

The Genetic Architecture of Response to Long-Term Artificial Selection for Oil Concentration in the Maize Kernel

Cathy C. Laurie,^{*,1} Scott D. Chasalow,^{*} John R. LeDeaux,^{*} Robert McCarroll,^{*} David Bush,^{*} Brian Hauge,^{*} Chaoqiang Lai,^{*} Darryl Clark,[†] Torbert R. Rocheford[†] and John W. Dudley[†]

^{*}Monsanto Company, St. Louis, Missouri 63167 and [†]Department of Crop Sciences, University of Illinois, Urbana, Illinois 61801

Manuscript received April 6, 2004
Accepted for publication August 16, 2004

ABSTRACT

In one of the longest-running experiments in biology, researchers at the University of Illinois have selected for altered composition of the maize kernel since 1896. Here we use an association study to infer the genetic basis of dramatic changes that occurred in response to selection for changes in oil concentration. The study population was produced by a cross between the high- and low-selection lines at generation 70, followed by 10 generations of random mating and the derivation of 500 lines by selfing. These lines were genotyped for 488 genetic markers and the oil concentration was evaluated in replicated field trials. Three methods of analysis were tested in simulations for ability to detect quantitative trait loci (QTL). The most effective method was model selection in multiple regression. This method detected ~50 QTL accounting for ~50% of the genetic variance, suggesting that >50 QTL are involved. The QTL effect estimates are small and largely additive. About 20% of the QTL have negative effects (*i.e.*, not predicted by the parental difference), which is consistent with hitchhiking and small population size during selection. The large number of QTL detected accounts for the smooth and sustained response to selection throughout the twentieth century.

THE genetic architecture of a quantitative trait consists of a set of parameters that explain the genetic component of trait variation within or among populations. These parameters include the number of quantitative trait loci (QTL) affecting the trait, their locations in the genome, the frequencies of alternative genotypes segregating at the QTL, the pattern of linkage disequilibrium among QTL, and the magnitudes of additive, dominance, and epistatic effects. Knowledge of genetic architecture has applications in two areas: (1) the identification of genes with utility in agriculture and/or treatment of disease and (2) making inferences about the evolutionary processes that maintain genetic variation and those that cause divergence between populations. Here we report a study of oil variation in maize that has both types of application.

The kernels of a modern maize (*Zea mays* L.) hybrid typically contain ~4% oil, 9% protein, 73% starch, and 14% other constituents (mostly fiber). The caloric content of oil is 2.25 times greater than that of starch on a weight basis and livestock feeding studies have shown a greater rate of weight gain per pound of feed for high-oil (>7%) than for normal maize (reviewed by LAMBERT 1994 and LAMBERT *et al.* 2004). Therefore, high-oil maize is in demand as a source of animal feed, which is the primary use of maize grown in developed coun-

tries. An objective of our study is to identify genes that may be used to increase the oil concentration of maize kernels through plant breeding or genetic engineering.

The experiment reported here originated in 1896 when C. G. Hopkins began the Illinois long-term selection lines, which have become a “textbook” example of the power of artificial selection (see review by DUDLEY and LAMBERT 2004). From an open-pollinated variety of maize, Hopkins started two populations that were selected divergently for the percentage of kernel dry weight that consists of oil (“oil concentration” or “percentage of oil”). These populations are called Illinois high oil (IHO) and Illinois low oil (ILO). In each generation and each population, bulked kernels from each of a number of ears (half-sib families) were analyzed and the highest (or lowest) 20% of ears were selected to be parents of the next generation. This selection was carried on throughout the twentieth century at the University of Illinois and continues to this day for IHO with no sign of a plateau in response. At generation 89, selection in ILO was discontinued because of poor viability and an oil concentration so low that it cannot be measured accurately. At this point, the populations had changed from 4.7% oil to 19.3% in IHO and 1.1% in ILO.

Currently, the oil concentration of IHO is ~20%, which is far greater than levels in commercial germplasm. Unfortunately, the yield and other agronomic characteristics of IHO are poor (DUDLEY *et al.* 1974; LAMBERT 1994), so the population itself is not used in

¹Corresponding author: 14031 Shadow Oaks Way, Saratoga, CA 95070.
E-mail: cathylaurie@pqgen.com

commercial production. However, the genes that confer high oil in IHO may have utility in two ways. First, it might be possible to separate the beneficial effects on oil from the negative effects on agronomic characteristics by focused introgression of high-oil alleles into high-yielding germplasm through plant breeding. Second, identification of the genes that actually cause oil variation may provide good candidates for genetic engineering of high-yielding germplasm through plant transformation. If variation in the sequence of a gene causes variation in oil, it suggests that the system of oil production and storage in maize is sensitive to the structure and/or amount of the corresponding gene product. Therefore, even if the natural alleles of these genes have small effects, it is possible that artificial changes in expression or amino acid sequence can produce larger effects of utility in agriculture.

Here we report on QTL identification and characterization in a population derived from the Illinois long-term selection lines. At generation 70 of selection, several individuals each of IHO and ILO (both polymorphic populations) were crossed to produce a hybrid population, which was randomly mated for 10 generations (RM10). The RM10 individuals were selfed to produce 500 RM10:S1 lines that constitute the study population. These lines were genotyped for a set of 488 genetic markers and oil was evaluated in replicated field trials in the partially inbred lines and in hybrids produced by crossing each line to an inbred tester. This design provides considerable power and resolution for QTL detection and localization because the sample size is relatively large, the trait was evaluated with replication on a line basis (*i.e.*, family mean rather than individual measurements), and the 10 generations of random mating provide multiple opportunities for recombination.

Standard methods of linkage analysis for quantitative traits, such as QTL interval mapping, are not directly applicable to this population because parental phases of the S1 individuals are unknown. However, this population is quite suitable for linkage disequilibrium (LD) mapping and we have analyzed the experiment as an association study (LANDER and SCHORK 1994). In this case, a single admixture event between IHO and ILO created LD between genes with different allelic frequencies. The subsequent 10 generations of random mating eliminated essentially all associations between unlinked markers and most of those between loosely linked markers, yet retained associations between closely linked marker/QTL pairs.

The study presented here has the following major objectives: (1) select a statistical method appropriate for multiple-QTL identification in an association study and evaluate its performance in simulations; (2) describe the genetic architecture of oil variation in sufficient detail to provide insight into the population genetic processes involved in sustained response to long-term artificial selection; and (3) identify candidate genes or

chromosomal regions that may be useful for increasing oil concentration in commercial maize germplasm.

MATERIALS AND METHODS

Notation: r_{LD}^2 is a measure of linkage disequilibrium, R_{TM}^2 is the coefficient of multiple determination for regression of trait value on the genotype scores of multiple markers, and R_{MM}^2 is the coefficient of multiple determination of the regression of the genotype score of one marker on the genotype scores of multiple other markers.

Germplasm and field trials: The association study population originated from a cross between IHO and ILO performed at generation 70 of selection when the oil concentrations were estimated for IHO as 16.7% and for ILO as 0.4% (DUDLEY and LAMBERT 2004). Approximately five to seven individuals of each parental population were crossed to produce a hybrid population, which was randomly mated for 10 generations. During random mating, ~200 plants per generation were used (100 as male and 100 as female). One RM10:S1 line was derived from each of 500 different RM10 plants in the following way. Each RM10 plant was selfed to produce an ear of S1 kernels (S1 ear). The kernels from one S1 ear were planted together in a single row and the plants were selfed to produce ears containing S2 kernels (S2 ears). One S2 ear was selected at random to represent an RM10 plant. The kernels on this S2 ear (and their progeny) constitute an RM10:S1-derived line. Kernels from each of the S2 ears were divided into two parts: 50 kernels to be used for DNA extraction and the remaining kernels (~50) for additional crosses to produce material for phenotypic analysis. For each line, the S2 kernels for phenotypic analysis were planted adjacent to a stiff stalk tester (Monsanto 7051) and two sets of crosses were made. The tester was fertilized with pollen from the S2 plants to produce hybrid offspring and the S2 plants were intermated within each S1 family to produce inbred offspring. The inbred and hybrid kernels from each of 500 lines were planted in replicated field trials.

The field layout was an $\alpha(0, 1)$ incomplete block design (PATTERSON and WILLIAMS 1976) with two replicates per location. Each replicate consisted of one row each of 500 lines arranged in 50 blocks of 10 lines each. The $\alpha(0, 1)$ design specifies that each line occurs together in a block with any other line either 0 or 1 times. To produce kernels for compositional analysis, ~10 sib-matings were made within each row, with one pollen parent per mating and each plant used as pollen parent only once. This procedure yielded six to eight ears per line per replicate, from which kernels were bulked for compositional analysis. Three locations were used in each of 2 years, for a total of 12 replicates each for inbred and hybrid seed. The locations were Urbana, Illinois; Monmouth, Illinois; and Williamsburg, Iowa.

Phenotypic trait measurements: Kernel composition was estimated by near-infrared spectroscopy (DYER and FENG 1997). For hybrids, whole-kernel samples were analyzed at Monsanto by near-infrared transmittance with an Infratec Grain Analyzer instrument. For the inbred samples, kernels were dried to a relatively uniform moisture level, ground, and then analyzed at the University of Illinois by near-infrared reflectance using a Dickey-John instrument (HYMOWITZ *et al.* 1974; DUDLEY and LAMBERT 1992). Both methods provide estimates of the percentage of kernel dry weight that consists of oil.

If there were no missing samples, the total number of replicates for each line would be 12. The mean (SD) replicate numbers are 11.9 (0.3) for hybrid and 11.0 (1.3) for inbred lines.

Markers and genotyping: For each RM10:S1 line, 50 S2 kernels were germinated and grown in the greenhouse. Then all

50 seedlings were lyophilized and bulked together for DNA extraction. The genotypes of individual S1 plants were inferred from this bulk DNA sample. In addition, DNA was extracted from individual plants of IHO and ILO of generation 70 (parents of the hybrid population). DNA was extracted by a standard procedure (DELLAPORTA *et al.* 1983).

DNA samples were genotyped for a set of biallelic single-nucleotide polymorphism (SNP) markers using the 5' nuclease Taqman assay described by LIVAK (2003). For quality control, standards of known genotype (six each of the three types) and six negative controls were included in every set of 192 samples analyzed. Such standards were used to estimate the accuracy of genotype calls in production runs at 0.996 ($n = 137,115$).

The markers assayed in this experiment are from a set of SNPs discovered in the parents of three mapping populations that are unrelated to the Illinois germplasm. The genetic map distances provided are from a Monsanto reference map, which is based on a composite of information from these three populations.

The primer and probe sequences for the Taqman genotyping assays are available on request of J. R. LeDeaux at Monsanto. Access requires agreement that the sequences will be used only for noncommercial research and will not be transferred to a third party.

To maximize the chances of detecting marker-trait associations, markers run on the association study population were selected to have a large allelic frequency difference between the parental populations, IHO and ILO. These frequencies were estimated from 25 individuals each of IHO and ILO and markers were selected that have a frequency difference of at least 0.60 or, if one parental population was fixed, a difference of at least 0.40. This selection resulted in 472 markers ($\sim 30\%$ of those tested), to which 16 markers with smaller frequency differences were added to fill some gaps in the genetic map. This brings the total to 488 markers (in 470 different genes), of which 440 are nonredundant in the sense that all pairs have $r_{LD}^2 < 0.99$, where r_{LD}^2 is defined below. The mean (SD) of the allelic frequency differences for the 440 nonredundant markers is 0.76 (0.22). The expected heterozygosity in the RM10:S1 population is high at most loci: the 0th, 25th, 50th, 75th, and 100th percentiles of the distribution of $2p(1-p)$, where p is the allelic frequency, are 0.04, 0.37, 0.45, 0.49, and 0.50, respectively.

Linkage disequilibrium estimation: A maximum-likelihood procedure was used to estimate linkage disequilibrium (LD) in the RM10:S1 from unphased diploid genotypes. Solutions to this problem for randomly mating populations have been well studied (WEIR and COCKERHAM 1979; EXCOFFIER and SLATKIN 1995; FALLIN and SCHORK 2000). We modified this approach to estimate LD for individuals obtained by selfing from a randomly mated population. For details, see Section A and Figure S1 of supplemental materials (<http://www.genetics.org/supplemental/>). Results are presented for a measure of LD, r_{LD}^2 , defined as follows. Consider two loci, locus A with alleles A and a, and locus B with alleles B and b. The allelic frequencies are P_x (where x represents A, a, B, or b), the frequency of the haplotype AB is P_{AB} , and $D = P_{AB} - P_A P_B$. Then $r_{LD}^2 = D^2 / (P_A P_a P_B P_b)$.

Imputation of missing genotypes: In the association study, 500 lines were genotyped for 488 SNP markers. One line had an excessive number of missing genotypes and was dropped from the analysis. Among the remaining 499 lines, 2.3% of scores were missing. The mean (SD) of the number of scores for each line is 477 (16).

Missing genotypes were imputed by using information from associated markers. To impute a missing genotype at locus A in individual X, we first identified the locus having the highest

level of association with A that is not missing in X (say locus B). Then the observed two-locus genotypic frequencies for loci A and B were used to estimate the frequencies of the three A locus genotypes conditional on the B locus genotype of individual X. The most frequent A locus genotype (conditional on B) was designated as the missing genotype. The complete genotypic data matrix with 97.7% observed and 2.3% imputed calls is referred to as genotypic matrix C.

Analysis of variance of phenotypic data: The model for analysis of variance of the phenotypic observations is

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_{k(j)} + \delta_{l(kj)} + \epsilon_{ijkl}, \quad (1)$$

where α_i is the effect of the i th line, β_j is effect of the j th combination of location and year (locyr), $(\alpha\beta)_{ij}$ is the line \times locyr interaction, $\gamma_{k(j)}$ is the effect of the k th replicate (rep) nested within locyr, $\delta_{l(kj)}$ is effect of the l th block nested within locyr \times rep, and ϵ_{ijkl} is the residual. All effects are considered to be random. Hybrid and inbred observations were analyzed separately. ANOVA of the ~ 6000 observations for each trait (500 lines \times 3 locations \times 2 replicates) was done using PROC MIXED in SAS with the REML option (SAS INSTITUTE 1999). This analysis provides variance component estimates and best linear unbiased predictors (BLUPs) of the line effects. Likelihood-ratio statistics were calculated to test the significance of the line \times locyr interaction and the main effect of lines, as described by LITTELL *et al.* (2002, p. 103).

A broad-sense heritability H^2 was estimated as the fraction of the variation among line means within an experiment that is due to genetic effects,

$$H^2 = \frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}_\alpha^2 + (1/b)\hat{\sigma}_{\alpha\beta}^2 + (1/bc)\hat{\sigma}_\epsilon^2}, \quad (2)$$

where $\hat{\sigma}^2$ is a variance component estimate, $b = 6$ location \times year combinations, and $c = 2$ replicates within each location \times year combination.

The relationship between phenotype and genotype was analyzed using the predicted line means (BLUP plus intercept from the random-model ANOVA).

Marker-trait associations: Initially, one-way ANOVAs were done to analyze the effect of each marker separately. The 500 predicted line means of one type (hybrid or inbred) were analyzed with the phenotypic value as dependent variable and the marker genotype (coded as a categorical variable with three levels) as independent variable. The significance of an additive effect was tested as a contrast between the two homozygous classes and that of a dominance effect as a contrast between the heterozygous class and the mean of the two homozygous classes. These single-marker ANOVA models were fit in S-PLUS with the "lm" function (S-PLUS 2000, 1999). Epistatic interactions between pairs of markers were tested in two-way ANOVA using PROC GLM of SAS.

The ANOVA indicated that dominance and epistatic interaction effects are very small compared with additive effects (see RESULTS). Therefore, subsequent analyses of marker-trait associations were done using regression analyses of strictly additive models (using PROC REG of SAS software). Three types of regression analyses, described in detail below, were done: (1) single-marker regression, (2) stepwise multiple regression with MAXR/BIC, and (3) covariate regression.

For stepwise regression, we used the option "MAXR" of PROC REG in SAS, which is described in the following quotation (where " R^2 " is R_{TM}^2). "The MAXR method begins by finding the one-variable model producing the highest R^2 . Then another variable, the one that yields the greatest increase in R^2 , is added. Once the two-variable model is obtained, each of the variables in the model is compared to each variable not in the model. For each comparison, the MAXR method

determines if removing one variable and replacing it with the other variable increases R^2 . After comparing all possible switches, the MAXR method makes the switch that produces the largest increase in R^2 . Comparisons begin again, and the process continues until the MAXR method finds that no switch could increase R^2 . Thus, the two-variable model achieved is considered the 'best' two-variable model the technique can find. Another variable is then added to the model, and the comparing-and-switching process is repeated to find the 'best' three-variable model, and so forth" (SAS INSTITUTE 1999, p. 2948). MAXR was used to identify the best model of a given dimension (number of markers) for the range of 1–120 markers.

The best model dimension was selected by minimizing a criterion that is equivalent to a maximum likelihood with a penalty on model complexity. In general, the criterion is $-2\log \text{lik} + p\gamma$, where $\log \text{lik}$ is the maximized log-likelihood, p is the number of parameters in the model (the number of markers plus one for the intercept), and γ is a penalty factor. Here we used the Bayesian information criterion (BIC; SCHWARZ 1978; RAWLINGS *et al.* 1998): $\text{BIC} = n \ln(SS_R) + p \ln(n) - n \ln(n)$, where n is the sample size and SS_R is the residual sum of squares. In this case $\gamma = \ln(n) = \ln(499) = 6.2$.

The covariate regression analysis provides a test focused on each marker (say marker x) by regressing phenotypic value on marker x and a set of covariate markers. The covariate markers were selected to account for variation in the trait, but have small correlation with marker x , as follows. Start with a potential set of covariates consisting of the set of markers selected by the MAXR/BIC procedure. Do multiple regression of the genotypic score of marker x on the genotypic scores of the set of potential covariates. If the coefficient of multiple determination of this regression (R_{MM}^2) is >0.10 , remove the marker that contributes the most to R_{MM}^2 and perform the regression again. The process continues until $R_{MM}^2 < 0.10$. The covariate regression procedure is similar in concept to composite interval mapping (JANSEN and STAM 1994; ZENG 1994).

Simulations: All simulated data sets are based on genotypic data matrix C (*i.e.*, on the actual, not simulated, genotypes). The phenotypic value of the i th individual ($i = 1, \dots, 499$) was simulated as

$$Y_i = \mu + \sum_{j=1}^m \beta_j X_{ij} + \varepsilon_i, \quad (3)$$

where μ is the mean, m is the number of QTL, β_j is the effect of the j th QTL (partial regression coefficient), X_{ij} is the genotypic score at the j th QTL ($-1, 0$, or 1) and the ε_i are independent and identically distributed $N(0, \sigma^2)$. For each model, 100 replicate data sets were simulated and analyzed.

The simulation models of phenotypic value are based on three key results of applying MAXR/BIC selection to the observed data set for the inbred lines: (1) 50 markers selected, (2) a set of 50 partial regression coefficients, and (3) a residual variance of 0.43. We refer to this set of results as model A. In all simulation models, the residual variance (σ^2) equals 0.43. Four different types of models were simulated (sets B, C, D, and E)—see Table 1. In all cases, a set of n markers ($n = 40, 50$, or 60) was chosen at random from the total set of 440 to be QTL. In models of type E, no further constraints were imposed. In types B, C, and D, random marker sets were rejected unless they met certain constraints regarding associations between QTL and/or the magnitude of the variance of phenotypic values. In models of type B, only 50 QTL were simulated and the effects of those QTL were selected at random without replacement from the set of 50 coefficients from model A. For all other models, the effects were assigned to QTL by random sampling with replacement from the coefficients of model A.

Each simulated data set was analyzed by single-marker regression, MAXR/BIC, and covariate regression. In each case, the analyses included m markers that are assigned to be true QTL ($m = 40, 50$, or 60) and $440 - m$ markers that are not QTL (but may be in linkage disequilibrium with QTL). For each simulated data set and each method of analysis, the best set of k markers was selected. In the case of single-marker and covariate regressions, the best set consists of the k markers with lowest P -value (in the t -test that the regression coefficient equals zero). In the case of MAXR, the best set consists of the k markers that maximize R_{TM}^2 according to the MAXR algorithm. A total of 16 values of k were used in the analysis of each simulated data set. These values were set by 15 different P -value thresholds and the BIC stopping rule. The P -value thresholds are 0.00001 (the Bonferroni criterion with $\alpha = 0.05$), 0.00005, 0.0001, 0.0005, 0.001, 0.003, 0.006, 0.01, 0.015, 0.02, 0.025, 0.03, 0.04, 0.05, and 0.10. The number of markers having a P -value from single-marker regression less than or equal to the threshold determined the value of k . Thus, for each simulated data set, 48 sets of markers were selected (3 methods \times 16 values of k).

In the simulations, all 440 markers analyzed were classified into one of five categories: (a) the marker is a true QTL, it was selected, and the estimated regression coefficient has the same sign as the effect of that QTL in the model; (b) the marker is a true QTL and it was selected, but the coefficient has the wrong sign; (c) the marker is a true QTL and was not selected; (d) the marker is a non-QTL and was selected; or (e) the marker is a non-QTL and was not selected. Let N_x be the number of markers in category x .

The quality of selected markers was assessed in four ways:

1. The fraction of QTL selected was calculated as $N_a/(N_a + N_b + N_c)$.
2. The fraction of markers selected that are considered as non-QTL was calculated as $(N_b + N_d)/(N_a + N_b + N_d)$. The contribution of N_b to this quantity is generally very small unless the total number selected is very large. For example, the average of $N_b/(N_b + N_d)$ is $<0.6\%$ for all three methods of regression when the number of markers selected is specified by BIC.
3. The distance between each selected marker and the nearest QTL on the genetic map was calculated. This distance is zero when the marker is a QTL.
4. The degree of association between each selected marker and the set of QTL was estimated as the R_{MM}^2 of the regression of the genotypic score of that marker on the genotypic scores of the QTL. The value of R_{MM}^2 is one when the marker is a QTL.

RESULTS

Linkage disequilibrium: Several different measures of two-locus LD have been described (DEVLIN and RISCH 1995; WEIR 1996). In the context of detecting marker-trait associations, we prefer the r_{LD}^2 measure because of its relationship to the additive genetic variance. In a randomly mating population, the additive genetic variance associated with a neutral marker equals the product of r_{LD}^2 (between marker and QTL) and the additive genetic variance due to the QTL. This relationship follows algebraically from equations for the additive genetic variance provided by NIELSEN and WEIR (1999). The RM10:S1 population studied here was derived by one generation of selfing from a randomly mating popu-

TABLE 1
Summary of simulation models

Model features	Model type			
	B	C	D	E
No. of QTL	50	40, 50, or 60	40, 50, or 60	40, 50, or 60
Regression coefficients sampled with replacement	No	Yes	Yes	Yes
Constraint on variance of predicted values	Yes ^a	Yes ^a	No	No
Constraint on associations between QTL pairs	Yes ^b	No	Yes ^c	No

See MATERIALS AND METHODS for similarities among the models and further explanation of differences.

^a The variance of predicted phenotypic values was constrained to be within 5% of the variance for model A (0.75).

^b The intrachromosomal distribution of LD between QTL pairs was constrained to match that of model A within bins of 0.1 (which has the effect of reducing the level of LD between QTL compared with randomly selected sets).

^c No pair of QTL was allowed to have $r_{LD}^2 \geq 0.25$.

lation. Although the relationship between additive genetic variance of a marker and that of an associated QTL does not hold exactly for this population, numerical calculations show that it is approximately correct (data not shown). If we assume that the pattern of r_{LD}^2 values between markers is similar to that between markers and QTL, the former may serve as an indication of our ability to detect QTL through marker associations.

The r_{LD}^2 was estimated for all pairs of the 488 markers run on the association study population. For the 110,529 pairs of unlinked markers (*i.e.*, on different chromosomes or >50 cM apart), the 0th, 25th, 50th, 75th, and 100th percentiles of the distribution of r_{LD}^2 are 0, 0.0002, 0.001, 0.003, and 0.05. So, there is very little, if any, LD between unlinked markers, as expected for a population that has undergone 10 generations of random mating. Therefore, any observed marker-trait association is almost certainly due to linkage between the marker and a QTL.

Figures 1 and 2 show the relationship between r_{LD}^2 and genetic map distance between pairs of linked markers (≤ 50 cM apart). LD declines with map distance, as expected, and is very small for markers >20 cM apart (maximum is 0.07). The median map distance between loci with very high r_{LD}^2 values (0.75–1.00) is 0 cM and 75% of all such r_{LD}^2 values occur between loci within 1.1 cM of each other. Among all pairs of linked markers, only 7.1% have $r_{LD}^2 \geq 0.25$ and the median and third quartile of map distance for this group are 1 and 2.6 cM, respectively. Thus, if an $r_{LD}^2 > 0.25$ were required to detect a significant marker-trait association, most of those associations would indicate a marker-QTL distance of less than a few centimorgans.

Phenotypic data analysis: Analyses of variance of the raw observations of percentage of oil were performed for inbreds and hybrids separately. In both cases, the

main effect of lines is extremely significant ($P < 10^{-16}$). The line \times locyr (genotype-by-environment) interaction is also highly significant in both cases ($P = 4.7 \times 10^{-5}$ for hybrids and $P = 2.0 \times 10^{-14}$ for inbreds).

Variance component and broad-sense heritability estimates (Table 2) have three notable features:

1. Despite being highly significant, the magnitude of the variance component for the genotype-by-environment interaction is less than one-tenth that of the line component.
2. The variance components for hybrids are much smaller than those for inbreds, but the heritability is about the same. Evidently, the phenotype of hybrids

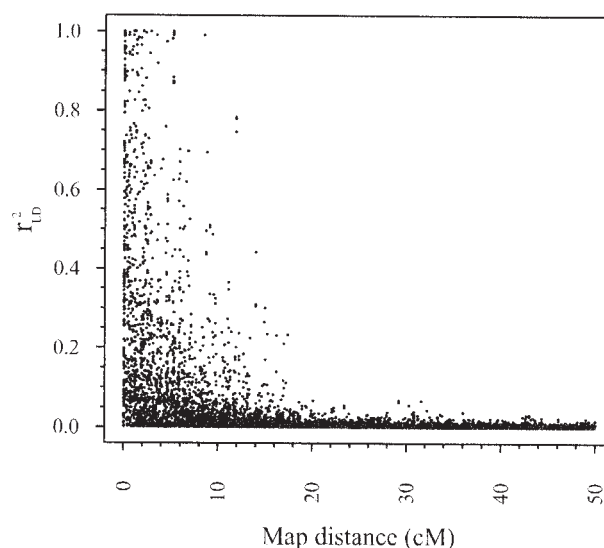


FIGURE 1.—Relationship between linkage disequilibrium (r_{LD}^2) and genetic map distance between 8299 linked marker pairs (<50 cM apart).

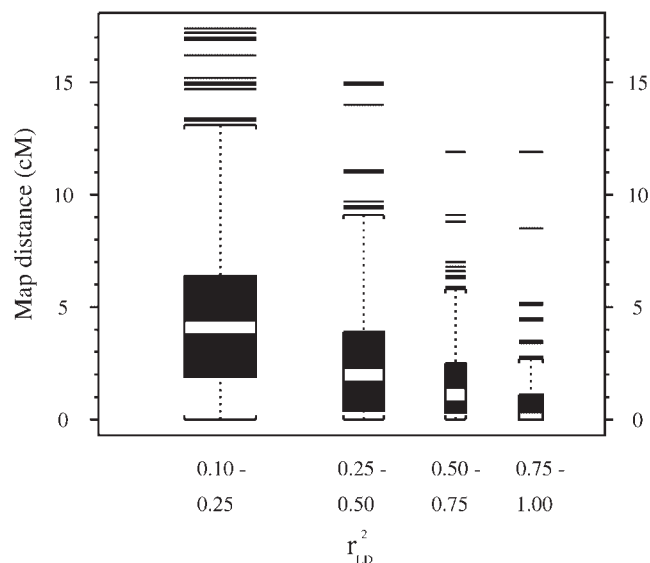


FIGURE 2.—Distribution of genetic map distance for linked marker pairs (<50 cM apart) that occur in four bins with respect to their r_{LD}^2 values. In this “box plot” format, the solid box extends from the first to the third quartiles (and thus contains 50% of the observations). The box is divided at the median by an open bar. A vertical line extends from the box to the greatest (and smallest) observation within 1.5 interquartile ranges. All observations outside these limits are plotted individually. The width of each box is proportional to sample size, which is 499, 281, 141, and 166 from left to right. The r_{LD}^2 bin from 0.0 to 0.10 (not shown) has 7212 pairs for which the first quartile equals 11.9 cM, the median equals 23.8 cM, and the third quartile equals 36.5 cM.

is more stable, but the total variance among line means is less, as expected since all hybrids have one of their parents in common.

3. The high heritability estimates (96% in both cases) are very favorable for detecting marker-trait associations.

The predicted line means of percentage of oil range

from 4.4 to 10.5% for inbreds and from 4.1 to 6.0% for hybrids. The corresponding ranges for most commercial inbreds and hybrids are from 3.4 to 5.4% and from 3.6 to 4.8%, respectively. This comparison indicates the potential of genes segregating in the study population to increase oil beyond the current norm of commercial germplasm.

Dominance and epistasis: The ordinary test for dominance is to contrast phenotypic values of the heterozygote with the mean of the two homozygotes. Such a test can be performed with the data from the RM10:S1 lines, but the interpretation is different from usual because the phenotypic data for each line are obtained from a pool of family members that will be a mixture of genotypes in families derived from a heterozygous parent. Therefore, the effect of any dominance that may exist is diluted. There is no true dominance in the hybrids because they are produced by crossing each line to a common inbred tester. Consequently, the heterozygous lines produce hybrids that are an equal mixture of the genotypes produced by the two types of homozygous lines.

A total of 440 *t*-tests for dominance, one for each marker, were performed with each set of phenotypic data (using the predicted line means). With $P < 0.05$, 26 tests were significant in the hybrids and 25 in the inbreds, with an overlap of 10 markers significant in both. The expected number of significant tests is 22 (5% of 440) if all tests are independent. Because of LD among markers, the tests are not all independent. Nevertheless, it appears that there is little or no evidence for dominance.

Two-way analyses of variance were used to detect interactions between pairs of markers. All pairs of 440 markers were tested in both inbreds and hybrids. In a total of 193,160 tests, 5.7% were significant at the 5% level, 1.2% at the 1% level, and 0.15% at the 0.1% level. These numbers are similar when the tests are limited to markers with significant additive effects on the trait.

TABLE 2
Variance components and summary statistics for percentage of oil in the kernel

Variance components	Inbreds		Hybrids	
	Estimate	95% C.I. ^a	Estimate	95% C.I. ^a
Locyr	1.05	0.40–6.92	0.029	0.011–0.191
Rep(locyr)	0.07	0.03–0.40	0.002	0.001–0.013
Block(locyr × rep)	0.20	0.18–0.24	0.010	0.009–0.013
Line	1.22	1.07–1.40	0.105	0.092–0.120
Locyr × line	0.10	0.08–0.13	0.004	0.003–0.008
Residual	0.47	0.44–0.50	0.049	0.047–0.052
Summary statistics				
Variance of BLUPs	1.16		0.10	
Heritability	0.96		0.96	
Mean	7.06		4.92	

^a C.I., confidence interval.

Because many of the pairwise tests are not independent, the expectation that a fraction α of tests are significant with P -value $< \alpha$ does not necessarily apply precisely. Nevertheless, these results show very little evidence for two-locus epistatic interactions.

The results described so far suggest little genotype-by-environment interaction, dominance, or epistasis. Therefore, subsequent analyses of marker-trait associations were done using regression analysis of the predicted line means with strictly additive models and marker genotypes coded as 1 for AA , 0 for Aa , and -1 for aa (where A is the allele with the higher frequency in IHO).

Single-marker regressions: A single-marker regression analysis was performed for each of 440 markers in both the hybrid and inbred data. A regression coefficient was estimated and the null hypothesis that the coefficient equals zero was evaluated with a standard t -test. Establishing a threshold for declaring significance of the t -test requires a consideration of the multiple testing issue. Here we use the “false discovery rate” (FDR) approach described by STOREY (2002) and STOREY and TIBSHIRANI (2003). The FDR is the expected proportion of null hypotheses declared false (*i.e.*, “significant”) that are not false. We use a threshold for significance that corresponds to an FDR of 5%, which means that we expect $\sim 5\%$ of the markers declared significant to be false leads (*i.e.*, not associated with QTL).

For the inbreds, 66 of 440 markers (15%) have an $FDR \leq 0.05$ (and $P \leq 0.014$). For the hybrids, 54 of 440 markers (12%) have an $FDR \leq 0.05$ (and $P \leq 0.010$). Thirty markers were significant in both inbreds and hybrids. The observation of a relatively large proportion of markers having significant association to the trait is not surprising, given that all markers were selected to show a large frequency difference between the parental populations, which are very divergent in trait value.

The additive effect for the trait associated with each marker is estimated by the regression coefficient. The magnitude of this effect is expected to be greater in inbreds than in hybrids because the contrast between homozygous lines in inbreds is AA vs. aa , whereas the corresponding contrast in hybrids is AA vs. Aa (for an AA tester) or Aa vs. aa (for an aa tester). Figure 3 shows that this is the case. Figure 3 also shows that the correlation between effects in inbreds and hybrids is high ($r = 0.75$, $P < 0.0001$). Similarly, the correlation between predicted line means in inbreds and hybrids is also high ($r = 0.73$, $P < 0.0001$). Consistency in effect between inbreds and hybrids is expected in the absence of dominance and epistasis.

Significant markers occur on all 10 chromosomes and many occur in clusters on the genetic map [Figure 4 and Figure S2 in supplemental materials (<http://www.genetics.org/supplemental/>)]. This clustering suggests that multiple markers may be associated with the same QTL.

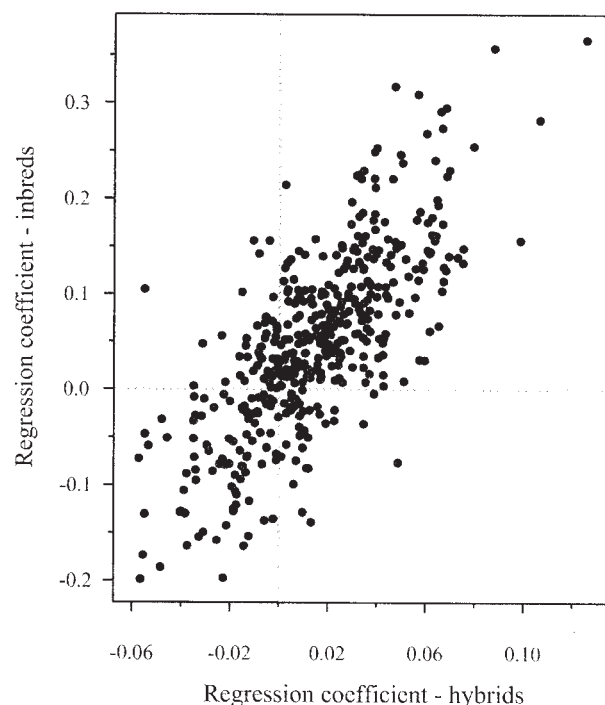


FIGURE 3.—Relationship between inbreds and hybrids with respect to the coefficient (slope) estimated by single-marker regression. Each point represents 1 of 440 markers.

MAXR stepwise regression: Model selection for multiple regression proceeded according to the following steps:

- A class of models was selected first—namely, linear models with only additive effects because of the weak evidence for nonadditive effects.
- The search for a set of models within this class was accomplished by the MAXR algorithm. Ideally, one would like to examine all possible combinations of a given number of regressors, but this is not feasible for the large number of potential regressors in this experiment. The MAXR method is a compromise between the exhaustive search and the more limited search provided by standard forward, backward, or stepwise methods (HOCKING 1976).
- Models of the same dimension (number of regressors) were compared and the one with maximum R_{TM}^2 was selected as the best model of a given dimension.
- Models with different dimensions were compared according to the value of the BIC and the model with the minimum BIC value was selected as the best model over all.

For the inbreds, 50 markers were selected with $r_{TM}^2 = 62\%$. Among the 50 coefficients, 38 are positive and 12 are negative. A positive coefficient indicates that the allele with higher frequency in IHO increases oil concentration. Therefore, a majority of coefficients are expected to be positive. For the hybrids, 39 markers were

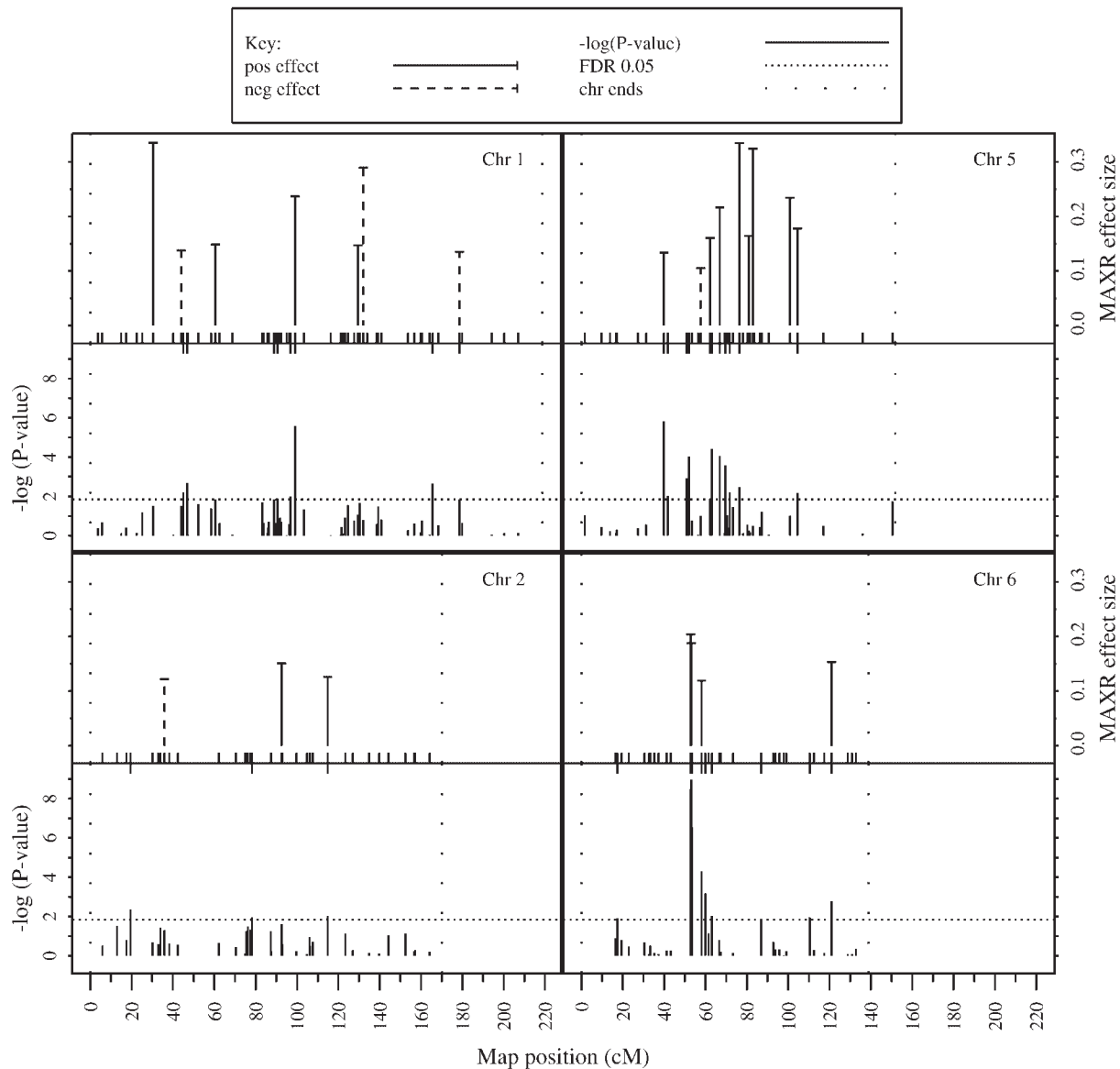


FIGURE 4.—Summary of single-marker regression and MAXR/BIC analysis of inbred line data. Each quadrant contains a pair of plots of data for one chromosome (“Chr”). The top plot shows the MAXR/BIC analysis and the bottom plot shows the single-marker regressions. The MAXR/BIC plots give the estimated effect size for each marker selected (*i.e.*, the magnitude of the partial regression coefficient). Some of the coefficients are positive (“pos effect”) and others are negative (“neg effect”). The single-marker regression plots give $-\log(P\text{-value})$ for all markers (where the P -value is for the t -test that the coefficient equals 0). The dotted lines labeled “chr ends” represent the first and last markers on the Monsanto composite genetic map. The short vertical lines at the bottom of the MAXR-effect size plots are the positions of all markers analyzed. The short vertical lines at the top of the $-\log(P\text{-value})$ plots are the positions of markers with a P -value that is less than the FDR 0.05 threshold.

selected, which account for 43% of the trait variance. Among the 39 coefficients, 33 are positive and 6 are negative. A total of 16 markers were selected in both inbreds and hybrids and, in each of those cases, the sign of the coefficient is the same. The distributions of the 50 inbred and 39 hybrid coefficients are shown in Figure 5. As expected, the range of values for hybrids (-0.09 to 0.08) is considerably smaller than the range for inbreds (-0.34 to 0.34). These ranges are similar to those for the coefficients from single-marker regression (Figure 3).

The markers selected by MAXR/BIC tend to be unassociated, since the procedure selects only markers that increase the r_{TM}^2 in combination with other markers. The level of association can be estimated by the r_{MM}^2 of regression of the genotype score of each selected marker on the others. The median value of r_{MM}^2 is 0.16 for the 50-marker inbred model and 0.11 for the 39-marker hybrid model. The selected markers are also well distributed on the genetic map. For the inbred data, the 50 selected markers are distributed across all 10 chromosomes with a range of 1–9 and a mean of 5 markers

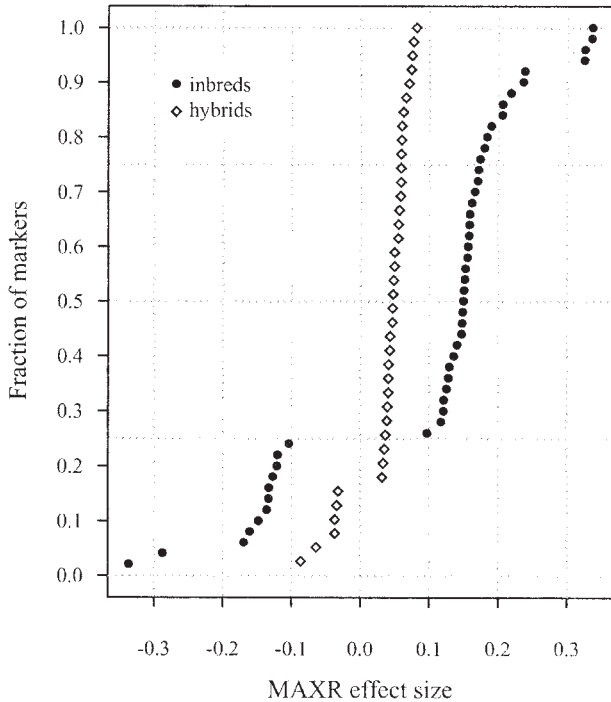


FIGURE 5.—Distribution of effect size estimates for markers selected by MAXR/BIC. The fraction of markers having a MAXR-effect size less than or equal to the value on the abscissa is given. Each point represents a marker. MAXR-effect size is the partial regression coefficient estimated by multiple regression of percentage of oil on the set of markers selected by MAXR/BIC analysis (50 markers for inbred and 39 for hybrid data).

per chromosome (Figure S2 of supplemental materials, <http://www.genetics.org/supplemental/>).

Three comparisons between the single-marker regression and MAXR/BIC results are notable:

1. For a given marker, the signs of the regression coefficients from MAXR/BIC and single-marker regression are nearly always the same (4 exceptions of 89). Thus, the two methods are quite consistent in estimating the direction of effect.
2. Only about one-half of the markers selected by MAXR/BIC are significant in the single-marker regression analysis (Figures 4 and S2). This observation may reflect the expectation that multiple regression has greater power.
3. Similarly, many of the markers that are significant in single-marker regression are not selected by MAXR/BIC, probably because the markers selected by MAXR/BIC tend to be unassociated with each other. If there are two highly associated markers that account for much the same variation in the trait, only one will be selected by MAXR/BIC, whereas both may be significant in single-marker regression.

Simulations: Determining whether the MAXR/BIC procedure selects an appropriate set of markers is a difficult problem that involves balancing the errors of excluding

important regressors and including extraneous ones (BROMAN and SPEED 2002). No statistical theory applies directly to solving this problem, but simulations can be used to evaluate the performance of model selection procedures. Of course, the usefulness of simulation results depends on the relevance of the simulation model to the biological situation at hand. Therefore, the models simulated here have several features of the observed data set. The observed genotypic data were used in all of the simulations so that the complex pattern of linkage disequilibria would be preserved. To simulate the phenotypic values, we assumed that the model derived by MAXR/BIC analysis of the observed data represents a rough approximation to the real biological situation. We then simulated variations of this model, analyzed the data by the MAXR/BIC and other methods, and compared results of the analysis to the model simulated. Four related questions are addressed:

1. Can the MAXR/BIC procedure be used to estimate the number of different QTL? This question can be addressed by asking whether the BIC stopping rule designates a number of markers similar to the actual number of QTL in the simulated model. Table 3 shows the results for 11 models in which 40, 50, or 60 QTL were “randomly placed” (in the sense that there were no constraints on the LD between QTL) and for 11 models in which the QTL were “spaced out” (in the sense that high values of LD were not allowed). Each model has a different set of markers selected to be QTL. When the QTL are spaced out, the match between the true number of QTL and the number selected is quite good. When they are randomly placed (so that QTL may be associated with one another), there is still a reasonable correspondence, but it appears that the number selected tends to be less than the actual number of QTL when there are 50 or 60 QTL. It should be noted also that the standard deviation of the number selected is fairly high. These results indicate that the number of markers selected by MAXR/BIC is similar to the number of QTL, but provides a fairly imprecise estimate.
2. What is the fraction of the QTL in a model that is selected by MAXR (“fraction of QTL selected”), and how does this fraction depend on the total number of markers selected? To address this question, a set of the best k markers was selected for each simulated data set and scored as QTL or non-QTL. Figure 6 shows plots of k vs. the mean fraction of QTL selected for 3 representative simulation models. The curves all have the same general shape [see also Figure 7 and Figure S3 in supplemental materials (<http://www.genetics.org/supplemental/>)]. The fraction of QTL selected increases and then plateaus at a point very close to the value of k specified by the BIC (and the true number of QTL). This result indicates that

TABLE 3
The number of markers selected by MAXR/BIC
from simulated data sets

No. of QTL	QTL spaced out		QTL randomly placed	
	Model	No. of markers selected ^a	Model	No. of markers selected ^a
50	B1	50 (7)	C1	45 (6)
50	B2	49 (6)	C2	45 (7)
40	D1	43 (6)	E1	41 (5)
40	D2	45 (6)	E2	43 (6)
40	D3	41 (6)	E3	44 (5)
50	D4	51 (5)	E4	48 (7)
50	D5	52 (6)	E5	43 (6)
50	D6	52 (7)	E6	47 (6)
60	D7	57 (7)	E7	53 (7)
60	D8	60 (7)	E8	56 (8)
60	D9	61 (6)	E9	57 (6)

^a The mean (SD) over 100 replicate simulations.

BIC is a good stopping rule in the sense that substantial gains in the fraction of QTL selected occur up to that point, but there is very little to gain by selecting additional markers. The fraction of QTL selected with MAXR/BIC varies somewhat among the 22 different simulation models. The mean (SE) over 100 replicates ranges from 0.56 (0.009) to 0.75 (0.007). The mean of the means for the 22 models is 0.65.

3. What is an appropriate balance between selection of QTL and non-QTL? The answer to this question depends on the goals of the experiment. In many cases, it is desirable to have the greatest possible difference between the fraction of QTL selected and the fraction of markers selected that are not QTL. This difference is maximal or nearly so when the number of markers selected is determined by the BIC, as suggested by Figure 6 and quantified in the following way. The mean difference (over 100 replicate simulations) was calculated for each of 16 k values in each model. The rank of the mean difference was averaged over the 22 models and the k value with highest mean rank is that specified by the BIC (14.7 out of 16).
4. How does the MAXR/BIC method compare with single-marker and covariate regressions? The fraction of QTL found in the selected set of markers is considerably greater for MAXR than for the other two methods for most values of k and appears to be greatest when k equals the BIC-determined value. This result is illustrated in Figure 7 for 3 models, which are very similar in this regard to all of the other 19 models analyzed (see also Figure S3). The performance of covariate regression is a little better than that of single-marker regression, but not markedly so. Figure S4 (supplemental materials, <http://www.genetics.org/supplemental/>) summarizes the

differences between the three methods at the BIC-determined value of k for all 22 models. The superiority of MAXR at the BIC-determined value of k is very consistent. Eventually, when a large number of markers is selected (well exceeding the number of QTL), the other methods find a greater fraction of the QTL (*e.g.*, model D3 in Figure 7). However, this occurs only when there is a great excess of non-QTL markers in the selected set. Similarly, the difference between the fraction of QTL selected and the fraction of markers selected that are not QTL is greater for MAXR than for the other two methods, except when the number of markers selected exceeds the number of QTL by a wide margin. Some additional comparisons among the regression methods are described in Section B and Figure S5 of supplemental materials (<http://www.genetics.org/supplemental/>).

The simulation results discussed so far are based on a complete genotypic data matrix with no missing values (*i.e.*, matrix C). Missing values are a serious problem for multiple regression because just one missing marker score eliminates all data for that individual. To perform multiple regression on real data sets, missing genotypes can be imputed. Here we used simulations to assess the potential impact of imputation on the results of multiple-regression analyses. In this study, only 2.3% of genotype calls were missing and the simulation results indicate that imputation of this small fraction of genotypes has very little effect on the results. Details are provided in Section C and Figure S3 of supplemental materials (<http://www.genetics.org/supplemental/>).

Simulations also were used to assess bias in the estimation of r_{TM}^2 and the regression coefficients for markers selected by MAXR/BIC. Details of these analyses are provided in Section D and Figure S6 of supplemental materials (<http://www.genetics.org/supplemental/>). As expected, r_{TM}^2 tends to be overestimated, but the simulation results allow the observed value of 0.62 for inbreds to be corrected to ~ 0.54 . Also, MAXR/BIC tends to underestimate the fraction of positive coefficients by a small amount. For the inbred data, 76% of the 50 markers selected have positive coefficient estimates, which can be corrected to a value of $\sim 83\%$. There is also evidence that the absolute values of the regression coefficients are overestimated to a small extent, but no corrections were attempted in this case because of other biases explained in the DISCUSSION.

DISCUSSION

Lessons from the simulation studies: The simulation studies reported here assess the performance of three regression methods for identifying QTL in an association study with a quantitative trait and a large number of markers. The MAXR/BIC model selection approach

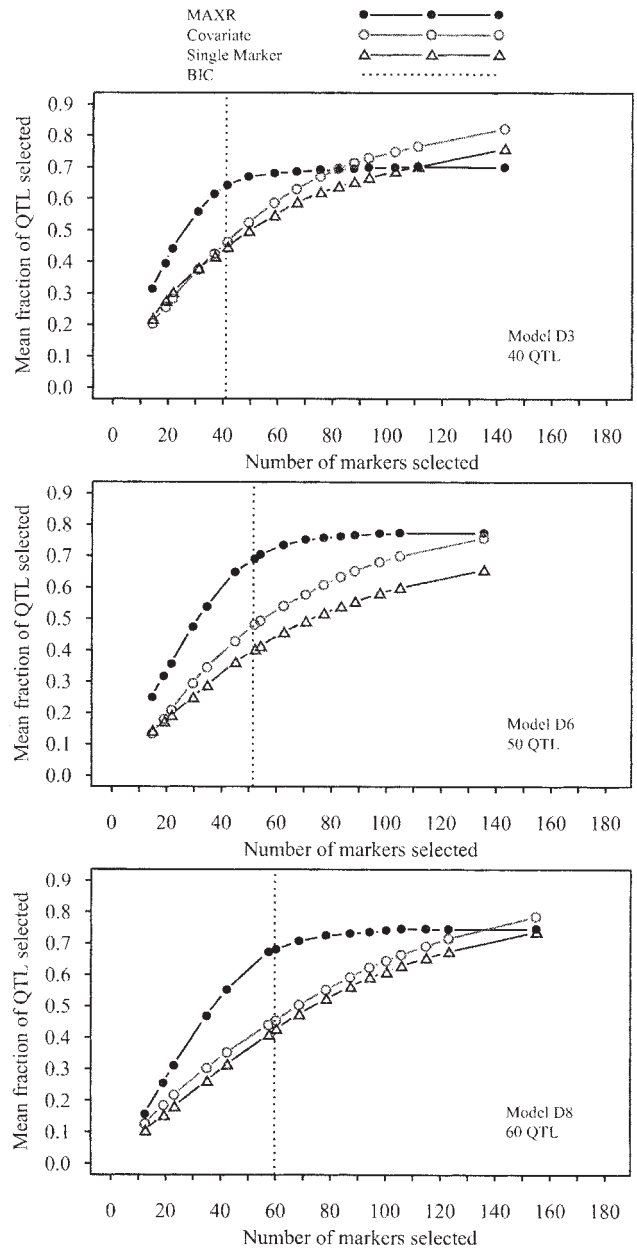
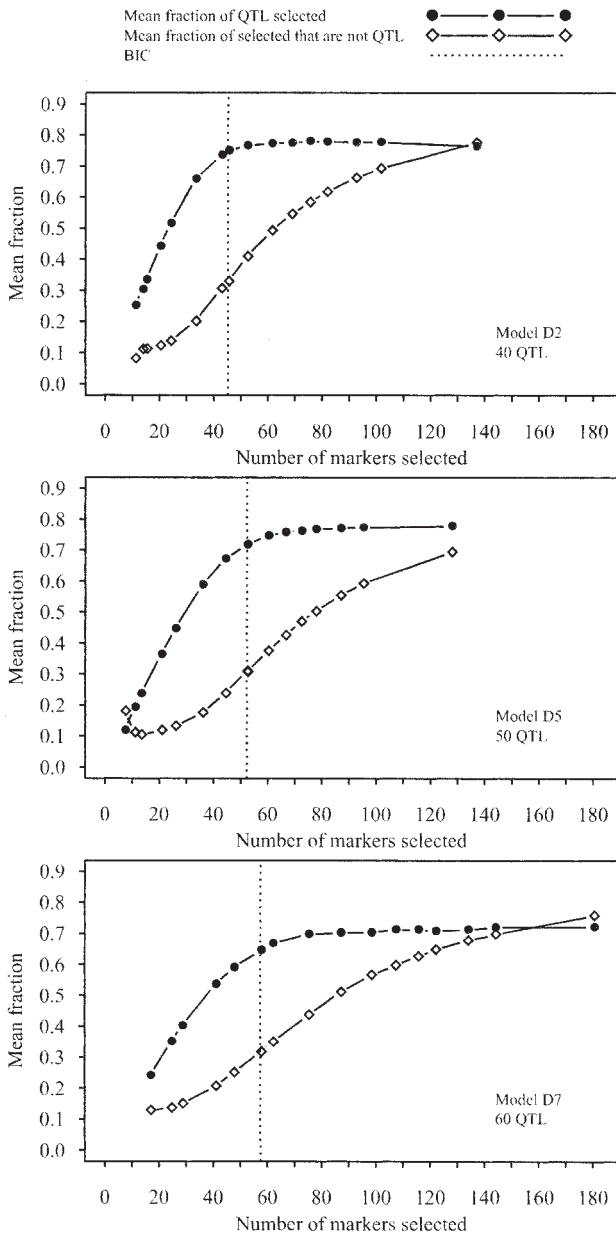


FIGURE 6.—MAXR analysis of simulated data for three models. These plots show the relationship between the mean number of markers in a selected set and two measures of the quality of those markers. One measure is the fraction of known QTL that are found in the set and the other measure is the fraction of markers in the set that are not QTL. The mean of each variable is over 100 replicate simulations. The standard errors (SE) of the means are small (maximum value of 0.01 for both quality measures and 1 for number selected). The dotted line is the mean number of markers selected by the BIC.

FIGURE 7.—Analysis by three methods of simulated data for three models. These plots show the relationship between the mean number of markers in a selected set and the fraction of known QTL that are found in the set. The mean of each variable is over 100 replicate simulations. The standard errors (SE) of the means are small (maximum value of 0.01 for all three methods for the fraction found and 1 for the number of markers).

has three key advantages over the other two methods (covariate regression and single-marker regression):

1. MAXR/BIC can be used to obtain a rough estimate of the number of QTL. It is not clear how to approach this problem with the other two methods.
2. MAXR/BIC detects a greater fraction of the QTL in

- the simulated model, except when the number of markers selected exceeds the number of QTL by a wide margin. MAXR/BIC also yields a greater difference between the fraction of QTL detected and the fraction of markers selected that are not QTL.
3. Because MAXR/BIC detects more QTL, the markers selected generally have better quality in terms of degree of association with QTL and map distance

to the nearest QTL. Therefore, we conclude that MAXR/BIC is the best of the three methods for QTL identification in this setting.

However, there are two issues to consider in the interpretation of the results of MAXR/BIC:

1. As noted earlier, the MAXR procedure adds very few new QTL to the model beyond the BIC stopping rule. Adding more markers tends to rapidly include ones that are not very closely related to QTL. This is not necessarily a problem, except that the number of markers specified by the BIC has a fairly large standard deviation (Table 3). Therefore, one may wish to select a number of markers less than the BIC specification if the goals of the experiment indicate that it is better to miss detecting some QTL than to include extraneous markers.
2. There is no established method to assess the statistical significance of markers selected by MAXR/BIC or other stepwise regression methods. BROMAN and SPEED (2002) suggest using a modified BIC criterion in which the model complexity penalty is multiplied by a factor δ , where the value of δ would be chosen (by simulation) to correspond to the LOD threshold for single-marker ANOVA under the hypothesis of no QTL. In this case, the modified BIC criterion should, in the case of no QTL, result in the selection of one or more “extraneous” loci $\sim\alpha\%$ of the time. However, Broman and Speed also note that the performance of a procedure in the presence of QTL may be rather different from its performance under the null hypothesis of no QTL. In the experiment reported here, there is no question that multiple QTL exist, so it is not clear that establishing the value of δ in the suggested fashion would be useful. Instead, we rely on the performance of MAXR/BIC in comparison with single-marker regression, for which statements of statistical significance can be made. For example, in the analysis of the inbred line data, MAXR/BIC selects 50 markers. The 50 markers with lowest probability in the single-marker regressions have a maximum P -value of 0.01, which corresponds to an FDR of 0.04. Our simulation studies indicate that a marker set selected by MAXR/BIC has better biological properties (*e.g.*, contains more QTL) than a set of the same size selected by the smallest P -values in single-marker regression. This comparison provides some indirect level of biological confidence in the importance of the markers selected by MAXR/BIC, despite our inability to provide a rigorous statement of statistical confidence.

Additional comments regarding the use of BIC as a stopping rule in model selection are given in Section E of supplemental materials (<http://www.genetics.org/supplemental/>).

The studies reported here represent an extensive set

of simulations of a highly polygenic trait in a population with a complex pattern of linkage disequilibrium. Furthermore, the number of markers analyzed is large relative to the sample size. These features apply to many association studies in man and perhaps eventually to other organisms, so the results may have some general implications. In this regard, there are two notable results:

1. Model selection by multiple-regression methods performs much better in QTL identification than does single-marker regression. In human association studies of complex disease traits, the standard approaches deal with one marker at a time and do not consider multiple-QTL models. Multilocus models of association may provide more reliable results.
2. When a trait is highly polygenic, with all QTL having similar magnitudes of effect, the QTL are difficult to detect reliably even under very favorable circumstances. Our simulation studies involved a large sample size, a high heritability, and the actual QTL included within the set of markers analyzed. Under these circumstances, and using the best method we found, $\sim 63\%$ of the QTL are detected and $\sim 33\%$ of markers selected are not QTL (although they may be associated with QTL). In human studies, the usual situation may be low heritability and very few of the actual QTL included as markers in the study. Thus, it is not surprising that more than half of the markers detected as significant in human association studies are not repeatable (LOHMUELLER *et al* 2003).

Genetic architecture of oil variation: The association study reported here leads to several conclusions about the genetic architecture of oil variation in the study population: the trait is highly polygenic, the inheritance is largely additive, the magnitudes of individual QTL effects appear to be relatively small, and most QTL have positive effects (*i.e.*, in the direction predicted by the parental difference). Discussion of these results and their relationship to other QTL studies focuses on the inbred line results, since the simulations were based on results from that data set and the hybrid lines appear to have very similar genetic effects.

Although oil variation in our study population is highly polygenic, it is difficult to estimate the number of QTL with confidence. Nevertheless, the following arguments suggest that >50 QTL may be involved.

The MAXR/BIC multiple-regression analysis selects 50 markers from the analysis of the inbred lines and the corresponding r_{TM}^2 is 0.62. The simulations suggest that the number of markers selected by MAXR/BIC provides a rough estimate of the actual number of QTL. Furthermore, it seems likely that most of the 50 markers selected represent different QTL because they are well dispersed on the genetic map and they tend to be unassociated with each other. The simulations also indicate that r_{TM}^2 tends to be overestimated by MAXR/BIC and a rough correction for the inbred lines suggests a value

of 0.54. The 96% heritability suggests that nearly all of the variance of predicted line means is genetic. So, there is direct evidence that markers associated with ~ 50 different QTL account for $\sim 50\%$ of the genetic variance in oil concentration.

Three types of factors may account for the $\sim 50\%$ of genetic variation that is not explained by the estimated number and magnitude of QTL effects:

1. Some of the unexplained variation may be due to the ~ 50 QTL detected in the experiment, but not accounted for in the fitted model because of imperfect associations between markers and QTL. This imperfect association causes the additive effects to be underestimated.
2. Epistatic interactions among the QTL detected could account for some of the unexplained variation. We noted earlier that $\sim \alpha\%$ of tests are significant with a P value $< \alpha\%$ in testing all pairs of markers for interaction in two-way ANOVA, suggesting very little evidence for epistasis. The same result is found when only those markers selected by MAXR/BIC are considered. Thus, it appears that epistatic effects are minimal at best.
3. Another possibility is QTL that were not detected because either their effects are too small or they are not associated with markers in this study. The markers were selected to show a large frequency difference between IHO and ILO, thereby increasing the probability of proximity to a QTL. However, only $\sim 60\%$ of the total map length lies within 2.5 cM of a marker and only $\sim 50\%$ of marker pairs within that distance have $r_{LD}^2 \geq 0.25$.

In conclusion, while it is formally possible that underestimation of individual QTL effects is responsible for all of the unexplained genetic variation, it seems likely that more than the ~ 50 QTL detected here are involved.

The number of QTL detected in this study is much larger than that in previous QTL mapping studies in maize or other species. For example, few plant studies report more than one QTL per chromosome (KEARSEY and FARQUHAR 1998), whereas the mean is 5 in our study. There are two reasons for this difference. First, our population is derived from parental lines that are very divergent for the trait because of 70 generations of artificial selection. This selection, and the subsequent cross, provided an opportunity for many QTL effects to be concentrated in one population. Second, the resolution of the study is considerably greater than that of most other QTL studies because of the large number of markers, the large sample size, and especially the 10 generations of random mating. Most QTL studies involve a single generation of recombination and have a resolution of 10–30 cM (KEARSEY and FARQUHAR 1998). In this study, the resolution is probably on the order of 2–3 cM, since pairs of markers any farther apart rarely have substantial levels of linkage disequilibrium.

Evidently, the improved resolution is most important, because an earlier study of an F_2 population derived from IHO and ILO detected 11 QTL regions with 80 markers (BERKE and ROCHEFORD 1995).

The absence of detectable epistasis in this study is notable. The literature contains many reports of epistatic interactions between QTL in various traits and species, but at least as many reports of no interaction (see reviews by MACKAY 2001; ORR 2001; BARTON and KEIGHTLEY 2002). It has been suggested that failure to find epistasis may represent low power or inadequate experimental design (FRANKEL and SCHORK 1996; MACKAY 2001). In our study, the large sample size, the very high heritability, and the large number of QTL detected indicate an unusually high power to detect genetic effects. Nevertheless, even large populations may contain few individuals in the least-frequent two-locus genotype classes and segregation of other QTL may interfere with detection of epistasis between individual pairs of loci (MACKAY 2001). In any case, because additive effects for oil variation predominate in our population, a simplified approach to modeling trait variation can be effective. With any luck, the same situation may prevail for many complex traits in human and other populations so that “fear of epistasis” (FRANKEL and SCHORK 1996) is unnecessary (but understandable because of the huge increase in complexity of models that must be considered).

The response to artificial selection: The response to artificial selection for oil concentration in IHO and ILO was very smooth and continuous for 89 generations (DUDLEY and LAMBERT 1992). BARTON and KEIGHTLEY (2002) note that sustained responses to selection must be based on a large number of minor variants present in the base population and/or new mutations. The large number of QTL detected and inferred by this study (> 50) can account easily for the smooth and prolonged selection response.

BARTON and KEIGHTLEY (2002) cite experimental evidence that spontaneous mutation rates at QTL in various species are high enough to make a substantial contribution to selection responses. Furthermore, WALSH (2004) used population genetics theory to suggest that the majority of response in IHO and ILO was due to new mutation, primarily because most of the initial variation would be rapidly fixed by selection and/or random drift in the small population sizes experienced by the selection lines ($N_e \sim 10$). However, KEIGHTLEY (2004) notes that responses due to new mutation often involve sudden changes in mean performance, while those due to standing variation tend to be relatively smooth (like the Illinois selection responses). Unfortunately, the genetic architecture results do not seem to have a direct bearing on the importance of new mutations.

WALSH (2004) also examined the question of whether selection on QTL can overpower random drift in IHO and ILO. He calculated the probability of fixation of a

favorable allele from estimates of the effective population size, initial frequency in the base population (from selection limit theory by DUDLEY 1977), and selection coefficient (which requires estimates of additive effect (a), selection intensity, and phenotypic standard deviation). He used $2a = 0.39$, which is based on a rough estimate of the number of effective factors (54) from a biometrical method (DUDLEY 1977), and an initial frequency estimate (0.20) based on response at 100 generations. If $N_e = 6$, the probability of fixation is 0.80 and if $N_e = 12$, it is 0.96. We get essentially the same answer using an estimate of $a = 0.16$ (the median value of positive coefficients in this study) and an initial frequency estimate (0.26) based on response at 70 generations. These probability estimates are consistent with our observation that $\sim 20\%$ of QTL effect estimates are negative (*i.e.*, a low-oil allele fixed or at higher frequency in IHO than in ILO). Hitchhiking, due to close linkage between QTL with opposite effects, may also contribute to this observation. Among the 12 QTL with negative effect estimates from MAXR/BIC, 4 are within 5 cM of a QTL with positive effect.

Applications in agriculture: A major objective of the study reported here was to identify oil QTL for improvement of commercial maize germplasm. One potential approach to improvement is focused introgression of high-oil alleles from IHO into high-yielding inbred lines using molecular markers. In this way, it might be possible to separate the beneficial effects on oil from negative effects on agronomic characteristics. The genetic architecture reported here suggests that this approach would be very difficult. A large number of genes contribute to high oil in IHO and none that we detected have a large effect.

Another approach is to identify individual genes that cause oil variation and then use genetic engineering with plant transformation to increase oil. The resolution of the study reported here is not at the level of single genes. We have suggested that the resolution may be on the order of a few centimorgans. In maize, there may be ~ 30 genes/cM (assuming 50,000 genes in 1600 cM). Although the degree of resolution varies in different chromosomal regions and may be <1 cM in certain regions, the genetic association evidence alone is not sufficient to identify individual QTL. However, in combination with other types of evidence, a significant genetic association may be sufficient to warrant transgenic testing. In this study, 60 of the 440 markers are in candidate genes identified by methods such as transcriptional profiling, metabolic pathway analysis, mutational effects, and functional genomics in Arabidopsis. Among the 60 candidate markers, 12 were selected by the MAXR/BIC method in inbreds and/or hybrids, and several of those are also significant in single-marker regression. Some of these candidates have been selected for genetic engineering to construct alleles that may have large effects on oil in maize.

The highly polygenic nature of oil variation in the study reported here may suggest that it will be difficult to make substantial improvements in oil concentration by engineering a single gene. However, an example from the evolution of pesticide resistance in insects is instructive (ROUSH and MCKENZIE 1987). The response to selection in small laboratory populations usually appears to involve many genes of small effect whereas the response in large natural populations often involves a single gene of major effect. This difference is expected if beneficial mutations with large effect are much rarer than those of small effect (FISHER 1930). In any case, the rare occurrence of major gene effects on resistance shows that it is possible to make large changes in the trait when the right mutation comes along. Finding genes that have some effect on oil in maize should help a great deal to focus the effort of finding the right mutation (or engineered sequence).

We gratefully acknowledge the assistance of Jason Bull, Yongwei Cao, Merce Crosas, Dawn Dionne, Fenggao Dong, Sam Eathington, Elizabeth Frias, Alistair Kerr, Jie-Yi Lin, Mark Magid, Don Nelson, Patricia Price, Christina Read, Randy Rich, James Rogers, Monica Ravanello, Thomas Savage, and Xiao Yang. We thank Lyle Crossland for his help in initiating and maintaining financial support. We also thank Zhao-Bang Zeng for some helpful suggestions on the simulation approach and Bruce Weir for pointing out the relationship between r_{ID}^2 and additive genetic variance. Zhao-Bang Zeng, Bruce Weir, and Bruce Walsh gave many useful comments on a draft of the manuscript. Financial support for this project was provided by Renessen LLC.

LITERATURE CITED

- BARTON, N. H., and P. D. KEIGHTLEY, 2002 Understanding quantitative genetic variation. *Nat. Rev. Genet.* **3**: 11–21.
- BERKE, T. G., and T. R. ROCHEFORD, 1995 Quantitative trait loci for flowering, plant and ear height, and kernel traits in maize. *Crop Sci.* **35**: 1542–1549.
- BROMAN, K. W., and T. P. SPEED, 2002 A model selection approach for the identification of quantitative trait loci in experimental crosses. *J. R. Stat. Soc. B* **64**: 641–656.
- DELLAPORTA, S. L., L. WOOD and J. B. HICKS, 1983 A plant DNA miniprep: version II. *Plant Mol. Biol. Rep.* **1**: 19–21.
- DEVLIN, B., and N. RISCH, 1995 A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* **29**: 311–322.
- DUDLEY, J. W., 1977 Seventy-six generations of selection for oil and protein percentage in maize, pp. 459–473 in *Proceedings of the International Conference on Quantitative Genetics*, edited by E. POLLAK, O. KEMPTHORNE and T. B. BAILEY, JR. Iowa State University Press, Ames, IA.
- DUDLEY, J. W., and R. J. LAMBERT, 1992 Ninety generations of selection for oil and protein in maize. *Maydica* **37**: 1–7.
- DUDLEY, J. W., and R. J. LAMBERT, 2004 100 Generations of selection for oil and protein in corn. *Plant Breed. Rev.* **24** (Pt. 1): 79–110.
- DUDLEY, J. W., R. J. LAMBERT and D. E. ALEXANDER, 1974 Seventy generations of selection for oil and protein concentration in the maize kernel, pp. 181–212 in *Seventy Generations of Selection for Oil and Protein in Maize*, edited by J. W. DUDLEY. Crop Science Society of America, Madison, WI.
- DYER, D., and P. FENG, 1997 NIR destined to be a major analytical influence. *Feedstuffs* **69**: 16–25.
- EXCOFFIER, L., and M. SLATKIN, 1995 Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.* **12**: 921–927.
- FALLIN, D., and N. J. SCHORK, 2000 Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization al-

- gorithm for unphased diploid genotype data. *Am. J. Hum. Genet.* **67**: 947–959.
- FISHER, R. A., 1930 *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.
- FRANKEL, W. N., and N. J. SCHORK, 1996 Who's afraid of epistasis? *Nat. Genet.* **14**: 371–373.
- HOCKING, R. R., 1976 The analysis and selection of variables in linear regression. *Biometrics* **32**: 1–49.
- HYMOWITZ, T., J. W. DUDLEY, F. I. COLLINS and C. M. BROWN, 1974 Estimations of protein and oil concentration in corn, soybean, and oat seed by near-infrared light reflectance. *Crop Sci.* **14**: 167–170.
- JANSEN, R. C., and P. STAM, 1994 High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**: 1447–1455.
- KEARSEY, M. J., and A. G. L. FARQUHAR, 1998 QTL analysis in plants; where are we now? *Heredity* **80**: 137–142.
- KEIGHTLEY, P. D., 2004 Mutational variation and long-term selection response. *Plant Breed. Rev.* **24** (Pt. 1): 227–247.
- LAMBERT, R. J., 1994 High-oil corn hybrids, pp. 123–145 in *Specialty Corns*, edited by A. R. HALLAUER. CRC Press, London.
- LAMBERT, R. J., D. E. ALEXANDER and I. J. MEJAYA, 2004 Single kernel selection for increased grain oil in maize synthetics and high-oil hybrid development. *Plant Breed. Rev.* **24** (Pt. 1): 153–175.
- LANDER, E. S., and N. J. SCHORK, 1994 Genetic dissection of complex traits. *Science* **265**: 2037–2048.
- LITTELL, R. C., W. W. STROUP and R. J. FREUND, 2002 *SAS for Linear Models*. SAS Publishing, Cary, NC.
- LIVAK, K. J., 2003 SNP genotyping by the 5'-nuclease reaction. *Methods Mol. Biol.* **212**: 129–147.
- LOHMUELLER, K. E., C. L. PEARCE, M. PIKE, E. S. LANDER and J. N. HIRSCHHORN, 2003 Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat. Genet.* **33**: 177–182.
- MACKAY, T. F. C., 2001 The genetic architecture of quantitative traits. *Annu. Rev. Genet.* **35**: 303–339.
- NIELSEN, D. M., and B. S. WEIR, 1999 A classical setting for associations between markers and loci affecting quantitative traits. *Genet. Res.* **74**: 271–277.
- ORR, H. A., 2001 The genetics of species differences. *Trends Ecol. Evol.* **16**: 343–350.
- PATTERSON, H. D., and E. R. WILLIAMS, 1976 A new class of resolvable incomplete block designs. *Biometrika* **63**: 83–92.
- RAWLINGS, J. O., S. G. PANTULA and D. A. DICKEY, 1998 *Applied Regression Analysis*. Springer-Verlag, New York.
- ROUSH, R. R., and J. A. MCKENZIE, 1987 Ecological genetics of insecticide and acaricide resistance. *Annu. Rev. Entomol.* **32**: 361–380.
- SAS INSTITUTE, 1999 *SAS Online Documentation, Version 8*. SAS Institute, Cary, NC.
- SCHWARZ, G., 1978 Estimating the dimension of a model. *Ann. Stat.* **6**: 461–464.
- S-PLUS 2000, 1999 *Guide to Statistics*. Vol. I. MathSoft, Seattle.
- STOREY, J. D., 2002 A direct approach to false discovery rates. *J. R. Stat. Soc. B* **64**: 479–498.
- STOREY, J. D., and R. TIBSHIRANI, 2003 Statistical significance for genome-wide experiments. *Proc. Natl. Acad. Sci. USA* **100**: 9440–9445.
- WALSH, B., 2004 Population- and quantitative-genetic models of selection limits. *Plant Breed. Rev.* **24** (Pt. 1): 177–225.
- WEIR, B. S., 1996 *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.
- WEIR, B. S., and C. C. COCKERHAM, 1979 Estimation of linkage disequilibrium in randomly mating populations. *Heredity* **42**: 105–111.
- ZENG, Z-B., 1994 Precision mapping of quantitative trait loci. *Genetics* **136**: 1457–1468.

Communicating editor: J. WAKELEY

