Buxton, R. Shibles, R.A. Forsberg, B.L. Blad, K.H. Asay, G.M. Paulsen, R.F. Wilson (Eds.), International Crop Science I. ASA, CSSA, and SSSA, Madison WI.

Melchinger A.E., 1999 Genetic diversity and heterosis. pp. 99-118. *In:* J.G. Coors, S. Pandey (Eds.), The genetics and exploitation of heterosis in crops. ASA, CSSA, and SSSA, Madison, WI.

Melchinger A.E., H.H. Geiger, G. Seitz, G.A. Schmidt, 1987 Optimum prediction of three-way crosses from single crosses in forage maize (*Zea mays* L.). Theor. Appl. Genet. **74:** 339-345.

Melchinger A.E., M. Lee, K.R. Lamkey, W.L. Woodman, 1990 Genetic diversity for restriction fragment length polymorphisms: relation to estimated genetic effects in maize inbreds. Crop Sci. **31:** 1033-1040.

Meuwissen T.H.E., B.J. Hayes, M.E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. Genetics **157:** 1819-1829.

Panter D.M., F.L. Allen, 1995 Using best linear unbiased predictions to enhance breeding for yield in soybean: II. Selection of superior crosses from a limited number of yield trials. Crop Sci. **35:** 405-410.

Piepho H.P., 2009 Ridge regression and extensions for genomewide selection in maize. Crop Sci. **49:** 1165-1176.

Röber F.K., G.A. Gordillo, H.H. Geiger, 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. Bericht über die 54. Tagung der Vereinigung der Pflanzenzüchter und Saatgutkaufleute Österreichs 2003, Gumpenstein. pp 1-6 (in German).

Schrag T.A., A.E. Melchinger, A.P. Sørensen, M. Frisch, 2006 Prediction of single-cross hybrid performance for grain yield and grain dry matter content in maize using AFLP markers associated with QTL. Theor. Appl. Genet. **113:** 1037-1047.

Schrag T.A., H.P. Maurer, A.E. Melchinger, H.-P. Piepho, J. Peleman, M. Frisch, 2007 Prediction of single-cross hybrid performance in maize using haplotype blocks associated with QTL for grain yield. Theor. Appl. Genet. **114:** 1345-1355.

SCHRAG T.A., J. MÖHRING, H.P. MAURER, B.S. DHILLON, A.E. MELCHINGER, H.-P. PIEPHO, A.P. SØRENSEN, M. FRISCH, 2009 Molecular marker-based prediction of hybrid performance in maize using unbalanced data from multiple experiments with factorial crosses. Theor. Appl. Genet. **118:** 741-751.

SCHRAG T.A., J. MÖHRING, B. KUSTERER, B.S. DHILLON, A.E. MELCHINGER, H.-P. PIEPHO, M. FRISCH, 2010 Prediction of hybrid performance in maize using molecular markers and joint analyses of hybrids and parental inbreds. Theor. Appl. Genet. (accepted).

SEITZ G., 2005 The use of doubled haploids in corn breeding. pp 1-7. Proc. 41st Annual Illinois Corn Breeders' School 2005, Urbana-Champaign.

SELLNER E.M., J.W. KIM, M.C. MCCLURE, K.H. TAYLOR, R.D. SCHNABEL, J.F. TAYLOR, 2007 Board-invited review: Applications of genomic information in livestock. J. Anim. Sci. **85:** 3148-3158.

SMITH O.S., 1986 Covariance between line per se and testcross performance. Crop Sci. **26:** 540-543.

SMITH O.S., K. HOARD, F. SHAW, R. SHAW, 1999 Prediction of single-cross performance. pp. 277-285. *In:* J.G. Coors, S. Pandey (Eds.), The Genetics and Exploitation of Heterosis in Crops. ASA, CSSA, and SSSA, Madison, WI.

SPRAGUE G.F., L.A. TATUM, 1942 General vs. specific combining ability in single crosses of corn. J. Am. Soc. Agron. **34:** 923-932.

VUYLSTEKE M., M. KUIPER, P. STAM, 2000 Chromosomal regions involved in hybrid performance and heterosis: their AFLP (R)-based identification and practical use in prediction models. Heredity **85:** 208-218.

XU W., S.S. VIRMANI, 2000 Prediction of hybrid performance in rice: comparisons among best linear unbiased prediction (BLUP) procedure, midparent value, and molecular marker distance. Int. Rice Res. Notes **25:** 12-13.

YU C.Y., S.W. HU, H.X. ZHAO, A.G. GUO, G.L. SUN, 2005 Genetic distances revealed by morphological characters, isozymes, proteins and RAPD markers and their relationships with hybrid performance in oilseed rape (*Brassica napus* L.). Theor. Appl. Genet. **110:** 511-518.

ZHANG Q.F., Y.J. GAO, S.H. YANG, R.A. RAGAB, M.A.S. MAROOF, Z.B. LI, 1994 A diallel analysis of heterosis in elite hybrid rice based on RFLPs and microsatellites. Theor. Appl. Genet. **89:** 185-192.

ZIEGLER A., I.R. KÖNIG, J.R. THOMPSON, 2008 Biostatistical aspects of genome-wide association studies. Biometrical J. **50:** 8-28.