Bias and Sampling Error of the Estimated Proportion of Genotypic Variance Explained by Quantitative Trait Loci Determined From Experimental Data in Maize Using Cross Validation and Validation With Independent Samples

H. Friedrich Utz,* Albrecht E. Melchinger* 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 April 27, 1999
Accepted for publication December 13, 1999

ABSTRACT

Cross validation (CV) was used to analyze the effects of different environments and different genotypic samples on estimates of the proportion of genotypic variance explained by QTL (p). Testcrosses of 344 F_3 maize lines grown in four environments were evaluated for a number of agronomic traits. In each of 200 replicated CV runs, this data set was subdivided into an estimation set (ES) and various test sets (TS). ES were used to map QTL and estimate p for each run ($\hat{p}_{ES}$) and its median ($\tilde{p}_{ES}$) across all runs. The bias of these estimates was assessed by comparison with the median ($\tilde{p}_{TS.ES}$) obtained from TS. We also used two independent validation samples derived from the same cross for further comparison. The median $\tilde{p}_{ES}$ showed a large upward bias compared to $\tilde{p}_{TS.ES}$. Environmental sampling generally had a smaller effect on the bias of $\tilde{p}_{ES}$ than genotypic sampling or both factors simultaneously. In independent validation, $\tilde{p}_{TS.ES}$ was on average only 50% of $\tilde{p}_{ES}$. A wide range among $\hat{p}_{ES}$ reflected a large sampling error of these estimates. QTL frequency distributions and comparison of estimated QTL effects indicated a low precision of QTL localization and an upward bias in the absolute values of estimated QTL effects from ES. CV with data from three QTL studies reported in the literature yielded similar results as those obtained with maize testcrosses. We therefore recommend CV for obtaining asymptotically unbiased estimates of p and consequently a realistic assessment of the prospects of MAS.

MOLECULAR markers are used by a great number of researchers to study quantitative traits of agronomic importance. The primary objective of these studies has been the identification of markers associated with quantitative trait loci (QTL) and their use in subsequent marker-assisted selection (MAS) programs.

In the statistical analysis of quantitatively inherited traits, the introduction of QTL interval mapping and maximum-likelihood estimation of effects by Lander and Botstein (1989) was a landmark. Simplicity and speed of QTL analyses were further increased by using multiple regression for determining significance of putative QTL and estimation of their genetic effects (Haley and Knott 1992; Martinez and Curnow 1992). An increase in power of QTL detection as well as accuracy and precision of estimated QTL positions and effects can be accomplished by including additional markers as cofactors in the statistical model (Jansen 1993; Zeng 1994). For review of the various statistical methods for QTL analyses, see Liu (1998).

Identification of significant QTL-marker associations forms the baseline for MAS. To be superior to classical phenotypic selection, several prerequisites must be satisfied: (1) QTL positions are estimated with high precision to choose markers showing a minimum of recombination with the QTL and to resolve linked QTL; (2) estimated QTL effects reflect their true genetic effects and, therefore, are estimated without bias due to genotypic or environmental sampling; (3) a sufficient proportion of the genotypic variance of the trait under study is explained by the detected QTL.

With respect to these prerequisites, the available statistical methods still have considerable shortcomings. Using computer simulations it was shown that (a) estimates of individual QTL effects and the proportion of genotypic variance explained by QTL can be severely inflated, leading to an overly optimistic assessment of the prospects of MAS (Utz and Melchinger 1994; Georges *et al.* 1995; Beavis 1998); and (b) confidence intervals for QTL positions are large for population sizes commonly used in QTL mapping experiments (van Ooijen 1992; Visscher *et al.* 1996).

In most experimental studies these limitations have been ignored even though with experimental data, the bias in QTL effects is expected to be even greater than in computer simulations that rely on simplifying assumptions. To overcome these pitfalls, Lande and Thompson (1990) suggested identifying QTL-marker associations in one data set and subsequently estimating genetic effects based on the *a priori* model in another, independent validation sample. However, owing to the

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

high costs of QTL studies their suggestion has not become common practice.

In an earlier study, we demonstrated for experimental data in maize that the magnitude of QTL effects and the proportion of the phenotypic and genotypic variance explained by QTL decreased substantially when estimated in independent validation samples (Melchinger *et al.* 1998). Because the samples used for mapping of QTL positions were evaluated in different environments than those for estimation of genetic effects, the observed decrease in variance explained by QTL had to be attributed to two confounded factors: environmental and genotypic sampling. In basic studies as well as in practical breeding, however, the contribution of each factor to the bias must be known for an optimal allocation of limited resources.

Several authors (*e.g.*, Beavis 1994; Visscher *et al.* 1996) recommended the use of resampling methods to determine the magnitude of bias caused by these factors and to get more realistic estimates of the amount of genotypic variance explained by QTL. In this study, we used cross validation (CV; Hjorth 1994) to elucidate the effects of environmental and genotypic sampling. Objectives of our research were to (1) obtain unbiased estimates of the proportion of the genotypic variance explained by all detected QTL (p); (2) analyze the influence of environmental and genotypic sampling on the magnitude of the bias and sampling error of estimates of p; and (3) compare the magnitude of the bias and sampling error of p determined by CV with results obtained with independent validation samples.

MATERIALS AND METHODS

Plant materials: The plant materials used for this study were partly identical to those described by Melchinger *et al.* (1998). 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 × P2 were selfed to produce 507 independently derived F_3 ($F_{2:3}$) lines. A subsample of these F_3 lines was advanced by single-seed descent to the F_4 generation to produce 71 independent F_5 ($F_{4:5}$) lines. Testcross (TC) seed was produced in isolation plots by mating the unrelated inbred tester KW5361 (T2 in Melchinger *et al.* 1998) to a random sample of 40 plants from each of the 507 F_3 lines, the 71 F_5 lines, as well as parents P1 and P2.

Field experiments: The TC progenies were evaluated in three different experiments. Experiment 1 (Exp. 1) comprised 380 TC of F_3 lines, TC of P1 and P2 included as quintuple entries, and 10 common check hybrids. Trials were conducted in 1990 and 1991 at two sites in Germany. 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 (Exp. 2) comprised TC of an independent 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 Exp.1. Trials were grown in 1992 and 1993 at two sites in Germany.

Experiment 3 (Exp. 3) comprised TC of 71 F_5 lines derived from 71 F_3 lines, whose TC were grown in Exp. 2, TC of P1 and P2 included as multiple entries, and the same check hybrids as in Exp. 1 and Exp. 2. Exp. 3 was grown in 1992 at four sites in Germany, two of which were in common with Exp. 2, and one additional site in France.

In all three experiments the experimental design was a generalized lattice with two replications (Patterson and Williams 1976). Two-row plots were overplanted and later thinned to reach a final stand of 8–11 plants m^{-2} depending on the location. All experiments were machine planted and harvested as grain trials with a combine.

Data were analyzed 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, and plant height (PH) measured in centimeters on a plot basis as the distance from the soil level to the lowest tassel branch.

RFLP marker genotyping and linkage map construction: The procedures for RFLP assays were described by Schön *et al.* (1994). Two linkage maps were constructed based on 89 RFLP marker loci using a subset of 344 parental F_2 plants of the 380 F_3 lines employed in Exp. 1 and a subset of 107 parental F_2 plants of the 127 F_3 lines employed in Exp. 2, respectively. A third linkage map was obtained from the 71 parental F_4 plants of the F_5 lines tested in Exp. 3, genotyped with 84 of the 89 RFLP marker loci used for the other two linkage maps. Software packages MAPMAKER 3.0 (Lander *et al.* 1987) and GMENDEL 3.0 (Holloway and Knapp 1993) were used for map construction.

Agronomic data analyses: For each experiment adjusted entry means and effective error mean squares derived from analyses of variance of each site-year combination were used to compute the combined analyses of variance across environments. For estimation of quantitative genetic parameters such as variance components and heritabilities, see Melchinger *et al.* (1998). Phenotypic ($\hat{r}_p$) and genotypic ($\hat{r}_g$) correlations between means of related TC progenies from Exp. 2 and Exp. 3 were calculated following standard procedures (Mode and Robinson 1959).

QTL analyses: QTL mapping and estimation of their effects were performed with PLABQTL (Utz and Melchinger 1996) employing composite interval mapping (CIM) by the regression approach (Haley and Knott 1992) in combination with the use of cofactors (Jansen and Stam 1994; Zeng 1994). Following Cowen (1988), an additive genetic model is appropriate for the analysis of TC progenies, because TC progenies from F_2 plants heterozygous at a given marker or QTL correspond to a 1:1 mixture of TC plants carrying alleles from P1 and P2. Accordingly, the underlying model for TC progenies can be written as

$$Y_j = \mu_{P1} + \alpha_\ell x^*_{j\ell} + \sum_m b_m x_{jm} + \varepsilon_j. \qquad (1)$$

Here, Y_j denotes the mean phenotypic trait value of the TC progeny of line j averaged across environments; μ_{P1} is the mean phenotypic trait value of TC progeny carrying the allele from P1 at the ℓth QTL; α_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 with flanking markers ℓ' and ℓ'' (subsequently denoted additive effect); $x^*_{j\ell}$ is the conditional expectation of the dummy variable $\theta_{j\ell}$ given the observed genotypes at flanking marker loci ℓ' and ℓ'', where $\theta_{j\ell}$ assumes values 0, 0.5, or 1, if the genotype of the F_2 individual j at the putative QTL is qq, Qq, or QQ, respectively; b_m is the partial regression coefficient of phenotype Y_j on the mth (selected) marker; x_{jm} 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 m is homozygous P1, heterozygous, or homozygous P2, respectively; and ε_j is a residual variable for the TC progeny of the jth F_3 line.

The selection of cofactors was described by Melchinger *et al.* (1998). Testing for presence of a putative QTL in an interval by a likelihood-ratio test was performed using a 2.5 LOD threshold. Estimates of QTL positions were obtained at the position where the LOD score assumed its maximum in the region under consideration. Following Draper and Smith (1981), the proportion of the phenotypic variance explained by QTL was determined by the unbiased estimator

$$R^2_{\text{adj}} = R^2 - \left(\frac{z}{N - z - 1}\right)(1 - R^2), \tag{2}$$

where R^2 is the coefficient of determination of regression fitting a model including z predictors (number of QTL positions and effects) and N is the number of phenotypic observations used in multiple regression. When R^2 is zero or small, R^2_{adj} can become negative. In our calculations negative values of R^2_{adj} were allowed, because when imposing a lower bound R^2_{adj} would no longer be unbiased (Kendall and Stuart 1961).

The proportion of the genotypic variance explained by all detected QTL was estimated from the ratio

$$\hat{p} = \frac{R^2_{\text{adj}}}{\hat{h}^2}, \tag{3}$$

where $\hat{h}^2$ is the heritability of the respective trait calculated on an entry-mean basis (Hallauer and Miranda 1981),

$$\hat{h}^2 = \frac{\hat{\sigma}^2_g}{\hat{\sigma}^2/re + \hat{\sigma}^2_{ge}/e + \hat{\sigma}^2_g}, \tag{4}$$

with σ^2 denoting the effective error variance, σ^2_{ge} the G × E interaction variance, σ^2_g the genotypic variance, r the number of replications, and e the number of test environments. All variance components were estimated from Exp. 1 unless stated otherwise.

Cross validation: One approach applied for evaluation of QTL mapping results was cross validation. Here, the entire data set (DS) is split into subsets. One or several subsets combined form the estimation set (ES) for QTL detection, localization, and estimation of genetic effects. The remaining subset(s) form the test set (TS) in which predictions derived from ES are tested for their validity by correlating predicted and observed data. For example, in fivefold CV, the DS comprising marker data from 344 F_2 plants and phenotypic data of their TC progenies from Exp. 1 was randomly subdivided into k = 5 genotypic subsamples, 4 with 69, 1 with 68 genotypes (Figure 1). Each of the 5 genotypic subsamples was evaluated in four environments, and consequently DS was divided into 20 disconnected subsets. For testing the effect of (i) environmental sampling (CV/E), (ii) genotypic sampling (CV/G), and (iii) both factors simultaneously (CV/GE), the same ES but different TS were used. The ES consisted of four genotypic subsamples with phenotypic data from three of the four environments. In CV/E, the TS consisted of the same four genotypic subsamples as in ES, but phenotypic data came from the fourth, disconnected environment. In CV/G, the fifth disconnected genotypic subsample with phenotypic data from the same three environments as in ES was used as TS. In CV/GE, estimates of QTL effects were obtained by using the one subset not connected with ES, either by environment or by genotypes, as TS. Consequently, by permutating the respective k subsets used for ES and TS, $4 \times k = 20$ different CV runs are possible for fivefold cross validation. To increase the precision of estimates of p, additional CV runs were generated by using 10 different randomizations for assigning genotypes to the respective subsamples, yielding a total of $10 \times 4 \times k = 200$ replicated CV runs.

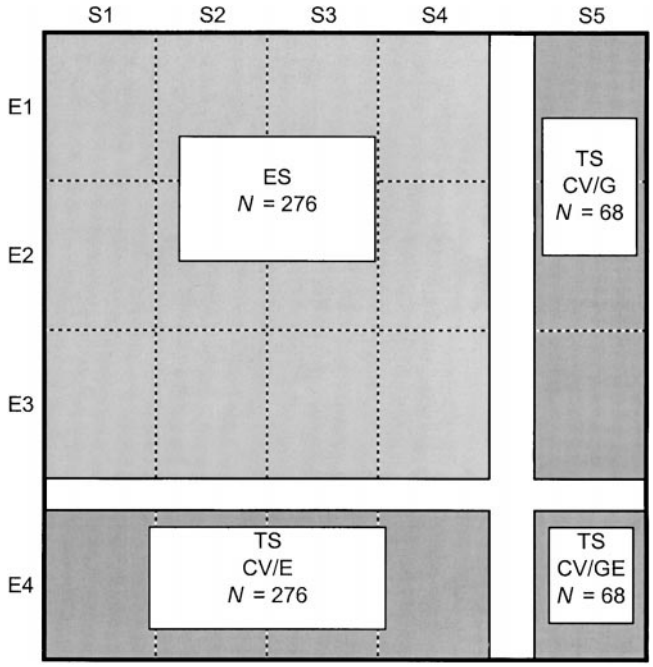

Figure 1.—Subdivision of the data set used for cross validation into five subsamples (S1–S5) evaluated in four environments (E1–E4). Testcross (TC) entry means from S1–S4 averaged across E1–E3 serve as estimation set (ES). TC entry means from S5 averaged across E1–E3 are used as test set (TS) accounting for genotypic sampling (CV/G). TC entry means from S1–S4 evaluated in E4 are used as TS accounting for environmental sampling (CV/E). TC entry means from S5 evaluated in E4 are used as TS for obtaining asymptotically unbiased estimates of the proportion of genotypic variance explained by QTL (CV/GE).

The effect of sample size in ES and TS was tested by varying the number of genotypic and the number of environmental subsamples used for estimation and testing. Genotypic subsampling was tested in five different CV schemes, dividing DS into k (k = 2, 3, 5, 9, and 16) genotypic subsamples containing N/k TC progenies each (Table 1). An additional cross validation scheme ($CV_{k=1}$) was created by randomly subdividing DS into subsamples of size $N_{ES} = 100$ and $N_{TS} = 244$. For all CV schemes, estimates of p were obtained as the median from a minimum of 200 CV runs originating from an appropriate number of different randomizations (Table 1).

For plant height, which was evaluated in nine environments, additional CV schemes were analyzed by varying the number of environments included in ES (u = 1, 2, 3, 4, and 8) for k = 5 (Table 1). The corresponding TS were based on TC progeny means averaged across the remaining $e = 9 - u$ disconnected environments. The number of possible CV runs obtained by a single randomization was $5\binom{9}{u}$ for the respective CV schemes (Table 1).

The magnitude of bias in estimates of genotypic variance explained by QTL due to genotypic and/or environmental sampling was obtained by comparing estimates of p obtained from the ES and TS. Based on QTL mapping results obtained with composite interval mapping (CIM) from the ES, the genotypic value of F_3 line j in TS $Q_{TS.ESj}$ can be predicted according to

$$Q_{TS.ESj} = \mathbf{X}^*_{\mathbf{TS}j}\,\hat{\boldsymbol{\beta}}^*_{\mathbf{ES}}, \tag{5}$$

TABLE 1

Description of cross validation schemes varying for the number of genotypic subsamples and the number of environments used for estimation sets and test sets

Genotypic subsamples[a]					Environmental samples[b]			
No. of subsamples k	Sample size N in		No. of		No. of environments in		No. of	
	ES	TS_{CV}	Randomizations	Replicated runs	ES (u)	TS (e)	Randomizations	Replicated runs
1	100	244	50	200	1	8	4	180
2	172	172	25	200	2	7	1	180
3	230	114	17	204	3	6	1	420
5	276	68	10	200	4	5	1	630
9	306	38	6	216	8	1	4	180
16	323	21	3	256				

CV, cross validation; ES, estimation set; TS, test set. Independent validation: (1) TS_{VS1}, $N = 107$, $e = 4$; (2) TS_{VS2}, $N = 71$, $e = 5$.

[a] Further assumptions: $u = 3$, $e = 1$.

[b] Further assumptions: $k = 5$; ES, $N = 276$; TS_{CV},$N = 68$.

with $\mathbf{X}^*_{TS_j}$ being the vector of conditional expectations of the dummy variable $\theta_{j\ell}$ given the observed genotypes at the flanking marker loci ℓ' and ℓ'' and a significant QTL (LOD > 2.5) in the ℓth marker interval in ES. For $\theta_{j\ell}$ values, 0, 0.5, or 1 are assumed, if the genotype of the corresponding F_2 plant j at the QTL significant in ES is qq, Qq, or QQ, respectively; $\hat{\boldsymbol{\beta}}^*_{ES}$ is the vector of genetic effects of all significant QTL detected in ES, estimated as partial regression coefficients from a simultaneous fit in ES.

The proportion of the genotypic variance explained by QTL in TS ($\hat{p}_{TS.ES}$) is calculated from the adjusted squared correlation coefficient (R^2_{adj}, see Equation 2) between the phenotypic means observed in TS ($\mathbf{Y}_{TS}$) and the predicted genotypic values $\mathbf{Q}_{TS.ES}$ on the basis of results derived from ES, divided by the heritability of the trait (see Equation 4) under study for the respective value of e:

$$\hat{p}_{TS.ES} = \frac{R^2_{adj}(\mathbf{Y}_{TS}, \mathbf{Q}_{TS.ES})}{\hat{h}^2}. \qquad (6)$$

The estimate $\hat{p}_{TS.ES}$ is asymptotically unbiased in the case of CV/GE because the data in $TS_{CV/GE}$ are independent from the data in ES from which the *a priori* model of prediction is determined (Lande and Thompson 1990). Estimates $\hat{p}_{TS.ES}$ calculated for CV/E and CV/G are still biased by genotypic and environmental sampling, respectively, because TS are not independent from ES.

Using a LOD threshold of 2.5 each CV run yielded different estimates for the number of QTL, their location, and genetic effects in ES. Estimates of p in ES and TS were calculated as the median $\tilde{p}$ over all replicated CV runs. For CV runs with no QTL detected in ES, $\hat{p}$ was assumed to be zero. The average number of QTL was determined as the mean across replicated CV runs.

A more detailed analysis was performed for putative QTL for GY and PH on chromosome 7. Precision of QTL localization was assessed by determining the relative frequency of detected QTL for 1376 replicated runs in 1-cM intervals along chromosome 7 from ES with $k = 5$, $u = 3$, and $e = 1$. In ES, allele substitution effects ($\hat{\alpha}_{ES}$) were estimated from a simultaneous fit of all significant QTL in each of the 1376 CV runs. The median allele substitution effect $\tilde{\alpha}_{ES}$ was calculated for each position along the chromosome. For each $\hat{\alpha}_{ES}$, the corresponding allele substitution effect from $TS_{CV/GE}$ ($\hat{\alpha}_{TS}$) was determined by multiple regression based on (a) the map positions of all QTL detected in ES and (b) the marker genotypes at the flanking markers of the F_2 plants in TS according to described procedures (Haley and Knott 1992; Utz and Melchinger 1996). Subsequently, the median $\tilde{\alpha}_{TS}$ was calculated across all CV runs.

Validation with independent samples: Statistical theory and procedures for the alternative approach of testing results obtained in ES with independent validation samples are equivalent to cross validation. The same estimation sets as in CV were used to predict genotypic values $\mathbf{Q}_{VS.ES}$ for validation sets (VS). The adjusted squared correlation coefficient (R^2_{adj}) between $\mathbf{Q}_{VS.ES}$ and the entry means ($\mathbf{Y}_{VS}$) from VS1 ($N = 107$) and VS2 ($N = 71$), divided by the heritabilities $\hat{h}^2_{VS1}$ and $\hat{h}^2_{VS2}$, served as an unbiased estimate of the genotypic variance explained by putative QTL in validation test sets. Variance components for the calculation of $\hat{h}^2_{VS1}$ and $\hat{h}^2_{VS2}$ were estimated from Exp. 2 and Exp. 3, respectively. Estimates of p were calculated as the median of replicated validation runs, the number of replications corresponding to the number of ES for the respective factor k (Table 1). For validation with VS2, calculations of conditional expectations of the genotype of F_5 lines at the putative QTL given flanking marker genotypes were adjusted to parental F_4 instead of F_2 plants.

RESULTS

Trait means, variances, and heritabilities: Quantitative genetic parameters for Exp. 1 and Exp. 2 were presented in detail by Melchinger *et al.* (1998). Genotypic variances among TC of F_3 and F_5 lines were significant for all traits in all three experiments. As anticipated from theory, genotypic variances among F_5 lines in Exp. 3 were greater than those among F_3 lines in Exp. 1 and Exp. 2 for all traits. Estimates of σ^2_{ge} were significantly greater than zero ($P < 0.01$) for all traits in Exp. 1 and Exp. 3. In Exp. 2 significant G × E interactions were found only for GM and KW but not for GY and PH. Heritabilities exceeded 0.70 for all traits except GY in Exp. 1 ($\hat{h}^2 = 0.48$) and were highest in Exp. 3 (0.70–0.92)

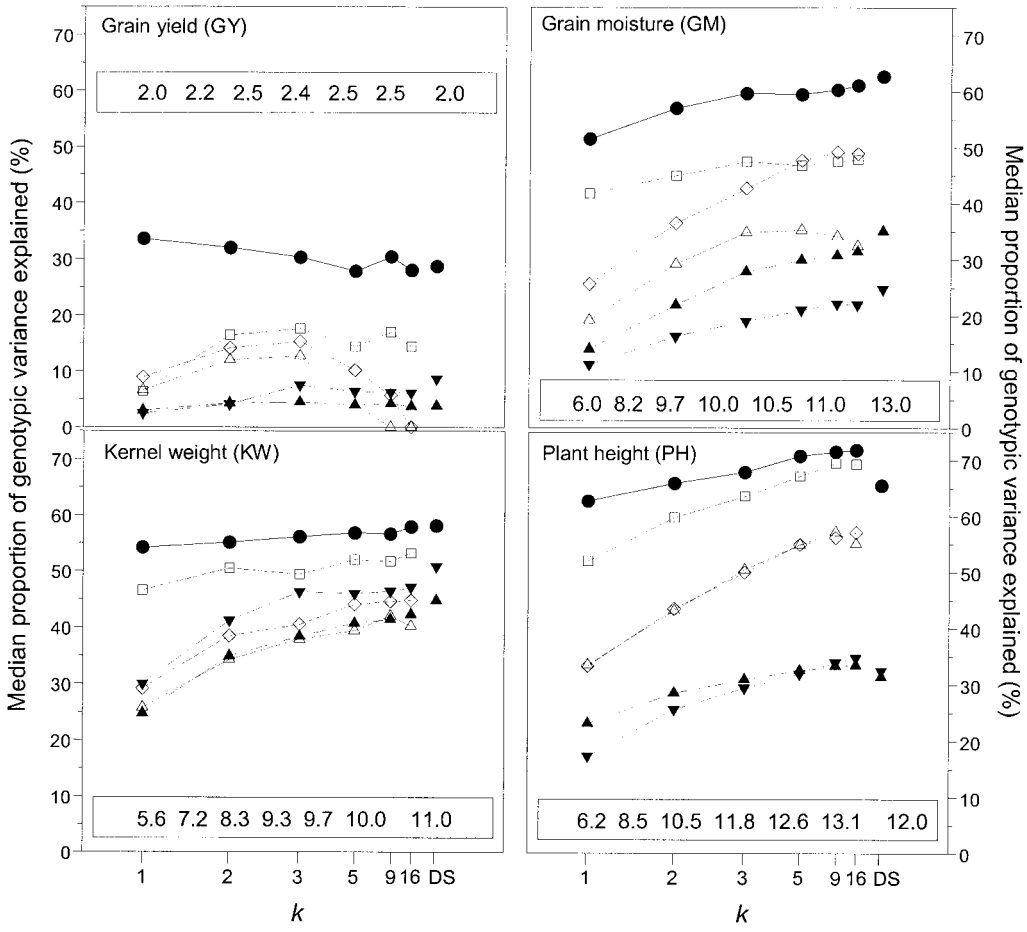

Figure 2.—Median proportion (percentage) of genotypic variance explained by detected QTL ($\tilde{p}$) determined from estimation (ES), cross validation (CV/E, CV/G, CV/GE), and validation with independent samples (VS1 and VS2) as well as the average number of QTL detected in estimation (see box, ascending order of k) for a varying number of genotypic subsamples (k) and the entire data set (DS) of Experiment 1 for grain yield, grain moisture, kernel weight, and plant height. (●) ES, (□) CV/E, (◇) CV/G, (△) CV/GE, (▲) V/VS1, (▼) V/VS2.

owing to larger genotypic variances and an additional test environment. Phenotypic ($\hat{r}_p$) and genotypic ($\hat{r}_g$) correlations between related TC progenies from early (F_3 lines) and advanced (F_5 lines) selfing generations ranged between 0.32 and 0.44 except for $\hat{r}_g = 0.62$ for GY.

QTL analyses: For detailed results from QTL analyses based on the entire DS (Exp. 1) see Melchinger *et al.* (1998). Briefly, the number of detected QTL in DS was 2 for GY, 13 for GM, 11 for KW, and 12 for PH, explaining between 28.7 and 65.5% of the genotypic variance (Figure 2).

For all traits but GY, the average number of QTL detected increased with increasing sample size N in ES and was almost twice as large for $k = 16$ ($N = 323$) as compared to $k = 1$ ($N = 100$; Figure 2). The median $\tilde{p}_{ES}$ increased only slightly (≤20%) with increasing N and even decreased for GY. In all three cross validation schemes (CV/E, CV/G, and CV/GE), $\tilde{p}_{TS.ES}$ was substantially reduced as compared to $\tilde{p}_{ES}$ for all values of k. The largest reduction in $\tilde{p}_{TS.ES}$ was found for GY in CV/GE for $k = 9$ and $k = 16$. In most cases, CV/G resulted in lower values for $\tilde{p}_{TS.ES}$ than CV/E. Except for GY, the difference between the two CV schemes was most pronounced for small k. For GM, $\tilde{p}_{TS.ES}$ was slightly larger for CV/G than for CV/E in half of the CV schemes ($k = 5$, 9, and 16). In cross validation CV/GE generally resulted in the smallest values for $\tilde{p}_{TS.ES}$. For PH however, values for $\tilde{p}_{TS.ES}$ from CV/GE were slightly greater than those for CV/G in some CV schemes.

Validation with the two independent samples VS1 and VS2 resulted in the lowest $\tilde{p}_{TS.ES}$ values, except for KW (Figure 2). On average only 50% of $\tilde{p}_{ES}$ could be confirmed. For all traits except PH, estimates of $\tilde{p}_{TS.ES}$ from VS1 and VS2 were comparable to those from CV/GE. For PH, $\tilde{p}_{TS.ES}$ for VS1 and VS2 was surprisingly small as compared to CV schemes, probably due to genotypic and environmental sampling in the validation experiments. Contrary to expectation, V/VS1 yielded smaller values for $\tilde{p}_{TS.ES}$ than V/VS2 for GY and KW.

For GM, KW, and PH, 95% confidence intervals for $\tilde{p}_{ES}$ span 1–3% for all k except for $k = 1$ with ~6% (data not shown). For GY, 95% confidence intervals for $\tilde{p}_{ES}$ ranged from 5 to 8%. In CV, 95% confidence intervals for $\tilde{p}_{TS.ES}$ were of similar size as for corresponding estimates of $\tilde{p}_{ES}$, but were quite large for CV/GE and $k = 16$ with a maximum of 13% for CV/GE with $k = 16$ of PH. Figure 3 shows the range in the number of QTL detected in ES and the variation in $\hat{p}_{ES}$ and $\hat{p}_{TS.ES}$ among the different CV runs for $k = 5$. For GY, the number of QTL detected in the 200 estimation runs varied from 0 to 8. In 11 of the 200 ES no QTL for GY was detected.

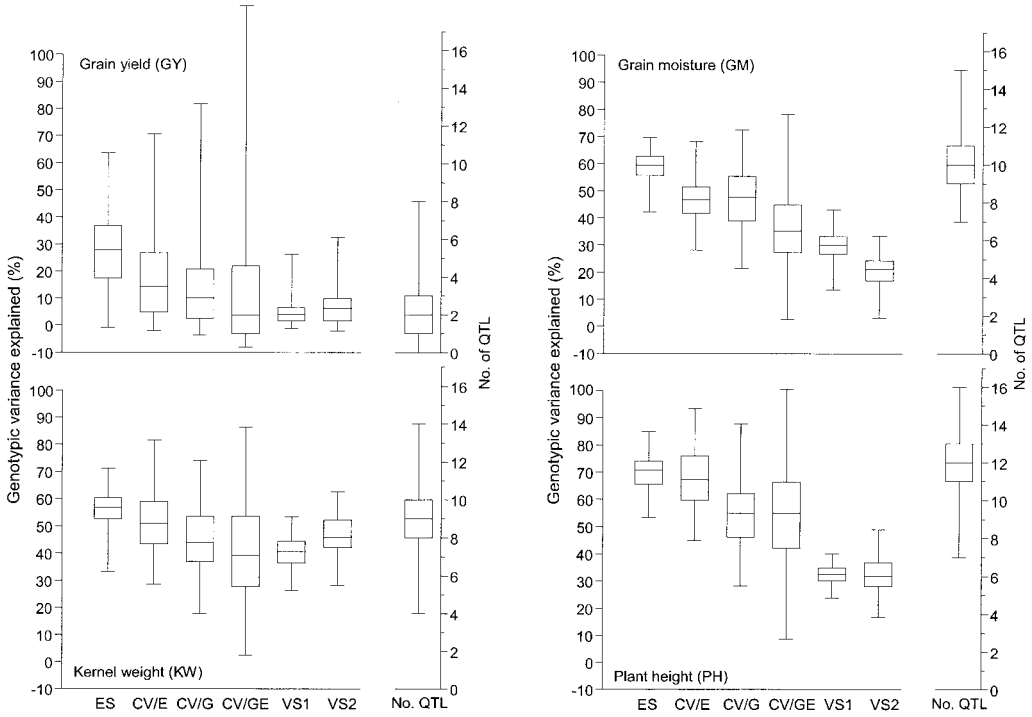

Figure 3.—Box-and-whisker plots of the genotypic variance explained by detected QTL ($\hat{p}$) determined from estimation (ES), cross validation (CV/E, CV/G, and CV/GE), and validation with independent samples (VS1 and VS2) for $k = 5$, $u = 3$, and $e = 1$ for grain yield, grain moisture, kernel weight, and plant height. Boxes indicate the median, 25, and 75% quartiles.

For GM, KW, and PH, between 4 and 16 QTL were significant in ES. A wide variation of $\hat{p}_{ES}$ and $\hat{p}_{TS.ES}$ was found for all three CV schemes and all traits, and the range for $\hat{p}_{TS.ES}$ was substantially larger than for $\hat{p}_{ES}$, the latter being generally of similar magnitude as for $\hat{p}_{TS.ES}$ with independent validation (VS1, VS2).

When varying the number of environments in ES (u) and TS (e) with $k = 5$ for PH, on average 7–14 QTL were detected in ES and $\tilde{p}_{ES}$ varied between 56.4 and 67.1% (Figure 4). Both the number of QTL detected and $\tilde{p}_{ES}$ increased with an increase in u. The median $\tilde{p}_{TS.ES}$ was considerably reduced for CV/G and CV/GE as compared to $\tilde{p}_{ES}$, the difference being smaller for greater values of u. In comparison with $\tilde{p}_{ES}$, the median $\tilde{p}_{TS.ES}$ for CV/E was smaller only for $u = 1$ and 2 but greater for $u = 3$, 4, and 8. In validation with independent samples (VS1 and VS2), $\tilde{p}_{TS.ES}$ was substantially smaller than in CV/GE, the largest reduction being found for VS2 with $u = 1$. For $u = 8$ and $k = 5$ the variation among results from replicated runs is presented in Figure 4. In ES, the analysis of PH yielded 10–17 QTL. The range in $\hat{p}_{TS.ES}$ from CV/E and CV/GE was considerably larger than for $\hat{p}_{ES}$. The variation in $\hat{p}_{TS.ES}$ values for independent validation (VS1 and VS2) was comparable to $\hat{p}_{ES}$.

Results of a more detailed analysis (1376 replicated runs) of one QTL for GY and two QTL for PH on chromosome 7 are presented in Figure 5. In the 1376

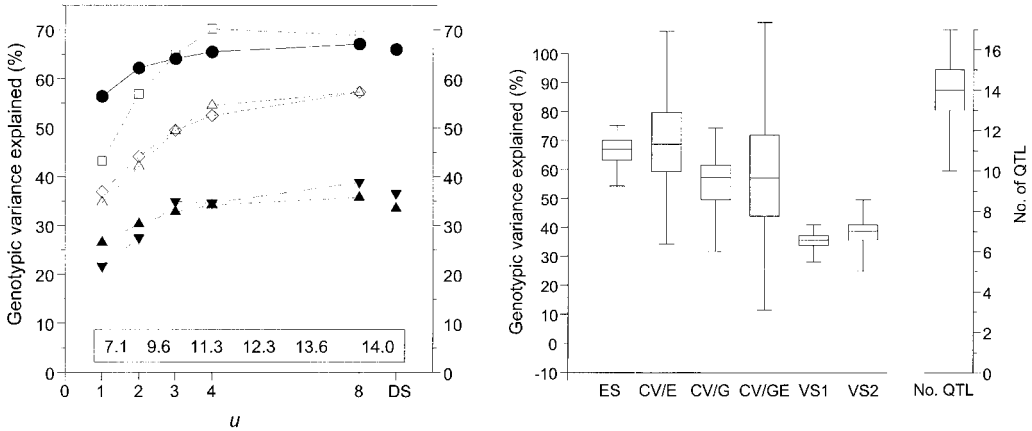

Figure 4.—(Left) Median proportion (percentage) of genotypic variance explained by detected QTL ($\tilde{p}$) determined from estimation (ES), cross validation (CV/E, CV/G, and CV/GE), and validation with independent samples (VS1 and VS2) as well as the average number of QTL detected in estimation (see box, ascending order of u) for $k = 5$ and a varying number of environments in estimation (u) and testing (e) as well as the entire data set (DS) of Experiment 1 for plant height evaluated in nine (CV, $e = 9 - u$), four (VS1, $e = 4$), and five (VS2, $e = 5$) environments, respectively. (Right) Box-and-whisker plots of the genotypic variance explained by detected QTL ($\hat{p}$) for $k = 5$ and $u = 8$ for plant height (see Figure 3 for details). (●) ES, (□) CV/E, (◇) CV/G, (△) CV/GE, (▲) V/VS1, (▼) V/VS2.

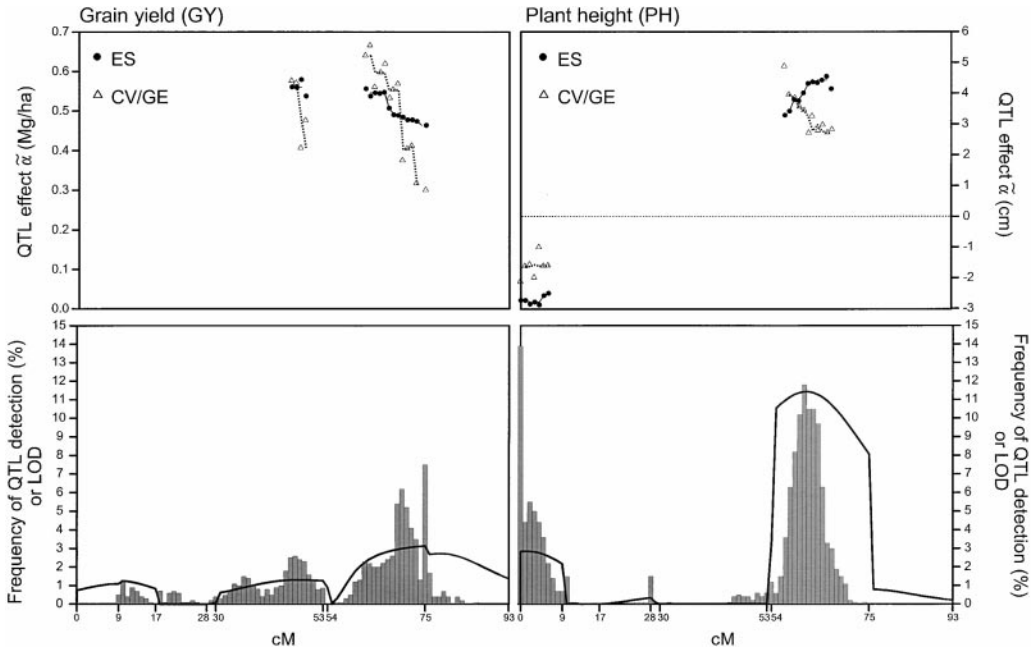

Figure 5.—(Bottom) QTL frequency distributions for grain yield and plant height at 1-cM intervals on chromosome 7 derived from 1376 replicated runs of cross validation with $k = 5$. The solid line indicates the LOD curve determined from the entire data set of Experiment 1 with composite interval mapping. (Top) If QTL frequencies exceeded 2.0 at the respective position on the chromosome, corresponding allele substitution effects were determined for the putative QTL. Median allele substitution effects ($\tilde{\alpha}$) from estimation (ES) and cross validation (CV/GE) are presented. Solid (ES) and dotted (CV/GE) lines are fitted according to the concept of running medians by Tukey (1977).

ES for $k = 5$, significant QTL for GY were detected at almost every position along the chromosome (Figure 5, bottom left). At position 75 cM the maximum of the distribution of relative QTL frequencies was reached (7.5%) but the distribution did not show a well-defined peak. The median allele substitution effects estimated from ES ($\tilde{\alpha}_{ES}$) and TS ($\tilde{\alpha}_{TS}$) are presented if the QTL frequency exceeded 2% (Figure 5, top). Otherwise sampling errors of estimates of effects were considered too large. At position 75 cM $\tilde{\alpha}_{ES}$ was 0.46 Mg ha^{-1} as compared to 0.30 Mg ha^{-1} for $\tilde{\alpha}_{TS}$. For PH the distribution of QTL frequencies along chromosome 7 was bimodal, showing distinct peaks at position 0 cM (13.9%) and 61 cM (11.4%; Figure 5, bottom right). Genetic effects at the two QTL were of opposite sign (Figure 5, top right). In the region 0–10 cM the absolute value of $\tilde{\alpha}_{ES}$ was larger than $\tilde{\alpha}_{TS}$ from CV/GE, amounting to $\tilde{\alpha}_{ES} = -2.7$ cm and $\tilde{\alpha}_{TS} = -2.1$ cm at position 1 cM. Accordingly, at 61 cM the median $\tilde{\alpha}_{ES}$ (4.0 cm) was greater than $\tilde{\alpha}_{TS}$ (3.4 cm).

DISCUSSION

Resampling methods: All statistical methods used for QTL analysis share the problem of model selection because the true number and position of QTL and, hence, the correct statistical model estimating their genetic effects, are unknown. With CIM, the general procedure is to identify among a large number of regressor variables x_i (coded marker genotypes or functions of them) those that account for the largest proportion in the variance of the response variable Y (phenotypic values), and use them for estimation of QTL effects and p. With a limited sample size, model selection leads to an overestimation of QTL effects and p due to sampling effects and consequently to a biased assessment of the prospects of MAS. In this experimental study, we tried to quantify the prediction error of our QTL models and to obtain unbiased estimates of the proportion of genetic variance explained by the detected QTL using resampling methods. CV was preferred over bootstrapping for two reasons: (1) CV/GE provides asymptotically unbiased estimates of p because the data in TS used for testing the prediction are stochastically independent from the data in ES from which the prediction rule is inferred (Davison and Hinkley 1997); (2) CV allows us to evaluate the effects of both genotypic and environmental sampling on estimates of p individually and simultaneously.

Cross validation: In the five CV and validation schemes, $\tilde{p}_{TS.ES}$ was considerably reduced as compared to $\tilde{p}_{ES}$, indicating a large upward bias in predictors of p inferred from estimation sets. The relative bias of estimation ($1 - \tilde{p}_{TS.ES}/\tilde{p}_{ES}$) was greatest for GY. The complex genetic architecture of the trait resulted in only few (two to three) QTL detected in ES with highly overestimated genetic effects owing to sampling and the relatively low heritability of the trait. Therefore, only a small proportion of $\tilde{p}_{ES}$ could be validated in the various TS (Figure 2). The median $\tilde{p}_{TS.ES}$ was 0.0 in two CV schemes ($k = 9$ and $k = 16$) indicating that in half of the CV runs no selection gain would have been achieved by choosing the respective markers for selection.

Charcosset and Gallais (1996) postulated that the adjusted R^2_{adj} instead of the ordinary R^2 from regression yields unbiased estimates of the proportion of phenotypic variance explained by markers. However, as shown

by CV in this study, this does not hold true because the problem of model selection is not taken into account. Knapp (1998) proposed circumventing the pitfalls of biased QTL estimates by including only bona fide QTL in a selection index, *i.e.*, only those QTL that are still significant when stringent significance thresholds are applied for identifying putative QTL. When increasing the LOD threshold to 5.0 for estimation with $k = 5$, only one QTL was detected for GY, when averaged over the 200 replicated CV runs, and about four QTL for GM, KW, and PH, respectively (data not shown). Naturally, the more stringent type I error was accompanied by reduced power for QTL detection and, therefore, fewer QTL explaining a smaller proportion of the genotypic variance were detected as compared to the results with LOD = 2.5. However, the relative bias ($1 - \tilde{p}_{TS.ES}/\tilde{p}_{ES}$) was almost identical for both threshold levels for GM, KW, and PH, demonstrating a certain robustness of CV results. Unless only few QTL were detected for a certain quantitative trait, we did not observe that the magnitude of the bias due to model selection in estimation of QTL effects was strongly influenced by the LOD threshold applied.

Choice of *k* in CV: When using CV, the value *k* for subdivision of the original DS is crucial for determining the bias of $\tilde{p}_{ES}$. Breiman and Spector (1992) showed for multiple regression ($N = 160$) that both 5-fold ($k = 5$) and 10-fold ($k = 10$) cross validation are well suited for model selection and estimation. Twofold ($k = 2$) cross validation tended to select models with too few variables, resulting in a lower accuracy of prediction. Similar results could be observed in our study. For all traits except GY, increasing values of *k*, *i.e.*, larger sample sizes in ES, resulted in an improved power of QTL detection in ES and a decrease in the relative bias ($1 - \tilde{p}_{TS.ES}/\tilde{p}_{ES}$) in CV. For GY, however, the relative bias did not decrease with increasing values of *k*. One reason could be that for a trait with low heritability and complex genetic architecture like GY, even a sample size of $N = 323$ in ES does not provide sufficient power for detection of "true" QTL. To allow in CV for both estimation with a minimum of bias and testing with a minimum of sampling error, the factor *k* for subdivision of DS must be chosen prudently depending on the size of the original DS. This is particularly important with a large number of predictor variables in the model, *e.g.*, for complex traits with a large number of detected QTL or when estimating and testing the effects of epistasis.

Choice of *u* in CV: For the highly heritable trait PH, $u = 3$ seemed sufficient to obtain an almost perfect agreement between $\tilde{p}_{ES}$ and $\tilde{p}_{TS.ES}$ for CV/E. However, for CV/G and CV/GE the closest agreement was obtained with $u = 8$. Therefore, we recommend for CV to include a maximum of environments in ES, leaving only one disconnected environment in TS.

Independent validation: As expected, $\tilde{p}_{TS.ES}$ in independent validation samples VS1 and VS2 increased when the power for QTL detection was improved and the estimation bias was reduced due to increased sample sizes used in ES. Best prediction generally was obtained from estimation in DS, except for PH. The median $\tilde{p}_{TS.ES}$ was smaller in validation samples than in CV/GE for all CV schemes and all traits except KW. Several confounded factors have probably contributed to this finding. For determining $\tilde{p}_{TS.ES}$ for VS1 and VS2, the same genotypic sample was used as TS in all replicated runs, while in CV/GE genotypic sampling was varied for TS. This is shown by the smaller range of $\hat{p}_{TS.ES}$ for VS1 and VS2 than for CV/GE ($k = 5$; Figure 3). Hence, for an assessment of the average gain from MAS, results from CV/GE are to be preferred over independent validation because the latter can be influenced considerably by the specific genotypic sample used for TS. A further reason for the differences in $\tilde{p}_{TS.ES}$ between VS and CV/GE could be the fact that results from environments of Exp. 1 were only partially valid for the environments of Exp. 2 (VS1) and Exp. 3 (VS2). It was surprising, however, that $\tilde{p}_{TS.ES}$ in VS2 was higher than in VS1 for GY and KW. The opposite was expected, because linkage disequilibrium between markers and QTL is reduced in advanced selfing generations. The slightly different genotypic sample and the higher heritability of Exp. 3 in comparison to Exp. 2 might have contributed to this discrepancy.

CV with data from the literature: To examine whether our conclusions concerning the magnitude of bias in $\tilde{p}_{ES}$ revealed by CV could be extended beyond the scope of this study, data from three published QTL experiments on agronomic traits in barley (Hayes *et al.* 1993; http://wheat.pw.usda.gov/ggpages/SxM) and insect resistance in maize (Schön *et al.* 1993; Bohn *et al.* 1996) were reanalyzed with CIM and fivefold CV (Table 2). All three CV schemes (CV/E, CV/G, and CV/GE) yielded very similar results as in our study on maize TC progenies. In all three studies, CV/G caused a greater decline from $\tilde{p}_{ES}$ to $\tilde{p}_{TS.ES}$ than CV/E, and CV/GE showed the largest reduction in $\tilde{p}_{TS.ES}$ except for plant height in barley. For GY of barley, the decrease from $\tilde{p}_{ES}$ to $\tilde{p}_{TS.ES}$ was considerable for CV/G and CV/GE despite the large number of environments used in estimation ($u = 15$). Environmental sampling had a fairly large effect on estimates of *p* for tunnel length in the study of Schön *et al.* (1993) even though G × E interactions were not significant. The reason was that one of the two environments had consistently smaller effects at most QTL than the other. This confirms the finding that data from two environments are probably not sufficient for obtaining accurate estimates of *p* even for traits with high heritability. The largest bias was found for insect resistance in both studies. This can be attributed at least partly to the fact that in both studies dominance effects were included in the model for estimation of QTL effects. In the TC progenies presented in this study and in the barley doubled haploid population, an additive model

TABLE 2

Comparison of results from three QTL studies (Hayes *et al.* 1993; Schön *et al.* 1993; Bohn *et al.* 1996) analyzed with composite interval mapping and cross validation (CV, $k = 5$)

	Parameter estimated[a]	Sampling	Hayes *et al.* 1993[b]		Schön *et al.* 1993[c]		Bohn *et al.* 1996[d]: Damage rating
			Yield	Plant height	Tunnel length	Plant height	
Data set	Heritability		0.77	0.96	0.63	0.87	0.64
	No. of QTL		6	11	10	8	10
	$\hat{p}_{DS}$(%)		65.3	90.3	52.3	60.5	84.0
Estimation set	No. of QTL		5.1	9.5	5.8	7.5	6.5
			1, 14	7, 14	1, 13	4, 11	1, 12
	$\tilde{p}_{ES}$(%)		62.0	89.2	48.3	60.7	88.1
			54.7, 66.6[e]	87.0, 90.3	35.4, 61.5	56.2, 64.3	66.9, 102.4
Test sets	$\tilde{p}_{TS.ES}$(%)	CV/E	59.3	98.3	26.5	53.7	41.1
			8.6, 112.9	73.2, 107.6	21.6, 32.5	50.4, 57.5	27.0, 59.2
		CV/G	33.3	84.8	15.5	42.7	31.6
			22.7, 47.3	79.4, 86.6	6.4, 26.5	36.5, 53.1	10.6, 54.2
		CV/GE	26.2	89.3	11.6	41.0	21.3
			−12.1, 81.9	68.3, 107.9	3.0, 26.1	29.4, 48.3	3.2, 52.3

DS, data set; ES, estimation set; TS, test set.

[a] Number of QTL in ES calculated as the mean; $\tilde{p}_{ES}$ and $\tilde{p}_{TS.ES}$ denote the median across all CV runs.

[b] Barley population consisting of 150 doubled haploid lines tested in 16 environments. Cross validation using $u = 15$ and $e = 1$ with 240 replicated runs.

[c] Maize population consisting of 300 $F_{2:3}$ lines tested in two environments. Cross validation using $u = 1$ and $e = 1$ with 200 replicated runs.

[d] Maize population consisting of 171 $F_{2:3}$ lines tested in three environments. Cross validation using $u = 2$ and $e = 1$ with 200 replicated runs.

[e] Quartiles of 25 and 75%.

was assumed. When dominance and/or epistatic effects are included in the model, the bias in $\tilde{p}_{ES}$ is likely to be increased due to the greater number of parameters to be estimated and their larger sampling error in comparison to additive effects.

Number of QTL and size of effects: The current knowledge about the efficiency of MAS has mainly been inferred from computer simulations (for review see Moreau *et al.* 1998). These investigations generally assumed that the quantitative trait under study is controlled by relatively few (≤10) genes of large effects that lead to a Gaussian normal distribution. These assumptions were supported by the results of numerous experimental studies, where QTL with large genotypic effects on quantitative traits were detected with small population sizes and few test environments (for review see Beavis 1998). However, recently published results from QTL analyses using large populations raise doubts on the validity of the assumptions of few QTL with large effects segregating for complex traits with agronomic importance. In a study with 976 maize testcross progenies evaluated in 19 environments, Openshaw and Frascaroli (1997) found 28 and 36 QTL for GY and PH. Despite this large number of QTL, they explained only 54 and 60% of the genotypic variance, respectively. Cross validation in our study corroborated these findings, with $\tilde{p}_{TS.ES}$ in CV/GE being <60% for all traits. According to Beavis (1998) these results might be indicative for the infinitesimal model upon which historical quantitative genetics was based, where quantitative traits are assumed to be controlled by a large number of genes with fairly small effects.

Presumably due to the negligible estimation bias with the large population size used and in accordance with the large number of QTL detected, Openshaw and Frascaroli (1997) did not find QTL with large effects. Hence, it seems legitimate to question the validity of results from QTL mapping studies with small population sizes. They run a high risk of overestimating genetic effects of QTL and *p* and, therefore, draw overly optimistic conclusions about the prospects of MAS. Consequently, if the expression of a quantitative trait is under the control of a large (>30) number of QTL with small effects it will be quite a challenge for breeders to combine them in one genotype by MAS and for molecular biologists to localize them precisely in the genome and clone them.

Precision of QTL localization: An additional assumption of simulation studies on MAS is that linkage between markers used for selection and the QTL is tight (Knapp 1998; Moreau *et al.* 1998). On the contrary, several researchers showed that precision of QTL localization is mostly poor (Visscher *et al.* 1996). Sillanpää and Arjas (1998) suggested the use of QTL intensity distributions for identification and detailed analysis of genomic regions with putative QTL. To obtain an idea

of the position of a QTL in different ES, we adopted the concept of QTL frequency distributions for cross validation ($k = 5$). As is obvious from Figure 5, the two QTL frequency distributions for GY and PH on chromosome 7 were in good agreement with LOD curves obtained with CIM in DS. For PH, but not for GY, the QTL frequency distribution yielded clear peaks and, hence, it was possible to identify the most likely position for the QTL on chromosome 7. The lack of a well-defined peak in the QTL frequency distribution of GY reflects the poor QTL fidelity in (cross) validation. If localization of the QTL is fairly vague in ES, there is little hope for unbiased estimation of the true genetic effects in TS.

Recommendations: From our experience with these experimental data, we recommend using all three CV schemes (CV/E, CV/G, and CV/GE) to evaluate the influence of environmental and genotypic sampling on the magnitude of the bias of estimates of *p*. With CIM based on multiple regression, CV should be computationally feasible on standard personal computers. In addition, for CV only little extra experimental expenditures are required in contrast to independent validation. If only limited computing resources are available and only small G × E interactions are observed, CV/G seems sufficient to assess the prospects of MAS. Accounting for both factors simultaneously, CV/GE is indispensable for obtaining asymptotically unbiased estimates of *p* for traits with complex genetic architecture and relatively low heritability such as grain yield. Nevertheless, there are still a number of open questions concerning the use of resampling methods for estimation of QTL positions and unbiased estimation of their genetic effects. Further research is needed on the optimum choice of the factors *k* and *u* and how to determine the true position of a QTL from the replicated CV runs. As already discussed above, the factors *k* and *u* need to be chosen as a function of the population size and the number of environments in the DS. The frequency distributions of detected QTL in estimation appeared to be a good tool for data interpretation and definition of QTL positions, at least for the more highly heritable traits. In addition, CV should be compared with other resampling methods such as bootstrapping with respect to desired properties in QTL analysis.

Conclusions: For MAS to be superior to classical phenotypic selection, QTL positions must be estimated with high precision, estimated QTL effects must reflect their true genetic effects, and a sufficient proportion of the genotypic variance of the trait under study must be explained by the detected QTL. Both cross validation and validation with independent samples revealed a large bias in *p* when estimated from the same data set that was also used for determining QTL positions by composite interval mapping. These results were also corroborated with data from a study on yield and plant height in barley and two studies on insect resistance in maize. In all three studies genotypic and environmental sampling had a significant effect on the bias of estimates of *p*. Evidence for a fairly poor precision of estimation of QTL effects was given by the large range of $\hat{p}_{ES}$ and $\hat{p}_{TS.ES}$ in all studies. The asymptotically unbiased estimate $\tilde{p}_{TS.ES}$ from CV/GE was <50% for all traits except PH in all studies, indicating that less than half of the genotypic variance could be explained by QTL, suggesting that quantitative traits are probably controlled by a large number of genes with fairly small effects.

By the construction of QTL frequency distributions we tested the precision of QTL localization. While for plant height the position of a QTL on chromosome 7 could be fairly well determined, the absence of a well-defined peak in the QTL frequency distribution of GY reflected the poor QTL fidelity in estimation. If localization of the QTL is fairly vague in ES, there is little hope for unbiased estimation of its true genetic effects in TS.

On the basis of these results, we recommend improving interpretation of QTL analyses by (1) using QTL frequency distributions for determining the position of a QTL and (2) using cross validation, accounting for environmental and genotypic sampling (CV/GE), to obtain unbiased estimates of the proportion of the genotypic variance explained by QTL and to draw realistic conclusions on the prospects of MAS.

LITERATURE CITED

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*. ASTA, Washington, DC.

Beavis, W. D., 1998 QTL analyses: power, precision and accuracy, pp. 145–162 in *Molecular Dissection of Complex Traits*, edited by A. H. Paterson. CRC Press, Boca Raton, FL.

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.

Breiman, L., and P. Spector, 1992 Submodel selection and evaluation in regression. The *X*-random case. Int. Stat. Rev. **60:** 291–319.

Charcosset, A., and A. Gallais, 1996 Estimation of the contribution of quantitative trait loci (QTL) to the variance of a quantitative trait by means of genetic markers. Theor. Appl. Genet. **93:** 1193–1201.

Cowen, N. M., 1988 The use of replicated progenies in marker-based mapping of QTL's. Theor. Appl. Genet. **75:** 857–862.

Davison, A. C., and D. V. Hinkley, 1997 Bootstrap methods and their application. Cambridge University Press, Cambridge, United Kingdom.

Draper, N. R., and H. Smith, 1981 *Applied Regression Analysis*, Ed. 2. Wiley & Sons, New York.

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.

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.

Hayes, P. M., B. H. Liu, S. J. Knapp, F. Chen, B. Jones *et al.*, 1993 Quantitative trait locus effects and environmental interaction in

a sample of North American barley germ plasm. Theor. Appl. Genet. **87:** 392–401.

Hjorth, J. S. U., 1994 *Computer Intensive Statistical Methods. Validation Model Selection and Bootstrap.* Chapman & Hall, London.

Holloway, J. L., and S. J. Knapp, 1993 G-MENDEL 3.0 software for the analysis of genetic markers and maps. Oregon State University, Corvallis.

Jansen, R. C., 1993 Interval mapping of multiple quantitative trait loci. Genetics **135:** 205–211.

Jansen, R. C., and P. Stam, 1994 High resolution of quantitative traits into multiple loci via interval mapping. Genetics **136:** 1447–1455.

Kendall, M. G., and A. Stuart, 1961 *The Advanced Theory of Statistics. Vol. 2. Inference and Relationship.* Charles Griffin & Company, London.

Knapp, S. J., 1998 Marker-assisted selection as a strategy for increasing the probability of selecting superior genotypes. Crop Sci. **38:** 1164–1174.

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.

Liu, B.-H., 1998 Computational tools for study of complex traits, pp. 43–79 in *Molecular Dissection of Complex Traits*, edited by A. H. Paterson. CRC Press, Boca Raton, FL.

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.

Martinez, O., and R. N. Curnow, 1992 Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor. Appl. Genet. **85:** 480–488.

Melchinger, A. E., H. F. Utz and C. C. Schön, 1998 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. Genetics **149:** 383–403.

Mode, C. J., and H. F. Robinson, 1959 Pleiotropism and the genetic variance and covariance. Biometrics **15:** 518–537.

Moreau, L., A. Charcosset, F. Hospital and A. Gallais, 1998 Marker-assisted selection efficiency in populations of finite size. Genetics **148:** 1353–1365.

Openshaw, S., and E. Frascaroli, 1997 QTL detection and marker-assisted selection for complex traits in maize, pp. 44–53 in *52nd Annual Corn and Sorghum Industry Research Conference*, Edited by ASTA, Washington, DC.

Patterson, H. D., and E. R. Williams, 1976 A new class of resolvable incomplete block designs. Biometrika **63:** 83–92.

Schön, C. C., M. Lee, A. E. Melchinger, W. D. Guthrie and W. L. Woodman, 1993 Mapping and characterization of quantitative trait loci affecting resistance against second-generation European corn borer in maize with the aid of RFLPs. Heredity **70:** 648–659.

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.

Sillanpää, M. J., and E. Arjas, 1998 Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. Genetics **148:** 1373–1388.

Tukey, J. W., 1977 *Exploratory Data Analysis.* Addison-Wesley Publishing Company, Reading, MA.

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*, edited by J. W. Van Ooijen and J. Jansen. Proceedings of the Ninth Meeting of the EUCARPIA Section Biometrics in Plant Breeding, Wageningen, The Netherlands. 6–8 July 1994.

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.

Zeng, Z.-B., 1994 Precision mapping of quantitative trait loci. Genetics **136:** 1457–1468.

Communicating editor: Z-B. Zeng