*Genet. Res., Camb.* (1999), **74**, *pp.* 271–277. Printed in the United Kingdom  

# A classical setting for associations between markers and loci affecting quantitative traits

DAHLIA M. NIELSEN* AND B. S. WEIR
*Program in Statistical Genetics, Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA*

(*Received 24 May 1999 and in revised form 2 August 1999*)

**Summary**

We examine the relationships between a genetic marker and a locus affecting a quantitative trait by decomposing the genetic effects of the marker locus into additive and dominance effects under a classical genetic model. We discuss the structure of the associations between the marker and the trait locus, paying attention to non-random union of gametes, multiple alleles at the marker and trait loci, and non-additivity of allelic effects at the trait locus. We consider that this greater-than-usual level of generality leads to additional insights, in a way reminiscent of Cockerham's decomposition of genetic variance into five terms: three terms in addition to the usual additive and dominance terms. Using our framework, we examine several common tests of association between a marker and a trait.

## 1. Introduction

Since the discovery of numerous polymorphic markers spread across the genomes of many species, linkage analysis has been highly successful in localizing regions of chromosomes containing genes affecting many traits of interest (Paterson *et al*., 1988; Georges *et al*., 1995; Comuzzie *et al*., 1997). Often, however, these chromosomal regions are very large, spanning many millions of bases and containing many genes. In order to narrow these regions to areas which are more amenable to molecular characterization, there has been an increasing amount of interest in fine-mapping techniques. These methods capitalize on evolutionary history and population genetics to capture the relationships between markers in very close proximity to genes affecting the trait of interest. The parameters of interest in these studies include the linkage disequilibria between the gene and a marker, which it is hoped can give an indication of degree of proximity between the two loci. The general theory behind this method is that markers which are very tightly linked to a gene should show high association with the trait, reflecting linkage disequilibria between the alleles at the marker and those at the gene affecting the trait. Markers which are less closely linked to the gene will have lost much of their association with the trait due to recombination over time.

In the analysis of dichotomous traits, two basic types of study designs are often used in fine-mapping experiments to detect association between a marker and a gene. The first is a case–control design, in which individuals are collected for both categories of the trait (i.e. 'affected' and 'unaffected') and then genotyped for the genetic markers. If marker allele frequencies among the two groups differ significantly, then it is concluded that the marker is associated with a gene affecting the trait; there is non-zero linkage disequilibrium between the gene and the marker. This type of analysis does not control for population dynamics such as admixture or selection, so association does not necessarily imply linkage.

The second study design which has been proposed for dichotomous traits is the transmission/disequilibrium design (Spielman *et al*., 1993; Kaplan *et al*., 1997). The methods utilize random population samples of small nuclear families, and test for non-equal segregation of marker alleles from heterozygous parents to affected offspring. If a marker is linked to a gene affecting the trait, then marker alleles which are in association with the alleles of the gene should be preferentially transmitted to affected offspring. A significant result for this test is evidence of both linkage and association, suggesting a more precise indication of location of the gene.

* Corresponding author.

Several extensions to the transmission/disequilibrium design have been proposed for fine-mapping of quantitative trait loci (QTL). Allison (1997) suggested five designs, each considering a different scheme for sampling based on phenotype data. Each of these designs utilizes data collected for trios of two parents with one offspring, where at least one of the parents is heterozygous for the marker locus being examined. Markers are assumed to be biallelic. Under the null hypothesis, the mean values of offspring within the three marker genotype classes are equal. If, however, the marker is associated with a gene affecting the trait, and the recombination rate between the two loci is less than 0·5, then the null hypothesis will not be true and the three means will not be equal. If a sampling scheme based on offspring phenotype is chosen, then under the alternative hypothesis, unequal transmission of marker alleles to offspring can also be used as an indication of association in the presence of linkage. Rabinowitz (1997) proposed a similar design to test for linkage in the presence of association, but allowed for larger nuclear families to be collected. In the case of families with a single offspring, his design reduces to the randomly sampled design of Allison (1997). Martin (1997) extended the work of Allison (1997) by allowing for an arbitrary number of marker alleles, all considered simultaneously. This test is based on the difference between the mean phenotypic value among offspring of parents who transmit a particular marker allele and the mean phenotypic value among offspring of parents who do not transmit that allele.

All these tests examine alleles individually, rather than as genotypes, though it is genotypes, rather than individual alleles, which generally affect phenotype. Here, we propose a classical genetics framework whereby the consequences of examining genetic data based on observations at the alleles alone can be determined. A key feature of our approach is that we do not restrict attention to biallelic loci. It is unlikely that genes affecting complex traits have only two alleles, and the restriction to biallelic markers seems unnecessary even though the currently used SNP markers seldom have more than two alleles.

## 2. Methods

We consider a trait locus with an arbitrary number of alleles which contributes to the genetic component of the quantitative trait of interest. Alleles at this locus are designated $A_r$, with population frequencies $p_r$. The genetic effect of genotype $A_r A_s$ on the trait, $G_{rs}$, can be described by the classical linear model

$$G_{rs} = \mu + \alpha_r + \alpha_s + d_{rs}, \tag{1}$$

where $\mu$ is the genotypic mean, $\alpha_r$ and $\alpha_s$ are the additive effects of the alleles, and $d_{rs}$ is the deviation from additivity. For a random mating reference population, the least squares solutions for the parameters are described in Weir & Cockerham (1977):

$$\mu = \sum_{r,s} p_r p_s G_{rs} = G\,.\,., \tag{2}$$

$$\alpha_r = \sum_{s} p_s G_{rs} - \mu = G_{r.} - G\,.\,., \tag{3}$$

$$\alpha_s = \sum_{r} p_r G_{rs} - \mu = G_{.s} - G\,.\,., \tag{4}$$

$$d_{rs} = G_{rs} - \alpha_r - \alpha_s - \mu = G_{rs} - G_{r.} - G_{.s} + G\,.\,. \tag{5}$$

These solutions embody the constraints $\Sigma_r p_r \alpha_r = \Sigma_r p_r d_{rs} = 0$.

The additive and dominance effects are generally regarded as being random, and the genetic variance contributed by the trait locus in a random mating population is

$$\sigma_G^2 = \sum_{r,s} p_r p_s (G_{rs} - \mu)^2 = \sigma_A^2 + \sigma_D^2,$$

where the additive and dominance variance components are

$$\sigma_A^2 = 2 \sum_{r} p_r (\alpha_r)^2, \tag{6}$$

$$\sigma_D^2 = \sum_{r,s} p_r p_s (d_{rs})^2. \tag{7}$$

Since the genotypes at the trait locus are generally not observable, we also consider a marker locus with an arbitrary number of alleles, designated $M_i$ and having population frequencies $q_i$. We are interested in determining the relationship between the genotypes at the marker locus and the gene affecting the trait. For this, we consider the same type of linear model as in (1), but now defined in terms of the marker genotypic classes, $M_i M_j$, and their genetic effects $G_{ij}^{(m)}$:

$$G_{ij}^{(m)} = \mu^{(m)} + \alpha_i^{(m)} + \alpha_j^{(m)} + d_{ij}^{(m)}. \tag{8}$$

Here $\alpha_i^{(m)}$ and $\alpha_j^{(m)}$ are the additive effects of marker alleles $M_i$ and $M_j$ on the trait, and $d_{ij}^{(m)}$ is the dominance deviation. Using the solutions of (2)–(5), we see that

$$\mu^{(m)} = \sum_{i,j} q_i q_j G_{ij}^{(m)} = \mu, \tag{9}$$

$$\alpha_i^{(m)} = \sum_{j} q_j G_{ij}^{(m)} - \mu^{(m)}, \tag{10}$$

$$d_{ij}^{(m)} = G_{ij}^{(m)} - \mu^{(m)} - \alpha_i^{(m)} - \alpha_j^{(m)}. \tag{11}$$

Each marker genotypic class $M_i M_j$ is composed of a mixture of elements from all the trait classes, $A_r A_s$, and the proportion of class $A_r A_s$ contained within class $M_i M_j$ is $\Pr(A_r A_s \mid M_i M_j)$:

$$G_{ij}^{(m)} = \sum_{r,s} \Pr(A_r A_s \mid M_i M_j)\, G_{rs}$$

$$= \sum_{r,s} \frac{P_{sj}^{ri}}{\Pr(M_i M_j)} G_{rs}.$$

https://doi.org/10.1017/S0016672399004231 Published online by Cambridge University Press

Here $P_{sj}^{ri}$ is the frequency of $A_r M_i / A_s M_s$ genotypes. With random union of gametes, and writing $D_{ri}$ for the linkage disequilibrium between marker allele $M_i$ and allele $A_r$ at the gene, this becomes

$$G_{ij}^{(m)} = \sum_{r,s} \frac{(p_r q_i + D_{ri})(p_s q_j + D_{sj})}{q_i q_j} G_{rs}$$

$$= \mu + \frac{1}{q_j}\delta_j + \frac{1}{q_i}\delta_i + \frac{1}{q_i q_j}\delta_{ij},$$

where $\delta_i = \Sigma_r \Sigma_s p_s G_{rs} D_{ri} = \Sigma_r \alpha_r D_{ri}$, and $\delta_{ij} = \Sigma_{r,s} D_{ri} D_{si} G_{rs} = \Sigma_{r,s} D_{ri} D_{si} d_{rs}$, since $\Sigma_i D_{ri} = 0$. Using this result and applying (10) and (11), it follows that

$$\alpha_i^{(m)} = \frac{1}{q_i}\delta_i = \frac{1}{q_i}\sum_r \alpha_r D_{ri}, \tag{12}$$

$$d_{ij}^{(m)} = \frac{1}{q_1 q_j}\delta_{ij} = \frac{1}{q_i q_j}\sum_{r,s} d_{rs} D_{ri} D_{si}. \tag{13}$$

This shows that the additive effect of a marker allele $M_i$ is the weighted sum of the additive effects of the alleles at the trait locus, where the weights are the linkage disequilibria between that marker allele and the alleles at the trait locus. Similarly, the dominance deviation of a marker genotype is the weighted sum of the dominance deviations of the genotypes at the trait locus.

It is important to distinguish the association measure $\delta_i$ between marker allele $M_i$ and a trait from the linkage disequilibrium measure $D_{ri}$ between marker allele $M_i$ and trait allele $A_r$. In the special case of the trait locus having only two alleles, $A_1$, $A_2$, we have $D_{1i} + D_{22i} = 0$ and the association parameter is proportional to linkage disequilibrium. More generally, non-zero association implies non-zero linkage disequilibrium but zero association does not imply zero linkage disequilibrium.

## 3. Tests for association

These relationships between marker and trait locus effects suggest several approaches to testing for association between the marker and a gene affecting the trait.

### (i) *Dichotomous trait case–control tests*

For a dichotomous trait, the genetic values $G_{rs}$ may be regarded as susceptibilities, or the probabilities that $A_r A_s$ individuals are affected (Nielsen *et al.*, 1998). These values are quantitative in nature, having a continuous distribution and relying on both genetic and environmental influences. The mean value $\mu$ is the probability of a random individual being affected, i.e. the population prevalence $\phi$ of the disease. The frequency of marker allele $M_i$ among affected individuals is found by taking the sum over all (unobserved) trait genotypes. Under the assumption of random union of gametes:

$$q_{i|\text{Aff.}} = \sum_j \sum_{r,s} \Pr(A_r M_i / A_s M_j | \text{Aff.})$$

$$= \sum_j \sum_{r,s} (p_r q_i + D_{ri})(p_s q_j + D_{sj}) G_{rs}/\phi$$

$$= q_i + \delta_i/\phi.$$

Among unaffected individuals the marker allele frequency is

$$q_{i|\text{Unaff.}} = q_i - \delta_i/(1-\phi),$$

suggesting that association, $\delta_i$, can be detected by comparing these two frequencies:

$$q_{i|\text{Aff.}} - q_{i|\text{Unaff.}} = \frac{\delta_i}{\phi(1-\phi)}.$$

A goodness-of-fit test statistic (Kaplan *et al.*, 1997) uses sample marker allele frequencies $\tilde{q}_i$ among affecteds and unaffecteds. For samples of $n$ affected and $n$ unaffected individuals:

$$X_c^2 = n \sum_i \frac{(\tilde{q}_{i|\text{Aff}} - \tilde{q}_{i|\text{Unaff.}})^2}{(\tilde{q}_{i|\text{Aff.}} + \tilde{q}_{i|\text{Unaff.}})}.$$

The allelic case–control test is therefore a test for additive effects at the trait locus, mediated by linkage disequilibrium between the trait and marker loci. Power to detect association, therefore, depends on both non-zero additive effects and non-zero linkage disequilibria, and that these terms do not cancel in the summary measure, $\delta_i$.

An alternative would be to compare marker genotype frequencies among affecteds and unaffecteds. With random union of gametes:

$$\Pr(M_i M_j | \text{Aff.} = \sum_{r,s} \Pr(A_r M_i / A_s M_j | \text{Aff.})$$

$$= \sum_{r,s} (p_r q_i + D_{ri})(p_s q_j + D_{si}) G_{rs}/\phi$$

$$= q_i q_j + (\delta_i + \delta_j)/\phi + \delta_{ij}/\phi,$$

$$\Pr(M_i M_j | \text{Unaff.}) = q_i q_j - (\delta_j + \delta_j)/(1-\phi)$$

$$- \delta_{ij}/(1-\phi).$$

The difference is

$$\Pr(M_i M_j | \text{Aff.}) - \Pr(M_i M_j | \text{Unaff.}) = \frac{\delta_i + \delta_j + \delta_{ij}}{\phi(1-\phi)}$$

and this contrast leads to a joint test of both additive and dominance components for the trait locus, mediated by linkage disequilibria.

https://doi.org/10.1017/S0016672399004231 Published online by Cambridge University Press

(ii) *Quantitative trait case–control tests*

For the continuous trait values of primary interest here, the viewpoint is reversed: conditioning is on marker type instead of on disease status (trait value). The simplest procedure is to compare trait means among individuals distinguished by their marker types. Suppose there are $n_{ij}$ individuals with marker genotype $M_i M_j$. The trait value for the $k$th of these individuals is

$$Y_{ijk} = G_{ij}^{(m)} + \epsilon_{ijk}, \quad k = 1, 2, \ldots, n_{ij}, \tag{14}$$

where $\epsilon_{ijk}$ is an error term. We assume the errors are independent of both marker and trait genotypes, and are distributed with a mean of zero and variance $\sigma_e^2$.

The mean squares between and within marker genotype classes have expected values of

$$\mathscr{E}(\mathrm{MSB}) = \sigma_e^2 + \frac{1}{m-1}\sum_{i,j} n_{ij}(G_{ij}^{(m)} - \overline{G^{(m)}})^2,$$

$$\mathscr{E}(\mathrm{MSW}) = \sigma_e^2,$$

where $m$ is the number of distinct marker genotypes in the data, and

$$\overline{G^{(m)}} = \sum_{i,j} n_{ij} G_{ij}^{(m)} / \sum_{i,j} n_{ij}.$$

For normally distributed trait values, the $F$-test will provide a test for association between trait and marker loci, but will not distinguish between additive and dominance effects at the trait locus.

Of course, it would be possible to find least-squares estimates of the effects $\alpha_i^{(m)}$ and $d_{ij}^{(m)}$ and form the summary statistics

$$\hat{\sigma}^2_{A^{(m)}} = 2\sum_i \hat{q}_i(\widehat{\alpha_i^{(m)}})^2,$$

$$\hat{\sigma}^2_{D^{(m)}} = \sum_{i,j} \hat{q}_i \hat{q}_j(\widehat{d_{ij}^{(m)}})^2,$$

and construct tests that the corresponding parameters were zero. Non-zero values of $\sigma^2_{A^{(m)}}$ and $\sigma^2_{D^{(m)}}$ imply non-zero additive and dominance variances for the trait, although the converse does not apply: zero values of $\sigma^2_{A^{(m)}}$ and $\sigma^2_{D^{(m)}}$ do not imply zero additive and dominance variances for the trait.

(iii) *Dichotomous trait transmission/disequilibrium tests*

A finding of association between a marker and the trait does not imply genetic linkage between marker and trait loci. In order to test for linkage, Spielman *et al.* (1993) introduced the TDT based on the marker allele transmitted from parent to child. For the basic design of this test, trios of parents and an affected offspring are collected, and the number of times an $M_i$ allele and not an $M_j$ allele are transmitted from a parent to an affected offspring is calculated. If no assumptions regarding the mating structure of the population are made, transmissions from parents to offspring must be considered jointly, as parents are not necessarily independent. The probability of a parent transmitting an $M_i$ and not an $M_j$ allele to an affected offspring, averaging over all trait alleles $A_t$ transmitted by the other parent is $T_{ij}$, where

$$\begin{aligned} T_{ij} &= \frac{1}{\phi}\sum_{r,t}[(1-c)\Pr(A_r M_i/M_j, A_t) \\ &\quad + c\Pr(A_r M_j/M_i, A_t)]G_{rt} \\ &= \frac{1}{\phi}\sum_{r,t}[(1-c)P^{ir}_{j|t} + cP^{jr}_{i|t}]G_{rt}, \end{aligned}$$

where $c$ is the recombination rate between the gene and the marker and $\Pr(A_r M_i/M_j, A_t) = P^{ir}_{j|t}$ is the joint probability that one parent has haplotype $A_r M_i$ and has $M_j$ as the other marker allele and the other parent carries the $A_t$ allele (Weir *et al.*, 1990). In this design, conditioning is based on the affection status of the offspring (only trios with affected offspring are chosen) instead of on the marker genotypes of the parents, which are not known in advance. The probability of a parent transmitting an $M_i$ and not an $M_j$ allele is equivalent to the probability of an $M_i M_j$ parent transmitting an $M_i$ allele.

The difference between the transmission probabilities for $M_i$ and $M_j$ from $M_i M_j$ parents to affected offspring is

$$T_{ij} - T_{ji} = \frac{(1-2c)}{\phi}\sum_{r,t}(P^{ir}_{j|t} - P^{jr}_{i|t})G_{rt},$$

which suggests a procedure for testing for linkage ($c \neq 0{\cdot}5$) in the presence of association, i.e. $P^{ir}_{j|t} \neq P^{jr}_{i|t}$. Ewens & Spielman (1995) noted that the TDT does not detect association in a structured population for which there is no linkage disequilibrium within each subpopulation and there has not been more than one generation of mating between subpopulations. This is an example where $P^{ir}_{j|t} = P^{jr}_{i|t}$ among parents in the whole population.

When there is random union of gametes and no higher-order disequilibria,

$$T_{ij} = q_i q_j + \frac{1-c}{\phi} q_j \delta_i + \frac{c}{\phi} q_i \delta_j$$

and

$$T_{ij} - T_{ji} = \frac{(1-2c)}{\phi}(q_j \delta_i - q_i \delta_j),$$

suggesting a test for association. In general, however, the difference depends also on the associations among

https://doi.org/10.1017/S0016672399004231 Published online by Cambridge University Press

all subsets of the four alleles $M_i, M_j, A_r, A_t$ (Cockerham & Weir, 1973; Weir, 1996), although disequilibria among non-gametic allele pairs such as $M_j, A_r$ and $M_j, A_t$ are likely to be small when there is Hardy–Weinberg equilibrium.

It is customary to concentrate only on the transmitted marker allele. The transmission probabilities $T_{ij}$ are summed over the non-transmitted allele and the contrast between transmitted and non-transmitted probabilities for a particular allele is

$$T_{\cdot i} - T_{i\cdot} = \frac{(1-2c)}{\phi}(\delta_i^* - \delta_i^{*c}),$$

where

$$\delta_i^* = \sum_{r,t}(D_{ri|t} + p_t D_{ri})\, G_{rt}$$

and

$$\delta_i^{*c} = \sum_{r,t}(D_{r/i|t} + p_t D_{r/i})\, G_{rt}$$

are necessary when there is not random union of gametes. With random union of gametes, the three-allele disequilibria $D_{ri|t}$, $D_{r/i|t}$ and the non-gametic disequilibrium $D_{r/i}$ are zero so that $T_{i\cdot} - T_{\cdot i} = (1-2c)\,\delta_i/\phi$.

Although rejection of the hypothesis $H_o: T_{i\cdot} = T_{\cdot i}$ implies that $(\delta_i^* - \delta_i^{*c}) \neq 0$, and therefore that there is linkage disequilibrium between trait and marker loci and/or Hardy–Weinberg disequilibrium in the population from which the $M_i M_j$ parent is drawn, the converse does not hold. There may be linkage and/or Hardy–Weinberg disequilibrium but little association. The various disequilibrium terms $D$ add to zero over any subscript, and weighting them by terms such as $\alpha_r$ or $G_{rt}$ can also give sums close to zero, especially for loci with little effect on the trait. With random union of gametes and no non-gametic disequilibria, this TDT is addressing the additive components of the trait-locus effects. Otherwise dominance at the trait locus also contributes to the term $(T_{i\cdot} - T_{\cdot i})$.

If there are $n_{ij}$ parent-affected offspring pairs where $M_i$ is transmitted and $M_j$ is not transmitted, then a test for equality of marginal totals for the $\{n_{ij}\}$ contingency table has test statistic

$$X^2_m = \sum_i \frac{(n_{i\cdot} - n_{\cdot i})^2}{n_{i\cdot} + n_{\cdot i}},$$

where $n_{i\cdot} = \Sigma_j n_{ij}, n_{\cdot i} = \Sigma_j n_{ji}$. It is more usual not to include transmissions from $M_i M_i$ homozygous parents in this calculation (Spielman & Ewens, 1996). The test statistic is modified to

$$X^2_{\text{mhet}} = \frac{m-1}{m}\sum_{i=1}^{m} \frac{(n_{i\cdot} - n_{\cdot i})^2}{n_{i\cdot} + n_{\cdot i} - 2n_{ii}}.$$

(iv) *Quantitative trait transmission/disequilibrium tests*

For a quantitative trait, in place of the marker allele transmission probabilities $T_{ij}$ to affected offspring, we consider the trait value $H_{ij}^{(m)}$ for an offspring that receives $M_i$ and not $M_j$ from parent $M_i M_j$. From the same argument as above, when there is random union of gametes:

$$\begin{aligned} H_{ij}^{(m)} &= \frac{1}{q_i q_j}\sum_{r,t}[(1-c)\Pr(A_r M_i/M_j, A_t) \\ &\quad + c\Pr(A_r M_j/M_i, A_t)]\, G_{rt} \\ &= \mu + (1-c)\frac{\delta_i}{q_i} + c\frac{\delta_j}{q_j}. \end{aligned}$$

The contrast between expected trait values of individuals receiving $M_i$ versus $M_j$ from an $M_i M_j$ parent is

$$H_{ij}^{(m)} - H_{ji}^{m} = (1-2c)\left(\frac{\delta_i}{q_i} - \frac{\delta_j}{q_j}\right),$$

suggesting a test statistic for the null hypothesis of no linkage or no association.

Allison (1997) and Martin (1997) worked with marginal expected trait values $H_{i\cdot}^{(m)}$ for offspring that received marker allele $M_i$ from a parent that carried that allele. The other parental marker allele does not need to be specified. Summing over $j$:

$$\begin{aligned} H_{i\cdot}^{(m)} &= \sum_j q_j H_{ij}^{(m)} \\ &= \mu + (1-c)\frac{\delta_i}{q_i}, \end{aligned}$$

so that the expected difference in trait values for individuals that either do or do not receive marker allele $M_i$ from a parent is

$$H_{i\cdot}^{(m)} - H_{\cdot i}^{(m)} = (1-2c)\frac{\delta_i}{q_i}. \tag{15}$$

The trait difference between marginals is expected to be zero if the recombination rate is 0·5 or if all the $\delta_i$ are zero. Since $\delta_i$ depends only on the additive effects of the trait, dominance at the trait locus does not affect these measures.

A statistical test for differences of expected trait values, $H_{i\cdot}^{(m)} - H_{\cdot i}^{(m)}$, is based on the set of observed phenotypes of offspring, $\mathbf{Y}$. A possible statistic to measure differences in marginal values is $\bar{Y}_{i\cdot} - \bar{Y}_{\cdot i}$, where $\bar{Y}_{i\cdot}$ is the mean phenotypic value of all offspring who have received an $M_i$ allele. When there is random mating in the population, the expected value of this contrast is the expression in (15). To calculate the variance of this contrast, both parents' transmissions

https://doi.org/10.1017/S0016672399004231 Published online by Cambridge University Press

to an offspring must be considered jointly. We find that the variances are functions of both the additive and dominance terms. This means that an expected value of zero for the contrasts does not imply that the test will behave as expected under the null hypothesis that all linkage disequilibria, $D_{ri}$, are equal to zero.

## 4. Discussion

We have proposed a classical linear model for a quantitative trait in terms of observable marker genotypes and we have shown that there is a simple relationship between the marker being examined and an associated locus which affects the trait of interest. Additive effects for the marker alleles are functions of the additive effects of the trait locus and the disequilibria existing between the marker and the trait locus. Dominance effects of the marker genotypes are functions of the dominance effects of the trait locus and the disequilibria between the loci. For the simplified case in which random mating within the population is assumed, the relationships between the marker and the trait locus involve only linkage disequilibria. For the general model, making no assumptions about random mating, higher-order disequilibria are involved.

Although our extension from biallelic to multiallelic loci is not profound, it does allow insights that may otherwise be lost. We found that it allowed a very natural distinction between linkage disequilibrium for specific pairs of trait and marker alleles, and association between a marker allele and a trait. The latter quantity, which is addressed by the data, depends on the magnitude of trait-locus effects and on marker allele frequencies in addition to linkage disequilibrium. It has been noted previously that there may be loss of insight in quantitative genetic theory when attention is restricted to two alleles per locus (Weir & Cockerham, 1977; Cockerham, 1983).

Since this genetic model can offer insight into the degree of association between the marker and the trait, a straightforward test of association could be performed using an analysis of variance. One method is to compare the expected mean squares between and within genotype classes. A significant *F*-test indicates the presence of association between the marker and the trait locus, but does not distinguish between additive and dominance effects.

Another possibility is to use the ANOVA estimates of additive terms of the marker. Estimates of additive effects which are significantly different from zero imply both non-zero additive effects at the trait locus and an association between the loci. Significantly non-zero dominance estimates at the marker imply non-zero dominance terms at the trait locus and associations between loci. This test, in a sense, provides a case–control type test for quantitative traits, indicating possible allele or genotype associations between the marker and the trait locus. The principle behind this method is similar to that of Luo (1998) and Luo & Suhai (1999); however, our model is the classical genetic model defined on marker genotypes rather than a combination of marker and trait genotypes.

We examined several common tests of association in light of our results for additive and dominance estimators at the marker locus, including the case–control test and the TDT for dichotomous and quantitative traits. For the tests based on allelic rather than genotypic associations, we find that the primary focus is additive genetic components of the trait locus, as expected. While the expected values of the statistics used in these tests are zero when the additive effects at the trait locus are zero or when linkage disequilibria are zero, the variances of the statistics are functions of both additive and dominance terms in conjunction with non-zero disequilibria. This implies that tests based on these statistics may have an increased or decreased variance due to dominance, even when the expected values of the statistics are zero. Thus, the null hypothesis being tested must be formulated correctly if the test is to have the proper size under the null hypothesis. Hypothesizing that all $D_{ri} = 0$ provides the proper null distribution. Conditions such as small additive effects at the trait locus may substantially reduce the power of these tests in spite of the existence of strong disequilibria and overall large genetic effects.

We have limited the present discussion to single-marker analyses. The search for loci affecting quantitative traits was greatly accelerated by the introduction of theory for marker-locus intervals (reviewed by Doerge *et al*., 1997), and the joint consideration of two or more marker loci may well offer an advantage in the present context. However, with the increasing density of markers on many genomes, there may still be a need to consider each marker in turn, provided proper attention is paid to multiple testing issues.

We thank Dr Stephanie Monks for very helpful discussions. This work was supported in part by NIH grant GM45344.

## References

Allison, D. B. (1997). Transmission-disequilibrium tests for quantitative traits. *American Journal of Human Genetics* **60**, 676–690.

Cockerham, C. C. (1983). Covariances of relatives from self-fertilization. *Crop Science* **23**, 1177–1180.

Cockerham, C. C. & Weir, B. S. (1973). Descent measures for two loci with some applications. *Theoretical Population Biology* **4**, 300–330.

Comuzzie, A. G., Hixson, J. E., Almasy, L., Mitchell, B. D., Mahaney, M. C., Dyer T. D., Stern, M. P., MacCluer, J. W. & Blangero, J. (1997). A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2. *Nature Genetics* **15**, 273–276.

Doerge, R. W., Zeng, Z. B. & Weir, B. S. (1997). Statistical issues in the search for genes affecting quantitative traits in experimental populations. *Statistical Science* **12**, 195–219.

Ewens, W. J. & Spielman, R. S. (1995). The transmission/disequilibrium test: history, subdivision, and admixture. *American Journal of Human Genetics* **57**, 455–464.

Falconer, D. S. & MacKay, T. F. C. (1996). *Introduction to Quantitative Genetics*, 4th edn. Harlow, Essex: Longman.

Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A. T., Sargeant, L. S., Sorensen, A., Steele, M. R., Zhao, X., Womack, J. E. & Hoeschele, I. (1995). Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* **139**, 907–920.

Kaplan, N. L., Martin, E. R. & Weir, B. S. (1997). Power studies for the transmission/disequilibrium tests with multiple alleles. *American Journal of Human Genetics* **60**, 691–702.

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. & Suhai, S. (1999). Estimating linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations. *Genetics* **151**, 359–371.

Martin, E. R. (1997). Extensions of the transmission/disequilibrium test for identifying human disease genes. PhD dissertation, North Carolina State University, Raleigh, NC.

Nielsen, D. M., Ehm, M. G. & Weir, B. S. (1998). Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *American Journal of Human Genetics* **63**, 1531–1540.

Paterson, A. H., Lander, E. S., Hewitt, J. D., Peterson, S., Lincoln, S. E. & Tanksley, S. D. (1988). Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* **335**, 721–726.

Rabinowitz, D. (1997). A transmission disequilibrium test for quantitative trait loci. *Human Heredity* **47**, 342–350.

Spielman, R. S. & Ewens, W. J. (1996). The TDT and other family-based test for linkage disequilibrium and association. *American Journal of Human Genetics* **59**, 983–989.

Spielman, R. S., McGinnis, R. E. & Ewens, W. J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (DDM). *American Journal of Human Genetics* **52**, 506–516.

Weir, B. S. (1996). *Genetic Data Analysis II*. Sunderland, MA: Sinauer.

Weir, B. S. & Cockerham, C. C. (1977). Two-locus theory in quantitative genetics. In *Proceedings of the International Conference on Quantitative Genetics* (ed. E. Pollak, O. Kempthorne & T. B. Bailey), pp. 247–269. Ames, Iowa: Iowa State University Press.

Weir, B. S., Reynolds, J. & Dodds, K. G. (1990). The variance of sample heterozygosity. *Theoretical Population Biology* **37**, 235–253.