The Annals of Statistics
2000, Vol. 28, No. 5, 1245–1278

MULTIPLE REGRESSION APPROACH TO MAPPING OF QUANTITATIVE TRAIT LOCI (QTL) BASED ON SIB-PAIR DATA: A THEORETICAL ANALYSIS

By Momiao Xiong[1] and Sunwei Guo

University of Texas and Medical College of Wisconsin

The interval mapping method has been shown to be a powerful tool for mapping QTL. However, it is still a challenge to perform a simultaneous analysis of several linked QTLs, and to isolate multiple linked QTLs. To circumvent these problems, multiple regression analysis has been suggested for experimental species. In this paper, the multiple regression approach is extended to human sib-pair data through multiple regression of the squared difference in trait values between two sibs on the proportions of alleles shared identical by descent by sib pairs at marker loci. We conduct an asymptotic analysis of the partial regression coefficients, which provide a basis for the estimation of the additive genetic variance and of locations of the QTLs. We demonstrate how the magnitude of the regression coefficients can be used to separate multiple linked QTLs. Further, we shall show that the multiple regression model using sib pairs is identifiable, and our proposed procedure for locating QTLs is robust in the sense that it can detect the number of QTLs and their locations in the presence of several linked (QTLs) in an interval, unlike a simple regression model which may find a "ghost" QTL with no effect on the trait in the interval with several linked QTLs. Moreover, we give procedures for computing the threshold values for prespecified significance levels and for computing the power for detecting (QTLs). Finally, we investigate the consistency of the estimator for QTL locations. Using the concept of epi-convergence and variation analysis theory, we shall prove the consistency of the estimator of map location in the framework of the multiple regression approach. Since the true IBD status is not always known, the multiple regression of the squared sib difference on the estimated IBD sharing is also considered.

1. Introduction. Mapping genes that influence quantitative traits such as blood pressure and weight is an important endeavor in genetics, and has received tremendous attention since the birth of genetics as a science. It is the first step towards the eventual identification of these genes, which would eventually lead to the understanding as to how genetic variability affects the variation of quantitative traits.

Statistical inference for gene mapping consists of locating quantitative trait loci (QTLs) relative to a set of DNA markers and of estimating their effects on trait values of interest. Our ability to map QTLs has been greatly enhanced by the rapid development in construction and refinement of genetic mapping combined with the development of relevant statistical methodology

Received April 1999; revised June 2000.

[1] Supported by Grants R29-GM52205, R01-GM56515 and MH59518 from the National Institute of Health.

AMS 1991 *subject classifications*. Primary 62F03; secondary 92D30, 60J20, 15A18.

Key words and phrases. Mapping QTL, linkage analysis, Sib pair analysis, multiple regression, variation analysis and epi-convergence.

[Haley, Knott and Elsen (1994), Wang et al. (1998)]. The interval mapping method [Lander and Botstein (1989)] has been shown theoretically to be a powerful tool for mapping QTL. The method uses markers that flank the chromosomal interval of interest to detect any QTL lying within. Compared with methods based on a single marker, the interval mapping approach has been shown to have greater statistical power and can provide much more accurate estimation of effects and positions of QTLs. It also has been shown that it is relatively robust [Knott and Haley (1992), Luo and Kearsey (1992)].

Despite these advantages, however, it is still difficult for the approach to search simultaneously several QTLs, linked or unlinked, and to distinguish multiple linked QTLs. In particular, when two or more QTLs are located on the same chromosomal region, the interval mapping approach may map these loci to wrong locations [Knott and Haley (1992), Martinez and Curnow (1992)].

The interval mapping combined with linear regression analysis in QTL mapping which is called composite interval mapping, mostly for experimental species, that has been proposed by several workers seems to provide some improvements [Haley and Knott (1992), Haley, Knott and Elsen (1994), Rodolphe and Lefort (1993), Jansen (1993), Zeng (1993, 1994)]. These workers demonstrate that, using multiple markers, the regression approach can detect the effects of multiple QTLs and separate multiple linked QTLs using both markers flanking the QTLs and markers in other regions. This approach is sensible, because quantitative traits are unlikely to be controlled by a single QTL, and because the use of multiple markers in different regions of chromosomes would help detect multiple QTLs. While one can still use the interval mapping method to search multiple QTLs simultaneously, the heavy computation burden makes this approach impractical. The advantage of composite interval mapping is that when testing for the putative QTL in an interval, one uses other markers as covariates to control for other QTL, and hence to separate multiple linked QTL effects and to reduce the residual variance [Kao, Zeng and Teasdale (1999)]. A great improvement in mapping QTLs by composite interval mapping has been reported in mice [Dragani et al. (1995)], *Drosophila* [Liu et al. (1996)] and *Arabidopsis thailiana* [Kuittinen, Sillanpää and Savolainen (1997)]. Recently, Bayesian inference for QTL mapping has been reported by Hoeschele and Varanden (1993), Satagopan, Yandell, Newton and Osborn (1996) and Sillanpää and Arjas (1998).

Understanding the genetic architecture of a quantitative trait is a major research focus in quantitative genetics [Templeton (1999)]. The genetic architecture of a trait refers in part to the number, genomic locations, frequencies and effects of QTL, as well as to the interactions of QTL alleles within (dominance) and between (epistasis) loci, pleiotropic effects of QTL, QTL by environment interactions and so forth. Multiple interval mapping was recently proposed for the identification and estimation of the genetic architecture parameters as well as simultaneous mapping multiple QTL [Kao, Zeng and Teasdale (1999)].

The interval mapping approach for experimental species apparently motivated Fulker and Cardon (1994) to propose an interval mapping method for human sib-pair data that uses information from a pair of flanking markers. Their simulations show that for nearly all modes of gene action, allele frequency and marker density, this approach provides greater power than traditional sib-pair analysis based on a single marker.

In view of the limitations of the interval mapping approach and advantages of the regression approach, it is logical to consider extending the regression approach that is suitable for experimental species to human data, for example, sib-pair data. Several multipoint variance components methods for mapping QTL have been developed recently that allow for marker-specific effects, residual additive genetic effects and random environmental effects [Goldga (1990), Schork (1993), Amos (1994), Xu and Atchley (1995), Blangero and Almasy (1997), Almasy and Blangero (1998)]. A number of applications of the variance components approach to QTL analysis in human pedigree data has appeared recently [Comuzzie et al. (1997), Wang et al. (1997), Begleiter et al. (1998), Duggirala et al. (1999)]. The goal of this paper is to further extend the multiple regression approach to human sib-pair data and provide a thorough theoretical analysis of the model.

Based on the proposed multiple regression model, we first consider the case in which the number of alleles shared identical by descent (IBD) by two siblings can be determined unequivocally. Under this situation, we consider the asymptotic properties of the partial regression coefficients, which provide a basis for the estimation of the additive genetic variance and of locations of the QTLs. We shall show how the magnitude of the regression coefficients can be used to separate multiple linked QTLs. Further, we shall show that the multiple regression model using sib pairs is identifiable, and our proposed procedure for locating QTLs is robust. Moreover, we give procedures for computing the threshold values for prespecified significance levels and for computing the power for detecting QTLs.

Once the theoretical framework is established for the case when IBD status can be determined unequivocally, we then turn to the more realistic case when the IBD status is estimated based on marker data. For simplicity, we only consider the case of diallelic markers. This may be suitable, for example, for markers such as single nucleotide polymorphism (SNP), which is diallelic.

Finally, we investigate the consistency of the estimator for QTL locations, an issue apparently ignored in the literature [Wright (1994)]. In fact, the traditional asymptotic theory is not sufficient to prove this consistency. Using the concept of epi-convergence and variation analysis theory, we shall prove the consistency of the estimator in the framework of the multiple regression approach.

2. A multiple linear regression model.

Let Y_i and $Y_{i'}$ be the trait values for a pair of siblings in the sibship i, respectively. We consider the following model:

$$Y_i = \mu + \sum_{l=1}^{k} g_{il} + e_i, \qquad Y_{i'} = \mu + \sum_{l=1}^{k} g_{i'l} + e_{i'}, \tag{1}$$

where μ is the grand mean, k is the number of QTLs that collectively influence the trait value, g_{il} and $g_{i'l}$ are genetic effects due to the lth QTL and $e_i, e_{i'}$ are residual environmental effects independent of g_{il} and $g_{i'l}$. It is assumed that e_i and e_i' are independent, and that $E(e_i) = E(e_i') = 0$ and $V(e_i) = V(e_i') = \sigma_e^2$.

Let $Z_i = (Y_i - Y_{i'})^2$ be the squared difference of trait values between the two sibs in sibship i. Let $\sigma^2_{a(l)}$ be the additive genetic variance due to the lth QTL. We assume that a sample of n sib pairs has been taken at random from the population, and that the number of markers to be considered is m. For ease of discussion, we further assume that there is no epistasis or dominance. Let $\bar{\pi}_{ij}$ be the proportion of alleles, at jth marker locus, shared identical by descent (IBD) by ith sib pair and $\pi_{ij} = \bar{\pi}_{ij} - \frac{1}{2}$, which is referred to as the IBD value throughout this paper. To detect which marker is linked to QTLs, the squared trait difference Z_i can be regressed onto $\pi_{ij}\,(j = 1, \dots, m)$. This is a multilocus generalization of the Haseman–Elston sib-pair method [Haseman and Elston (1972)]. Therefore, we have the following regression model:

$$Z_i = \alpha + \pi_{i1}\beta_1 + \cdots + \pi_{im}\beta_m + \varepsilon_i, \tag{2}$$

where ε_is are independent random variables with $E[\varepsilon_i] = 0$ and $V(\varepsilon_i) = \sigma^2$. Note that σ^2 should not be confused with the residual environmental variance σ_e^2. In the following, we shall see that

$$\sigma^2 = 4\left[\sum_{l=1}^{k} \sigma^2_{a(l)}\right]\sigma_e^2 + 2\sigma_e^4.$$

If we let

$$R_i = [1, \pi_{i1}, \dots, \pi_{im}], \qquad Z = [Z_1, \dots, Z_n]^T,$$
$$\varepsilon = [\varepsilon_1, \dots, \varepsilon_n]^T, \qquad \beta = [\alpha, \beta_1, \dots, \beta_m]^T,$$

and

$$R = \left[R_1^T, \dots, R_n^T\right]^T,$$

then the above model can be written in a matrix form:

$$Z = R\beta + \varepsilon, \tag{3}$$

with $E[\varepsilon] = 0$ and $V(\varepsilon) = \sigma^2 I$, where I is an $n \times n$ identity matrix.

Using the standard least squares method, we obtain the following estimator for β:

$$\hat{\beta}_n = (R^T R)^{-1} R^T Z.$$

3. Asymptotic properties. In this section, we investigate the asymptotic properties of the estimator $\hat{\beta}$, which are the basis for further theoretical analysis of the model. We assume that data on n sib pairs are obtained independently. We also assume Haldane's mapping function (i.e., no crossing-over interference).

3.1. *Consistency.* Under these assumptions, using the same argument as that of Rodolphe and Lefort (1993), we have the following asymptotic results (Appendix A).

THEOREM 1. *Under the assumed model (3) and the above assumptions,*

$$\sqrt{n}(\hat{\beta}_n - \beta) \overset{\mathcal{L}}{\to} N(0, \sigma^2 U^{-1}),$$

where $U = E[R_1^T R_1]$ *is given by*

$$U = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & A_1 & 0 & \cdots & 0 \\ 0 & 0 & A_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & A_\nu, \end{pmatrix} \tag{4}$$

with $A_i = [\frac{1}{8}\exp(-4\Delta_{jj'})]$ *for the ith chromosome, and* $\Delta_{jj'}$ *representing the genetic distance between the markers* M_j *and* $M_{j'}$, *where* ν *is the number of chromosomes on which the markers are placed, and the dimension of* U *is* $m+1$.

This theorem establishes the consistency of the estimator $\hat{\beta}_n$.

3.2. *Asymptotic variance of* $\hat{\beta}$. If we let $a_j = \exp(-4\Delta_{j,\,j+1})$, we can show, after some algebra, that

$$U^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & A_1^{-1} & 0 & \cdots & 0 \\ 0 & 0 & A_2^{-1} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & A_\nu^{-1}, \end{pmatrix}$$

where

$$A_i^{-1} = 8 \begin{pmatrix} \dfrac{1}{1-a_1^2} & \dfrac{-a_1}{1-a_1^2} & 0 & \cdots & 0 \\ \dfrac{-a_1}{1-a_1^2} & \dfrac{1-a_1^2 a_2^2}{(1-a_1^2)(1-a_2^2)} & \dfrac{-a_2}{1-a_2^2} & \cdots & 0 \\ 0 & \dfrac{-a_2}{1-a_2^2} & \dfrac{1-a_2^2 a_3^2}{(1-a_2^2)(1-a_3^2)} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \dfrac{1}{1-a_{m_i-1}^2} \end{pmatrix}.$$

From the proof of Theorem 1 (Appendix A), we know that the variance matrix is given by $V(\hat{\beta}_n) = \sigma^2(R^T R)^{-1}$ and that $\left(\frac{1}{n}R^T R\right)^{-1}$ is almost surely convergent to U^{-1}. From the structure of the matrix U^{-1}, we can further see that the asymptotic correlation structure of $\hat{\beta}_n$ is simple: the estimator $\hat{\beta}_j$ is correlated only with the estimators of the effect of flanking markers. More specifically, we have

$$\begin{aligned} V(\hat{\beta}_j) &= \frac{8\sigma^2}{n}\frac{1-\exp(-8\Delta_{j,\,j+1})}{(1-\exp(-8\Delta_{j-1,\,j}))(1-\exp(-8\Delta_{j,\,j+1}))} \\ &\approx \frac{\sigma^2}{n}\frac{\Delta_{j-1,\,j+1}}{\Delta_{j-1,\,j}\Delta_{j,\,j+1}} + o(\Delta_{j-1,\,j}, \Delta_{j,\,j+1}) \end{aligned} \tag{5}$$

and

$$\begin{aligned} \operatorname{Cov}(\hat{\beta}_j, \hat{\beta}_{j+1}) &= \frac{8\sigma^2}{n}\frac{-\exp(-4\Delta_{j,\,j+1})}{1-\exp(-8\Delta_{j,\,j+1})} \\ &\approx -\frac{\sigma^2}{n}\frac{1-4\Delta_{j,\,j+1}}{\Delta_{j,\,j+1}} + o(\Delta_{j,\,j+1}). \end{aligned} \tag{6}$$

It can be seen from the above equation that, as the distance between two adjacent markers becomes smaller, that is, the marker density increases, $V(\hat{\beta}_j)$ increases. This suggests that although we can have a dense map, only those markers that are close to QTL are worth fitting in the model. Therefore, to select an optimal subset of markers is very helpful in mapping QTL [Kao, Zeng and Teasdale (1999)]. Furthermore, in order to maintain the accuracy of the estimation, one needs to increase the sample size.

3.3. *Asymptotic partial regression coefficients.* In this section, we shall examine asymptotic properties of the estimated partial regression coefficients. These asymptotic results allow us to distinguish multiple linked QTLs, to estimate the number of the QTL and to detect their true locations.

To increase the reliability and accuracy of QTL mapping, the effects of possible multiple linked QTLs on the same chromosome should be adequately separated in testing and estimation. In asymptotic terms, it is desirable that the asymptotic partial regression coefficient of the trait associated with the marker depends only on those QTLs which are located on the interval flanked by the two neighboring markers, and are independent of the effects of QTL located outside of the interval.

If we assume the existence of k QTLs, distributed all over the genome, and with linkage equilibrium and no epistasis, we can show that this is the case. Using arguments similar to those of Rodolphe and Lefort (1993) and Zeng (1993, 1994), we can show (Appendix B)

THEOREM 2. *Let t_l be the location of the lth QTL and $\sigma^2_{a(l)}$ be its additive genetic variance. Then*

$$\hat{\beta}_j \overset{\text{a.s.}}{\to} \beta^\star_j = -\frac{a_p}{1-a_p^2}x_{j-1} + \frac{1-a_p^2a_r^2}{(1-a_p^2)(1-a_r^2)}x_j - \frac{a_r}{1-a_r^2}x_{j+1}, \tag{7}$$

where $a_p = \exp(-4\Delta_{j-1,\,j})$, $a_r = \exp(-4\Delta_{j,\,j+1})$ *and* $x_j = -2\sum_{l=1}^{k}\sigma^2_{a(l)} \times \exp(-4\Delta_{j,\,t_l})$.

Theorem 2 gives an explicit asymptotic formula for the estimate of the partial regression coefficient $\hat{\beta}_j$, which allows us to obtain some analytic results and provides the bases for gaining insight into the multiple regression model for mapping QTL. As the following corollary shows, similar to the case of experimental organisms [Zeng (1993, 1994)], $\beta^\star_j$ depends only on those QTLs that are located within the interval that is flanked by the two neighboring markers, and is independent of effects of QTLs located outside the interval (Appendix C).

COROLLARY 1. (i) *If a subset of QTLs are located in the left-hand side of marker* M_{j-1}, *then their contribution to* $\beta^\star_j$ *is zero.*

(ii) *If a subset of QTLs are located in the right-hand side of marker* M_{j+1}, *then their contribution to* $\beta^\star_j$ *is also zero.*

(iii) *If a subset of QTLs are located between markers* M_{j-1} *and* M_j, *then their contribution to* $\beta^\star_j$ *and* $\beta^\star_{j-1}+\beta^\star_j$ *is*

$$\beta^\star_j = -2\sum_l \sigma^2_{a(t_l)} \exp(-4\Delta_{j,t_l}) \frac{1-\exp(-8\Delta_{j-1,t_l})}{1-\exp(-8\Delta_{j-1,j})} \tag{8}$$

and

$$\beta^\star_{j-1}+\beta^\star_j = -2\sum_l \sigma^2_{a(t_l)} \frac{\exp(-4\Delta_{j-1,t_l})+\exp(-4\Delta_{j,t_l})}{1+\exp(-4\Delta_{j-1,j})}, \tag{9}$$

respectively.

From Corollary 1, we can conclude that, roughly speaking, the effect of each QTL is shared by its flanking markers. Suppose that in the interval flanked by markers M_{j-1} and M_j, there is only one QTL, which is located at t_l. If $\Delta_{j-1,j} \to 0$, then $\beta^\star_{j-1}+\beta^\star_j \to -2\sigma^2_{a(l)}$; that is, when the interval harboring the QTL shrinks to the true QTL t_l, $\beta^\star_{j-1}+\beta^\star_j$ converges to $\beta = -2\sigma^2_{a(l)}$. Hence, if there is only one QTL in the interval flanked by markers M_{j-1} and M_j, then $-\frac{1}{2}(\beta^\star_{j-1}+\beta^\star_j)$ can be taken as an estimate of $\sigma^2_{a(l)}$.

An immediate implication of Theorem 2 is that QTLs located in other chromosomes do not contribute to the partial regression coefficient β_j, nor do the effects of any subset of QTLs located outside the interval flanked by markers M_{j-1} and M_{j+1}. This very nice property suggests that a conditional (interval) test or an estimation procedure for locating QTLs can be constructed based on the partial regression coefficient, and such a test or estimate of the true QTL location can be used to detect the linkage of those QTLs which are located within the defined interval of interest.

The idea of the composite interval mapping of Zeng (1993, 1994) and Jansen (1993, 1994) for experimental organisms also can be extended to human data, as explained as follows. Assume that the interval of interest is flanked by markers M_1 and M_2, arranged in that order. We then use two additional markers: one marker M_L is placed further left of marker M_1 and another marker M_R, further right of M_2. Denote the recombination fraction between marker M_i and the putative QTL by $c_i (i = 1, 2)$, and the recombination fraction between the markers M_1 and M_2 by c_{12}. The composite interval mapping approach based on sib pairs is to regress the squared difference of trait values on the IBD values π_L, π_R and π_q at marker M_L, M_R and the putative QTL [Fulker and Cardon (1994), Xu and Atchley (1995)]; that is,

$$Z_i = \alpha + \pi_L\beta_L + \pi_q\beta_q + \pi_R\beta_R + \varepsilon_i,$$

where

$$\pi_q = \beta_1\pi_1 + \beta_2\pi_2,$$
$$\hat{\beta}_1 = \frac{[(1-2c_1)^2 - (1-2c_2)^2(1-2c_{12})^2]}{1-(1-2c_{12})^4},$$
$$\hat{\beta}_2 = \frac{[(1-2c_2)^2 - (1-2c_1)^2(1-2c_{12})^2]}{1-(1-2c_{12})^4}.$$

Since π_q is unknown, it is estimated through IBD values π_1 and π_2 calculated at the two flanking markers M_1 and M_2. The QTLs located outside of the interval flanked by M_L and M_R have no contribution to $\beta_q^\star$. Only QTLs located in the interval contribute to $\beta_q^\star$. Therefore, with an appropriate dense map, we should be able to separate the multiple linked QTLs and to narrow down the chromosomal region to localize QTL. These results are similar to the results of Zeng (1993, 1994) and Jansen (1993, 1994) for experimental species.

3.4. *Identifiability and robustness.* As far as identifiability, robustness and consistency are concerned, we only discuss the dense marker case in order to facilitate the discussion. In this case, it is assumed that at any point in the genome, there is a marker. This is admittedly an ideal case, but we point out that the advance of molecular genetics renders this assumption possible.

In the previous discussion, we presented an algorithm which can be used to detect the region in which the true QTL is located, but we still do not know the exact locations of the true QTL in the region. Now we shall discuss how to identify the exact locations of these QTLs. Assume that a marker is located at t in the chromosome of interest and that its corresponding coefficient in the multiple regression is denoted by $\beta_t^\star$ for an infinitely large sample size. We also assume that a true QTL is located at $t^\star$ in the chromosome. Intuitively, we can expect that the partial regression coefficient β_t will tend to have a large negative peak in the neighborhood of the true location $t^\star$ of the QTL. Indeed, this is true. The following theorem further states that, at $t^\star$, $\beta_t^\star$ will reach the local minimum (Appendix D).

THEOREM 3. *Suppose that there is only one QTL, located at $t^\star$, in the region $[l_t, r_t]$, and the two flanking markers of the QTL are located at l_t and r_t, respectively. Then*

(i) $\beta_{t^\star}^\star < \beta_t^\star$ *when* $t \neq t^\star$ *and* $t \in [l_t, r_t]$ *that is,* $\beta_t^\star$ *reaches its minimum over the region* $[l_t, r_t]$ *at* $t^\star$.

Furthermore,

$$\frac{d\beta_t^\star}{dt} < 0 \qquad \textit{when } t \in (t^\star - \varepsilon, t^\star) \quad \textit{for some } \varepsilon > 0,$$
$$\frac{d\beta_t^\star}{dt} > 0 \qquad \textit{when } t \in (t^\star, t^\star + \varepsilon) \quad \textit{for some } \varepsilon > 0,$$

and $d\beta_t^\star/dt$ *does not exist at* $t = t^\star$.

(ii) *As the distance* $|r_t - l_t|$ *goes to zero, then* $\beta^\star_{l_t}$ *and* $\beta^\star_{r_t}$ *collapse into one variable.*

$$\beta^\star_{t\star} = \sigma^2_{a(t^\star)}.$$

Since the number of QTLs is typically unknown a priori, it may be difficult to fit, at first, a multiple regression model. However, Theorem 3 says that when all true QTLs are separated, we can always narrow regions (intervals) so that in each region (or interval) there exists only one QTL. Further, the minimum of the partial regression coefficient in such regions (or intervals) is uniquely determined and corresponds to the true QTL location. Of course, it should be pointed out that this theorem holds asymptotically. In reality, especially in human genetics research, sample sizes are typically not very large.

To demonstrate the robustness of the model for the localization of QTLs, we present three figures. Figure 1 shows the regression coefficient $\beta^\star_t$ of a simple regression model and the partial regression coefficient of a multiple regression

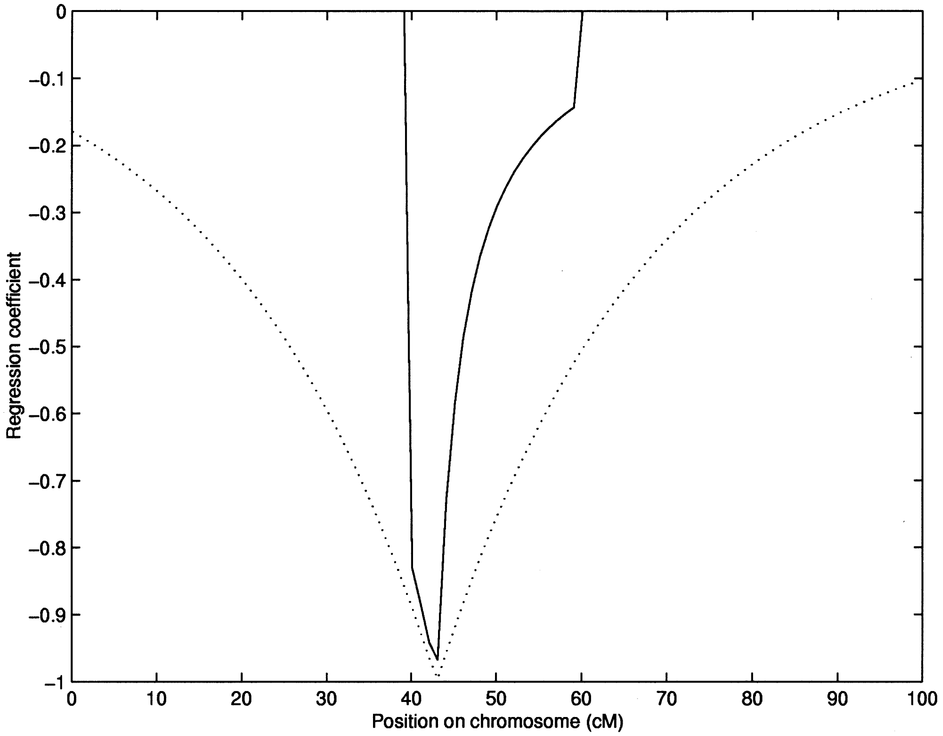

FIG. 1. *Profile of partial regression coefficients of simple and multiple regression models at marker loci. A single QTL with an additive genetic variance of* 0.25 *is assumed. The QTL is located at* 43 *cM from one end of the chromosome. The solid curve represents the multiple regression model, and the dotted, the simple regression model.*

for a single QTL with its true location at 43 cM in a chromosome with 100 cM in length containing six equally spaced markers. The additive genetic variance for the QTL is set to be 0.25. Suppose that the putative QTL is located at $t^\star$, the marker information is available at t and t varies from 0 to 100 cM on the chromosome. It is clear that for the multiple regression model, $\beta_t^\star = 0$ outside the interval $[40cM, 60cM]$ and reaches the minimum -0.25 at the true location within the interval $[40cM, 60cM]$. For the simple regression model, although the regression coefficient $\beta_t^\star$ is not equal to zero outside the interval, it also reaches the minimum at the true location. Thus, for the single QTL, both simple regression and multiple regression models are able to localize the true QTL when the map is dense.

Now consider the case when there are two QTLs, one located at 44 cM and the other, 73 cM from one end of the chromosome. The regression coefficients for the simple and the partial regression coefficient for the multiple regression models are plotted in Figure 2. Although there are actually two QTLs in the same chromosomal region, we do not know the number of QTLs a priori. From

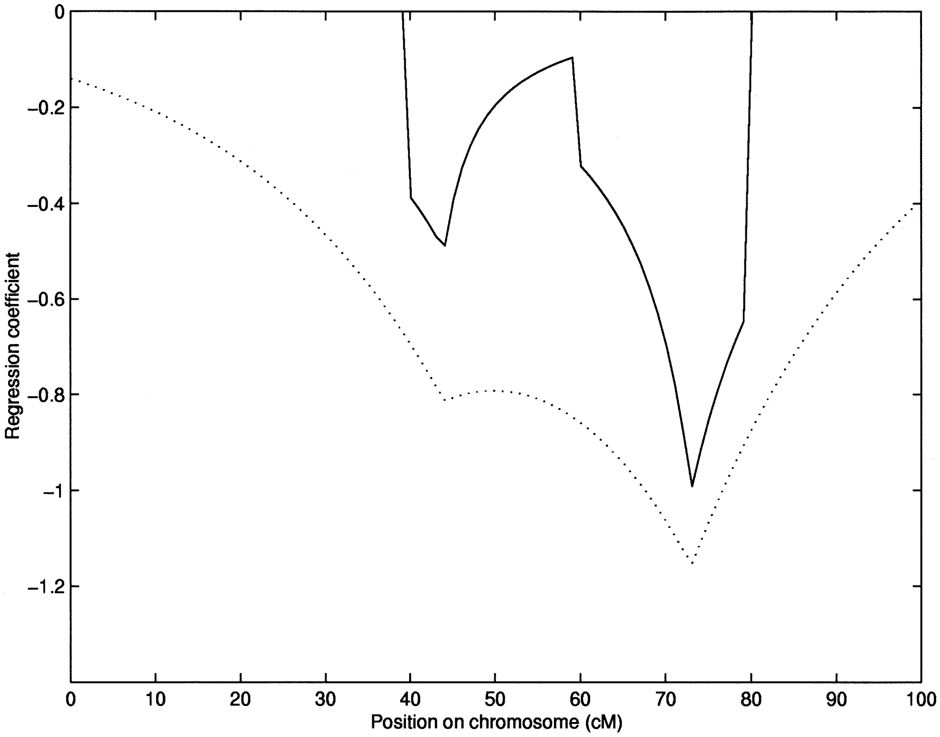

FIG. 2. *Profile of partial regression coefficients of simple and multiple regression at marker loci. Two QTLs, with additive genetic variances of* 0.25 *and* 0.5, *respectively, are located at* 44 *cM and* 73 *cM. The solid curve represents the multiple regression model, and the dotted one, the simple regression model.*

Figure 2 we can see that the regression coefficients of both simple and multiple regression models have two peaks at 44 cM and 73 cM positions. However, the regression coefficient of the simple regression model has only one sharp peak at 73 cM location. It is clear that its peak at the location of 44 cM is less prominent. It shows that if there are two QTLs, the multiple regression model is more robust and has better ability to distinguish multiple linked QTLs than the simple regression model.

Figure 3 shows the regression coefficient of the simple regression and partial regression coefficient of the multiple regression for two QTLs, located at 54 cM and 68 cM, respectively. It is evident that, for the two linked QTLs in the interval $[40cM,\ 80cM]$, they can hardly be separated by a simple regression model. However, the two loci can be distinguished very clearly by the multiple regression model. Again, Figure 3 demonstrates that even in the case where the two QTLs are closely linked, by searching the local minimum of the partial regression coefficient, the multiple regression model is still able to localize QTLs when we have a dense map.

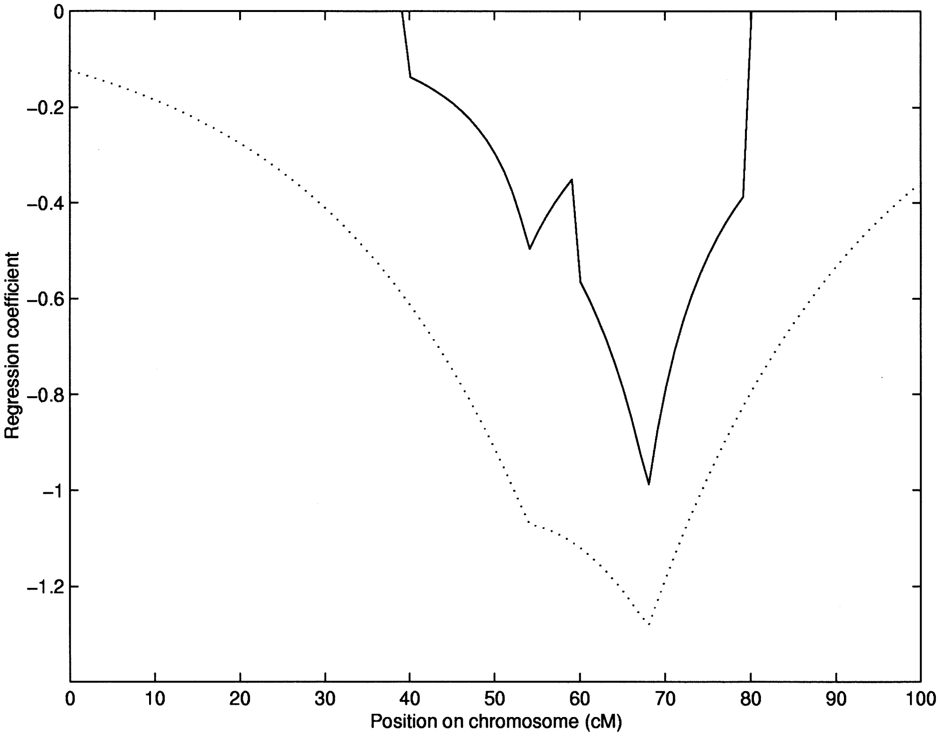

FIG. 3. *Profiles of partial regression coefficients of simple and multiple regression models. Two QTLs, with additive genetic variances of* 0.25 *and* 0.5, *respectively, are located at* 54 *cM and* 68 *cM. The solid curve represents the multiple regression model, and the dotted one, the simple regression model.*

4. Thresholds and power.

4.1. *The thresholds of the test.* To implement the proposed mapping procedure, it is critical to determine the threshold for a given significance level, so that one can reject or accept the null hypothesis H_0: $\sigma_a^2 = 0$ depending on whether or not the statistic exceeds the threshold. In this section, we give procedures for computing the thresholds.

For a particular chromosomal interval flanked by two markers, the estimated partial regression coefficient of the multiple regression depends only on those QTL located within the interval. It is thus natural to test whether or not there exists a QTL in a given marker interval. Suppose that we want to test the interval $[M_{j-1}, M_j]$. Let

$$X_d = -\sqrt{n}\frac{\hat{\beta}(d)}{\sqrt{2}\hat{\sigma}},$$

Where $\hat{\beta}(d)$ is associated with a marker located at locus d in the interval $[M_{j-1}, M_j]$ and σ^2 is the variance of the residuals, and are both estimated by the multiple regression method. The test statistic is then taken as

$$\max_{d\varepsilon[M_{j-1}, M_j]} X_d.$$

We can show that (Appendix E)

$$\sigma^2 = 4\left(\sum_{l=1}^{k} \sigma_{a(l)}^2\right)\sigma_e^2 + 2\sigma_e^4.$$

It can be seen that under the null hypothesis H_0: $\sigma_a^2 = 0$, X_d is asymptotically a Gaussian process with mean 0 and a complicated covariance function, which can be approximated by the function $R(u) = e^{-|u|}$, as $n \to \infty$. Therefore, X_d can still be approximated by an Ornstein–Uhlenbeck process.

Using the results of Feingold, Brown and Siegmund (1993), we have, under the null hypothesis,

$$P_0\left(\max_d X_d > b\right) \approx 1 - \Phi(b) + tlb\phi(b), \tag{10}$$

where l is the length of the tested interval, and $\phi(x)$ and $\Phi(x)$ are the density and cumulative functions of the standard normal distribution, respectively.

When the genetic map is not dense, that is, markers are not available at some locations, it is usually assumed that $x_i(d)$ is known at equispaced distances of Δ centimorgons. For this case, (10) becomes

$$P_0\Big(\max_k X_{k\Delta} > b\Big) \approx 1 - \Phi(b) + tlb\phi(b)\nu(b\sqrt{2t\Delta}), \tag{11}$$

where $\nu(x) \approx e^{-0.583x}$ [Feingold, Brown and Siegmund (1993)]. Note that this equation is equivalent to (10) when $\Delta = 0$. Here, the function $\nu(x)$ is a discreteness correction factor to account for the fact that we are computing the likelihood ratio statistic at discrete points on the chromosome instead of continuously as is the case for a dense map.

Another test statistic $T = \hat{\beta}/\sqrt{V(\hat{\beta})}$ also can be used to detect the existence of QTL [Fulker and Cardon (1994)]. We shall show an asymptotic relationship between X_d and T. Suppose that the markers are equally spaced along the chromosome with the length l of the interval. It has been shown above that

$$V(\hat{\beta}) \approx \frac{\sigma^2}{n}\frac{2}{l}.$$

Therefore,

$$\begin{aligned} T &\approx \frac{\sqrt{n}\hat{\beta}}{\sqrt{2}\sigma}\sqrt{l} \\ &\approx -\sqrt{l}X_d. \end{aligned}$$

When $l \le 1$ Morgan, $|T|$ and X_d may be very close.

This result makes sense intuitively. When we search QTL in a longer chromosomal region, there is a higher probability to make errors. Therefore, to maintain a prespecified significance level of the test, we need to increase the critical value of the test.

To illustrate this point graphically, we calculated the threshold as a function of l for the statistic X_d at 0.05, 0.01 and 0.001 level. The results are shown in Figure 4. Thus, we can see that for a fixed significance level α, increasing l will decrease $1 - \Phi(b)$ and hence increase the threshold b.

4.2. *The power of the test.* Assume that in the interval $[M_{j-1}, M_j]$ there exists only one QTL, located at l, with additive genetic variance $\sigma^2_{a(l)}$. Under the null hypothesis H_0: $\sigma^2_{a(l)} = 0$, we have

$$\begin{aligned} E[\hat{\beta}(d)] &\approx -2\sigma^2_{a(l)} \exp(-4\Delta_{dl}) \frac{1 - \exp(-8\Delta_{l,\,j-1})}{1 - \exp(-8\Delta_{j-1,\,d})} \\ &\approx -2\sigma^2_{a(l)} \frac{\Delta_{l,\,j-1}}{\Delta_{j-1,\,d}} \exp(-4\Delta_{d,l}). \end{aligned}$$

The coefficient of $\exp(-4\Delta_{dl})$ depends in general on the genetic distance between the marker and the QTL. However, if $\Delta_{d,\,l}$ is small, $E[\hat{\beta}(d)]$ can be approximated by

$$E[\hat{\beta}(d)] \approx -2\sigma^2_{a(l)} \exp(-4\Delta_{d,\,l}).$$

In this case,

$$\begin{aligned} E[X_d] &\approx \sqrt{\frac{n}{2}}\frac{2\sigma^2_{a(l)}}{\sigma} \exp(-4\Delta_{d,\,l}) \\ &\approx \xi e^{-4|u|}, \end{aligned}$$

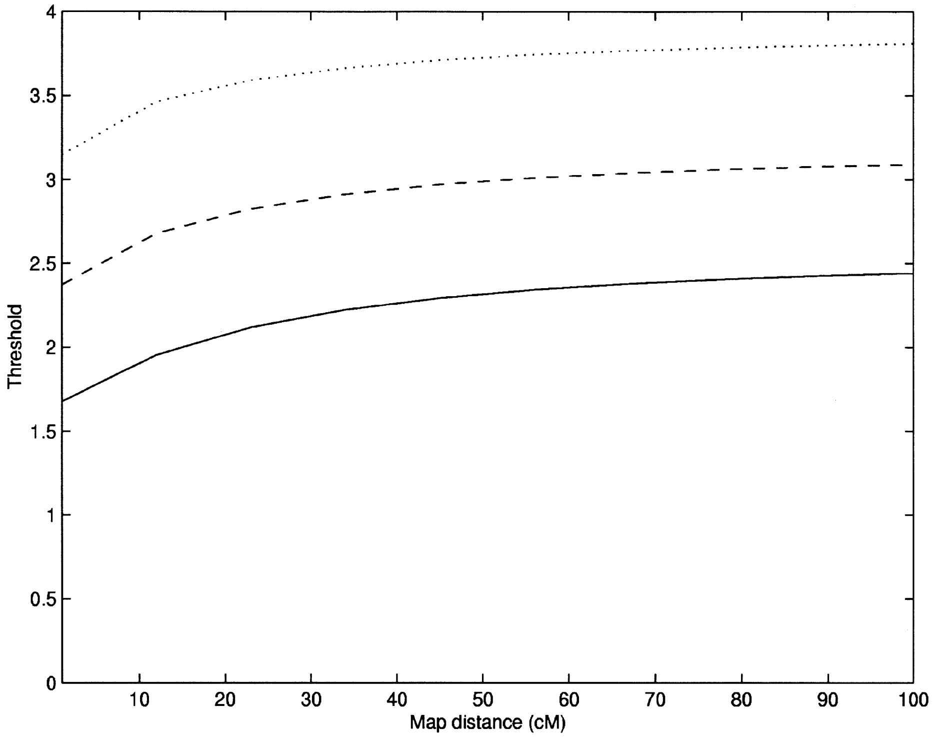

FIG. 4. *Thresholds of test statistic X_d as a function of length of chromosomal region to be searched. The solid curve is for a significance level of* 5%, *the dashed one, a significance level of* 1%, *and the dotted one, a significance level of* 0.1%.

where $\xi = \sqrt{2n}(\sigma^2_{a(l)}/\sigma)$, and $|u|$ is the distance between the marker and the QTL. Using the results of Feingold, Brown and Siegmund (1993), we can obtain the following approximations:

(i) for a dense map,

$$P_{d,\xi}\Big(\max_d X_d > b\Big) \approx 1 - \Phi(b-\xi) + \phi(b-\xi)[2\xi^{-1} - (b+\xi)^{-1}], \tag{12}$$

(ii) for an equispaced map,

$$\begin{aligned} P_{d,\xi}\Big(\max_k X_{k\Delta} > b\Big) \approx 1 &- \Phi(b-\xi) \\ &+ \phi(b-\xi)[2\xi^{-1}\nu - (b+\xi)^{-1}\nu^2], \end{aligned} \tag{13}$$

where $\nu = \nu(b\sqrt{2t\Delta})$.

Figure 5 demonstrates the power of the test statistic X_d as a function of genetic distance between the marker and trait loci for a significance level of 0.05 with n = 100, 200, 500 and 1000. As expected, as sample size or heritability increases, so does the power of the test.

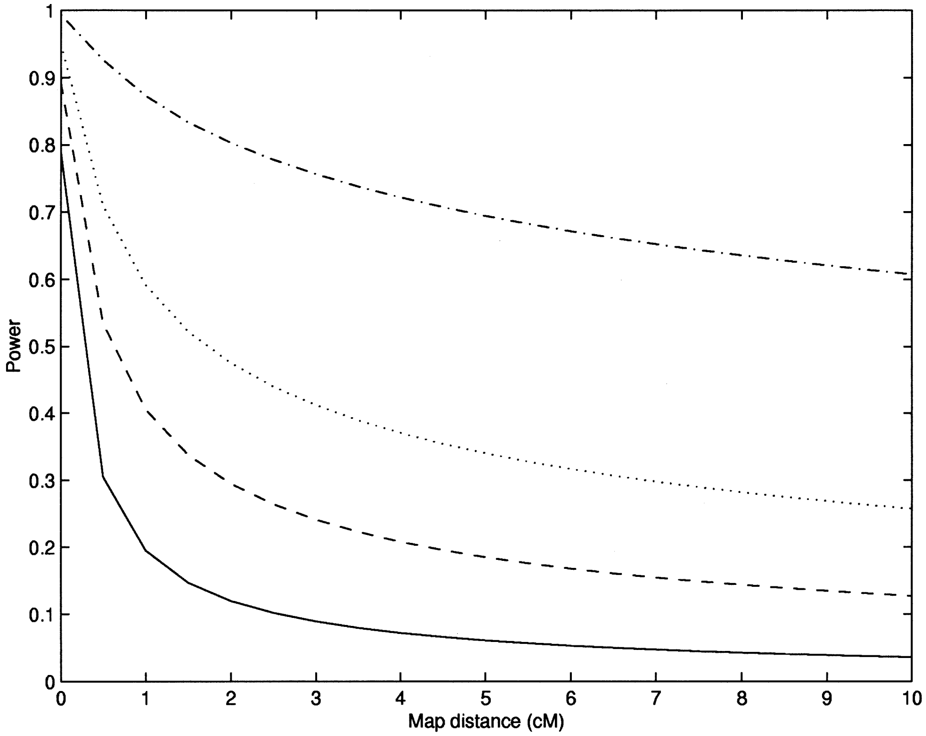

FIG. 5. *Statistical power as a function of genetic distance between the marker and trait loci. The power is calculated assuming a heritability of* 0.15, *with a significance level of* 5% *and an additive genetic variance of* 0.5. *The four curves, from upper left-hand side to the lower right-hand side, represent powers for sample sizes* $n = 100, 200, 500$ *and* 1000, *respectively, where* n *is the number of sib pairs.*

5. Regression on estimated IBD proportions. For human data, the proportions $\bar{\pi}$ of alleles shared IBD cannot always be scored with certainty and thus need to be estimated. We consider only estimation of $\bar{\pi}$ for each marker individually. The estimation of $\bar{\pi}$ by the joint marker information will be discussed elsewhere.

Let f_i be the probability that the sib pair has i alleles IBD at the marker locus. Then, for any given marker locus, the estimated $\bar{\pi}$ is given by

$$\hat{\bar{\pi}} = f_2 + \tfrac{1}{2} f_1.$$

Let $\hat{\bar{\pi}}_{i,\,j}$ be the estimation of $\bar{\pi}_{ij}$ and $\pi^\star_{ij} = \hat{\bar{\pi}}_{ij} - \frac{1}{2}$. The squared difference of trait values can be regressed onto $\pi^\star_{ij}$ as follows:

$$(14) \qquad Z_i = \alpha^\star + \pi^\star_{i1}\beta^\star_1 + \cdots + \pi^\star_{im}\beta^\star_m + \varepsilon_i, \qquad i = 1, \dots, n,$$

where ε_i's are independent random variables with $E[\varepsilon_i] = 0$ and $V(\varepsilon_i) = \sigma^2$, the same as defined in (2). Let

$$R^\star_i = [1, \pi^\star_{i1}, \dots, \pi^\star_{im}], \qquad R^\star = [R^{\star T}_1, \dots, R^{\star T}_n]^T,$$

and

$$\beta^\star = [\alpha^\star, \beta_1^\star, \dots, \beta_m^\star]^T.$$

The least squares estimate of $\beta^\star$ is given by

$$\hat{\beta}^\star = (R^{\star T} R^\star)^{-1} R^{\star T} Z.$$

Under the same assumption as that in Theorem 1, we can show (Appendix F)

THEOREM 4. *Under model* (14) *and the above assumptions,*

$$\sqrt{n}(\hat{\beta}_n^\star - \beta^\star) \overset{\mathcal{L}}{\to} N(0, \sigma^2 U_\star^{-1})$$

and

$$\hat{\beta}_n^\star \overset{\text{a.s.}}{\to} U_\star^{-1} W_\star,$$

where

$$W_\star = \left[x_0, x_1^\star, \dots, x_m^\star\right]^T, \qquad x_j^\star = -\frac{5}{4}\sum_{l=1}^{k} \sigma_{a(l)}^2 \exp(-4\Delta_{jl}),$$

and $U_\star = E[R_1^{\star T} R_1^\star]$ *is given by*

$$U_\star = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & B_1 & 0 & \cdots & 0 \\ 0 & 0 & B_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & B_\nu, \end{pmatrix} \tag{15}$$

with $B_i = [b_{j,l}]$ *for the ith chromosome,* $b_{j,j} = \frac{1}{4} p_j q_j (1 - p_j q_j), b_{j,l} = \frac{1}{2} p_j q_j p_l q_l (1 - 2 p_l q_l + p_j q_j p_l q_l) \exp(-4\Delta_{j,l})$, p_j *and* q_j, p_l *and* q_l *being frequencies of the marker alleles* 1 *and* 2 *at the marker locus* M_j *and* M_l, *respectively, and* $\Delta_{j,l}$ *representing the genetic distance between the markers* M_j *and* M_l.

It is clear from above that the matrix $U_\star$ does not have the same nice structures as matrix U does. This feature, unfortunately, does not lead to a simple form of the inverse matrix of B_i and hence makes it difficult to obtain an explicit expression for the asymptotic estimation of the partial regression coefficient $\beta_j^\star$. Therefore, Theorem 2 and Corollary 1 no longer hold in this case.

As we can see from the above discussion, Theorem 2 and Corollary 1 form the foundation to distinguish multiple linked QTLs using the multiple regression approach. The question naturally arises as whether or not it is still possible for the multiple regression model to separate the linked QTLs using $\pi^\star$ instead of π. To investigate this question, we consider a numerical example. Assume that all alleles at all loci have an equal frequency of 0.5, with all the other parameters the same as that of Figure 2, that is, two QTLs at 44 cM and 73 cM from one end of the chromosome, respectively. We can see

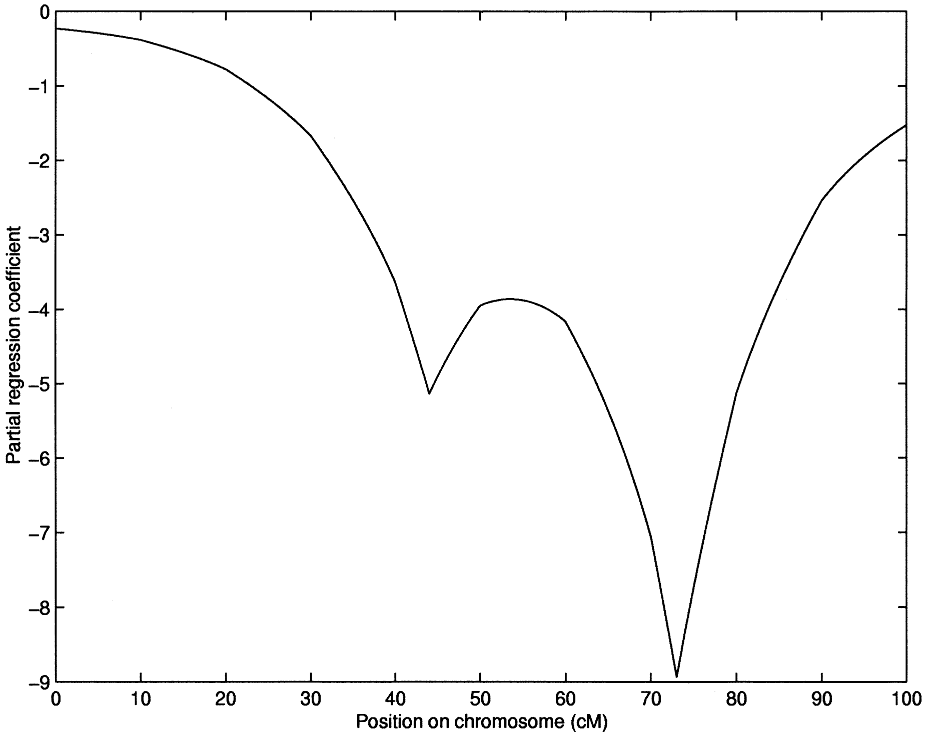

FIG. 6. *Profile of partial regression coefficients of multiple regression based on estimated IBD values $\pi^\star$. Two QTLs, with the additive genetic variances* 0.25 *and* 0.5, *respectively, are located at* 44 *cM and* 73 *cM from one end of the chromosome.*

from Figure 6 that the asymptotic partial regression coefficient of the multiple regression is no longer zero outside the interval $[40cM, 80cM]$. This implies that multiple regression of the squared difference onto the estimation $\pi^\star$ of π has unfortunately less power to separate the multiple linked QTLs than based on π. We can also see that, fortunately, the partial regression coefficient curve of the multiple regression model, based on $\pi^\star$, does not depart significantly from the curve of the multiple regression model based on π. This shows that in many cases, replacing π by $\pi^\star$, the multiple regression model may still be able to distinguish multiple linked QTLs.

6. Consistency of map locations. Throughout this section, we have assumed a dense map. Given a random sample of n sib pairs from the population, we can obtain estimates of the partial regression coefficient $\hat{\beta}_t$ for marker locus t. It is clear that $\hat{\beta}_t$ depends on sample size n. To explicitly express this dependency, we denote $\hat{\beta}_t$ by $\hat{\beta}_t(n)$.

As we discussed earlier, to detect the true QTL, we seek the local minimum of $\hat{\beta}_t(n)$. Suppose that in interval Ω_0, the local minimum of $\hat{\beta}_t(n)$ is obtained. Let $t_n = \arg\min_{t\in\Omega_0} \hat{\beta}_t(n)$. Ideally, we hope that when the sample size n goes to infinity, the estimated location of the putative QTL, t_n, will converge to the true QTL location $t^\star$ in the interval Ω_0, as is taken implicitly in the literature. In Section 3, we have proved that $\hat{\beta}_t(n) \rightarrow^{\text{a.s.}} \beta_t$ and in the dense marker case, $t^\star = \arg\min_{t \in \Omega_0} \beta_t$ is exactly the location of the true QTL. Now the question is whether $t_n \to t^\star$ or would converge at all. It should be pointed out that, in general, $\lim_{n\to\infty} \min_{t\in\Omega_0} \hat{\beta}_t(n) \neq \min_{t\in\Omega_0} \lim_{n\to\infty} \hat{\beta}_t(n)$, that is, almost sure convergence of $\hat{\beta}_t(n)$ does not guarantee automatically $t_n \to t^\star$. To circumvent this problem, we resort to the concept of epi-convergence [see Aubin and Frankowska (1990)].

We begin with a brief introduction of the concept of epi-convergence and some related basic results. Interested readers should consult Dupačová and Wets (1988) for more details.

A sequence of function $\{g^\nu: R^n \to \overline{R}, \nu = 1, \ldots\}$ is said to be epi-convergent to g: $R \to \overline{R}$ if, for all t in R, we have

$$\varliminf_{\nu\to\infty} g^\nu(t^\nu) \geq g(t) \quad \text{for all } \{t^\nu\}_{\nu=1}^\infty \quad \text{converging to } t, \tag{16}$$

and for some $\{t^\nu\}_{\nu=1}^\infty$ converging to t,

$$\varlimsup_{\nu\to\infty} g^\nu(t^\nu) \leq g(t), \tag{17}$$

we then say that g is the epi-limit of the g^ν and write $g = \text{epi-}\lim_{\nu\to\infty} g^\nu$. The following theorem establishes that if g^ν is epi-convergent to g, then this ensures the sequence of the minimizer of g^ν will converge to the minimizer of g.

THEOREM 5 [Wets (1991)]. *Suppose $\{g, g^\nu: R^n \to \overline{R}, \nu = 1, \ldots\}$ is a collection of functions such that $g = \text{epi-}\lim_{\nu\to\infty} g^\nu$. Then if $t^k \in \arg\min g^{\nu_k}$ for some subsequence $\{\nu_k, k = 1, \ldots\}$ and $t = \lim_{k\to\infty} t^k$, it follows that*

$$t \in \arg\min g$$

and

$$\lim_{k\to\infty} (\inf g^{\nu_k}) = \inf g.$$

Hence, in particular, if there exists a bounded set $D \subset R^n$ such that, for some subsequence $\{\nu_k, k = 1, \ldots\}$,

$$\arg\min g^{\nu_k} \bigcap D \neq \varnothing,$$

then the minimum of g is attained at some point in the closure of D.

To invoke Theorem 5, we first need to show that $\hat{\beta}_t(n)$ is epi-convergent.

LEMMA 1. *A sequence of the estimator $\hat{\beta}_t(n)$ of the partial regression coefficient is epi-convergent to $\beta_t^\star$ with probability* 1.

The proof is given in Appendix G.

We are now in a position to prove the consistency of the estimator t_n of the true location of the QTL.

THEOREM 6. *Assume that the region Ω_0 of the chromosome contains a unique QTL, located at $t^\star$, and the marker map is dense. Then the estimator $t_n = \arg\min_{t\in\Omega_0} \hat{\beta}_t(n)$ of the location $t^\star$ of the true QTL is consistent, that is, $t_n \to^{\text{a.s.}} t^\star$.*

PROOF. Since $t_n = \arg\min_{t\in\Omega_0} \hat{\beta}_t(n)$ is bounded in the set Ω_0, then by Lemma 1 and Theorem 5, for every sequence t_n, there exists a subsequence t_{n_k} such that it converges to, say, t_0 with probability 1 and

$$t_0 \in \arg\min_{t\in\Omega_0} \beta_t.$$

Because of the assumption of the unique QTL in Ω_0, we conclude that

$$t_0 = t^\star.$$

Therefore

$$\lim_{n\to\infty} t_n = t^\star \qquad \text{a.s.}$$

The proof is complete. □

Note that the above consistency of the estimator t_n of the true location of the QTL is proved under the assumption of a dense marker map. For a nondense marker map, the IBD values π_j in the chromosome which are not at marker loci are estimated from the IBD values at flanking markers. The limit of the sequence of the minimum of the partial regression coefficient of multiple regression of the squared sib trait difference on such estimated IBD values may no longer correspond to the true QTL location. Therefore, in general, in the nondense map case, the above consistency theorem may not hold. The detailed analysis is beyond the scope of this paper.

7. Discussion.

A great challenge to all existing QTL mapping methods is how to separate multiple linked QTLs. Although a simultaneous search using interval mapping approach can in principle be helpful, heavy computational burdens would practically preclude this approach. In addition, computing the threshold for declaring linkage also can be difficult. Under assumptions of no epistasis or dominance, the multiple regression approach has been shown to be statistically more powerful and more precise in separating multiple linked QTLs in experimental organisms [Haley and Knott (1992), Haley, Knott and Elsen (1994), Rodolph and Lefort (1993), Jansen (1993), Zeng (1993, 1994)].

In this paper, we have extended the multiple regression approach to QTL mapping using human sib-pair data, which regresses squared difference in trait values of two siblings onto the proportions of genes shared IBD scored at multiple marker loci. For the first time, we investigated the asymptotic properties of this approach under the ideal situation in which the marker density is high and IBD status can be scored unequivocally. Parallel to the case of experimental organisms, we have shown that, for sib-pair data, the expected partial regression coefficient of the regression model at any particular marker depends on the effects of QTLs which are located in the nearby interval flanked by two neighboring markers, and is unaffected by the effects of other QTLs located outside the interval. This feature alone enables us to improve both the precision as well as the accuracy of QTL mapping. Obviously, our model and results can be extended for relative pairs other than siblings.

For the dense marker case, we have further proved the consistency of the estimators of the partial regression coefficients when IBD status can be determined unequivocally. We also have shown that the multiple regression method is identifiable and is fairly robust, and that QTLs can be mapped through detecting the minimum of the partial regression coefficients. In addition, we have provided methods for computing the thresholds for linkage declaration and for power calculation.

One important contribution of this paper is the proof of consistency of the estimator t_n of the true map location to the QTL using the multiple regression method. Since the estimator t_n is obtained by minimizing $\hat{\beta}_t(n)$, the proof of convergence of t_n to the true QTL location $t^\star$ of the QTL requires the exchange of limiting process with minimization process. However, the traditional pointwise convergence does not guarantee this exchange to be legal. To justify this exchange, we have used the concept of epi-convergence and the theory of variation analysis, and have proved the epi-convergence of the sequence of the estimators t_n. By doing so, we have proved the consistency of the estimator t_n under the dense map assumption.

It should be noted that the major focus of our theoretical investigation of the proposed multiple regression model is based on the assumption that IBD information can be scored unequivocally for all sib pairs. This is mainly for the sake of mathematical tractability. In practice, of course, IBD sharing is often estimated due to missing parental data, or uninformative matings. If this is the case, we have demonstrated that the use of estimated IBD information would result in a decrease in statistical power in detecting multiple linked QTLs. This is to be expected, since the estimated IBD information obscures the linkage information contained in the IBD configurations.

Given rapid advances in molecular genetics, especially in map refinement and genotyping technology [see, e.g., Wang, Fan and Siao (1998)], the assumption of unequivocal IBD information may not be too far off, since the map density in humans can be made practically very high, and the IBD information at a particular locus can be extracted in most cases through the use of multiple polymorphic, closely linked markers. In view of this, the major conclusions reached in this paper should hold in general for currently available data.

Another assumption, somewhat related with the assumption of unequivocal IBD information, is the dense map assumption. This, again, is mainly for mathematical tractability, as in the case of Feingold, Brown and Siegmund (1993). With today's molecular technology, the genetic map is not dense in the sense assumed in this paper. Therefore, the asymptotic behaviors of all estimators in the nondense marker case, including the identifiability, robustness and consistency of the multiple regression model, and the impact of the density of the markers on QTL mapping should be further investigated. However, the current molecular technology (e.g., SNP markers) can provide us a practically dense map. Therefore, although the asymptotic behaviors of the proposed multiple regression model may not hold exactly as in the dense map case, we believe all major conclusions would be correct in general.

For the proposed multiple regression model, one may find an apparent paradox concerning the desired marker density. On one hand, our results demonstrate that a denser map is better than a sparser map, for example, for isolation of multiple QTLs. On the other hand, however, the variance of the estimates of the partial regression coefficients increases as the marker density increases. How should one compromise these two conflicting demands?

The variance of the estimated partial regression coefficients increases as the marker density increases. This is due to the fact that the advantage of a dense map can only be taken fully if there are enough recombinations in the data. For experimental organisms, the shortage of recombinations in the collected data can be compensated by the use of historical recombinations when experiments are carefully designed [Xiong and Guo (1997)]. In humans, obviously, experimentation is out of the question. However, the use of linkage disequilibrium for fine-scale mapping of QTLs is clearly worth investigating in the future. An alternative approach is to select an optimal subset of markers for mapping QTL. Once we have a dense map, only the markers that are close to QTL are worth fitting in the model [Kao, Zeng and Teasdale (1999)]. The strategy for selection of an optimal set of markers will be also studied in the future.

APPENDICES

Appendix A. Let $U = E[R_1^T R_1]$. Then

$$U = \begin{pmatrix} 1 & E[\pi_{11}] & \cdots & E[\pi_{1j}] & \cdots & E[\pi_{1m}] \\ E[\pi_{11}] & E[\pi_{11}^2] & \cdots & E[\pi_{11}\pi_{1j}] & \cdots & E[\pi_{11}\pi_{1m}] \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ E[\pi_{im}] & E[\pi_{1m}\pi_{11}] & \cdots & E[\pi_{1m}\pi_{1j}] & \cdots & E[\pi_{im}^2] \end{pmatrix}.$$

To determine the matrix U, we need to calculate the elements of the matrix U.

When $\bar{\pi}_{1j} = 0$, the two sibs inherited different alleles from their parents. Thus, $P(\bar{\pi}_{ij} = 0) = \frac{1}{2}\frac{1}{2}$. By symmetry, we have $P(\bar{\pi}_{1j} = 1) = \frac{1}{4}$. When $\bar{\pi}_{1j} = \frac{1}{2}$, the two sibs share one allele IBD. Therefore, $P(\bar{\pi}_{1j} = \frac{1}{2}) = \frac{1}{2}$. Since $\pi_{1j} = \bar{\pi}_{1j} - \frac{1}{2}$,

$$\begin{aligned} E[\pi_{1j}] &= E[\bar{\pi}_{1j}] - \tfrac{1}{2} \\ &= 0, \qquad j = 1, \ldots, m. \end{aligned}$$

Similarly, we have

$$\begin{aligned} E[\pi_{1j}^2] &= \left(-\tfrac{1}{2}\right)^2 \tfrac{1}{4} + 0^2 \tfrac{1}{2} + \left(\tfrac{1}{2}\right)^2 \left(\tfrac{1}{4}\right) \\ &= \tfrac{1}{8}, \qquad j = 1, \ldots, m. \end{aligned}$$

Using results in Haseman and Elston (1972, Table IV), we have

$$\begin{aligned} E[\pi_{1j}\pi_{1j'}] &= \left(-\tfrac{1}{2}\right)\left(\tfrac{1}{2}\right) P(\bar{\pi}_{1j} = 0, \bar{\pi}_{1j'} = 0) \\ &\quad + \left(-\tfrac{1}{2}\right)\left(\tfrac{1}{2}\right) P(\bar{\pi}_{1j} = 0, \bar{\pi}_{1j'} = 1) \\ &\quad + \left(\tfrac{1}{2}\right)\left(-\tfrac{1}{2}\right) P(\bar{\pi}_{1j} = 1, \bar{\pi}_{1j'} = 0) \\ &\quad + \tfrac{1}{2}\tfrac{1}{2} P(\bar{\pi}_{1j} = 1, \bar{\pi}_{1j'} = 1) \\ &= \tfrac{1}{8}(2\Psi - 1) \\ &= \tfrac{1}{8}\exp(-4\Delta_{j,j'}), \quad j, j' = 1, \ldots, m, \quad j \neq j', \end{aligned} \tag{A.1}$$

where $\Psi = c_{j,j'}^2 + (1 - c_{j,j'})^2$ and $c_{jj'}$ is the recombination fraction between marker j and marker j'. Thus, U has form (4).

Since each element of U is finite, by the strong law of large numbers, we have

$$\frac{1}{n}R^T R = \frac{1}{n}\sum_{i=1}^{n} R_i^T R_i \overset{\text{a.s.}}{\to} E[R_1^T R_1] = U. \tag{A.2}$$

Recall that

$$\hat{\beta}_n = (R^T R)^{-1} R^T Z. \tag{A.3}$$

Substituting Z in (3) into (A.3), we obtain

$$\hat{\beta}_n = \beta + (R^T R)^{-1} R^T \varepsilon.$$

Thus,

$$\sqrt{n}(\hat{\beta}_n - \beta) = \sqrt{n}(R^T R)^{-1} R^T \varepsilon.$$

By the central limit theorem and the assumption that ε_i are iid, we obtain

$$\sqrt{n} R^T \varepsilon \overset{\mathcal{L}}{\to} N(0, \sigma^2 U). \tag{A.4}$$

From (A.2) and (A.4), and using Slutsky's theorem, it follows that

$$\sqrt{n}(\hat{\beta}_n - \beta) \overset{\mathcal{L}}{\to} N(0, \sigma^2 U^{-1}) \quad \text{as } n \to \infty.$$

The proof is complete.

Appendix B. From the definitions of R and Z, it follows that

$$\frac{1}{n}R^T Z = \left[\frac{1}{n}\sum_{i=1}^{n} z_i, \frac{1}{n}\sum_{i=1}^{n} \pi_{i1} z_i, \ldots, \frac{1}{n}\sum_{i=1}^{n} \pi_{im} z_i\right]^T. \tag{A.5}$$

To apply the strong law of large numbers, we need to calculate $E[\pi_{1j} z_1]$. By conditioning, we have

$$\begin{aligned} E[\pi_{1j} Z_1] &= -\tfrac{1}{2} P(\bar{\pi}_{1j} = 0) E[Z_1 | \bar{\pi}_{1j} = 0] \\ &\quad + \tfrac{1}{2} P(\bar{\pi}_{1j} = 1) E[Z_1 | \bar{\pi}_{1j} = 1] \\ &= \tfrac{1}{8}(E[Z_1 | \bar{\pi}_{1j} = 1] - E[Z_1 | \bar{\pi}_{1j} = 0]). \end{aligned} \tag{A.6}$$

Let $\bar{\pi}_{t_l}$ be the proportion of alleles shared IBD at the lth trait locus and $g_{t_l} = g_{it_l} - g_{i't_l}$. Therefore, the squared difference is

$$Z_i = \sum_{l=1}^{k} g_{t_l}^2 + 2\sum_{u=1}^{k}\sum_{v=1}^{k} g_{t_u} g_{t_v} + 2\sum_{l=1}^{k} g_{t_l}(e_i - e_{i'}) + (e_i - e_{i'})^2. \tag{A.7}$$

Now we first calculate $E[z_1 | \bar{\pi}_{1j} = 1]$. Conditioning on all QTLs, we have

$$\begin{aligned} E[Z_1 | \bar{\pi}_{1j} = 1] &= \sum_{\pi_{t_1}} \cdots \sum_{\pi_{t_k}} E[Z_1 | \bar{\pi}_{t_1} \cdots \bar{\pi}_{t_k}] P(\bar{\pi}_{t_1} \cdots \bar{\pi}_{t_k} | \bar{\pi}_{1j} = 1) \\ &= \sigma_e^2 + \sum_{l=1}^{k} \left[\sigma_{a(l)}^2 2\Psi_{t_l}(1 - \Psi_{t_l}) + 2\sigma_{a(l)}^2 (1 - \Psi_{t_l})^2\right] \\ &= \sigma_e^2 + 2\sum_{l=1}^{k} \sigma_{a(l)}^2 (1 - \Psi_{t_l}), \end{aligned} \tag{A.8}$$

where $\Psi_{t_l} = c_{jt_l}^2 + (1 - c_{jt_l})^2$ and c_{jt_l} is the recombination fraction between the marker M_j and the lth true QTL. Similarly, we have

$$E[Z_1 | \bar{\pi}_{1j} = 0] = \sigma_e^2 + 2\sum_{l=1}^{k} \sigma_{a(l)}^2 \Psi_{t_l}. \tag{A.9}$$

Thus, it follows from (A.6), (A.8) and (A.9) that

$$x_j = E[\pi_{1j} Z_1] = -\frac{1}{4}\sum_{l=1}^{k} \sigma_{a(l)}^2 \exp(-4\Delta_{jt_l}), \qquad j = 1, \ldots, m. \tag{A.10}$$

Similarly, from (A.7) it follows that

$$x_0 = E[Z_1] = \sum_{l=1}^{k} \sigma^2_{a(l)} + \sigma^2_e. \tag{A.11}$$

By the strong law of large numbers, we obtain

$$\frac{1}{n}\sum_{i=1}^{n} \pi_{ij} Z_i \overset{\text{a.s.}}{\to} E[\pi_{1j} Z_1].$$

Thus,

$$\frac{1}{n} R^T Z \overset{\text{a.s.}}{\to} W = [x_0, x_1, \ldots, x_m]^T.$$

In Theorem 1, we have proved that

$$\frac{1}{n} R^T R \overset{\text{a.s.}}{\to} U$$

or

$$\left(\frac{1}{n} R^T R\right)^{-1} \overset{\text{a.s.}}{\to} U^{-1}.$$

Therefore,

$$\begin{aligned}\hat{\beta} &= (R^T R)^{-1} R^T Z \\ &= \Big(\frac{1}{n} R^T R\Big)^{-1} \frac{1}{n} R^T Z \overset{\text{a.s.}}{\to} U^{-1} W.\end{aligned}$$

Using the tri-diagonal structure of inverse A_i, we obtain

$$\hat{\beta}_j \overset{\text{a.s.}}{\to} \beta^{\star}_j = -\frac{a_p}{1-a_p^2} x_{j-1} + \frac{1-a_p^2 a_r^2}{(1-a_p^2)(1-a_r^2)} x_j - \frac{a_r}{1-a_r^2} x_{j+1},$$

where $x_j = 8x'_j, a_p = \exp(-4\Delta_{j-1,\,j}), a_r = \exp(-4\Delta_{j,\,j+1})$ and $A_i = [\frac{1}{8} \times \exp(-4\Delta_{j,\,j'})]$. This completes the proof.

Appendix C. (i) Under the assumption that a subset of QTLs is located on the left-hand side of marker M_{j-1}, we have

$$\Delta_{jt_l} = \Delta_{j-1,\,t_l} + \Delta_{j-1,\,j}$$

and

$$\Delta_{j+1 t_l} = \Delta_{j,\,t_l} + \Delta_{j,\,j+1}.$$

It follows from (7) that

$$\beta_j^\star = -2\sum_l \sigma_{a(t_l)}^2 \exp(-4\Delta_{jt_l})\left[-\frac{a_p}{1-a_p^2}\frac{1}{a_p} + \frac{1-a_p^2a_r^2}{(1-a_p^2)(1-a_r^2)} - \frac{a_r}{1-a_r^2}a_r\right]$$
$$= 0.$$

(ii) can be similarly proved.

(iii) From the above discussions, we have

$$\Delta_{j+1,\,t_l} = \Delta_{j,\,t_l} + \Delta_{j,\,j+1}.$$

Note that if all QTLs are located between markers M_{j-1} and M_j, then, from (7), it follows that

$$\beta_j^\star = -2\sum_l \sigma_{a(t_l)}^2\left[\frac{-a_p}{1-a_p^2}\exp(-4\Delta_{j-1,\,t_l}) + \frac{1-a_p^2a_r^2}{(1-a_p^2)(1-a_r^2)}\exp(-4\Delta_{j,\,t_l})\right.$$
$$\left.- \frac{a_r}{1-a_r^2}\exp(-4\Delta_{j+1,\,t_l})\right]$$
$$= -2\sum_l \sigma_{a(t_l)}^2\left[\frac{-a_p}{1-a_p^2}\exp(-4\Delta_{j-1,\,t_l}) + \frac{1}{1-a_p^2}\exp(-4\Delta_{j,\,t_l})\right]$$
$$= -2\sum_l \sigma_{a(t_l)}^2 \frac{\exp(-4\Delta_{j,\,t_l})(1-\exp(-8\Delta_{j-1,\,t_l}))}{1-\exp(-8\Delta_{j-1,\,j})}.$$

Similarly, we have

$$\beta_{j-1}^\star + \beta_j^\star = -2\sum_l \sigma_{a(t_l)}^2\left[\frac{1}{1-a_p^2}\exp(-4\Delta_{j-1,\,t_l})\frac{a_p}{1-a_p^2}\exp(-4\Delta_{j,\,t_l})\right.$$
$$\left.+ \frac{-a_p}{1-a_p^2}\exp(-4\Delta_{j-1,\,t_l}) + \frac{1}{1-a_p^2}\exp(-4\Delta_{j,\,t_l})\right]$$
$$= -2\sum_l \sigma_{a(t_l)}^2 \frac{\exp(-4\Delta_{j-1,\,t_l}) + \exp(-4\Delta_{jt_l})}{1+\exp(-4\Delta_{j-1,\,j})}.$$

This completes the proof.

Appendix D. (i) From the proof of Theorem 2, we know that

$$\beta_t^\star = -2\sigma_{a(t^\star)}^2 \frac{\exp(-4|t-t^\star|)(1-\exp(-8(t^\star-l_t)))}{1-\exp(-8(r_t-l_t))}. \tag{A.12}$$

Simple calculations yield that, when $t > t^\star$,

$$\frac{d\beta_t^\star}{dt} = \frac{8\sigma_{a(t^\star)}^2(1-\exp(-8(t^\star-l_t)))}{1-\exp(-8(r_t-l_t))}\exp(-4(t-t^\star)), \tag{A.13}$$

and when $t < t^\star$,

$$\text{(A.14)} \qquad \frac{d\beta_t^\star}{dt} = -\frac{8\sigma^2_{a(t^\star)}(1-\exp(-8(t^\star - l_t)))}{1-\exp(-8(r_t - l_t))}\exp(-4(t^\star - t)).$$

Thus, it follows from (A.13) and (A.14) that, when $l_t \le t < t^\star$, $d\beta_t^\star/dt < 0$ and $t^\star < t \le r_t$, $d\beta_t^\star/dt > 0$, which implies that

$$\beta^\star_{t^\star} < \beta^\star_t, \quad \forall\, t \in [l_t, r_t] \quad \text{and} \quad t \neq t^\star,$$

that is, $\beta_t^\star$ reaches its local minimum at $t^\star$ in the region $[l_t, r_t]$. It follows from (A.13) that

$$\lim_{t\to t^\star_+} \frac{d\beta_t^\star}{dt} = \frac{8\sigma^2_{a(t^\star)}(1-\exp(-8(t^\star - l_t)))}{1-\exp(-8(r_t - l_t))}.$$

Similarly, from (A.14) we have

$$\lim_{t\to t^\star_-} \frac{d\beta_t^\star}{dt} = -\frac{8\sigma^2_{a(t^\star)}(1-\exp(-8(t^\star - l_t)))}{1-\exp(-8(r_t - l_t))}.$$

Taken together, it is obvious that $d\beta_t/dt$ does not exist.

(ii) Recall from Corollary 1,

$$\beta^\star_{l_t} + \beta^\star_{r_t} = -2\sigma^2_{a(t^\star)}\frac{\exp(-4\Delta_{l_t t^\star}) + \exp(-4\Delta_{r_t t^\star})}{1+\exp(-4\Delta_{l_t r_t})},$$

which can be approximated by

$$\begin{aligned}\beta^\star_{l_t} + \beta^\star_{r_t} &\approx -2\sigma^2_{a(t^\star)}\frac{2-4(\Delta_{l_t t^\star} + \Delta_{r_t t^\star})}{2-4\Delta_{l_t r_t}} \\ &= -2\sigma^2_{a(t^\star)}\end{aligned}$$

when $\Delta_{l_t r_t}$ is small.

This completes the proof.

Appendix E. In Appendix B, we show that

$$\begin{aligned}Z_1 &= \sum_{l=1}^{k} g^2_{t_l} + 2\sum_{u=1}^{k}\sum_{v=1}^{k} g_{t_u} g_{t_v} + 2\sum_{l=1}^{k} g_{t_l}\varepsilon_1 + \varepsilon_1^2 \\ &= \alpha + g + 2\left(\sum_{l=1}^{k} g_{t_l}\right)\varepsilon_1 + \varepsilon_1^2 - \sigma_e^2,\end{aligned}$$

where

$$\alpha = \sum_{l=1}^{k} \sigma^2_{a(l)} + \sigma^2_e,$$

$$g = \sum_{l=1}^{k} \big(g^2_{t_l} - \sigma^2_{a(l)}\big) + 2 \sum_{u=1}^{k} \sum_{v=1}^{k} g t_u g t_v.$$

Thus,

$$\begin{aligned} Z_1^2 = {} & (\alpha + g)^2 + 4\left(\sum_{l=1}^{k} g_{t_l}\right)^2 \varepsilon_1^2 + (\varepsilon_1^2 - \sigma_e^2)^2 \\ & + 4(\alpha + g)\left(\sum_{l=1}^{k} g_{t_l}\right)\varepsilon_1 + 2(\alpha + g)\big(\varepsilon_1^2 - \sigma_e^2\big) \\ & + 4\left(\sum_{l=1}^{k} g_{t_l}\right)\varepsilon_1\big(\varepsilon_1^2 - \sigma_e^2\big). \end{aligned}$$

Denote the terms in Z_1^2 involving ε_1 by γ. Then,

$$\begin{aligned} \gamma = {} & 4\left(\sum_{l=1}^{k} g_{t_l}\right)^2 \varepsilon_1^2 + (\varepsilon_1^2 - \sigma_e^2)^2 + 4(\alpha + g)\sum_{l=1}^{k} g_{t_l}\varepsilon_1 \\ & + 2(\alpha + g)(\varepsilon_1^2 - \sigma_e^2) + 4\left(\sum_{l=1}^{k} g_{t_l}\right)\varepsilon_1(\varepsilon_1^2 - \sigma_e^2). \end{aligned}$$

since ε_1 is independent of all genetic effects, we have

$$\begin{aligned} E[\gamma] &= 4\left(\sum_{l=1}^{k} \sigma^2_{a(l)}\right)\sigma_e^2 + 3\sigma_e^4 - 2\sigma_e^4 + \sigma_e^4 \\ &= 4\left(\sum_{l=1}^{k} \sigma^2_{a(l)}\right)\sigma_e^2 + 2\sigma_e^4. \end{aligned}$$

The proof is complete. □

Appendix F. Now we first calculate the matrix $U_\star$. Recall that $\pi^\star_{1j} = \hat{\pi}_{1j} - \frac{1}{2}$. The estimation $\hat{\pi}_{1j}$ of IBD values $\bar{\pi}_{1j}$ can take values 0, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$ or 1 with the probabilities listed in Haseman and Elston [(1972), Table V]. Thus,

$$\begin{aligned} E\big[\pi^{\star 2}_{1j}\big] &= \tfrac{1}{4}P(\hat{\pi}_{1j} = 0) + \tfrac{1}{16}P\Big(\hat{\pi}_{1j} = \tfrac{1}{4}\Big) + \tfrac{1}{16}P\Big(\hat{\pi}_{1j} = \tfrac{3}{4}\Big) + \tfrac{1}{4}P(\hat{\pi}_{1j} = 1) \\ &= \tfrac{1}{4}p_j q_j(1 - p_j q_j). \end{aligned}$$

To calculate $E[\pi^\star_{1j}\pi^\star_{1l}]$, we need to calculate $P(\hat{\pi}_{1j}\hat{\pi}_{1l})$. Clearly,

$$\begin{aligned} P(\hat{\pi}_{1j} = 1, \hat{\pi}_{1l} = 1) &= \sum_m \sum_n P(\hat{\pi}_{1j} = 1, \hat{\pi}_{1l} = 1, \bar{\pi}_{1m}\bar{\pi}_{1n}) \\ &= p_j^2 q_j^2 p_l^2 q_l^2 \Psi^2, \end{aligned}$$

where $\Psi = \theta_{jl}^2 + (1 - \theta_{jl})^2$ and θ_{jl} is the recombination fraction between the marker locus M_j and the marker locus M_l. Arguing similarly above, we obtain

$$\begin{aligned} P\big(\hat{\pi}_{1j} = \tfrac{3}{4}, \hat{\pi}_{1l} = 1\big) &= 2p_j^2 q_j^2 (p_l^3 q_l + p_l q_l^3)\Psi, \\ P\big(\hat{\pi}_{1j} = \tfrac{3}{4}, \hat{\pi}_{1l} = \tfrac{3}{4}\big) &= 2(p_j^3 q_j + p_j q_j^3)(p_l^3 q_l + p_l q_l^3)(2\Psi + 1), \\ P\big(\hat{\pi}_{1j} = \tfrac{1}{4}, \hat{\pi}_{1l} = 1\big) &= 2P_j^2 q_j^2 (p_l^3 q_l + p_l q_l^3)(1 - \Psi), \\ P\big(\hat{\pi}_{1j} = \tfrac{1}{4}, \hat{\pi}_{1l} = \tfrac{3}{4}\big) &= 2(p_j^3 q_j + p_j q_j^3)(p_l^3 q_l + p_l q_l^3)(3 - 2\Psi), \\ P\big(\hat{\pi}_{1j} = \tfrac{1}{4}, \hat{\pi}_{1l} = \tfrac{1}{4}\big) &= 2(p_j^3 q_j + p_j q_j^3)(p_l^3 q_l + p_l q_l^3)(2\Psi + 1), \\ P\big(\hat{\pi}_{1j} = \tfrac{1}{4}, \hat{\pi}_{1l} = 0\big) &= 2p_j^2 q_j^2 (p_l^3 q_l + p_l q_l^3)\Psi, \\ P\big(\hat{\pi}_{1j} = 0, \hat{\pi}_{1l} = 1\big) &= p_j^2 q_j^2 p_l^2 q_l^2 (1 - \Psi)^2, \\ P\big(\hat{\pi}_{1j} = 0, \hat{\pi}_{1l} = \tfrac{3}{4}\big) &= 2p_j^2 q_j^2 (p_l^3 q_l + p_l q_l^3)(1 - \Psi), \\ P\big(\hat{\pi}_{1j} = 0, \hat{\pi}_{1l} = \tfrac{1}{4}\big) &= 2p_j^2 q_j^2 (p_l^3 q_l + p_l q_l^3)\Psi, \\ P\big(\hat{\pi}_{1j} = 0, \hat{\pi}_{1l} = 0\big) &= p_j^2 q_j^2 q_l^2 p_l^2 \Psi^2. \end{aligned}$$

Using the above results and computing $E[\pi^\star_{1j}\pi^\star_{1l}]$ by conditioning on the valves of $P(\hat{\pi}_{1j}, \hat{\pi}_{1l})$, we have

$$\begin{aligned} E[\pi^\star_{1j}\pi^\star_{1l}] &= \tfrac{1}{2} p_j q_j p_l q_l (1 - 2p_l q_l + p_j q_j p_l q_l)(2\Psi - 1) \\ &= \tfrac{1}{2} p_j q_j p_l q_l (1 - 2p_l q_l + p_j q_j p_l q_l) \exp(-4\Delta_{jl}). \end{aligned}$$

Similar to the proof in Appendix B, we can show that

$$E[Z_1 | \hat{\pi}_{1j} = 1] = \sigma_e^2 + 2\sum_{l=1}^{k} \sigma_{a(l)}^2 \Psi,$$

$$E\Big[Z_1 | \hat{\pi}_{1j} = \tfrac{1}{4}\Big] = \sigma_e^2 + \frac{1}{2}\sum_{l=1}^{k} \sigma_{a(l)}^2 (2\Psi + 1),$$

$$E\Big[Z_1 | \hat{\pi}_{1j} = \tfrac{3}{4}\Big] = \sigma_e^2 + \frac{1}{2}\sum_{l=1}^{k} \sigma_{a(l)}^2 (3 - 2\Psi),$$

$$E[Z_1 | \hat{\pi}_{1j} = 1] = \sigma_e^2 + 2\sum_{l=1}^{k} \sigma_{a(l)}^2 (1 - \Psi),$$

which implies

$$x_j = E[Z_1 \pi_{1j}^\star] = -\frac{5}{4}\sum_{l=1}^{k} \sigma_{a(l)}^2 \exp(-4\Delta_{jl}).$$

Therefore,

$$\hat{\beta}_n^\star \overset{\text{a.s.}}{\to} U_\star^{-1} W_\star.$$

Appendix G. In Theorem 1, we have shown that

$$\hat{\beta}_t(n) \overset{\text{a.s.}}{\to} \beta_t^\star.$$

Taking $\{t_n = t\}_{n=1}^{\infty}$, then condition (17) of the epi-convergence is satisfied. Now our main task is to verify condition (16).

Since it is difficult to obtain the explicit formula for $\hat{\beta}_{t_n}(n)$, is difficult to verify directly condition (16) by using the explicit formula for $\hat{\beta}_{t_n}(n)$. To indicate that the elements of the matrix R depend on t_n, we denote R by $R(t_n)$. For the same reason, we denote the matrix U and the vector W by $U(t)$ and $W(t)$ to emphasize their dependence on the marker locus t.

To verify that condition (16) is satisfied, we first show that

$$\lim_{n\to\infty}\left[\frac{1}{n}R^T(t_n)R(t_n)\right]^{-1} = U^{-1}(t) \tag{A.15}$$

and

$$\lim_{n\to\infty}\frac{1}{n}R^T(t_n)Z = [E[z_1], E[\pi_{11}Z_1]\cdots E[\pi_{1t}z_1]\cdots E[\pi_{1m}z_1]]. \tag{A.16}$$

Let $A(t_n) = \frac{1}{n}R^T(t_n)R(t_n)$. If we can show

$$\lim_{n\to\infty} A(t_n) = U(t), \tag{A.17}$$

then

$$\lim_{n\to\infty} A^{-1}(t_n) = U^{-1}(t) \tag{A.18}$$

will hold. To see this, let $A(t_n) = (a_{ij}(t_n))_{m\times m}$, $B(t_n) = (b_{ij}(t_n))_{m\times m} = A^{-1}(t_n)$ and $U(t) = (u_{ij}(t))_{m\times m}$; then $(b_{ij}(t_n) = f(a_{11}(t_n), \dots, a_{mn}(t_n)), i, j = 1, \dots, m)$ is a continuous function of $(a_{ij}(t_n), i, j = 1, \dots, m)$ because both $A^{-1}(t_n)$ and $U^{-1}(t)$ exist. Therefore,

$$\begin{aligned}\lim_{n\to\infty} b_{ij}(t_n) &= f\left(\lim_{n\to\infty} a_{11}(t_n), \dots, \lim_{n\to\infty} a_{mm}(t_n)\right) \quad \text{(by continuity of function } f\text{)}\\ &= f(u_{11}(t), \dots, u_{mm}(t)) \quad \text{(by (A.17))}\\ &= U^{-1}(t).\end{aligned}$$

Now we prove (A.17) is satisfied.

Recall that

$$\frac{1}{n}R^T(t_n)R(t_n) = \begin{pmatrix} 1 & \cdots & \frac{1}{n}\sum_{i=1}^{n}\pi_{it_n} & \cdots & \frac{1}{n}\sum_{i=1}^{n}\pi_{im} \\ \frac{1}{n}\sum_{i=1}^{n}\pi_{i1} & \cdots & \frac{1}{n}\sum_{i=1}^{n}\pi_{i1}\pi_{it_n} & \cdots & \frac{1}{n}\sum_{i=1}^{n}\pi_{i1}\pi_{im} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{n}\sum_{i=1}^{n}\pi_{im} & \cdots & \frac{1}{n}\sum_{i=1}^{n}\pi_{im}\pi_{it_n} & \cdots & \frac{1}{n}\sum_{i=1}^{n}\pi_{im}^2 \end{pmatrix}. \tag{A.19}$$

Thus, we need only to prove

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\pi_{it_n} = E[\pi_{1t}] \tag{A.20}$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\pi_{ij}\pi_{it_n} = E[\pi_{1j}\pi_{1t}], \quad j = 1,\dots,m, \tag{A.21}$$

and

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\pi_{it_n}^2 = E[\pi_{1t}^2]. \tag{A.22}$$

Since proving that (A.20)–(A.22) are satisfied is similar to proving that (A.20) is satisfied, and since (A.22) is easier to prove than (A.21), here we only sketch the proof of (A.21).

In Theorem 1, we have proved that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\pi_{ij}\pi_{it} = E[\pi_{1j}\pi_{1t}] \qquad \text{a.s.} \tag{A.23}$$

Thus,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\pi_{ij}(\pi_{it_n} - \pi_{it}) = 0 \qquad \text{a.s.} \tag{A.24}$$

implies (A.21). Now we prove (A.24). Note that by the strong law of large numbers, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\pi_{ij}\pi_{it_n} = E[\pi_{1j}\pi_{1t}] \qquad \text{a.s.} \quad \text{(by (A.23))} \tag{A.25}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}(\pi_{ij}\pi_{it_n} - E[\pi_{1j}\pi_{1t_n}]) \overset{\text{a.s.}}{\to} 0. \tag{A.26}$$

From (A.1), it follows that

$$\begin{aligned} E[\pi_{1j}\pi_{1t_n}] - E[\pi_{1j}\pi_{1t}] &= \tfrac{1}{8}\exp(-4\Delta_{j,t_n}) - \tfrac{1}{8}\exp(-4\Delta_{j,t}) \\ &\le \tfrac{1}{8}\big|\exp(-4|\Delta_{j,t_n} - \Delta_{jt}|) - 1\big| \\ &= \tfrac{1}{8}\big|\exp(-4|t_n - t|) - 1\big| \\ &\to 0 \quad (\text{as } t_n \to t). \end{aligned} \tag{A.27}$$

Combining (A.25), (A.26) and (A.27), we have

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(\pi_{ij}\pi_{it_n} - \pi_{ij}\pi_{it}) &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(\pi_{ij}\pi_{it_n} - E[\pi_{1j}\pi_{1t_n}]) \\ &\quad - \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(\pi_{ij}\pi_{it} - E[\pi_{1j}\pi_{1t}]) \\ &\quad + \lim_{n\to\infty}(E[\pi_{1j}\pi_{1t_n}] - E[\pi_{1j}\pi_{1t}]) \\ &\to 0, \end{aligned}$$

which proves (A.24), and hence (A.23) and (A.15).

Now we prove (A.16). In (A.16), only one term $\frac{1}{n}\sum_{i=1}^{n}\pi_{it_n}z_1$ which involves t_n needs to be dealt with because the other terms in the equation have already been considered in Theorem 1. By the same argument as that used in proving (A.24), we have

$$\begin{aligned} &\lim_{n\to\infty}\left(\sum_{i=1}^{n}\pi_{it_n}z_i - E[\pi_{1t}z_1]\right) \\ &\quad = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\pi_{it_n}z_i - \pi_{it}z_i) + \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\pi_{it}z_i - E[\pi_{1t}z_1]) \\ &\quad = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\pi_{it_n}z_i - E[\pi_{1t_n}z_1]) - \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\pi_{it}z_i - E[\pi_{1t}z_1]) \\ &\qquad + \lim_{n\to\infty}(E[\pi_{1t_n}z_1] - E[\pi_{1t}z_1]) + \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\pi_{it}z_i - E[\pi_{1t}z_1]) \\ &\quad \to 0, \end{aligned} \tag{A.28}$$

which implies (A.16). Note that in proving that the penultimate term in (A.28) converges to zero we also used the following fact:

$$\begin{aligned} \lim_{n\to\infty}(E[\pi_{1t_n}z_1] - E[\pi_{1t}z_1]) &= -\tfrac{1}{4}\sigma_a^2(\exp(-4\Delta_{t_n t^*}) - \exp(-4\Delta_{tt^*})) \\ &\le \tfrac{1}{4}\sigma_a^2\big|\exp(-4|\Delta_{t_n t^*} - \Delta_{tt^*}|) - 1\big| \\ &\le \tfrac{1}{4}\sigma_a^2\big|\exp(-4|t_n - t|) - 1\big| \\ &\to 0 \quad (\text{as } t_n \to t). \end{aligned} \tag{A.29}$$

Let $a(t_n)$ denote the row of $A^{-1}(t_n)$ corresponding to the putative QTL t. Then (A.15) and (A.16) imply that

$$\lim_{n\to\infty}\hat{\beta}_{t_n}(n) = \lim_{n\to\infty} a(t_n)\lim_{n\to\infty}\frac{1}{n}R^T(t_n)Z, \tag{A.30}$$

which suggests that the condition (16) is satisfied since $\lim_{n\to\infty}\hat{\beta}_{t_n}(n)$ exists and hence

$$\underline{\lim}_{n\to\infty}\hat{\beta}_{t_n}(n) = \lim_{n\to\infty}\hat{\beta}_{t_n}(n).$$

REFERENCES

ALMASY, L. and BLANGERO, J. (1998). Multipoint quantitative-trait linkage analysis in general pedigrees. *Amer. J. Hum. Genetics* **62** 1198–1211.

AMOS, C. I. (1994). Robust variance-components approach for assessing genetic linkage in pedigress. *Amer. J. Hum. Genetics* **54** 535–543.

AUBIN, J. P. and FRANKOWSKA, H. (1990). *Set-Valued Analysis.* Birkhäuser, Boston.

BEGLEITER, H., PORJESZ, B., REICH, T., EDENBERG, H. J., GOATE, A., and BLANGERO, J. et al. (1998). Quantitative trait loci analysis of human event-related brain potentials: P3 voltage. *Electroencephologr. Clin. Neurophysiol.* **108** 244–250.

BLANGERO, J. and ALMASY, L. (1997). Multipoint oligogenic linkage analysis of quantitative traits. *Genet. Epidemiol*. **14** 959–964.

COMUZZIE, A. G., HIXSON, J. E., ALMASY, L., MITCHELL, B. D., MAHANEY, M. C., DYER, T. D., STEN, M. P., MACCLUER, J. W. and BLANGERO, J. (1997). A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2. *Nat. Genetics* **15** 273–276.

DRAGANI, T. A., ZENG, Z.-B., CANZIAN, F., GARIBOLDI, M., GHILARDUCCI, M. T., MANENTI, G. and PIEROTTI, M. A. (1995). Mapping of body weight loci on mouse chromosome X. *Mammalian Genome* **6** 778–781.

DUGGIRALA, R., BLANGERO, J., ALMASY, L., DYER, T. D., WILLIAMS, K. L., LEACH, R. J., O'CONNELL, P. and STERN, M. P. (1999). Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. *Amer. J. Hum. Genetics* **64** 1127–1140.

DUPAČOVA, J. and WETS, R. (1988). Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *Ann. Statist.* **16** 1517–1549.

FEINGOLD, S., BROWN, O. P. and SIEGMUND, D. (1993). Gaussian models for genetic linkage analysis using complete high-resolution maps of identity by descent. *Amer. J. Hum. Genetics* **53** 234–251.

FULKER, D. W. and CARDON, L. R. (1994). A sib-pair approach to interval mapping of quantitative trait loci. *Amer. J. Hum. Genetics* **54** 1092–1103.

GOLDGAR, D. E. (1990). Multipoint analysis of human quantitative genetic variation. *Amer. J. Hum. Genetics* **47** 957–967.

HALEY, C. S., KNOTT, S. A. and ELSEN, J. M. (1994). Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* **136** 1195–1207.

HALEY, C. S. and KNOTT, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69** 315–324.

HASEMAN, J. K. and ELSTON, R. C. (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behavior Genetics* **2** 3–19.

HOESCHELE, I. and VANRANDEN, P. (1993). Bayesian analysis of linkage between genetic markers and quantitative trait loci: II. Combining prior knowledge with experimental evidence. *Theor. Appl. Genetics* **85** 946–952.

JANSEN, R. C. (1993). Interval mapping of multiple quantitative trait loci. *Genetics* **135** 205–211.

JANSEN, R. C. (1994). Controlling the Type I and Type II errors in mapping quantitative trait loci. *Genetics* **138** 871–881.

KAO, C.-H., ZENG, Z.-B. and TEASDALE, R. D. (1999). Multiple interval mapping for quantitative trait loci. *Genetics* **132** 1203–1216.

KNOTT, S. A. and HALEY, C. S. (1992). Maximum likelihood mapping of quantitative trait loci using full-sib families. *Genetics* **132** 1211–1222.

KUITTINEN, H., SILLANPÄÄ, M. J. and SAVOLAINEN, O. (1997). Genetic basis of adaptation, flowering time in *Arabidopsis Thaliana. Theor. Appl. Genetics* **95** 573–583.

LANDER, E. S. and BOTSTEIN, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121** 185–199.

LANDER, E. and SCHORK, N. J. (1994). Genetic dissection of complex traits. *Science* **265** 2037–2048.

LIU, J., MERCER, J. M., STEM, L. F., GIBSON, G. C., ZENG, Z.-B. and LAURIE, C. C. (1996). Genetic analysis of a morphological shape difference in the male genitalia of *Drosiphite simulans and D. mauritiana. Genetics* **142** 1129–1145.

LUO, Z. W. and KEARSEY, M. J. (1992). Interval mapping of quantitative trait loci in an F_2 population. *Heredity* **69** 236–242.

MARTINEZ, O. and CURNOW, R. N. (1992). Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theor. Appl. Genetics* **85** 480–488.

OLSON, J. M. (1995). Robust multipoint linkage analysis: an extension of the Haseman–Elston method. *Genetic Epidemiology* **12** 177–193.

RODOLPHE, F. and LEFORT, M. (1993). A multi-marker model for detecting chromosomal segments displaying QTL activity. *Genetics* **134** 1277–1288.

SATAGOPAN, J. M., YANDELL, B. S., NEWTON, M. A. and OSBORN, T. C. (1996). A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics* **144** 805–816.

SCHORK, N. J. (1993). Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations. *Amer. J. Hum. Genetics* **53** 1306–1319.

SILLANPÄÄ, M. J. and ARJAS, E. (1998). Bayesian mapping of multiple quantitative trait loci from incomplete line cross data. *Genetics* **148** 1373–1388.

TEMPLETON, A. R. (1999). Uses of evolutionary theory in the human genome project. *Annu. Rev. Ecol. Syst.* **30** 23–49.

WANG, D. G., FAN, J. B. and SIAO, C. J., et al. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280** 1082–1086.

WANG, X. L., MAHANEY, M. C., SIM, A. S., WANG, J., WANG, J., BLANGERO, J., ALMASY, L., BADENHOP, R. B. and WILCKEN, D. E. L. (1997). Genetic contribution of the endothelial constitutive nitric oxide synthase gene to plasma nitric oxide levels. *Arterioscler. Thromb. Vasc. Biol.* **17** 3147–3153.

WRIGHT, F. (1994). Asymptotics and robustness for genetic linkage mapping. Ph.D. dissertation, Univ. Chicago.

WETS, R. (1991). Constraint estimation: consistency and asymptotics. *Appl. Stochastic Models Data Anal.* **7** 17–32.

XIONG, M. and GUO, S. W. (1997). Fine-scale mapping based on linkage disequilibrium: theory and applications. *Amer. J. Hum. Genetics* **60** 1513–1531.

XU, S. and ATCHLEY, W. R. (1995). A random model approach to interval mapping of quantitative trait loci. *Genetics* **141** 1189–1197.

ZENG, Z. B. (1993). Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc. Nat. Acad. Sci. U.S.A.* **90** 10972–10976.

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

HUMAN GENETICS CENTER
UNIVERSITY OF TEXAS—HOUSTON HEALTH
SCIENCE CENTER
HOUSTON, TEXAS 77225
E-MAIL: mxiong@sph.uth.tmc.edu

DEPARTMENT OF PEDIATRICS
MEDICAL COLLEGE OF WISCONSIN
MILWAUKEE, WISCONSIN 53226