

# Chromosomal regions involved in hybrid performance and heterosis: their AFLP®-based identification and practical use in prediction models

M. VUYLSTEKE\*†, M. KUIPER†<sup>1</sup> & P. STAM‡

†Keygene N.V., PO Box 216, Wageningen, The Netherlands; ‡Department of Plant Breeding, Wageningen Agricultural University, PO Box 386, 6700 AJ Wageningen, The Netherlands

In this paper, a novel approach towards the prediction of hybrid performance and heterosis is presented. Here, we describe an approach based on: (i) the assessment of associations between AFLP®<sup>2,3</sup> markers and hybrid performance and specific combining ability (SCA) across a set of hybrids; and (ii) the assumption that the joint effect of genetic factors (loci) determined this way can be obtained by addition. Estimated gene effects for grain yield varied from additive, partial dominance to overdominance. This procedure was applied to 53 interheterotic hybrids out of a 13 by 13 half-diallel among maize inbreds, evaluated for grain yield. The hybrid value, representing the joint effect of the genetic factors, accounted for up to 62.4% of the variation in the hybrid performance observed, whereas the corresponding efficiency of the SCA model was 36.8%. Efficiency of the prediction for hybrid performance was evaluated by means of a cross-validation procedure for grain yield of (i) the 53 interheterotic hybrids and (ii) 16 hybrids partly related to the 13 by 13 half-diallel. Comparisons in prediction efficiency with the ‘distance’ model were made. Because the map position of the selected markers is known, putative quantitative trait loci (QTL) affecting grain yield, in terms of hybrid performance or heterosis, are identified. Some QTL of grain yield detected in the present study were located in the vicinity of loci reported earlier as having quantitative effects on grain yield.

**Keywords:** AFLP®, heterosis, hybrid performance, QTL, *Zea mays*.

## Introduction

Heterosis is still the main reason for the success of the commercial maize (*Zea mays* L.) industry. Therefore, it is of particular interest to identify those genetic factors contributing to hybrid performance (HP) and/or heterosis, and a suitable method that could predict HP and/or heterosis with some accuracy before field evaluation of test hybrids. In maize, the main strategy that has been followed towards new ways of hybrid prediction during the last decade is based on the ‘distance’ model: heterosis, defined and measured as the superiority of

the hybrid over the midparent (the average performance of the two parents of the hybrid), is related to the genetic divergence between its parental lines (Lee *et al.*, 1989). The potential of this ‘distance’ model-strategy has been extensively tested in maize, where genetic distances were computed from RFLP data on parental inbreds (Lee *et al.*, 1989; Melchinger *et al.*, 1990a,b, 1992; Smith *et al.*, 1990; Dudley *et al.*, 1991; Boppenmaier *et al.*, 1993; Burstin *et al.*, 1995) and recently from PCR-based AFLP® markers (Ajmone Marsan *et al.*, 1998). The general tendency found was that the prediction efficiency of the ‘distance’ model is high when (i) hybrids between related lines (intraheterotic crosses) and (ii) hybrids between both related and unrelated lines (intra- and interheterotic crosses) are considered. However, correlations between genetic distances of unrelated lines only and their respective interheterotic crosses, were of low practical predictive value. This tendency is in good agreement with quantitative-genetic expectations (Charcosset & Essioux, 1994), ascribing the failure of the ‘distance’

\*Correspondence and present address: Aventis CropScience N.V., Nazarethsesteenweg 77, 9800 Astene, Belgium.

E-mail: marnik.vuyksteke@aventis.com

<sup>1</sup>Present address: Aventis CropScience N.V., J. Plateaustraat, 22, 9000 Gent, Belgium.

<sup>2</sup>AFLP® is a registered trademark of Keygene N.V.

<sup>3</sup>The methylation AFLP® method is subject to a patent application filed by Keygene N.V.

model for interheterotic crosses to the fact that linkage associations between markers and quantitative trait loci (QTL) generally differ randomly from one heterotic group to the other. Because only the interheterotic crosses are of commercial importance and of interest to the breeder, the practical value of the ‘distance’ model-approach is limited.

Other, more recent strategies for predicting HP, especially between unrelated lines, were proposed by Bernardo (1994) and Charcosset *et al.* (1998). The first method, based on best linear unbiased prediction (BLUP), uses covariances between HPs, estimated with marker data on parental inbreds, to predict the performance of an untested hybrid from the performance of related, tested hybrids. The second method is based on the principle that two hybrids with parents similar at the marker level, should display similar specific combining ability (SCA) values. Markers are used to generate covariates for SCA by means of principal components analysis.

In maize, many studies have been conducted to identify and map QTL for grain yield (GY) and yield components (Edwards *et al.*, 1987; Stuber *et al.*, 1992; Zehr *et al.*, 1992; Beavis *et al.*, 1994; Veldboom *et al.*, 1994; Ajmone Marsan *et al.*, 1995; Austin & Lee, 1996; Cockerham & Zeng, 1996; Eathington *et al.*, 1997; Austin & Lee, 1998). These studies suggested strongly that there are multiple QTL affecting GY throughout the genome. The results are generally in favour of the hypothesis of dominance of favourable alleles to explain the observed heterosis in GY, although overdominance at individual QTL (Stuber *et al.*, 1992) and epistasis cannot be ruled out.

In this paper, we present a novel approach towards the prediction of HP and heterosis. This approach is based on (i) the assessment of associations between AFLP markers and HP, respectively, SCA across a set of hybrids, and (ii) the assumption that the joint effect of genetic factors determined this way can be obtained by addition. The chromosomal position of the loci involved

in HP or heterosis is assumed to be in tight linkage with the marker locus as loose trait locus–marker associations will be broken up by accumulated recombination events during the establishment of the inbred lines. At the same time, because the map position of the selected markers is known, putative QTL affecting the trait of interest are identified.

## Materials and methods

### Plant material

Six inbred lines from Iowa Stiff Stalk Synthetic (BSSS), five from Lancaster Sure Crop (LSC) and two of miscellaneous origin (MO) were chosen as parents for a half-diallel mating design (Fig. 1). The six BSSS and the five LSC inbred lines from the half-diallel, were also chosen to be tested against Lo881 and C103 (LSC testers) and B14A and B73 (BSSS testers), respectively. Another 16 parental testcrosses were obtained by testing eight BSSS inbreds (Lo999, N28, A1, A2, A3, A4, A5 and A8) against Lo881 and C103. The pedigree backgrounds of all inbreds are given in Vuylsteke *et al.* (2000).

### Field trials

The 78 single-cross hybrids from the half-diallel and the 38 parental testcrosses were evaluated in 1994 at three different sites (Bergamo, Luignano and Turano) for GY (t ha<sup>-1</sup> at 15.5% moisture). The experimental design is described in Ajmone Marsan *et al.* (1998).

### Data handling

The half-diallel and the testcrosses had 22 interheterotic F<sub>1</sub> data in common. These duplicate and reciprocal F<sub>1</sub> data were averaged. For the application of the ‘distance’ model, all the 78 F<sub>1</sub> data were considered. Diallel analysis was performed according to Griffing (1956)

**Fig. 1** Schematic representation of the half-diallel mating design, involving six inbred lines from Iowa Stiff Stalk Synthetic (BSSS), five inbred lines from Lancaster Sure Crop (LSC) and two of miscellaneous origin (MO) chosen as parents. Intraheterotic crosses are marked by ‘x’, interheterotic crosses are marked by ‘⊗’.

	BSSS						LSC					MO	
	B14A	B37	B73	B84	Lo950	Lo951	C103	C123	Lo881	Mo17	Va59	H55	Pa91
B S S S	B14A	x	x	x	x	x	⊗	⊗	⊗	⊗	⊗	⊗	⊗
	B37		x	x	x	x	⊗	⊗	⊗	⊗	⊗	⊗	⊗
	B73			x	x	x	⊗	⊗	⊗	⊗	⊗	⊗	⊗
	B84				x	x	⊗	⊗	⊗	⊗	⊗	⊗	⊗
	Lo950					x	⊗	⊗	⊗	⊗	⊗	⊗	⊗
	Lo951						⊗	⊗	⊗	⊗	⊗	⊗	⊗
L S C	C103						x	x	x	x	⊗	⊗	
	C123							x	x	x	⊗	⊗	
	Lo881								x	x	⊗	⊗	
	Mo17									x	⊗	⊗	
	Va59										⊗	⊗	
M O	H55											⊗	
	Pa91												

Model I of Method 4 excluding parents and reciprocals, partitioning the performance of the hybrid ( $Y_{ij}$ ) between inbreds  $i$  and  $j$  classically as:

$$Y_{ij} = \mu + gca_i + gca_j + sca_{ij},$$

where  $\mu$  is the mean of the HPs,  $gca_i$  and  $gca_j$  are the GCAs of the inbreds  $i$  and  $j$ , respectively, and  $sca_{ij}$  is the SCA between inbreds  $i$  and  $j$ . GCA and SCA variances were highly significant ( $P < 0.01$ ) and of a similar order of magnitude.

Because the breeder is interested only in interheterotic crosses, and in order to remove the influence of group effects (intra- vs. interheterotic groups) on further analyses, e.g. marker selection, the intraheterotic group crosses (BSSS  $\times$  BSSS; LSC  $\times$  LSC) are excluded, reducing the dataset to 53 hybrids (Fig. 1). A simple ANOVA-test revealed a significant difference ( $P < 0.001$ ) between the intra- and interheterotic HP and SCA for GY (data not shown).

Considering the 53 interheterotic  $F_1$  data only, the GCAs of the parental lines were adjusted for the contribution of the other lines to the mean of the line in question, as there are a small number of parental lines (Falconer, 1989):

$$gca_i = \left( \frac{n-1}{n} \right) (\bar{i} - \mu),$$

where  $gca_i$  is the GCA of the inbred  $i$ ,  $n$  is the number of lines crossed with inbred  $i$ ,  $\bar{i}$  is the mean performance of parental line  $i$  and  $\mu$  is the mean of the HPs of (i) BSSS  $\times$  LSC and BSSS  $\times$  MO crosses, when inbred  $i$  is a BSSS inbred line (ii) LSC  $\times$  BSSS and LSC  $\times$  MO crosses, when inbred  $i$  is a LSC inbred line, and (iii) BSSS  $\times$  MO and LSC  $\times$  MO crosses, when inbred  $i$  is from miscellaneous origin (Fig. 1). If there is no dominance or epistasis, the performance of the hybrid from a cross between the  $i$ th female and the  $j$ th male is predicted by

$$EY_{ij} = \mu + gca_i + gca_j.$$

Any significant deviation from the observed  $Y_{ij}$ , referred to as SCA, must be caused by dominance or epistatic effects.

#### AFLP® and methylation AFLP® analysis

The 13 inbred lines were assayed for their respective AFLP and methylation AFLP profiles as described in Vuylsteke *et al.* (in press). A total of 1385 AFLP markers (592 *EcoRI/MseI* (E/M), 532 *PstI/MseI* (P/M) and 261 <sup>m</sup>*PstI/MseI* (<sup>m</sup>P/M) markers) out of 1539 AFLP

markers mapped on the B73  $\times$  Mo17 Recombinant Inbred (RI) high-density AFLP linkage map (Vuylsteke *et al.*, 1999), were chosen for further analysis.

#### 'Distance' model

Genetic Distances (GD) between pairs of inbred lines were calculated from AFLP data as complement to the genetic similarity coefficient originally devised by Jaccard (1908). Analogous to the partitioning of HP into GCA and SCA of parents, the GD values associated with 78  $F_1$  hybrids were partitioned as:

$$GD_{ij} = \mu + ggd_i + ggd_j + sgd_{ij},$$

with the analogous interpretation of general and specific genetic distances (Melchinger *et al.*, 1990b). Linear correlations were calculated for various combinations of HP, SCA, GD and SGD for the 78  $F_1$  hybrids.

#### Selection of markers using the Kruskal–Wallis test

To find markers that are, across the 53  $F_1$ s, significantly associated with HP, the rank sum test of Kruskal–Wallis, one of the three QTL mapping methods handled by MapQTL™ (van Ooijen & Maliepaard, 1996) has been used as a nonparametric statistical method. Given the parental molecular genotypes, the molecular genotypes of the 53 hybrid combinations were inferred, and converted and structured in a way in order to meet the input file structure of MapQTL™ (van Ooijen & Maliepaard, 1996): A for homozygous absence, H for heterozygosity and B for homozygous presence of the marker allele. Besides the genotype information at each locus, the map position of the loci and the quantitative data are needed as input for the Kruskal–Wallis test as performed by MapQTL. The output of the Kruskal–Wallis test lists for every locus (sorted according to the map) the name of the locus and its map position, the number of informative individuals, the Kruskal–Wallis test statistic and corresponding  $P$ -value, and, subsequently, for each class, respectively, the genotype, the mean rank, the arithmetic mean and the number of individuals in the class. B73 and Mo17 will be referred to as 'origin' of a selected marker allele, when the AFLP marker has been identified as a B73 or Mo17 marker, respectively, mapped on the B73  $\times$  Mo17 RI linkage map (Vuylsteke *et al.*, 1999).

In order to keep the overall false positive rate low, a stringent significance level of 0.001 and 0.005 was used in the selection of markers significantly associated with HP and SCA, respectively, across the 53  $F_1$ s. Only

those loci at the 0.001 and 0.005 significance level, respectively, for which all individuals are informative and the three genotypic classes are represented with at least one individual, were retained for further analysis.

**Model for the prediction of HP**

For each selected marker, the additive (*a*) and dominance (*d*) effects are estimated from the arithmetic means  $\bar{x}_{MM}$ ,  $\bar{x}_{mm}$  and  $\bar{x}_{Mm}$  of the genotypic classes B, A and H, respectively. *M* is considered as the marker allele represented by an AFLP fragment, while *m* indicates the absence of that marker allele (i.e. *m* encodes one or different other alleles at the same marker locus).

If the trait is controlled by *nl* loci acting independently (no epistasis), the genotypic value of the *F*<sub>1</sub> from the cross *i* × *j* can be written, using the notation of Hayman (1954), as:

$$Y_{ij} = C + \sum_{l=1}^{nl} a_l(\theta_l^i + \theta_l^j)/2 + d_l(1 - \theta_l^i/\theta_l^j)/2.$$

with  $C = \sum_{l=1}^{nl} c_l$ , the mean of all homozygotes over all loci, *a*<sub>*l*</sub> and *d*<sub>*l*</sub> the additive and dominance effects, respectively, for each locus *l*, and  $\theta_l^i$  representing the genotype of hybrid *ij* at locus *l*, which takes the value -1, 0 and +1 for genotypes *mm*, *Mm* and *MM*, respectively.

Finally, a hybrid value *TCSM*<sub>*ij*</sub> can be calculated for any hybrid as *TCSM*<sub>*ij*</sub> = *Y*<sub>*ij*</sub> - *C*, representing the total contribution of the selected markers (TCSM) in terms of their *a*<sub>*l*</sub> and *d*<sub>*l*</sub> estimates. Different TCSMs can be calculated as a function of the significance level used in selecting markers, resulting in a *TCSM*<sub>0.001</sub>, *TCSM*<sub>0.0005</sub> and *TCSM*<sub>0.0001</sub>. Linear regression of the HP on the TCSM results in a model for the prediction of the HP.

Note that, as a parental line is supposed to be either homozygous for the absence or homozygous for the presence of the marker alleles showing significant association with QTL of the trait of interest, its TCSM *per se* reduces to

$$TCSM_{ii} = \sum_{l=1}^{nl} a_l \theta_l^i.$$

**Model for the prediction of GCA**

Analogous to the partitioning of HP into GCA and SCA of parents, the *TCSM*<sub>*ij*</sub> of the hybrid *Y*<sub>*ij*</sub> between inbreds *i* and *j* can be written as:

$$TCSM_{ij} = \mu + GCSM_i + GCSM_j + SCMSM_{ij},$$

with the analogous interpretation of general and specific contributions of selected markers. The GCSM

of a line *i* is calculated as the deviation of its mean from the overall mean  $\mu$ , adjusted for the contribution of the other lines to the mean of the line in question (Falconer, 1989) in a analogous way as for GCA (Fig. 1).

Linear regression of the GCAs of the parental lines calculated for the trait of interest on the GCSMs results in an additive model for the prediction of the GCA.

An ‘expected’ TCSM of the hybrid from a cross between the *i*th female and the *j*th male can now be calculated as

$$ETCSM_{ij} = \mu + GCSM_i + GCSM_j.$$

**Models for the prediction of SCA**

There are two alternative models for the prediction of SCA based on selected markers.

**1** The difference between the calculated and ‘expected’ TCSM for a hybrid results in an estimation of the SCMSM of the two parental lines in combination. Linear regression of the SCAs of the hybrids calculated for the trait of interest on the SCMSMs results in a **first model for the prediction of the SCA**.

**2** In a way similar to finding markers significantly associated with HP, markers associated with SCA can be selected. The estimates of *a*<sub>*l*</sub> and *d*<sub>*l*</sub> of the marker alleles selected as being significantly associated with SCA are used to calculate a TCSM value of any hybrid. Linear regression of the SCAs on the TCSMs results in a **second model for the prediction of the SCA**.

**Allelic divergence among groups**

Allelic divergence (*ald*) among groups of inbreds at the marker loci and the QTL produces linkage disequilibrium between marker loci and QTL involved in SCA (Charcosset & Essioux, 1994). Because specific heterotic groups like BSSS and LSC have been classified on the basis of intra- and interheterotic heterosis, these groups should differ for their allelic frequencies at the QTL that exhibit dominance effects. As we were able to determine group membership of the parental inbreds, allelic divergence among the two major groups for the markers showing significant association with SCA has been calculated as follows:

$$ald = (|f_1 - f_2|),$$

where *f*<sub>1</sub> and *f*<sub>2</sub> are the allelic frequencies at the marker locus in group 1 (BSSS) and group 2 (LSC), respectively. High *ald* values must (i) provide evidence for the correlation between SCA and heterozygosity at marker loci, and (ii) support the linkage association between the

selected marker loci and QTL exhibiting dominance effects. The allelic frequency at the marker locus in the third group containing the two parental lines of miscellaneous origin, was left out of consideration.

#### Evaluation of the model for the prediction of HP

A first type of cross-validation performed to evaluate the additive model for prediction of HP is by a jack-knife sampling procedure. The jack-knife sampling procedure requires the partition of the initial set of  $N$  hybrids into (i) a set of  $N - 1$  predictor hybrids used for parameter estimation and (ii) one 'removed' hybrid used to compare predicted HP with observed HP. At each iteration, the selection of markers associated with HP and the calculation of the corresponding TCSM was repeated. Evaluation of prediction efficiency was made by examining plots of observed vs. predicted values and two synthetic statistics: (i) the standard error (SE) estimated as:

$$SE = \sqrt{s^2} \times \left( 1 + \frac{1}{n} + (x_0 - \bar{x})^2 / \sum (x_i - \bar{x})^2 \right),$$

where  $s^2$  is the sample variance,  $n$  is the number of observations,  $x_0$  is the predicted value,  $\bar{x}$  is the mean of the observed values and  $x_i$  is the observed value  $i$ ; (ii) the coefficient of determination ( $r^2$ , squared correlation coefficient) between observed and predicted values. The minimum value SE can reach is equal to  $\sigma$ , because a new hybrid will show some variation around the regression, equal to at least the residual variance  $\sigma^2$ . A second source of variation to be taken into account is the inaccuracy of the regression line: the estimates of the regression coefficients are stochastic, as they are based on a limited set of observations.

A second type of cross-validation to evaluate the additive models for prediction of HP is by linear regression of the HP of additional single crosses on their corresponding TCSM. In this study, the 16 parental test crosses were chosen. Evaluation of predic-

tion efficiencies was made by examining plots of observed vs. predicted values and the corresponding  $r^2$  values.

All computations in modelling HP, GCA and SCA, and in cross-validating HP were performed using the GENSTAT program (Genstat-5-Committee, 1993).

## Results

#### Relationship of genetic distance to HP and SCA

The estimates of linear correlations ( $r$ ) of GD and SGD calculated from the total marker dataset with HP and SCA for GY, respectively, are presented in Table 1. It must be emphasized that the results obtained from the BSSS  $\times$  BSSS and LSC  $\times$  LSC groups of crosses, although these groups are of minor interest, should be interpreted with caution because of their small number of observations. The  $r$ -value of GD with HP for the entire set of 78 hybrids was highly significant ( $P < 0.001$ ) but of moderate size (0.48). By contrast, a lack of relationship was noted between GD and HP in the three subsets of crosses. The  $r$ -value of SGD with the SCA effect was for the entire set of crosses and the BSSS  $\times$  BSSS subset highly positive (0.81 and 0.86, respectively) and highly significant ( $P < 0.001$ ). In addition, highly significant ( $P < 0.001$ ) and of a high magnitude was the correlation found in the subset of unrelated lines (0.64). The  $r$ -values of GD and SGD calculated from the total marker data set with HP and SCA for GY calculated from the 78  $F_1$  data from the half-diallel only, were similar to those reported by Ajmone Marsan *et al.* (1998) (data not shown).

#### Prediction of Hybrid Performance, GCA and SCA for GY

Table 2 gives an output list of the 20 marker alleles selected as being significantly ( $P < 0.001$ ) associated with QTL alleles contributing to GY, as well as their corresponding map position, 'origin', Kruskal-Wallis

Variables	Crosses ( $n$ )			
	All (78)	BSSS $\times$ BSSS (15)	LSC $\times$ LSC (10)	Unrelated lines (53)
GD - Tot	HP 0.48***	0.30	0.28	0.19
SGD - Tot	SCA 0.81***	0.86***	0.59	0.64***

\*\*\* $P < 0.001$ .

**Table 1** Linear correlations of genetic distance (GD) and specific genetic distance (SGD) based on the total marker data set, with hybrid performance (HP) and specific combining ability (SCA) for grain yield, for the total set of 78 single crosses and for different subsets of single crosses

**Table 2** Map position, 'origin', Kruskal–Wallis test statistic ( $K$ ), the means for the three genotypic classes and the  $a$ - and  $d$ -effects for the 20 marker alleles selected as being significantly ( $P < 0.001$ ) associated with QTL alleles contributing to the hybrid performance for grain yield, across the 53 interheterotic crosses

No.	Map position	Marker	Origin <sup>a</sup>	$K$	$a$	$d$
1	1/94.5	E42M4750-396.5	Mo17	15.248****	1.04	0.37
2	2/24.5	P13M60-59.0	Mo17	16.06****	-1.3	-0.1
3	2/30.4	P12M50-146.4	Mo17	15.527****	-1.05	-0.26
4	2/53.4	P18M60-419.7	Mo17	13.882***	-1.23	-0.04
5	3/49.5	<sup>m</sup> P12M48-587.1	Mo17	17.21****	-0.88	-0.56
6	4/56.2	E35M5154-82.0	B73	19.283*****	1.48	-0.08
7	4/58.0	E38M51-139.7	Mo17	19.338*****	-1.53	0.24
8	4/58.0	E39M54-314.1	Mo17	19.338*****	-1.53	0.24
9	4/58.0	E35M50-415.8	B73	17.045****	1.14	-0.28
10	5/20.1	P12M61-93.5	B73	18.192****	1.11	-0.35
11	6/10.3	E39M62-168.6	B73	15.226****	0.96	-0.13
12	6/10.8	E39M47-413.1	B73	23.814*****	1.09	-0.58
13	6/10.8	E45M60-582.5	B73	14.222***	0.99	-0.38
14	6/64.7	P13M62-473.2	B73	21.219*****	1.47	-0.12
15	6/68.4	E42M6061-158.4	B73	16.892****	1.12	-0.33
16	8/124.6	E33M50-148.9	B73	13.990***	0.85	-0.65
17	8/124.7	E33M50-148.1	Mo17	17.210****	-0.88	-0.56
18	9/0.0	E39M50-174.0	Mo17	18.861*****	-1.17	-0.28
19	9/54.1	E38M51-71.6	B73	20.229*****	1.31	0.32
20	9/58.6	E35M50-228.1	Mo17	13.223***	-1.52	0.48

\*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0005$ , \*\*\*\*\* $P < 0.0001$ .

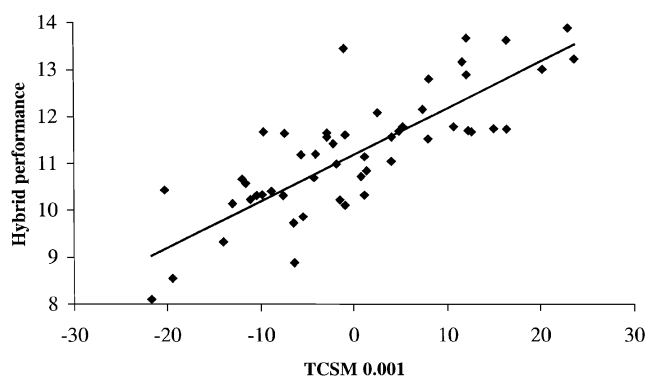
<sup>a</sup> B73 and Mo17 will be referred to as 'origin' of a selected marker allele, when the AFLP marker has been identified as a B73 or Mo17 marker, respectively, mapped on the B73 × Mo17 RI linkage map (Vuylsteke *et al.*, 1999).

test statistic and the corresponding  $a$ - and  $d$ -values. It is clear from the  $a$ - and  $d$ -values that the marker alleles selected for HP for GY fit single-gene models with additive and partial dominance effects ( $0 \leq |d| < |a|$ ).

The selected markers are clearly confined to particular regions of chromosomes, rather than being evenly distributed across the entire maize genome (residing on eight of the 10 chromosomes). Only one putative QTL of GY was revealed on 1/94 where Mo17 contributed the superior allele. In contrast, B73 contributed the superior allele at the putative QTL on 4/56.2-58.0, 5/20.1, 6/10.3-10.8, 6/64.7-68.4, 8/124.6 and 9/54.1.

The  $r$ -values of TCSM with HP for GY (0.79, 0.78 and 0.77, respectively) calculated for different numbers of selected markers (20, 16 and 7, respectively) are very highly significant ( $P < 0.001$ ) and of a much higher magnitude than the  $r$ -values of HP with GD (Fig. 2).

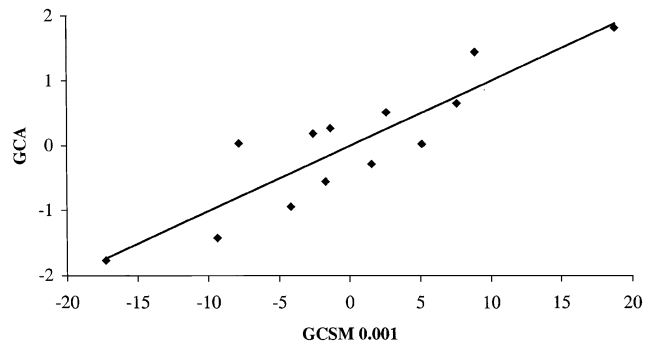
The cross Lo881 × Lo951, of which GY is amongst the highest (12.94 t ha<sup>-1</sup> at 15.5% moisture), has the highest TCSM<sub>0.001</sub> value (23.65) for GY among the 53 hybrids (data not shown). The maximal TCSM<sub>0.001</sub> value that can be reached, based on the maximal contribution of each selected marker listed in Table 2 equals 23.65 ( $\sum_{i=1}^{20} |a_i|$ ). This means there is no additional gain in GY possible using the QTL detected in the germplasm under consideration.



**Fig. 2** Observed hybrid performance (t ha<sup>-1</sup> at 15.5% moisture) vs. the total contribution of the markers selected at a significance level of 0.001 (TCSM<sub>0.001</sub>). The 53 interheterotic crosses are considered. The straight line represents the linear regression of hybrid performance on TCSM<sub>0.001</sub>.

The  $r$ -values of GCSM with GCA for GY (0.88, 0.88 and 0.87, respectively) for the 13 inbred lines, calculated for different numbers of selected markers (20, 16 and 7, respectively) are very highly significant ( $P < 0.001$ ) (Fig. 3).

The  $r$ -values of SCSM with SCA for GY (first model for the prediction of the SCA) calculated for different



**Fig. 3** General combining ability (GCA) ( $\text{t ha}^{-1}$  at 15.5% moisture) vs. the general contribution of the markers selected at a significance level of 0.001 ( $\text{GCSM}_{0.001}$ ). The 13 inbred lines are considered. The straight line represents the linear regression of GCA on  $\text{GCSM}_{0.001}$ .

numbers of selected markers (20, 16 and 7, respectively) are very highly significant ( $P < 0.001$ ) but of moderate size (0.49, 0.47 and 0.48, respectively) and of a lower magnitude than the  $r$ -values of SGD with SCA.

Table 3 gives an output list of the 25 marker alleles selected as being significantly ( $P < 0.005$ ) associated with QTL alleles contributing to SCA for GY as well as their corresponding map locus, 'origin', Kruskal–Wallis test statistic and the corresponding  $a$ - and  $d$ -values. It is clear from the  $a$ - and  $d$ -values that the marker alleles selected for SCA for GY fit single-gene models with only overdominance effects ( $|d| > |a|$ ). Again, the selected markers are confined to particular regions of the chromosomes, rather than being evenly distributed across the entire maize genome (they reside on seven of the 10 chromosomes). The selected markers are showing positive overdominance (e.g. 1/74.7), as well as negative overdominance (e.g. 1/53.1). Where positive overdominance occurs, the superior allele originates evenly from B73 and Mo17. Simultaneous fit of the 25 selected marker alleles (second model for the prediction of the SCA) accounted for 36.8% of the SCA variance among the 53 hybrids which is of a higher extent than explained by SCSM, but of a lower extent than explained by SGD.

**Table 3** Map position, 'origin', Kruskal–Wallis test statistic ( $K$ ), the means for the three genotypic classes, the  $a$ - and  $d$ -effects and the allelic divergence ( $ald$ ) for the 25 marker alleles selected as being significantly ( $P < 0.005$ ) associated with QTL alleles contributing to the specific combining ability for grain yield across the 53 interheterotic crosses

No.	Map position	Marker	Origin <sup>a</sup>	$K$	$a$	$d$	$ald$
1	1/53.1	E33M50-169.1	B73	11.420**	0.125	-0.495	80
2	1/69.0	P12M50-278.2	Mo17	11.420**	-0.125	-0.495	80
3	1/74.7	P18M49-117.7	B73	11.093**	0.453	0.734	100
4	2/70.4	E45M60-137.3	Mo17	11.093**	-0.453	0.734	100
5	3/12.2	P12M61-154.0	Mo17	11.206**	-0.473	-0.702	100
6	3/34.6	E35M49-58.9	B73	11.093**	0.453	0.734	100
7	3/70.4	E45M48-257.8	Mo17	10.965**	-0.351	-0.673	100
8	3/91.4	P12M61-159.3	Mo17	10.965**	-0.351	-0.673	100
9	3/96.6	P18M49-261.7	Mo17	11.093**	-0.453	0.734	100
10	6/0.0	P18M48-136.8	B73	10.965**	0.351	-0.673	100
11	6/1.7	P18M48-142.1	Mo17	11.420**	-0.125	-0.495	80
12	6/11.4	<sup>m</sup> P12M49-148.9	B73	11.093**	0.453	0.734	100
13	8/38.8	E39M60-394.1	Mo17	12.304**	-0.527	0.926	60
14	8/55.7	<sup>m</sup> P12M61-507.8	Mo17	10.965**	-0.351	-0.673	100
15	8/57.1	E33M47-92.5	Mo17	10.965**	-0.351	-0.673	100
16	8/67.4	P18M48-196.9	B73	12.304**	0.527	0.926	60
17	8/72.2	P13M49-226.9	Mo17	10.965**	-0.351	-0.673	100
18	8/74.6	P12M47-181.6	Mo17	10.965**	-0.351	-0.673	100
19	8/74.6	P12M47-178.4	B73	10.965**	0.351	-0.673	100
20	9/20.9	E38M51-153.0	Mo17	11.093**	-0.453	0.734	100
21	9/58.6	E33M50-69.4	B73	10.965**	0.351	-0.673	100
22	9/58.6	E35M50-228.1	Mo17	12.304**	-0.527	0.926	60
23	10/25.4	E35M58-103.0	Mo17	10.965**	-0.351	-0.673	100
24	10/37.7	P12M61-185.6	Mo17	11.420**	-0.125	-0.495	80
25	10/54.0	P13M60-96.1	Mo17	11.420**	-0.125	-0.495	80

\*\* $P < 0.005$ .

<sup>a</sup> B73 and Mo17 will be referred to as 'origin' of a selected marker allele, when the AFLP marker has been identified as a B73 or Mo17 marker, respectively, mapped on the B73 × Mo17 RI linkage map (Vuylsteke *et al.*, 1999).

All of the selected markers but three show a high *ald* value (either 80 or 100). These high *ald* values are in good agreement with the hypothesis that heterotic groups should differ for their allelic frequencies at the QTL that exhibit dominance effects, or at the marker loci tightly linked to these QTL (Charcosset & Essioux, 1994).

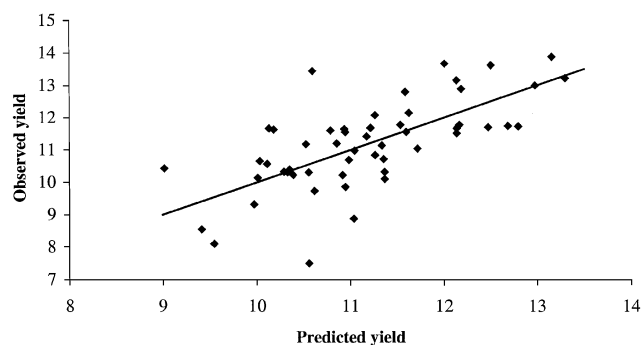
#### Evaluation of the additive model for prediction of HP

Performing the cross-validation of the additive models for prediction of HP by a jack-knife sampling procedure, shows that the highest efficiency for prediction of HP of GY was reached when considering the selected markers at a significance level of  $P < 0.001$  only (Table 4; Fig. 4). In this situation, HP predicted by the model explained 45.1% of the variation observed. The 13 hybrids with a predicted HP value  $\geq 12.0$  t ha<sup>-1</sup> at 15.5% moisture, had a observed value  $\geq 11.5$  t ha<sup>-1</sup> at 15.5% moisture. The corresponding mean SE of predicted vs. observed values was 0.88 t ha<sup>-1</sup> at 15.5% moisture. The ratio between SE and the total range of variation that was observed (7.5–13.88 t ha<sup>-1</sup> at 15.5% moisture), and the fact that most of the best single crosses are identified, suggests that prediction based on the TCSM model is highly efficient for a preliminary screening of test hybrids before field evaluation. Note that cross-validation of the prediction model by a jack-knife procedure involved the 53 interheterotic hybrids of the analysed 13 by 13 half-diallel.

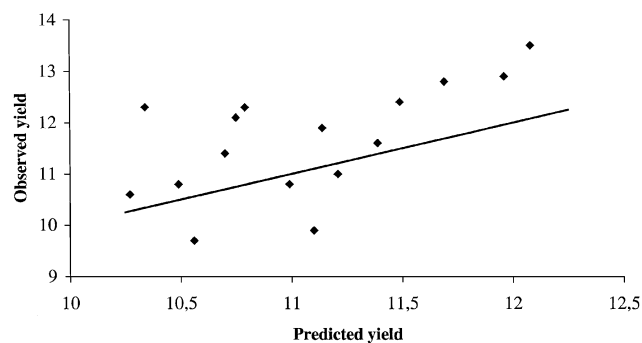
Cross-validating the additive model for prediction of HP by linear regression of the HP of the 16 parental testcrosses on their corresponding TCSM, the highest efficiency ( $r^2 = 33.0\%$ ; Table 4; Fig. 5) was reached by simultaneous fit of the 20 selected marker alleles, given in Table 2. Note that here, in contrast with the cross-

**Table 4** Coefficient of determination ( $r^2$ ) between observed and predicted hybrid performance (HP) of (a) the 53 hybrids forming part of the 13 by 13 half-diallel and (b) 16 hybrids partly related to the 13 by 13 half-diallel; the corresponding mean standard error (SE) and the empirical standard deviation of SE (within brackets) over the 53 and 16 cross-validations

	HP			
	53 hybrids		16 hybrids	
	$r^2$ (%)	SE	$r^2$ (%)	SE
$P < 0.001$	45.1	0.88 (0.03)	33.0	0.842 (0.002)
$P < 0.0005$	42.1	0.90 (0.03)	15.7	0.868 (0.008)
$P < 0.0001$	35.3	0.92 (0.02)	11.2	0.883 (0.009)



**Fig. 4** Observed vs. predicted performance (t ha<sup>-1</sup> at 15.5% moisture) of maize hybrids based on 52 predictor hybrids used for parameter estimation (jack-knife sampling procedure), considering the selected markers at a significance level of  $P < 0.001$  (Table 2). The 53 interheterotic hybrids are considered. The straight line represents the predicted = observed equation.



**Fig. 5** Observed vs. predicted performance (t ha<sup>-1</sup> at 15.5% moisture) of the 16 hybrids partly related to the 13 by 13 half-diallel, based on the selected markers at a significance level of  $P < 0.001$  (Table 2). The straight line represents the predicted = observed equation.

validation by jack-knifing, the prediction model is evaluated using hybrids of which only one parent (the LSC tester) is forming part of the 13 by 13 half-diallel. Despite a moderate  $r^2$  value, Fig. 5 shows that the best single crosses are identified. This suggests that prediction, based on the  $a_i$  and  $d_i$  estimates of the 20 markers selected at  $P < 0.001$  (Table 2), is efficient as a preliminary screening of related test hybrids before field evaluation.

## Discussion

In good accordance with published results (Lee *et al.*, 1989; Melchinger *et al.*, 1990a,b, 1992; Smith *et al.*, 1990; Dudley *et al.*, 1991; Boppenmaier *et al.*, 1993; Burstin *et al.*, 1995; Ajmone Marsan *et al.*, 1998), estimates of the GD between parents did not consistently



identify the best crosses, particularly not when the two parents are nonrelated lines. On the contrary, correlations between SGD and SCA found in the subset of unrelated lines were highly significant ( $P < 0.001$ ) and of a high magnitude suggesting practical utility in predicting SCA effects. Differences between our results and those reported by Ajmone Marsan *et al.* (1998) are caused solely by differences in field data.

The segregational resolution of conventional segregating populations (e.g.  $F_2$ , BC) is too low to distinguish tightly from less tightly QTL-marker combinations. The approach to QTL identification followed in the present study has the following potential advantages over QTL detection in a segregating population. First, marker-trait associations are only expected to be found in case a marker is tightly linked to a QTL. This is because across a set of lines, associations between QTL and loosely linked markers will be nonexistent because of accumulated recombination events during the establishment of the lines. Basically, the type of associations we have identified are caused by identity by descent of QTL and marker alleles across lines. Secondly, only a limited number of lines representing the gene pool used by the breeder need to be genotyped. And thirdly, it may allow the detection of QTL that vary across a wide spectrum of the germplasm used. The possible advantages are not easily generalized, especially because the joint identity by descent of alleles of linked loci depends on factors that are largely unknown for most germplasm collections, i.e. the number of generations since the descent from a common ancestor and the amount of exchange between lines of descents by crossing in the past.

The  $a$ - and  $d$ -values of the markers selected for HP for GY indicate that QTL with additive to partial dominance effects are prevalent. Although the magnitude of the genetic effects for any single QTL contributing to GY can vary considerably, the joint added contribution of single QTL involved in GY explains 59.3–62.4% of the HP variance. Cross-validation of the prediction efficiency of the TCSM model for HP showed that the best crosses were identified, suggesting that the TCSM-approach is efficient as a preliminary screening of test hybrids before field evaluation. The higher prediction efficiency of the TCSM model in comparison with the ‘distance’ model can be explained as follows: rather than converting molecular polymorphism between inbred lines, having direct or no direct effect on the trait of interest, into the metric GD, only specific markers were selected that were supposed to be linked to loci that affect the quantitative trait of interest. Hybrids heterozygous for marker loci significantly associated with SCA, often show a higher GY than hybrids homozygous for those marker loci. This pattern may result from either true overdominance (i.e. particular single loci at

which the heterozygote phenotype exceeds that of either homozygote), epistasis or pseudo-overdominance (i.e. closely linked loci at which alleles with dominant or partially dominant advantageous effects are in repulsion phase). With more than one QTL linked to the marker, epistatic effects modify the additive and dominance effects or pseudo-overdominance results. Although all QTL were detected at marker loci and deliver (in this way) the maximal genetic information, and an extensive dissociation of alleles at linked loci is most likely represented in the inbred lines, our results still cannot distinguish these possibilities.

If the joint effect of multiple QTL involved in the heterotic response of GY is additive, 36.8% of the SCA variance among the 53 hybrids can be explained, which is of a higher extent than explained by SCSM (22.1–24.0%), and almost equal to what is explained by SGD (40.1%). The high  $ald$  value of the marker alleles showing significant association with the SCA effects is consistent with the fact that the process of inbreeding and selection by which lines are commonly developed (i) generates allelic divergence among groups at the marker loci and the QTL involved in the trait of interest, and (ii) produces linkage disequilibrium between marker loci and QTL involved in SCA (Charcosset & Essioux, 1994).

The aim of the breeder is to accumulate in the same genotype the maximum number of favourable genes. Because all putative QTL mentioned in Table 2 show additive to partial dominance gene effects, fixing one or more of the favourable QTL alleles in the inbred line is desirable. More than GCSM, the TCSM value *per se* of an inbred line is suited to monitoring the improvement of an inbred line by fixation of favourable alleles and, subsequently, marker-assisted selection, as the TCSM *per se* value can be calculated directly from the genotype of the inbred line.

Although direct comparisons of QTL are complicated by differences in parental lines, design of the cross, number of progeny and the environments in which the progeny was assessed, as well as by different marker loci and QTL detection methods, other reports have identified some of the same regions detected in the present study to be associated with GY. Austin & Lee (1996, 1998) detected GY QTL on  $5S/umc72$ ,  $6L/bnl5.47-npi280$  and  $8L/umc7$  that were also associated with GY in the present study ( $5S/20.1$ ;  $6L/64.7-68.4$ ;  $8L/124.6-124.7$ ). Also Zehr *et al.* (1992) reported the GY QTL on  $6L$ , showing marker association with  $umc38a$ . Another GY QTL reported by Zehr *et al.* (1992) was associated with  $umc44$  on  $2S$ , likely to coincide with  $2S/53.4$  in our study. Ajmone Marsan *et al.* (1995) reported a major GY QTL associated with  $umc051$  on chromosome 6, which is in the vicinity of

the putative QTL on 6C/10.3-10.8 found in our study. Another GY QTL found on 9C/54.1-58.6, is likely to coincide with the GY QTL found on 9C by Stuber *et al.* (1992); Zehr *et al.* (1992) and Ajmone Marsan *et al.* (1995). Finally, Ajmone Marsan *et al.* (1995) also detected a GY QTL in the interval 4L/umc42-umc19, associated with GY in the present study (4C/56.2-58.0). Besides the agreement in chromosomal location for a few GY QTL, agreement in origin of the superior and inferior allele is present only for the QTL on 4C, 9C, and on 2S, respectively. Of the chromosomal regions selected as being significantly associated with QTL contributing to the SCA for GY, one (9/58.6:E35/M50-228.1) was also selected as being significantly ( $P < 0.001$ ) associated with a QTL contributing to the HP for GY in this study, and four (1/53.1, 1/69.0, 2/70.4 and 10/37.7) were associated with heterosis for GY in Stuber *et al.* (1992).

The efficiency of the TCSM method as a prediction method and/or as a QTL screening method may be increased by fulfilling the following requirements. (i) A higher marker density on the map will allow the detection of more marker alleles tightly linked to specific QTL, consolidating already identified QTL or identifying new putative QTL. Where a higher marker density can be obtained by intensifying the mapping efforts, a higher number of specific alleles per locus can be obtained by integrating linkage maps covering different genomes. (ii) A more reliable and easier evaluation of the effect of a QTL allele will be obtained when the three genotypic classes are represented more equally. This balance can be obtained by enlarging the half-diallel. And (iii) yield data of hybrids, available from multiple trials carried out across different locations and years, are highly desirable in order to reduce the phenotypic variance, representing a gain in accuracy.

## Acknowledgements

We would like to thank Dr F. Salamini and Dr M. Zabeau for the initial idea for this work, and Dr M. Motto for providing us seeds of the inbred lines and the grain yield data of the hybrids. This work was financed by Keygene N.V. and five Dutch breeding companies: Cebeco Zaden, De Ruiter, ENZA Zaden, Rijk Zwaan and Van der Have.

## References

AJMONE MARSAN, P., CASTIGLIONI, P., FUSARI, F., KUIPER, M. AND MOTTO, M. 1998. Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. *Theor. Appl. Genet.*, **96**, 219-227.

- AJMONE MARSAN, P., MONFREDINI, G., LUDWIG, R. F., MELCHINGER, A. E., FRANCESCHINI, P., PAGNOTTO, G. *ET AL.* 1995. In an elite cross of maize a major quantitative trait locus controls one-fourth of the genetic variation for grain yield. *Theor. Appl. Genet.*, **90**, 415-424.
- AUSTIN, D. F. AND LEE, M. 1996. Comparative mapping in F<sub>2:3</sub> and F<sub>6:7</sub> generations of quantitative trait loci for grain yield and yield components in maize. *Theor. Appl. Genet.*, **2**, 817-826.
- AUSTIN, D. F. AND LEE, M. 1998. Detection of quantitative trait loci for grain yield and yield components in maize across generations in stress and nonstress environments. *Crop Sci.*, **38**, 1296-1308.
- BEAVIS, W. D., SMITH, O. S., GRANT, D. AND FINCHER, R. 1994. Identification of quantitative trait loci using a small sample of topcrossed and F<sub>4</sub> progeny from maize. *Crop Sci.*, **34**, 882-896.
- BERNARDO, R. 1994. Prediction of maize single-cross performance using RFLPs and information from related hybrids. *Crop Sci.*, **34**, 20-25.
- BOPPENMAIER, J., MELCHINGER, A. E., SEITZ, G., GEIGER, H. H. AND HERRMANN, R. G. 1993. Genetic Diversity for RFLPs in European Inbreds. III. Performance of crosses within versus between heterotic groups for grain traits. *Pl. Breed.*, **111**, 217-226.
- BURSTIN, J., CHARCOSSET, A., BARRIÈRE, Y., HÉBERT, Y., DE VIENNE, D. AND DAMERVAL, C. 1995. Molecular markers and protein quantities as genetic descriptors in maize. II. Prediction of performance of hybrids for forage traits. *Pl. Breed.*, **114**, 427-433.
- CHARCOSSET, A., BONNISEAU, B., TOUCHÉBEUF, O., BURSTIN, J., DUBREUIL, P., BARRIÈRE, Y. *ET AL.* 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. AND ESSIUX, L. 1994. The effect of population structure on the relationship between heterosis and heterozygosity at marker loci. *Theor. Appl. Genet.*, **81**, 571-575.
- COCKERHAM, C. C. AND ZENG, Z.-B. 1996. Design III with marker loci. *Genetics*, **143**, 1437-1456.
- DUDLEY, J. W., SAGHAI MAROOF, M. A. AND RUFENER, G. K. 1991. Molecular markers and grouping of parents in maize breeding programs. *Crop Sci.*, **31**, 718-723.
- EATHINGTON, S. R., DUDLEY, J. W. AND RUFENER, G. K. 1997. Marker effects estimated from testcrosses of early and late generations of inbreeding in maize. *Crop Sci.*, **37**, 1679-1685.
- EDWARDS, M. D., STUBER, C. W. AND WENDEL, J. F. 1987. Molecular-marker facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. *Genetics*, **116**, 113-125.
- FALCONER, D. S. 1989. *Introduction to Quantitative Genetics*, 3rd edn. John Wiley, New York.
- GENSTAT-5-COMMITTEE. 1993. *Genstat 5: Reference Manual*. Oxford University Press, New York.
- GRIFFING, B. 1956. Concept of general and specific combining ability in relation to the diallel crossing system. *Aust. J. Biol. Sci.*, **9**, 463-493.

- HAYMAN, B. I. 1954. The theory and analysis of diallel crosses. *Genetics*, **39**, 789–809.
- JACCARD, P. 1908. Nouvelles recherches sur la distribution florale. *Bull. Soc. Vaudoise Sci. Nat.*, **44**, 223–270.
- LEE, M., GODSHALK, E. B., LAMKEY, K. R. AND WOODMAN, W. L. 1989. Association of restriction length polymorphism among maize inbreds with agronomic performance of their crosses. *Crop Sci.*, **29**, 1067–1071.
- MELCHINGER, A. E., BOPPENMAIER, J., DHILLON, B. S., POLLMER, W. G. AND HERRMANN, R. G. 1992. Genetic diversity for RFLPs in European maize inbreds: II. Relation to performance of hybrids within versus between heterotic groups for forage traits. *Theor. Appl. Genet.*, **84**, 672–681.
- MELCHINGER, A. E., LEE, M., LAMKEY, K. R., HALLAUER, A. R. AND WOODMAN, W. L. 1990a. Genetic diversity for restriction length polymorphisms and heterosis for two diallel sets of maize inbreds. *Theor. Appl. Genet.*, **80**, 488–496.
- MELCHINGER, A. E., LEE, M., LAMKEY, K. R. AND WOODMAN, W. L. 1990b. Genetic diversity for restriction length polymorphisms: relation to estimated genetic effects in maize inbreds. *Crop Sci.*, **31**, 1033–1040.
- SMITH, O. S., SMITH, J. S. C., BOWEN, S. L., TENBORG, R. A. AND WALL, S. J. 1990. Similarities among a group of elite maize inbreds as measured by pedigree, F<sub>1</sub> grain yield, grain yield, heterosis and RFLPs. *Theor. Appl. Genet.*, **80**, 833–840.
- STUBER, C. W., LINCOLN, S. E., WOLFF, D. W., HELENTJARIS, T. AND LANDER, E. S. 1992. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics*, **132**, 823–839.
- VAN OOIJEN, J. W. AND MALIEPAARD, C. 1996. *MapQTL*<sup>TM</sup>. Software for the calculation of QTL positions on genetic maps, Version 3.0. CPRO-DLO, Wageningen.
- VELDBOOM, L. R., LEE, M. AND WOODMAN, W. L. 1994. Molecular marker-facilitated studies in an elite maize population: I. Linkage analysis and determination of QTL for morphological traits. *Theor. Appl. Genet.*, **88**, 7–16.
- VUYLSTEKE, M., MANK, R., ANTONISE, R., BASTIAANS, E., SENIOR, M. L., STUBER, C. W. ET AL. 1999. Two high-density AFLP® linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. *Theor. Appl. Genet.*, **99**, 921–935.
- VUYLSTEKE, M., MANK, R., BRUGMANS, B., STAM, P. AND KUIPER, M. 2000. Further characterization of AFLP® data as a tool in genetic diversity assessments among maize (*Zea mays* L.) inbred lines. *Mol. Breeding*, **6**, 265–276.
- ZEHR, B. E., DUDLEY, J. W., CHOJECKI, J., SAGHAI-MAROOF, M. A. AND MOWERS, R. P. 1992. Use of RFLP markers to search for alleles in a maize population for improvement of an elite hybrid. *Theor. Appl. Genet.*, **83**, 903–911.