Heredity **72** (1994) 175–192

Received 30 June 1993

Effect of genetic architecture on the power of human linkage studies to resolve the contribution of quantitative trait loci

LINDON J. EAVES

Department of Human Genetics, Virginia Commonwealth University, P O Box 3, Richmond, VA 23298-0003, USA

The effect of genetic architecture (linkage relationships, dominance and two forms of non-allelic interaction) on the power of marker studies to detect, locate and analyse the contributions of specific quantitative trait loci (QTLs) to continuous human traits is considered for randomly mating populations in linkage equilibrium under a two-locus model. The expected regression of the within-sibling-pair mean-square on number of alleles identical by descent (IBD) at two marker loci is explored for every possible pair of markers over a region of the genome containing two QTLs linked loosely (50 CM) or more tightly (20 CM).

For the cases examined, it is shown that epistasis between the pair of QTLs reduces considerably the total amount of information available for the location and analysis of the QTL effects. The overall effects of epistasis are more marked when there are duplicate gene interactions (i.e. genes operate in parallel) than when there are complementary interactions (i.e. genes operate in series). However, when there are complementary interactions, the regression approach is almost certain to fail to detect any evidence of epistasis. The numerical analysis suggests that methods of QTL analysis based on IBD in humans are unlikely to offer the resolving power that is desirable if QTLs are to be located precisely unless inheritance is very simple or prohibitively large numbers of highly selected individuals are available.

Keywords: epistasis, linkage, power, QTLs, RFLPs, sib-pairs.

Introduction

Breeding studies in infra-human species have typically shown that sites of polygenic activity influencing continuous traits are numerous and scattered widely across the genome (see e.g. Mather & Jinks, 1982, Chaps 1, 11 and references). Until recently, the quest for specific sites of polygenic activity in humans has been frustrated by the inability of human geneticists to manipulate their material in ways comparable to those of microbial, plant and animal geneticists. Although there have been mathematical models for the contribution of loci of large effect to quantitative traits ('QTLs') for a long time (e.g. Lalouel *et al.*, 1983 and references) the opportunities for locating them have been few for lack of adequate methods for marking the human genome. The recent explosion in the technology for generating and exploiting highly polymorphic markers (e.g. Lander & Botstein, 1989) in any species, allied to the demonstration that QTLs can be located and analysed in experimental organisms (Paterson *et al.*, 1988, 1990, 1991), has lent renewed impetus to the analysis of QTLs in humans and to the identification of individual loci contributing to risk for human disease (e.g. Lander & Botstein, 1986).

Typically, the genetic models undergirding such attempts have assumed that the phenotypic effects of QTLs are relatively simple. Thus, for example, Haseman & Elston's (1972) classical regression treatment of sib-pair methods for locating sites of polygenic activity assumed that the effects of QTLs were independent, i.e. that there was no epistasis. Since then, numerous approaches have been presented for analysing linkage to QTLs in experimental organisms where it is possible to secure crosses and backcrosses between inbred lines in which the allele frequencies are equal and the degree of linkage disequilibrium fixed in the parents (e.g. Luo & Kearsey, 1989, 1991) or in large numbers of sibling pairs (e.g. Knott & Haley, 1992). Usually, these approaches have used the method of maximum-likelihood (ML). Recently Haley & Knott (1992) have examined the application of standard regression models rather than ML and showed, for the case of F_2s between a pair of marked inbred lines, that

regression methods were far quicker, easier and scarcely less informative for locating QTLs between flanking markers.

The failure of most human genetic diseases to show simple Mendelian segregation has led to a growing interest in the behaviour of two locus and multilocus models. There is considerable evidence for the role of epistatic interactions between QTLs in experimental organisms (see e.g. Mather & Jinks, 1982; McGuire, 1992) even for variables of potential clinical significance (e.g. Schlager & Chao, 1991). Breeding studies have even been able to resolve the digenic and trigenic interactions of linked and unlinked loci (e.g. Perkins & Jinks, 1970). Such studies invite caution, even skepticism, about what can be achieved with the relatively blunt instruments still available to the human geneticist. Several authors have noted the potential impact of multiple loci and non-allelic interactions on patterns of segregation and relative risk in human kinships (e.g. Risch, 1990; Neuman & Rice, 1992; Greenberg, 1993). Mather (1974) derived the expected contributions of digenic interactions to the covariances between relatives in randomly mating populations. Eaves (1988) showed how one particular form of epistasis (duplicate gene interaction) could lead to a very marked reduction in the correlation of dizygotic twins compared with monozygotic twins.

The availability of highly polymorphic loci provides, in theory at least, the opportunity to extend human biometrical genetics to allow specification of the contributions of individual loci and pairs of loci to the additive, dominance and epistatic components of variance in randomly mating populations (Mather, 1974; Mather & Jinks, 1982) and thus pave the way for a systematic exploration of the feasibility of resolving their contributions in humans. This paper examines how far certain types of biologically significant epistatic interactions may facilitate or inhibit the detection, location and analysis of the gene action underlying human continuous variation.

Materials and methods

We begin by considering the contributions of the additive, dominance and epistatic effects of two QTLs to the covariance of sibling pairs sharing none, one and two alleles identical by descent (IBD) at each QTL. Then we consider the four-locus system comprising the two QTLs and two potentially linked (fully informative) marker loci employed as part of a genome search for QTLs. We express the covariance of the sibling pairs for the quantitative trait conditional on the pairs sharing none, one and two alleles IBD at the marker loci. For a given set of markers in which the two QTLs are embedded, we then express the expected variance within-sib pairs as a function of the effects of the QTLs and the recombination fractions between the four relevant loci.

Using the method of iterative weighted least squares applied to the expected within-pair variances for a notional total sample size we conduct a regression analysis of within-pair variance on the linear and non-linear functions of the IBD values at the two marker loci. The terms of the regression yield tests for the additive and dominance effects of the two QTLs and for the four possible digenic epistatic interactions between the hypothetical QTLs. Comparison of the weighted residual sums of squares under various nested submodels yields expectations for the non-centrality parameter of the non-central chi-square distribution corresponding to salient hypotheses about the action and interaction of the QTLs. The non-centrality parameters ('chi-squares') can be compared for different hypothetical situations to yield quantitative indices of the relative amounts of information available for the detection and analysis of QTL effects under different allele frequencies, kinds of gene action and linkage relationships.

Model for gene effects

Following Robson (see Van der Veen, 1959; Mather & Jinks, 1982) we specify the additive deviations (d), heterozygous (dominance) effects (h) and the four possible digenic epistatic interactions at a pair of diallelic QTLs, A/a and B/b. The effects are defined as follows:

d_a, d_b = the additive deviations at locus A/a and B/b;

h_a, h_b = the heterozygous effects at A/a and B/b;

i_{ab} = the interaction between the additive effects at the two loci;

j_{ab} = the interaction between the additive effect at A/a and the heterozygous effect at B/b;

j_{ba} = the interaction between the additive effect at B/b and the heterozygous effect at A/a;

l_{ab} = the interaction between the heterozygous effects at the two loci.

The contributions of these genic effects to the phenotypes of all nine possible genotypes (c.f. Mather & Jinks, 1982) at the two QTLs are given in Table 1.

Contributions of genic effects to total phenotypic variation

Mather (1974; see Mather & Jinks, 1982) derived the contributions of digenic interactions to the population

Table 1 Expected phenotypic values of nine genotypes at two epistatic loci

Genotype	Expected phenotype
AA BB	$m + d_a + d_b + i_{ab}$
AA Bb	$m + d_a + h_b + j_{ab}$
AA bb	$m + d_a - d_b - i_{ab}$
Aa BB	$m + h_a + d_b + j_{ba}$
Aa Bb	$m + h_a + h_b + l_{ab}$
Aa bb	$m + h_a - d_b - j_{ba}$
aa BB	$m - d_a + d_b - i_{ab}$
aa Bb	$m - d_a + h_b - j_{ab}$
aa bb	$m - d_a - d_b + i_{ab}$

Note: see Mather & Jinks (1982).

variance and covariances between relatives in a randomly mating population in which the loci were in linkage equilibrium. We make the same assumptions throughout the subsequent treatment. Mather treated the aggregate effects of all possible pairs of loci and could not distinguish, simply by looking at the covariances between relatives, the two possible 'j-type' interactions between gene pairs. As we are interested with QTL resolution by pairs of linked markers, however, we consider only one pair of loci and keep separate the two homozygote × heterozygote interactions.

Following Mather (1974) we define the contribution of the pair of QTLs to the total genetic variance in a randomly mating population thus:

$$\sigma^2_{QTL} = \tfrac{1}{2}D_{RA} + \tfrac{1}{2}D_{RB} + \tfrac{1}{4}H_{RA} + \tfrac{1}{4}H_{RB} + \tfrac{1}{4}I_{RAB} + \tfrac{1}{8}J_{RAB} + \tfrac{1}{8}J_{RBA} + \tfrac{1}{16}L_{RAB}.$$

The upper case letters D, H etc. represent the components of variance corresponding to the additive (d), dominance (h) and epistatic effects (i, j and l) of the genotypes at the two loci. The subscripts 'A' and 'B' refer to the contributions of the diallelic loci A/a and B/b, respectively. The subscript 'R' was introduced by Mather (1949) to distinguish the components of variance in a randomly mating population (with unequal allele frequencies) from those in generations derived from crosses between pairs inbred lines in which the allele frequencies are equal. One advantage of this notation, with its slightly more awkward coefficients in expected variances and covariances, is the fact that the different parameters take the same value when allele frequencies are equal and the different kinds of gene effects are equal.

If we write u_a and u_b for the frequencies of the increasing alleles, A and B, at the two loci and $v_a = 1 - u_a$ and $v_b = 1 - u_b$ for the frequencies of the decreasing alleles, $\Pi_a = u_a v_a$ and $\Delta_a = u_a - v_a$, the components of variance are expected to be as follows in terms of the gene effects and frequencies:

$$D_{RA} = 4\Pi_a[(d_a + 2\Pi_b j_{ab} + \Delta_b i_{ab}) - \Delta_a(h_a + \Delta_b j_{ba} + 2\Pi_b l_{ab})]^2,$$

$$D_{RB} = 4\Pi_b[(d_b + 2\Pi_a j_{ba} + \Delta_a i_{ab}) - \Delta_b(h_b + \Delta_a j_{ab} + 2\Pi_a l_{ab})]^2,$$

$$H_{RA} = 16\Pi_a^2(h_a + \Delta_b j_{ba} + 2\Pi_b l_{ab})^2,$$

$$H_{RB} = 16\Pi_b^2(h_b + \Delta_a j_{ab} + 2\Pi_a l_{ab})^2,$$

$$I_{RAB} = 16\Pi_a\Pi_b(i_{ab} - \Delta_b j_{ab} - \Delta_a j_{ba} + \Delta_a\Delta_b l_{ab})^2,$$

$$J_{RAB} = 64\Pi_a\Pi_b^2(j_{ab} - \Delta_a l_{ab})^2,$$

$$J_{RBA} = 64\Pi_a^2\Pi_b(j_{ba} - \Delta_b l_{ab})^2 \text{ and}$$

$$L_{RAB} = 256\Pi_a^2\Pi_b^2 l_{ab}^2$$

(Mather, 1974; Mather & Jinks, 1982, p. 222).

Note that, in the original formulation which represents the effects of all (unmarked) loci and pairs of loci the components are summed over all loci and pairs of loci to provide a series of composite genetic variance components, D_R, H_R, I_R, J_R and L_R. It is the potential to identify the contribution of individual loci through their linkage to informative markers that allows and necessitates the separation of the additive or dominance contributions of each locus and requires that we separate the contributions of the two homozygote × heterozygote interactions. The parameters D_{RA}, etc. are convenient estimable functions of the gene effects and frequencies. It should also be noted (Mather & Jinks, 1982) that when allele frequencies are unequal, as they will generally be in natural populations, the additive and dominance components of variance at the individual loci are not pure functions of the additive and dominance effects at the contributing loci but may be inflated or reduced (depending on allele frequencies and the direction of the effects) by various epistatic interactions. Thus, the effect of dominance and epistasis at the loci is expected to affect our ability even to detect the additive genetic contribution of individual QTLs when the allele frequencies at the QTL are unequal.

Expected contributions of the QTLs to the covariance of siblings

Sibling pairs may be IBD at none, one or two alleles at each of the two QTLs. Table 2 summarizes the contributions of the additive, dominance and epistatic components of the QTL effects to the covariance of sibling pairs stratified by the IBD status at the two QTLs.

Table 2 Coefficients of expected contribution of additive, dominance and epistatic components of genetic variance to covariance of sibling pairs sharing none, one and two alleles identical by descent at two QTLs

No. of alleles IBD		Component							
Locus A/a	Locus B/b	D_{RA}	D_{RB}	H_{RA}	H_{RB}	I_{RAB}	J_{RAB}	J_{RBA}	L_{RAB}
2	2	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$
2	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	0	$\frac{1}{8}$	0	$\frac{1}{16}$	0
2	0	$\frac{1}{2}$	0	$\frac{1}{4}$	0	0	0	0	0
1	2	$\frac{1}{4}$	$\frac{1}{2}$	0	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	0	0
1	1	$\frac{1}{4}$	$\frac{1}{4}$	0	0	$\frac{1}{16}$	0	0	0
1	0	$\frac{1}{4}$	0	0	0	0	0	0	0
0	2	0	$\frac{1}{2}$	0	$\frac{1}{4}$	0	0	0	0
0	1	0	$\frac{1}{4}$	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Contributions of the QTLs to covariances of siblings conditional on IBD at linked marker loci

In practice, we do not know the IBD status at the QTLs but only at markers which are hypothesized to be linked to one or both QTLs. We thus require the expected covariance for the quantitative trait of sibling pairs who share none, one and two alleles IBD at the (linked) marker loci in terms of the additive, dominance and epistatic components of variance defined above for the effects of the two QTLs. For the most general epistatic model the exact algebraic expectations are extremely tedious and not especially helpful. However, the algorithm for the derivation is relatively straightforward and is easily implemented in computer programs for simulation and analysis.

We write a_{klm} for the coefficient of the mth genetic component of variance (D_{RA} etc) in the expectation of the sibling covariance for pairs sharing k and l alleles IBD at the first and second QTL, respectively. If $p_{kl|ij}$ is the probability that a pair of siblings sharing i and j alleles IBD at the first and second marker loci share k and l alleles IBD at the first and second QTLs. The coefficient of the mth genetic parameter in the expected covariance of siblings sharing i and j alleles IBD at the markers is thus:

$$b_{ijm} = \sum_k \sum_l p_{kl|ij} a_{klm} .$$

The conditional probability, $p_{kl|ij}$, is obtained by dividing the probability that sib pairs share k, l, i and j alleles IBD at the four loci, markers and QTLs, given their linkage relationships by the probability that the same pairs share i and j alleles IBD at the markers.

If $0 < P_{mn} < \frac{1}{2}$ is the recombination fraction between loci m and n, with no distinction being made here between markers and QTLs, and $Q_{mn} = 1 - P_{mn}$, the probabilities, $w_{j|i}$, that a sib pair shares j alleles IBD at the second locus given they share i alleles at the first are given in Table 3.

Considering the four loci, without regard to which are markers and which are QTLs, the probability that a sib pair shares i, j, k and l alleles IBD at the four loci reading, say, from left to right along the genome is:

$$p_{ijkl} = w_i w_{j|i} w_{k|j} w_{l|k} ,$$

where w_i is the probability that sibs share i alleles IBD at the (leftmost) locus. This probability is $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ for $i = 0, 1, 2$, respectively. The conditional probabilities between adjacent locus pairs are those given in Table 3, with substitution of the appropriate recombination fraction.

Similarly, the probability a pair of siblings shares m and n alleles IBD at the two marker loci is:

$$p_{mn} = w_m w_{n|m} ,$$

so the probability that a sib shares k and l alleles IBD at the pair of QTLs, which can occur at any place among the four loci under consideration, given they share i and j IBD at the markers is, for example:

$$p_{kl|ij} = p_{ikjl} / p_{ij} .$$

In this case, the markers are assumed to be the second and fourth in order on the genome and the QTLs are assumed to occupy the first and third positions. Appropriate permutation of the subscripts will yield the conditional probabilities for other sequences of markers and QTLs.

Table 3 Probability that sib pair shares, none, one, two alleles IBD at a second locus, given that the pair shares none, one or two alleles IBD at a first locus

		Second locus		
		0	1	2
First locus	0	$Q^4+2Q^2P^2+P^4$	$4(Q^3P+QP^3)$	$4Q^2P^2$
	1	$2(Q^3P+QP^3)$	$Q^4+6Q^2P^2+P^4$	$2(Q^3P+QP^3)$
	2	$4Q^2P^2$	$4(Q^3P+QP^3)$	$Q^4+2Q^2P^2+P^4$

The recombination fraction is P; Q = 1 − P.

These relatively simple formulae may be used to generate the expected covariances of sibling pairs sharing none, one and two alleles IBD at each of the marker loci considered jointly for given values of the allele frequencies and additive, dominance and epistatic effects at the QTLs, given specified linkage relationships among the markers and QTLs.

Mapping sites of QTL activity by regression analysis

We now imagine conducting a large study of sibling pairs in which we are able to mark the genome every (say) 10 CM with fully informative markers. We consider an arbitrary pair of markers and stratify the sib pairs into nine groups with respect to their IBD status at both of the two markers. The analysis comprises computation of the average variance within-sibling pairs for each of the nine marker groups and obtaining the regression of within-pair variance on IBD status at the two markers considered jointly and severally. The model is:

$$\sigma^2_{mn} = c + b_1 m + b_2 n + b_3 n^2 + b_4 m^2 + b_5 mn + b_6 mn^2 + b_7 m^2 n + b_8 m^2 n2,$$

where σ^2_{mn} is the average within pair variance for siblings sharing m alleles IBD at the first marker and n alleles IBD at the second ($m = 0, 1, 2$; $n = 0, 1, 2$). The (linear) regressions, b_1 and b_2, of σ^2_{mn} on m and n yield tests of the additive components of QTLs linked to the first and second marker, respectively (D_{RA} and D_{RB}). The (quadratic) regressions, b_3 and b_4 on m^2 and n^2 test for the dominance contributions of QTLs linked to the marker. The four regressions on products of m and n, $b_5 \ldots b_8$, test for epistatic effects. Regression on mn test for the additive × additive interaction (I_{Rab}) between these loci. Higher order regressions, on $(mn)^2$, m^2n and mn^2, test for the heterozygote × heterozygote (L_{Rab}), heterozygote × homozygote (J_{Rba}) and homozygote × heterozygote (J_{Rab}) epistatic interactions, respectively.

The regression model may be fitted to the within-pair variances by iterative weighted least squares, recovering the set of regression coefficients (variance component estimates) which minimizes the weighted residual sum of squares:

$$S^2 = \sum_m \sum_n w_{mn}(s^2_{mn} - \sigma^2_{mn})^2.$$

The weights, w_{mn}, are the expected amounts of information about the observed variances, s^2_{mn}:

$$w_{mn} = N_{mn}/2\sigma^4_{mn},$$

N_{mn} being the degree of freedom for σ^2_{mn} (Nelder & Wedderburn, 1972).

The regression model may be fitted using one of several packages for general linear modelling which allows the observed data points to be sampled from a gamma distribution (e.g. Numerical Algorithms Group, 1990).

In practice, a variety of submodels of the full two-locus epistatic model can be fitted. Comparison of the S^2 values under a variety of (nested) submodels allows us to test for genetic effects of increasing complexity. Starting with a model (the 'no-linkage' model) which includes only the constant term, c, we can add only the two linear regression terms. The reduction in the residual sum of squares provides a chi-square test, for two degrees of freedom, of the additive genetic effects of QTLs linked to the two marker loci. Addition of the regressions on m^2 and n^2 provides a joint test (also for two degrees of freedom) of the contributions of dominance at QTLs linked to the markers. Finally, further addition of the four interaction terms provides a further reduction in the residual sum of squares which provides a joint test of epistasis between QTLs linked to the two markers. In theory, the analysis can be conducted very quickly for large numbers of pairs of marker loci to generate quite rapidly a picture of the sites of additive and non-additive activity at loci affecting a given quantitative trait.

Assessing the impact of genetic architecture on QTL analysis

In practice, it is well known that things are more difficult. Early simulation studies (e.g. MacLean *et al.*, 1975) under the 'mixed model' (Morton & MacLean, 1974; Lalouel *et al.*, 1983) for detecting major locus inheritance against a background of polygenic inheritance and environmental influences suggest that chances of detecting a single locus which accounts for less than 20–30 per cent of the total variance in a quan-

titative trait are likely to be small unless a candidate locus can be identified the alleles of which have direct and distinct effects on the phenotype of interest. Typically, such power studies have assumed fairly simple genetic models. In the remaining part of this paper we explore how far the above approach can be used to analyse the impact of various more complex (and possibly more 'realistic') forms of genetic architecture on the power of any analysis of QTL activity.

We assume a 140 CM segment of the human genome is being assayed for QTL activity. The genome is assumed to be marked by 15 fully informative markers known to be located at 10 CM intervals. Somewhere along the segment lie two QTLs which explain all the variation in a given quantitative trait. This unrealistic concession to the major locus model is made because this study is primarily concerned with the effect of genetic architecture on the relative ability to resolve QTLs as a function of allele frequencies, effects and linkage relationships. It is less concerned to analyse the overall power of the QTL analysis. In every case, we assumed all the loci were in linkage equilibrium and that there was no interference over the map distances in question.

We considered two possible configurations of QTLs: loose linkage in which the QTLs are assumed to be 50 CM apart; closer, but not 'tight', linkage in which the QTLs are 20 CM apart. In both cases it is assumed that the QTLs are equidistant between the two adjacent markers. Given these two basic conditions, a number of instructive sets of conditions were considered for the genic effects at the two loci under each set of linkage relationships.

In both sets of analyses, the baseline for comparison is the additive model in which the heterozygotes are intermediate between the homozygotes ($h_a = h_b = 0$) and all the epistatic effects (i, j, k, l) are zero. The additive deviations are assumed to be unity at both QTLs. The frequencies of the increasing alleles were set at 0.9 at both QTLs to reflect what might be anticipated for traits related to fitness. Further parameter sets superimposed different kinds of non-additive gene action on this basic model. Three main conditions were considered: (1) complete dominance at both loci for increasing trait value ($h_a = h_b = 1$) but no epistasis, (2) complete dominance at both loci and duplicate gene epistasis ($h_a = h_b = 1$, $i_{ab} = j_{ab} = j_{ba} = l_{ab} = -1$), (3) complete dominance and complementary gene interaction ($h_a = h_b = 1$, $i_{ab} = j_{ab} = j_{ba} = l_{ab} = 1$). The last two possibilities are especially significant biologically from both a biochemical and evolutionary perspective (see e.g. Mather, 1966; Mather & Caligari, 1975). The 'duplicate gene' model represents the expected pattern of epistatic interactions when separate loci control parallel ('duplicate') pathways between genotype and phenotype (Figure 1a) and reflects a system with built-in redundancy so that a (recessive) mutation at one locus is buffered in its impact on the phenotype by the presence of an alternative pathway. The 'complementary gene' model is the mathematical equivalent of two pathways in sequence in which a (recessive) mutation at either locus will block the pathway to the adaptive outcome (Figure 1b). The various conditions explored in this study are summarized in Table 4.

The values of the allele frequencies and gene effects were substituted in the formulae for the genetic parameters and the total variance given above to provide numerical values for the components of variance. The components were divided by the total genetic variance to provide a common scale of unit genetic variance to allow the different models for the contribution of the QTLs to the genetic variance to be compared. In this study no additional genetic and environmental factors were incorporated as the main purpose was to explore the impact of different kinds of gene action relative to one another.

Expected sibling covariances were computed conditional on IBD status at every pairwise combination of marker loci. The expected variances within pairs are then simply the difference between the total variance (in this case scaled to unity) and the sibling covariance. The above series of regression models was then fitted by iterative weighted least-squares to the intra-pair variances. Differences between residual sums of squares were computed under various nested submodels as a basis for comparing the amounts of information about different aspects of QTL activity under the two-locus model. Because the analysis does not allow for stochastic error in the variances, the differences in residual sums of squares correspond to the

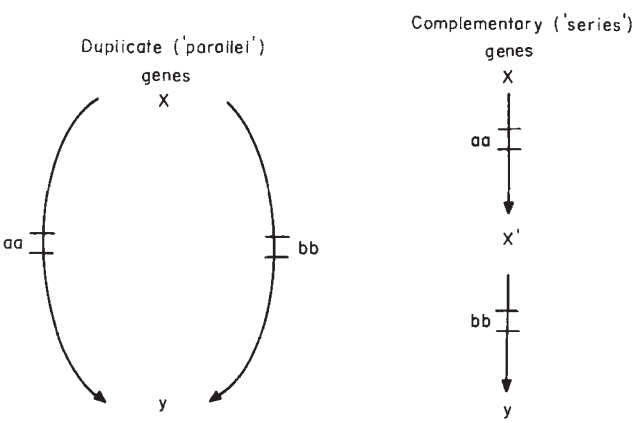

Fig. 1 Diagrammatic pathways for duplicate and complementary epistasis.

non-centrality parameters of the non-central chi-square distributions which would be obtained in testing the corresponding null hypotheses were true with a given set of population parameter values (Martin *et al.*, 1978). A total sample size of 10 000 random sibling pairs was assumed throughout to provide a convenient scale for the S^2's. These pairs were assumed to be divided among the various IBD categories in proportion to their expected frequencies in a random sample of sibling pairs. Clearly, such a random sampling strategy does not necessarily correspond to optimal practice as it is more likely that pairs will be genotyped in some systematic fashion as a function of their phenotype for the quantitative trait. For example, it may be more appropriate to begin by genotyping only pairs where the differences are extremely small. However, in these circumstances, given an optimal strategy for stratification, the resulting power is likely to approach that which can be achieved with the target random population size that has to be screened before genotyping can begin.

The computations were programmed in FORTRAN. The NAG FORTRAN library subroutine G04GDF (Numerical Algorithms Group, 1990) was used to perform the iterative weighted least squares analyses.

Results

Results are tabulated in the form of (triangular) matrices. The elements correspond to the expected regression sum of squares (non-centrality parameter) for testing particular aspects of QTL activity linked to every pairwise combination of markers. The elements of the matrix are the ordinates of the surface which characterizes the contribution of the mapped region to the genetic variation in the quantitative trait. It would be hoped that the surface shows sharp peaks corresponding to the sites of the two QTLs known to account for the genetic variation in the trait.

Table 5 summarizes the results for a population in which the two QTLs are assumed to be 50 CM apart and to show no dominance or epistasis. Given this very large sample, the non-centrality parameters for testing the additive contribution of the two loci are all comfortingly large when the markers are within striking distance of the QTLs. The model assumes that the markers are fully informative and that there are no additional QTL effects outside the region in question. For this reason, additional markers to the left of a marker itself to the left of the first QTL yield no additional information. For example, the non-centrality parameters in the row of Table 5 corresponding to marker 5 are all identical. The same effect is true for partial regressions involving markers further to the right than any marker already to the right of the second QTL.

It must be remembered that the absolute values are based on the extreme assumption of no other effects apart from the two QTLs specified in the model. We note, however, that even when the two QTLs are far apart (50 CM) there are marked 'shoulders' to the amounts of information about the additive genetic effects on either side of the two markers (5 and 11) which flank the region containing the two QTLs and a very wide 'plateau' in the intervening region. As long as we only deal with 'expectations' (i.e. very large samples under the ideal circumstances of no other QTLs in the region it is possible to identify two peaks in surface defined by the non-centrality parameters for pairwise combinations of markers. But when we allow for the fact that sample sizes are likely to be far smaller, that markers are likely to be less than fully informative and that there may be additional genetic and environmental activity apart from the two postulated QTLs, we must

Table 4 Summary of genetic systems compared in analyses

Gene action	$u_a = u_b$	$\delta^{\dagger}_{ab}$	d_a	d_b	h_a	h_b	i_{ab}	j_{ab}	j_{ba}	l_{ab}
Additive	0.9	50	1	1	0	0	0	0	0	0
Dominant	0.9	50	1	1	1	1	0	0	0	0
Duplicate	0.9	50	1	1	1	1	−1	−1	−1	−1
Complementary	0.9	50	1	1	1	1	1	1	1	1
Additive	0.9	20	1	1	0	0	0	0	0	0
Dominant	0.9	20	1	1	1	1	0	0	0	0
Duplicate	0.9	20	1	1	1	1	−1	−1	−1	−1
Complementary	0.9	20	1	1	1	1	1	1	1	1

†Map distance in CM.

expect, at best, to detect genetic activity over a relatively wide region of the genome without being able to locate the QTLs with much precision. Reducing the distance between markers will obviously increase the resolving power by sharpening the 'peaks' in principle but only at the expense of increasing the number of individuals who will have to be screened to secure sufficient numbers of recombination events.

Tables 6 and 7 give the parameters for testing the additive and dominance contributions in the presence

Table 5 Non-centrality parameters (2 d.f.) for additive contribution of identity by descent at pairs of markers linked to two loosely linked QTLs with additive effects on the phenotype. The QTLs are situated 5 CM to the right of markers 5 and 10, respectively

Markers	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	49													
3	109	109												
4	243	243	243											
5	540	540	540	540										
6	607	610	615	631	688									
7	404	414	438	497	652	626								
8	351	367	403	485	683	673	442							
9	251	429	471	565	781	776	546	442						
10	621	641	686	785	1009	1007	776	673	626					
11	558	581	630	741	990	1009	781	683	652	688				
12	263	288	343	466	741	785	565	485	497	631	540			
13	130	156	214	343	630	686	471	403	438	615	540	243		
14	71	97	156	288	581	641	429	367	414	610	540	243	109	
15	44	71	130	263	558	621	411	351	404	607	540	243	109	49

In this and all subsequent tables, the markers are assumed to be fully informative. Adjacent markers are 10 CM apart. Column numbers refer to the first ('leftmost') marker. Row numbers refer to markers situated to the right of the first marker. The QTLs have equal effects, with the frequency of the increasing alleles being 0.9. For comparative purposes, there are assumed to be no additional genetic or environmental effects on the phenotype. A total sample size of 10 000 randomly sampled sibling pairs is assumed throughout.

Table 6 Non-centrality parameters (2 d.f.) for additive contribution of identity by descent at pairs of markers linked to two loosely linked QTLs with completely dominant increasing effects on the phenotype. The QTLs are situated 5 CM to the right of markers 5 and 10, respectively

Marker	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	25													
3	55	55												
4	123	123	123											
5	276	276	276	276										
6	311	312	314	322	351									
7	204	209	221	249	330	320								
8	177	185	203	243	345	343	223							
9	127	216	237	282	391	392	275	223						
10	318	327	348	393	496	499	392	343	320					
11	285	296	319	371	484	496	391	345	330	351				
12	133	145	172	233	371	393	282	243	249	322	276			
13	66	79	108	172	319	348	237	203	221	314	276	123		
14	36	49	79	145	296	327	216	185	209	312	276	123	55	
15	22	36	66	133	285	318	207	177	204	311	276	123	55	25

Table 7 Non-centrality parameters (2 d.f.) for dominant contributions of pairs of marker loci linked to two completely dominant QTLs (see Table 6)

Marker	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	0													
3	2	2												
4	10	10	10											
5	45	45	45	45										
6	46	46	47	50	64									
7	13	14	15	22	52	47								
8	8	8	10	18	52	49	15							
9	4	14	16	25	63	60	23	15						
10	47	47	50	63	111	106	60	49	47					
11	45	46	49	61	112	111	63	52	52	64				
12	10	10	12	21	61	63	25	18	22	50	45			
13	2	2	4	12	49	50	16	10	15	47	45	10		
14	0	1	2	10	46	47	14	8	14	46	45	10	2	
15	0	0	2	10	45	47	13	8	13	46	45	10	2	0

Table 8 Apparent epistatic interaction effects (non-centrality parameters, 4 d.f.) generated by dominant QTLs not completely linked to marker loci

Marker	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	0													
3	0	0												
4	0	0	0											
5	0	0	0	0										
6	0	1	1	2	2									
7	0	1	2	3	2	0								
8	0	1	1	2	1	0	0							
9	0	0	1	1	1	0	0	0						
10	0	0	0	1	0	0	0	0	0					
11	0	0	0	1	1	0	1	1	2	2				
12	0	0	0	0	1	1	1	2	3	2	0			
13	0	0	0	0	0	0	1	1	2	1	0	0		
14	0	0	0	0	0	0	0	1	1	1	0	0	0	
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0

of complete dominance for increasing trait value. Although the hypothetical trait is still completely heritable, the effects of the dominance deviations lead to a substantial erosion of the additive genetic contribution (with the increasing alleles being more frequent) such that the maximum amount of information available for detecting additive genetic effects (Table 6) is reduced to approximately 50 per cent of that available when there are no heterozygous effects (Table 5). Addition of the information about dominance (Table 6) still yields regressions which extract only about 60 per cent of the information available when the QTLs are purely additive.

Table 8 shows that there is a small bias, even when the true gene effects are dominant but not epistatic, which appears as a small epistatic interaction. The bias presumably arises because the markers are not exactly on top of the QTLs. We note, however, that the effects of this bias are so small as to preclude their ever reaching statistical significance in actual studies.

Table 9, 10 and 11 show the expected results for tests of the additive, dominant and epistatic interactions in the case of digenic duplicate gene epistasis.

The total genetic effects, adding over cells reflecting all three sources of variation, yield only about 30 per cent of the total information recovered about genetic effects in the two-locus additive case with the same additive deviations and allele frequencies (Table 5). At best, when there are duplicate gene interactions, the additive genetic effects are expected to yield chi-squares for the test of additive effects which are only about 8–9 per cent of the values expected when there is no dominance or epistasis. When there are duplicate gene interactions, the effects of epistasis yield a large part of the information available about genetic effects (Table 11). The effects of interaction will also by most significant in the region between the two markers flanking the pair of QTLs. In theory, the presence of epistasis, although reducing the overall power of the

Table 9 Non-centrality parameters (2 d.f.) for additive contribution of identity by descent at pairs of markers linked to two loosely linked QTLs with duplicate effects on the phenotype. The QTLs are situated 5 CM to the right of markers 5 and 10, respectively

Marker	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	7													
3	15	15												
4	33	33	33											
5	73	73	73	73										
6	82	82	83	85	92									
7	54	56	59	66	85	83								
8	47	49	53	63	84	85	59							
9	31	57	61	69	82	86	69	59						
10	83	83	84	81	33	47	86	85	83					
11	74	75	77	76	24	33	82	84	85	92				
12	35	38	44	57	76	81	69	63	66	85	73			
13	17	21	28	44	77	84	61	53	59	83	73	33		
14	9	13	21	38	75	83	57	49	56	82	73	33	15	
15	6	9	17	35	74	83	55	47	54	82	73	33	15	7

Table 10 Tests for dominance effects (2 d.f.) at two loosely linked QTLs showing duplicate gene epistasis (see Table 9)

Marker	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	0													
3	1	1												
4	4	4	4											
5	17	17	17	17										
6	22	22	22	23	28									
7	9	9	10	12	24	24								
8	7	7	8	11	25	27	12							
9	3	10	10	14	31	34	17	12						
10	22	22	24	29	62	65	34	27	24					
11	18	18	19	24	57	62	31	25	24	28				
12	4	4	4	8	24	29	14	11	12	23	17			
13	1	1	1	4	19	24	10	8	10	22	17	4		
14	0	0	1	4	18	22	10	7	9	22	17	4	1	
15	0	0	1	4	18	22	9	7	9	22	17	4	1	0

QTL analysis, may provide the best clue that there is more than one locus operating within a particular region. It should be cautioned, however, that a non-centrality parameter of a little more than 100, for four degrees of freedom based on 10 000 pairs, fully informative markers and with only two QTLs needed to account for all the variance does not betoken much hope that this will ever be exploited in practice.

The effects of complementary gene epistasis are especially disconcerting (Tables 12, 13 and 14). The results indicate that the total genetic effects yield non-centrality parameters as high as 60 per cent of those obtained under the baseline case of purely additive QTL effects. Of this information, about 80 per cent is realized in tests of the additive genetic component (comparing the largest elements of Table 12 with those

Table 11 Tests for epistatic interactions (4 d.f.) between two loosely linked QTLs showing duplicate gene interaction

Marker	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	0													
3	0	0												
4	0	0	0											
5	0	0	0	0										
6	0	0	1	1	1									
7	0	1	2	4	6	3								
8	1	2	4	8	17	12	3							
9	1	4	8	19	47	39	11	3						
10	4	9	20	52	150	132	39	12	3					
11	4	8	20	52	160	150	47	17	6	1				
12	1	3	7	18	52	52	19	8	4	1	0			
13	1	1	3	7	20	20	8	4	2	1	0	0		
14	0	1	1	3	8	9	4	2	1	0	0	0	0	
15	0	0	1	1	4	4	2	1	0	0	0	0	0	0

Table 12 Non-centrality parameters (2 d.f.) for additive contribution of identity by descent at pairs of markers linked to two loosely linked QTLs with complementary effects on the phenotype. The QTLs are situated 5 CM to the right of markers 5 and 10, respectively

Marker	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	25													
3	55	55												
4	122	122	122											
5	275	275	275	275										
6	309	310	312	320	349									
7	203	208	219	248	328	318								
8	176	184	201	242	343	340	222							
9	126	215	235	281	388	389	273	222						
10	316	325	346	391	492	496	389	340	318					
11	283	294	317	368	480	492	388	343	328	349				
12	132	144	171	232	368	391	281	242	248	320	275			
13	65	78	107	171	317	346	235	201	219	312	275	122		
14	35	49	78	144	294	325	215	184	208	310	275	122	55	
15	22	35	65	132	283	316	206	176	203	309	275	122	55	25

in Table 5). However, the tests for two-locus epistatic interactions show that these effects are almost inperceptible in comparison with other effects. These findings echo the theoretical results reported for the expected pattern of twin correlations (Eaves, 1988) under two-locus epistatic models. Although it was shown that duplicate gene effects would probably be detectable against the background of additive genetic effects, those of complementary genes were virtually indistinguishable from additive effects. There is thus no 'generic' answer to the expected consequences of epistasis for a genetic analysis.

The remaining tables show, for comparison, the results for a pair of QTLs only 20 CM apart against the background of the same known marker map. Table 15 summarizes the results for the purely additive case.

Table 13 Tests for dominance effects (2 d.f.) at two loosely linked QTLs showing complementary gene epistasis (see Table 12)

Marker	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	0													
3	2	2												
4	10	10	10											
5	45	45	45	45										
6	46	46	47	50	64									
7	13	14	15	22	52	47								
8	8	8	10	18	52	49	15							
9	4	14	16	25	63	60	23	15						
10	46	47	50	63	111	106	60	49	47					
11	45	46	49	61	112	111	63	52	52	64				
12	10	10	12	21	61	63	25	18	22	50	45			
13	2	2	4	12	49	50	16	10	15	47	45	10		
14	0	1	2	10	46	47	14	8	14	46	45	10	2	
15	0	0	2	10	45	46	13	8	13	46	45	10	2	0

Table 14 Tests for epistatic effects (2 d.f.) at two loosely linked QTLs showing complementary gene epistasis (see Table 12)

Marker	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	0													
3	0	0												
4	0	0	0											
5	0	0	0	0										
6	0	1	1	2	2									
7	0	1	2	3	2	0								
8	0	1	1	2	1	0	0							
9	0	0	1	1	1	0	0	0						
10	0	0	0	1	0	0	0	0	0					
11	0	0	0	1	1	0	1	1	2	2				
12	0	0	0	0	1	1	1	2	3	2	0			
13	0	0	0	0	0	0	1	1	2	1	0	0		
14	0	0	0	0	0	0	0	1	1	1	0	0	0	
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tables 16 and 17 show the expected pattern of results for two more closely linked completely dominant loci.

The effects of duplicate gene epistasis and complementary gene interaction are summarized in Tables 18–20 and Tables 21–23, respectively.

The results for the more tightly linked loci resemble those for the conditions of looser linkage, i.e. any non-additive genetic effects tend to reduce the power when the increasing alleles are more frequent than the decreasing alleles. When there are complementary genes the effects are virtually indistinguishable from the classical additive effects of quantitative trait loci.

As expected, tighter linkage does not have a major impact on the ability of marker studies to detect genetic effects or to identify a broad region of the human

Table 15 Non-centrality parameters (2 d.f.) for additive contribution of identity by descent at pairs of markers linked to two more tightly linked QTLs with additive effects on the phenotype. The QTLs are situated 5 CM to the right of markers 7 and 9, respectively (see Tables 5–14)

Marker	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	16													
3	36	36												
4	80	80	80											
5	178	178	178	178										
6	395	395	395	395	395									
7	880	880	880	880	880	880								
8	1169	1170	1171	1173	1179	1194	1252							
9	1171	1173	1177	1188	1212	1270	1425	1400						
10	884	888	898	919	968	1081	1353	1425	1252					
11	401	407	422	455	529	697	1081	1270	1194	880				
12	184	192	209	248	335	529	968	1212	1179	880	395			
13	87	95	114	155	248	455	919	1188	1173	880	395	178		
14	43	52	71	114	209	422	898	1177	1171	880	395	178	80	
15	23	32	32	95	192	407	888	1173	1170	880	395	178	80	36

Table 16 Non-centrality parameters (2 d.f.) for additive contribution of identity by descent at pairs of markers linked to two more tightly linked QTLs with dominant increasing effects on the phenotype. The QTLs are situated 5 CM to the right of markers 7 and 9, respectively

Marker	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	8													
3	18	18												
4	40	40	40											
5	90	90	90	90										
6	200	200	200	200	200									
7	453	453	453	453	453	453								
8	608	609	609	610	612	619	648							
9	609	610	612	616	627	651	729	726						
10	455	457	461	471	493	543	676	729	648					
11	203	206	213	229	265	347	543	651	619	453				
12	93	97	105	125	168	265	493	627	612	453	200			
13	44	48	57	78	125	229	471	616	610	453	200	90		
14	22	26	36	57	105	213	461	612	609	453	200	90	40	
15	12	16	16	48	97	206	457	610	609	453	200	90	40	18

genome which might contain sites of significant QTL activity. However, the results show that the clarity with which the component loci may be resolved is uniformly reduced as the QTLs are closer together. This is no surprise. However, the discouraging fact is that the power is likely to be extremely poor even in optimal conditions. Reducing the interval between markers is only a partial solution to this problem because of the need to sample large numbers of gametes to obtain recombinants between closely linked loci.

Discussion

Before the current technological revolution in genetics, Jinks (1977, p. 353) argued that 'the number of genes

Table 17 Non-centrality parameters (2 d.f.) for dominant contributions of pairs of marker loci linked to two completely dominant QTLs (see Table 16)

Marker	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	0													
3	0	0												
4	1	1	1											
5	3	3	3	3										
6	13	13	13	13	13									
7	59	59	59	59	59	59								
8	80	80	80	80	81	84	99							
9	80	80	80	81	84	94	129	116						
10	59	59	59	60	65	80	128	129	99					
11	13	13	13	14	17	30	80	94	84	59				
12	3	3	3	3	6	17	65	84	81	59	13			
13	1	1	1	1	3	14	60	81	80	59	13	3		
14	0	0	0	1	3	13	59	80	80	59	13	3	1	
15	0	0	0	1	3	13	59	80	80	59	13	3	1	0

Table 18 Non-centrality parameters (2 d.f.) for additive contribution of identity by descent at pairs of markers linked to two more tightly linked QTLs with duplicate effects on the phenotype. The QTLs are situated 5 CM to the right of markers 7 and 9, respectively

Marker	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	3													
3	8	8												
4	17	17	17											
5	38	38	38	38										
6	84	84	84	84	84									
7	190	190	190	190	190	190								
8	257	257	257	257	257	257	259							
9	257	257	256	254	247	229	171	227						
10	190	191	191	191	188	172	85	171	259					
11	85	86	89	95	108	133	172	229	257	190				
12	39	41	44	52	70	108	188	247	257	190	84			
13	18	20	24	33	52	95	191	254	257	190	84	38		
14	9	11	15	24	44	89	191	256	257	190	84	38	17	
15	5	7	7	20	41	86	191	257	257	190	84	38	17	8

found is proportional to the patience and effort which the experimenter is willing to put into their detection'. Breeding studies in *Drosophila melanogaster* (e.g. Spickett & Thoday, 1966) and *Nicotiana rustica* (Jinks & Towey, 1976) have confirmed that sites of polygenic activity which were originally relatively few in number can readily be subdivided into ever smaller units as the opportunities for recombination are multiplied and more refined genetic studies are conducted. In fungi, where environmental conditions can be carefully controlled and the (haploid) effects of individual loci be analysed with greater precision, studies summarized by Caten (1979) have shown that a very large number of individual loci may have effects on a quantitative trait

Table 19 Tests for dominance effects (2 d.f.) at two more tightly linked QTLs showing duplicate gene epistasis (see Table 18)

	First marker locus													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	0													
3	0	0												
4	0	0	0											
5	2	2	2	2										
6	10	10	10	10	10									
7	49	49	49	49	49	49								
8	83	83	83	83	84	88	104							
9	83	83	84	86	92	113	191	166						
10	49	49	50	52	58	82	200	191	104					
11	10	10	11	11	14	25	82	113	88	49				
12	2	2	2	3	4	14	58	92	84	49	10			
13	0	0	1	1	3	11	52	86	83	49	10	2		
14	0	0	0	1	2	11	50	84	83	49	10	2	0	
15	0	0	0	0	2	10	49	83	83	49	10	2	0	0

Table 20 Tests for epistatic effects (2 d.f) at two more tightly linked QTLs showing duplicate gene epistasis (see Table 18)

Marker	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	0													
3	0	0												
4	0	0	0											
5	0	0	0	0										
6	0	0	0	0	0									
7	0	0	0	0	0	0								
8	0	0	1	1	3	7	14							
9	1	2	5	11	24	54	124	79						
10	1	3	6	13	31	74	178	124	14					
11	0	1	2	5	13	30	74	54	7	0				
12	0	0	1	2	5	13	31	24	3	0	0			
13	0	0	0	1	2	5	13	11	1	0	0	0		
14	0	0	0	0	1	2	6	5	1	0	0	0	0	
15	0	0	0	0	0	1	3	2	0	0	0	0	0	0

that, on *a priori* grounds, might be considered far 'simpler' than many human traits. Although no specific consideration of sample sizes has been given, the non-centrality parameters tabulated could be used to determine the sample sizes likely to be necessary to detect particular effects of interest as the non-centrality parameters are linearly related to sample size (Martin *et al.*, 1978). However, the situations simulated are likely to be far simpler than almost any encountered in practice because they do not allow for the resolution of genes whose expression depends on sex, age or environment. Clearly, linkage studies are increasingly recognized as a relatively blunt instrument for the genetic analysis of anything but the simplest human traits. The experience

Table 21 Non-centrality parameters (2 d.f) for additive contribution of identity by descent at pairs of markers linked to two more tightly linked QTLs with complementary epistatic effects on the phenotype. The QTLs are situated 5 CM to the right of markers 7 and 9, respectively

Marker	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	8													
3	18	18												
4	40	40	40											
5	89	89	89	89										
6	199	199	199	199	199									
7	451	451	451	451	451	451								
8	606	606	606	607	610	616	646							
9	606	607	609	614	624	648	725	723						
10	453	455	459	469	490	540	672	725	646					
11	202	205	212	228	264	345	540	648	616	451				
12	92	96	105	124	167	264	490	624	610	451	199			
13	43	48	57	78	124	228	469	614	607	451	199	89		
14	22	26	36	57	105	212	459	609	606	451	199	89	40	
15	12	16	16	48	96	205	455	607	606	451	199	89	40	18

Table 22 Tests for dominance effects (2 d.f.) at two more tightly linked QTLs showing complementary gene epistasis (see Table 21)

Marker	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	0													
3	0	0												
4	1	1	1											
5	3	3	3	3										
6	13	13	13	13	13									
7	59	59	59	59	59	59								
8	80	80	80	80	81	84	99							
9	80	80	80	81	84	95	129	116						
10	59	59	59	60	65	80	129	129	99					
11	13	13	13	14	17	30	80	95	84	59				
12	3	3	3	3	6	17	65	84	81	59	13			
13	1	1	1	1	3	14	60	81	80	59	13	3		
14	0	0	0	1	3	13	59	80	80	59	13	3	1	
15	0	0	0	1	3	13	59	80	80	59	13	3	1	0

of geneticists working on infra-human species is that we cannot know in advance which these traits will be. The further ambiguity introduced by alternative models of gene action to those commonly assumed in human quantitative genetics only enhances the need to develop and evaluate methods of genetic analysis which do not depend on methods based on identity by descent. Approaches which require genotyping only a select sample of the population such as affected sibling pairs (e.g. Motro & Thomson, 1985) or individuals from the tails of a distribution (e.g. Carey & Williamson, 1991) preserve much of the power of the linkage analysis but still require the screening of large numbers of families to identify those individuals who should be

Table 23 Tests for epistatic effects (4 d.f.) at two more tightly linked QTLs showing complementary gene epistasis (see Table 21)

Marker	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	0													
3	0	0												
4	0	0	0											
5	0	0	0	0										
6	0	0	0	0	0									
7	0	0	0	0	0	0								
8	0	0	0	1	1	2	2							
9	0	0	1	1	3	4	2	0						
10	0	0	1	2	3	6	6	2	2					
11	0	0	1	1	2	5	6	4	2	0				
12	0	0	0	1	1	2	3	3	1	0	0			
13	0	0	0	0	1	1	2	1	1	0	0	0		
14	0	0	0	0	0	1	1	1	0	0	0	0	0	
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0

genotyped. Methods which capitalize on linkage disequilibrium (e.g. Spielman *et al.*, 1993) or exploit alternative approaches to locating specific alleles (e.g. Paterson *et al.*, 1990) will probably be necessary to provide a more powerful and fine-grained resolution. The implications of epistatic and other interactions for these methods have still to be explored.

Acknowledgements

This work is supported by grants AG04945, MH45268 and HL48148 from the National Institutes of Aging, Mental Health and Heart Lung and Blood, respectively.

References

CAREY, G. AND WILLIAMSON, J. 1991. Linkage analysis of quantitative traits: increased power by using selected samples. *Am. J. Hum. Genet.*, **49**, 786–796.

CATEN, C. E. 1979. Quantitative genetic variation in fungi. In: Thompson, J. N., Jr and Thoday, J. M. (eds) *Quantitative Genetic Variation*, pp. 35–60. Academic Press, New York.

EAVES, L. J. 1988. Dominance alone is not enough. *Behav. Genet.*, **18**, 27–33.

GREENBERG, D. A. 1993. Linkage analysis of 'necessary' disease loci versus 'susceptibility' loci. *Am. J. Hum. Genet.*, **52**, 135–143.

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. *Behav. Genet.*, **2**, 3–19.

JINKS, J. L. 1977. Discussion of Inferring causes of human variation' *J. R. Stat. Soc. Lond. (A)*, **140**, 353.

JINKS, J. L. AND TOWEY, P. M. 1976. Estimating the number of genes in a polygenic system by genotype assay. *Heredity*, **37**, 69–81.

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

LALOUEL, J. M., RAO, D. C., MORTON, N. E. AND ELSTON, R. C. 1983. A unified model for complex segregation analysis. *Am. J. Hum. Genet.*, **35**, 816–826.

LANDER, E. S. AND BOTSTEIN, D. 1986. Strategies for studying heterogeneous genetic traits in humans by using a linkage map of restriction fragment length polymorphisms. *Proc. Nat. Acad. Sci. U.S.A.*, **83**, 7353–7357.

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

LUO, Z. W. AND KEARSEY, M. J. 1989. Maximum likelihood estimation of linkage between a marker gene and a quantitative locus. *Heredity*, **63**, 401–408.

LUO, Z. W. AND KEARSEY, M. J. 1991. Maximum likelihood estimation of linkage between a marker gene and a quantitative locus. II. Application to backcross and doubled haploid populations. *Heredity*, **66**, 117–124.

MACLEAN, C. J., MORTON, N. E. AND LEW, R. 1975. Analysis of family resemblance. IV. Operational characteristics of segregation analysis. *Am. J. Hum. Genet.*, **27**, 365–384.

MCGUIRE, T. R. 1992. A biometrical genetic approach to chromosome analysis of *Drosophila*: detection of epistatic interactions in geotaxis. *Behav. Genet.*, **22**, 453–467.

MARTIN, N. G., EAVES, L. J., KEARSEY, M. J. AND DAVIES, P. 1978. The power of the classical twin study. *Heredity*, **9**, 539–552.

MATHER, K. 1949. *Biometrical Genetics*, 1st edn. Methuen, London.

MATHER, K. 1974. Non-allelic interaction in continuous variation of randomly breeding populations. *Heredity*, **32**, 414–419.

MATHER, K. AND JINKS, J. L. 1982. *Biometrical Genetics: The Study of Continuous Variation*, 3rd edn. Chapman and Hall, London.

MORTON, N. E. AND MACLEAN, C. J. 1974. Analysis of family resemblance. III. Complex segregation of quantitative traits. *Am. J. Hum. Genet.*, **26**, 489–503.

MOTRO, U. AND THOMSON, G. 1985. The affected sib pair method. I. Statistical features of the affected sib pair method. *Genetics*, **110**, 525–538.

NELDER, J. A. AND WEDDERBURN, R. W. M. 1972. Generalized linear models. *J. R. Stat. Soc. Lond. (A)*, **135**, 370–384.

NEUMAN, R. J. AND RICE, J. P. 1992. Two locus models of disease. *Genet. Epidemiol.*, **9**, 347–365.

NUMERICAL ALGORITHMS GROUP. 1990. NAG FORTRAN Library: Mark 14. Numerical Algorithms Group, Oxford.

PATERSON, A. H., DAMON, S., HEWITT, J. D., ZAMIR, D., RABINOWITCH, H., LINCOLN, S. E., LANDER, E. S. AND TANKSLEY, S. D. 1991. Mendelian factors underlying quantitative triats in tomato: comparison across species, generations and environments. *Genetics*, **127**, 181–197.

PATERSON, A. H., DEVERNA, J. W., LANINI, B. AND TANKSLEY, S. D. 1990. Fine mapping of quantitative trait loci using

selected overlapping recombinant chromosomes in an interspecies cross of tomato. *Genetics*, **124**, 735–742.

PATERSON, A. H., LANDER, E. S., HEWITT, J. D., PETERSON, S., LINCOLN, S. E. AND TANKSLEY, S. D. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature*, **335**, 721–725.

PERKINS, J. M. AND JINKS, J. L. 1970. Detection and estimation of genotype-environmental, linkage and epistatic components of variation for a metrical trait. *Heredity*, **26**, 203–209.

RISCH, N. 1990. Linkage strategies for genetically complex traits. I. Multilocus models. *Am. J. Hum. Genet.*, **46**, 222–228.

SCHLAGER, G. AND CHAO, C. S. 1991. The role of dominance and epistasis in the genetic control of blood pressure in rodent models of hypertension. *Clin. Exp. Hypertens.*, **13**, 947–953.

SPICKETT, S. G. AND THODAY, J. M. 1966. Regular responses to selection: 3. Interaction between located polygenes. *Genet. Res.*, **7**, 96–121.

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

VAN DER VEEN, J. H. 1959. Tests of non-allelic interaction and linkage for quantitative characters in generations derived from two diploid pure lines. *Genetica*, **30**, 201–232.