# Quantitative Trait Locus Mapping Based on Resampling in a Vast Maize Testcross Experiment and Its Relevance to Quantitative Genetics for Complex Traits

**Chris C. Schön,\* H. Friedrich Utz,[†] Susanne Groh,[‡] Bernd Truberg,[§] Steve Openshaw\*\* and Albrecht E. Melchinger[†,1]**

*\*State Plant Breeding Institute and [†]Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, 70593 Stuttgart, Germany, [‡]Pioneer Génétique, 68740 Nambsheim, France, [§]Pioneer Hi-Bred Northern Europe GmbH, 48268 Greven, Germany and \*\*Syngenta Seeds, Stanton, Minnesota 55018*

## ABSTRACT

From simulation studies it is known that the allocation of experimental resources has a crucial effect on power of QTL detection as well as on accuracy and precision of QTL estimates. In this study, we used a very large experimental data set composed of 976 $F_5$ maize testcross progenies evaluated in 19 environments and cross-validation to assess the effect of sample size ($N$), number of test environments ($E$), and significance threshold on the number of detected QTL, the proportion of the genotypic variance explained by them, and the corresponding bias of estimates for grain yield, grain moisture, and plant height. In addition, we used computer simulations to compare the usefulness of two cross-validation schemes for obtaining unbiased estimates of QTL effects. The maximum, validated genotypic variance explained by QTL in this study was 52.3% for grain moisture despite the large number of detected QTL, thus confirming the infinitesimal model of quantitative genetics. In both simulated and experimental data, the effect of sample size on power of QTL detection as well as on accuracy and precision of QTL estimates was large. The number of detected QTL and the proportion of genotypic variance explained by QTL generally increased more with increasing $N$ than with increasing $E$. The average bias of QTL estimates and its range were reduced by increasing $N$ and $E$. Cross-validation performed well with respect to yielding asymptotically unbiased estimates of the genotypic variance explained by QTL. On the basis of our findings, recommendations for planning of QTL mapping experiments and allocation of experimental resources are given.

DURING the past 15 years a large number of studies have identified molecular markers linked to quantitative trait loci (QTL) involved in the inheritance of agronomically important traits. These QTL generally explained a significant proportion of the phenotypic variance of the respective trait and, therefore, gave rise to an optimistic assessment of the prospects of marker-assisted selection (MAS; for review see Lynch and Walsh 1998). On the basis of results from these studies, MAS programs were initiated, leading to controversial results. While some authors succeeded in applying MAS to improve their breeding populations (*e.g.*, Yousef and Juvik 2001) or even clone QTL controlling quantitative traits (*e.g.*, Fridman *et al.* 2000), others reported that no substantial genetic progress was achieved by using MAS (*e.g.*, Openshaw and Frascaroli 1997) or that only a fraction of the putative QTL actually contributed to the inheritance of the trait of interest in a selected population (*e.g.*, Bouchez *et al.* 2002).

An explanation for the latter results could be found in theoretical studies (Lande and Thompson 1990) and computer simulations (Utz and Melchinger 1994; Beavis 1998; Göring *et al.* 2001; Allison *et al.* 2002), which demonstrated especially for small samples that estimates of the proportion of genotypic variance explained by QTL were severely inflated irrespective of the statistical method used for analysis. Reasons are that QTL effects are generally estimated from the same data set used for model selection and factors such as epistasis and QTL × environment interactions additionally bias upward. For marker-assisted breeding this has severe consequences: (i) power calculations for experiments trying to replicate earlier findings in independent samples are based on false assumptions and, therefore, are subject to error; (ii) weights given to individual marker-trait associations as components of selection indices could be severely biased and have a large sampling error; (iii) prospects of MAS are overrated; and (iv) prospects of fine mapping and cloning of a QTL might be misjudged if very small or spurious QTL are chosen on account of their overestimated effects.

The effect of experimental dimensions such as sample size and number of test environments on the power of QTL detection as well as accuracy and precision of QTL estimates has been investigated in simulation studies, generally with the assumption of few ($\leq$10) segregating QTL. On the basis of simulations with 40 segregating

QTL, BEAVIS (1998) raised the question of whether the infinitesimal model, upon which quantitative genetics is based (FISHER 1918), could be confirmed if larger experimental populations were evaluated. He also recommended the use of resampling techniques to obtain asymptotically unbiased estimates of QTL effects. First results with experimental data have been reported by BENNEWITZ *et al.* (2002) for bootstrapping and UTZ *et al.* (2000) for cross-validation. However, when testing the efficiency of resampling techniques in experimental studies a limitation has been their relatively small sample size. Consequently, the question remained, how efficiently resampling techniques could be applied to large populations. Here, data from a vast experimental study composed of almost 1000 $F_5$ maize testcross progenies were used to: (1) estimate the number of QTL involved in expression of grain yield, grain moisture, and plant height; (2) assess the effect of sample size, heritability, and significance threshold on the power of QTL detection, the proportion of the genotypic variance explained by the detected QTL, and the corresponding bias in estimates from experimental data; (3) analyze the behavior of cross-validation-derived estimates for number of QTL, genotypic variance explained, and the magnitude of bias across an array of environmental and genotypic subpopulations; and (4) give recommendations about the sample size and number of environments to be used in QTL mapping experiments for complex quantitative traits. In addition, we used computer simulations to test the usefulness of cross-validation for obtaining unbiased estimates of QTL effects.

## MATERIALS AND METHODS

**Plant materials:** Two elite dent inbred lines, subsequently referred to as P1 and P2, were used as parents. They belonged to the same heterotic pool but were known to be genetically diverse with a coefficient of coancestry (FALCONER and MACKAY 1996) of 0.21. Randomly chosen $F_2$ plants from the cross P1 × P2 were selfed to produce 990 independently derived $F_5$ ($F_{4:5}$) lines. Testcross seed was produced by controlled hand pollinations using each of the 990 $F_5$ lines as male parent and crossing to an unrelated inbred tester line from a complementary heterotic pool. Check inbreds including parents P1 and P2 as well as the $F_1$ between P1 and P2 were also crossed to the inbred tester. All plant materials used in this study are proprietary to Pioneer Hi-Bred International.

**Field experiments:** The testcross progenies were evaluated in 1994 and 1995 in 7 and 12 locations, respectively. The experiments were located in Illinois (3 locations), Indiana (2), Iowa (3), Kansas (1), Nebraska (2), and Italy (1). In each of the 19 environments the experimental design consisted of 18 blocks with 60 entries. Each block contained testcrosses of 55 $F_5$ lines, P1, P2, their $F_1$, and two checks. Trials were performed with one replication per environment. Two-row plots (8.2 m²) were machine planted (5.5–7.0 plants m⁻²) and harvested as grain trials with a combine.

Data were recorded for grain yield in megagrams per hectare, adjusted to 155 g kg⁻¹ grain moisture, and grain moisture in grams per kilogram at harvest. Plant height was measured in centimeters on a plot basis as the distance from the soil level to the uppermost leaf in 16 of the 19 environments.

**RFLP marker genotyping and linkage map construction:** DNA extraction, restriction enzyme digestions, gel electrophoresis, transfer of DNA to nylon membranes, and DNA hybridizations were performed by standard procedures (SAMBROOK *et al.* 1989). Each $F_4$ plant was represented by 20 bulked $F_5$ plants. Observed genotype frequencies at each marker locus were checked for deviations from Mendelian segregation ratios and allele frequency 0.5 using a $\chi^2$ test. Appropriate type I error rates were determined by the sequentially rejective Bonferroni test (HOLM 1979). High-quality molecular data were produced for 976 of the 990 analyzed $F_4$ plants and 172 restriction fragment length polymorphism (RFLP) markers. Therefore, the construction of the linkage map and subsequent QTL analyses are based on 976 genotypes. The software package GMENDEL 3.0 (HOLLOWAY and KNAPP 1993) was used for map construction.

**Agronomic data analyses:** All quantitative genetic parameters were estimated on the basis of the 976 testcross progenies of $F_5$ lines for which high-quality molecular data were available. Each site-year combination was treated as an environment in the analysis. Trait values were adjusted for block effects. For each environment, block effects were calculated as the deviation of the 55 $F_5$ testcrosses in that block from the mean of all $F_5$ testcrosses. An analysis of variance (ANOVA) combined across environments was computed. Components of variance were estimated considering all effects in the statistical model as random. Estimates of variance components $\sigma_{ge}^2$ [genotype-by-environment ($G \times E$) interaction variance confounded with experimental error] and $\sigma_g^2$ (genotypic variance) of testcross progenies of $F_5$ lines and their standard errors (SE) were calculated as described by SEARLE (1971, p. 475). Heritabilities ($h^2$) on a testcross progeny-mean basis were estimated as

$$\hat{h}^2 = \frac{\hat{\sigma}_g^2}{(\hat{\sigma}_{ge}^2/E) + \hat{\sigma}_g^2}, \tag{1}$$

where $E$ is the number of environments. Exact 95% confidence intervals (C.I.) of $\hat{h}^2$ were calculated according to KNAPP *et al.* (1985). Heritabilites on a plot basis ($h_{plot}^2$) were estimated using $E = 1$ in Equation 1.

**QTL analyses:** QTL mapping and estimation of their effects were performed with software PLABQTL (UTZ and MELCHINGER 1996), employing composite interval mapping by the regression approach (HALEY and KNOTT 1992) in combination with the use of cofactors (JANSEN and STAM 1994; ZENG 1994). An additive genetic model was chosen for the analysis of testcross progenies as described by UTZ *et al.* (2000). Cofactors were selected by stepwise regression according to MILLER (1990, p.49). Two different levels of significance were used:

1. Cofactors were chosen with an "*F*-to-enter" and an "*F*-to-delete" value of 12.4 and testing for presence of a putative QTL in an interval by the likelihood-ratio test was performed using a LOD threshold of 3.21. The experimentwise type I error was determined to be $P_e < 0.02$, using 1000 permutation runs (DOERGE and CHURCHILL 1996).
2. *F*-to-enter and *F*-to-delete values were set to 3.5 and the LOD threshold to 2.5. The latter combination corresponds to an experimentwise type I error of $P_e < 0.35$. Estimates of QTL positions were obtained at the position, where the LOD score assumed its maximum in the region under consideration.

The proportion of the phenotypic variance explained by QTL was determined by the estimator $R_{adj}^2$ as described by

TABLE 1

**Number of possible genotypic subpopulations × environmental subpopulations × randomizations as well as the actual number of data sets (in parentheses) used for cross-validation for each of 12 possible combinations of sample size ($N$) and number of test environments ($E$) for experimental data and six possible combinations for simulated data**

| | | | $N$ | | |
|---|---|---|---|---|---|
| $E$ | 976 | 488 | 244 | 122 |
| | | Experimental data | | |
| 19 or 161 | | $2 \times 1 \times 60$ | $4 \times 1 \times 30$ | $8 \times 1 \times 15$ |
| | (1) | (120) | (120) | (120) |
| 4 | $1 \times 4 \times 30$ | $2 \times 4 \times 15$ | $4 \times 4 \times 8$ | $8 \times 4 \times 4$ |
| | (120) | (120) | (128) | (128) |
| 2 | $1 \times 8 \times 15$ | $2 \times 8 \times 8$ | $4 \times 8 \times 4$ | $8 \times 8 \times 2$ |
| | (120) | (128) | (128) | (128) |

| | $N$ | |
|---|---|---|
| $E$ | 488 | 122 |
| | Simulated data | |
| 16 | $10 \times 1 \times 16$ | $40 \times 1 \times 4$ |
| | (160) | (160) |
| 4 | $10 \times 4 \times 4$ | $40 \times 4 \times 1$ |
| | (160) | (160) |
| 2 | $10 \times 8 \times 2$ | $40 \times 8 \times 1$ |
| | (160) | (320) |

Abbreviations are as follows for all tables: CV, cross-validation; CV/GE, cross-validation accounting for genotypic and environmental sampling; DS, data set; ES, estimation set; LOD, likelihood odds ratio; MAS, marker-assisted selection; $P_{ED}$, experimental data reference population; $P_{SD}$, simulated data reference population; QTL, quantitative trait locus/loci; TS, test set.

UTZ *et al.* (2000). The proportion of the genotypic variance explained by all detected QTL was estimated from the ratio

$$\hat{p} = \frac{\hat{R}^2_{adj}}{\hat{h}^2}. \qquad (2)$$

Both parameters, $R^2_{adj}$ and $h^2$, are estimated with an experimental error and estimates of $p$ can exceed 100% or become negative. We did not restrict estimates of $p$ to the parameter space [0, 100], because additional bias is introduced if estimates are constrained to lie within theoretical boundaries (ALLISON *et al.* 2002).

**Subdivision and analysis of experimental data:** From the experimental reference population $P_{ED}$ ($N = 976$, $E = 19$; grain yield and grain moisture) or $P_{ED}$ ($N = 976$, $E = 16$; plant height) an array of (a) genotypic subpopulations of size $N$ ($N = 976$, 488, 244, 122) and (b) environmental subpopulations of size $E$ ($E = 19$ or 16, 4, 2) was sampled without replacement. After randomization of the genotypes and environments in $P_{ED}$ ($N$, $E$), this procedure was repeated 2–60 times to result in a total of 120 or 128 different data sets (DS) per $P_{ED}$ ($N$, $E$) except for $P_{ED}$ (976, 19/16), where only 1 DS exists (Table 1).

Within each $P_{ED}$ ($N$, $E$) estimation of quantitative genetic parameters such as variance components and heritabilities as well as QTL analyses were performed for each DS individually at two levels of significance as described earlier. Fivefold cross-validation (fivefold CV; HJORTH 1994) accounting for genotypic sampling was applied. Each DS was randomly subdivided into five genotypic samples without replacement. Means across all environments of four genotypic samples were used as an estimation data set (ES) for localization of QTL and estimation

of their effects and means across environments of the fifth independent sample were used as the test data set (TS). The TS was used to validate QTL detected in the ES and to obtain asymptotically unbiased estimates of QTL effects and the genotypic variance explained by QTL. For each DS 5 different ES and corresponding TS are possible. The randomization step of assigning genotypes to the five subsamples was repeated 24 times, resulting in 120 different ES and corresponding TS per DS. The CV described in this article (subsequently denoted as "standard CV") deviates slightly from the CV described by UTZ *et al.* (2000), accounting for genotypic sampling, where the ES and TS comprised all but one environment from the DS.

The following parameters were estimated:

The heritability of each trait for each DS ($\hat{h}^2_{DS}$) and averaged over all DS ($\bar{h}^2_{DS}$) for a given $P_{ED}$ ($N$, $E$).

The number of QTL ($\hat{m}_{DS}$) and the proportion of the genotypic variance explained by QTL ($\hat{p}_{DS}$) in each DS as well as their arithmetic mean over all DS ($\overline{m}_{DS}$ and $\overline{p}_{DS}$) for a given $P_{ED}$ ($N$, $E$).

The number of QTL ($\hat{m}_{ES}$) and the proportion of the genotypic variance explained by QTL in each ES ($\hat{p}_{ES}$), their arithmetic mean over all ES for a given DS ($\ddot{m}_{ES}$ and $\ddot{p}_{ES}$), and their grand arithmetic mean over all DS ($\overline{m}_{ES}$ and $\overline{p}_{ES}$) for a given $P_{ED}$ ($N$, $E$).

The proportion of the genotypic variance explained by QTL in each TS ($\hat{p}_{TS}$), the arithmetic mean over all TS for a given DS ($\ddot{p}_{TS}$), the grand arithmetic mean over all DS ($\overline{p}_{TS}$) for a given $P_{ED}$ ($N$, $E$), and the median ($\tilde{p}_{TS}$) and 12.5% [$\dot{p}_{TS}$

(12.5%)] and 87.5% [ $\ddot{p}_{TS}$ (87.5%)] quantiles of the proportion of the genotypic variance for all DS of a given $P_{ED}$ $(N, E)$.

The magnitude of the bias in estimates of the genotypic variance explained by QTL due to genotypic sampling calculated as the difference between average estimates of $p$ obtained from ES and corresponding TS $(\hat{p}_{ES} - \hat{p}_{TS})$, averaged over all ES and TS for a given DS $(\ddot{p}_{ES} - \ddot{p}_{TS})$, the grand arithmetic mean over all DS $(\bar{p}_{ES} - \bar{p}_{TS})$ for a given $P_{ED}$ $(N, E)$, and the median and 12.5 and 87.5% quantiles of the bias for all DS of a given $P_{ED}$ $(N, E)$.

For DS with $\hat{h}_{DS}^2 < 0.002$ the parameters $\hat{p}_{ES}$, $\hat{p}_{TS}$, and the bias were set to zero due to the problem of dividing by a number close to zero.

**Simulation data set:** In the QTL analysis of the experimental data $P_{ED}$ (976, 16) with LOD 3.21, 21 QTL for plant height were detected, explaining an estimated 56.1% of the genotypic variance. On the basis of the linkage map with 172 RFLP markers and assuming the estimated positions and effects of these 21 QTL to be the true QTL parameters, a simulated reference population $P_{SD}$ ($N = 4880$, $E = 16$) consisting of 4880 $F_4$ individuals was generated using software PLABSIM (Frisch *et al.* 2000). The genotypic value of each $F_4$ individual was determined by the known effects at the 21 QTL and a random normal deviate accounting for 43.9% of the genotypic variance attributable to undetected QTL. Moreover, for each individual, 16 random normal deviates were generated for simulation of $G \times E$ interactions plus experimental error under the assumption of $h^2 = 0.3318$ for a single environment, resulting in 16 phenotypic values per $F_4$ genotype with $h^2 = 0.8882$ for 16 environments.

**Subdivision and analysis of simulated data:** The simulated reference population $P_{SD}$ (4880, 16) was partitioned into 10 or 40 genotypic subpopulations of size $N = 488$ or 122, respectively. Environmental subpopulations of size $E = 16$, 4, and 2 were also generated. After randomization of the genotypes and environments in $P_{SD}$ $(N, E)$, this partitioning was conducted 1–16 times to result in a total of 160 DS per $P_{SD}$ $(N, E)$ except for $P_{SD}$ (122, 2) with 320 possible DS (Table 1).

In general, estimation of quantitative genetic parameters such as variance components and heritabilities as well as QTL mapping and QTL parameter estimation ($m_{DS}$, $m_{ES}$, $p_{DS}$, $p_{ES}$, and $p_{TS}$) were conducted as described for the experimental subpopulations, but with simulated data only one threshold for declaring significant QTL (LOD = 2.5 and *F*-to-enter = 3.5) was used. Two CV schemes, standard CV and a second CV analysis accounting for genotypic and environmental sampling simultaneously (CV/GE) as described by Utz *et al.* (2000), were performed for all DS within each of the six $P_{SD}$ $(N, E)$. Twenty CV/GE runs were conducted for each DS.

For each DS QTL positions and effects estimated with composite interval mapping were used to predict the genotypic value of each of the 4880 $F_4$ individuals from the simulated reference population on the basis of its marker genotype ($\hat{G}_{DS}$). Pearson's correlation coefficient $r$ (Snedecor and Cochran 1989, p. 177) was calculated for the predicted and the known, simulated genotypic value of the 4880 $F_4$ individuals ($G$) and the proportion of the genotypic variance attributable to simulated QTL was estimated as $\hat{p}_{SD:DS} = \hat{r}_{adj}^2(\hat{G}_{DS}, G)$. Subsequently, the arithmetic mean over all DS ($\bar{p}_{SD:DS}$) was calculated.

Analogously, for each of the two CV schemes QTL positions and effects estimated in ES were used to predict the genotypic value of each of the 4880 $F_4$ individuals from the simulated reference population on the basis of its marker genotype ($\hat{G}_{ES}$). Likewise, the proportion of the genotypic variance attributable to simulated QTL was estimated as $\hat{p}_{SD:ES} = \hat{r}_{adj}^2(\hat{G}_{ES}, G)$. Subsequently, the arithmetic mean over all ES for a given DS and

the grand arithmetic mean over all DS ($\bar{p}_{SD:ES}$) for a given $P_{SD}$ $(N, E)$ were calculated.

In CV with experimental data, the magnitude of the bias of QTL estimates is determined from the difference $\bar{p}_{ES} - \bar{p}_{TS}$. However, in fivefold CV, estimation of QTL is based on 20% fewer individuals than in DS and, therefore, power of QTL detection is reduced, affecting the estimation of the number of QTL, $p$, and the bias. For the simulated data, the QTL genotypes are (partly) known and an estimate of the truly accountable proportion of the genotypic variance ($\bar{p}_{SD:DS}$ and $\bar{p}_{SD:ES}$) can be obtained. Consequently, an estimate of the bias of QTL effects due to model selection in DS can be calculated from the difference $\bar{p}_{DS} - \bar{p}_{SD:DS}$. The unbiasedness of $\bar{p}_{TS}$ can be derived from the congruency between $\bar{p}_{TS}$ and $\bar{p}_{SD:ES}$. To be useful for assessing the prospects of MAS, $\bar{p}_{TS}$ should be of similar size as $\bar{p}_{SD:DS}$ despite the 20% fewer individuals in ES as compared to DS.

**Consistency of QTL estimates across subpopulations:** QTL consistency across subpopulations was assessed. All QTL detected for plant height in the experimental reference population $P_{ED}$ (976, 16) with LOD 3.21 were assumed as reference QTL. Around their position on the genome, 20-cM intervals (10 cM downstream and upstream) were constructed. Subsequently, QTL mapping results of all DS within a given $P_{SD}$ $(N, E)$ and $P_{ED}$ $(N, E)$ were scanned with LOD 3.21 and the number of QTL positioned within one of the 21 intervals and of the same sign as the reference QTL (matching) was counted. The number of newly occurring QTL (not matching) was also assessed. The same analysis for determining matching and nonmatching QTL was performed for all DS of a given $P_{ED}$ $(N, E)$ on the basis of LOD 2.5 with an extended set of 30 QTL detected in the reference population $P_{ED}$ (976, 16) with a LOD threshold of 2.5.

## RESULTS

**Analysis of the experimental population $P_{ED}$ (976, 19/16):** *Molecular data:* Three chromosomal regions on chromosomes 2, 5, and 8 showed allele and genotype frequencies deviating highly significantly from Mendelian expectations ($P < 0.0001$). The 172 RFLP marker loci spanned a map distance of 1818 cM with an average interval length of 11.2 cM. One hundred percent of the genome was located within a 20-cM distance to the nearest marker.

*Trait means, variances, and heritabilities:* Climatic conditions were favorable for maize production in all environments. Phenotypic correlations between environments based on performance of the testcrosses of the 976 $F_5$ progenies varied between 0.03 and 0.24 for grain yield, between 0.09 and 0.65 for grain moisture, and between 0.22 and 0.44 for plant height. For all three traits, the nonsignificant orthogonal contrast between the average testcross performance of the two parental lines and the testcross mean of the $F_5$ lines indicated the absence of epistasis (Table 2), supporting an additive model for QTL analyses. The range in testcross performance of $F_5$ lines considerably transgressed the testcross means of the parents and $\hat{\sigma}_g^2$ was significantly greater than zero ($P < 0.01$). Heritability on a progeny-mean basis was high ($\hat{h}^2 \geq 0.89$) for grain moisture and plant height but medium for grain yield ($\hat{h}^2 = 0.64$). Heritability on

TABLE 2

**First- and second-degree statistics for maize testcross progenies of parent lines (P1 and P2) and 976 $F_5$ lines from cross P1 $\times$ P2 for grain yield and grain moisture evaluated in 19 environments and plant height evaluated in 16 environments**

| Parameter | Entry nos. | Grain yield (Mg ha$^{-1}$) | Grain moisture (g kg$^{-1}$) | Plant height (cm) |
|---|---|---|---|---|
| Testcross means[a] | | | | |
| P1 | 21 | $10.78 \pm 0.10$[b] | $196.2 \pm 1.37$ | $266.5 \pm 2.33$ |
| P2 | 18 | $10.94 \pm 0.12$ | $211.7 \pm 1.61$ | $274.8 \pm 2.43$ |
| $\bar{P}$ | 39 | $10.85 \pm 0.08$ | $203.4 \pm 1.07$ | $270.8 \pm 1.68$ |
| $F_5$ lines | 976 | $10.66 \pm 0.01$ | $205.4 \pm 0.24$ | $270.7 \pm 0.20$ |
| Range of $F_5$ lines | | 9.46–11.80 | 178.3–230.9 | 251.5–290.8 |
| Variance components | | | | |
| $\sigma_g^2$ | | $0.080 \pm 0.006$** | $51.47 \pm 2.53$** | $34.78 \pm 1.77$** |
| $\sigma_{ge}^2$ | | $0.861 \pm 0.009$ | $85.19 \pm 0.91$ | $70.00 \pm 0.82$ |
| Heritability | | | | |
| $h_{\text{plot}}^2$ | | 0.085 | 0.38 | 0.33 |
| $h^2$ | | 0.64 | 0.92 | 0.89 |
| 95% C.I.[c] on $h^2$ | | 0.60, 0.67 | 0.91, 0.93 | 0.88, 0.90 |

**Significant at the 0.01 probability level.
[a] $\bar{P}$ = TC mean of P1 and P2.
[b] Standard errors are attached.
[c] Confidence interval.

a plot basis was 0.38 and 0.33 for grain moisture and plant height, respectively, and 0.085 for grain yield.

*QTL analyses:* Results from QTL analyses in the experimental population $P_{\text{ED}}$ (976, 19/16) are presented in Table 3 for both significance thresholds. Increasing the LOD and *F*-to-enter threshold (decreasing the type I error rate) decreased the power of QTL detection, as reflected by the number of detected QTL ($\hat{m}_{\text{DS}}$) and the proportion of genotypic variance explained by them in the DS ($\hat{p}_{\text{DS}}$) and averaged over ES ($\ddot{p}_{\text{ES}}$) for all traits. Averages over TS ($\ddot{p}_{\text{TS}}$) were between 3.1% (grain yield) and 7.7% (plant height) higher for LOD 2.5 than for LOD 3.21. Fivefold standard CV revealed no major difference in absolute bias ($\ddot{p}_{\text{ES}} - \ddot{p}_{\text{TS}}$) between the two thresholds of QTL detection for grain moisture and plant height; however, a slightly increased bias was ob-

served for grain yield and LOD 2.5 as compared to LOD 3.21.

**Analysis of experimental subpopulations:** *Heritability estimates* on an entry-mean basis were calculated for each DS within a given $P_{\text{ED}}$ (*N, E*). The average estimates ($\bar{h}_{\text{DS}}^2$) ranged from 17.3% [$P_{\text{ED}}$ (976, 2)] to 63.8% [$P_{\text{ED}}$ (976, 19)] for grain yield, from 54.0% [$P_{\text{ED}}$ (488, 2)] to 92.0% [$P_{\text{ED}}$ (976, 19)] for grain moisture, and from 48.4% [$P_{\text{ED}}$ (244, 2)] to 88.8% [$P_{\text{ED}}$ (976, 16)] for plant height. For all traits, $\hat{h}_{\text{DS}}^2$ showed a wide range across DS especially for small *N* and *E* (data not shown). For grain yield, DS with $\hat{h}_{\text{DS}}^2 < 0.002$ were observed for $P_{\text{ED}}$ (244, 4), $P_{\text{ED}}$ (122, 4), and all $P_{\text{ED}}$ (*N*, 2).

*Power of QTL detection* was affected by the trait under study, the significance threshold, the sample size, and the number of environments (Table 4). The average

TABLE 3

**Number of QTL ($\hat{m}_{\text{DS}}$) detected in the experimental population $P_{\text{ED}}$ (*N* = 976, *E* = 19/16), the proportion of the genotypic variance explained by QTL (%) in the data set ($\hat{p}_{\text{DS}}$) and averaged over estimation sets ($\ddot{p}_{\text{ES}}$) and test sets ($\ddot{p}_{\text{TS}}$), as well as the bias ($\ddot{p}_{\text{ES}} - \ddot{p}_{\text{TS}}$) using fivefold standard cross-validation for two significance levels of QTL detection**

| Parameter estimated | LOD 2.5 and *F*-to-enter = 3.5 | | | LOD 3.21 and *F*-to-enter = 12.4 | | |
|---|---|---|---|---|---|---|
| | Grain yield | Grain moisture | Plant height | Grain yield | Grain moisture | Plant height |
| $\hat{m}_{\text{DS}}$ (no.) | 18 | 32 | 30 | 8 | 23 | 21 |
| $\hat{p}_{\text{DS}}$ | 42.9 | 61.9 | 60.8 | 36.1 | 57.5 | 56.1 |
| $\ddot{p}_{\text{ES}}$ | 42.8 | 59.9 | 58.1 | 35.2 | 55.7 | 51.0 |
| $\ddot{p}_{\text{TS}}$ | 29.8 | 52.3 | 49.8 | 26.7 | 49.0 | 42.1 |
| Bias | 13.0 | 7.6 | 8.3 | 8.5 | 6.7 | 8.9 |

## TABLE 4

**Number of QTL detected in estimation sets averaged over data sets ($\overline{m}_{ES}$) and the range (Min. $\dddot{m}_{ES}$ − Max. $\dddot{m}_{ES}$) of the number of QTL detected in estimation sets for each of 12 possible combinations of sample size ($N$) and number of test environments ($E$) of experimental data using fivefold standard cross-validation and two significance levels for grain yield, grain moisture, and plant height**

| | | LOD 2.5 and $F$-to-enter = 3.5 | | | | LOD 3.21 and $F$-to-enter = 12.4 | | | |
|---|---|---|---|---|---|---|---|---|---|
| $E$ | $N$: | 976 | 488 | 244 | 122 | 976 | 488 | 244 | 122 |
| | | | | | Grain yield | | | | |
| 19 | | 15.3 | 8.9 | 6.4 | 6.2 | 7.4 | 3.4 | 1.6 | 0.5 |
| | | | 5.7–12.4 | 3.4–10.3 | 2.9–9.8 | | 2.0–5.1 | 0.3–3.3 | 0.0–2.3 |
| 4 | | 8.2 | 4.8 | 3.8 | 4.4 | 3.9 | 1.7 | 0.6 | 0.2 |
| | | 4.1–13.7 | 2.3–8.8 | 1.5–7.7 | 1.2–9.4 | 1.4–7.3 | 0.1–4.6 | 0.0–2.3 | 0.0–1.7 |
| 2 | | 6.0 | 3.5 | 2.9 | 4.0 | 3.0 | 1.0 | 0.3 | 0.1 |
| | | 1.6–10.3 | 0.5–7.9 | 0.7–6.4 | 1.1–8.2 | 0.1–7.0 | 0.0–4.5 | 0.0–1.6 | 0.0–1.7 |
| | | | | | Grain moisture | | | | |
| 19 | | 28.3 | 19.2 | 12.8 | 9.7 | 18.4 | 9.3 | 3.2 | 0.7 |
| | | | 14.1–23.5 | 8.3–16.6 | 6.1–14.7 | | 6.5–12.3 | 0.5–5.6 | 0.0–3.2 |
| 4 | | 22.6 | 15.2 | 10.3 | 8.3 | 14.6 | 7.1 | 2.2 | 0.6 |
| | | 19.7–26.3 | 11.1–20.2 | 5.3–14.7 | 4.3–13.0 | 12.0–19.9 | 4.1–11.7 | 0.3–4.9 | 0.0–2.0 |
| 2 | | 19.2 | 12.6 | 8.7 | 7.6 | 12.5 | 5.6 | 1.7 | 0.5 |
| | | 13.0–23.8 | 7.9–16.7 | 4.3–13.9 | 2.1–13.0 | 7.5–17.7 | 1.7–9.8 | 0.0–5.2 | 0.0–2.0 |
| | | | | | Plant height | | | | |
| 16 | | 26.4 | 17.6 | 12.0 | 9.1 | 14.8 | 8.4 | 3.6 | 1.0 |
| | | | 14.6–22.9 | 8.2–17.1 | 5.5–13.7 | | 6.3–10.3 | 1.6–5.8 | 0.0–2.9 |
| 4 | | 19.0 | 12.5 | 8.7 | 7.2 | 10.8 | 5.9 | 2.0 | 0.6 |
| | | 15.2–24.2 | 7.9–16.6 | 4.3–14.7 | 4.1–12.8 | 8.8–13.0 | 4.0–8.8 | 0.2–4.6 | 0.0–2.1 |
| 2 | | 14.5 | 9.1 | 6.5 | 6.1 | 8.3 | 3.9 | 1.1 | 0.4 |
| | | 7.7–19.7 | 2.1–14.5 | 2.7–12.1 | 2.5–11.2 | 5.3–10.9 | 1.1–6.4 | 0.0–3.3 | 0.0–2.0 |

number of QTL in ES ($\overline{m}_{ES}$) was highest for grain moisture in $P_{ED}$ (976, 19) with LOD 2.5 (28.3) and lowest for $P_{ED}$ (122, 2) and grain yield with LOD 3.21 (0.1). With increasing $N$ and $E$, $\overline{m}_{ES}$ increased for all traits and both significance levels, except for grain yield with more QTL detected for $P_{ED}$ (122, 2, and 4) as compared to $P_{ED}$ (244, 2, and 4) with LOD 2.5. In small samples only few QTL were detected for LOD 3.21; on average $\overline{m}_{ES} \leq$ 1.0 with $N$ = 122 for all traits and numbers of test environments. With LOD 3.21, only in few DS were >10 QTL detected even for large samples. For $N$ = 122 most DS yielded ES with no detected QTL ($\dddot{m}_{ES}$ = 0).

*The average proportion of the genotypic variance explained by the detected QTL* in TS ($\overline{p}_{TS}$) generally increased with increasing $N$ for all traits and both significance thresholds (Figure 1) and was always greater for LOD 2.5 than for LOD 3.21. The number of test environments had only a small effect on both $\overline{p}_{TS}$ and the bias ($\overline{p}_{ES} - \overline{p}_{TS}$). A dramatic increase in bias could be observed for small $N$, LOD 2.5, and grain yield. As a consequence, $\overline{p}_{ES}$ was greatest for small $N$ and grain yield with an estimated average bias close to 100% for $P_{ED}$ (122, 2). For grain moisture and plant height and LOD 2.5, $\overline{p}_{ES}$ was almost constant for all $P_{ED}$ ($N$, $E$). For LOD 3.21, average bias was similar for most $P_{ED}$ ($N$, $E$) in all traits. Deviating

from expectations, in some cases the average bias even increased from smaller to larger subpopulations. This can be attributed to the fact that (i) for small subpopulations, more ES with 0 detected QTL occurred and (ii) for $E$ = 2, samples with no significant genotypic variance ($\hat{h}_{DS}^2$ = 0) were observed, in which case $\hat{p}_{ES}$ and $\hat{p}_{TS}$ were zero, resulting in zero bias.

The mean ($\overline{p}_{TS}$), median ($\tilde{p}_{TS}$), and 12.5 and 87.5% quantiles [$\dddot{p}_{TS}$ (12.5%), $\dddot{p}_{TS}$ (87.5%)] of the proportion of the genotypic variance explained in TS is shown for each combination of $P_{ED}$ ($N$, $E$) and LOD 2.5 in Figure 2. Results with LOD 3.21 were similar and are therefore not shown. For grain moisture and plant height, variation of $\dddot{p}_{TS}$ among DS was increased for small $N$. For grain yield and $E$ = 4 and $E$ = 2, however, $N$ had no clear effect on the range of $\dddot{p}_{TS}$. Increasing the number of test environments generally decreased the range of $\dddot{p}_{TS}$. The mean ($\overline{p}_{TS}$) and the median ($\tilde{p}_{TS}$) proportions of genotypic variance explained by QTL were in good agreement except for grain yield and $E$ = 2 and 4.

A similar picture was observed when analyzing the variation in the bias ($\dddot{p}_{ES} - \dddot{p}_{TS}$) among DS for a given $P_{ED}$ ($N$, $E$) (Figure 3). For small $N$ and $E$, the variation of bias was large for grain yield and LOD 2.5. Estimates of the bias of over 100% occurred in some DS with a
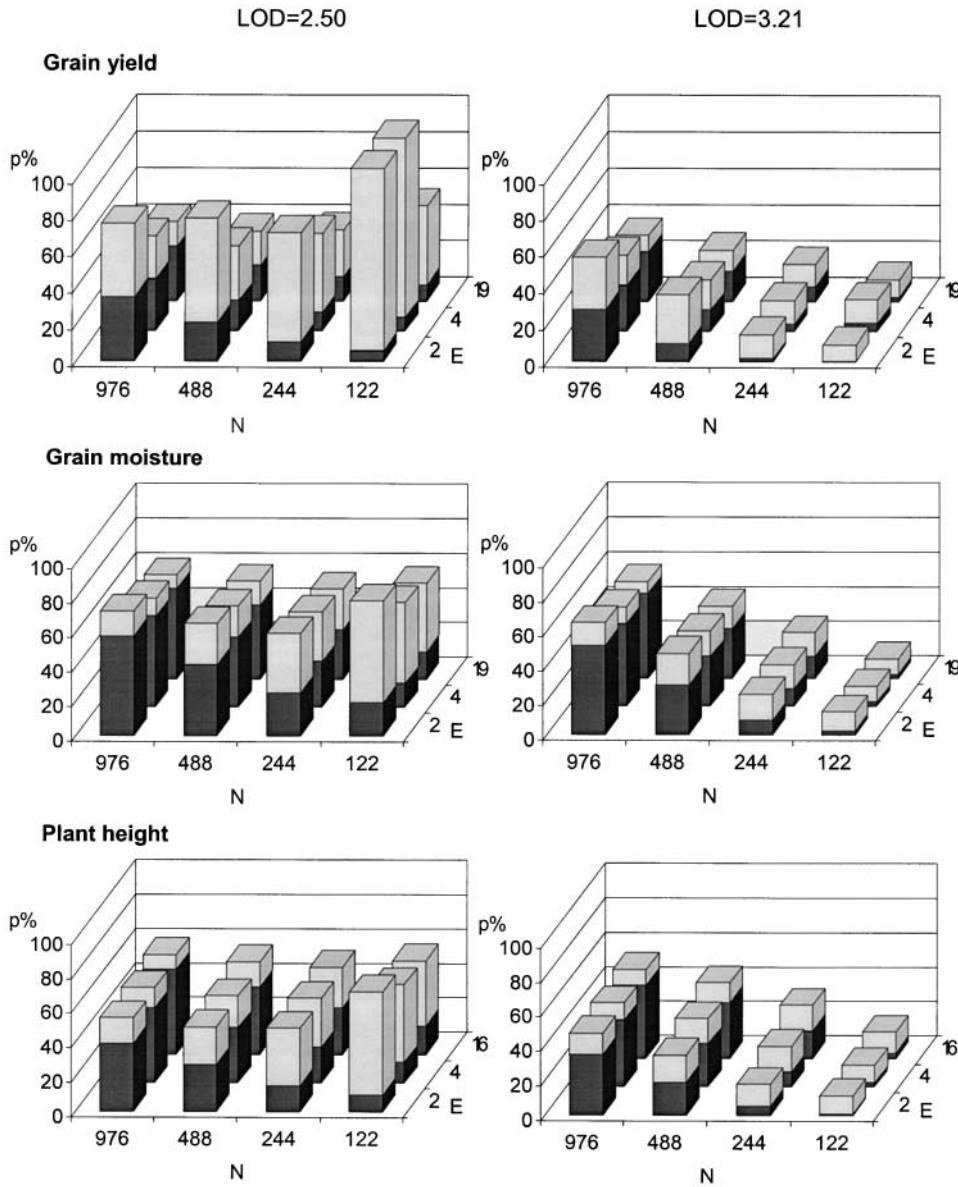
FIGURE 1.—Proportion of the genotypic variance explained by detected QTL in estimation sets averaged over all data sets ($\bar{p}_{ES}$) for 12 combinations of experimental data $P_{ED}$ ($N, E$), using fivefold standard cross-validation and two significance levels for grain yield, grain moisture, and plant height. Individual columns are partitioned into the genotypic variance explained in test sets ($\bar{p}_{TS}$, solid bottom) and the bias calculated as the difference $\bar{p}_{ES} - \bar{p}_{TS}$ (shaded top).

largely overestimated proportion of genotypic variance explained in ES divided by a small heritability estimate. The range of the 12.5 and 87.5% quantiles was considerably reduced for LOD 3.21 as compared to LOD 2.5. As shown for $\ddot{p}_{TS}$, the mean bias and the median bias across DS differed considerably for grain yield and $E = 2$. For LOD 2.5, the median bias decreased with increasing $N$ and $E$ for all traits and combinations of $P_{ED}$ ($N$, $E$) according to expectations. However, with LOD 3.21 the median and the mean bias increased from $N = 122$ to $N = 244$ for all three traits due to a high number of ES with 0 detected QTL.

**Analysis of simulated subpopulations:** Results for plant height in each of the six $P_{SD}$ ($N$, $E$) are given in Table 5. Heritability estimates agreed well with those underlying the simulated values. In none of the 160 DS samples (320 for $N = 122$, $E = 2$) could all 21 QTL be detected. Power of QTL detection and the proportion

of estimated QTL positions lying within a 20-cM interval of the simulated QTL (matching) increased with increasing $N$ and $E$ in DS analyzed with composite interval mapping and in ES analyzed with standard CV and CV/GE. With $N = 122$ and $E = 2$, on average only 2.5 simulated QTL (4.7 × 0.53) were detected with composite interval mapping. The proportion of genotypic variance explained by QTL estimated in DS ($\bar{p}_{DS}$) was generally greater than the proportion of the genotypic variance attributable to simulated QTL ($\bar{p}_{SD:DS}$), revealing considerable bias of QTL estimation (7.6–37.4%), especially for the small sample size. The number of matching QTL was smaller for the two CV schemes than for composite interval mapping due to the 20% smaller population size and/or reduced number of environments in ES as compared to DS. As a consequence, estimates of the bias increased from composite interval mapping ($\bar{p}_{DS} - \bar{p}_{SD:DS}$) to standard CV ($\bar{p}_{ES} -$
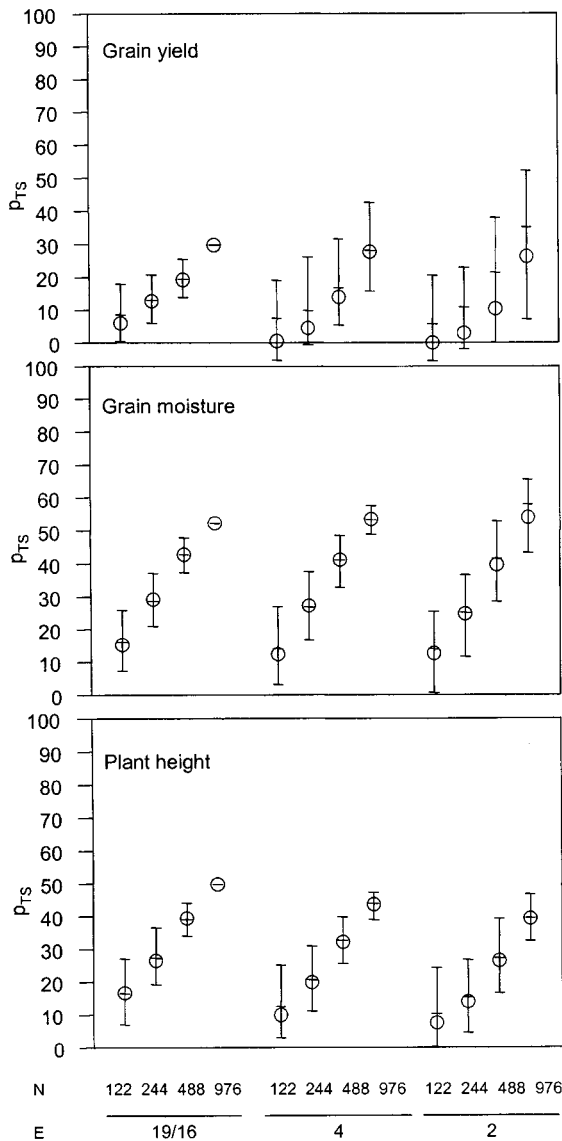
FIGURE 2.—Mean (–), median (o), and 12.5 and 87.5% quantiles of the proportion of the genotypic variance explained in test sets [$\bar{p}_{TS}$, $\tilde{p}_{TS}$, $\ddot{p}_{TS}$ (12.5%), and $\ddot{p}_{TS}$ (87.5%)] calculated for individual data sets for 12 combinations of experimental data $P_{ED}$ $(N, E)$ using fivefold standard cross-validation and LOD 2.5 for grain yield, grain moisture, and plant height.

$\bar{p}_{TS}$) and were greatest for CV/GE and $P_{SD}$ (122, 2). For all combinations of $N$ and $E$ a better agreement of $\bar{p}_{TS}$ with $\bar{p}_{SD:ES}$ and $\bar{p}_{SD:DS}$ was found for standard CV than for CV/GE. For standard CV, a close agreement between $\bar{p}_{TS}$ and $\bar{p}_{SD:ES}$ and only a slight decrease in magnitude from $\bar{p}_{TS}$ to $\bar{p}_{SD:DS}$ (<5%) for all combinations of $N$ and $E$ indicated that $\bar{p}_{TS}$ can be used as an unbiased estimate of the genotypic variance explained by QTL with finite population sizes. For CV/GE and $N = 122$, $\bar{p}_{TS}$ deviated markedly from $\bar{p}_{SD:ES}$ and $\bar{p}_{SD:DS}$, indicating that more research is needed to examine the small sample properties of $\bar{p}_{TS}$ from CV/GE.

**Consistency of QTL estimates across experimental subpopulations:** The 21 intervals of size 20 cM constructed around the QTL detected for plant height in the DS of $P_{ED}$ (976, 16) with LOD 3.21 did not overlap. On average, the number of detected QTL matching the reference QTL was increased considerably for DS with large $N$ [$P_{ED}$ (976, $E$)] as compared to the smaller subpopulations [*e.g.*, $P_{ED}$ (122, $E$)] (Table 6). The number of unmatched QTL was increased to a much smaller extent from smaller to larger populations, indicating a higher ratio of true:false QTL for the larger subpopulations. Doubling the number of environments did not improve the ratio of true:false QTL as much as doubling the number of testcross progenies in the subpopulation. Similar results were obtained with 30 intervals of size 20 cM constructed around QTL detected in DS of $P_{ED}$ (976, 16) with a LOD threshold of 2.5.

## DISCUSSION

**Influence of sample size and number of test environments:** To our knowledge, phenotypic and molecular data on testcross progenies of 976 $F_5$ lines evaluated in 19 environments is by far the largest QTL mapping experiment ever published in plants. The dimensions of the experiment had been designed to meet assumptions from a simulation study performed by BEAVIS (1998), who had inferred that with 40 QTL with additive effects of equal size, a heritability of 63%, and a sample of 1000 $F_2$ progenies it should be possible to obtain a power of QTL detection of ~60% and consequently explain ~60% of the genotypic variance with QTL. With the high-density genetic map used in this study (average interval length 11.2 cM), the power of QTL detection was expected to be even higher compared to the simulation study by BEAVIS (1998) because of the higher heritability (92% *vs.* 63%) and the different population type (testcross progenies of $F_5$ lines *vs.* $F_2$ progenies). However, the maximum, validated genotypic variance explained by QTL in this study was 52.3% for grain moisture (Table 3), which is fairly small considering the expenditures that had to be undertaken for testing almost 1000 unselected testcross progenies in 19 environments. A substantial bias was found for estimates of the proportion of genotypic variance explained by the detected QTL even with $N = 976$, irrespective of the trait, the heritability, and the significance threshold. This corroborates results from the study by BEAVIS (1998), who pointed out that the bias of QTL estimates could not be ignored even for $N > 500$. Results obtained with simulated data in this study support these findings. With $N = 488$ and $E = 16$ an absolute bias of 7.6% was observed when estimating $\bar{p}_{DS}$, indicating that QTL mapping results need to be interpreted with caution and strategies are needed for their validation.

In simulated and experimental data, the effect of sample size on QTL parameter estimation was large.
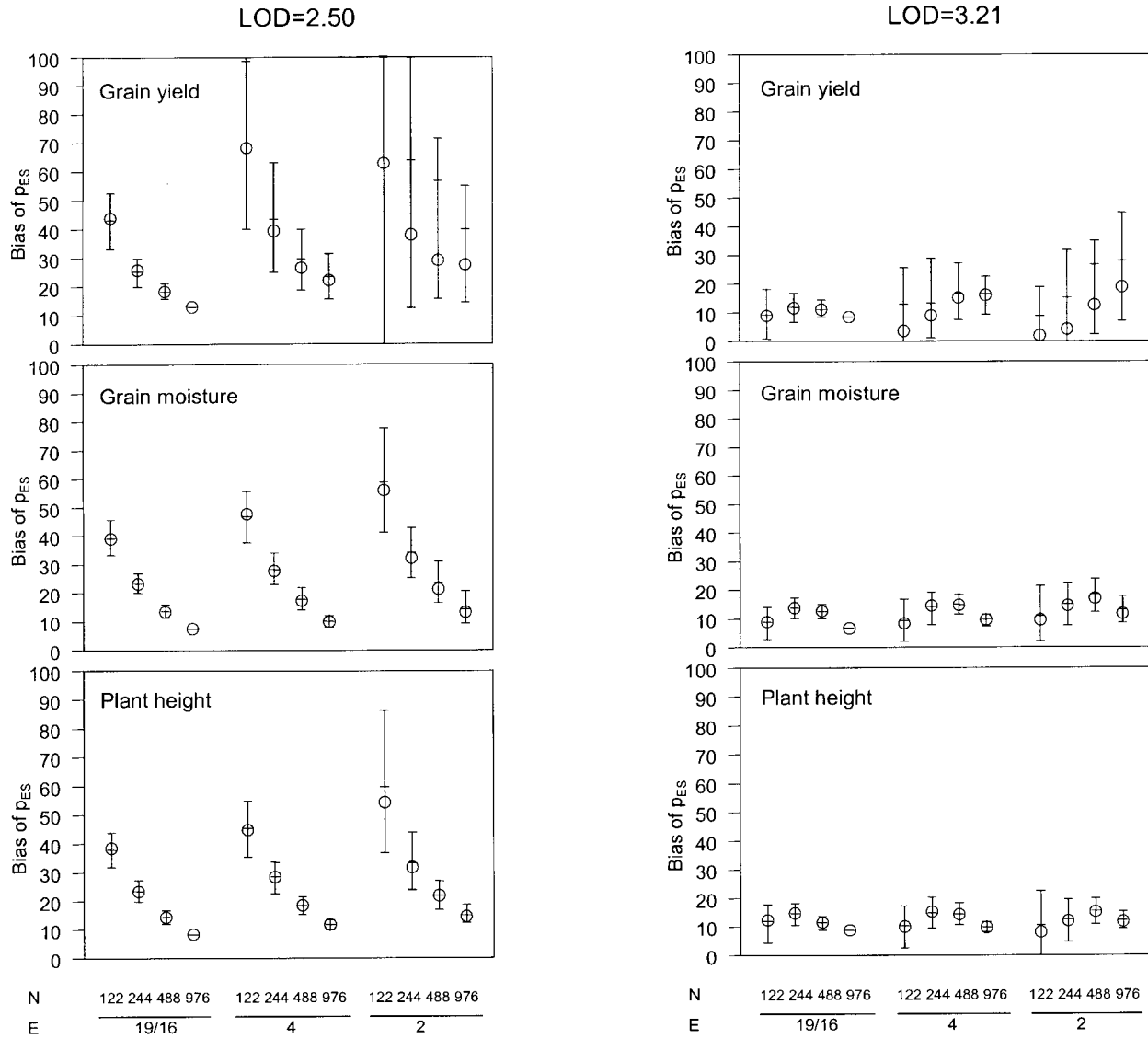
LOD=2.50

LOD=3.21



FIGURE 3.—Mean (–), median (o), and 12.5 and 87.5% quantiles of the bias ($\ddot{\bar{p}}_{ES} - \ddot{\bar{p}}_{TS}$) calculated for individual data sets for 12 combinations of experimental data $P_{ED}$ (*N, E*), using fivefold standard cross-validation and two significance levels for grain yield, grain moisture, and plant height.

As expected, the number of detected QTL generally increased with increasing sample size. The comparison of subpopulations with the same plot capacities for phenotypic evaluation revealed that increasing the number of progenies generally increased the power of QTL detection ($\overline{m}_{ES}$) and the proportion of the genotypic variance explained by QTL ($\overline{p}_{TS}$) and reduced the bias more efficiently than did increasing the number of test environments. For grain yield and LOD 2.5, however, the number of detected QTL was higher when doubling plot capacities from $P_{ED}$ (122, 2) to $P_{ED}$ (122, 4) as compared to $P_{ED}$ (244, 2), probably due to the fact that the estimated average heritability for grain yield and $E = 2$ ($\overline{h}_{DS}^2 < 0.18$) was too low for detecting significant QTL for small *N*. For grain moisture, the number of test environments had only little effect on $\overline{p}_{TS}$ for all popula-

tion sizes, presumably because few QTL showed significant interactions with environments. As pointed out by MOREAU *et al.* (1998) and KNAPP and BRIDGES (1990), it is therefore advisable in a MAS program to increase population size rather than the number of test environments or replications for most traits unless plot heritabilities are very low and/or the expenditures for molecular analyses of additional genotypes are much higher than those for additional testing of phenotypes.

When increasing the population size from $N = 488$ to $N = 976$, the increase in the proportion of genotypic variance explained by QTL ($\overline{p}_{TS}$) per additionally tested genotype was always smaller as compared to increasing *N* from 244 to 488. This diminishing return per additional test unit was expected due to the nonlinear relationship of sample size and power of QTL detection

TABLE 5

Heritability, average number of detected QTL ($\overline{m}_{DS}$ and $\overline{m}_{ES}$), proportion of detected QTL matching simulated QTL (%), and proportion of genotypic variance explained by QTL ($p$ in %) for plant height calculated using composite interval mapping (CIM) and two cross-validation schemes (standard CV and CV/GE) averaged over 160 or 320 simulated data sets (DS) for six combinations of sample size ($N$) and number of test environments ($E$)

| Parameter [a] | $N$: | $E$ [b] | | | | | |
| | | 16 | | 4 | | 2 | |
| | | 488 | 122 | 488 | 122 | 488 | 122 |
|---|---|---|---|---|---|---|---|
| $\hat{h}^2_{DS}$ | | 0.887 | 0.886 | 0.663 | 0.659 | 0.497 | 0.489 |
| | | CIM | | | | | |
| $\overline{m}_{DS}$ | | 12.7 | 7.6 | 10.0 | 5.7 | 7.5 | 4.7 |
| Matching QTL | | 87.3 | 63.6 | 84.8 | 56.6 | 83.1 | 53.0 |
| $\overline{p}_{DS}$ | | 43.6 | 41.7 | 41.2 | 40.9 | 37.2 | 44.6 |
| $\overline{p}_{SD:DS}$ | | 36.0 | 15.0 | 30.3 | 10.2 | 23.8 | 7.2 |
| | | Standard CV | | | | | |
| $\overline{m}_{ES}$ | | 11.5 | 7.2 | 8.7 | 5.8 | 6.6 | 4.9 |
| Matching QTL | | 84.9 | 57.4 | 82.0 | 51.5 | 80.1 | 46.8 |
| $\overline{p}_{ES}$ | | 42.5 | 44.0 | 39.3 | 46.1 | 35.7 | 52.7 |
| $\overline{p}_{TS}$ | | 31.7 | 10.6 | 25.7 | 7.3 | 19.1 | 5.4 |
| $\overline{p}_{SD:ES}$ | | 33.1 | 12.4 | 26.6 | 8.4 | 20.3 | 6.0 |
| | | CV/GE | | | | | |
| $\overline{m}_{ES}$ | | 11.4 | 7.2 | 7.9 | 5.5 | 4.3 | 4.1 |
| Matching QTL | | 84.9 | 56.5 | 81.6 | 50.0 | 74.6 | 41.1 |
| $\overline{p}_{ES}$ | | 42.7 | 44.2 | 38.0 | 47.6 | 30.5 | 60.0 |
| $\overline{p}_{TS}$ | | 29.5 | 2.1 | 21.3 | −0.6 | 9.8 | −2.4 |
| $\overline{p}_{SD:ES}$ | | 32.9 | 12.3 | 24.1 | 7.5 | 12.6 | 3.8 |

[a] For details of parameter estimation see MATERIALS AND METHODS. Matching QTL (%) are given as the proportion of detected QTL found within a 20-cM interval flanking one of the 21 simulated QTL.
[b] $E = 16$ corresponds to $h^2 = 0.888$, $E = 4$ to $h^2 = 0.665$, and $E = 2$ to $h^2 = 0.489$.

(LYNCH and WALSH, 1998). In addition, BOST *et al.* (2001) pointed out that genetic factors, such as enzyme variation in metabolic pathways, can lead to an L-shaped distribution of QTL effects for a given quantitative trait. The distribution of the standardized genetic effects found for QTL in the experimental reference popula-

tion $P_{ED}$ (976, 19/16) is shown in Figure 4. The largest genetic effect was detected for grain moisture ($\alpha_{max} = 0.49\hat{\sigma}_P$). The median genetic effect was small ($0.1\hat{\sigma}_P < \tilde{\alpha} < 0.2\hat{\sigma}_P$) and the distributions were skewed toward smaller values (L-shaped) for all traits. These findings corroborate the hypothesis that polygenic traits are reg-

TABLE 6

Number of matching and nonmatching (in parentheses) QTL for plant height with respect to the QTL detected in the experimental reference population $P_{ED}$ (976, 16) with two levels of significance averaged across data sets for each of 12 possible combinations of sample size ($N$) and number of test environments ($E$) of experimental data

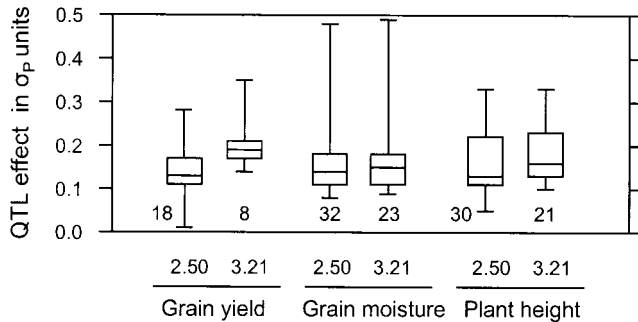| $E$ | $N$: | LOD 2.5 and $F$-to-enter = 3.5 | | | | LOD 3.21 and $F$-to-enter = 12.4 | | | |
| | | 976 | 488 | 244 | 122 | 976 | 488 | 244 | 122 |
|---|---|---|---|---|---|---|---|---|---|
| 16 | | 30.0 | 17.4 (2.4) | 11.0 (3.0) | 6.5 (3.3) | 21.0 | 8.6 (1.6) | 3.8 (1.1) | 1.0 (0.6) |
| 4 | | 19.7 (1.9) | 11.7 (2.5) | 6.8 (2.6) | 4.3 (2.8) | 10.5 (2.2) | 5.9 (1.6) | 2.3 (0.9) | 0.5 (0.4) |
| 2 | | 14.5 (2.1) | 8.5 (1.9) | 5.2 (2.1) | 3.3 (2.7) | 8.2 (1.9) | 4.2 (1.3) | 1.1 (0.5) | 0.3 (0.3) |

Figure 4.—Boxplot with median, extremes, and quartiles of standardized absolute QTL effects estimated from the experimental reference population $P_{ED}$ (976, 19/16), using two significance thresholds for grain yield, grain moisture, and plant height. The number of detected QTL is given below the box.

ulated by a large number of genes with small effects that follow approximately a geometric distribution. Hence, it seems questionable if simulation studies on QTL mapping and MAS assuming 5–10 QTL with equal effects of up to or $>1.0\sigma_P$ are reflecting the true inheritance of polygenic quantitative traits such as grain yield.

Depending on the genetic architecture of the trait and its environmental stability, the extra input of resources for explaining a small additional proportion of the genotypic variance by markers can be vast (MAS for QTL with effects of $0.1\sigma_P$ and even more, so their cloning seems an idle undertaking). Therefore, trait-specific strategies for MAS have to be developed. MAS seems promising only if alleles with large effects are segregating for the trait of interest. Falconer and Mackay (1996) gave examples for such traits in animal breeding. They pointed out that a "large" effect in this context would be $0.5$–$1.0\sigma_P$. In plant breeding experiments, QTL with effects of this size have been reported. However, results from QTL studies indicating the presence of major genes with large effects have to be interpreted with caution due to the problem of model selection. In a large number of published QTL studies with small sample size ($100 < N < 200$) a considerable proportion of the genotypic variance could be explained by few QTL (Lynch and Walsh 1998). Probably few of these QTL would hold in a MAS program what they promised. As can be shown from the experimental data presented here, $\bar{p}_{ES}$ was almost constant for grain moisture and plant height for all combinations of $P_{ED}$ ($N$, $E$) and LOD 2.5 but for small $N$ most of $\bar{p}_{ES}$ must be attributed to bias and not to the effects of real QTL. These findings were also corroborated by results from simulated data on plant height. When performing a linear regression of the proportion of genotypic variance explained in ES on the number of detected QTL, the correlation ($r$) between $\bar{p}_{ES}$ and $\overline{m}_{ES}$ was relatively high, amounting to $r = 0.74$ for $P_{SD}$ (488, 4) and $r = 0.80$ for $P_{SD}$ (122, 4). When the dependent variable was $\bar{p}_{TS}$, however, the cor-

relation dropped to $r = 0.57$ for $N = 488$ and even more for $N = 122$ ($r = 0.37$), indicating that especially for small $N$ quite a few QTL detected in ES were false positives and did not contribute to $\bar{p}_{TS}$.

**Influence of significance threshold:** Knapp (1998) suggested using a conservative significance threshold to improve the accuracy of the selection index in MAS. As expected, the power of QTL detection was increased in experimental data with LOD 2.5 and $F$-to-enter $= 3.5$ as compared to LOD 3.21 and $F$-to-enter $= 12.4$. It was surprising, however, that increasing the type I error rate also increased $\bar{p}_{TS}$ for all three traits and most combinations of $P_{ED}$ ($N$, $E$). The effect was also reflected in the higher percentage of QTL matching those of the reference population for LOD 2.5 as compared to LOD 3.21 (Table 6). This corroborates results described by Moreau et al. (1998), who found for low $h^2$ ($<0.2$) that increasing the type I error rate can lead to a higher relative efficiency of MAS because the power of QTL detection increased more than the risk of detecting false positives. However, for higher estimates of $h^2$ they reported this relationship to be vice versa. Reasons for the discrepancies between their simulation study and our results could be the assumption of only few segregating QTL (5 and 10) in the study by Moreau et al. (1998) as compared to a much higher number in our study. With few QTL and a high heritability, the power of QTL detection seems sufficiently high even with a more conservative threshold.

The choice of significance threshold depends on the goals of the breeder and the cost of marker analyses. For construction of an ideal genotype a more conservative threshold should be chosen to minimize the risk of false positives. As shown here, for complex quantitative traits the number of putative QTL is very large, but due to the L-shaped distribution of detected QTL effects each additional marker linked to a putative QTL would produce diminishing returns but equal costs. Hence, even if a large population was available from which to select optimal genotypes, the optimal number of putative QTL to be used for genotype construction would be much smaller than the total number of QTL detected in a mapping experiment of reasonable sample size, especially when considering more than one trait simultaneously.

When constructing a selection index for combined marker-assisted and phenotypic selection, the type I error rate determines the stop criterion for including additional markers in the model. Depending on marker costs and the magnitude of the detected QTL effects, i.e., the genetic architecture of the trait, an optimum type I error rate should exist. If marker costs are neglected, the experimentwise type I error of $P_e \leq 0.35$ used in this study for grain yield, grain moisture, and plant height yielded a higher efficiency of MAS than did the more conservative threshold of $P_e \leq 0.02$. In this study, we cannot draw conclusions about the opti-

mum type I error rate to be used for the construction of a selection index, because only two different significance thresholds have been used. However, we believe that the choice of type I error rate warrants further research.

**Variation of cross-validation-derived estimates of $p_{TS}$ and bias:** As pointed out earlier, estimates of $h^2_{DS}$, $\ddot{p}_{TS}$, and the bias varied tremendously among DS. Most authors use the coefficient of determination from regression $R^2$ or $R^2_{adj}$ (DRAPER and SMITH 1981) to present results from QTL mapping studies and give an indication of the phenotypic variance explained by markers. However, the proportion of the phenotypic variance explained by markers is a function of the allocation of resources and the trait under study. To obtain results comparable across experiments with a varying number of test environments, different sample sizes, and different traits, the proportion of genotypic variance explained ($p$) is more appropriate. For each DS and each trait, the heritability was calculated and used for obtaining estimates of $p_{TS}$ ($\hat{p}_{TS} = \hat{R}^2_{adj}/\hat{h}^2_{DS}$). Because $\hat{R}^2_{adj}$ and $\hat{h}^2_{DS}$ are both subject to sampling errors, estimates of $p_{TS}$ beyond theoretical boundaries [0, 100] can occur. To warrant accurate estimation of $\bar{p}_{TS}$, values of $\hat{p}_{TS}$ exceeding theoretical boundaries were accepted (ALLISON *et al.* 2002). In the experimental data of this study, estimates of $h^2_{DS}$ for grain yield and grain moisture showed a large variation among DS as a consequence of sampling. For grain yield with a plot heritability of $h^2 = 0.085$, even four test environments ($E = 4$) were insufficient to obtain estimates of $\ddot{p}_{TS}$ within theoretical boundaries [0, 100] in all DS when the population size was low ($N = 122$). If individual DS with extreme estimated values of $\ddot{p}_{TS}$ occur, the mean over all DS ($\bar{p}_{TS}$) is affected and the distribution of $\ddot{p}_{TS}$ is skewed toward larger values. This circumstance is reflected in the difference between the mean ($\bar{p}_{TS}$) and the median ($\tilde{p}_{TS}$) proportion of genotypic variance explained by QTL for grain yield and small $N$ and $E$ (Figure 2). On the basis of these findings we conclude that cross-validation yields best results when a minimum sample size ($N > 200$) and a minimum number of test environments ($E > 4$) are available for analysis. Hence, QTL experiments need to be designed with special consideration of the population size and the number of test environments depending on the genetic architecture and the heritability of the trait.

The sampling error of the heritability estimates also affected estimates of the magnitude of the bias. Since the same estimate of $h^2_{DS}$ is used for calculation of $\hat{p}_{ES}$ and $\hat{p}_{TS}$, the difference between the genotypic variance explained in ES as compared to TS was also inflated for DS with low $h^2_{DS}$ estimates, but the effect was not as pronounced as for estimates of $p_{TS}$. As expected, the population size had an effect not only on the average bias but also on the range of the bias for different DS (Figure 3). This needs to be taken into account when evaluating the prospects of MAS. When MAS for a specific trait is tested, the success of MAS is predicted on the basis of results from a QTL mapping study generally performed with one or few segregating populations and compared to the actual selection gain achieved in another independent population from the same or a different cross subjected to MAS. The difference between the predicted and the realized selection gain corresponds to the difference between $\ddot{p}_{ES}$ and $\ddot{p}_{TS}$ in this study and therefore must be attributed to the bias of estimating the genotypic variance explained in the QTL mapping experiment. As can be seen in Figure 3, there are quite a few DS with bias of 100%, where QTL explaining a large proportion of the genotypic variance could be identified in ES but no gain from selection would be realized in the TS sample. This was especially pronounced for grain yield. In other DS, however, a considerable proportion of the variance explained by markers in ES could also be found in TS and would result in gain from selection if the TS had been used for MAS. With LOD 3.21 there were quite a few DS with zero bias for all traits, indicating that the entire genotypic variance explained in the mapping experiment would be useful in selection. This has to be interpreted with caution, however, because it could be the result of ES with 0 QTL detected. With LOD 2.5 and grain moisture and plant height it can be seen that the minimum bias for most combinations of $N$ and $E$ was ~10%. This means that, depending on the sample used for mapping, the maximum amount of genotypic variance explained by markers usable for selection would be only 10% less than the genotypic variance explained in the mapping experiment. These findings explain the controversial results from published experiments on MAS. Those MAS experiments that were successful might have estimated QTL effects with little or no bias in the mapping experiment. Those that were not successful might have had biased QTL estimates or many false positive QTL due to sampling. Increasing $N$ and $E$ helps to reduce the average bias and its range and, thus, provides a more realistic assessment of the prospects of MAS. Moreover, it also increases $\bar{p}_{TS}$ and consequently the efficiency of MAS as compared to phenotypic selection.

**Choice of resampling method:** The necessity for correction of bias in estimates of QTL effects has been shown in this study and has been pointed out by several researchers performing simulation studies on QTL mapping and the relative efficiency of MAS as compared to phenotypic selection. BEAVIS (1994) suggested the use of resampling methods for a realistic assessment of the prospects of MAS and to obtain unbiased estimates of QTL effects and of the proportion of genotypic variance explained by QTL. On the basis of experimental results UTZ *et al.* (2000) proposed three different cross-validation schemes for assessing the effect of environmental and genotypic sampling on QTL estimation. In this

study, the properties of cross-validation-derived estimates of the genotypic variance explained by QTL have been investigated using computer simulation data. For the trait plant height, 21 QTL (list of QTL can be retrieved from http://www.maizegdb.org) were assumed to be the "true" QTL explaining 56% of the genotypic variance. Six combinations of varying population size and number of test environments were evaluated. The good agreement between estimates of $\bar{p}_{TS}$ and $\bar{p}_{SD:ES}$ indicated that standard CV yielded almost unbiased estimates of the true genotypic variance explained by QTL ($p$) even for moderate sample sizes. Standard fivefold CV showed slightly lower power for QTL detection in ES as compared to DS due to the 20% fewer individuals used for QTL estimation but on average the loss of power in QTL estimation from DS to ES was reflected in only a slight underestimation of $\bar{p}_{SD:DS}$ by $\bar{p}_{TS}$. In CV/GE the underestimation of the true genotypic variance explained by QTL was more pronounced, especially for small $N$ and $E$. This can be attributable to QTL $\times$ environment interactions. The sum of all estimated QTL $\times$ environment interaction effects adds up to zero and, therefore, these estimates are not stochastically independent. With few test environments, the correlation between effects cannot be ignored.

A comparison of our results from CV with other resampling methods yielded similar findings (MELCHINGER *et al.* 2003). The first 100 sampled DS were analyzed with three different bootstrapping (BS) methods: (i) standard BS with bias correction as in EFRON and TIBSHIRANI (1993), (ii) bias estimation as in BREIMAN and SPECTOR (1992), and (iii) leave-one-out BS or 0.368 BS as in EFRON (1983). Results from all three methods showed underestimation of the true genotypic variance explained by QTL similar to CV/GE. In addition to its preferable statistical properties, standard CV is computationally less resource demanding than CV/GE and BS. We therefore recommend standard CV for analysis of QTL mapping data to obtain asymptotically unbiased estimates of the true QTL effects and the genotypic variance explained by markers.

**Recommendations for QTL mapping experiments:** Our results from cross-validation of experimental data agreed well with the results from simulation experiments on MAS (BEAVIS 1998; MOREAU *et al.* 1998) concerning the effect of population size, number of test environments, and LOD threshold. When performing QTL mapping experiments to identify QTL and assess the prospects of MAS, attention must be given to the genetic architecture of a quantitative trait. For resistance or quality traits, a model assuming few (<10) QTL with relatively large effects and high probability of explaining a considerable proportion of the genotypic variance with markers might be suitable. For polygenic traits like grain yield, however, FISHER's (1918) infinitesimal model seems more realistic especially when analyzing progenies of elite germplasm. In the latter case, it is questionable whether >60% of the genotypic variance can be explained by markers with reasonable input because only QTL with small effects are likely to be segregating while at QTL with large effects favorable alleles are expected to be fixed. As a consequence, pure MAS for complex traits without additional phenotypic selection does not seem promising.

The proportion of genotypic variance that can be explained by markers is trait specific. On the basis of our results, some general recommendations can be given for QTL mapping studies to maximize the proportion of genotypic variance explained:

i. With limited resources, adding more genotypes is more efficient than replicating the same genotypes.

ii. Depending on the trait of interest, a minimum number of environments is necessary ($E > 4$ with unreplicated trials) to allow for reliable estimation of $h^2$.

iii. Optimization of the population size to be used in QTL mapping experiments has to be trait specific depending on the genetic architecture and plot heritability of the trait as well as the total resources available. Increasing population size from 488 to 976 was beneficial and had a relatively large effect on the amount of genotypic variance explained, although the impact of increasing the population size was more dramatic for smaller $N$.

iv. The choice of significance threshold depends on the population size and also on the trait of interest. If the aim of a study is to identify the few large QTL regulating a limited proportion of the genetic variance, a more conservative threshold is recommended because the frequency of QTL detection is correlated with the size of QTL effects and the reliability of finding the large QTL is improved.

For traits regulated by a few QTL with large ($>0.2\sigma_P$) effects, for which phenotypic selection is expensive or hampered due to rare occurrence in the field, MAS can be efficiently used. The finesse of the breeder will be to find the optimum allocation of resources for detecting QTL and to obtain a realistic assessment of the genotypic variance explained by them for combining MAS with phenotypic selection.

## LITERATURE CITED

ALLISON, D. B., J. R. FERNANDEZ, H. MOONSEONG, Z. SHANKUAN, C. ETZEL *et al.*, 2002 Bias in estimates of quantitative-trait-locus effect in genome scans: demonstration of the phenomenon and a method-of-moments procedure for reducing bias. Am. J. Hum. Genet. **70:** 575–585.

BEAVIS, W. D., 1994 The power and deceit of QTL experiments: lessons from comparative QTL studies. 49th Annual Corn and Sorghum Industry Research Conference. ASTA, Washington, DC, pp. 250–266.

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.

BENNEWITZ, J., N. REINSCH and E. KALM, 2002 Improved confidence

intervals in quantitative trait loci mapping by permutation bootstrapping. Genetics **160:** 1673–1686.

BOST, B., D. DE VIENNE, F. HOSPITAL, L. MOREAU and C. DILLMANN, 2001 Genetic and nongenetic bases for the L-shaped distribution of quantitative trait loci effects. Genetics **157:** 1773–1787.

BOUCHEZ, A., F. HOSPITAL, M. CAUSSE, A. GALLAIS and A. CHARCOSSET, 2002 Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. Genetics **162:** 1945–1959.

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

DOERGE, R. W., and G. A. CHURCHILL, 1996 Permutation tests for multiple loci affecting a quantitative character. Genetics **142:** 285–294.

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

EFRON, B., 1983 Estimating the error rate of a prediction rule: improvement on cross-validation. J. Am. Stat. Assoc. **78** (382): 316–330.

EFRON, B., and R. J. TIBSHIRANI, 1993 *An Introduction to the Bootstrap.* Chapman & Hall, New York.

FALCONER, D. S., and T. F. C. MACKAY, 1996 *Introduction to Quantitative Genetics.* Longman, London.

FISHER, R. A., 1918 The correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edinb. **52:** 399–433.

FRIDMAN, E., T. PLEBAN and D. ZAMIR, 2000 A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc. Natl. Acad. Sci. USA **97** (9): 4718–4723.

FRISCH, M., M. BOHN and A. E. MELCHINGER, 2000 Plabsim: software for simulation of marker-assisted backcrossing. J. Hered. **91:** 86–87.

GÖRING, H. H. H., J. D. TERWILLIGER and J. BLANGERO, 2001 Large upward bias in estimation of locus-specific effects from genomewide scans. Am. J. Hum. Genet. **69:** 1357–1369.

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.

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, OR.

HOLM, S., 1979 A simple sequentially rejective multiple test procedure. Scand. J. Stat. **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.

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

KNAPP, S. J., and W. C. BRIDGES, 1990 Using molecular markers to estimate quantitative trait locus parameters: power and genetic variances for unreplicated and replicated progeny. Genetics **126:** 769–777.

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.

LANDE, R., and R. THOMPSON, 1990 Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics **124:** 743–756.

LYNCH, M., and B. WALSH, 1998 *Genetics and Analysis of Quantitative Traits.* Sinauer Associates, Sunderland, MA.

MELCHINGER, A. E., H. F. UTZ and C. C. SCHÖN, 2003 New developments in the biometric analysis of complex traits. Proceedings of 12th Meeting of the EUCARPIA Section Biometrics in Plant Breeding. September 3–5, 2003, A Coruña, Spain (in press).

MILLER, A. J., 1990 *Subset Selection in Regression.* Chapman & Hall, London.

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. 52nd Annual Corn and Sorghum Industry Research Conference. ASTA, Washington, DC, pp. 44–53.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

SEARLE, S. R., 1971 *Linear Models.* Wiley & Sons, New York.

SNEDECOR, G. W., and W. G. COCHRAN, 1989 *Statistical Methods.* Iowa State University Press, Ames, IA.

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. July 6–8, 1994, CPRO-DLO, Wageningen, The Netherlands.

UTZ, H. F., and A. E. MELCHINGER, 1996 PLABQTL: a program for composite interval mapping of QTL. J. Quant. Trait Loci **2** (1): 1–5.

UTZ, H. F., A. E. MELCHINGER and C. C. SCHÖN, 2000 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. Genetics **154:** 1839–1849.

YOUSEF, G. G., and J. A. JUVIK, 2001 Comparison of phenotypic and marker-assisted selection for quantitative traits in sweet corn. Crop Sci. **41:** 645–655.

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

Communicating editor: R. W. DOERGE