

# Quantitative Trait Locus (QTL) Mapping Using Different Testers and Independent Population Samples in Maize Reveals Low Power of QTL Detection and Large Bias in Estimates of QTL Effects

Albrecht E. Melchinger,\* H. Friedrich Utz\* and Chris C. Schön†

\**Institute of Plant Breeding, Seed Science and Population Genetics and* †*State Plant Breeding Institute, University of Hohenheim, 70593 Stuttgart, Germany*

Manuscript received August 7, 1997  
Accepted for publication February 2, 1998

## ABSTRACT

The efficiency of marker-assisted selection (MAS) depends on the power of quantitative trait locus (QTL) detection and unbiased estimation of QTL effects. Two independent samples ( $N = 344$  and  $107$ ) of  $F_2$  plants were genotyped for 89 RFLP markers. For each sample, testcross (TC) progenies of the corresponding  $F_3$  lines with two testers were evaluated in four environments. QTL for grain yield and other agronomically important traits were mapped in both samples. QTL effects were estimated from the same data as used for detection and mapping of QTL (calibration) and, based on QTL positions from calibration, from the second, independent sample (validation). For all traits and both testers we detected a total of 107 QTL with  $N = 344$ , and 39 QTL with  $N = 107$ , of which only 20 were in common. Consistency of QTL effects across testers was in agreement with corresponding genotypic correlations between the two TC series. Most QTL displayed no significant QTL  $\times$  environment nor epistatic interactions. Estimates of the proportion of the phenotypic and genetic variance explained by QTL were considerably reduced when derived from the independent validation sample as opposed to estimates from the calibration sample. We conclude that, unless QTL effects are estimated from an independent sample, they can be inflated, resulting in an overly optimistic assessment of the efficiency of MAS.

**M**OLECULAR marker technologies allow plant geneticists to construct high density genetic maps for any species of interest and use them for detecting, mapping, and estimating the effects of quantitative trait loci (QTL). While the basic idea of this approach was published more than 70 years ago (Sax 1923), new interest was generated when studies with maize and tomatoes successfully demonstrated that some markers explained a substantial proportion of the phenotypic variance of complex characters (for review, see Tanksley 1993). As a consequence, vigorous research on QTL mapping for quantitative traits such as yield, quality, maturity, and resistance to biotic and abiotic stress was initiated in many crop species (for review, see Lee 1995). Based on first results, it was anticipated that identification of important QTL regions could enhance plant breeding efficiency by marker-assisted selection (MAS). However, the prospects of this approach depend strongly upon the expenditures required for QTL mapping experiments, because their high costs reduce or even nullify the advantages of MAS schemes in a comprehensive economic assessment.

An important consideration in this context relates to the sample size ( $N$ ) needed for QTL mapping. Most

published experiments with replicated trials have employed between 100 and 200 progenies (for review, see Melchinger 1997), this choice being mainly dictated by the excessive labor and costs required for phenotyping and genotyping large populations. According to theoretical investigations (Lande and Thompson 1990), the proportion,  $p$ , of the additive genetic variance explained by the detected QTL is inversely related to the product,  $h^2N$ , where  $h^2$  is the heritability of the trait. Consequently, for traits with moderate or low  $h^2$ , where MAS should be most efficient, the chances of QTL detection with the above sample sizes are fairly low unless the QTL explains a substantial proportion of the genetic variance. A comparison of QTL detected in large versus small samples from the same population should give some insight into the power of QTL detection. So far the only experimental study that has been published on such a comparison is by Beavis (1994) on the highly heritable trait plant height using a limited data set of 20 markers only.

In view of the high costs of QTL studies, it has been common practice to estimate QTL effects from the same data as used for QTL mapping. With this approach, however, QTL effects generally are overestimated (Lande and Thompson 1990). As demonstrated by computer simulations (Beavis 1994; Utz and Melchinger 1994), the upward bias can be severe for low  $N$  and  $h^2$ . To resolve this problem, Lande and Thompson (1990) suggested

*Corresponding author:* Albrecht E. Melchinger, Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany. E-mail: melchinger@uni-hohenheim.de

obtaining unbiased estimates of QTL effects by mapping QTL with one data set and based on this information estimating QTL effects in an independent data set. No experimental data have been presented so far on this approach even though knowledge about the magnitude of the bias of estimated QTL effects may strongly affect the conclusions concerning the prospects of MAS.

An indication that the bias of estimated QTL effects can be fairly large stems from the comparison of different QTL mapping studies. Beavis (1994) found little congruency of QTL locations and estimated effects for QTL on plant height and grain yield in different samples of progeny from the same cross (B73  $\times$  Mo17). However, as he pointed out, the comparison of results was confounded by a number of factors. Different sets of genetic markers were used in the studies and seed sources of parental lines were not the same. The progeny were evaluated in different environments and the level of inbreeding varied. Another confounding aspect was the evaluation of lines *per se* as opposed to testcross (TC) performance. Most QTL mapping studies have concentrated on line *per se* performance, even though in hybrid breeding it is essential to test new lines for their TC performance in combination with unrelated testers. Because testers and potential hybrid partners of new lines are often not fixed or may change over time, an important question concerns the consistency of QTL for TC performance with different testers.

In this study, we evaluated TC progenies of 344  $F_3$  lines in combination with two unrelated testers plus additional TC progenies from an independent but smaller sample ( $N = 107$ ) of  $F_3$  lines from the same cross in combination with the same two testers for grain yield and four other important agronomic traits. Objectives of our research were to (i) assess the magnitude of the bias of estimated QTL effects by mapping QTL with one data set (calibration) and, based on this information, estimate QTL effects in an independent data set (validation), (ii) compare the power of QTL detection in samples of different size, (iii) investigate the consistency of QTL across testers, and (iv) assess the importance of epistatic and QTL-by-environment interactions.

## MATERIALS AND METHODS

**Plant materials:** The plant materials used for this study were partly identical to those employed and described in previous studies on kernel weight, protein concentration, plant height (Schön *et al.* 1994) and forage traits in maize (Lübberstedt *et al.* 1997). Briefly, two early maturing elite European flint inbreds, KW1265 and D146 (subsequently referred to as P1 and P2), were used as parents. Randomly chosen  $F_2$  plants from the cross P1  $\times$  P2 were selfed to produce 507 independently derived  $F_3$  lines. Subsequently, TC seed was produced in two separate isolation plots by mating each of two inbred testers as pollinators (KW4115 and KW5361, subsequently referred to as T1 and T2) to a random sample of 40  $F_3$  plants from each of the 507  $F_3$  lines as well as parents P1 and P2. Both

testers were elite inbreds from two diverse European dent heterotic pools and unrelated by pedigree.

**Field experiments:** The TC progenies of  $F_3$  lines and parents P1 and P2 were evaluated in two series of experiments. Experiment 1 comprised two adjacent subexperiments each with 400 entries (Subexperiment 1T1 = TC with tester T1, Subexperiment 1T2 = TC with tester T2) conducted in 1990 and 1991 at two sites in Germany (Gondelsheim and Grucking) with diverse agroecological conditions and representing two main maize growing areas in Germany, the Upper Rhine valley and Lower Bavaria. Data on plant height were additionally available from forage trials conducted at five environments in Germany described in detail by Lübberstedt *et al.* (1997).

Experiment 2 also comprised two subexperiments, each with 150 entries (Subexperiment 2T1 = TC with tester T1, Subexperiment 2T2 = TC with tester T2) conducted in four environments. Two of the trials were grown adjacent to each other in the same environments (Eckartsweier 1993, Bad Krozingen 1993) and two environments were only used for one subexperiment (Subexperiment 2T1: Hochburg 1993, Zell 1993; Subexperiment 2T2: Eckartsweier 1992, Bad Krozingen 1992).

The 400 entries in Subexperiments 1T1 and 1T2 comprised 380 TC of  $F_3$  lines, TC of P1 and P2 included as quintuple entries, and 10 common check hybrids. The 150 entries in Subexperiments 2T1 and 2T2 comprised TC from a different set of 127  $F_3$  lines, TC of P1 and P2 included as six and seven entries, respectively, and the same set of 10 check hybrids as in Experiment 1. The experimental design was a 40-by-10 alpha design (Patterson and Williams 1976) for Experiment 1 and a 15-by-10 alpha design for Experiment 2 with two replications each. Two-row plots were overplanted and later thinned to 45 plants per row in 1990, 50 plants per row in 1991, and 52 plants per row in 1992 and 1993 to reach a final stand of 10, 11, and 8.7 plants  $m^{-2}$ , respectively. All experiments were machine planted and harvested as grain trials with a combine.

Data were collected for the following traits: grain yield (GY) in  $Mg\ ha^{-1}$ , adjusted to 155  $g\ kg^{-1}$  grain moisture, grain moisture (GM) in  $g\ kg^{-1}$  at harvest, kernel weight (KW) in  $mg\ kernel^{-1}$  determined from four samples of 50 kernels from each plot, protein concentration (PC) in grain ( $g\ kg^{-1}$ ) measured by near infrared reflectance spectroscopy as described by Melchinger *et al.* (1986), and plant height (PH) measured in cm on a plot basis as the distance from the soil level to the lowest tassel branch. In Experiment 1, PC could not be determined in 1991 at Grucking because of technical problems.

**RFLP marker genotyping and linkage map construction:** The procedures for RFLP assays, segregation analysis of individual markers, and construction of an RFLP linkage map for cross P1  $\times$  P2 were described in detail by Schön *et al.* (1994). A subset of 344 parental  $F_2$  plants of the 380  $F_3$  lines employed in Experiment 1 and a second subset of 107 parental  $F_2$  plants of the 127  $F_3$  lines employed in Experiment 2, all chosen for good DNA yield and showing no evidence of contamination, were genotyped for a total of 89 RFLP marker loci, 82 of them showing a codominant and seven a dominant inheritance pattern. Observed genotype frequencies at each marker locus were checked for deviations from Mendelian segregation ratios and allele frequency 0.5 by ordinary  $\chi^2$  tests. Owing to multiple tests, appropriate type I error rates were determined by the sequentially rejective Bonferroni procedure (Holm 1979). A linkage map was constructed for the combined set of 451  $F_2$  plants using MAPMAKER version 3.0b (Lander *et al.* 1987) and a LOD threshold of 3.0 in two-point analyses. Recombination frequencies between marker loci were estimated by multi-point analyses and transformed into map distances (cM) by Haldane's mapping function.

**Data analyses:** Each site-year combination was treated as

an environment in the statistical analyses. First, analyses of variance were performed on the data from each subexperiment and environment. Adjusted entry means and effective error mean squares were then used to compute the combined analyses of variance and covariance across environments for each subexperiment. The sums of squares for entries (399 d.f. each in Subexperiments 1T1 and 1T2 and 149 d.f. each in Subexperiments 2T1 and 2T2) were subdivided into the variation among TC of  $F_3$  lines (379 d.f. in Subexperiments 1T1 and 1T2 and 126 d.f. in Subexperiments 2T1 and 2T2) and orthogonal contrasts among the TC means of P1, P2,  $F_3$  lines and the hybrid checks. A corresponding subdivision was conducted on the entry-by-environment interaction sums of squares.

Components of variance for the TC of  $F_3$  lines in each subexperiment were computed considering all effects (environments,  $F_3$  lines) in the statistical model as random. Estimates of variance components  $\sigma^2$  (error variance),  $\sigma_{ge}^2$  (genotype-by-environment ( $G \times E$ ) interaction variance), and  $\sigma_g^2$  (genotypic variance) of  $F_3$  TC progenies and their standard errors (SE) were calculated as described by Searle (1971, p. 475). Heritabilities ( $h^2$ ) on a TC progeny mean basis were estimated as described by Hallauer and Miranda (1981)

$$\hat{h}^2 = \frac{\hat{\sigma}_g^2}{\frac{\hat{\sigma}^2}{r} + \frac{\hat{\sigma}_{ge}^2}{e} + \hat{\sigma}_g^2},$$

where  $r$  = number of replications and  $e$  = number of environments. Exact 90% confidence intervals of  $\hat{h}^2$  were calculated according to Knapp *et al.* (1985).  $F$ -tests were employed for testing the homogeneity of  $\hat{\sigma}_g^2$  between (1) the two TC series in each experiment and (2) the two experiments for each tester according to the approximation given by Satterthwaite (1946). Phenotypic ( $r_p$ ) and genotypic ( $r_g$ ) correlations were calculated between TC of  $F_3$  lines with T1 and T2 for each trait in Experiments 1 and 2 by standard procedures (Mode and Robinson 1959).

All QTL analyses were performed using the linkage information given in Figure 1. While Schön *et al.* (1994) used interval mapping according to Lander and Botstein (1989) for QTL analyses, in this study, the method of composite interval mapping (CIM) (Jansen and Stam 1994; Zeng 1994) was employed for mapping of QTL and estimation of their effects in each of the four subexperiments. All necessary computations were performed with PLABQTL (Utz and Melchinger 1996), which employs interval mapping by the regression approach (Haley and Knott 1992) in combination with the use of selected markers as cofactors. The underlying model for TC progenies with a given tester can be written as

$$Y_{jz} = \mu_{P1} + \alpha_l x_{jl} + \sum_k b_k x_{jk} + \varepsilon_{jz}. \quad (1)$$

Here,  $Y_{jz}$  denotes the mean phenotypic trait value of the TC progeny of line  $j$  with tester  $z$  ( $z = 1, 2$ ) averaged across all environments;  $\mu_{P1}$  is the mean phenotypic trait value of TC progeny carrying the allele from P1 at the QTL,  $\alpha_l$  is the average effect of substituting allele  $q$  in P1 by allele  $Q$  in P2 at the putative QTL in the marker interval ( $l, l + 1$ ) under consideration;  $x_{jl}$  is the conditional expectation of the dummy variable  $\theta_l$  given the observed genotypes at the flanking marker loci, where  $\theta_l$  assumes values 0, 0.5 or 1, if the genotype of the  $F_2$  plant at the putative QTL is  $qq$ ,  $Qq$ , or  $QQ$ , respectively;  $b_k$  is the partial regression coefficient of phenotype  $Y_{jz}$  on the  $k$ th (selected) marker;  $x_{jk}$  is a dummy variable (cofactor) taking values 0, 0.5, or 1 depending on whether the marker genotype of the parental  $F_2$  individual  $j$  at marker locus  $k$  is homozygous P1, heterozygous, or homozygous P2, respectively;  $\varepsilon_{jz}$  is a residual variable for the TC progeny of the  $j$ th  $F_3$  line with tester  $z$ .

Cofactors were selected by stepwise regression according to

Miller (1990, p. 49) with an "F-to-enter" and an "F-to-delete" value of 3.5. Testing for presence of a putative QTL in an interval by a likelihood ratio (LR) test (yielding so-called LOD scores) was performed as described by Lübberstedt *et al.* (1997). We chose a LOD (=0.217 LR) threshold of 2.5 for declaring a putative QTL. Given that the LR test statistic follows in our analysis of TC progenies approximately a  $\chi^2$  distribution with 2 df (1 df for the  $\alpha$ -effect and 1 df for the position of the QTL; Zeng 1994), this approximates a comparisonwise type I error  $P_c < 0.0032$  or a genomewise type I error  $P_g = MP_c < 0.25$  ( $M = 78$  being the total number of intervals tested). Estimates of QTL positions were obtained at the point where the LOD score assumed its maximum in the region under consideration. Under CIM, computation of confidence intervals for the QTL position is still an unsolved problem (Visscher *et al.* 1996). Therefore, QTL detected with different testers or in different experiments were regarded as common if their estimated map position was within a 20-cM distance and the estimated  $\alpha$ -effects had identical sign. Presence of QTL-by-environment (QTL  $\times$  E) interactions and digenic epistatic interactions between the detected QTL were tested in combined analyses of variance across environments by  $F$ -tests described by Bohn *et al.* (1996). A detailed list of expected mean squares for the analysis of QTL experiments from multi-environments is given by Melchinger (1998).

The proportion of the phenotypic variance ( $\sigma_p^2$ ) explained by a single QTL was determined as the square of the partial correlation coefficient ( $\beta^2$ ). Estimates of the allele substitution ( $\alpha$ ) effect of each putative QTL, the total LOD score as well as the total proportion ( $R^2$ ) of  $\sigma_p^2$  explained were obtained by fitting a model including all QTL for the respective trait simultaneously. This model was also used to estimate  $p$ , the proportion of the genotypic variance ( $\sigma_g^2$ ) explained by all detected QTL, according to the procedures described by Bohn *et al.* (1996).

Two approaches were applied in calculating estimates of QTL effects: (1) Following common practice, QTL effects were estimated from exactly the same experiment as used for QTL detection; (2) QTL detection was performed in one experiment (subsequently denoted as calibration) and, based on this information, QTL effects were estimated from the data of the other experiment with the same tester (subsequently referred to as validation). In the latter case, the design matrix  $X$  in multiple regression was calculated on the basis of (a) the map position of the QTL detected in the calibration and (b) the marker genotype at the flanking markers of the  $F_2$  plants in the validation according to described procedures (Haley and Knott 1992; Utz and Melchinger 1996).

Finally, for each  $F_2$  genotype  $j$  the marker index score  $M_{jz}$  of its TC progeny with tester  $Tz$  was calculated from its marker genotype and the  $X$  matrix from the multiple regression in calibration as outlined by Lande and Thompson (1990). According to standard procedures (Mode and Robinson 1959), the  $M_{jz}$  values were subsequently used to estimate the genotypic correlation  $r_g(Y_{jz}, M_{jz})$  of the observed TC performance  $Y_{jz}$  with tester  $Tz'$  ( $z' = 1, 2; z' \neq z$ ) and the marker index score based on results with tester  $Tz$ . Standard errors of  $\hat{r}_g(Y_{jz}, M_{jz})$  were determined according to Kendall and Stuart (1961, Equation 27.90), under the assumption that heritabilities of  $Y_{jz}$  are known.

## RESULTS

**Segregation and linkage of RFLP markers:** In the data analysis of the combined set of 451  $F_2$  plants (344 from Experiment 1 and 107 from Experiment 2), observed genotype frequencies were consistent with the expected

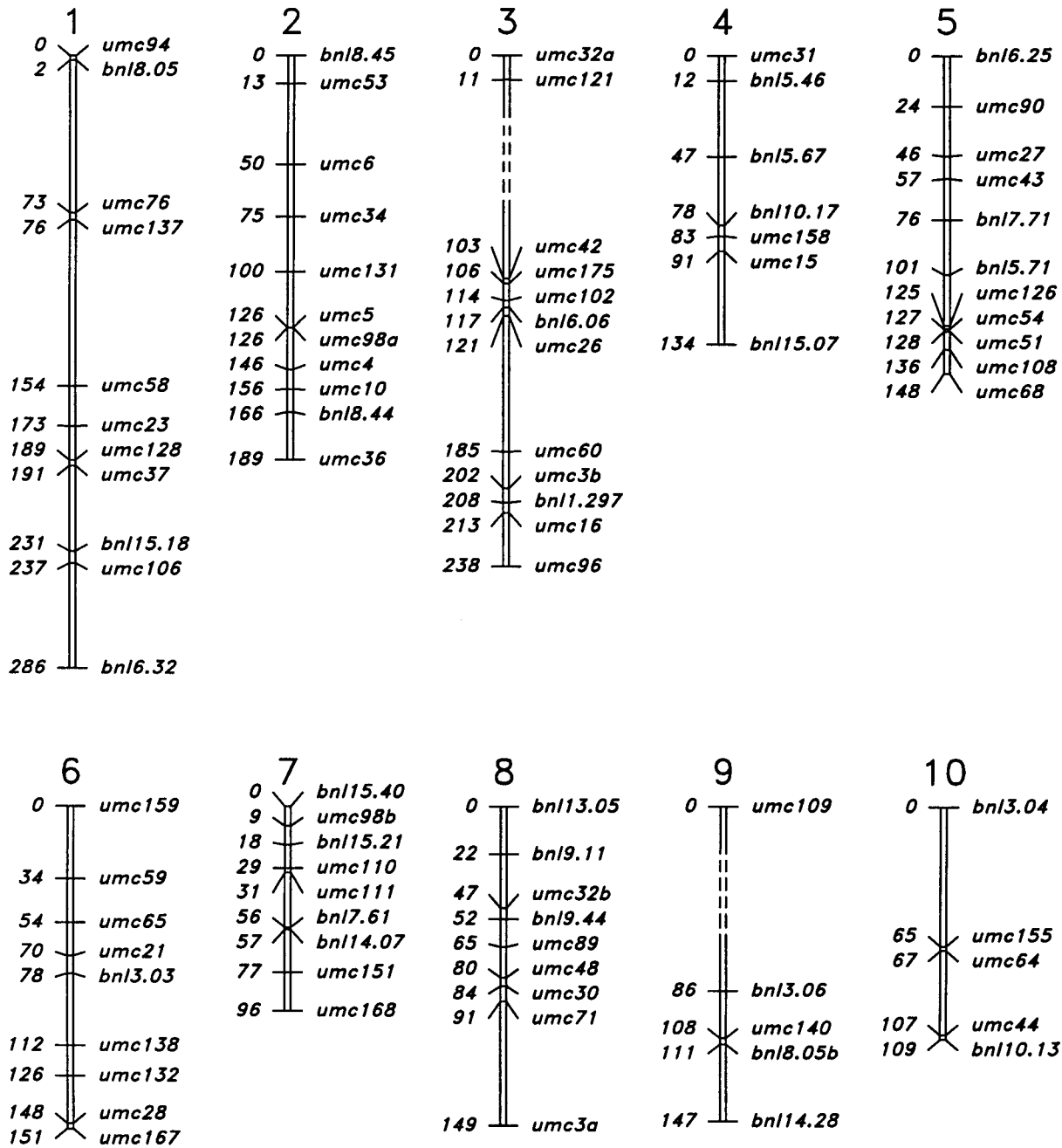


Figure 1.—The RFLP map with 89 markers constructed from 451  $F_2$  plants of maize cross  $P1 \times P2$ .

Mendelian segregation ratios for all 89 RFLP markers assayed (data not shown). The 89 marker loci spanned a map distance of 1647 cM with an average interval length of 24 cM (Figure 1). About 90% of the genome was located within a 20-cM distance to the nearest marker.

**Trait means, variances, heritabilities, and correlations:** Climatic conditions were favorable for maize grain production in all 10 test environments. Means and phenotypic variances of the 10 check varieties included in each of the four subexperiments varied considerably between environments for all traits exhibiting rather diverse growing conditions. Average yield of the 10 checks ranged

from 7.5 to 12.8  $Mg\ ha^{-1}$ . Phenotypic correlations based on performance of the 10 check varieties and averaged over traits and subexperiments were medium when calculated separately for the four environments of Experiment 1 ( $r_p = 0.65$ ) and also for the six environments of Experiment 2 ( $r_p = 0.79$ ). Using performance of check varieties in environments of Experiment 1 and correlating it with their performance in environments of Experiment 2 resulted in slightly lower phenotypic correlations ( $r_p = 0.53$ ).

In Experiment 1, TC means of  $F_3$  progenies with tester T1 were significantly ( $P < 0.01$ ) smaller than with tester

T2 for GY and KW but greater for GM (Table 1). For Experiment 2, the respective comparison is not meaningful because the two TC series were not evaluated in the same environments. The TC means of P1 and P2 differed significantly ( $P < 0.01$ ) for all traits with both testers in both experiments. Parent P1 generally had higher TC means than P2 except for GY and GM (tester T2 in Experiment 2) (Tables 1 and 2). The orthogonal contrast between the average TC performance of the parent lines ( $\bar{P}$ ) and the TC mean of the  $F_3$  lines ( $\bar{F}_3$ ) was significant ( $P < 0.05$ ) only for KW in Experiment 1 and PH in both experiments. The range in TC performance of  $F_3$  lines considerably transgressed the TC means of the parents for all traits but KW.

Genotypic variances among TC of  $F_3$  lines ( $\hat{\sigma}_g^2$ ) were highly significant ( $P < 0.01$ ) for all traits with both testers in both experiments (Tables 1 and 2). Estimates of  $\sigma_g^2$  for TC with T1 and T2 were heterogeneous ( $P < 0.01$ ) only for GM in Experiment 1. Estimates of  $\sigma_{ge}^2$  were significantly greater than zero ( $P < 0.01$ ) except in Experiment 2 for GY and PH (both testers) and GM (tester T1). Estimates of  $\sigma_g^2$  and  $\sigma_{ge}^2$  for both testers were significantly ( $P < 0.01$ ) greater in Experiment 1 than Experiment 2 for GY and GM. Heritability was medium for GY ( $0.48 < \hat{h}^2 < 0.74$ ) but relatively high for the other traits ( $0.64 < \hat{h}^2 < 0.91$ ) with similar estimates for both testers and mostly overlapping confidence intervals for the two experiments (Tables 1 and 2).

Phenotypic correlations between TC of  $F_3$  lines with tester T1 and T2 were greater than 0.53 except for GY ( $r_p < 0.39$ ) yet highly significant ( $P < 0.01$ ) for all traits in both experiments (Tables 1 and 2). Genotypic correlations ( $r_g$ ) varied between 0.60 and 0.88 and were in good agreement between both experiments for all traits.

**Identification of QTL:** Results from QTL analyses are presented for means across environments. For Experiments 1 and 2 estimates of the QTL position in the genome, the level of significance, the size of the phenotypic variance explained, the substitution effects and the significance of QTL-by-environment interactions are shown in Tables 3 and 4, respectively. The number of selected cofactors was higher in Experiment 1 (14–28) than in Experiment 2 (6–14) and more significant cofactors were found for traits with higher heritability than *e.g.*, for GY. A complete list of the number of selected cofactors used for each trait, tester, and experiment can be obtained upon request from the corresponding author.

#### Comparison of QTL effects between experiments:

**Grain yield:** For GY, seven putative QTL were identified in Experiment 1 in TC with T1 (Table 3). A simultaneous fit accounted for  $R^2 = 30.8\%$  of  $\hat{\sigma}_p^2$  and  $\hat{p} = 51.5\%$  of  $\hat{\sigma}_g^2$  (Figure 2). Only two QTL were detected in TC with T2. Collectively, they accounted for  $R^2 = 14.8\%$  and  $\hat{p} = 30.6\%$ . In Experiment 2, one QTL on chromosome 3 explaining 4.1% of  $\hat{\sigma}_p^2$  and 4.0% of  $\hat{\sigma}_g^2$  was detected in TC with T1 (Table 4 and Figure 2). In contrast, four

QTL were found in TC with T2. They explained collectively  $R^2 = 32.1\%$  and  $\hat{p} = 40.6\%$ . In each experiment, none of the QTL were in common between testers or displayed significant ( $P < 0.05$ ) QTL  $\times$  E interactions.

There was one common QTL for GY between Experiment 1 and 2 (Table 5). For QTL positions identified in Experiment 1 (calibration),  $\hat{\alpha}$ -effects estimated from Experiment 2 (validation) were on average about half as large yet of the same sign as those obtained from calibration (Table 3). An exception was the QTL on chromosome 1 with similar  $\hat{\alpha}$ -effects of opposite sign in calibration and validation. Collectively, the QTL effects from validation accounted for  $R^2 = 12.2\%$  and  $\hat{p} = 7.6\%$  for T1 and  $R^2 = 3.8\%$  and  $\hat{p} = 5.1\%$  for T2. When calibration was performed with Experiment 2 and validation with Experiment 1, the estimates dropped to  $R^2 = 0.7\%$  and  $\hat{p} = 0.7\%$  for TC with T1 and  $R^2 = 11.6\%$  and  $\hat{p} = 22.6\%$  for TC with T2 (Table 4).

**Grain moisture:** In Experiment 1, 12 and 13 QTL influencing GM in TC with tester T1 and T2, respectively, were detected. A simultaneous fit yielded  $R^2 = 45.6\%$  and  $\hat{p} = 58.2\%$  for TC with T1 and  $R^2 = 55.1\%$  and  $\hat{p} = 61.3\%$  for TC with T2. About half of the detected QTL displayed significant ( $P < 0.05$ ) QTL  $\times$  E interactions. Seven QTL were in common for both testers with similar  $\hat{\alpha}$ -effects.

In Experiment 2, three QTL were found for GM in TC with T1 ( $R^2 = 17.4\%$  and  $\hat{p} = 17.6\%$ ) and nine in TC with T2 ( $R^2 = 57.9\%$  and  $\hat{p} = 63.4\%$ ). One of the QTL was in common between testers, and none displayed significant QTL  $\times$  E interactions.

Two and six QTL were in common between Experiment 1 and 2 for tester T1 and T2, respectively (Table 5). Estimates of  $\alpha$ -effects from validation in Experiment 2 were in four cases larger but otherwise much smaller than those from calibration in Experiment 1 (Table 3). If significant, both estimates of  $\alpha$  had identical sign except for one QTL on chromosome 9 for tester T1 and one on chromosome 2 for T2. Collectively, the QTL effects from validation in Experiment 2 accounted for  $R^2 = 24.3\%$  and  $\hat{p} = 15.5\%$  for TC with T1 and  $R^2 = 46.4\%$  and  $\hat{p} = 44.8\%$  for TC with T2. Estimates of  $\alpha$ -effects from calibration in Experiment 2 generally agreed well with those from validation in Experiment 1 (Table 4), where a simultaneous fit explained  $R^2 = 6.2\%$  and  $\hat{p} = 10.1\%$  for TC with T1 and  $R^2 = 31.3\%$  and  $\hat{p} = 34.7\%$  for TC with T2.

**Kernel weight:** In Experiment 1, 12 QTL in TC with T1 and 11 QTL in TC with T2 were found for KW. The 12 QTL accounted for  $R^2 = 63.7\%$  and  $\hat{p} = 71.2\%$  for TC with T1 and  $R^2 = 52.5\%$  and  $\hat{p} = 58.9\%$  for TC with T2. About one quarter of the QTL showed significant ( $P < 0.05$ ) QTL  $\times$  E interactions. Ten QTL were in common and had similar  $\hat{\alpha}$ -effects for both testers.

In Experiment 2, four QTL were detected for TC with T1 and five QTL for TC with T2. Collectively, these QTL explained  $R^2 = 41.5\%$  and  $\hat{p} = 47.5\%$  for TC with

**TABLE 1**  
**First and second degree statistics for maize TC progenies from parent lines (P1 and P2) and 380 F<sub>3</sub> lines (from cross P1 × P2) with inbred testers T1 and T2 in Experiment 1 for GY, GM, KW, PC (three environments), PC (four environments), PC (three environments), and PH (nine environments)**

Parameter	Entries (N)	Grain yield (GY) (Mg ha <sup>-1</sup> )		Grain moisture (GM) (g kg <sup>-1</sup> )		Kernel weight (KW) (mg)		Protein concentration (PC) (g kg <sup>-1</sup> )		Plant height (PH) (cm)	
		T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
<b>Testcross means<sup>a</sup></b>											
P1	5	9.91 ± 0.21 <sup>b</sup>	10.84 ± 0.24	328.1 ± 2.5	292.7 ± 2.6	337.9 ± 2.4	336.8 ± 2.6	108.3 ± 0.80	112.0 ± 0.93	252.4 ± 1.4	254.0 ± 1.3
P2	5	10.56 ± 0.21	11.33 ± 0.24	315.9 ± 2.5	299.8 ± 2.6	283.6 ± 2.4	297.5 ± 2.6	102.0 ± 0.80	105.5 ± 0.93	244.9 ± 1.4	243.0 ± 1.3
P	10	10.23 ± 0.15	11.09 ± 0.17	322.0 ± 1.7	296.2 ± 1.8	310.7 ± 1.7	317.2 ± 1.8	105.1 ± 0.57	108.8 ± 0.66	248.7 ± 1.0	248.5 ± 0.9
F <sub>3</sub>	380	10.28 ± 0.03	11.39 ± 0.03	319.6 ± 0.4	293.2 ± 0.4	306.8 ± 0.5	311.2 ± 0.5	104.9 ± 0.12	108.1 ± 0.13	251.1 ± 0.3	251.4 ± 0.3
Range of F <sub>3</sub> lines		8.83, 11.64	9.88, 13.08	297.8, 342.2	261.7, 315.1	282.2, 339.8	279.7, 335.7	95.7, 111.5	99.1, 115.0	235.2, 264.3	231.5, 271.3
<b>Variance components</b>											
σ <sub>P</sub> <sup>2</sup>		0.142 ± 0.020**	0.129 ± 0.020**	35.33 ± 3.56**	59.44 ± 5.30**	95.39 ± 8.03**	76.85 ± 6.62**	3.58 ± 0.37**	5.10 ± 0.50**	26.64 ± 2.18**	33.21 ± 2.64**
σ <sub>PE</sub> <sup>2</sup>		0.106 ± 0.025**	0.155 ± 0.028**	17.45 ± 2.61**	20.53 ± 2.57**	13.81 ± 3.10**	19.03 ± 2.77**	1.32 ± 0.26**	2.01 ± 0.28**	5.07 ± 1.03**	4.69 ± 0.97**
σ <sub>E</sub> <sup>2</sup>		0.751 ± 0.030	0.811 ± 0.032	70.22 ± 2.78	65.10 ± 2.56	92.18 ± 3.64	74.14 ± 2.93	6.06 ± 0.28	5.78 ± 0.26	50.75 ± 1.34	47.71 ± 1.26
<b>Heritability</b>											
h <sup>2</sup>		0.54	0.48	0.73	0.82	0.86	0.85	0.71	0.76	0.89	0.91
95% C.I. on h <sup>2</sup> <sup>c</sup>		0.46, 0.61	0.38, 0.56	0.68, 0.77	0.78, 0.84	0.84, 0.88	0.82, 0.87	0.66, 0.76	0.71, 0.80	0.87, 0.90	0.90, 0.92
<b>Correlation coefficients<sup>d</sup></b>											
r <sub>P</sub>		0.33**		0.62**		0.76**		0.64**		0.79**	
r <sub>E</sub>		0.63 ± 0.09 <sup>b</sup>		0.73 ± 0.04		0.88 ± 0.02		0.80 ± 0.04		0.87 ± 0.02	

\*\* Significant at P < 0.01.

<sup>a</sup> P = TC mean of P1 and P2; F<sub>3</sub> = TC mean of F<sub>3</sub> lines.

<sup>b</sup> Standard errors are attached.

<sup>c</sup> Confidence intervals of h<sup>2</sup> were calculated by the method of Knapp et al. (1985).

<sup>d</sup> Phenotype (r<sub>P</sub>) and genotype (r<sub>E</sub>) correlation coefficients between TC progenies with T1 and T2.

**TABLE 2**  
**First and second degree statistics for maize TC progenies from parent lines (P1 and P2) and 127 F<sub>3</sub> lines (from cross P1 × P2) with inbred testers T1 and T2 in Experiment 2 for five agronomic traits measured in four environments**

Parameter	Grain yield (GY) (Mg ha <sup>-1</sup> )		Grain moisture (GM) (g kg <sup>-1</sup> )		Kernel weight (KW) (mg)		Protein concentration (PC) (g kg <sup>-1</sup> )		Plant height (PH) (cm)		
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	
<b>Testcross means<sup>a</sup></b>											
P1	7	9.48 ± 0.11 <sup>b</sup>	8.80 ± 0.07	327.1 ± 1.0	322.2 ± 1.5	356.7 ± 2.2	344.9 ± 2.8	105.9 ± 1.11	111.6 ± 1.02	255.5 ± 1.0	244.9 ± 0.9
P2	6	10.45 ± 0.11	9.62 ± 0.08	365.3 ± 1.1	329.0 ± 1.6	304.2 ± 2.2	298.4 ± 2.8	97.5 ± 1.13	105.8 ± 1.04	246.6 ± 1.1	236.3 ± 0.9
$\bar{P}$	13	9.97 ± 0.08	9.21 ± 0.06	368.7 ± 0.8	325.6 ± 0.7	330.5 ± 1.6	321.7 ± 2.0	101.7 ± 0.80	108.7 ± 0.74	251.0 ± 0.8	240.6 ± 0.6
F <sub>3</sub>	127	10.10 ± 0.04	9.34 ± 0.03	366.7 ± 0.5	323.9 ± 0.5	330.1 ± 0.9	321.2 ± 0.9	100.9 ± 0.36	109.6 ± 0.23	252.7 ± 0.6	244.9 ± 0.5
Range of F <sub>3</sub> lines		8.78, 11.06	8.01, 10.20	354.6, 388.7	308.7, 341.9	306.1, 356.4	295.4, 348.3	95.5, 108.5	103.4, 116.9	234.7, 270.5	227.5, 265.3
<b>Variance components</b>											
$\sigma_g^2$		0.112 ± 0.022**	0.097 ± 0.017**	21.27 ± 3.74**	30.56 ± 4.71**	75.27 ± 12.45**	74.98 ± 11.89**	4.67 ± 0.99**	4.87 ± 0.89**	29.42 ± 4.89**	33.29 ± 4.76**
$\sigma_{ge}^2$		0.027 ± 0.020	-0.006 ± 0.013	3.16 ± 1.89	10.60 ± 2.29**	13.92 ± 7.32**	32.65 ± 6.57**	5.54 ± 0.92**	4.78 ± 0.68**	-1.35 ± 3.03	1.01 ± 1.78
$\sigma_e^2$		0.369 ± 0.024	0.283 ± 0.018	30.80 ± 2.02	34.26 ± 2.20	119.87 ± 8.91	90.51 ± 6.71	10.28 ± 0.76	7.50 ± 0.56	62.17 ± 4.00	35.70 ± 2.29
<b>Heritability</b>											
$h^2$		0.68	0.74	0.82	0.82	0.80	0.79	0.64	0.70	0.80	0.88
95% C.I. on $h^2$ <sup>c</sup>		0.56, 0.76	0.65, 0.80	0.75, 0.87	0.75, 0.86	0.73, 0.85	0.72, 0.84	0.50, 0.73	0.59, 0.77	0.72, 0.85	0.83, 0.91
<b>Correlation coefficients<sup>d</sup></b>											
$r_p$		0.39**		0.53**		0.60**		0.57**		0.65**	
$r_g$		0.62 ± 0.12 <sup>b</sup>		0.60 ± 0.07		0.73 ± 0.08		0.76 ± 0.28		0.79 ± 0.07	

\*\* Significant at P < 0.01.

<sup>a</sup>  $\bar{P}$  = TC mean of P1 and P2;  $\bar{F}_3$  = TC mean of F<sub>3</sub> lines.

<sup>b</sup> Standard errors are attached.

<sup>c</sup> Confidence intervals of  $h^2$  were calculated by the method of Knapp et al. (1985).

<sup>d</sup> Phenotype ( $r_p$ ) and genotype ( $r_g$ ) correlation coefficients between TC progenies with T1 and T2.

**TABLE 3**  
**QTL affecting GY, GM, KW, PC, and PH detected in maize TC progenies of F<sub>3</sub> lines from cross P1 × P2 in Experiment 1**  
**(calibration, N = 344) and respective QTL effects estimated in Experiment 2 (validation, N = 107)**

Chromosome	Experiment 1										Experiment 2			
	Position on chromosome (cM)		LOD score		R <sup>2</sup> a (%)		α-effect <sup>b</sup> (Mg ha <sup>-1</sup> )		R <sup>2</sup> a (%)		α-effect <sup>b</sup> (Mg ha <sup>-1</sup> )			
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2		
<b>Grain yield (GY)</b>														
1	268	—	3.2	—	2.3	—	-0.19	—	3.3	—	0.24	—		
2	22	—	5.7	—	6.0	—	-0.38	—	2.3	—	-0.22	—		
5	40	—	3.8	—	4.6	—	-0.45	—	0.1	—	-0.07	—		
5	64	—	4.0	—	4.0	—	0.43	—	0.1	—	0.07	—		
5	140	—	5.0	—	5.5	—	0.30	—	0.0	—	0.02	—		
7	—	74	—	3.1	—	9.8	—	0.46	—	2.1	—	0.15		
8	54	—	5.9	—	8.6	—	-0.40	—	1.4	—	-0.15	—		
8	—	80	—	3.5	—	6.4	—	-0.35	—	1.7	—	-0.15		
10	70	—	3.7	—	7.9	—	0.38	—	4.2	—	0.26	—		
Total <sup>c</sup>	—	—	26.7	11.9	30.8	14.8	2.53	0.82	12.2	3.8	1.04	0.30		
<b>Grain moisture (GM)</b>														
1	—	46	—	12.8	—	15.8	—	-9.8	—	9.7	—	-7.7		
1	68	—	3.4	—	6.0	—	-3.6	—	1.6	—	-2.0	—		
1	240	238	4.7	7.5	5.9	8.5	-4.7	-6.3	1.9	3.5	-2.4	-3.6		
2	22	—	5.2	—	5.8	—	4.2	—	1.3	—	1.8	—		
2	—	102	—	5.1	—	6.2	—	5.7	—	15.1	—	8.3		
2	130	138	8.9	13.5	9.8	10.5	5.8	6.8	0.2	2.7	0.7	-3.0		
2	180	—	4.2	—	6.5	—	-4.2	—	1.7	—	-1.9	—		
3	—	114	—	9.6	—	12.1	—	-5.8	—	1.3	—	-1.7		
3	—	174	—	4.4	—	6.1	—	4.4	—	0.1	—	-0.5		
4	130	128	7.2	2.7	9.6	4.3	-5.9	-4.4	7.1	0.3	-4.4	-1.0		
6	—	70	—	3.0	—	5.1	—	-3.6	—	2.8	—	-2.3		
7	2	0	4.8	3.8	6.0	3.7	4.4	3.5	0.0	0.1	0.1	0.5		
7	30	32	2.6	2.6	2.8	2.4	-3.2	-3.1	0.4	0.5	-1.1	-1.2		
7	74	—	3.9	—	4.5	—	3.4	—	6.4	—	3.7	—		
8	52	52	4.5	4.2	14.8	13.4	-6.8	-7.0	0.6	3.0	-1.1	-1.7		
8	106	114	12.6	6.4	17.2	8.3	9.6	7.3	2.8	17.2	3.5	10.3		
9	138	—	3.6	—	3.9	—	-3.6	—	7.1	—	4.3	—		
10	—	84	—	12.1	—	11.6	—	7.4	—	19.2	—	8.6		
Total	—	—	45.2	58.2	45.6	55.1	59.2	74.9	24.3	46.4	26.9	51.3		

Continued



**TABLE 3**  
**Continued**

Chromosome	Experiment 2										Experiment 1			
	Position on chromosome (cM)		LOD score		$\mathcal{H}^{2a}$ (%)		$\alpha$ -effect <sup>b</sup> (mg)		$\mathcal{H}^{2a}$ (%)		$\alpha$ -effect <sup>b</sup> (mg)			
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2		
<b>Kernel weight (KW)</b>														
1	246	232	5.4	6.9	7.1	7.0	-6.8	-6.4	3.1	3.6	-5.5	-5.5		
2	74	66	9.4	3.0	11.2	2.4	-7.9	-3.2	9.2	0.0	-8.8**	-0.3		
2	122	122	3.1	5.9	5.1	7.2	5.9	6.8	0.1	0.1	0.8	1.0		
2	148	156	4.7	5.1	6.8	4.8	-6.3	-5.3	0.3	0.0	1.6	0.4		
3	—	100	—	3.6	—	3.6	—	-3.7	—	0.0	—	-0.3		
5	50	56	8.2	7.8	11.9	8.7	-6.8	-5.8	14.0	8.6	-8.6**	-6.8**		
5	132	128	5.8	8.4	16.4	10.2	-8.3	-6.8	3.5	4.0	-4.2	-5.1*		
6	70	—	4.5	—	4.8	—	-3.9	—	0.0	—	-0.4	—		
7	58	60	15.9	13.5	21.1	14.9	-10.7	-9.2	6.5	14.5	-5.6*	-9.5**		
8	48	48	22.2	7.3	25.5	17.0	-11.0	-8.9	24.2	8.9	13.3**	-7.9**		
9	0	—	5.3	—	8.1	—	-5.3	—	0.0	—	-0.5	—		
9	102	120	2.9	3.4	5.6	13.8	5.1	8.7	1.9	6.7	3.4	7.0**		
10	112	112	6.7	5.4	8.5	4.6	5.8	4.3	4.0	7.0	4.2	6.1**		
Total	—	—	73.7	54.2	63.7	52.5	83.8	69.1	47.7	39.8	57.0	49.9		
<b>Protein concentration (PC)</b>														
1	—	2	—	3.7	—	4.0	—	1.2	—	1.8	—	1.0		
1	238	220	9.1	2.6	12.4	1.2	-2.4	-0.6	11.6	0.8	-3.5**	-0.6		
2	146	136	5.2	2.8	3.8	5.3	-1.3	-1.6	4.0	6.0	-1.9	-1.9*		
2	180	180	2.8	3.3	3.7	2.6	-1.2	-1.0	0.2	0.1	0.4	0.2		
3	110	—	4.7	—	1.7	—	0.7	—	2.2	—	-1.1	—		
3	—	164	—	4.8	—	4.7	—	-1.4	—	11.9	—	-2.7**		
3	198	—	3.1	—	1.4	—	0.7	—	0.0	—	0.1	—		
4	0	0	4.8	3.0	5.7	4.2	-1.3	-1.2	1.9	0.0	-1.1	-0.0		
5	38	36	3.9	12.2	2.8	11.1	0.9	2.1	0.0	4.7	0.1	1.6*		
6	78	74	5.8	11.7	5.8	13.4	-1.3	-2.3	0.3	0.2	-0.4	0.4		
6	130	132	6.4	4.5	7.0	6.5	1.5	1.6	0.0	1.1	-0.1	-0.8		
7	—	62	—	5.1	—	10.4	—	-2.1	—	8.8	—	-2.2**		
Total	—	—	35.2	41.9	37.7	43.0	11.2	15.0	19.5	25.7	8.6	11.4		

Continued

**TABLE 3**  
**Continued**

Chromosome	Experiment 1										Experiment 2			
	Position on chromosome (cM)		LOD score		R <sup>2</sup> a (%)		α-effect <sup>b</sup> (cm)		R <sup>2</sup> a (%)		α-effect <sup>b</sup> (cm)			
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2		
Plant height (PH)														
1	142	126	6.2	3.6	15.2	5.5	-4.3 <sup>..</sup>	-3.8 <sup>..</sup>	1.7	2.6	-1.8	-3.5		
1	—	184	—	5.5	—	6.4	—	-4.1 <sup>..</sup>	—	0.7	—	1.8		
1	236	230	17.6	27.0	27.9	25.2	-7.6 <sup>..</sup>	-8.1 <sup>..</sup>	24.7	19.1	-8.9 <sup>**</sup>	-9.0 <sup>**</sup>		
2	—	20	—	2.6	—	4.0	—	2.5	—	0.2	—	-0.7		
2	136	144	5.9	5.3	5.9	6.2	2.6	2.8	2.6	0.7	2.3	1.2		
3	112	116	3.2	6.8	1.8	8.1	1.4	3.2 <sup>..</sup>	7.0	3.0	3.4 <sup>*</sup>	2.6		
3	154	164	4.9	7.3	6.3	8.8	3.4	4.0 <sup>..</sup>	1.3	2.6	1.9	2.8		
4	118	114	10.7	9.6	12.7	12.3	-5.0	-5.1 <sup>..</sup>	6.9	3.2	-4.4 <sup>*</sup>	-3.2		
5	72	90	4.0	3.3	9.3	5.7	-3.4	-3.1	20.8	0.3	-6.8 <sup>**</sup>	0.9		
5	130	—	3.1	—	3.1	—	-2.4	—	1.5	—	2.2	—		
5	152	—	4.5	—	5.9	—	3.2	—	0.0	—	0.3	—		
6	154	156	5.9	8.0	8.1	8.1	-2.9	-3.1	1.6	0.7	-1.8	-1.3		
7	—	0	—	4.5	—	5.7	—	-2.6	—	2.3	—	-2.1		
7	40	50	3.0	3.1	2.0	12.4	2.2	4.6	6.0	1.1	4.7 <sup>*</sup>	1.5		
7	68	—	6.2	—	3.4	—	2.7	—	0.5	—	-1.2	—		
8	—	22	—	4.3	—	4.3	—	-2.3	—	2.3	—	-2.3		
8	52	—	7.1	—	8.5	—	-3.0	—	8.7	—	-4.0 <sup>**</sup>	—		
8	144	—	5.1	—	4.5	—	2.1	—	0.0	—	-0.2	—		
9	56	—	8.5	—	8.2	—	5.4	—	3.8	—	5.0	—		
9	136	—	2.7	—	4.0	—	2.7	—	0.8	—	-1.5	—		
10	82	70	6.0	6.6	10.1	6.2	4.1	2.8 <sup>..</sup>	4.4	0.0	3.2	0.2		
Total	—	—	72.7	73.5	63.2	63.6	58.4	52.1	55.7	37.5	53.5	33.1		

<sup>\*</sup>, <sup>\*\*</sup> Significant at P < 0.05 and P < 0.01, respectively.

<sup>..</sup> QTL × E interaction significant at P < 0.05 and P < 0.01, respectively.

<sup>a</sup> Percent phenotypic variance explained by QTL. For details see materials and methods.

<sup>b</sup> Average effect of substituting allele P1 by allele P2 in the TC progenies.

<sup>c</sup> Estimates of total LOD score and coefficient of determination obtained from a simultaneous fit of all putative QTL for the respective trait.

**TABLE 4**  
**QTL affecting GY, GM, KW, PC, and PH detected in maize TC progenies of F<sub>3</sub> lines from cross P1 × P2 in Experiment 2**  
**(calibration, N = 107) and respective QTL effects estimated in Experiment 1 (validation, N = 344)**

Chromosome	Experiment 2						Experiment 1					
	Position on chromosome (cM)		LOD score		R <sup>2</sup> a (%)		α-effect <sup>b</sup> (Mg ha <sup>-1</sup> )		R <sup>2</sup> a (%)		α-effect <sup>b</sup> (Mg ha <sup>-1</sup> )	
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
<b>Grain yield (GY)</b>												
2	—	0	—	4.3	—	11.3	—	-0.31	—	2.0	—	-0.20**
3	148	—	2.9	—	4.1	—	0.29	—	0.7	—	-0.16	—
6	—	50	—	4.8	—	16.5	—	0.43	—	0.0	—	-0.00
6	—	148	—	2.6	—	9.3	—	-0.33	—	0.1	—	0.05
7	—	56	—	2.7	—	11.1	—	0.29	—	9.8	—	0.49**
Total <sup>c</sup>	—	—	0.9	9.2	4.1	32.1	0.29	1.35	0.7	11.6	0.16	0.74
<b>Grain moisture (GM)</b>												
1	—	66	—	6.6	—	16.3	—	-6.2	—	14.0	—	-7.8**
2	—	98	—	5.6	—	18.5	—	6.7	—	9.7	—	7.1**
2	—	180	—	3.8	—	7.8	—	-4.0	—	0.3	—	1.1
3	—	102	—	2.5	—	8.6	—	-3.8	—	3.1	—	-3.6**
4	140	—	3.6	—	8.5	—	-4.2	—	6.2	—	-5.1**	—
5	—	108	—	2.8	—	7.5	—	3.7	—	1.0	—	2.3
5	—	152	—	3.6	—	9.0	—	-4.5	—	0.2	—	-1.0
8	—	64	—	3.0	—	11.1	—	-5.1	—	5.5	—	-5.7**
8	—	114	—	5.6	—	21.1	—	11.5	—	6.6	—	8.3**
9	154	—	5.5	—	10.3	—	-4.4	—	2.0	—	-2.6	—
10	90	78	3.1	10.7	0.1	25.7	-0.4	8.5	0.4	7.4	1.6	6.5**
Total	—	—	4.0	20.3	17.4	57.9	9.1	53.9	6.2	31.3	9.2	43.4
<b>Kernel weight (KW)</b>												
4	—	134	—	3.0	—	8.4	—	-7.6	—	1.5	—	-3.3*
5	30	—	3.0	—	7.4	—	-7.3	—	3.4	—	-4.9**	—
5	110	94	2.6	2.8	6.7	18.8	-7.0	-11.8	11.7	9.3	-9.9**	-8.8**
7	74	74	3.1	4.3	7.4	15.7	-6.5	-9.7	7.5	8.6	-7.1**	-7.2**
8	38	52	5.1	4.2	25.8	15.2	-15.4	-10.1	20.9	16.4	-14.4**	-10.1**
9	—	92	—	4.3	—	12.4	—	9.0	—	8.3	—	7.3**
Total	—	—	11.3	13.6	41.5	43.7	36.1	48.3	35.9	33.0	36.2	36.7

Continued

**TABLE 4**  
**Continued**

Chromosome	Experiment 2										Experiment 1			
	Position on chromosome (cM)		LOD score		R <sup>2a</sup> (%)		α-effect <sup>b</sup> (g kg <sup>-1</sup> )		R <sup>2a</sup> (%)		α-effect <sup>b</sup> (g kg <sup>-1</sup> )			
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2		
<b>Protein concentration (PC)</b>														
1	258	—	3.3	—	12.4	—	-3.2	—	7.5	—	-2.0**	—		
2	118	114	3.8	3.1	15.9	11.5	-3.1	-2.5	5.7	4.8	-1.6**	-1.8**		
7	—	72	—	2.9	—	9.0	—	-2.0	—	3.9	—	-1.5**		
8	50	50	2.5	7.7	10.0	24.9	-2.4	-5.1	2.2	0.2	-0.9**	-0.5		
8	—	64	—	4.7	—	12.2	—	3.3	—	0.0	—	0.0		
Total	—	—	8.0	10.1	31.6	34.8	8.6	12.9	15.3	9.3	4.5	3.8		
<b>Plant height (PH)</b>														
1	238	232	5.9	6.3	27.9	23.5	-10.4	-9.8	11.7	20.7	-6.7**	-9.3**		
3	118	—	3.3	—	7.8	—	3.7	—	1.0	—	1.7	—		
4	36	—	2.8	—	10.3	—	-5.4	—	0.2	—	-0.9	—		
5	56	—	6.0	—	16.6	—	-5.1	—	4.2	—	-2.9**	—		
8	—	64	—	3.3	—	3.7	—	-3.0	—	0.0	12.2	-0.0		
Total	—	—	12.1	7.3	43.6	26.5	24.7	12.8	16.1	20.8	12.2	9.3		

\*, \*\* Significant at  $P < 0.05$  and  $P < 0.01$ , respectively.

\*, \*\* QTL × E interaction significant at  $P < 0.05$  and  $P < 0.01$ , respectively.

<sup>a</sup> Percent phenotypic variance explained by QTL. For details see materials and methods.

<sup>b</sup> Average effect of substituting allele P1 by allele P2 in the TC progenies.

<sup>c</sup> Estimates of total LOD score and total variance explained obtained from simultaneous fit of all putative QTL for the respective trait.

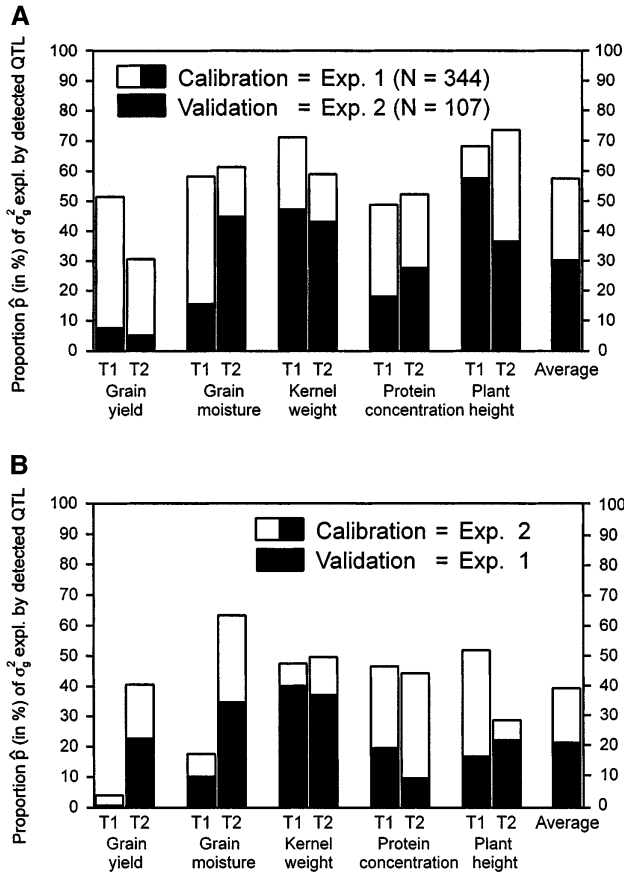


Figure 2.—Proportion ( $\hat{p}$ ) of  $\hat{\sigma}_g^2$  explained by the detected QTL in maize TC progenies of  $F_3$  lines from cross  $P_1 \times P_2$ . (A) Calibration in Experiment 1 ( $N = 344$ ), validation in Experiment 2 ( $N = 107$ ). (B) Calibration in Experiment 2, validation in Experiment 1.

T1 and  $R^2 = 43.7\%$  and  $\hat{p} = 49.6\%$  for TC with T2. Three QTL were in common between testers and none showed significant QTL  $\times$  E interactions.

Three QTL for T1 and two QTL for T2 were in common between Experiment 1 and 2 (Table 5), including the largest QTL explaining about 25.5% of  $\hat{\sigma}_p^2$  in both

experiments (Tables 3 and 4). Estimates of  $\alpha$ -effects from validation in Experiment 2 were on average almost as large as those from calibration in Experiment 1 and agreed in sign except for one QTL on chromosome 2 (Table 3). Collectively, the QTL from validation in Experiment 2 explained  $R^2 = 47.7\%$  and  $\hat{p} = 47.3\%$  for TC with T1 and  $R^2 = 39.8\%$  and  $\hat{p} = 43.1\%$  for TC with T2. Likewise,  $\hat{\alpha}$ -effects from calibration in Experiment 2 were mostly in close agreement with those obtained from validation in Experiment 1 (Table 4). Collectively, the QTL from validation in Experiment 1 explained  $R^2 = 35.9\%$  and  $\hat{p} = 40.2\%$  for TC with T1 and  $R^2 = 33.0\%$  and  $\hat{p} = 37.1\%$  for TC with T2.

**Protein concentration:** In Experiment 1, nine and ten QTL influencing PC in TC with T1 and T2, respectively, were mapped. A simultaneous fit yielded  $R^2 = 37.7\%$  and  $\hat{p} = 48.8\%$  for TC with T1 and  $R^2 = 43.0\%$  and  $\hat{p} = 52.3\%$  for TC with T2. Altogether, five QTL showed significant ( $P < 0.05$ ) QTL  $\times$  E interactions. Seven QTL were in common for both testers with  $\hat{\alpha}$ -effects of similar size and same sign.

In Experiment 2, three QTL affected PC in TC with T1 ( $R^2 = 31.6\%$  and  $\hat{p} = 46.6\%$ ) and four QTL with T2 ( $R^2 = 34.8\%$  and  $\hat{p} = 44.4\%$ ). Two QTL were in common between both testers. None of the QTL displayed significant QTL  $\times$  E interactions.

Only one QTL was in common between Experiment 1 and 2 for each tester (Table 5). In several instances,  $\hat{\alpha}$ -effects from calibration in Experiment 1 differed in sign (six QTL) or deviated in magnitude from those estimated from validation in Experiment 2 (Table 3). In the latter analysis, we obtained  $R^2 = 19.5\%$  and  $\hat{p} = 18.0\%$  for TC with T1 and  $R^2 = 25.7\%$  and  $\hat{p} = 27.7\%$  for TC with T2. Likewise,  $\hat{\alpha}$ -effects from validation in Experiment 1 were consistently smaller than those obtained from calibration in Experiment 2 and resulted in reduced estimates of  $R^2 = 15.3\%$  and  $\hat{p} = 19.5\%$  for TC with T1 and  $R^2 = 9.3\%$  and  $\hat{p} = 9.6\%$  for TC with T2 (Table 4).

**Plant height:** In Experiment 1, 17 and 14 QTL affecting PH in TC with T1 and T2, respectively, were identified

TABLE 5  
Number of putative QTL detected and in common<sup>a</sup> (∩) for maize TC progenies of  $F_3$  lines with testers T<sub>1</sub> and T<sub>2</sub> in Experiments 1 and 2

Experiment/tester	Grain yield	Grain moisture	Kernel weight	Protein concentration	Plant height	Sum
1T1	7	12	12	9	17	57
1T2	2	13	11	10	14	50
1T1 ∩ 1T2 <sup>a</sup>	0	7	10	7	10	34
2T1	1	3	4	3	4	15
2T2	4	9	5	4	2	24
2T1 ∩ 2T2	0	1	3	2	1	7
1T1 ∩ 2T1	0	2	3	1	3	9
1T2 ∩ 2T2	1	6	2	1	1	11

<sup>a</sup> See materials and methods.

on all 10 chromosomes (Table 3). A simultaneous fit with all QTL accounted for  $R^2 = 63.2\%$  and  $\hat{p} = 68.2\%$  in TC with T1 and  $R^2 = 63.6\%$  and  $\hat{p} = 73.6\%$  in TC with T2. Ten QTL were in common with similar  $\hat{\alpha}$ -effects for both testers. Nine QTL displayed significant ( $P < 0.05$ ) QTL  $\times$  E interactions.

In Experiment 2, four QTL were found in TC with T1. Collectively, these QTL explained  $R^2 = 43.6\%$  and  $\hat{p} = 51.9\%$ . Two QTL were found in TC with T2 with  $R^2 = 26.5\%$  and  $\hat{p} = 28.7\%$  in a simultaneous fit. The largest QTL on chromosome 1 explaining more than 23% of  $\hat{\sigma}_p^2$  was in common between both testers.

Three QTL for TC with T1 and one QTL for TC with T2 were in common between Experiment 1 and 2, including the largest QTL found for both testers on chromosome 1 (Table 5). Estimates of  $\alpha$ -effects from validation in Experiment 2 were largely consistent in sign and magnitude with those from calibration in Experiment 1 (Table 3). Collectively, the former explained  $R^2 = 55.7\%$  and  $\hat{p} = 57.6\%$  for TC with T1 and  $R^2 = 37.5\%$  and  $\hat{p} = 36.6\%$  for TC with T2. Validation in Experiment 1 based on calibration in Experiment 2 yielded reduced  $\hat{\alpha}$ -effects with  $R^2 = 16.1\%$  and  $\hat{p} = 16.8\%$  for TC with T1 and  $R^2 = 20.8\%$  and  $\hat{p} = 22.1\%$  for TC with T2.

**Digenic epistasis between detected QTL:** In Experiment 1, the test for digenic epistatic interactions ( $\alpha\alpha$ -effects) among detected QTL was significant ( $P < 0.05$ ) in few instances. In TC with T1, we found epistasis only for GM between the QTL on chromosome 2 (position 180 cM) and chromosome 8 (position 106 cM) with  $\hat{\alpha}\hat{\alpha} = -2.0 \text{ g kg}^{-1}$  and the QTL on chromosome 7 (position 2 cM) and chromosome 8 (position 52 cM) with  $\hat{\alpha}\hat{\alpha} = -1.7 \text{ g kg}^{-1}$ . In TC with T2, epistasis was indicated for GM between the QTL on chromosome 4 (position 128 cM) and chromosome 8 (position 114 cM) with  $\hat{\alpha}\hat{\alpha} = -3.9 \text{ g kg}^{-1}$  and for KW between two linked QTL on chromosome 2 (position 122 cM and position 156 cM) with  $\hat{\alpha}\hat{\alpha} = -2.30 \text{ mg}$  and between the QTL on chromosome 5 (position 56 cM) and chromosome 8 (position 48 cM) with  $\hat{\alpha}\hat{\alpha} = 2.35 \text{ mg}$ . Including the  $\hat{\alpha}\hat{\alpha}$ -effects for these pairs of QTL in the model for the simultaneous fit increased the  $R^2$  values only by 2–3% compared to the model without epistasis. None of the epistatic interactions were confirmed by validation in Experiment 2 and the  $R^2$  values for the epistatic model decreased usually by 1–2% in comparison to the model without epistasis. In Experiment 2, we found no significant ( $P < 0.05$ ) digenic epistasis between any of the detected QTL.

**Correlation between predicted and observed TC performance:** In Experiment 1, estimates of the genotypic correlation  $r_g(Y_{jz}, M_{jz})$  exceeded 0.70 for KW and PH, and ranged between 0.60 and 0.53 for GM and PC, but were below 0.39 for GY (Table 6). In Experiment 2,  $r_g(Y_{jz}, M_{jz})$  was below 0.50 in most cases, except for KW where it ranged from 0.62 to 0.69.

## DISCUSSION

**Advantages of CIM:** A comparison of our results in Experiment 1 for PC and KW with those of Schön *et al.* (1994) clearly demonstrates the advantages of CIM over simple interval mapping. Both investigations relied on the same data set and employed the same LOD threshold for QTL detection. For both traits, we found about twice the number of QTL and a much better agreement between testers than reported by Schön *et al.* (1994). This was due to the detection of additional QTL and a better resolution of linked QTL with CIM, as expected from theory and simulation results (Zeng 1994). The  $R^2$  values for the simultaneous fit were only marginally increased with CIM. However, this comparison is confounded with the difference in  $R^2$  estimates obtained by the regression and maximum likelihood approach (Xu 1995) implemented in software packages PLABQTL and MAPMAKER/QTL, respectively, employed in the two studies.

**Estimation of QTL effects from independent samples:** Estimates of individual QTL effects were in most cases considerably smaller when estimated from an independent validation experiment in lieu of the calibration experiment (Tables 3 and 4). In some cases effects of opposite sign were found in the validation experiment, suggesting the occurrence of a type III error (*i.e.*, a significant association is correctly declared but the marker allele is associated with the wrong QTL allele; Dudley 1993). For all traits and both testers,  $\hat{p}$  from the simultaneous fit with all detected QTL was considerably smaller for validation than for calibration (Figure 2). Averaged over all traits and both testers,  $\hat{p}$  dropped from 57.5% for calibration in Experiment 1 to 30.3% for validation in Experiment 2, and from 39.4% for calibration in Experiment 2 to 21.3% for validation in Experiment 1. The decrease in  $\hat{p}$  was particularly pronounced for GY, probably due to its complex genetic architecture.

In our opinion, the decrease in  $\hat{p}$  can mainly be attributed to two factors: (i) the effect of different samples and (ii) the effect of different environments. Both factors are confounded and with currently available statistical models it is not possible to separate them. However, the majority of the detected QTL showed no significant QTL  $\times$  E interactions, suggesting that different test environments for the two experiments were not the major cause for identification of different QTL. Additionally, the environments used in Experiments 1 and 2 were assumed a random sample of environments available for testing performance of maize in Germany. This assumption was corroborated by the similar magnitude of phenotypic correlations for performance of the 10 check varieties when calculated for environments within Experiments 1 and 2 as compared to correlations between environments of Experiments 1 and 2. Furthermore, when practicing MAS, gain from selection will

TABLE 6

Genotypic correlation  $r_g(Y_{jz}, M_{jz})$  between observed performance ( $Y_{jz}$ ) of maize TC progenies of  $F_3$  lines  $j$  with tester  $T_z$ , and their prediction ( $M_{jz}$ ) from QTL mapping results of TC progenies with tester  $T_z$  in Experiments 1 and 2

Experiment	Tester <sup>a</sup>		Grain yield	Grain moisture	Kernel weight	Protein concentration	Plant height
	$Y_{jz}$	$M_{jz}$					
1	T2	T1	$0.38 \pm 0.07$	$0.59 \pm 0.04$	$0.73 \pm 0.03$	$0.53 \pm 0.05$	$0.84 \pm 0.02$
	T1	T2	$0.26 \pm 0.07$	$0.59 \pm 0.05$	$0.75 \pm 0.03$	$0.58 \pm 0.05$	$0.73 \pm 0.03$
2	T2	T1	$0.29 \pm 0.11$	$0.12 \pm 0.11$	$0.62 \pm 0.08$	$0.48 \pm 0.10$	$0.44 \pm 0.08$
	T1	T2	$0.29 \pm 0.12$	$0.35 \pm 0.10$	$0.69 \pm 0.07$	$0.57 \pm 0.10$	$0.51 \pm 0.09$

Standard errors are attached.

<sup>a</sup> Tester employed for determining  $Y_{jz}$  (TC performance) and  $M_{jz}$  (marker index score).

usually be assessed in environments (years) different from those in which calibration was performed.

On the other hand, computer simulations (Beavis 1994; Utz and Melchinger 1994; Georges *et al.* 1995) demonstrated that statistical sampling has a strong impact on QTL analyses and that the bias in QTL effects estimated from calibration can be severe, the most important factors being the sample size  $N$ , the magnitude of the QTL effect, and  $h^2$ . Utz and Melchinger (1994) showed that with CIM the  $\mathfrak{R}^2$  value of a QTL explaining 8% of  $\sigma_g^2$  for a trait with  $h^2 = 0.4$  can be overestimated up to 388% for  $N = 100$  and up to 44% for  $N = 300$ . With experimental data the bias in QTL effects is expected to be even greater than in computer simulations given the uncertainties in the selection of cofactors and the obscuring effects of missing marker data and QTL  $\times$  E interactions.

The inflation in the  $R^2$  and  $\hat{p}$  values of QTL estimated directly from calibration can be attributed to several reasons. All are related to the fact that QTL mapping can be considered as a problem of model selection in multiple linear regression (Haley and Knott 1992; Whittaker *et al.* 1996). Using results of Haley and Knott (1992), the partial  $\mathfrak{R}^2$  value of a putative QTL in regression according to Equation 1 can be linked with its LOD score and the sample size  $N$  (see appendix a)

$$\mathfrak{R}^2 = 1 - e^{-\text{LOD}/(0.2171 * N)}. \quad (2)$$

Therefore, a QTL search based on the LOD score criterion is according to Equation 2 equivalent to selecting for those regressor variables, which account for the largest proportion ( $\mathfrak{R}^2$ ) of the variance in the response variable ( $\sigma_p^2$ ) and consequently, faces the same problems of model selection as does multiple linear regression. As is well-known from the statistical literature (see *e.g.*, Rencher and Pun 1980; Freedman 1983), model selection leads to an inflation in the  $\mathfrak{R}^2$  values of the selected explanatory variables, the bias being very severe when the number of observations is small and close to the number of predictor variables. Furthermore, the presence of closely linked markers can introduce multicollinearity among the regressors with negative impacts

on the quality and stability of the selected model. In particular, it increases the variance of the estimated regression coefficients and can also strongly affect their magnitude (Jobson 1991).

Suggestions in the statistical literature (for review, see Miller 1990) to diminish these problems include (1) model validation with an additional sample as proposed by Lande and Thompson (1990) or (2) cross-validation in the case of larger sample sizes. For validation we calculated the regressors without model selection based on QTL positions identified in the calibration and estimated the partial regression coefficients based on the *a priori* chosen model. Hence, the estimated regression coefficients are unbiased in the validation based on standard linear model theory. Beavis (1994) proposed the use of resampling strategies for reducing the bias by using results from experiments with multiple independent samples of progeny. Other resampling strategies such as bootstrapping may be a further alternative for eliminating the bias. However, when using CIM, it is not obvious from which pool to draw the bootstrap samples for estimating the QTL parameters (Visscher *et al.* 1996).

While the absolute proportion of  $\sigma_g^2$  explained by QTL in validation differed substantially, depending upon whether Experiment 1 or 2 was used for calibration, the relative decrease in  $\hat{p}$  from calibration to validation was largely independent of the sample size used for calibration. This could be attributable to the fact that with a larger sample size additional QTL with smaller effects are detected in calibration. Estimates of the effects of these QTL are very likely subject to considerable sampling bias and, therefore, contribute substantially to the inflation in  $\hat{p}$  values estimated from calibration even with large sample sizes.

The lack of consistency between QTL effect estimates obtained from calibration and validation has several important consequences for QTL mapping and MAS for polygenic traits: (1) It demonstrates that, due to sampling and QTL  $\times$  E interactions, individual QTL effects estimated directly from calibration can be inflated, especially for smaller values of  $N$  and complexly

inherited traits such as GY. Inferences about the relative magnitude of QTL effects estimated from previous experimental studies should be reexamined under this aspect. (2) The distribution of estimated QTL effects may not reflect the distribution of true QTL effects. A large estimate may reflect either a large QTL or a small QTL estimated with a large bias. (3) The decision of which QTL regions to transfer with MAS and/or to consider in a selection index should be based on QTL effects verified in an independent validation sample. (4) For a correct assessment of the prospects of MAS, the key parameter  $p$  must not be determined from calibration, but from an independent validation sample or by using cross-validation.

**Comparison of QTL detected in samples of different size:** We evaluated the power of QTL detection by comparing results from QTL mapping in two independent samples of different size from the same population. The smaller sample size ( $N = 107$ ) in Experiment 2 was chosen in accordance with (1) most experimental QTL studies reported in the literature and (2) the maximum number of progenies generally employed per cross for early testing in recycling breeding (Beavis *et al.* 1994). In Experiment 1 we chose, from a breeder's point of view, a large sample size ( $N = 344$ ) to meet the minimum requirements for the detection of smaller QTL, as suggested by theory (Lander and Botstein 1989; Lande and Thompson 1990). From Equation 2 it can be shown that with a LOD threshold of 2.5 we were able to detect a QTL accounting for at least 10.2% of  $\sigma_p^2$  in Experiment 2 ( $N = 107$ ), but as little as 3.3% of  $\sigma_p^2$  in Experiment 1 ( $N = 344$ ). (Smaller values found in Tables 3 and 4 are due to the fact that these estimates refer to partial  $\mathfrak{R}^2$  values from a simultaneous fit of all detected QTL, which can deviate from the  $\mathfrak{R}^2$  values calculated in multiple regression according to Equation 1 due to confounding effects of undetected minor QTL linked in repulsion phase). As a consequence, the total number of QTL detected for all traits and both testers in Experiment 1 was almost triple the number detected in Experiment 2.

Only about half (20) of the putative QTL detected in Experiment 2 were in common with QTL identified in Experiment 1 and the poorest agreement was observed for GY (Table 5). As pointed out earlier, the comparison of results between Experiments 1 and 2 is confounded by the different test environments and very likely both factors, sampling and QTL  $\times$  E interactions, contributed to the lack of congruency of QTL found for the two experiments. In a comparison of QTL mapping results from two independent studies with elite cross B73  $\times$  Mo17, Beavis *et al.* (1994) found similar results for GY. These authors concluded that the lack of congruency was mainly attributable to sampling of progeny because the sample sizes used in their study were small. However, even with large sample sizes the statistical power of QTL detection is only moderate for QTL with

smaller effects as demonstrated by various simulation studies (Van Ooijen 1992; Beavis 1994; Utz and Melchinger 1994). For example, the power for detecting a QTL explaining 3.5% of  $\sigma_p^2$  in Experiment 1 is only about 0.5. Consequently, if such a QTL is detected in one experiment, it has only an even chance of being identified in an independent set of progeny. This argument applies to (1) small QTL and large values of  $N$  or (2) large QTL and small values of  $N$ , but it does not suffice to explain why half of the large QTL detected in Experiment 2 were not recovered in Experiment 1. This is because with  $N = 344$ ,  $h^2 > 0.5$ , and  $\text{LOD} > 2.5$ , the power for detecting a QTL which supposedly accounts for 10% or more of  $\sigma_p^2$ , exceeds 0.90 (H. F. Utz, unpublished results).

This apparent gap of explanation can be closed by considering that many of the QTL effects estimated in Experiment 2 had a large upward bias, as discussed earlier. Assuming their true effects were often much smaller, it follows in combination with the previous argument that there was only a moderate chance of detecting them simultaneously in Experiment 1. In addition, we cannot rule out that a few of the putative QTL were either environment-specific or "false positives" and therefore occurred only in one experiment, given that a LOD threshold of 2.5 corresponds in our study to a genomewide Type I error rate  $P_g \leq 0.25$ .

For testing congruency of QTL it was not possible to adopt a criterion based on overlapping confidence intervals, because with CIM their computation is still an unsolved problem (Visscher *et al.* 1996). We declared two QTL as being in common if they had the same sign and were within a 20-cM distance. Using a wider interval length (*e.g.*, 40 cM) would have increased the proportion of common QTL only marginally but entailed a high risk that well-separated different QTL are declared as common, if they have by chance the same sign. This applies particularly for such traits as PH, where a large number of QTL was detected.

The "genetic architecture" of a trait characterized by the number of effective factors (Wright 1968) has an impact on both the power of QTL detection and the magnitude of the bias when estimating QTL effects. With a large number of minor QTL influencing a quantitative trait, the power of QTL detection and consequently the number of common QTL should be smaller than for a trait governed by a small number of major QTL. Likewise, the relative bias in  $R^2$  and  $\hat{p}$  is expected to be smaller with a small number of major QTL explaining a substantial proportion of  $\sigma_g^2$  than for a trait with a large number of minor QTL. A comparison of our results for GY and the other traits supports this hypothesis. It was interesting to observe, however, that the highest number of QTL was found for PH, a trait with presumably oligogenic inheritance. Before the advent of molecular markers, estimates on the number of genes involved in expression of quantitative traits were



mainly based on Wright's (1968) formula, which severely underestimates the number of effective factors involved in trait expression if certain assumptions such as purely additive gene action and independent segregation of genes with equal effects are not met. Based on results from QTL mapping studies, it seems very likely that even highly heritable traits like PH are regulated by a large number of genes and that assumptions about the inheritance of these traits need to be revised. Similar findings have recently been reported by Cheverud *et al.* (1996) for murine growth.

The lack of congruency among QTL detected in two samples from the same cross provides the baseline for comparisons of QTL detected in populations derived from different crosses. Therefore, it was not surprising that most QTL regions reported here were either unique or found in just one or two comparable studies in the literature. Only one QTL region adjacent to marker *umc89* on chromosome 8 affecting both GY and GM was identified in several other investigations (Stuber *et al.* 1992; Beavis *et al.* 1994; Bohn *et al.* 1996). Likewise, the QTL region adjacent to *umc140* on chromosome 9 was repeatedly shown to have a large effect on KW (Beavis *et al.* 1994; Austin and Lee 1996). Three of the seven QTL detected for PC in cross IHP  $\times$  ILP (Goldman *et al.* 1993) were also active in Experiments 1 and 2 with both testers.

In addition to the confounding factor of sampling, two features of our experimental materials could explain the singular set of QTL reported here: (1) Our mapping population was generated from a cross of two elite European flint lines, whereas all other QTL studies in maize have employed wide crosses between North American dent lines or tropical germplasm. Because flint and dent are fairly distinct germplasm groups, we hypothesize that they have only a small subset of polymorphic QTL in common. (2) In practical breeding programs, elite lines from the same heterotic group are crossed and early selfing generations ( $F_2$  plants or  $F_3$  lines) are evaluated for their TC performance in combination with testers from the opposite heterotic group. We therefore mapped QTL for TC performance as opposed to line *per se* performance commonly determined in most previous QTL studies. This can result in largely different sets of QTL as demonstrated in a comparison of both features in cross B73  $\times$  Mo17 (Beavis *et al.* 1994) and expected from theory (see appendix b).

**Comparison among testers:** According to theory (see appendix b), consistent QTL mapping results across testers are expected in the absence of epistasis if both testers have identical alleles at the QTL, or additive gene action prevails, or the dominance effects satisfy the condition

$$d_{1T1} - d_{2T1} = d_{1T2} - d_{2T2}. \quad (3)$$

In all these cases, interactions in the two-way table of TC means with parents representing one factor and

testers the second factor are absent and  $r_g$  between different TC series is 1.0. Conversely, inconsistent results can arise when (a) the two testers have different alleles at the QTL ( $T1 \neq T2$ ) and Equation 3 is not satisfied (*i.e.*, the alleles in P1 and P2 show different dominance relationships with each tester allele) or (b) the QTL alleles of P1 and P2 display different epistatic interactions with the tester alleles at other loci. Obviously, both cases also result in lower estimates for  $r_g$ .

With the exception of GY, our QTL mapping results in Experiment 1 agreed well across testers for all traits: more than half of the QTL detected with one tester were also found with the other tester and the proportion of common QTL was in close agreement with the magnitude of  $\hat{r}_g$ . This is consistent with the preponderance of additive gene action found for these traits in classic quantitative-genetic experiments (Hallauer and Miranda 1981, Chapter 5) and recent QTL mapping studies for *per se* performance in maize (Beavis *et al.* 1994; Berke and Rocheford 1995; Bohn *et al.* 1996).

The absence of common QTL between both testers observed for GY can be explained by several causes, the most important being related to gene action. Studies on GY exhibited a high degree of dominance (Hallauer and Miranda 1981, Chapter 5) and a large proportion of QTL with dominance and overdominance (or pseudooverdominance) (Stuber *et al.* 1992; Beavis *et al.* 1994; Bohn *et al.* 1996; Cockerham and Zeng 1996). Under this supposition, inconsistent QTL results among testers can be explained by masking effects of the tester allele. If a QTL is detected for tester T1, but tester T2 carries an allele fully dominant over the alleles carried by P1 and P2, no QTL will be detected in its TC progenies. Epistasis between unlinked QTL and QTL  $\times$  E interactions were presumably not important causes for the inconsistencies between testers, as they were generally of minor importance.

Given that the tester may change over time in a hybrid breeding program, we examined whether a marker index score  $M_{jz}$  based on QTL mapping results with tester  $T_z$  would be effective for improving TC performance  $Y_{jz}$  with tester  $T'_z$ . For this purpose, we estimated the genotypic correlation  $r_g(Y_{jz}, M_{jz})$ , which represents the key parameter in the formula for the selection response in  $Y_j$  from indirect selection for  $M_j$  (Falconer and Mackay 1996). Our results showed for all traits except GY high enough estimates of  $r_g$  so that *for a given sample*, QTL-marker associations determined for one tester would be effective for improving TC performance with other testers. For GY, however, separate QTL mapping experiments would be required for each tester. In Experiment 2,  $\hat{r}_g(Y_{jz}, M_{jz})$  was relatively small in most cases. In combination with the low proportion of  $\sigma_g^2$  explained by the detected QTL in this experiment, these results corroborate that for complex polygenic traits a sample size  $N \approx 100$  is not sufficient to obtain reliable QTL estimates for MAS.

**Epistasis among QTL:** The comparison of TC generation means for parents ( $\bar{P}$ ) and  $F_3$  lines ( $\bar{F}_3$ ) provides a test for the net effect of epistasis across the entire genome (Melchinger 1987). In agreement with comparable studies in maize (Melchinger *et al.* 1988; Lamkey *et al.* 1995), contrasts between TC generation means were nonsignificant for most traits, leaving two possible explanations open: (1) absence of epistasis or (2) canceling of positive and negative epistatic effects among QTL in the sum. Our results from QTL analyses supported the first hypothesis because no significant digenic epistatic interactions were found among QTL detected for each trait, particularly when reassessed by validation. Stuber *et al.* (1992) also obtained no evidence for epistasis among pairs of detected QTL in their analyses of cross B73  $\times$  Mo17. However, a reanalysis of their data by Cockerham and Zeng (1996) based on single marker analyses gave evidence for substantial epistatic effects between linked QTL.

One reason for the absence of significant epistasis in our study could be that we investigated a genetically narrow cross between elite lines from the same germplasm group. In this case, there should be less opportunity to disrupt coadapted epistatic gene complexes in the parents as might be expected for wide or interspecific crosses oftentimes employed in QTL mapping studies. Furthermore, the power for detecting epistatic interactions among QTL is lower for TC performance than line *per se* performance due to masking effects of the tester (Gallais and Rives 1993).

In our analysis, those QTL with significant epistatic but insignificant main effects would remain undetected. A recent QTL study on grain yield components in rice identified a large number of QTL regions of this type (Li *et al.* 1997). However, the genome-wide search for epistatic effects among QTL employed by these authors is expected to aggravate the problems associated with model selection discussed earlier, because the number of regressor variables and multicollinearity among them increase tremendously. Thus, the need for validation with an independent sample is even more compelling for epistatic than for main effects of QTL.

**QTL  $\times$  environment interactions:** In Experiment 1, about one third of the detected QTL displayed significant QTL  $\times$  E interactions. The smallest fraction (one out of nine) was observed for GY, although estimates of  $\sigma_{ge}^2$  for this trait were highly significant and of the same magnitude as  $\hat{\sigma}_g^2$  (Table 1). For the other traits, the proportion of QTL with significant QTL  $\times$  E interactions was approximately proportional to the ratio  $\hat{\sigma}_{ge}^2:\hat{\sigma}_g^2$  and by far greatest for GM. Interestingly, reducing the number of test environments for PH in Experiment 1, from nine to four neither altered the number of detected QTL nor reduced the number of significant QTL  $\times$  E interactions (data not shown). In Experiment 2, the ratio  $\hat{\sigma}_{ge}^2:\hat{\sigma}_g^2$  was generally smaller than in Experi-

ment 1, which was reflected in fewer QTL showing significant QTL  $\times$  E interactions.

Most QTL studies reported in the literature (*e.g.*, Stuber *et al.* 1992; Ragot *et al.* 1995; Cockerham and Zeng 1996), including ours, found rarely significant QTL  $\times$  E interactions despite the presence of significant G  $\times$  E interactions at the phenotypic level. The reasons for this apparent discrepancy are not clear but two possible explanations include: (1) the detected (major) QTL display smaller QTL  $\times$  E interactions than the smaller undetected (minor) QTL (Tanksley 1993), and (2) the test procedure for detection of QTL  $\times$  E interactions is less powerful than that for detection of G  $\times$  E interactions. Our results did not support the first hypothesis because, among the QTL detected in Experiment 1, presence or absence of QTL  $\times$  E interactions was not associated with the magnitude of  $\hat{\alpha}$ -effects. However, the second hypothesis cannot be ruled out with the statistical analysis followed here because our search for QTL started with an analysis of means across environments, which favors the detection of QTL with large main effects over those with small main effects and large QTL  $\times$  E interactions. Only after all putative QTL had been mapped, we applied a combined analysis across environments for testing the presence of QTL  $\times$  E interactions. In contrast, the new method of multi-trait analysis devised by Jiang and Zeng (1995) starts QTL search with a likelihood ratio test for the presence of QTL activity in at least one environment and subsequently tests for the presence of QTL  $\times$  E interactions. It was not adopted in the present study because it assumes that environments are fixed, and therefore limits the scope of inference to the array of environments used in each experiment.

In general, varying the statistical analysis for CIM had only little impact on our findings. A comparison of our method used for QTL and QTL  $\times$  E analysis with that of Jiang and Zeng (1995) applied to GY, GM and PH showed that a larger number of QTL  $\times$  E interactions were detected with the latter approach, but results from both types of analyses hardly differed with respect to agreement of results between the two experiments. Likewise, only marginal deviations from original results were found when varying the strategy or threshold for cofactor selection in the QTL analysis. We infer from these findings that the particular choice of statistical analysis used for CIM has only little influence on the congruency between calibration and validation.

**Conclusions:** Identification of QTL affecting TC performance of agronomically important traits and accurate estimation of their genetic effects, including epistasis and QTL  $\times$  E interactions, are essential requirements for application of MAS in hybrid breeding of maize. Here, we used independent samples of TC progenies from the same population to (1) assess the magnitude of the bias of estimated QTL effects and (2) compare the power of QTL detection in samples of different size.

Our results suggest that inferences drawn from QTL mapping studies about the efficiency of MAS should be verified in an independent validation sample. When QTL effects are estimated from the same data as used for detection and mapping of QTL positions, they can be inflated due to statistical sampling and  $G \times E$  interactions. The relative magnitude of the bias can be substantial for sample sizes typically used in QTL mapping experiments ( $N < 200$ ) especially for traits with moderate heritability and a complex genetic architecture such as grain yield. As a consequence, the key factor determining the efficiency of MAS in comparison with classical phenotypic selection, the proportion,  $p$ , of the genotypic variance explained by QTL-marker associations, is overestimated. Moreover, if in this study the magnitude of estimated QTL effects had been used as a criterion for the choice of important QTL regions to be transferred by MAS or to be considered in a selection index, selection response would have been smaller than expected, because QTL effects estimated from calibration were biased.

With currently available statistical methods it was not possible to separate the effects of statistical sampling and QTL  $\times$  E interactions in this study, but we believe that at least the bias owing to sampling effects can be reduced by validation or cross-validation. For a correct assessment of the prospects of MAS as compared to classical phenotypic selection, more research efforts need to be dedicated to the analysis of the different factors leading to the inflation of QTL effect estimates.

The moderate agreement among the QTL detected in each sample provides evidence for a low power of QTL detection for most traits, especially GY. Only a small fraction of the detected QTL showed significant QTL  $\times$  E interactions for all traits except GM, suggesting that field testing of experimental materials could be limited to few environments known to provide good differentiation.

The consistency of QTL mapping results across testers largely reflected the genotypic correlations among testers and the predominant type of gene action for each trait. Thus, for a given sample, selection response from MAS for TC performance of traits with mainly additive gene action should be comparable for TC progenies with the tester used in QTL mapping and TC progenies with other unrelated testers. For all traits, we found little evidence for digenic epistasis among the detected QTL, particularly when reexamined in an independent sample. On the contrary, differences in the TC performance of  $F_3$  lines with each tester were due to the presence or absence of common QTL. This suggests that non-epistatic gene effects are major determinants of general and specific combining ability in hybrid performance, as was also concluded from numerous classic quantitative-genetic experiments (Hallauer and Miranda 1981, Chapter 8).

The RFLP data of this study were produced in the lab of R. G.

Herrmann, Ludwig-Maximilians-Universität, Munich, by E. Brunklaus-Jung and J. Boppenmaier. The contributions of D. Klein and W. Schmidt in conducting the field trials are gratefully acknowledged. We thank T. Lübberstedt, S. Groh and two anonymous reviewers for helpful suggestions and comments on the manuscript. The present study was part of EUREKA project 290. This research was supported by grants from the German Ministry of Research and Technology (BMFT) and KWS Kleinwanzlebener Saatzucht AG, grant 0319233A. This article is dedicated to F. W. Schnell on the occasion of his 85th birthday.

#### LITERATURE CITED

- Austin, D. F., and M. Lee, 1996 Comparative mapping in  $F_{2,3}$  and  $F_{6,7}$  generations of quantitative trait loci for grain yield and yield components in maize. *Theor. Appl. Genet.* **92**: 817–826.
- Beavis, W. D., 1994 The power and deceit of QTL experiments: Lessons from comparative QTL studies, pp. 250–266 in *49th Annual Corn and Sorghum Industry Research Conference*. American Seed Trade Association, Washington, DC.
- Beavis, W. D., O. S. Smith, D. Grant and R. Fincher, 1994 Identification of quantitative trait loci using a small sample of topcrossed and  $F_4$  progeny from maize. *Crop Sci.* **34**: 882–896.
- Berke, T. G., and T. R. Rocheford, 1995 Quantitative trait loci for flowering, plant and ear height, and kernel traits in maize. *Crop Sci.* **35**: 1542–1549.
- Bohn, M., M. M. Khairallah, D. Gonzalez-De-Leon, D. A. Hoisington, H. F. Utz *et al.*, 1996 QTL mapping in tropical maize: I. Genomic regions affecting leaf feeding resistance to sugarcane borer and other traits. *Crop Sci.* **36**: 1352–1361.
- Cheverud, J. M., E. J. Routman, F. A. M. Duarte, B. Van Swinderen, K. Cothran *et al.*, 1996 Quantitative trait loci for murine growth. *Genetics* **142**: 1305–1319.
- Cockerham, C. C., and Z.-B. Zeng, 1996 Design III with marker loci. *Genetics* **143**: 1437–1456.
- Dudley, J. W., 1993 Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. *Crop Sci.* **33**: 660–668.
- Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics*, 4th Ed., Longman Group, London.
- Freedman, D. A., 1983 A note on screening regression equations. *Am. Statist.* **37**: 152–155.
- Gallais, A., and M. Rives, 1993 Detection, number and effects of QTLs for a complex character. *Agronomie* **13**: 723–738.
- Georges, M., D. Nielsen, M. Mackinnon, A. Mishra, R. Okimoto *et al.*, 1995 Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* **139**: 907–920.
- Goldman, I. L., T. R. Rocheford and J. W. Dudley, 1993 Quantitative trait loci influencing protein and starch concentration in the Illinois long term selection maize strains. *Theor. Appl. Genet.* **87**: 217–224.
- Haley, C. S., and S. A. Knott, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**: 315–324.
- Hallauer, A. R., and J. B. Miranda, 1981 *Quantitative Genetics in Maize Breeding*. Iowa State University Press, Ames, IA.
- Holm, S., 1979 A simple sequentially rejective multiple test procedure. *Scand. J. Statist.* **6**: 65–70.
- Jansen, R. C., and P. Stam, 1994 High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**: 1447–1455.
- Jiang, C., and Z.-B. Zeng, 1995 Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* **140**: 1111–1127.
- Jobson, J. D., 1991 *Applied Multivariate Data Analysis*, Vol. I, Springer-Verlag, New York.
- Kendall, M. G., and A. Stuart, 1961 *The Advanced Theory of Statistics*, Vol. II, Griffin Press, London.
- Knapp, S. J., W. W. Stroup and W. M. Ross, 1985 Exact confidence intervals for heritability on a progeny mean basis. *Crop Sci.* **25**: 192–194.
- Lamkey, K. R., B. J. Schnicker and A. E. Melchinger, 1995 Epista-

- sis in an elite maize hybrid and choice of generation for inbred line development. *Crop Sci.* **35**: 1272–1281.
- Lande, R., and R. Thompson, 1990 Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* **124**: 743–756.
- Lander, E. S., and D. Botstein, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.
- Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly *et al.*, 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**: 174–181.
- Lee, M., 1995 DNA markers and plant breeding programs. *Adv. Agron.* **55**: 265–344.
- Li, Z., S. R. M. Pinson, W. D. Park, A. H. Paterson and J. W. Stansel, 1997 Epistasis for three grain yield components in rice (*Oryza sativa* L.). *Genetics* **145**: 453–465.
- Lübberstedt, T., A. E. Melchinger, C. C. Schön, H. F. Utz and D. Klein, 1997 QTL mapping in testcrosses of European flint lines of maize: I. Comparison of different testers for forage yield traits. *Crop Sci.* **37**: 921–931.
- Lynch, M., and B. Walsh, 1997 *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.
- Melchinger, A. E., 1987 Expectation of means and variances of testcrosses produced from  $F_2$  and backcross individuals and their selfed progenies. *Heredity* **59**: 105–115.
- Melchinger, A. E., 1988 Means, variances, and covariances between relatives in hybrid populations with disequilibrium in the parent population, pp. 400–415 in *Proceedings of the 2nd International Conference on Quantitative Genetics*, edited by B. S. Weir, E. J. Eisen, M. M. Goodman and G. Namkoong. Sinauer Associates, Sunderland, MA.
- Melchinger, A. E., 1998 Advances in the analysis of data on quantitative trait loci. *Proceedings of the 2nd International Crop Science Congress*, New Delhi, India (in press).
- Melchinger, A. E., G. A. Schmidt and H. H. Geiger, 1986 Evaluation of near infrared reflectance spectroscopy for predicting grain and stover quality traits in maize. *Plant Breed.* **97**: 20–29.
- Melchinger, A. E., W. Schmidt and H. H. Geiger, 1988 Comparison of testcrosses produced from  $F_2$  and first backcross populations in maize. *Crop Sci.* **28**: 743–749.
- Miller, A. J., 1990 *Subset Selection in Regression*. Chapman and Hall, London.
- Mode, C. J., and H. F. Robinson, 1959 Pleiotropism and the genetic variance and covariance. *Biometrics* **15**: 518–537.
- Patterson, H. D., and E. R. Williams, 1976 A new class of resolvable incomplete block designs. *Biometrika* **63**: 83–92.
- Ragot, M., P. H. Sisco, D. A. Hoisington and C. W. Stuber, 1995 Molecular-marker-mediated characterization of favorable exotic alleles at quantitative trait loci in maize. *Crop Sci.* **35**: 1306–1315.
- Rao, C. R., 1973 *Linear Statistical Inference and Its Applications*, 2nd Ed., John Wiley & Sons, New York.
- Rencher, A. C., and F. C. Pun, 1980 Inflation of  $R^2$  in best subset regression. *Technometrics* **22**: 49–53.
- Satterthwaite, F. E., 1946 An approximate distribution of estimates of variance components. *Biometrics* **2**: 110–114.
- Sax, K., 1923 The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* **8**: 552–560.
- Schön, C. C., A. E. Melchinger, J. Boppenmaier, E. Brunklaus-Jung, R. G. Herrmann *et al.*, 1994 RFLP mapping in maize: Quantitative trait loci affecting testcross performance of elite European flint lines. *Crop Sci.* **34**: 378–389.
- Searle, S. R., 1971 *Linear Models*. John Wiley & Sons, New York.
- Stuber, C. W., S. E. Lincoln, D. W. Wolff, T. Helentjaris and E. S. Lander, 1992 Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics* **132**: 823–839.
- Tanksley, S. D., 1993 Mapping polygenes. *Annu. Rev. Genet.* **27**: 205–233.
- Utz, H. F., and A. E. Melchinger, 1994 Comparison of different approaches to interval mapping of quantitative trait loci, pp. 195–204 in *Biometrics in Plant Breeding: Applications of Molecular Markers, Proceedings of the Ninth Meeting of the EUCARPIA Section Biometrics in Plant Breeding*, edited by J. W. Van Ooijen and J. Jansen. CPRO-DLO, Wageningen, Netherlands.
- Utz, H. F., and A. E. Melchinger, 1996 PLABQTL: A program for composite interval mapping of QTL. *J. Quant. Trait Loci* **2**(1).
- Van Ooijen, J. W., 1992 Accuracy of mapping quantitative trait loci in autogamous species. *Theor. Appl. Genet.* **84**: 803–811.
- Visscher, P. M., R. Thompson and C. S. Haley, 1996 Confidence intervals in QTL mapping by bootstrapping. *Genetics* **143**: 1013–1020.
- Whittaker, J. C., R. Thompson and P. M. Visscher, 1996 On the mapping of QTL by regression of phenotype on marker-type. *Heredity* **77**: 23–32.
- Wright, S., 1968 *Evolution and the Genetics of Populations, Vol. I: Genetic and Biometric Foundations*. University of Chicago Press, Chicago, IL.
- Xu, S., 1995 A comment on the simple regression method for interval mapping. *Genetics* **141**: 1657–1659.
- Zeng, Z.-B., 1994 Precision mapping of quantitative trait loci. *Genetics* **136**: 1457–1468.

Communicating editor: Z.-B. Zeng

#### APPENDIX A: RELATIONSHIP BETWEEN $R^2$ , $\mathfrak{R}^2$ , AND THE LOD SCORE

According to Haley and Knott (1992, p. 317) and Searle (1971, p. 125), the likelihood ratio test (LR) for presence of a putative QTL in composite interval mapping (Zeng 1994) can be written in terms of the residual sum of squares of the full model (fitting the cofactors plus the  $\alpha$ -effect of the putative QTL in Equation 1) and the reduced model (fitting only the cofactors), and the number of observations (SSE,  $SSE_R$ , and  $N$ , respectively)

$$LR = N \ln \left( \frac{SSE_R}{SSE} \right). \quad (A1)$$

This can be expressed in terms of the coefficient of multiple correlation for the full model ( $R^2$ ) and reduced model ( $R_R^2$ )

$$LR = N \ln \left( \frac{1 - R_R^2}{1 - R^2} \right) = -N \ln \left( \frac{1 - R^2}{1 - R_R^2} \right). \quad (A2)$$

Let  $\mathfrak{R}^2$  denote the square of the partial correlation coefficient between  $y$  and  $x_i$  (the  $i$ th QTL) controlling for the cofactors ( $x_1, x_2, \dots, x_k$ ), then we have according to Rao (1973, p. 268)

$$\mathfrak{R}^2 = \frac{R^2 - R_R^2}{1 - R_R^2} = 1 - \frac{1 - R^2}{1 - R_R^2}. \quad (A3)$$

Inserting Equation A3 into Equation A2, we obtain

$$LOD = \frac{1}{2 \ln 10} LR = -0.21715 N \ln (1 - \mathfrak{R}^2) \quad (A4)$$

or

$$\mathfrak{R}^2 = 1 - e^{-LOD/(0.21715 * N)}. \quad (A5)$$

For simple interval mapping (no cofactors  $x_k$  in Equation 1, *i.e.*,  $k = 0$ ), Equation A4 coincides with the following result given by Lynch and Walsh (1997):

$$LOD = -0.21715 N \ln (1 - R^2), \quad (A6)$$

from which we obtain

$$R^2 = 1 - e^{-\text{LOD}/(0.21715 * N)}. \quad (\text{A7})$$

#### APPENDIX B: AVERAGE EFFECT OF ALLELE SUBSTITUTION IN TESTCROSSES

For a given QTL and tester Tz, the average effect of substituting the allele from parent P1 by the allele from parent P2 can be expressed as (Melchinger 1988)

$$\alpha_{Tz} = a + d_{2Tz} - d_{1Tz}. \quad (\text{B1})$$

Here,  $a$  is the additive effect of alleles  $A_1$  and  $A_2$  in

parents P1 and P2, respectively, *i.e.*,  $a = (A_2 A_2 - A_1 A_1)/2$ , and  $d_{iTz} = A_i A_{Tz} - (A_i A_i + A_{Tz} A_{Tz})/2$  is the dominance effect between allele  $A_i$  from parent  $P_i$  ( $i = 1, 2$ ) and allele  $A_{Tz}$  from tester Tz. Hence, we have

$$\begin{aligned} \alpha_{T1} = \alpha_{T2} \quad &\text{if and only if} \\ d_{2T1} - d_{1T1} = d_{2T2} - d_{1T2}. \end{aligned} \quad (\text{B2})$$

Furthermore, for the comparison of QTL effects determined from evaluation of TC performance with tester Tz and line *per se* performance, we have

$$\alpha_{Tz} = a \quad \text{if and only if} \quad d_{1Tz} = d_{2Tz}. \quad (\text{B3})$$