Inferring Linkage Disequilibrium Between a Polymorphic Marker Locus and a Trait Locus in Natural Populations

Z. W. Luo*, S. H. Tao* and Z-B. Zeng†

**Laboratory of Population and Quantitative Genetics, Institute of Genetics, Fudan University, Shanghai 200433, People's Republic of China and †Program in Statistical Genetics, Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695-8203*

Manuscript received January 10, 2000
Accepted for publication May 26, 2000

ABSTRACT

Three approaches are proposed in this study for detecting or estimating linkage disequilibrium between a polymorphic marker locus and a locus affecting quantitative genetic variation using the sample from random mating populations. It is shown that the disequilibrium over a wide range of circumstances may be detected with a power of 80% by using phenotypic records and marker genotypes of a few hundred individuals. Comparison of ANOVA and regression methods in this article to the transmission disequilibrium test (TDT) shows that, given the genetic variance explained by the trait locus, the power of TDT depends on the trait allele frequency, whereas the power of ANOVA and regression analyses is relatively independent from the allelic frequency. The TDT method is more powerful when the trait allele frequency is low, but much less powerful when it is high. The likelihood analysis provides reliable estimation of the model parameters when the QTL variance is at least 10% of the phenotypic variance and the sample size of a few hundred is used. Potential use of these estimates in mapping the trait locus is also discussed.

LINKAGE disequilibrium between pairs of loci may be created by a variety of population genetic processes. When a new mutant occurs, it has a strong positive disequilibrium with the carrier haplotype and negative disequilibria with other haplotypes in the population. Genetic drift, population subdivision or admixture, assortative mating, and selection can also introduce disequilibria between different loci. Genetic linkage generally serves as a major mechanism that maintains the disequilibrium in a population over a long period of time, regardless of how the disequilibrium was introduced into the population (Hill and Robertson 1968; Luo *et al.* 1997). The relationship between the recombination rate and the rate of dissipation in the disequilibrium supports the use of population-based linkage disequilibrium analysis for fine-scale mapping of disease genes (Chakraborty and Weiss 1988; Spileman *et al.* 1993; Hill and Weir 1994; Jorde 1995; Collins and Morton 1998; de la Chapelle and Wright 1998; Laan and Paabo 1998; George *et al.* 1999). This population-based linkage disequilibrium analysis has been used to map mutant genes that cause a number of human inherited disorders, such as myotonic dystrophy (Harley *et al.* 1991), diastrophic dysplasia (Hastbacka *et al.* 1992), adenomatous polyposis coli (Jorde 1995), and inclusion body myopathy (Eisenberg *et al.* 1999). A common assumption made in these theoretical and experimental studies is that all affected individuals must share a common disease-causing mutation that is inherited by descent, and utility of the methods is limited to the circumstances where gametic or genotypic data are available at the involved loci.

There has been a significant shift in human disease gene mapping in recent years, from simple Mendelian diseases to complex diseases (Lander and Schork 1994; Fulker *et al.* 1995; Risch and Zhang 1996; Allison 1997; Luo 1998; Allison *et al.* 1999; Service *et al.* 1999). Many complex diseases are affected by multiple genes, which sometimes are called quantitative trait loci (QTL). A major challenge in analyzing the genetic basis of complex diseases is to uncover the genotypic information at the QTL. Neimann-Sorenson and Robertson (1961) were probably the first to attempt to formulate quantitative genetic variation associated with molecular DNA markers in segregating populations. A theoretical analysis by Luo *et al.* (1997) showed that the key component of the marker-associated polygenic variance can be parameterized in term of the linkage disequilibrium between marker gene and QTL in populations with various genetic structures. Spileman *et al.* (1993) developed the transmission disequilibrium test (TDT) for detecting compound association between a genetic marker and disease susceptibility due to genetic linkage and linkage disequilibrium between the marker and disease loci. The TDT is for analyzing transmission of marker alleles from parents to affected offspring for those markers that are heterozygotes. Allison (1997) extended the TDT approach to analyze the association between a trait locus and a marker locus. A more general model was proposed in Luo (1998) to detect linkage

Corresponding author: Z. W. Luo, Laboratory of Population and Quantitative Genetics, Institute of Genetics, Fudan University, Shanghai 200433, People's Republic of China. E-mail: zwluo@fudan.edu.cn

disequilibrium between a polymorphic marker locus and a trait locus from a natural population. The population-based method has certain advantages over the family-based method, as the use of recombination events that accumulated during the evolutionary history of a population may improve resolution of gene mapping, and the feasibility in collecting a much larger sample size can improve the statistical power to detect the linkage disequilibrium. However, it must be pointed out that there could be some uncertainty involved in the population-based analysis without taking the appropriate evolutionary process into account. Luo and Suhai (1999) recently proposed a likelihood approach to estimate the coefficient of linkage disequilibrium on the basis of reparameterization of the disequilibrium coefficient. The method may provide accurate estimation of the genetic parameters where the trait locus explains at least 25% of phenotypic variation of the trait and a sample size >500 is required.

In this study, a theoretical model is developed to formulate joint distribution of marker and QTL genotypes in a natural population with random mating, which is a function of the frequencies of genes at the two loci and the disequilibrium coefficient. This allows a closed form of maximum-likelihood estimation of genetic parameters that define the model. We study various statistical properties of this method and compare it with the analysis of variance and regression on markers on the statistical power to detect the linkage disequilibrium. We also compare the power of the analysis of variance and regression with the family-based TDT analysis.

THEORY AND METHODS

A quantitative genetic model is usually defined on the basis of the least-squares principle that specifies genetic effects as a deviation from the population mean and other lower-order genetic effects (Weir and Cockerham 1977). Applying the model to association studies between a marker and a trait, Nielsen and Weir (1999) showed that there is a simple relationship between the marker being examined and the trait loci. Additive effects of the marker alleles are functions of the additive effects of the trait loci and the disequilibria between the marker and the trait loci, and dominance effects of the marker genotypes are functions of the dominance effects of the trait loci and the disequilibria between the marker and trait loci. This model is a very appropriate framework to specify the structure of the associations between loci and to examine the genetic context of many marker-based statistical tests. Here, however, our purpose is to directly estimate the genetic parameters of QTL and the linkage disequilibrium through the marker information.

Consider two autosomal loci: one affects a quantitative trait and the other is a codominant marker locus that is devoid of effect on the trait. The two alleles are denoted by M and m at the marker locus and by A and a at the QTL. The three genotypes at the QTL, say, AA, Aa, and aa, are assumed to have the genotypic values

$$\begin{aligned} G_{AA} &= \mu + a - d/2, \\ G_{Aa} &= \mu + d/2, \\ G_{aa} &= \mu - a - d/2, \end{aligned} \tag{1}$$

where a and d are the additive and dominance effects of QTL, respectively. The association between the two loci in a natural population is quantified by linkage disequilibrium with D, the coefficient of the disequilibrium, defined as $D = f_{MA} - pq$, where f_{MA} is the frequency of the MA haplotype and p and q denote frequencies of alleles M and A accordingly. With the assumption of random mating, the joint distribution of genotypes at the marker locus and QTL can be expressed as a function of D, p, and q and is shown in Table 1. For example, frequency of genotype $MMAA$ is calculated as $f_{MA}^2 = (D + pq)^2$.

Maximum-likelihood analysis: Consider a random sample of size n from the population defined above and let z_j and x_j be the phenotypic value and marker genotype of the jth individual in the sample. Under the above model, the likelihood of the observed trait data (Z) and the marker genotypic data (X), given the population genetic parameters $\Omega = (p, q, D, \mu, a, d, v)$, can be written as

$$L(\Omega, Z, X) = \prod_{j=1}^{n} \Pr\{z_j|x_j, \Omega\}\Pr\{x_j|\Omega\},$$

where $\Pr\{z_j|x_j, \Omega\}$ is the conditional probability of z_j given the marker genotype x_j and the parameters Ω, and $\Pr\{x_j|\Omega\}$ the probability of the marker genotype x_j given the parameters. Let y_{ij} be the phenotypic value of the jth individual with the ith marker genotype ($i = 1$, 2, and 3 corresponding to MM, Mm, and mm, respectively) and assume y_{ij} follows a normal distribution with mean μ_k and variance v; the log-likelihood can be simplified into a form given by

$$\log[L(\Omega, Y, X)] = \sum_{i=1}^{3}\sum_{j=1}^{n_i} \log\left[\sum_{k=1}^{3} h_{ik}e_{ijk}\right] - \frac{n}{2}\log(2\pi v), \tag{2}$$

where $e_{ijk} = \exp[-(y_{ij} - \mu_k)^2/2v]$, $\mu_k = \mu + (2 - k)a + (-1)^k d/2$ being the phenotypic mean for the individuals with the kth QTL genotype ($k = 1$, 2, and 3 corresponding to AA, Aa, and aa, respectively), and $n = n_1 + n_2 + n_3$ with n_i being the number of the ith marker genotype in the sample. Note that h_{ik} is the joint frequency of the ith marker genotype and the kth QTL genotype (Table 1). To obtain the maximum-likelihood estimates of the unknown parameters, we use the expectation-maximization (EM) algorithm. It should be noted that data for estimating marker allele frequency contain no missing information, and the estimate of p is calculated directly

TABLE 1

Joint distribution of marker and QTL genotypes in a random mating population

	QTL genotype		
Marker genotype	*AA*	*Aa*	*aa*
MM	$(D + pq)^2$	$2(D + pq)[p(1 - q) - D]$	$[D - p(1 - q)]^2$
Mm	$2(D + pq)[(1 - p)q - D]$	$2[2D^2 + (1 - 2p)(1 - 2q)D + 2pq(1 - p)(1 - q)]$	$2[D + (1 - p)(1 - q)][p(1 - p) - D]$
Mm	$[D - (1 - p)q]^2$	$2[(1 - p)q - D][D + (1 - p)(1 - q)]$	$[D + (1 - p)(1 - q)]^2$
Genotypic value	$\mu + a - d/2$	$\mu + d/2$	$\mu - a - d/2$

D, the coefficient of linkage disequilibrium between the marker and QTL; p and q, the frequencies of marker allele M and QTL allele A, respectively; a and d, the additive and dominance effects of QTL.

from marker genotype frequencies as $p = (n_1 + n_2/2)/n$. The algorithm involves the iteration of the following two steps.

E-step: Calculate the posterior probability of the jth individual having the kth QTL genotype given its phenotypic value y_{ij}, the marker genotype i, and the values of the parameters at the tth iteration $\Omega^{(t)} = (p, q^{(t)}, D^{(t)}, \mu^{(t)}, a^{(t)}, d^{(t)}, v^{(t)})$ as

$$w_{ijk} = \frac{h_{ik}\exp[-(y_{ij} - \mu_k^{(t)})^2/2v^{(t)}]}{\Sigma_{l=1}^{3} h_{il}\exp[-(y_{ij} - \mu_l^{(t)})^2/2v^{(t)}]}. \tag{3}$$

M-step: Substitute w_{ijk} and $\Omega^{(t)}$ into the following equations. The updated estimates, $D^{(t+1)}$ and $q^{(t+1)}$, can be obtained numerically (*e.g.*, Press *et al.* 1992) by solving

$$\begin{aligned}
\sum_{i=1}^{3}\sum_{j=1}^{n_i}\sum_{k=1}^{3} w_{ijk}\frac{\partial}{\partial D}[\log h_{ik}] = \sum_{i=1}^{3}\sum_{k=1}^{3} \hat{w}_{ik}\frac{\partial}{\partial D}[\log h_{ik}] &= \frac{2\hat{w}_{11}}{D + pq} \\
&+ \frac{[2D - p(1 - 2q)]\hat{w}_{12}}{(D + pq)[D - p(1 - q)]} + \frac{2\hat{w}_{13}}{D - p(1 - q)} \\
&+ \frac{[2D - (1 - 2p)q]\hat{w}_{21}}{(D + pq)[D - (1 - p)q]} \\
&+ \frac{[4D + (1 - 2p)(1 - 2q)]\hat{w}_{22}}{2D^2 + (1 - 2p)(1 - 2q)D + 2pq(1 - p)(1 - q)} \\
&+ \frac{[2D + (1 - 2p)(1 - 2q)]\hat{w}_{23}}{[D + (1 - p)(1 - q)][D - p(1 - q)]} \\
&+ \frac{2\hat{w}_{31}}{D - (1 - p)q} \\
&+ \frac{[2D + (1 - p)(1 - 2q)]\hat{w}_{32}}{[D + (1 - p)(1 - q)][D - (1 - p)q]} \\
&+ \frac{2\hat{w}_{33}}{D + (1 - p)(1 - q)} \\
&= 0
\end{aligned}$$

and

$$\begin{aligned}
\sum_{i=1}^{3}\sum_{j=1}^{n_i}\sum_{k=1}^{3} w_{ijk}\frac{\partial}{\partial D}[\log h_{ik}] = \sum_{i=1}^{3}\sum_{k=1}^{3} \hat{w}_{ik}\frac{\partial}{\partial D}[\log h_{ik}] &= \frac{2p\hat{w}_{11}}{D + pq} \\
&+ \frac{p[2D - p(1 - 2q)]\hat{w}_{12}}{(D + pq)[D - p(1 - q)]} + \frac{2p\hat{w}_{13}}{D - p(1 - q)} \\
&- \frac{[(1 - 2p)D + 2p(1 - p)q]\hat{w}_{21}}{(D + pq)[D - (1 - p)q]} \\
&- \frac{2[(1 - 2p)D - p(1 - p)(1 - 2q)]\hat{w}_{22}}{2D^2 + (1 - 2p)(1 - 2q)D + 2pq(1 - p)(1 - q)} \\
&- \frac{[(1 - 2p)D - 2p(1 - p)(1 - q)]\hat{w}_{23}}{[D + (1 - p)(1 - q)][D - p(1 - q)]} \\
&+ \frac{2(1 - p)\hat{w}_{31}}{D - (1 - p)q} \\
&- \frac{(1 - p)[2D + (1 - p)(1 - 2q)]\hat{w}_{32}}{[D + (1 - p)(1 - q)][D - (1 - p)q]} \\
&+ \frac{2(1 - p)\hat{w}_{33}}{D + (1 - p)(1 - q)} \\
&= 0,
\end{aligned}$$

where $\hat{w}_{ik} = \Sigma_{j=1}^{n_i} w_{ijk}$.

The updated estimates of the parameters μ, a, d, and v are

$$\mu^{(t+1)} = \frac{1}{n}\left[\sum_{i=1}^{3}\sum_{j=1}^{ni} y_{ij} + \sum_{i=1}^{3}\sum_{j=1}^{ni}(w_{ij3} - w_{ij1})\,a^{(t)} + \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{ni}(w_{ij1} - w_{ij2} + w_{ij3})\,d^{(t)}\right] \quad (4)$$

$$a^{(t+1)} = \frac{1}{\sum_{i=1}^{3}\sum_{j=1}^{ni}(w_{ij1} + w_{ij3})} \times \sum_{i=1}^{3}\sum_{j=1}^{ni}(w_{ij1} - w_{ij3})(y_{ij} - \mu^{(t+1)} + d^{(t)}/2) \quad (5)$$

$$d^{(t+1)} = \frac{2}{n}\left[\sum_{i=1}^{3}\sum_{j=1}^{ni}(w_{ij2} - w_{ij1} - w_{ij3})(y_{ij} - \mu^{(t+1)}) + \sum_{i=1}^{3}\sum_{j=1}^{ni}(w_{ij1} - w_{ij3})\,a^{(t+1)}\right] \quad (6)$$

$$v^{(t+1)} = \frac{1}{n}\sum_{i=1}^{3}\sum_{j=1}^{ni}\sum_{k=1}^{3}[w_{ij1}(y_{ij} - \mu_k^{(t+1)})^2], \quad (7)$$

where $\mu_k^{(t+1)} = \mu^{(t+1)} + (2 - k)\,a^{(t+1)} + (-1)^k d^{(t+1)}/2$. The maximum-likelihood estimate of p is $(n_1 + n_2/2)/n$.

The E and M steps are iterated until a convergence criterion is satisfied. The converged values are the maximum-likelihood estimates of the parameters. The test for the null hypothesis, H_0, $D = 0$ *vs.* the alternative hypothesis, H_1, $D \neq 0$ is performed through a likelihood ratio,

$$\mathrm{LR} = 2[L(\hat{\Omega}, Y) - L(\hat{\Omega}, Y)|_{D=0}], \quad (8)$$

where $L(\hat{\Omega}, Y)$ is the maximum log-likelihood under the full model (the alternative hypothesis) and $L(\hat{\Omega}, Y)|_{d=0}$ is that under the reduced model with D constrained to zero (the null hypothesis).

Analysis of variance under an unbalanced nested design: The likelihood analysis has an advantage in that it provides not only a test for the linkage disequilibrium, D, but also direct estimates of the genetic parameters such as D, q, a, and d. There are also many other methods that have been used for testing the linkage disequilibrium, notably the analysis of variance and regression analysis on a marker (Martin *et al.* 1997; Excoffier and Slatkin 1998; Fulker *et al.* 1999; George *et al.* 1999; Service *et al.* 1999). Here we compare the statistical power of the likelihood analysis with ANOVA and regression analysis in testing the linkage disequilibrium.

In ANOVA based on a marker, we compare the variances between and within the marker classes. For that, we consider the following nested statistical model and define the phenotypic value of individual k that has marker genotype i and the underlying QTL genotype j as

$$y_{ijk} = \mu + \beta_i + \omega_{ij} + \varepsilon_{ijk}, \quad (9)$$

where $i = 1, 2, 3$ refers to the marker genotypes *MM*, *Mm*, and *mm*, and $j = 1, 2, 3$ to the QTL genotypes *AA*, *Aa*, and *aa*, respectively. Individuals in a sample of size n are classified into three subgroups according to their marker genotypes. In the above model, μ is the population mean, β_i is the QTL effect associated with marker genotype i, ω_{ijk} is the residual QTL effect in the nested model, and ε_{ij} is the environmental effect that is assumed to be normal distributed with mean zero and variance v.

Expressed in terms of parameters in model 1, it has been shown (Luo 1998) that

$$\beta_1 = \frac{2D[pa - (D - p + 2pq)\,d]}{p^2}$$

$$\beta_2 = \frac{D\{(1 - 2p)\,a + [2D + (1 - 2p)(1 - 2q)]\,d\}}{p(1 - p)}$$

$$\beta_3 = -\frac{2D\{(1 - p)\,a + [D + (1 - p)(1 - 2q)]\,d\}}{(1 - p)^2}. \quad (10)$$

The values of ω_{ij} (i, $j = 1, 2, 3$) have been given in Luo (1998). Taking into account that $h_{ij}n$ is the expected value of n_{ij} in an unbalanced nested model (Searle 1987, p. 111ff), the expected mean square (EMS) between the marker genotypes is

$$\mathrm{EMS}_\beta = \frac{1}{2}\left\{\sum_{i=1}^{3} h_i n\beta_i^2 - \left[\sum_{i=1}^{3} h_i[1 + (n-1)h_i]\beta_i^2 + 2(n-1)\sum_{i<j\leq 3} h_i h_j \beta_i\beta_j\right] + \sum_{i=1}^{3}\frac{1}{h_i}\left[\sum_{j=1}^{3} h_{ij}[1 + (n-1)h_{ij}]\omega_{ij}^2 + 2(n-1)\sum_{j<k\leq 3} h_{ij}h_{ik}\omega_{ij}\omega_{ik}\right] - \left[\sum_{i=1}^{3}\sum_{j=1}^{3} h_{ij}[1 + (n-1)h_{ij}]\omega_{ij}^2 + 2(n-1)\sum_{j<k\leq 3}\sum_{j<k\leq 3} h_{ij}h_{ik}\omega_{ij}\omega_{kl}\right]\right\} + v \quad (11)$$

and the expected mean square within the marker genotypes is

$$\mathrm{EMS}_\omega = \frac{1}{n-3}\left\{\sum_{i=1}^{3}\sum_{j=1}^{3} h_{ij}\omega_{ij}^3 - \sum_{i=1}^{3}\frac{1}{h_i} \times \left[\sum_{j=1}^{3} h_{ij}[1 + (n-1)h_{ij}]\omega_{ij}^2 + 2(n-1)\sum_{j<k\leq 3} h_{ij}h_{ik}\omega_{ij}\omega_{ik}\right]\right\} + v. \quad (12)$$

The method can be found in Searle (1987) for numerical calculation of the mean squares in the ANOVA under an unbalanced nested design. The ANOVA test statistic for the between-marker variance component is expected to follow a central F distribution under the null hypothesis $D = 0$. The power of the statistical test at a significance level α is then

$$\eta_F = \Pr\{F_{v_1,v_2}(\lambda_F) > F_{\alpha;v_1,v_2}\}, \quad (13)$$

where $F_{\alpha;v1,v2}$ is the $(1 - \alpha)$ 100th percentile of the central F statistic with v_1 and v_2 d.f. In this case, $v_1 = 2$ and

$v_2 = n - 3$. $F_{v1,v2}(\lambda_F)$ is a noncentral F statistic with the noncentral parameter, λ_F, that can be calculated (Johnson and Kotz 1970, p. 189) as

$$\lambda_F = \left[\frac{\text{EMS}_\beta}{\text{EMS}_\omega}\right] \frac{v_1(v_2 - 1)}{v_2} - v_1. \quad (14)$$

Regression analysis: If we regress the trait value of individual k, y_k, against the number of marker alleles M, x_k, we have the regression model

$$y_k = \mu + bx_k + e_k \quad (15)$$

with the regression coefficient

$$b = \frac{D[a + (1 - 2q)d]}{p(1 - p)}. \quad (16)$$

A test for b is essentially a test for D. The power of such a test is provided by

$$\eta_t = \Pr\{t_\nu(\lambda_t) > t_{\alpha/2;\nu}\} + \Pr\{t_\nu(\lambda_t) < t_{1-\alpha/2;\nu}\}, \quad (17)$$

where $t_{\alpha/2;\nu}$ and $t_{1-\alpha/2;\nu}$ are the upper and lower $\alpha/2$ points of a central t-statistic with $\nu = n - 2$ d.f. $t_\nu(\lambda_t)$ is a noncentral t-statistic with the noncentral parameter

$$\lambda_t = \frac{\Gamma[\nu/2]b}{\sqrt{\nu/2}\Gamma[(\nu - 1)/2]\sigma_b} \quad (18)$$

(Johnson and Kotz 1970, p. 201), where $\Gamma()$ is a gamma function and $\sigma_b^2 = [(1 - r^2)\sigma_y^2]/n\sigma_x^2$ is the sampling variance of the regression coefficient, $r = 2D[a + (1 - 2q)d]/\sigma_x\sigma_y$ is the correlation coefficient between y and x, and $\sigma_y^2 = 2q(1 - q)[a^2 + 2(1 - q)ad + (1 - 2q + 2q^2)d^2] + v$ and $\sigma_x^2 = 2p(1 - p)$ are the phenotypic and additive marker variances.

SIMULATION STUDY AND NUMERICAL ANALYSES

Simulation model: The strategy has been described elsewhere (Luo 1998) for simulating the linkage disequilibrium between a polymorphic marker locus and a QTL in natural populations. The simulation programs allow simulation parameters, n, p, q, D, μ, a, d, and v to be varied readily. For the purpose of comparing the three different approaches presented here, 13 sets of parameters were considered in the simulation study and are listed in Table 2. The number of simulation replicates varied from 200 to 500 depending on the purpose of a specific analysis.

Behavior of the test statistics under the null hypothesis: We first compare the theoretical and simulated distributions of the test statistics under the null hypothesis for the determination of critical values. Under the null hypothesis, the test statistic of the unbalanced nested model for ANOVA is F distributed with mean $v_2/(v_2 - 2)$ and variance $2v_2^2(v_1 + v_2 - 2)/[v_1(v_2 - 2)^2(v_2 - 4)]$ (Johnson and Kotz 1970, p. 78). The test statistic of the regression model has a mean deviation given by $\sqrt{v}\Gamma[(v - 1)/2]/[\sqrt{\pi}\Gamma(v/2)]$ and variance given by $v/(v - 2)$ (Johnson and Kotz 1970, p. 96). The percentiles can be obtained from the corresponding F or t distributions.

It is not clear what the distribution of the likelihood-ratio test statistic is under the null hypothesis. Many likelihood-ratio test statistics under a null hypothesis might be asymptotically χ^2 distributed. However, our simulation clearly shows that this, in the present context, may not be the case.

Table 3 illustrates the expected and simulated values of mean, variance, and 95th percentile of the test statistics under the null hypothesis, $D = 0$, for the 13 configurations of simulation parameters listed in Table 2. For the likelihood-ratio test, we list here the mean, variance, and 95th percentile of the χ^2 distribution with 1 d.f. for comparison, which are 1, 2, and 3.84, respectively. The simulated values were obtained on the basis of 500 simulation runs. The table shows that the simulated values for the F- and t-tests were in good agreement with their corresponding theoretical expectations. However, the simulated values of mean, variance, and 95th percentile of the likelihood-ratio test statistic differ significantly from the χ^2 distribution values, particularly when the sample size was small. Because of this, we suggest using the permutation test, which is regularly used for mapping QTL in planned experiments (Churchill and Doerge 1994), to empirically estimate the threshold from the data.

Power calculation for the statistical tests: Power of the statistical tests was evaluated both theoretically and numerically. Theoretical predictions of the power for the F and t tests were evaluated according to Equations 13 and 17. The numerical evaluation was based on the frequency of the significant tests in 200 simulation trials. For the likelihood-ratio test, the critical value used for the power evaluation was obtained from 300 permutations at the null hypothesis.

Table 2 gives the theoretically predicted and numerically observed power of the unbalanced nested model (η_{FT} and η_{FS}) and the regression model (η_{RT} and η_{RS}), as well as the empirical power for the likelihood-ratio test (η_{LR}), for detecting the linkage disequilibrium between a polymorphic marker locus and a QTL whose segregation explains 10 or 20% phenotypic variation of the trait. The power was evaluated at the significant level $\alpha = 5\%$. In the table, n_F and n_R represent sample size required for 80% power at the same significant level under the unbalanced nested model and the regression model, respectively.

Table 2 shows a good agreement between the theoretical prediction and the simulated observation of the power for the first two models. Comparison of the three approaches suggests that the regression model is probably the preferred test as far as the power is concerned,

TABLE 2

Parameter values for the 13 populations considered in the numerical analyses

Population	n	p	q	D	μ	a	d	v	β_{FT}	β_{FS}	β_{RT}	β_{RS}	β_{LR}	n_F	n_R
1	200	0.5	0.5	0.10	0.0	0.47	0.00	1.0	0.33	0.25	0.43	0.37	0.30	586	475
2	200	0.5	0.5	0.10	0.0	0.71	0.00	1.0	0.62	0.61	0.72	0.71	0.64	291	235
3	500	0.5	0.5	0.10	0.0	0.47	0.00	1.0	0.71	0.69	0.81	0.79	0.65	586	495
4	500	0.5	0.5	0.00	0.0	0.71	0.00	1.0	0.05	0.06	0.05	0.04	0.05	—	—
5	500	0.5	0.5	0.20	0.0	0.47	0.00	1.0	1.00	1.00	1.00	1.00	1.00	184	170
6	500	0.5	0.5	0.10	0.0	0.44	0.22	1.0	0.67	0.65	0.76	0.78	0.65	645	534
7	500	0.5	0.5	0.10	0.0	0.38	0.38	1.0	0.56	0.54	0.64	0.65	0.55	814	714
8	500	0.3	0.3	0.09	0.0	0.51	0.00	1.0	0.86	0.85	0.92	0.94	0.82	410	334
9	500	0.7	0.7	0.09	0.0	0.51	0.00	1.0	0.87	0.85	0.92	0.93	0.82	410	334
10	500	0.3	0.5	0.10	0.0	0.47	0.00	1.0	0.70	0.67	0.87	0.80	0.67	492	398
11	500	0.3	0.5	0.10	0.0	0.38	0.38	1.0	0.80	0.65	0.87	0.74	0.64	491	398
12	500	0.5	0.3	0.10	0.0	0.51	0.00	1.0	0.64	0.72	0.76	0.82	0.76	678	599
13	500	0.5	0.3	0.10	0.0	0.33	0.33	1.0	0.72	0.70	0.83	0.80	0.71	575	484

n, the sample size; p and q, the frequencies of alleles M and A; D, the coefficient of linkage disequilibrium between the marker locus and QTL; a and d, the additive and dominant effects of QTL; μ and v, the mean and residual variances for the trait; β_{FT} and β_{FS}, the theoretically predicted and empirically observed powers for ANOVA; β_{RT} and β_{RS}, the theoretically predicted and empirically observed powers for the regression analysis; β_{LR}, the empirically observed power for the likelihood ratio test; n_F and n_R, the sample sizes required to obtain a statistical power of 80% at $\alpha = 0.05$ for ANOVA and regression analysis.

and the remaining two tests are almost equally efficient in detecting the disequilibrium under the circumstances considered here. From a statistical point of view, the test statistic in the ANOVA essentially tests the significance of the correlation ratio of a continuous quantitative variate (z_k in the present context) on a discrete variate (x_k in the present study), whereas the test statistic in the regression analysis virtually tests linearity of the regression of z_k on x_k. It is generally known (*e.g.*, KENDALL and STUART 1961, pp. 296–300) that the regression test has higher power than the correlation ratio test (*i.e.*, the test of the ANOVA). Less efficiency of the likelihood-ratio test is hard to explain but may be partially because the EM algorithm is not guaranteed to converge to the global maxima of the likelihood function (MENG and VAN DYKE 1997). The likelihood-analysis method was

TABLE 3

Theoretical and simulated distribution characteristics of the three test statistics under the null hypothesis ($D = 0$) for the 13 populations considered

	F-test			*t*-test			Likelihood ratio		
Population	$E(F)$	Var(F)	95%	$E[\|t\|]$	Var(t)	95%	Mean	Variance	95%
Expectation	1.00	1.04	3.04	0.80	1.01	1.97	1.00[a]	2.00[a]	3.84[a]
1	0.97	1.06	2.85	0.75	1.10	2.01	2.05	4.36	6.06
2	0.91	0.94	2.74	0.76	1.01	2.00	1.72	3.10	5.09
Expectation	1.00	1.02	3.01	0.80	1.00	1.96	1.00[a]	2.00[a]	3.84[a]
3	1.00	0.89	2.86	0.78	1.14	1.95	1.57	4.57	5.72
4	1.10	1.11	3.25	0.84	1.01	1.75	1.77	4.77	6.18
5	1.08	1.19	3.22	0.76	1.13	2.01	1.64	7.12	6.74
6	0.97	1.18	2.82	0.79	1.03	1.85	2.00	6.13	6.84
7	0.98	1.03	3.28	0.80	1.10	1.90	2.07	4.99	6.32
8	1.03	1.01	3.29	0.76	1.09	1.79	1.44	4.80	4.62
9	0.94	0.97	2.77	0.73	1.02	1.76	1.60	4.09	5.91
10	0.99	0.95	2.66	0.88	1.04	1.98	1.92	4.08	5.60
11	1.05	1.10	2.94	0.84	1.12	1.89	2.11	5.88	6.81
12	1.08	1.15	2.99	0.85	0.94	1.87	1.73	4.95	6.67
13	1.00	1.05	3.00	0.85	0.96	2.01	2.08	5.88	6.74

For the likelihood ratio, the distribution values of χ^2_1 are given for comparison. $E(F)$ and Var(F), the mean and variance of the *F*-test statistic; $E[|t|]$ and Var(t), the mean deviation and variance of the *t*-test statistic; 95%, the 95th percentile of a distribution.

[a] χ^2_1 distribution values.

also observed to be less powerful than analysis of variance in detecting linkage between marker and QTL in Le Roy and Nelen (1995). When the marker and QTL are in linkage equilibrium (population 4), all three tests show a significance frequency that approximately equals the given significant level α, indicating a well-controlled type I error for the tests. In addition to the major influence due to the sample size and the level of the disequilibrium on the power, the dominance effect of QTL also has some effect on the power, decreasing the power (populations 3, 6, and 7). The effect of allelic frequencies of the marker and QTL on the power is difficult to interpret because they are compounded with other parameters, such as the disequilibrium and additive effect of QTL. However, Table 2 clearly shows that for a wide spectrum of inheritance models, there is a very good chance of detecting a QTL, explaining 10% of phenotypic variation with a sample size of a few hundred.

Relative efficiency of the transmission disequilibrium test: Allison (1997) extended the TDT, originally developed by Spileman *et al.* (1993), to QTL analysis and demonstrated that the linkage disequilibrium between a polymorphic marker locus and a QTL can be detected by using data of family trios consisting of parents, at least one of them being heterozygous at the marker locus, and their child. The analysis of Allison (1997) is, however, basically a marker analysis and was not developed to estimate the disequilibrium. Here we compare the statistical power of the TDT procedure with the unbalanced nested model and the regression model discussed in this article. For the purpose of comparison, we simulate the data on the basis of the models defined in this article with equal frequency for the marker and QTL alleles (*i.e.*, $p = q$) and the maximum disequilibrium for the given allelic frequency, that is, $p(1 - p)$ or $q(1 - q)$ (Weir 1995).

Table 4 gives the theoretical prediction of sample size requirement by the three approaches with 80% power at $\alpha = 0.0001$ and the observed power from 500 simulation runs using the predicted sample size. It must be noted that the sample size required for the TDT analysis is the number of the trio families, but for the ANOVA and regression analyses the size is the number of individuals. It can be seen that the power of TDT decreases as the allelic frequency increases, whereas the power of the ANOVA and regression analyses is basically unaffected by the allelic frequency given that the QTL effect is additive ($f = 0.0$). At a low frequency (*e.g.*, $p = q = 0.1$), the TDT approach requires a smaller number of trios than the number of individuals required by the other two approaches for achieving the same power. However, when the allelic frequency is high (*e.g.*, $p = q = 0.5$), the ANOVA and regression analyses require a sample size that is about half of the number of trio families needed by the TDT. The effect of QTL dominance on the power for TDT depends on the allelic frequency. At a low frequency (*e.g.*, 0.1 or 0.3), an increase in QTL dominance (comparing $f = 1.0$ with $f = 0.0$) tends to slightly increase the power. However, at a high frequency (*e.g.*, 0.5), dominance decreases the power significantly, thus increasing the sample size requirement. In a sharp contrast, the power of the ANOVA model is independent of both allelic frequency and dominance. However, for the regression analysis, the QTL dominance decreases the power, particularly with a high allelic frequency. The results generally show that a sample of 500 individuals from a natural population might be sufficient to detect a QTL explaining 5% of the phenotypic variance with 80% power at $\alpha = 0.0001$ significance level by using the ANOVA and regression analyses. When the marker locus is the trait locus itself as assumed here, the variance between the marker genotypes is equivalent to the proportion (say, h^2) of the trait phenotypic variance explained by the QTL. Thus, the test power of the ANOVA is determined by h^2 and independent of the gene frequency and the dominance level at the trait locus. Moreover, it can be seen from Equation 18 that with $d = 0$ (*i.e.*, absence of dominance at the QTL) and $\mathrm{Dip}(1 - p) = q(1 - q)$, the noncentral parameter λ_t can be simplified into a form of $\{\Gamma[v/2]/\sqrt{v/2}\Gamma[(v - 1)/2]\}\sqrt{nh^2/(1 - h^2)}$. This indicates that the power of the regression analysis is entirely determined by h^2 and independent of allelic frequency when there is no dominance at the QTL.

Table 4 also shows the observed power of the ANOVA and regression analyses by using the predicted sample size for simulation. A consistent agreement between the simulated observations and the expected value (80%) confirms the reliability of the power prediction for the two analyses in the extreme case of complete linkage disequilibrium considered here.

Maximum-likelihood estimates of the model parameters: Table 5 shows the observed mean and standard error for the maximum-likelihood estimates of the genetic parameters from 100 replicates of simulation. The parameter values are given in Table 2.

All parameters except the residual variance seem to be well estimated by maximum likelihood when the sample size is 500 and the QTL variance is at least 10% of the phenotypic variance. The residual variance is generally underestimated, particularly when the sample size or QTL variance is small. However, when the sample size and QTL variance are large, the bias tends to be minor. The bias may be explained by slow convergence in estimating the variance parameter and by a low stability in the derivative of the likelihood function with regard to this parameter as pointed out in Meng and van Dyke (1997). Moreover, it should be noted that the standard errors represented here were calculated on the basis of repeated simulation trials. In practice, the standard deviation of the maximum-likelihood estimates may be obtained by bootstrapping, as proposed

TABLE 4

Theoretically predicted sample size required for 80% power at $\alpha = 0.0001$ and empirical power observed using the predicted sample size

	Sample size			Empirical power	
Parameters	TDT	ANOVA	Regression	ANOVA	Regression
$h^2 = 0.05$ $f = 0.0$					
$p = q = 0.10$	308	485	429	0.754	0.778
$p = q = 0.30$	727	485	429	0.796	0.770
$p = q = 0.50$	873	485	429	0.802	0.782
$h^2 = 0.10$ $f = 0.0$					
$p = q = 0.10$	147	214	208	0.728	0.800
$p = q = 0.30$	351	214	208	0.778	0.780
$p = q = 0.50$	426	214	208	0.780	0.796
$h^2 = 0.15$ $f = 0.0$					
$p = q = 0.10$	93	156	134	0.744	0.768
$p = q = 0.30$	225	156	134	0.822	0.802
$p = q = 0.50$	277	156	134	0.840	0.774
$h^2 = 0.05$ $f = 1.0$					
$p = q = 0.10$	269	485	443	0.800	0.796
$p = q = 0.30$	604	485	511	0.796	0.790
$p = q = 0.50$	1,321	485	634	0.796	0.778
$h^2 = 0.10$ $f = 1.0$					
$p = q = 0.10$	128	214	221	0.826	0.784
$p = q = 0.30$	291	214	256	0.836	0.814
$p = q = 0.50$	649	214	319	0.782	0.778
$h^2 = 0.15$ $f = 1.0$					
$p = q = 0.10$	81	156	143	0.838	0.780
$p = q = 0.30$	186	156	166	0.804	0.782
$p = q = 0.50$	426	156	208	0.778	0.794

The comparison was made between ANOVA, regression, and the transmission disequilibrium test by Allison (1997).

for mapping QTL in planned experiments (Visscher *et al.* 1996).

DISCUSSION

Positional cloning efforts have been very successful in isolating mutant genes for many rare monogenic diseases. The success of gene cloning depends critically on the mapping of a gene in high resolution. Recent theoretical and experimental studies have shown that linkage disequilibrium analysis between DNA molecular polymorphism and disease susceptibility from genetically isolated populations can provide an effective way to map disease susceptibility genes into a <1-cM genome interval in the best cases (see de la Chapelle and Wright 1998 for a review on the topic). Given the high-throughput identification and genotyping of polymorphisms, whole-genome linkage disequilibrium studies have recently been proposed as a powerful approach for detecting many subtle genetic effects that underlie susceptibility to common diseases (Kruglyak 1999). The basic idea behind linkage disequilibrium analysis for gene mapping has been demonstrated in Kaplan and Weir (1997). With the assumption that the disease mutation occurred many years ago and, perhaps through a founder effect, propagated in the population, the marker alleles on the original mutant haplotype may still be in linkage disequilibrium with the mutant allele in the current population, and the level of the disequilibrium is likely to be proportional to the number of generations since the mutation occurred and to the genetic distance between the marker and QTL alleles. Given that a significant disequilibrium is detected in the current population, a longer generation number may imply a closer linkage between the marker and mutant alleles.

However, most genetic diseases, such as idiopathic epilepsy, essential hypertension, atopic diathesis, and coronary artery disease, are quantitative traits. Successful extrapolation of the positional cloning methodology to genes affecting complex traits depends critically on careful study design and application of sophisticated analytical and computational tools (Ghosh and Collins 1996). Our study addresses statistical issues on detecting and estimating linkage disequilibrium between a polymorphic genetic marker and a trait locus from natural populations. The results show that, with a sample size of a few hundred, the statistical power to detect the disequilibrium seems to be reasonably high for a wide parameter region. We also presented a likelihood

TABLE 5

Mean and standard error (in parentheses) of the maximum-likelihood estimates of the genetic parameters for the 13 populations of Table 2 from 100 simulation replicates

Population	$\hat{p}$	$\hat{q}$	$\hat{D}$	$\hat{\mu}$	$\hat{a}$	$\hat{d}$	$\hat{v}$
1	0.4946	0.5055	0.1003	−0.1081	0.5042	−0.0074	0.8381
	(0.0027)	(0.0118)	(0.0039)	(0.0193)	(0.0319)	(0.0699)	(0.0169)
2	0.5022	0.4828	0.1085	0.0210	0.7040	0.0585	0.8488
	(0.0022)	(0.0097)	(0.0043)	(0.0160)	(0.0314)	(0.0622)	(0.0221)
3	0.5017	0.4896	0.1083	0.0245	0.4485	0.0486	0.9288
	(0.0017)	(0.0106)	(0.0039)	(0.0109)	(0.0219)	(0.0455)	(0.0137)
4	0.4985	0.4995	−0.0004	0.0088	0.7367	0.0152	0.9311
	(0.0015)	(0.0076)	(0.0043)	(0.0104)	(0.0239)	(0.0449)	(0.0150)
5	0.4986	0.4927	0.1973	0.0144	0.5080	0.0186	0.9441
	(0.0015)	(0.0061)	(0.0037)	(0.0100)	(0.0172)	(0.0200)	(0.0135)
6	0.5001	0.5053	0.0992	−0.0240	0.4851	0.2502	0.9047
	(0.0041)	(0.0122)	(0.0037)	(0.0149)	(0.0295)	(0.0445)	(0.0136)
7	0.4991	0.5188	0.1045	−0.0143	0.4263	0.4115	0.9255
	(0.0015)	(0.0110)	(0.0037)	(0.0116)	(0.0211)	(0.0425)	(0.0127)
8	0.3000	0.3245	0.1030	−0.0195	0.4989	0.0956	0.9363
	(0.0015)	(0.0104)	(0.0032)	(0.0146)	(0.0217)	(0.0453)	(0.0121)
9	0.6963	0.6957	0.1022	0.0017	0.5094	−0.0676	0.9453
	(0.0016)	(0.0085)	(0.0034)	(0.0137)	(0.0242)	(0.0418)	(0.0119)
10	0.2996	0.4783	0.0976	0.0247	0.4627	0.0604	0.9210
	(0.0015)	(0.0098)	(0.0028)	(0.0145)	(0.0228)	(0.0489)	(0.0132)
11	0.3000	0.4616	0.0976	0.0166	0.3878	0.4207	0.9226
	(0.0015)	(0.0102)	(0.0025)	(0.0104)	(0.0234)	(0.0385)	(0.0143)
12	0.4997	0.3504	0.0992	−0.0587	0.5063	0.0198	0.9215
	(0.0015)	(0.0103)	(0.0031)	(0.0143)	(0.0234)	(0.0457)	(0.0149)
13	0.4991	0.3537	0.0981	−0.0370	0.3670	0.3372	0.9162
	(0.0017)	(0.0104)	(0.0032)	(0.0125)	(0.0276)	(0.0487)	(0.0145)

$\hat{p}$ and $\hat{q}$, the estimates of the allelic frequency of *M* and *A*; $\hat{D}$, the linkage disequilibrium; $\hat{\mu}$, the population mean; $\hat{a}$ and $\hat{d}$, the additive and dominance effects of QTL; $\hat{v}$, the residual variance.

analysis to directly estimate the QTL parameters, such as the frequency and effects of QTL alleles and the linkage disequilibrium between the marker and QTL. These estimates may be useful in interpreting the demographic history of the natural populations under question (Thompson and Neel 1997) and, in turn, in inferring the mapping information of the trait locus on the basis of the disequilibrium analysis (Devlin *et al.* 1996).

Allison (1997) extended the TDT method to detect association between the marker locus and a trait locus using family data. Allison's method is designed to detect linkage between a marker locus in the presence of linkage disequilibrium from nuclear families. We also compared the statistical power in detecting the trait locus between our methods and the TDT method. It must be pointed out that the sample size presented in Table 4 for the TDT analysis is the number of trio families, whereas for the ANOVA and regression analyses it is the number of individuals. The results show that, given the genetic variance explained by a trait locus, the power of TDT depends on the trait allele frequency, whereas the power of ANOVA and regression is relatively independent from the allelic frequency. The TDT method is more powerful when the trait allele frequency is low, but much less powerful when it is high. It is worthwhile to note that samples of individuals under the present study are assumed genetically unrelated. If some individuals in the sample are genetically related, the power of the methods could be reduced.

There have been many studies on detecting and/or estimating linkage disequilibrium between marker and trait loci using either nuclear families (Spileman *et al.* 1993; Terwilliger 1995; Martin *et al.* 1997; Excoffier and Slatkin 1998) or samples from natural populations (Hill and Robertson 1968; Hill 1974; Chakraborty and Weiss 1988; Hastbacka *et al.* 1992; Xiong and Guo 1997). Many of these studies directly analyze marker effects through designed experiments or choosing appropriate family data for analysis. In this article, we outlined several methods for testing the linkage disequilibrium between a polymorphic marker and a trait locus from a natural population. More importantly, we presented a likelihood analysis to directly estimate the trait locus parameters and the linkage disequilibrium with the marker. This study gives some insight into the feasibility of extending the principle of linkage disequilibrium-based mapping to complex disease traits.

The major difficulty in linkage disequilibrium-based

mapping is to quantify the relationship between recombination fraction and linkage disequilibrium measure. Since recombinant events are not observed, the recombination fraction between the marker and trait locus must be estimated on the basis of a population genetic model. Several methods have been suggested to address this problem. One of these has attempted to search for the reparameterization by which the disequilibrium measure can be directly related to the recombination fraction. For instance, Devlin and Risch (1995) found that, in the present notations, the measure of the disequilibrium

$$\delta = \frac{D}{q[(1-p)(1-q) + D]}$$

has some interesting properties. In certain situations, the disequilibrium measure is related to the recombination fraction r as $\delta = (1-r)^T$, where T represents the generation number since the creation of the initial disequilibrium and may be estimated either from an epidemiological survey (*i.e.*, in Hastbacka *et al.* 1992) or directly from the sampled data (*i.e.*, Kaplan *et al.* 1995; Thompson and Neel 1996). Kaplan and Weir (1997) proposed a simulation-based approach, which allows the maximum-likelihood estimate of the recombination fraction and its confidence interval to be estimated entirely on the basis of the observation of linkage disequilibrium between a polymorphic marker locus and a simple monogenic disease locus. These analyses were confined to the circumstances where the genotypes at the trait locus can be observed. However, the basic idea may be extended to the case where the genotypes at the trait locus are not observed as considered in our study. In fact, the information can be, at least partially, uncovered for the joint distribution of genotypes at both the marker and trait loci using the maximum-likelihood estimates of the parameters p, q, and D. It can be anticipated that the predicted recombination fraction will bear a larger sampling variation since q and D have to be estimated from data with incomplete information.

The present theoretical analysis is developed on the basis of the assumption of two alleles at the trait locus and the marker locus. Multiple alleles at marker loci may be common in natural populations. In principle, it is straightforward to extend the ANOVA and regression methods to analyze the multiple-allele markers. In fact, the methods compare the between-marker genotype variation to the within-marker genotype variation. With multiple marker alleles, calculation of the test statistics for these two methods will remain the same algorithm. In theory, the power of these analyses will be improved as the number of marker alleles increases because larger degrees of freedom are expected under the multiple marker allele model. Care must be taken when the likelihood method is extended to analyze multiple allelic marker data. With multiple alleles, we suggest assigning the most frequent marker allele as the allele M considered in our model on the basis of the intuition that the higher frequent allele tends to have a higher chance to be in linkage disequilibrium with its tightly linked loci.

The model discussed in our study is still too simplistic for complex traits that may be affected by multiple genes. We recognize that there are several multilocus linkage disequilibrium mapping methods. Some of them are based on comparison of pairwise linkage disequilibria between the single disease locus and a set of marker polymorphisms and use the peak value of disequilibrium measured over several marker loci as evidence for location of the hypothesized QTL (*e.g.*, Terwilliger 1995; Rannala and Slatkin 1998; Slatkin 1999). In addition, some theoretical effort has been made to combine the pairwise disequilibrium between the disease locus and each of a set of marker loci by use of the composite likelihood principle (Devlin *et al.* 1996; Collins and Morton 1998). Our study provides statistical inference of linkage disequilibrium between a single marker locus and a locus affecting complex genetic variation; hence, the multiple-locus analysis could be built upon the two-locus model. However, more work is needed to extend the method to appropriately take multiple marker information into account and to analyze multiple trait loci. Such an extension would certainly need to take into account the complex structure of the disequilibria among multiple loci (Weir 1979) or at least a major part of it. With the availability of many densely distributed molecular markers, such as SNP, the opportunity is there to uncover the genetic architecture of complex traits in natural populations.

We are indebted to two anonymous reviewers and Dr. Y. X. Fu for their constructive comments and criticisms that have been helpful in improving presentation and clarifying several ambiguities in an earlier version of this article. Z.W.L. is grateful for discussion with Dr. X. L. Meng on the EM algorithm. Z.W.L. was supported by China's Basic Research Program "973," the National Science Foundation, the Qiu-Shi Foundation, and the Changjiang Scholarship; and Z-B.Z. by U.S. Public Health Service grant GM-45344.

LITERATURE CITED

Allison, D. B., 1997 Transmission disequilibrium tests for quantitative traits. Am. J. Hum. Genet. **60:** 676–690.

Allison, D. B., M. Heo, N. Kaplan and E. R. Martin, 1999 Sibling based tests of linkage and association for quantitative traits. Am. J. Hum. Genet. **64:** 1754–1764.

Chakraborty, R., and K. M. Weiss, 1988 Admixture a tool for finding linked genes and detecting that difference from allelic association between loci. Proc. Natl. Acad. Sci. USA **85:** 9119–9123.

Churchill, G. A., and R. W. Doerge, 1994 Empirical threshold for quantitative trait mapping. Genetics **128:** 963–971.

Collins, A., and N. E. Morton, 1998 Mapping a disease locus by allelic association. Proc. Natl. Acad. Sci. USA **95:** 1741–1745.

de la Chapelle, A., and F. A. Wright, 1998 Linkage disequilibrium mapping in isolated populations: the example of Finland revisited. Proc. Natl. Acad. Sci. USA **95:** 12416–12423.

Devlin, B., and N. Risch, 1995 A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics **29:** 311–322.

Devlin, B., N. Risch and K. Roeder, 1996 Disequilibrium mapping:

composite likelihood for pairwise disequilibrium. Genomics **36:** 1–16.

Eisenberg, I., C. Thiel, T. Levi, A. Sanilevich, M. Shemesh *et al.*, 1999 Fine-structure mapping of the hereditary inclusion body myopathy locus. Genomics **55:** 43–48.

Excoffier, L., and M. Slatkin, 1998 Incorporating genotypes of relatives into a test of linkage disequilibrium. Am. J. Hum. Genet. **62:** 171–180.

Fulker, D. W., S. S. Cherny and L. R. Cardon, 1995 Multipoint interval mapping of quantitative trait loci, using sib pairs. Am. J. Hum. Genet. **56:** 1224–1233.

Fulker, D. W., S. S. Cherny, P. C. Sham and J. K. Hewitt, 1999 Combined linkage and association sib-pair analysis for quantitative traits. Am. J. Hum. Genet. **64:** 259–267.

George, V., H. K. Tiwari, X. Zhu and R. C. Elston, 1999 A test of transmission/disequilibrium for quantitative trait loci in pedigree data, by multiple regression. Am. J. Hum. Genet. **65:** 236–245.

Ghosh, S., and F. S. Collins, 1996 The geneticist's approach to complex disease. Annu. Rev. Med. **47:** 333–353.

Harley, H. G., J. D. Brook, J. Floyd, S. Crow, M. C. Thibault *et al.*, 1991 Detection of linkage disequilibrium between the myotonic dystrophy locus and a new polymorphic DNA marker. Am. J. Hum. Genet. **49:** 68–75.

Hastbacka, J., A. de la Chapelle, I. Kaitil, P. Sistonen, A. Weaver *et al.*, 1992 Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nat. Genet. **2:** 204–211.

Hill, W. G., 1974 Estimation of linkage disequilibrium in randomly mating populations. Heredity **33:** 229–239.

Hill, W. G., and A. Robertson, 1968 Linkage disequilibrium in finite populations. Theor. Appl. Genet. **38:** 226–231.

Hill, W. G., and B. S. Weir, 1994 Maximum likelihood estimation of gene location by linkage disequilibrium. Am. J. Hum. Genet. **54:** 705–714.

Johnson, N. L., and N. Kotz, 1970 *Distributions in Statistics: Continuous Univariate Distributions.* Houghton Mifflin, Boston.

Jorde, L. B., 1995 Linkage disequilibrium as a gene mapping tool. Am. J. Hum. Genet. **56:** 11–14.

Kaplan, N. L., and B. S. Weir, 1997 The use of linkage disequilibrium for estimating the recombination fraction between a marker and a disease gene, pp. 207–219 in *Progress in Population and Human Evolution*, edited by P. I. Donnelly and S. Tavare. Springer-Verlag, New York.

Kaplan, N., W. G. Hill and B. S. Weir, 1995 Likelihood methods for locating disease genes in nonequilibrium populations. Am. J. Hum. Genet. **56:** 18–32.

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

Kruglyak, L., 1999 Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat. Genet. **22:** 139–144.

Laan, M., and S. Paabo, 1998 Mapping genes by drift generated linkage disequilibrium. Am. J. Hum. Genet. **63:** 654–656.

Lander, E. S., and N. J. Schork, 1994 Genetic dissection of complex traits. Science **265:** 2037–2048.

Le Roy, R., and J. M. Nelen, 1995 Numerical comparison between powers of maximum likelihood and analysis of variance methods for QTL detection in progeny test designs: the case of monogenic inheritance. Theor. Appl. Genet. **90:** 65–72.

Luo, Z. W., 1998 Detecting linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations. Heredity **80:** 198–208.

Luo, Z. W., and S. Suhai, 1999 Estimating linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations. Genetics **151:** 359–371.

Luo, Z. W., R. Thompson and J. A. Woolliams, 1997 A population genetics model of marker-assisted selection. Genetics **146:** 1173–1183.

Martin, E. R., N. L. Kaplan and B. S. Weir, 1997 Tests for linkage association in nuclear families. Am. J. Hum. Genet. **61:** 439–448.

Meng, X. L., and D. van Dyke, 1997 The EM algorithm—an old-song sung to a fast new tune. J. R. Stat. Soc. B **59:** 511–540.

Neimann-Sorenson, A., and A. Robertson, 1961 The association between blood groups and several production characters in three Danish cattle breeds. Acta Agric. Scand. **11:** 163–196.

Nielsen, D. M., and B. S. Weir, 1999 A classical setting for associations between markers and loci affecting quantitative traits. Genet. Res. **74:** 271–277.

Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, 1992 *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press, New York.

Rannala, B., and M. Slatkin, 1998 Likelihood analysis of disequilibrium mapping and related problems. Am. J. Hum. Genet. **62:** 459–473.

Risch, N., and H. Zhang, 1996 Mapping quantitative trait loci with extreme discordant sib pairs: sampling consideration. Am. J. Hum. Genet. **58:** 819–830.

Searle, S. R., 1987 *Linear Models for Unbalanced Data.* John Wiley and Sons, New York.

Service, S. K., D. W. Lang, N. B. Freimer and L. A. Sandkuijl, 1999 Linkage disequilibrium mapping of disease genes by reconstruction of ancestral haplotypes in founder population. Am. J. Hum. Genet. **64:** 1728–1738.

Slatkin, M., 1999 Disequilibrium mapping of a quantitative trait locus in an expanding population. Am. J. Hum. Genet. **64:** 1765–1773.

Spileman, R. S., R. E. McGinnis and W. J. Ewens, 1993 Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am. J. Hum. Genet. **52:** 506–516.

Terwilliger, J. D., 1995 A powerful likelihood method for the analysis of linkage disequilibrium between trait locus and one or more polymorphic marker loci. Am. J. Hum. Genet. **49:** 31–41.

Thompson, E. A., and J. V. Neel, 1997 Allelic disequilibrium and allele frequency distribution as a function of social and demographic history. Am. J. Hum. Genet. **60:** 197–204.

Visscher, P. M., R. Thompson and C. S. Haley, 1996 Confidence intervals in QTL mapping by bootstrapping. Genetics **143:** 1013–1020.

Weir, B. S., 1979 Inferences about linkage disequilibrium. Biometrics **35:** 235–254.

Weir, B. S., 1995 *Genetic Data Analysis II.* Sinauer, Sunderland, MA.

Weir, B. S., and C. C. Cockerham, 1977 Two locus theory in quantitative genetics, pp. 247–269 in *Proceedings of the International Conference on Quantitative Genetics*, edited by E. Pollak, O. Kempthorne and T. B. Bailey. Iowa State University Press, Ames, IA.

Xiong, M., and S. W. Guo, 1997 Fine-scale genetic mapping based on linkage disequilibrium: theory and application. Am. J. Hum. Genet. **60:** 1513–1531.

Communicating editor: Y.-X. Fu