

Precision and High-Resolution Mapping of Quantitative Trait Loci by Use of Recurrent Selection, Backcross or Intercross Schemes

Z. W. Luo,^{*,†,1} Chung-I Wu[‡] and M. J. Kearsey^{*}

^{*}*School of Biosciences, The University of Birmingham, Edgbaston, Birmingham B15 2TT, England, †Laboratory of Population and Quantitative Genetics, Institute of Genetics, Fudan University, Shanghai 200433, China and*

[‡]*Department of Ecology and Evolution, The University of Chicago, Chicago, Illinois 60637*

Manuscript received October 10, 2001

Accepted for publication March 13, 2002

ABSTRACT

Dissecting quantitative genetic variation into genes at the molecular level has been recognized as the greatest challenge facing geneticists in the twenty-first century. Tremendous efforts in the last two decades were invested to map a wide spectrum of quantitative genetic variation in nearly all important organisms onto their genome regions that may contain genes underlying the variation, but the candidate regions predicted so far are too coarse for accurate gene targeting. In this article, the recurrent selection and backcross (RSB) schemes were investigated theoretically and by simulation for their potential in mapping quantitative trait loci (QTL). In the RSB schemes, selection plays the role of maintaining the recipient genome in the vicinity of the QTL, which, at the same time, are rapidly narrowed down over multiple generations of backcrossing. With a high-density linkage map of DNA polymorphisms, the RSB approach has the potential of dissecting the complex genetic architecture of quantitative traits and enabling the underlying QTL to be mapped with the precision and resolution needed for their map-based cloning to be attempted. The factors affecting efficiency of the mapping method were investigated, suggesting guidelines under which experimental designs of the RSB schemes can be optimized. Comparison was made between the RSB schemes and the two popular QTL mapping methods, interval mapping and composite interval mapping, and showed that the scenario of genomic distribution of QTL that was unlocked by the RSB-based mapping method is qualitatively distinguished from those unlocked by the interval mapping-based methods.

THE benchmark article by LANDER and BOTSTEIN (1989) stimulated enormous interest in locating quantitative trait loci (QTL) in experimental and natural populations. Research efforts in the last decade were focused on mass production of high-throughput DNA polymorphic markers (DIB *et al.* 1996; WANG *et al.* 1998; MARTH *et al.* 2001) and development of analytical methods for detecting the presence and inferring the locations of QTL in marker linkage maps (LANDER and BOTSTEIN 1989; HALEY and KNOTT 1992; LUO and KEARSEY 1992; ZENG 1994; RABINOWITZ 1997; MOTT *et al.* 2000). A recent comprehensive review based on 47 experimental studies of QTL mapping in plants, however, revealed that the current QTL mapping practice entails tremendous research effort and financial investment but yields QTL map localizations that are far from being satisfactory for identifying and isolating the quantitative trait genes at the molecular level (KEARSEY and FARQUHAR 1998). The analysis showed that QTL were usually mapped with low accuracy and poor resolution (~10–30 cM) and that the proportion of quantitative genetic variation determined by the QTL detected was

very low (~5%). Little progress has been made so far in cloning quantitative trait genes on the basis of inferred map location of QTL despite the claim in ALPERT and TANKSLEY (1996) that a yeast artificial chromosome (YAC) clone bearing a major QTL affecting fruit weight in tomato was successfully obtained. However, the gene (*fw2.2*) was finally identified after 20 years' journey in narrowing down the candidate genomic region that contains *fw2.2* (FRARY *et al.* 2000).

Theoretical investigations (BOEHNKE 1994; GUO and LANGE 2000) suggested that the major bottleneck in narrowing down the confidence interval of QTL location is the limited number of informative meioses obtainable in most mapping populations in the literature. Experimental strategies using historically accumulated recombinations between markers and QTL have been suggested as an efficient approach to improving map resolution of QTL. These include several alternatives. First, DARVASI and SOLLER (1995) demonstrated that the confidence interval of QTL location inferred from a conventional F₂ mapping population might be reduced by up to fivefold if the F₂ population is expanded into a so-called advanced intercross line (AIL) by continued intercross. Improvement in the mapping resolution in an AIL is due to breakdown of linkage disequilibrium between the QTL and their linked marker loci. How-

¹*Corresponding author:* School of Biosciences, The University of Birmingham, Edgbaston, Birmingham B15 2TT, England.
E-mail: z.luo@bham.ac.uk

ever, an appropriate statistical method still needs to be developed to model and analyze the data from an AIL experiment (MANLY and OLSON 1999). Second, the rate of dissipation in linkage disequilibrium between QTL and nearby marker loci in genetically isolated natural populations with good genealogical records may be modeled in terms of the recombination fraction between the loci and demographic parameters defining the evolutionary history of the populations. This approach may, in the best case, enable QTL to be located in a region of <1 cM (HILL and WEIR 1994; DE LA CHAPELLE and WRIGHT 1998; LUO *et al.* 2000; LUO and WU 2001). However, it must be pointed out that much uncertainty exists in this population-based analysis if evolutionary details of the populations are not appropriately taken into account (ZOLLNER and VON HAESELER 2000). Statistical modeling of linkage disequilibrium involved with QTL has arisen as a new challenge to modern quantitative genetics (LUO *et al.* 2000; LUO and WU 2001). Third, use of congenic lines was shown to be effective in narrowing intervals of inferred QTL location, providing the QTL effect was so significant that genotypes at the QTL could be assigned accurately (DARVASI 1997; NADEAU and FRANKEL 2000). Another major limitation of this approach is that tightly linked QTL would not be resolved if their genetic effects act in the same direction in the congenic strains. Fourth, construction of chromosome substitution lines enables precise identification of the chromosomes, which carry QTL. Recombinant progeny from backcrossing the appropriate chromosome substitution strain to its host strain may be used to test whether more than one QTL accounts for trait phenotypic difference among the substitution chromosomes and to locate each QTL with considerable map resolution (NADEAU *et al.* 2000).

A breeding scheme with repeated backcrossing and selection was proposed long ago by WRIGHT (1952) for isolating quantitative trait genes of large effect, but little progress was made in QTL analysis until a recent series of elegant theoretical studies by HILL (1997, 1998). On the basis of his theory of directional selection for quantitative traits in finite populations (HILL 1969), he formulated the probability that quantitative trait genes of specified effects remain segregating in a backcross family undergoing one generation of truncation selection for nonrecurrent parental phenotype. The results were extended to multiple generations by an approximation that did not take into account the change in gene frequency under repeated selection and random drift. Use of multiple generations in the recurrent selection backcross (RSB) scheme is essential for accumulating a sufficient amount of recombination both between closely linked QTL and between the QTL and nearby markers. It has been clear that the RSB is an effective approach only in isolating QTL of large effect from their surrounding genome regions; QTL with small to medium effect have a high probability of being lost during the

breeding scheme. An alternative way, suggested in HILL (1998), is to intersperse one generation of intercrossing among the selected individuals between consecutive backcrossings. However, it is less clear what impact the recurrent selection backcross *inter se* intercross (RSBI) scheme will have on maintaining QTL of small effect on the one hand and on separating the QTL from their surrounding marker loci on the other. Moreover, many important questions remain to be answered. How can the basic idea of RSB or RSBI be extended to dissect complex genetic variation into QTL? How robust is the strategy to various models of QTL effect? What precision and resolution in the QTL mapping may be expected by use of the RSB or RSBI schemes if advantage is fully taken of the fast development of single nucleotide polymorphic (SNP) markers? The extremely dense distribution of SNP markers over the whole genome may provide at least one polymorphic site in each of the functional genes in the genome (MARTH *et al.* 2001), thus allowing full control of the genetic architecture that underlies complex quantitative genetic variation. What are the major factors affecting the RSB or RSBI for their efficiency in QTL mapping? In an attempt to address these questions, this article provides an exact theoretical prediction of mean and variance of heterozygosity at a marker locus linked to one or two QTL with any degree of recombination for any number of generations of the RSB schemes. This builds up a theoretical basis for the RSB-based QTL mapping. More complicated models were investigated and the above questions were explored by numerical evaluation of the theoretical predictions and by extensive computer simulation. Comparisons were made between the RSB(I) strategy and the routine methods of QTL mapping for their precision and resolution in identifying locations of multiple linked QTL.

THEORY AND METHODS

The breeding scheme: The theoretical analysis considers a breeding scheme initiated from two inbred lines P_1 and P_2 that are assumed to be fixed for different alleles at m marker loci and q loci affecting a quantitative trait. The QTL alleles increasing the phenotype are fixed in P_1 and those decreasing the phenotype in P_2 . Effects of individual QTL are scaled in units of standard deviation of the residual variation of the trait, which are assumed to be normally distributed with mean zero and variance 1.0. Various models for genetic effects of the QTL are considered in the following analysis.

The two inbred lines are crossed to generate an F_1 family and a random sample of F_1 individuals are backcrossed to recurrent parental line P_2 to produce F independent backcross families with a constant size of N . These families are defined as the first generation of the breeding program. In each of these backcross families, n individuals with the top-scoring phenotype for a quan-

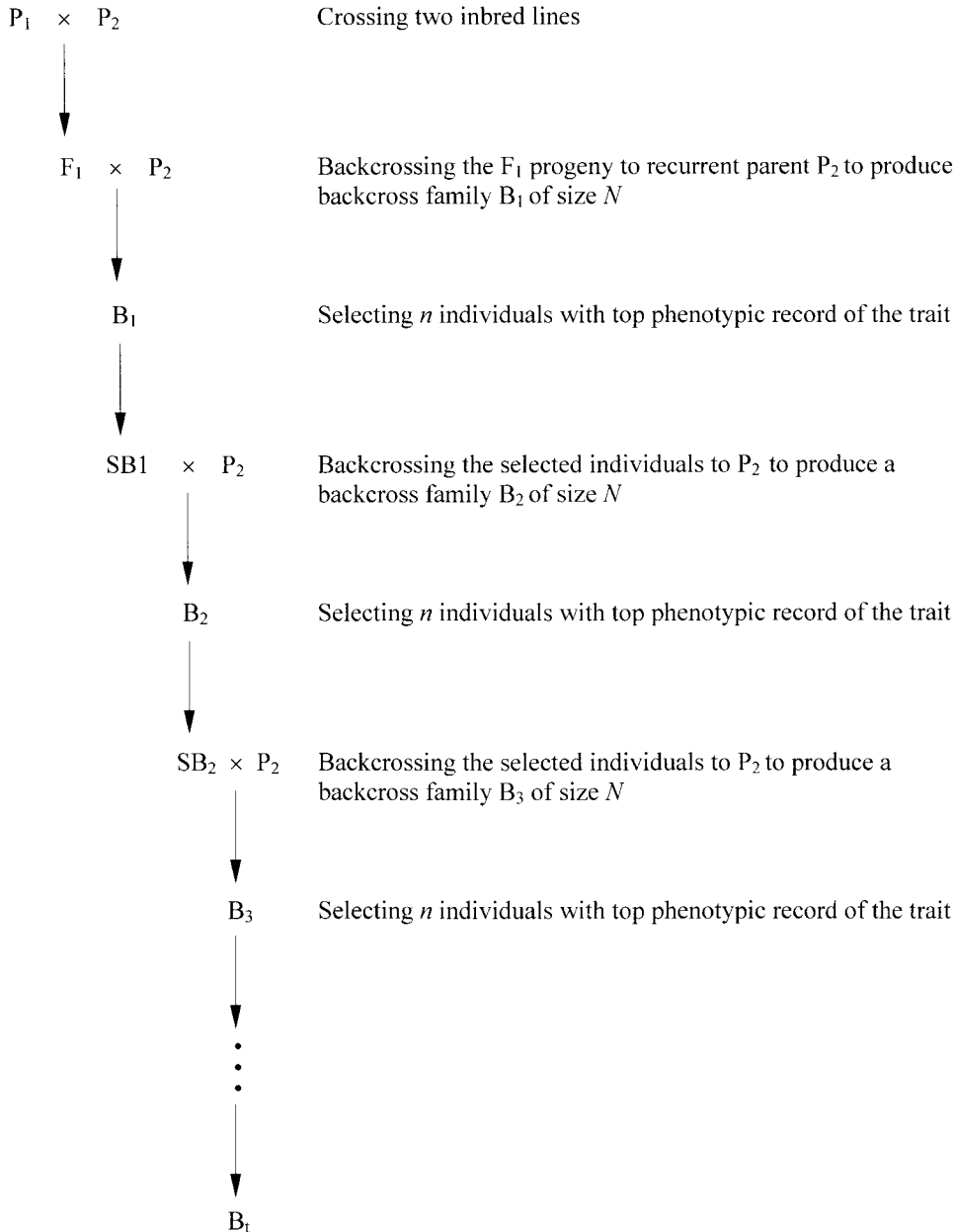


FIGURE 1.—A diagram illustrating a recurrent selection and backcross (RSB) breeding scheme.

titative trait are selected. The selected individuals are either backcrossed to the recurrent parent or randomly intercrossed to produce the next generation of the families. The breeding program lasts for T generations. Let B_{it} (SB_{it}) or I_{it} (SI_{it}) denote the i th backcross or intercross family ($i = 1, 2, \dots, F$) at generation t before (or after) selection. A simple diagram describing the RSB breeding scheme is given in Figure 1.

HILL (1997, 1998) exploited the probability that genes at one or two QTL of specified effects remain segregating after one generation of truncation selection. This provides a direct comparison with the situation where the probability is reduced by one-half, on average, if the locus is free of the selection pressure. Instead of calculating the segregating probability at the selected locus, we consider heterozygosity maintained

at any marker loci, which may or may not be linked to the QTL in any generation of the breeding program. In all QTL mapping strategies, map location of a putative QTL is inferred from information of marker genotype and trait phenotype observed from a specified mapping population. In the RSB scheme, the genetic contribution of QTL to the trait phenotype is reflected as efficiency of selection in counterbalancing against the loss of the QTL allele of the nonrecurrent parent due to repeated backcross and genetic drift, while the genetic marker in the system provides information about genome location and extent to which the selection affects it. The closer the marker locus is to the selected locus (QTL), the more efficient the selection will be in maintaining the recipient allele at the marker locus, and thus the more likely the marker locus is in

heterozygous status. Therefore, the marker heterozygosity serves as a natural and rational measure for its location relative to the QTL. When the genome regions bearing QTL are covered with densely distributed markers, the closest approximation for map location of the QTL may be inferred as the location at which the marker heterozygosity reaches its peak value. Thus, the theoretical analyses below are focused on heterozygosity of a marker locus that is linked to QTL with an arbitrary value of recombination frequency.

The theoretical analyses comprise four sections. The first two relate to the dynamic change in genetic structure of the RSB breeding populations under two loci (one marker and one QTL) and three loci (one marker and two QTL) models. The third section develops the calculation of the mean and variance of the marker heterozygosity under the above two models. When the RSB scheme is interrupted by incorporation of intercrossing, prediction of the dynamic change in linkage disequilibria between multiple loci becomes intractable. To investigate the impact of the RSBI, theoretical analysis is restrained to one QTL only and described in the last section.

Two loci model: The model considers one marker locus and one QTL. The two alleles at the marker locus are denoted by M and m , respectively, and those at the QTL by A and a . For each of the F independent backcross families, let $X_t = (x_{t1} \ x_{t2} \ x_{t3} \ x_{t4})^T$ and $Y_t = (y_{t1} \ y_{t2} \ y_{t3} \ y_{t4})'$ be the two vectors that represent the distribution of four possible genotypes, at generation t , in B_{it} (before selection) and in SB_{it} (after selection), respectively. In other words, element x_{ij} (or y_{ij}) is a random variable for the number of individuals with genotype j in B_{it} (or SB_{it}), where $j = 1, 2, 3, 4$ corresponding, respectively, to joint genotypes MA/ma , ma/ma , mA/ma , and Ma/ma at the marker and QTL.

The stochastic change in population genetic structure during a RSB program is described fully by the following probability distributions:

$$\begin{aligned} p_{IR} &= \Pr\{X_t = R\} = \Pr\{x_{tk} = r_k, k = 1, 2, 3, 4\} \\ q_{IS} &= \Pr\{Y_t = S\} = \Pr\{y_{tk} = s_k, k = 1, 2, 3, 4\} \\ \xi_{ISR} &= \Pr\{Y_t = S | X_t = R\} \\ &= \Pr\{y_{tk} = s_k, k = 1, 2, 3, 4 | x_{tk} = r_k, k = 1, 2, 3, 4\} \\ \eta_{IRS} &= \Pr\{X_t = R | Y_{t-1} = S\} \\ &= \Pr\{x_{tk} = r_k, k = 1, 2, 3, 4 | y_{t-1k} = s_k, k = 1, 2, 3, 4\}. \end{aligned}$$

These can be evaluated in a recursive formulation that is initiated from

$$\begin{aligned} p_{1R} &= \Pr\{X_1 = R\} = \frac{N!}{2^N \prod_{k=1}^4 r_k!} (1-c)^{r_1+r_2} c^{r_3+r_4} \\ q_{1S} &= \Pr\{Y_1 = S\} = \sum_R \Pr\{X_1 = R\} \Pr\{Y_1 = S | X_1 = R\} = \sum_R p_{1R} \xi_{1SR}. \end{aligned} \quad (1)$$

In general,

$$\begin{aligned} p_{IR} &= \Pr\{X_t = R\} = \sum_S \Pr\{Y_{t-1} = S\} \Pr\{X_t = R | Y_{t-1} = S\} = \sum_S q_{I-1S} \eta_{IRS} \\ q_{IS} &= \Pr\{Y_t = S\} = \sum_R \Pr\{X_t = R\} \Pr\{Y_t = S | X_t = R\} = \sum_R p_{IR} \xi_{ISR}. \end{aligned} \quad (2)$$

In the above, \sum_R (or \sum_S) represents summation over all possible r_k under the constraint $\sum_{k=1}^4 r_k = N$ (or s_k under $\sum_{k=1}^4 s_k = n$). The conditional probability can be calculated as

$$\xi_{ISR} = \Pr\{Y_t = S | X_t = R\} = \prod_{k=1}^4 \binom{r_k}{s_k} \sum_{j=1}^4 s_j I_j(S, R; \mu) \quad (3)$$

in which

$$I_i(S, R; \mu) = \int_{-\infty}^{\infty} [1 - \Phi_i(x)]^{r_i-1} \prod_{j=1}^4 \Phi_j(x)^{r_j-s_j} \prod_{j \neq i}^4 [1 - \Phi_j(x)]^{s_j} f_i(x) dx, \quad (4)$$

where $\mu = (\mu_1 \ \mu_2 \ \mu_1 \ \mu_2)'$ and μ_1 and μ_2 are means of QTL genotypes Aa and aa , respectively, and $f_i(x)$ and $\Phi_i(x)$ are, respectively, the probability density function and the probability distribution function of a normal distribution with mean μ_i and variance 1.0,

$$\eta_{IRS} = \Pr\{X_t = R | Y_{t-1} = S\} = \frac{N!}{\prod_{k=1}^4 r_k!} \prod_{k=1}^4 \phi_k^k, \quad (5)$$

where $\phi_1 = (1-c)s_1/2N$, $\phi_2 = [(1-c)s_1 + 2s_2 + s_3 + s_4]/2N$, $\phi_3 = [cs_1 + s_3]/2N$, and $\phi_4 = [cs_1 + s_4]/2N$.

Three loci model: There are three possible patterns of relative locations of one marker locus and two QTL under this model. Let c_1 and c_2 be recombination frequencies between loci 1 and 2 and between loci 2 and 3, respectively. Assuming there is no recombination interference, the recombination frequency between the first and the third loci will be $c = c_1(1-c_2) + c_2(1-c_1)$. Analogous to the above two loci model, the breeding scheme at generation t can be described by two random vectors: $X_t = (x_{t1} \ x_{t2} \ \dots \ x_{t8})'$ or $Y_t = (y_{t1} \ y_{t2} \ \dots \ y_{t8})'$ whose component x_{ij} (or y_{ij}) denotes the number of individuals with the marker-QTL genotype i ($i = 1, 2, \dots, 8$ corresponding to AMB/amb , amb/amb , Amb/amb , aMB/amb , AMb/amb , $ambB/amb$, $AmbB/amb$, and aMb/amb , accordingly) in B_{it} (or SB_{it}). The probability distributions of these random vectors are given by

$$\begin{aligned} p_{IR} &= \Pr\{X_t = R\} = \Pr\{x_{tk} = r_k, k = 1, 2, \dots, 8\} \\ &= \sum_S \Pr\{Y_{t-1} = S\} \Pr\{X_t = R | Y_{t-1} = S\} = \sum_S q_{I-1S} \eta_{IRS} \end{aligned} \quad (6)$$

$$\begin{aligned} q_{IS} &= \Pr\{Y_t = S\} = \Pr\{y_{tk} = s_k, k = 1, 2, \dots, 8\} \\ &= \sum_R \Pr\{X_t = R\} \Pr\{Y_t = S | X_t = R\} = \sum_R p_{IR} \xi_{ISR}. \end{aligned} \quad (7)$$

These can be evaluated from the initial condition

$$p_{1R} = \frac{N!}{2^N \prod_{k=1}^8 r_k!} [(1 - c_1)(1 - c_2)]^{r_1+r_2} \times [c_1(1 - c_2)]^{r_3+r_4} [(1 - c_1)c_2]^{r_5+r_6} (c_1c_2)^{r_7+r_8} \quad (8)$$

and the conditional probabilities

$$\xi_{sR} = \Pr\{Y_i = s | X_i = R\} = \prod_{k=1}^8 \binom{r_k}{s_k} \sum_{i=1}^8 s_i \int_{-\infty}^{\infty} [1 - \Phi_i(x)]^{s_i-1} \prod_{j=1}^8 \Phi_j(x)^{s_j-1} \prod_{j \neq i} [1 - \Phi_j(x)]^{s_j} f_i(x) dx \quad (9)$$

$$\eta_{RS} = \Pr\{X_i = R | Y_{i-1} = s\} = \frac{N!}{\prod_{k=1}^8 r_k!} \prod_{k=1}^8 \lambda_k^{s_k} \quad (10)$$

where $f_i(x)$ and $\Phi_i(x)$ are defined similarly to those in Equation 4. $\Lambda = (\lambda_1 \lambda_2 \dots \lambda_8)'$ can be calculated from product of matrices \mathbf{G} and $\mathbf{S} = (s_1 s_2 \dots s_8)'$ as

$$\Lambda = \frac{1}{N} \mathbf{G} \mathbf{S}', \quad (11)$$

where the matrix \mathbf{G} has a form that depends on the relative location of the marker to the QTL. If the marker locates at the left side of the linked QTL,

$$\mathbf{G} = \begin{pmatrix} (1 - c_1)(1 - c_2)/2 & 0 & 0 & 0 \\ (1 - c_1)(1 - c_2)/2 & 1 & 1/2 & [(1 - c_1)(1 - c_2) + c_1c_2]/2 \\ c_1c_2/2 & 0 & 1/2 & [c_1(1 - c_2) + c_2(1 - c_1)]/2 \\ c_1c_2/2 & 0 & 0 & [(1 - c_1)(1 - c_2) + c_1c_2]/2 \\ (1 - c_1)c_2/2 & 0 & 0 & 0 \\ (1 - c_1)c_2/2 & 0 & 0 & [c_1(1 - c_2) + c_2(1 - c_1)]/2 \\ c_1(1 - c_2)/2 & 0 & 0 & 0 \\ c_1(1 - c_2)/2 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ (1 - c_1)/2 & 1/2 & (1 - c_2)/2 & 1/2 \\ c_1/2 & 0 & c_2/2 & 0 \\ 0 & 0 & 0 & 0 \\ (1 - c_1)/2 & 0 & 0 & 0 \\ 0 & 1/2 & c_2/2 & 0 \\ 0 & 0 & (1 - c_2)/2 & 0 \\ c_1/2 & 0 & 0 & 1/2 \end{pmatrix};$$

when the marker is between the linked QTL,

$$\mathbf{G} = \begin{pmatrix} (1 - c_1)(1 - c_2)/2 & 0 & 0 & 0 \\ (1 - c_1)(1 - c_2)/2 & 1 & 1/2 & (1 - c_2)/2 \\ c_1(1 - c_2)/2 & 0 & 1/2 & 0 \\ c_1(1 - c_2)/2 & 0 & 0 & (1 - c_2)/2 \\ (1 - c_1)c_2/2 & 0 & 0 & 0 \\ (1 - c_1)c_2/2 & 0 & 0 & c_2/2 \\ c_1c_2/2 & 0 & 0 & 0 \\ c_1c_2/2 & 0 & 0 & c_2/2 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ (1 - c_1)/2 & 1/2 & [(1 - c_1)(1 - c_2) + c_1c_2]/2 & 1/2 \\ c_1/2 & 0 & [c_1(1 - c_2) + (1 - c_1)c_2]/2 & 0 \\ 0 & 0 & 0 & 0 \\ (1 - c_1)/2 & 0 & 0 & 0 \\ 0 & 1/2 & [c_1(1 - c_2) + (1 - c_1)c_2]/2 & 0 \\ 0 & 0 & [(1 - c_1)(1 - c_2) + c_1c_2]/2 & 0 \\ c_1/2 & 0 & 0 & 1/2 \end{pmatrix};$$

and if the marker locates at the right of the linked QTL,

$$\mathbf{G} = \begin{pmatrix} (1 - c_1)(1 - c_2)/2 & 0 & 0 & 0 \\ (1 - c_1)(1 - c_2)/2 & 1 & 1/2 & (1 - c_2)/2 \\ c_1(1 - c_2)/2 & 0 & 1/2 & 0 \\ c_1(1 - c_2)/2 & 0 & 0 & (1 - c_2)/2 \\ c_1c_2/2 & 0 & 0 & 0 \\ c_1c_2/2 & 0 & 0 & c_2/2 \\ (1 - c_1)c_2/2 & 0 & 0 & 0 \\ (1 - c_1)c_2/2 & 0 & 0 & c_2/2 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ [(1 - c_1)(1 - c_2) + c_1c_2]/2 & 1/2 & (1 - c_1)/2 & 1/2 \\ [c_1(1 - c_2) + (1 - c_1)c_2]/2 & 0 & c_1/2 & 0 \\ 0 & 0 & 0 & 0 \\ [(1 - c_1)(1 - c_2) + c_1c_2]/2 & 0 & 0 & 0 \\ 0 & 1/2 & c_1/2 & 0 \\ 0 & 0 & (1 - c_1)/2 & 0 \\ [c_1(1 - c_2) + (1 - c_1)c_2]/2 & 0 & 0 & 1/2 \end{pmatrix}.$$

Mean and variance of the marker heterozygosity: The marginal probabilities $\pi_{ik} = \Pr\{x_{ik} = r_{ik}\}$ and $\pi_{ij} = \Pr\{x_{ij} = r_{ij}\}$ can be computed from the above joint probability distributions as $\pi_{ik} = \sum_{l \neq k} \Pr\{x_{il} = r_{il}, l = 1, 2, \dots, K | \sum_{i=1}^K r_i = N\}$ and $\pi_{ij} = \sum_{l \neq i, j} \Pr\{x_{il} = r_{il}, l = 1, 2, \dots, K | \sum_{i=1}^K r_i = N\}$ with $K = 4$ or 8 for the two or three loci model, respectively. Thus, the expected mean and variance of the marker heterozygosity can be calculated as

$$H_i = \begin{cases} \frac{1}{N} \sum_{i=1}^N (\pi_{i1} + \pi_{i3}) i & \text{for the two loci model} \\ \frac{1}{N} \sum_{i=1}^N (\pi_{i1} + \pi_{i4} + \pi_{i5} + \pi_{i8}) i & \text{for the three loci model} \end{cases} \quad (12)$$

and

$$V_i = \frac{F-1}{N^2 F} \left\{ \sum_{i=1}^N (\pi_{i1} + \pi_{i3}) i^2 + 2 \sum_{i=0}^N \sum_{j=0}^N \pi_{i3} i j - \left[\sum_{i=0}^N (\pi_{i1} + \pi_{i3}) i \right]^2 \right\} \quad (13)$$

for the two loci model or

$$V_i = \frac{F-1}{N^2 F} \left\{ \sum_{i=1}^N (\pi_{i1} + \pi_{i4} + \pi_{i5} + \pi_{i8}) i^2 + 2 \sum_{i=0}^N \sum_{j=0}^N (\pi_{i4} + \pi_{i5} + \pi_{i8} + \pi_{i5} + \pi_{i8} + \pi_{i8}) i j - \left[\sum_{i=0}^N (\pi_{i1} + \pi_{i4} + \pi_{i5} + \pi_{i8}) i \right]^2 \right\} \quad (14)$$

for the three loci model. In theory, the above analysis may be extended to any number of marker loci and QTL but the algebra involved would be very tedious.

Recurrent selection and backcross or intercross scheme: Here we exploited the impact of introducing intercross among the selected individuals into the RSB scheme on maintenance of heterozygosity at a single QTL. To model the dynamics of the RSBI, let $X_i = (x_{i1} x_{i2} x_{i3})'$ and $Y_i = (y_{i1} y_{i2} y_{i3})'$ be two vectors whose elements

x_{ik} (or y_{ik}) are a random number of individuals with genotype i in one of the independent families before (or after) selection ($i = 1, 2, 3$ corresponding to the QTL genotypes AA, Aa , and aa respectively). Probability distributions of these random vectors are given by

$$p_{iR} = \Pr\{X_i = R\} = \Pr\{x_{ik} = r_k, k = 1, 2, 3\} \\ = \sum_S \Pr\{Y_{t-1} = S\} \Pr\{X_i = R | Y_{t-1} = S\} = \sum_S q_{t-1} \eta_{iRS}^{(X)} \quad (15)$$

$$q_{iS} = \Pr\{Y_i = S\} = \Pr\{y_{ik} = s_k, k = 1, 2, 3\} \\ = \sum_R p_{iR} \prod_{k=1}^3 \binom{r_k}{s_k} \sum_{i=1}^3 s_i \int_{-\infty}^{\infty} [1 - \Phi_i(x)]^{s_i-1} \prod_{j=1}^3 \Phi_j(x)^{r_j-s_j} \\ \times \prod_{j \neq i}^3 [1 - \Phi_j(x)]^{r_j} f_i(x) dx. \quad (16)$$

The conditional probability $\eta_{iRS}^{(X)}$ depends on whether the selected individuals are backcrossed to the recurrent line ($X = B$) or intercrossed to each other ($X = I$) and has forms as

$$\eta_{iRS}^{(B)} = \binom{n}{r_2} [(2s_1 + s_2)/n]^{r_2} [(s_2 + 2s_3)/n]^{r_3} \quad (17)$$

$$\eta_{iRS}^{(I)} = \frac{N!}{\prod_{k=1}^3 r_k!} \prod_{k=1}^3 \phi_k^{r_k}, \quad (18)$$

where $\phi_1 = [4s_1(s_1 - 1) + 2s_1s_2 + s_2(s_2 - 1)]/4n(n - 1)$, $\phi_2 = [2s_1s_2 + 4s_1s_3 + s_2(s_2 - 1) + 2s_2s_3]/2n(n - 1)$ and $\phi_3 = [s_2(s_2 - 1) + 2s_2s_3 + 4s_3(s_3 - 1)]/4n(n - 1)$. The dynamics of the probability distributions can be readily evaluated using the initial condition $p_{iR} = \Pr\{X_i = R\} = (n!/r_2!r_3!)(1/2)^{(r_2+r_3)}$. Similarly, the mean and variance of heterozygosity at the QTL in generation t of the RSBI scheme are given by $H_t = \sum_{i=1}^N \pi_{i2}i/N$ and $V_t = [\sum_{i=1}^N \pi_{i2}i^2 - (\sum_{i=1}^N \pi_{i2}i)^2] \times (F - 1)/N^2F$, respectively.

NUMERICAL ANALYSES AND SIMULATION STUDY

For simplicity but without loss of generality, we considered one chromosome on which there were 100 evenly distributed marker loci. Map distance between a pair of adjacent loci is 1 cM (approximately equivalent to a recombination frequency of 0.01 under Haldane's mapping function). Among the 100 loci, 1 or 2 were assigned to be within the genes underlying a quantitative trait in the numerical analyses of the theoretical model represented in the previous section. In the simulation study discussed below, 3 of the polymorphic marker loci were assumed to locate at the same positions as 3 QTL on the chromosome. The rest of the markers were devoid of effect on the quantitative trait. Assignment of some marker loci to the same polymorphic sites within QTL was based on two considerations. First, single nucleotide polymorphisms within coding sequences of functional genes have proved abundant in genomes

whose complete sequence data are available (*i.e.*, yeast, *Caenorhabditis elegans*, *Drosophila*, *Arabidopsis*, etc.) or are to be available (*i.e.*, human, mice, pig, cattle, rice, etc.). Second, use of extremely dense marker maps allows us to exploit the maximum efficiency of the RSB mapping strategy in identifying the QTL map locations.

Numerical analyses: For a given set of parameters defined in the previous theoretical analysis, the mean and variance of heterozygosity maintained at a marker locus can be worked out numerically. The only technical difficulty in the numerical analysis is the limitation in computer time and memory for evaluating Equations 2, 6, and 7. These equations involve summation of a huge number of terms. The total number of terms in the summation is equivalent to the number of different configurations of r_i ($i = 1, 2, \dots, K$ with $K = 4$ or 8 under the two and three loci model, respectively) such that $N = \sum_{i=1}^K r_i$. A general formula for this number is given by $c(K, N) = \prod_{i=2}^K (N + i - 1)/(K - 1)!$, which takes a value of 264,385,836 when $K = 8$ and $N = 50$. Thus, numerical analysis demonstrated here has to be restricted to the cases with a small family size.

Figure 2 illustrates distributions of the means and variances of heterozygosity at 100 marker loci that were linked with (a) one, (b) two, or (c) three QTL in the various RSB breeding programs. Figure 2, a and b, was obtained from theoretical predictions and Figure 2c was calculated from averages of 100 repeated simulations of the RSB scheme whose parameters were given as the scheme 1 in Table 1. The pattern of change in the mean and variance of heterozygosity maintained at the marker loci directly reflected their locations relative to that of the QTL. The peak of the mean curves always occurred at the same locations as the QTL regardless of the QTL effects and the other parameters. On the other hand, there was a rapid decline in the marker heterozygosity as its mapping distance from the QTL increased. However, the change in the variance over the marker loci showed two different patterns. Whenever there was a substantial level of heterozygosity maintained around the selected loci, the variance was observed to be lower at the selected loci than their nearby markers (scheme 2 in Figure 2a and the scheme in Figure 2c), while the variance took a peak value at the selected loci to those of their neighboring markers if the heterozygosity had drifted to a low level (schemes 1–4 in Figure 2a and all schemes in Figure 2b). The difference in the pattern of the variance curves can be explained by noting three facts. First, the variance of heterozygosity in the present context is equivalent to the variance in frequency of the gene favored by selection. The variance in gene frequency due to genetic drift is proportional to the gene frequency. Second, selection efficiency (*i.e.*, the absolute change in frequency of the gene under selection) is inversely proportional to the allele frequency. The smaller the frequency, the more efficient the selection will be in driving up the frequency. Third, the hitchhik-

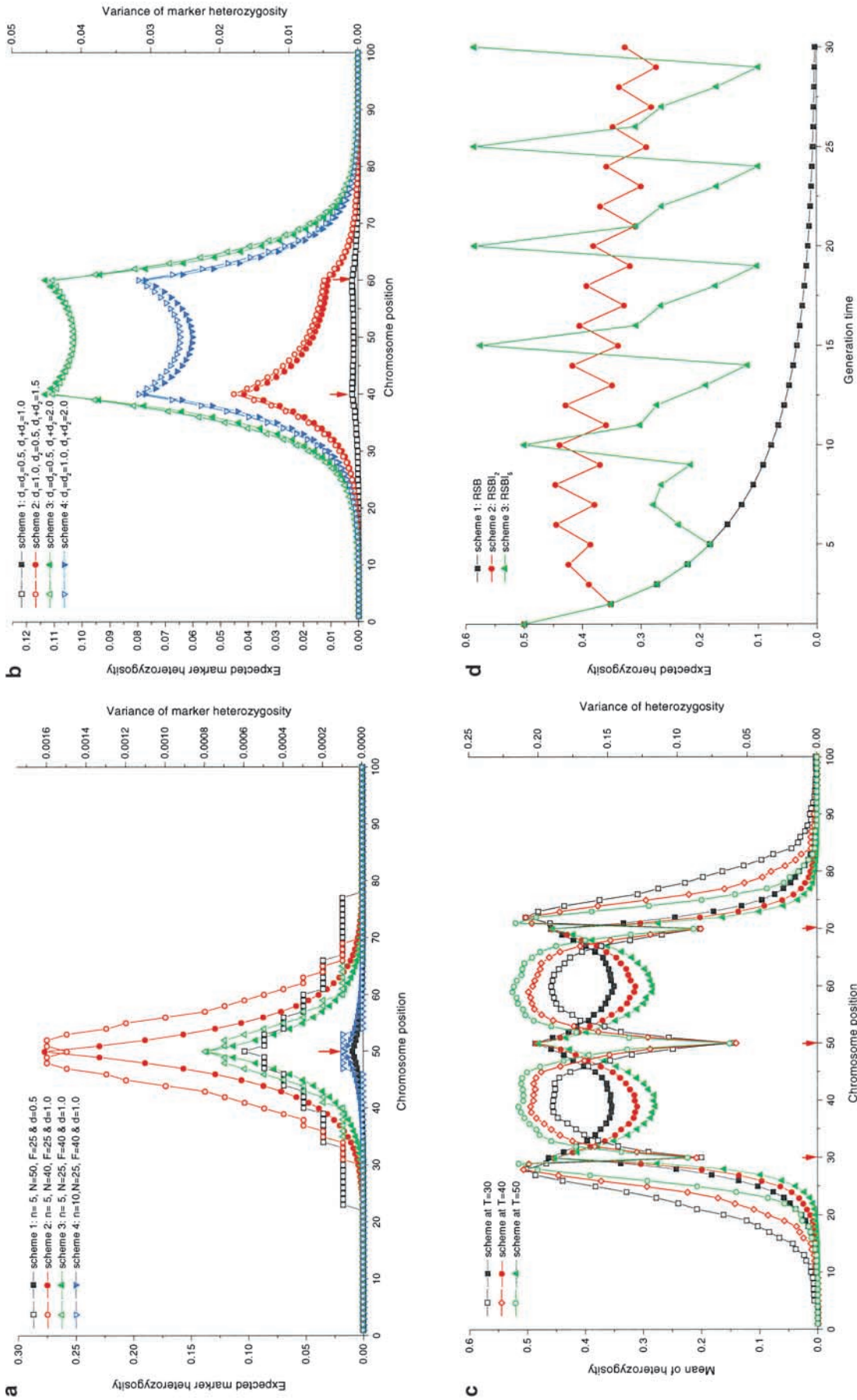


FIGURE 2.—Expected value (solid symbols) and the variance (open symbols) of heterozygosity at the marker loci linked to varying numbers of QTL under different design parameters: (a) one QTL and all schemes with $T = 20$; (b) two QTL and all schemes with $n = 2, N = 10, F = 50$, and $T = 20$; (c) three QTL and all schemes with $n = 5, N = 50, F = 20$, and $h^2 = 0.5$; and (d) one QTL and all schemes with $n = 5, N = 50, F = 20$, and $d = 0.5$, where n, N, F, T, h^2 , and d are accordingly the number of individuals selected, the family size, the number of generations of consecutive selection and backcrossing, heritability, and additive effect of QTL. Red arrows indicate location of the QTL.

TABLE 1
The parameters defining the 16 simulated breeding schemes

Scheme	n	N	F	MD	h^2	GEM	T_1	T_2	T_3	QTL _L	QTL _M	QTL _R
1	5	50	20	RSB	0.50	AEE	30	40	50	30	50	70
2	5	50	20	RSB	0.50	AUE	30	40	50	30	50	70
3	5	50	20	RSB	0.50	AEE	30	40	50	30	50	60
4	5	100	10	RSB	0.50	AEE	30	40	50	30	50	70
5	5	50	20	RSB	0.30	AEE	10	20	30	30	50	70
6	10	50	20	RSB	0.30	AEE	10	20	30	30	50	70
7	5	50	20	RSB	0.30	AUE	10	20	30	30	50	70
8	5	50	20	RSB	0.30	AEE	10	20	30	40	50	60
9	5	50	20	RSB	0.30	EEM	10	20	30	40	50	60
10	5	50	20	RSBI ₃	0.30	AEE	10	20	30	40	50	60
11	5	50	20	RSBI ₅	0.30	AEE	10	20	30	40	50	60
12	5	50	20	RSBI _p	0.30	AEE	10	20	30	40	50	60
13	5	50	20	RSBI _p	0.30	AUE	10	20	30	40	50	60
14	5	50	20	RSBI _p	0.30	EEM	10	20	30	40	50	60
15	5	50	20	RSB	0.15	AEE	10	20	30	30	50	70
16	5	50	30	RSBI _p	0.15	AEE	10	20	30	30	50	70

n , N , and F are the number of selected individuals, the size of each family, and the number of families. The three mating designs (MD) are RSB, recurrent selection and backcross; RSBI _{k} , recurrent selection, backcross at every k generation otherwise intercross; and RSBI _{p} , recurrent selection, backcross whenever family phenotypic mean reaches that of backcross family at the first generation otherwise intercross. h^2 , heritability of the trait. The three QTL genetic effect models (GEM) are AEE, additive and equal effect; AUE, additive unequal effect; and EEM, epistatic effect model. T_i ($i = 1, 2, 3$) is the generation time at which the marker heterozygosity was surveyed; QTL _{X} ($X = L, M, R$) are the locations of the left, middle, and right QTL on the simulated chromosome.

ing effect of the selected locus on the nearby neutral markers depends on the selection efficiency at the selected locus and the linkage disequilibrium between the selected and the neutral loci. The low heterozygosity maintained at the selected locus indicates low effectiveness of selection at the locus and thus the dynamics of heterozygosity at the nearby marker loci were less affected by hitchhiking but dominated by genetic drift. This resulted in a much lower allele frequency at the marker loci and thus a smaller variance in their allele frequencies. In contrast, a high gene frequency at the QTL reflects that the influence of genetic drift on frequency is effectively counterbalanced by selection. The strong hitchhiking effect maintained a high gene frequency for those markers in the vicinity of the selected loci before complete breakage of linkage disequilibria between the QTL and markers, while drift in the marker gene frequencies was balanced less effectively than that in the QTL gene frequencies. Thus, a larger variance in gene frequency might be observed at these markers than at the nearby selected loci. The analysis revealed that not only the mean but also the variance of the marker heterozygosity provide essential information regarding locations of the QTL and regarding how genes at both QTL and marker loci in the RSB populations evolve under a complicated combination of evolutionary factors of recombination, genetic drift, and selection.

It is clear from the figures that there was rapid fixation

at the QTL with small effect in the RSB breeding program. However, the program may be improved in properly designed experiments. For a given number of individuals involved in the experiment, the breeding scheme with larger family sizes but smaller number of families (scheme 2 in Figure 2a) was superior in maintaining a high level of heterozygosity to that with smaller family size but larger number of families (scheme 3 in Figure 2a). Selection intensity played a significant role in slowing down the allele loss at QTL due to genetic drift, particularly when the QTL had a low genetic contribution to the trait. Figure 2b presents the analysis of two linked QTL under three models of genetic effects: (i) the additive and equal effect model (schemes 1 and 4), where the two QTL contributed equally and additively to the trait; (ii) the additive and unequal effect model (scheme 2) under which the two QTL affected the trait additively but the first QTL had twice as large an effect as the second QTL; and (iii) the epistatic model (scheme 3), where the individuals carrying the trait-increasing allele at each of the two QTL had a genotypic value that was fourfold those carrying only one such allele. A distinct feature observed from comparing the three models is that epistasis had remarkable influence on the marker heterozygosity. Under the epistatic model, the individuals carrying the increasing allele at every QTL had a much better chance of being selected in comparison to the corresponding additive model. This anticipates a strong trend for the nonre-

combinant gametes carrying all increasing alleles to be selected and in turn results in a substantial increase in the level of heterozygosity at these loci on the one hand and a noticeable decrease in recombination between the loci on the other.

Loss of heterozygosity at the QTL with small effect was very fast in the RSB program. Reducing the amount of backcrossing by incorporation of intercrossing into the program was effective in slowing down the allelic loss particularly when intercrossing was frequent (scheme 2 in Figure 2d), but this caused a large fluctuation in the gene frequency. The effect of the RSBI on inferring the QTL locations is discussed in the following simulation study.

Simulation study: We developed a series of computer simulation programs that offer a high degree of flexibility in mimicking the RSB or RSBI schemes specified with various mating design parameters, different genetic architectures of quantitative traits, and arbitrary linkage relationships between the marker loci and QTL. In a single meiosis, the “random walk” procedure has been described elsewhere (LUO and KEARSEY 1992) to simulate genetic recombination between linked loci. Chiasmata interference, sexual differentiation in recombination frequency, and segregation distortion were assumed to be absent in the simulation model.

As described in the above numerical analysis, we considered only one chromosome in the simulation study. There were 100 evenly distributed marker loci on the chromosome, 3 of them affecting a quantitative trait. Map distance between adjacent loci is constantly 1 cM. For a given proportion of quantitative genetic variation explained by the 3 QTL (h^2), three models under which h^2 was resolved into genetic effect of QTL were considered to investigate robustness of the mapping strategy to various QTL effect models. These include (i) the additive equal effect (AEE) model, under which each allele increasing the trait delivered an additive contribution of $d = [2h^2/3(1 - h^2)]^{1/2}$ to the trait phenotype; (ii) the additive unequal effect (AUE) model, under which genetic effects of the increasing alleles at the three QTL are, respectively, d , $d/2$, and $d/4$ with $d = [32h^2/21(1 - h^2)]^{1/2}$; and (iii) the epistatic effect model (EEM), under which any individual carrying $k(\leq 2)$ increasing alleles would have a genotypic effect of $[2kh^2/3(1 - h^2)]^{1/2}$, but the genotypic effect was 2.0 when it carried all 3 increasing alleles (*i.e.*, $k = 3$). The individual phenotype was determined by its genotypic effect plus a number that was randomly sampled from a standard normal distribution.

The breeding program was initiated with F backcross families obtained from two homozygous inbred parental lines, which were fixed with different alleles at the 100 loci. After the start of the breeding program, three mating strategies might be performed: (i) RSB, (ii) recurrent selection and either intercrossing or backcrossing in every Δt generation(s) (RSBI $_{\Delta t}$), and (iii) recurrent

selection and either intercrossing or backcrossing whenever the family mean at generation t was not lower than that at generation 1 (RSBI $_p$). Intercrossing was simulated as random mating among the selected individuals. This includes the possibility of selfing of some selected individuals, but this would not seriously influence our results. The parameters defining the 16 simulated schemes are listed in Table 1. Each of the simulated schemes was repeated 100 times unless otherwise stated.

Means and standard deviations of heterozygosity maintained at the QTL were calculated from the repeated simulation trials for all these simulated schemes and are illustrated in Table 2. It can be seen from the table that there was a general trend in loss of heterozygosity at the QTL as the breeding schemes evolved (from T_1 to T_3). However, the rate of loss of the heterozygosity was influenced by almost all parameters defining the breeding program. Heritability was a dominant factor influencing the RSB schemes. Given the other parameters, the selected alleles at the QTL, which contributed 50% of phenotypic variance, were maintained at an unchanged frequency after 50 generations of RSB (scheme 1), whereas the alleles had almost completely vanished after 30 generations of RSB if the QTL explained only 15% of the phenotypic variance (scheme 15). However, incorporation of intercrossing in the RSB breeding schemes was effective in reducing loss of the increasing alleles at the QTL (scheme 16). Comparison of heterozygosity at the QTL between schemes 15 and 16 revealed that use of family phenotypic mean was an effective way to determine the switch between backcrossing and intercrossing during the breeding program such that the selected alleles may be effectively protected from being lost. As has been shown in the previous numerical analysis, the genetic model of the QTL effect influenced the heterozygosity loss remarkably. Epistasis in the QTL effects dramatically slowed down the loss of the QTL heterozygosity (scheme 9) compared to the corresponding additive models (schemes 7 and 8). A higher level of heterozygosity was maintained at more closely linked QTL (scheme 9) than at the less closely linked QTL (scheme 5). For a given experimental size ($N \times F$), the scheme with larger family size but smaller number of families (scheme 4) was more effective than the scheme with smaller family size but larger number of families (scheme 1) for maintenance of the heterozygosity at the selected loci.

Locations of the QTL were inferred in the simulation study as locations of the marker loci at which the heterozygosity curve at generation 30 reached the peak values, and the accuracy in locating the QTL was evaluated as the percentage of the peak values of the same curve occurring at the simulated QTL locations in repeated simulations. Tabulated in Table 3 were the percentages of the correct locations of the QTL and means and standard deviations of the inferred QTL locations. It can be seen that a major factor determining accuracy

TABLE 2
Means and standard deviations (in parentheses) of heterozygosity maintained at the QTL at different generation times in the simulated breeding schemes

Scheme	T_1			T_2			T_3		
	QIL _L	QIL _M	QIL _R	QIL _L	QIL _M	QIL _R	QIL _L	QIL _M	QIL _R
1	0.45 (0.03)	0.49 (0.02)	0.46 (0.03)	0.45 (0.03)	0.48 (0.02)	0.46 (0.03)	0.45 (0.03)	0.49 (0.02)	0.46 (0.04)
2	0.48 (0.02)	0.43 (0.04)	0.13 (0.06)	0.49 (0.02)	0.44 (0.04)	0.09 (0.06)	0.48 (0.02)	0.04 (0.06)	0.06 (0.04)
3	0.48 (0.02)	0.50 (0.02)	0.48 (0.02)	0.48 (0.02)	0.50 (0.02)	0.48 (0.02)	0.48 (0.02)	0.49 (0.02)	0.48 (0.02)
4	0.39 (0.04)	0.45 (0.02)	0.38 (0.04)	0.37 (0.04)	0.44 (0.03)	0.36 (0.04)	0.35 (0.05)	0.43 (0.03)	0.34 (0.05)
5	0.37 (0.05)	0.43 (0.03)	0.38 (0.05)	0.32 (0.06)	0.39 (0.05)	0.33 (0.06)	0.28 (0.07)	0.36 (0.06)	0.29 (0.06)
6	0.35 (0.04)	0.40 (0.03)	0.34 (0.04)	0.31 (0.04)	0.38 (0.03)	0.31 (0.05)	0.28 (0.05)	0.35 (0.04)	0.28 (0.05)
7	0.43 (0.03)	0.36 (0.05)	0.17 (0.06)	0.41 (0.04)	0.30 (0.06)	0.07 (0.04)	0.39 (0.04)	0.26 (0.06)	0.03 (0.03)
8	0.44 (0.04)	0.46 (0.03)	0.44 (0.03)	0.43 (0.04)	0.46 (0.03)	0.43 (0.04)	0.41 (0.05)	0.45 (0.03)	0.41 (0.04)
9	0.47 (0.02)	0.47 (0.02)	0.47 (0.02)	0.47 (0.02)	0.47 (0.02)	0.47 (0.02)	0.48 (0.02)	0.47 (0.02)	0.47 (0.02)
10	0.87 (0.04)	0.91 (0.03)	0.87 (0.04)	0.49 (0.02)	0.50 (0.02)	0.48 (0.02)	0.49 (0.02)	0.50 (0.01)	0.49 (0.02)
11	0.48 (0.02)	0.49 (0.02)	0.48 (0.02)	0.48 (0.02)	0.49 (0.02)	0.47 (0.03)	0.47 (0.02)	0.49 (0.02)	0.47 (0.03)
12	0.27 (0.08)	0.28 (0.08)	0.27 (0.08)	0.16 (0.07)	0.17 (0.08)	0.15 (0.07)	0.10 (0.06)	0.11 (0.07)	0.10 (0.06)
13	0.31 (0.09)	0.30 (0.08)	0.23 (0.07)	0.22 (0.08)	0.20 (0.07)	0.11 (0.06)	0.17 (0.07)	0.15 (0.06)	0.06 (0.05)
14	0.40 (0.05)	0.40 (0.05)	0.40 (0.04)	0.33 (0.06)	0.33 (0.06)	0.33 (0.06)	0.30 (0.07)	0.30 (0.07)	0.30 (0.07)
15	0.19 (0.05)	0.23 (0.05)	0.18 (0.04)	0.08 (0.04)	0.11 (0.05)	0.08 (0.05)	0.03 (0.03)	0.05 (0.04)	0.03 (0.03)
16	0.28 (0.06)	0.33 (0.06)	0.27 (0.06)	0.21 (0.08)	0.28 (0.08)	0.22 (0.07)	0.18 (0.07)	0.25 (0.10)	0.18 (0.07)

of QTL mapping was size of the QTL genetic effect. The QTL with effect as large as 1.23 units of the residual standard deviation was located correctly in 87% of the repeated simulation trials, but the accuracy dropped to only 15% if the effect was quartered (scheme 17). As the trait had a very low heritability, increasing alleles at the QTL were almost completely lost after 30 generations of RSB (refer to scheme 15 in Table 2) and the QTL in this scheme was poorly mapped. However, the loss in mapping accuracy was substantially recovered in scheme 16 in which the RSBI_p was performed. The actual map distance between linked QTL showed an obvious influence on their mapping accuracy: The closer the QTL were linked, the poorer their mapping accuracy (scheme 1 *vs.* 3). For a given experimental size, the QTL in the scheme with a larger family size but smaller number of families (scheme 4) were located more accurately than those with a smaller family size but larger number of families (scheme 1). Comparison between schemes 5 and 6 showed that the QTL in the scheme under a stronger selection had resulted in a worse rather than a better mapping precision of QTL. Epistasis in the QTL genetic effects was preferred in the RSB schemes for a better maintenance of genetic heterozygosity at the loci (Table 2), but it hindered rather than improved accumulation of recombination between the QTL and between them and the nearby marker loci and thus resulted in a reduced accuracy.

Although there was variation in the percentage of correct identification of the QTL locations over the various schemes considered in the simulation, the QTL in all these schemes were, on average, mapped to locations that were not significantly different from their actual map locations. The standard deviations of the estimated QTL locations were in the range of 0.49–4.14 cM, and the change in the standard deviation among the different breeding schemes was consistent with change in the percentage of correct QTL locations inferred.

The above discussion represented evaluation of precision in mapping QTL under the RSB or RSBI schemes. It is important to exploit resolution of QTL mapping by use of the marker heterozygosity distribution. An *ad hoc* measure of resolution for an estimated QTL location at the i th marker locus can be $\lambda_i = H_i - (H_{i-1} + H_{i+1})/2$, where H_i denotes the heterozygosity of the i th marker locus. Means and standard errors of the resolution estimates over repeated simulations of all the schemes are tabulated in Table 4. It can be seen that all parameters considered here influenced the mapping resolution. As expected, there was a trend of improvement in the mapping resolution as the breeding schemes progressed, provided a substantial amount of heterozygosity was maintained at the stage when the QTL locations were examined. When the other parameters were fixed, the trait heritability played an important role in determining the mapping resolution. The higher the heri-

TABLE 3

The percentage of correctly inferred QTL locations and means and standard deviations (in parentheses) of estimated QTL map locations

Scheme	% of correct QTL locations			Inferred QTL locations		
	QTL _L	QTL _M	QTL _R	QTL _L	QTL _M	QTL _R
1	67	60	67	30.73 (1.28)	49.88 (1.34)	69.46 (1.08)
2	78	63	24	30.39 (0.84)	49.19 (1.45)	68.02 (2.23)
3	44	41	45	41.87 (1.88)	49.75 (2.30)	58.13 (1.92)
4	88	87	82	30.31 (0.74)	50.11 (0.49)	69.68 (0.73)
5	51	68	59	31.24 (1.67)	49.72 (1.24)	69.16 (1.49)
6	63	76	64	30.91 (1.50)	49.97 (0.85)	69.15 (1.43)
7	87	55	15	30.18 (0.50)	48.72 (1.87)	70.38 (4.14)
8	32	47	35	41.93 (1.86)	50.02 (2.07)	58.06 (1.92)
9	54	29	54	41.32 (1.67)	50.07 (2.90)	58.66 (1.63)
10	38	28	55	41.74 (1.87)	49.80 (2.55)	58.80 (1.71)
11	35	29	37	41.96 (1.88)	49.58 (2.65)	58.20 (1.82)
12	17	28	17	41.57 (3.28)	50.36 (3.58)	59.62 (3.07)
13	36	23	6	40.66 (2.14)	48.49 (2.78)	58.05 (3.68)
14	42	13	33	41.62 (1.99)	49.70 (3.24)	58.25 (1.97)
15	30	31	32	32.60 (2.79)	51.18 (3.15)	70.60 (3.60)
16	47	50	47	31.05 (2.08)	49.94 (1.98)	69.14 (1.79)

tability, the better the QTL was resolved. In contrast to its positive effect on maintenance of heterozygosity, epistasis in the QTL genetic effect showed a negative influence on the mapping resolution of the QTL due to a reduced number of recombinants between the linked QTL and between them and their nearby marker loci during the selection and backcrossing process. This effect was more obvious with the QTL surrounded by other QTL. Population designs with smaller family size but larger number of families (*i.e.*, scheme 1) were more effective in achieving a better mapping resolution. Better resolution observed in such designs may be explained by the fact that with larger numbers of families there is a better chance of maintaining a wider range of different recombinants between the linked QTL themselves and between the QTL and the marker. Selection intensity was less important for mapping resolution. Comparison of the mapping resolution between the RSB schemes and the RSBI schemes showed that reducing the number of backcrossing generations and allowing intercrossing in a RSB scheme had reduced the mapping resolution. More frequent intercrossing tended to worsen the mapping resolution providing frequent backcrossing had not driven the increasing allele at the QTL to a very low frequency.

Comparison of the RSB schemes to the interval mapping-based methods: Interval mapping and its later extended versions have been the most popular methods in QTL mapping in man, plants, and animals. This section compares the RSB-based QTL mapping approach with the two most popularly cited interval mapping methods in the literature: the interval mapping (IM; LANDER and BOTSTEIN 1989) and the composite interval mapping

(CIM; ZENG 1994) methods under the constraint of a constant capacity of genotyping 1000 individuals for 100 linked and evenly spaced marker loci. Of the marker loci, 3 were assumed to be QTL, which explained 30% of phenotypic variance of a quantitative trait. Map distance between any pair of adjacent marker loci was 1 cM ($\sim 1\%$ recombination frequency). For the IM and CIM analyses, a backcross family with 1000 individuals was generated and analyzed by use of the QTL cartographer (BASTEN *et al.* 1994), the computer software for carrying out the IM and CIM analyses. The RSB breeding schemes were simulated with 20 independent families and 50 individuals for each of these families, yielding the same population size as the interval mapping analyses. Selection in the RSB program was for 50 generations at a constant intensity of 10%. It has already been pointed out in the above simulation study that both mean and variance of the marker heterozygosity are informative about relative locations of the marker loci to the selected QTL. To combine the information from these two statistics, a measure for the QTL presence was calculated as $\tau_i = H_{it}/\sqrt{V_{it}}$ when $V_{it} > 0$ otherwise 0, where H_{it} and V_{it} are, respectively, the mean and variance of the i th marker heterozygosity at generation t .

Illustrated in Figure 3 are the distribution of the likelihood-ratio test statistics from the IM and CIM analyses and the distribution of the measure of the QTL presence in the RSB schemes. It was very clear from analysis of the RSB schemes that the QTL locations were accurately and unambiguously identified as the chromosome locations at which τ_i reached its peak value. In addition, the method was very robust to various models of the QTL effects on the trait. The presence of the QTL on the

TABLE 4
Means and standard errors (in parentheses) of the resolution for the QTL mapping at different generation times in the simulated breeding schemes

Scheme	T_1			T_2			T_3		
	QTL _L	QTL _M	QTL _R	QTL _L	QTL _M	QTL _R	QTL _L	QTL _M	QTL _R
1	0.072 (0.003)	0.030 (0.002)	0.070 (0.004)	0.091 (0.003)	0.038 (0.003)	0.092 (0.004)	0.108 (0.004)	0.048 (0.003)	0.109 (0.004)
2	0.077 (0.003)	0.048 (0.003)	0.017 (0.002)	0.100 (0.004)	0.066 (0.003)	0.018 (0.002)	0.116 (0.004)	0.083 (0.004)	0.013 (0.002)
3	0.070 (0.004)	0.016 (0.002)	0.072 (0.004)	0.089 (0.004)	0.020 (0.002)	0.087 (0.004)	0.105 (0.004)	0.021 (0.002)	0.108 (0.004)
4	0.062 (0.002)	0.032 (0.002)	0.061 (0.002)	0.076 (0.003)	0.044 (0.002)	0.077 (0.003)	0.085 (0.003)	0.053 (0.002)	0.086 (0.003)
5	0.022 (0.002)	0.011 (0.001)	0.018 (0.002)	0.035 (0.003)	0.022 (0.002)	0.029 (0.002)	0.040 (0.003)	0.027 (0.002)	0.040 (0.003)
6	0.018 (0.001)	0.012 (0.001)	0.019 (0.002)	0.033 (0.002)	0.021 (0.002)	0.032 (0.002)	0.042 (0.002)	0.030 (0.002)	0.045 (0.003)
7	0.028 (0.002)	0.015 (0.002)	0.008 (0.001)	0.048 (0.003)	0.026 (0.002)	0.006 (0.001)	0.072 (0.003)	0.032 (0.003)	0.003 (0.001)
8	0.024 (0.002)	0.008 (0.001)	0.022 (0.002)	0.047 (0.003)	0.013 (0.002)	0.042 (0.003)	0.060 (0.004)	0.020 (0.002)	0.058 (0.003)
9	0.027 (0.002)	0.004 (0.001)	0.021 (0.002)	0.048 (0.003)	0.008 (0.001)	0.043 (0.003)	0.069 (0.004)	0.013 (0.002)	0.065 (0.003)
10	0.041 (0.003)	0.011 (0.001)	0.038 (0.003)	0.039 (0.002)	0.011 (0.002)	0.039 (0.003)	0.048 (0.003)	0.012 (0.002)	0.051 (0.003)
11	0.020 (0.002)	0.005 (0.001)	0.014 (0.002)	0.041 (0.003)	0.010 (0.001)	0.032 (0.002)	0.057 (0.003)	0.016 (0.002)	0.052 (0.003)
12	0.005 (0.002)	0.001 (0.001)	0.003 (0.002)	0.005 (0.002)	0.005 (0.002)	0.009 (0.002)	0.008 (0.002)	0.007 (0.002)	0.009 (0.002)
13	0.007 (0.002)	0.004 (0.002)	0.006 (0.002)	0.014 (0.003)	0.008 (0.002)	0.008 (0.002)	0.017 (0.002)	0.011 (0.002)	0.004 (0.002)
14	0.017 (0.002)	0.003 (0.001)	0.015 (0.002)	0.030 (0.003)	0.005 (0.001)	0.026 (0.003)	0.041 (0.003)	0.005 (0.002)	0.041 (0.003)
15	0.008 (0.001)	0.008 (0.001)	0.008 (0.001)	0.008 (0.001)	0.009 (0.001)	0.005 (0.001)	0.005 (0.001)	0.006 (0.001)	0.004 (0.001)
16	0.010 (0.002)	0.008 (0.002)	0.011 (0.002)	0.019 (0.003)	0.021 (0.003)	0.019 (0.003)	0.026 (0.003)	0.030 (0.003)	0.023 (0.003)

simulated chromosome was strongly evident from the interval mapping methods, but in sharp contrast, it did not provide clear-cut inference of the locations of the QTL. The interval mapping predictions of the QTL locations worsened when there was epistasis in the QTL effects.

DISCUSSION

The present article develops a theoretical framework for predicting the mean and variance of heterozygosity maintained at marker loci linked to one or two QTL for any number of generations using the recurrent selection and backcross schemes previously proposed by WRIGHT (1952) and studied by HILL (1998). The theoretical prediction takes appropriate account of the dynamic change in linkage disequilibria between the QTL themselves and between the QTL and the marker loci due to selection, recombination, and genetic drift during the breeding program. In principle, it is tractable to extend the analysis to more than the three loci modeled here because the distribution of multiple loci linkage disequilibria under the present setting is equivalent to that under a multiple loci haplotype model of linkage disequilibria. Nevertheless, numerical evaluation of the multiple loci system will be computationally very demanding when the experimental size is large. More complicated models were investigated in the simulation study. The analyses demonstrated that distributions of mean and variance of heterozygosity at different marker loci in the RSB breeding program provide sufficient information regarding their relative locations to the QTL under selection in the program and regarding the evolutionary driving factors behind the breeding populations. Appropriate use of these statistics may provide a simple but efficient alternative approach for mapping complex quantitative genetic variation at a substantially improved precision and resolution. The major features of the RSB-based QTL mapping schemes can be summarized as follows:

1. The mapping strategy is powerful in identifying the polymorphic sites, which are in close linkage (*i.e.*, 1 or 2 cM) to the QTL. Use of the dense marker maps that include the genomic polymorphisms within the QTL (cSNP, for instance) enables precise identification of the map locations of the QTL. In general, the error in the inference of the QTL locations will not be beyond one or two times the coverage density of the marker maps.
2. Maintenance of the selected genes at a nontrivial frequency is a prerequisite for achieving both precision and resolution in inferring their map locations under the RSB framework. For QTL with large genetic effects on the trait phenotype, selection on the trait is usually efficient enough to counterbalance dilution of the recipient genome regions around the selected loci over repeated backcrossing and genetic

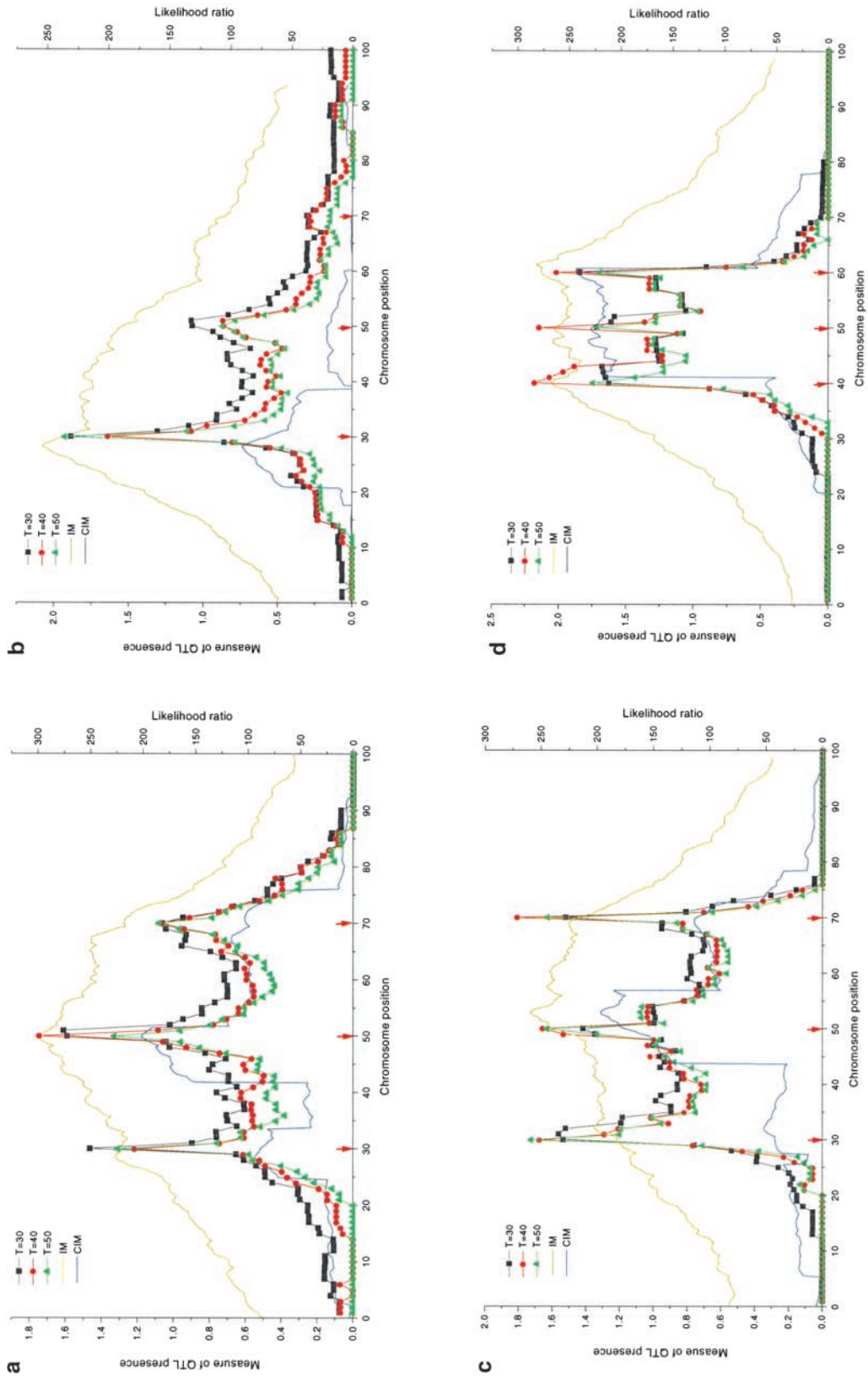


FIGURE 3.—Comparison of the RSB schemes to interval mapping (IM) and composite interval mapping (CIM) of a trait controlled by three QTL, which explained 30% of phenotypic variation under different models of their genetic effects and varying mapping distances between linked QTL: (a) additive equal QTL effect and 20 cM between adjacent QTL, (b) additive unequal QTL effect and 20 cM between adjacent QTL, (c) epistatic QTL effect and 20 cM between adjacent QTL, and (d) epistatic QTL effect and 10 cM between adjacent QTL. The red arrow indicates the simulated location of the QTL.

drift that causes loss of the selected genes over the breeding process. Selection in the RSB schemes is usually not able to prevent QTL with small effects from quick gene fixation. Instead of repeatedly backcrossing the selected individuals to the recurrent parental line, incorporation of intercrossing among the selected individuals at some stages of the breeding program is effective in maintaining the selected QTL alleles for long enough to break down linkage disequilibrium between the QTL and the nearby marker loci so that the QTL may be isolated from their closely linked genome regions.

3. Many factors that can be managed by experimentalists play an important role in determining success of the RSB-based approach in the QTL mapping. The number of generations of the breeding program affects the degree of breakage in linkage disequilibrium between the QTL and between them and their nearby marker loci. A prerequisite for precision and high-resolution mapping of the QTL using this program is sufficient breakage of the linkage disequilibria. In principle, the theory developed in this study can easily be extended to provide the estimate of the number of generations required to reach independent segregation between the linked QTL and between the QTL and their linked marker loci. After establishment of linkage equilibrium between these loci, selection will maintain a slow change in gene frequency at the QTL and, at the same time, genetic drift will drive the marker gene to be fixed very quickly, yielding a clear scenario of the QTL locations as showed in Figure 3. For a given size of experiment and whenever possible, experimental designs with a larger family but smaller number of families are superior in reducing the effect of genetic drift and thus maintaining higher allele frequencies at the selected loci, but may be inferior in obtaining a better mapping resolution of the QTL in comparison to the designs with smaller family size but a larger number of families.
4. The RSB-based QTL mapping approach is quite robust to various models of genetic effects of QTL even though there is a difference in the mapping resolution for the QTL affecting quantitative traits under different models of genetic effects. Positive epistasis in the QTL effects enhanced effectiveness of their selection. This resulted in two consequences: an increased heterozygosity at the QTL but a decreased resolution in their mapping when compared to the corresponding additive model.
5. The QTL mapping approach discussed above has several advantages over the interval mapping method and its extended versions. It takes full advantage of a dense marker map and uses accumulated recombinations between the QTL and their nearby marker loci. This yields precise identification of the QTL locations or accurate inference of their locations at

a mapping resolution as close as 1 cM or less. In contrast, there is an upper limit in marker density by which the efficiency of the interval mapping methods will not be improved further (LANDER and SCHORK 1994). On the other hand, the informative meioses are practically limited in the segregating populations for which the interval mapping methods were developed. These make the QTL mapping inferred from the methods far from satisfactory at the same criterion as the RSB mapping results (Figure 3). The RSB mapping approach is robust to errors in estimation of map distances between the marker loci whereas methods based on interval mapping are sensitive to these errors (LUO and KEARSEY 1992). The practical implementation of the RSB mapping schemes does not need any complicated statistical modeling of the experimental data. Given that many years of considerable research efforts to isolate genes affecting complex traits have resulted in slow progress, we would not consider the long duration of the RSB breeding program to be an expensive investment for significant improvement in mapping precision and resolution in the QTL locations that may lead directly to cloning of QTL. It may not be strictly appropriate to compare the multiple generation approach of RSB-based QTL mapping to interval mapping analysis, which is based on the populations of a single segregating generation. Among the several multiple generation approaches suggested in the literature, use of the advanced intercross line has been shown theoretically to have potential for improving mapping resolution of QTL localization (DARVASI and SOLLER 1995; XIONG and GUO 1997). These researchers consistently revealed a diminished return in the resolution improvement after eight generations of continued intercross in an AIL scheme. This property of the AIL approach is in sharp contrast to that of the RSB scheme in which the mapping resolution has been observed in our study to increase steadily as the breeding scheme progresses until complete breakage of linkage disequilibrium between the QTL and their linked marker loci. However, a direct and comprehensive comparison between these two schemes requires that an appropriate statistical method be established to analyze experimental data from these breeding schemes.

Selection in the RSB breeding program is based entirely on the trait phenotype. Appropriate use of marker information will be an effective way of improving the efficiency of selection for the QTL with small effects (LUO *et al.* 1997). The marker-assisted RSB mapping will enable not only the QTL of large effects but those with small effects to be maintained during the breeding procedure, and thus the complete genetic architecture underlying the polygenic variation will, in principle, be uncovered. Other useful information that can be ex-

tracted during the breeding program is the change in pattern of association between the different recombinant marker genotypes and the trait phenotype. A careful assay of the different marker recombinants for their trait phenotype enables direct assessment of whether or not a specific recombinant is informative for resolving the QTL from their nearby marker loci (MACKAY 2001). Taking this information into account in selection allows an efficient accumulation of these informative recombinants and thus further improvement in resolution of the QTL mapping. Although this study serves as only an initial step in illuminating the potential of appropriate use of genetic tools in dissecting complex genetic variation at a molecular level, the analyses and results demonstrated in this article suggest a potentially better future for map-based cloning of genes affecting most evolutionarily, agronomically, and medicinally important complex traits as has been done successfully in FRARY *et al.* (2000).

We thank Professor W. G. Hill for his discussion and encouragement to the research presented here and Dr. Zhao-Bang Zeng for his advice on using the QTL cartographer software. This work was initiated when Z.W.L. was supported by a collaboration grant from the University of Chicago. Z.W.L. is supported by a start-up research grant from the School of Biosciences and a grant from the Medical Research Committee of the University of Birmingham, by China's Basic Research Program "973", the National Science Foundation of China, the QiuShi Foundation, and the Changjiang Scholarship. C.-I.W. is supported by National Institutes of Health and National Science Foundation grants. M.J.K. is supported by research grants from the United Kingdom Biotechnology and Biological Sciences Research Council.

LITERATURE CITED

- ALPERT, K. D., and S. D. TANKSLEY, 1996 High-resolution mapping and isolation of a yeast artificial chromosome containing fw2.2: a major fruit weight quantitative trait locus in tomato. *Proc. Natl. Acad. Sci. USA* **93**: 15503–15507.
- BASTEN, C. J., B. S. WEIR and Z.-B. ZENG, 1994 ZMAP—A QTL cartographer. *Proc. 5th World Congr. Genet. Appl. Livest. Prod.* **22**: 65–66.
- BOEHNKE, M., 1994 Limits of resolution of genetic linkage studies: implications for the positional cloning of human disease genes. *Am. J. Hum. Genet.* **55**: 379–390.
- DARVASI, A., 1997 Interval-specific congenic strain (ISCS): an experimental design for mapping a QTL into a 1-centimorgan interval. *Mamm. Genome* **8**: 163–167.
- DARVASI, A., and M. SOLLER, 1995 Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* **141**: 1199–1207.
- DE LA CHAPPELLE, A., and F. A. WRIGHT, 1998 Linkage disequilibrium mapping in isolated populations: the example of Finland revisited. *Proc. Natl. Acad. Sci. USA* **24**: 12416–12423.
- DIB, C., S. FAURE, C. FIZAMES, D. SAMSON, N. DROUOT *et al.*, 1996 A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**: iii–v.
- FRARY, A., T. C. NESTBITT, A. FRARY, S. GRANDILLO, E. VAN DER KNAPP *et al.*, 2000 *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science* **28**: 85–88.
- GUO, S. W., and K. LANGE, 2000 Genetic mapping of complex traits: promises, problems, and prospects. *Theor. Popul. Biol.* **57**: 1–11.
- HALEY, C., and S. A. KNOTT, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**: 315–324.
- HILL, W. G., 1969 On the theory of artificial selection in finite populations. *Genet. Res.* **13**: 143–163.
- HILL, W. G., 1997 The use of selection experiments for detecting quantitative trait loci. *Genet. Res.* **69**: 227–232.
- HILL, W. G., 1998 Selection with recurrent backcrossing to develop congenic lines for quantitative trait loci analysis. *Genetics* **149**: 1341–1352.
- HILL, W. G., and B. S. WEIR, 1994 Maximum likelihood estimation of gene location by linkage disequilibrium. *Am. J. Hum. Genet.* **54**: 379–390.
- KEARSEY, M. J., and A. G. L. FARQUHAR, 1998 QTL analysis in plants; where are we now? *Heredity* **80**: 137–142.
- LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.
- LANDER, E. S., and N. J. SCHORK, 1994 Genetic dissection of complex traits. *Science* **265**: 2037–2047.
- LUO, Z. W., and M. J. KEARSEY, 1992 Interval mapping of quantitative trait loci in an F₂ population. *Heredity* **69**: 236–242.
- LUO, Z. W., and C.-I. WU, 2001 Modeling linkage disequilibrium between a polymorphic marker locus and a locus affecting complex dichotomous traits in natural populations. *Genetics* **158**: 1785–1800.
- LUO, Z. W., R. THOMPSON and J. A. WOOLLIAMS, 1997 A population genetics model of marker-assisted selection. *Genetics* **146**: 1173–1183.
- LUO, Z. W., S. H. TAO and Z.-B. ZENG, 2000 Inferring linkage disequilibrium between a marker locus and a trait locus in natural populations. *Genetics* **156**: 457–467.
- MACKAY, T. F. C., 2001 The genetic architecture of quantitative traits. *Annu. Rev. Genet.* **35**: 303–339.
- MANLY, K. F., and J. M. OLSON, 1999 Overview of QTL mapping software and introduction to map manager QTL. *Mamm. Genome* **10**: 327–334.
- MARTH, G., R. YEH, M. MINTON, R. DONALDSON, Q. LI *et al.*, 2001 Single-nucleotide polymorphisms in the public domain: how useful are they? *Nat. Genet.* **27**: 371–372.
- MOTT, R., C. J. TALBOT, M. G. TURRI, A. C. COLLINS and J. FLINT, 2000 A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc. Natl. Acad. Sci. USA* **97**: 12649–12654.
- NADEAU, J. H., and W. N. FRANKEL, 2000 The roads from phenotypic variation to gene discovery: mutagenesis versus QTLs. *Nat. Genet.* **24**: 221–225.
- NADEAU, J. H., J. B. SINGER, A. MARTIN and E. S. LANDER, 2000 Analysing complex genetic traits with chromosome substitution strains. *Nat. Genet.* **24**: 381–384.
- RABINOWITZ, A., 1997 A transmission disequilibrium test for quantitative trait loci. *Hum. Hered.* **47**: 342–350.
- WANG, D. G., J. B. FAN, C. J. SIAO, A. BERNO, P. YOUNG *et al.*, 1998 Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**: 1077–1082.
- WRIGHT, S., 1952 *Quantitative Inheritance*. Her Majesty's Stationery Office, London.
- XIONG, M., and S. W. GUO, 1997 Fine-scale mapping of quantitative trait loci using historical recombinations. *Genetics* **145**: 1201–1218.
- ZENG, Z.-B., 1994 Precision mapping of quantitative trait loci. *Genetics* **136**: 1457–1468.
- ZOLLNER, S., and A. VON HAESELER, 2000 A coalescent approach to study linkage disequilibrium between single-nucleotide polymorphisms. *Am. J. Hum. Genet.* **66**: 615–628.

