npg

ORIGINAL ARTICLE

Efficiency of triple test cross for detecting epistasis with marker information

C Zhu and R Zhang

Laboratory of Population and Quantitative Genetics, The State Key Laboratory of Genetic Engineering, Institute of Biostatistics, School of Life Science, Fudan University, Shanghai, PR China

The triple test cross (TTC) is an experimental design for detecting epistasis and estimating the components of genetic variance for quantitative traits. In this paper, we extend the analysis to include molecular information. The statistical power of the mating design was assessed under a model assuming that a finite number of loci affect the trait in question. Formulae are developed for the analysis with or without marker information relating to the recombination fraction between loci, the genetical properties of quantitative trait controlled by the quantitative trait loci (QTL), the linkage phases of the parents and population size. Application of these formulae showed that the recombination fraction between genes and the magnitude and the types of epistasis have important interactions in their effects on power. The results demonstrate that the TTC may have increased power to detect epistasis when marker information is present. However, the simulation experiments show that the standard deviation of the estimated expected mean square was higher with one marker than that with two, whereas the corresponding value without marker information was the lowest. In addition, we demonstrate that the relative position of QTL and markers and the number of markers can both affect the power of epistatic detection.

Heredity (2007) **98,** 401–410; doi:10.1038/sj.hdy.6800956; published online 28 March 2007

Introduction

In any attempt to make appropriate use of quantitative genetic variation in animal and plant breeding programs, one needs to appraise the genetic architecture of the traits, using specific mating designs to statistically evaluate the relevant parameters. Most such designs offer statistical tests for the significance of, and estimation of, the additive and dominant components of the polygenic variation (Zeng, 1999; Mackay, 2001). More complicated designs are needed to obtain efficient statistical inference about the epistatic component of the genetic variation (in addition to the additive and dominant parts). The triple test cross (TTC), which was originally proposed by Kearsey and Jinks (1968), provides not only a direct test for significance of the epistatic variance component but also unbiased estimates of additive and dominant components whenever epistasis among polygenes is absent. Pooni and Jinks (1976) demonstrated the distinct superiority of the TTC over the alternative strategies, in statistical power for detecting complementary and duplicate epistasis based on a random model of polygenic effects. Since its innovation, various modifications or extensions have been made and popularized its applications in both animal and plant breeding (Jinks *et al.*, 1969; Jinks and Perkins, 1970; Pooni *et al.*, 1980; Goldringer *et al.*, 1997).

The recent advances in molecular biology have allowed construction of fine-scale genetic marker maps for dissecting quantitative genetic variation into chromosomal loci (QTL) (Lander and Botstein, 1989; Haley and Knott, 1992; Luo and Kearsey, 1992; Zeng, 1994; Satagopan and Yandell, 1996; Kao *et al.*, 1999; Sen and Churchill, 2001). QTL analysis opens the opportunity to characterize epistatic effects between QTL as well as effects at individual QTL (Holland, 1998; Boer *et al.*, 2002; Kao and Zeng, 2002; Yi and Xu, 2003; Yi *et al.*, 2005), and reveals ubiquitous evidence for epistatic effects detected in both animal and plant species (Fijneman *et al.*, 1996; McMullen *et al.*, 1998; Hua *et al.*, 2003; Moore, 2003). However, many statistical problems and issues of experimental design remain to be resolved for improving statistical inference of epistasis (Flint and Mott, 2001; Doerge, 2002; Jansen, 2003). For instance, Kao and Zeng (2002) recently pointed out that a two-way ANOVA exploiting genetic marker and trait phenotype data from an F_2 segregating population was, in principle, inappropriate for testing for pairwise epistasis, even though this approach has been widely used in analyses of such data sets (Yu *et al.*, 1997; Li *et al.*, 2001; Hua *et al.*, 2003).

The present paper aims at developing a quantitative genetics model and method for detecting epistasis by making use of the TTC mating design with marker information and exploring statistical power of the experimental design with or without incorporating marker information.

Correspondence: *Dr R Zhang, Laboratory of Population and Quantitative Genetics, The State Key Laboratory of Genetic Engineering, School of Life Science, Fudan University, 220 Handan Road, Shanghai 200433, PR China.*
E-mail: rmzhang@fudan.edu.cn

Theory and analysis

A finite loci model of TTC without marker information

We start by re-describing the TTC mating design, for the benefit of readers unfamiliar with it. A random sample of m individuals from the F_2 generation obtained by crossing two inbred lines P_1 and P_2 are backcrossed to three testers, that is, the parental lines P_1, P_2 and F_1. This generates $3m$ families, each of which is replicated by raising either n plots or n individuals in a randomized block experiment. Kearsey and Jinks (1968) demonstrated that to test for epistasis was equivalent to testing if $\bar{L}_{1i} + \bar{L}_{2i} - 2\bar{L}_{3i} = 0$, with $\bar{L}_{1i}$, $\bar{L}_{2i}$ and $\bar{L}_{3i}$ being progeny means of the three families from crossing the testers to the ith F_2 individual ($i = 1, 2, …, m$).

In the present study, we consider two QTL, A and B, each with two alleles (Aa and Bb). They are linked with a recombination fraction r. The genotypes of the three testers are denoted as AABB, aabb and AaBb, respectively. There are nine possible genotypes, AABB, AABb, AAbb, AaBB, AaBb, Aabb, aaBB, aaBb and aabb, in an F_2 population. Their genetic effects can be written as

$$g_i = \mu + a_1 x_1 + a_2 x_2 + d_1(0.5 - x_1^2) \\ + d_2(0.5 - x_2^2) + i_{aa} x_1 x_2 + i_{ad} x_1(0.5 - x_2^2) \quad (1) \\ + i_{da} x_2(0.5 - x_1^2) + i_{dd}(0.5 - x_1^2)(0.5 - x_2^2)$$

where μ is the population mean, a_1 and a_2 (d_1 and d_2) are the additive (dominant) genetic effects at loci A and B; i_{aa}, i_{ad} (i_{da}) and i_{dd} are additive × additive, additive × dominant and dominance × dominance epistatic effects, respectively. The indicator variables are defined as

$$x_1 = \begin{cases} 1 & \text{if genotype at A locus is AA} \\ 0 & \text{if genotype at A locus is Aa} \\ -1 & \text{if genotype at A locus is aa} \end{cases}$$

$$x_2 = \begin{cases} 1 & \text{if genotype at B locus is BB} \\ 0 & \text{if genotype at B locus is Bb} \\ -1 & \text{if genotype at B locus is bb} \end{cases}$$

Under a random genetic effect model, the expected variance component between $\bar{L}_{1i} + \bar{L}_{2i} - 2\bar{L}_{3i}$ can be worked out as

$$\sigma_I^2 = \frac{r^2}{36}\left[i_{aa}^2 + \frac{1}{2} i_{ad}^2 + (1 - 2r) i_{ad} i_{da} + \frac{1}{2} i_{da}^2 \\ + (1 - 2r + 2r^2)^2 i_{dd}^2 + 2(1 - r)(1 - 2r) i_{aa} i_{dd} \right] \quad (2)$$

It can be seen that epistatic effects and linkage between the two loci determine the above epistatic component. However, the estimate of the variance component could be biased downwards because of the multinomial variance of sample means (Falconer and Mackay, 1996, pp 51–56) and a highly unbalanced hierarchical structure of the data (Luo, 1993; Knott, 1994). This problem may disappear if the QTL effects are assumed to be fixed such as in Knott (1994) and Luo (1998).

Let f_t be the probability of the ith genotypes at QTL in the F_2 population ($t = 1, 2, …, 10$), and f_{tjk} be the frequency of the kth QTL genotype ($k = 1, 2, …, 10$) within the jth full-sib family ($j = 1, 2, 3$) from the tth F_2 parent ($t = 1, 2, …, 10$). If we let m_t be the number of the tth genotype in the F_2 samples and n_{tjk} be the number of individuals with the kth QTL genotype within the jth

full-sib family from the tth F_2 parent, they may be considered to be random variables following multinomial distributions with two sets of parameters: f_t and ($m = \sum_{t=1} m_t$) and f_{tjk} and n, respectively.

Let $\bar{\beta}_t$ ($t = 1, 2, …, 10$) be the effect of family of $\bar{L}_{1i} + \bar{L}_{2i} - 2\bar{L}_{3i}$, with $\bar{L}_{1i}$, $\bar{L}_{2i}$ and $\bar{L}_{3i}$ being progeny means of the three families from crossing the testers to the ith F_2 individual ($i = 1, 2, …, m$) and γ_{ijk} be the fixed effect of the kth QTL genotype ($k = 1, 2, …, 10$) within the jth full-sib family ($j = 1, 2, 3$) from the ith F_2 parent ($i = 1, 2, …, m$). The expected value of β_t and γ_{tjk} can be written in terms of the effect of the QTL. According to the method of analysis of variance under an unbalanced two-way nested design as described in Searle (1987, p 74), we can work out the following statistics:

The expected mean square between $\bar{L}_{1i} + \bar{L}_{2i} - 2\bar{L}_{3i}$

$$EMS_\beta = \frac{1}{6(m-1)} \left\{ \sum_{t=1}^{10} mn f_t \beta_t^2 \right. \\ - n\left[\sum_{t=1}^{10} f_t(1 + (m-1)f_t)\beta_t^2 \right. \\ \left. + 2(m-1)\sum_{t<j\leqslant 10} f_t f_j \beta_t \beta_j \right] \\ + \sum_{t=1}^{10} mf_t \sum_{j=1}^{3} T_j^2 \left[\sum_{k=1}^{10} f_{tjk}(1 + (n-1)f_{tjk})\gamma_{tjk}^2 \right. \\ \left. + 2(n-1)\sum_{k<l} f_{tjk} f_{tjl} \gamma_{tjk} \gamma_{tjl} \right] \\ - \sum_{j=1}^{3} T_j^2 \left[\sum_{t=1}^{10}\sum_{k} f_t f_{tjk}[1 + (mn-1)f_t f_{tjk}]\gamma_{tjk}^2 \right. \\ \left. \left. + 2(mn-1)\sum_{t<s\leqslant 10}\sum_{k<l} f_t f_s f_{tjk} f_{sjl} \gamma_{tjk} \gamma_{sjl} \right] \right\} + \sigma_e^2 \quad (3)$$

which has $m-1$ degrees of freedom and its significance indicates the presence of dominance × dominance and additive × dominance epistasis (Jinks and Perkins, 1970), The expected sum of squares of $\bar{L}_{1i} + \bar{L}_{2i} - 2\bar{L}_{3i}$

$$ESS_\beta = \frac{1}{6} \left\{ \sum_{t=1}^{10} n f_t \beta_t^2 \right. \\ + \sum_{t=1}^{10} f_t \sum_{j=1}^{3} T_j^2 \left[\sum_{k=1}^{10} f_{tjk}(1 + (n-1)f_{tjk})\gamma_{tjk}^2 \right. \\ \left. \left. + 2(n-1)\sum_{k<l} f_{tjk} f_{tjl} \gamma_{tjk} \gamma_{tjl} \right] \right\} + \sigma_e^2 \quad (4)$$

which has m degrees of freedom and its significance infers the presence of all epistatic effects (Jinks and Perkins, 1970), and the expected mean square within full-sib families

$$EMS_w = \sum_{t=1}^{10} f_t \sum_{j=1}^{3} \frac{T_j^2}{6} \left[\sum_{k=1}^{10} f_{tjk}(1 - f_{tjk})\gamma_{tjk}^2 \right. \\ \left. - 2\sum_{k<l} f_{tjk} f_{tjl} \gamma_{tjk} \gamma_{tjl} \right] + \sigma_e^2 \quad (5)$$

Table 1 Genotypic frequencies at two linked loci given genotypes at the marker locus in an F_2 population

Genotype	MM	Mm	mm
AABB	$(1-r_{12})^2(1-r_{23})^2$	$r_{12}(1-r_{12})r_{23}(1-r_{23})$	$r_{12}^2r_{23}^2$
AABb	$2(1-r_{12})^2r_{23}(1-r_{23})$	$r_{12}(1-r_{12})(1-2r_{23}+2r_{23}^2)$	$2r_{12}^2r_{23}(1-r_{23})$
AAbb	$(1-r_{12})^2r_{23}^2$	$r_{12}(1-r_{12})r_{23}(1-r_{23})$	$r_{12}^2(1-r_{23})^2$
AaBB	$2r_{12}(1-r_{12})(1-r_{23})^2$	$(1-2r_{12}+2r_{12}^2)r_{23}(1-r_{23})$	$2r_{12}(1-r_{12})r_{23}^2$
$A^B//a^b$	$2r_{12}(1-r_{12})r_{23}(1-r_{23})$	$r_{12}^2r_{23}^2+(1-r_{12})^2(1-r_{23})^2$	$2r_{12}(1-r_{12})r_{23}(1-r_{23})$
$A^b//a^B$	$2r_{12}(1-r_{12})r_{23}(1-r_{23})$	$r_{12}^2(1-r_{23})^2+r_{23}^2(1-r_{12})^2$	$2r_{12}(1-r_{12})r_{23}(1-r_{23})$
Aabb	$2r_{12}(1-r_{12})r_{23}^2$	$(1-2r_{12}+2r_{12}^2)r_{23}(1-r_{23})$	$2r_{12}(1-r_{12})(1-r_{23})^2$
aaBB	$r_{12}^2(1-r_{23})^2$	$r_{12}(1-r_{12})r_{23}(1-r_{23})$	$(1-r_{12})^2r_{23}^2$
aaBb	$2r_{12}^2r_{23}(1-r_{23})$	$r_{12}(1-r_{12})(1-2r_{23}+2r_{23}^2)$	$2(1-r_{12})^2r_{23}(1-r_{23})$
aabb	$r_{12}^2r_{23}^2$	$r_{12}(1-r_{12})r_{23}(1-r_{23})$	$(1-r_{12})^2(1-r_{23})^2$

The marker locus is assumed to locate between the QTL with recombination frequencies of r_{12} to its left flanking marker and r_{23} to its right flanking QTL, respectively.

which has $3m(n-1)$ degrees of freedom and is used to test for significance of the above two expected statistics. In equations (3)–(5), T_j ($j=1, 2, 3$) is the coefficient of orthogonal contrast and takes values $T_1=T_2=1$ and $T_3=-2$ (Snedecor and Cochran, 1989, p 257). It must be noted that covariances in the number of progeny with any given QTL genotype between different marker genotype classes are equal to zero and this holds throughout the paper.

A model of TTC in the case of finite loci with marker information

We incorporate marker information into the above TTC analysis by considering two scenarios – one- or two-marker loci. Firstly, we consider a marker locus linked to the two QTL.

We consider a marker locus lying between two linked QTL with genotypes MM, Mm, mm observed for each F_2 parent. The genotypes of the parental lines P_1, P_2 and their offspring F_1 are denoted as AAMMBB, aammbb and AaMmBb, respectively. Let r_{12} be the recombination fraction between loci A and M, and r_{23} between M and B. Without interference, the recombination fraction between A and B is given by $r_{13}=r_{12}(1-r_{23})+r_{23}(1-r_{12})$. The m families of L_1, L_2 and L_3 populations can be divided into three categories according to marker genotype of F_2 parent. The number of individuals with each of the marker genotypes, m_s ($s=1, 2, 3$), follows a trinomial distribution. Under the QTL effect model given by equation (1), frequencies of F_2 QTL genotypes given marker genotypes are described in Table 1.

It can be readily shown that the effect of the sth marker genotype α_s of $\bar{L}_{1s}+\bar{L}_{2s}-2\bar{L}_{3s}$ ($s=1, 2, 3$) is

$$\alpha_s = \sum_{t=1}^{10} f_{st}\beta_t \qquad (6)$$

where β_t describes the progeny means of the tth family group ($t=1, \ldots, 10$) of $\bar{L}_{1s}+\bar{L}_{2s}-2\bar{L}_{3s}$ within the sth marker genotype, f_{st} describes the conditional probabilities of the tth family group given sth marker genotype shown in Table 1. We can work out

$$\alpha_1 = r_{13}[i_{aa}+(1+2r_{23})i_{ad}+(1+2r_{12})i_{da}+(1+2r_{13})i_{dd}]$$

$$\alpha_2 = r_{13}\left[i_{aa}+(1+2r_{13})^2i_{dd}\right]$$

$$\alpha_3 = r_{13}[i_{aa}+(1+2r_{23})i_{ad}+(1+2r_{12})i_{da}+(1+2r_{13})i_{dd}]$$

$$(7)$$

When there are two molecular loci linked to the two QTL, we consider one of all possible orders of marker-QTL loci, M_1AM_2B, to demonstrate the following analysis. Let r_{ij} be the recombination fraction between the ith and jth loci and assume there is no recombination interference. Under the two-marker model, we can work out the effect of the ith maker genotype α_s ($s=1, \ldots, 9$)

$$\alpha_1 = r_{24}\{i_{aa}-(1-2r_{34})i_{ad}-(1-r_{12}-r_{23}) \\ \times[i_{da}-(1-2r_{34})i_{dd}]/(1-r_{13})\}$$

$$\alpha_2 = r_{24}\{i_{aa}-[(1-2r_{12})r_{23}(1-r_{23})i_{da} \\ -(1-2r_{24})^2r_{12}(1-r_{12})i_{dd}]/[r_{13}(1-r_{13})]\}$$

$$\alpha_3 = r_{24}\{i_{aa}+(1-2r_{34})i_{ad} \\ +(r_{12}-r_{23})[i_{da}+(1-2r_{34})i_{dd}]/r_{13}\}$$

$$\alpha_4 = r_{24}\{i_{aa}-(1-2r_{34})i_{ad}-r_{12}(1-r_{12})[(1-2r_{23})i_{da} \\ -(1-2r_{24})i_{dd}]/[r_{13}(1-r_{13})]\}$$

$$\alpha_5 = r_{24}\{i_{aa}+[(1-2r_{24})^2 \\ \times(1-2r_{12}+2r_{12}^2)/(1-2r_{13}+2r_{13}^2)]i_{dd}\}$$

$$\alpha_6 = r_{24}\{i_{aa}+(1-2r_{34})i_{ad}+r_{12}(1-r_{12})[(1-2r_{23})i_{da} \\ +(1-2r_{24})i_{dd}]/[r_{13}(1-r_{13})]\}$$

$$\alpha_7 = r_{24}\{i_{aa}-(1-2r_{34})i_{ad} \\ +(r_{23}-r_{12})[i_{da}-(1-2r_{34})i_{dd}]/r_{13}\}$$

$$\alpha_8 = r_{24}\{i_{aa}+[(1-2r_{12})r_{23}(1-r_{23})i_{da} \\ +(1-2r_{24})^2r_{12}(1-r_{12})i_{dd}]/[r_{13}(1-r_{13})]\}$$

$$\alpha_9 = r_{24}\{i_{aa}+(1-2r_{34})i_{ad} \\ +(1-r_{12}-r_{23})[i_{da}+(1-2r_{34})i_{dd}]/(1-r_{13})\}$$

$$(8)$$

Equations (7) and (8) show that the marker-associated quantitative genetic effects are entirely determined by epistatic effects and linkage parameters and that significant variation between the marker effects is an indicator of the presence of epistasis. It should be noted that r_{24} represents recombination between two QTL. When $r_{24}=0$, the model degenerates to a single QTL.

Let f_s be the frequency of the sth marker genotype, f_{st} the conditional probability of the tth family group given the sth marker genotype of $\bar{L}_{1s}+\bar{L}_{2s}-2\bar{L}_{3s}$ and f_{stjk} the conditional probability of the kth QTL genotype ($k=1, 2, \ldots, 10$) within the jth full-sib family ($j=1, 2, 3$) from the tth F_2 parent within the sth marker genotype. The expected mean squares must be calculated following an

analysis of variance under an unbalanced linear model such as that described in Searle (1987, p 73).

The expected mean square between-marker genotypes is

$$
\begin{aligned}
EMS_\alpha = \frac{1}{6\times(G-1)} &\left\{ n\sum_{s=1}^{G} mf_s\alpha_s^2 \right. \\
&- n\left[\sum_{s=1}^{G} f_s[1+(m-1)f_i]\alpha_i^2 \right. \\
&\left. +2(m-1)\sum_{s<t} f_sf_t\alpha_s\alpha_t\right] \\
&+ n\sum_{s=1}^{G}\left[\sum_{t=1}^{10} f_{st}[1+(mf_s-1)f_{st}]\omega_{st}^2 \right.\\
&\left. + 2(mf_s-1)\sum_{t<k\leqslant 10} f_{st}f_{sk}\omega_{st}\omega_{sk}\right] \\
&- n\left[\sum_{s=1}^{G}\sum_{t=1}^{10} f_sf_{st}[1+(m-1)f_sf_{st}]\omega_{st}^2 \right.\\
&\left. +2(m-1)\sum_{s<k} \sum_{t<l\leqslant 10} f_sf_{st}f_kf_{kl}\omega_{st}\omega_{kl}\right] \\
&+ \sum_{s=1}^{G}\sum_{j=1}^{3} T_j^2\left[\sum_{t=1}^{10}\sum_{k=1}f_{st}f_{stjk}(1+(mnf_s-1)f_{st}f_{stjk})\gamma_{stjk}^2 \right.\\
&\left. +2(mnf_s-1)\sum_{t'<s'\leqslant 10}\sum_{k'<t'} f_{st}f_{stjk}f_{s't'}f_{st'jk'}\gamma_{stjk}\gamma_{stjk'}\right] \\
&- \sum_{j=1}^{3} T_j^2\left[\sum_{s=1}^{G}\sum_{t=1}^{10}\sum_{k=1} f_sf_{st}f_{stjk}(1+(mn-1)f_sf_{st}f_{stjk})\gamma_{stjk}^2 \right.\\
&+ 2(mn-1) \\
&\left.\left. \times \sum_{s<s'\leqslant G}\sum_{t<t'\leqslant 10}\sum_{k<k'} f_sf_{st}f_{stjk}f_{s'}f_{s't'}f_{s't'j'k'}\gamma_{stjk}\gamma_{s't'j'k'}\right]\right\} \\
&+ \sigma_e^2
\end{aligned}
$$

(9)

and the expected mean square within-marker genotypes is

$$
\begin{aligned}
EMS_w = \frac{1}{6(m-G)} &\left\{ \sum_{s=1}^{G}\sum_{t=1}^{10} mnf_sf_{st}\omega_{st}^2 \right. \\
&- n\sum_{s=1}^{G}\left[\sum_{t=1}^{10} f_sf_{st}(1+(m-1)f_sf_{st}\omega_{st}^2 \right.\\
&\left. +2(m-1)\sum_{t<t'\leqslant 10} f_sf_{st}f_sf_{st'}\omega_{st}\omega_{st'}\right] \\
&+ \sum_{s=1}^{G}\sum_{t=1}^{10}\sum_{j=1}^{3} T_j^2 mnf_sf_{st}\left[\sum_{k=1} f_{stjk}(n-1)f_{stjk}\gamma_{stjk}^2 \right.\\
&\left. +2(n-1)\sum_{k<k'} f_{stjk}f_{stjk'}\gamma_{stjk}\gamma_{stjk'}\right]
\end{aligned}
$$

$$
\begin{aligned}
&+ \sum_{s=1}^{G}\sum_{t=1}^{10}\sum_{j=1}^{3} T_j^2 mf_sf_{st}\left[\sum_{k=1} f_{stjk}(n-1)f_{stjk}\gamma_{stjk}^2 \right.\\
&\left. +2(n-1)\sum_{k<k'} f_{stjk}f_{stjk'}\gamma_{stjk}\gamma_{stjk'}\right] \\
&- \sum_{s=1}^{G}\sum_{j=1}^{3} T_j^2\left[\sum_{t=1}^{10}\sum_{k=1} f_{st}f_{stjk}(1+(mnf_s-1)f_{st}f_{stjk})\gamma_{stjk}^2 \right.\\
&\left.\left. +2(mnf_s-1)\sum_{t<t'\leqslant 10}\sum_{k<k'} f_{st}f_{stjk}f_{st'}f_{stj'k'}g_{stjk}g_{st'jk'}\right]\right\} \\
&+ \sigma_e^2
\end{aligned}
$$

(10)

where G is the number of marker genotypes ($G=3$ and 9 for the one- and two-marker model, respectively), ω_{st} is the effect of the tth family group within the sth marker genotype and γ_{stjk} is the effect of the kth QTL genotype ($k=1, 2, ..., 10$) within the jth full-sib family ($j=1, 2, 3$) from the tth F_2 parent within the sth marker genotype as described before.

Power prediction

The above analysis demonstrates that the significance of the epistatic variance can be evaluated by testing for significance of the expected mean square between family (EMS_β) against that within full-sib family (EMS_w) with or without use of marker information.

As both the between- and the within-marker genotype mean squares follows a noncentral χ^2 distribution with degree of freedom predefined, the F statistic for significance test of the between-marker variances given by equation (10) follows a doubly noncentral F distribution. The power of the F test statistic can be calculated from the probability as follows:

$$
\text{power} = \Pr[F(v_1, v_2, \lambda, \delta) > F(v_1, v_2, \lambda', \delta', 1-\alpha)] \quad (11)
$$

where $F(v_1, v_2, \lambda, \delta)$ represents a doubly noncentral F variable with degrees of freedom v_1, v_2 and the noncentral parameters λ and δ for the numerator and denominator mean squares, respectively (Bulgren, 1971).

Calculation of noncentral parameters

The distribution parameters can be determined by following Johnson et al. (1995, vol 2, p 131) for the situations with or without incorporating marker information.

When no marker information is involved in the analysis, the noncentral parameter of the numerator of the F-statistic is given by

$$
\lambda = \sum_{t=1}^{10} \frac{mf_t\left[n\beta_t^2+\sum_{j=1}^{3} T_j^2[\sum_{k=1} f_{tjk}(1-f_{tjk})\gamma_{tjk}^2-2\sum_{k<l} f_{tjk}f_{tjl}\gamma_{tjk}\gamma_{tjl}]\right]}{\sum_{j=1}^{3} T_j^2\left[\sum_{k=1} f_{tjk}(1-f_{tjk})\gamma_{tjk}^2-2\sum_{k<k'} f_{tjk}f_{tjk'}\gamma_{tjk}\gamma_{tjk'}\right]+6\sigma_e^2}
$$

(12)

which has m degrees of freedom, when significance of the expected sum squares given by equation (4) is tested.

$$\lambda = \sum_{t=1}^{10} mf_t \frac{A_1}{A_2} \tag{13}$$

which has $m-1$ degree of freedom, when significance of the expected mean squares given by equation (3) is tested, where

$$A_1 = n(\beta_t - \bar{\beta}_t)^2 - \frac{1}{m}\left[\sum_{t=1} f_t(1-f_t)\beta_t^2 - 2\sum_{t<j} f_t f_j \beta_t \beta_j\right]$$
$$+ \sum_{j=1}^{3} T_j^2\left[\sum_{k=1} f_{tjk}(1-f_{tjk})\gamma_{tjk}^2 - \sum_{k<k'} 2f_{tjk}f_{tjk'}\gamma_{tjk}\gamma_{tjk'}\right]$$
$$- \frac{1}{m}\left[\sum_{t=1}^{10}\sum_{j=1}^{3} T_j^2 \sum_{k=1} f_t f_{tjk}(1-f_t f_{tjk})\gamma_{tjk}^2\right.$$
$$\left. -2\sum_{t<t'\leqslant 10}\sum_{j=1}^{3} T_j^2 \sum_{k=k'} f_t f_{tjk} f_{t'} f_{t'jk'}\gamma_{tjk}\gamma_{t'jk'}\right]$$

and

$$A_2 = \sum_{j=1}^{3} T_j^2\left[\sum_{k=1} f_{tjk}(1-f_{tjk})\gamma_{tjk}^2 - 2\sum_{k<k'} f_{tjk}f_{tjk'}\gamma_{tjk}\gamma_{tjk'}\right]$$
$$+ 6\sigma_e^2$$

and the noncentral parameter of denominator is

$$\delta = \sum_{t=1}^{10}\sum_{j=1}^{3}\left(mf_t T_j^2\left\{\sum_{k=1} nf_{tjk}\gamma_{tjk}^2\right.\right.$$
$$-\left[\sum_{k=1} f_{tjk}(1+(n-1)f_{tjk})\gamma_{tjk}^2\right.$$
$$\left.\left.\left. +2(n-1)\sum_{k<k'} f_{tjk}f_{tjl}\gamma_{tjk}\gamma_{tjk'}\right]\right\}\right)\bigg/6\sigma_e^2 \tag{14}$$

Under the marker-QTL model, the noncentral parameter of the numerator statistic is

$$\lambda = \sum_{s=1}^{G} \frac{B_1}{B_2} \tag{15}$$

where

$$B_1 = mnf_s(\alpha_s - \bar{\alpha}_s)^2 + n\left[f_s(1-f_s)\omega_{st}^2 - 2\sum_{t<t'} f_{st}f_{st'}\omega_{st}\omega_{st'}\right]$$
$$- nf_s\left[\sum_{s=1}^{G}\sum_{t=1}^{10}\omega_{st}^2 - 2\sum_{s<s'}\sum_{t<t'} f_s f_{st} f_{s'} f_{s't'}\omega_{st}\omega_{s't'}\right]$$
$$+ \sum_{j=1}^{3} T_j^2\left[\sum_{t=1}^{10}\sum_{k=1} f_{st}f_{stjk}(1-f_{st}f_{stjk})\gamma_{stjk}^2\right.$$
$$\left. -2\sum_{t<t'\leqslant 10}\sum_{k<k'} f_{st}f_{stjk}f_{st'}f_{st'jk'}\gamma_{stjk}\gamma_{st'jk'}\right]$$

$$- f_s\sum_{j=1}^{3} T_j^2\left[\sum_{s=1}^{G}\sum_{t=1}^{10}\sum_{k=1} f_s f_{st}f_{stjk}(1-f_s f_{st}f_{stjk})\gamma_{stjk}^2\right.$$
$$\left. -2\sum_{s<s'}\sum_{t<t'\leqslant 10}\sum_{k<k'} f_s f_{st}f_{stjk}f_{s'}f_{s't'}f_{s't'jk'}\gamma_{stjk}\gamma_{s't'jk'}\right]$$

and

$$B_2 = \frac{1}{6(mf_s-1)}\left[n\sum_{t=1}^{10} mf_s f_{st}\omega_{st}^2\right.$$
$$- nf_{st}\sum_{t=1}^{10} f_{st}(1+(mf_s-1)f_{st}\omega_{st}^2$$
$$\left. +2(mf_s-1)\sum_{t<t'\leqslant 10} f_{st}f_{st'}\omega_{st}\omega_{st'}\right]$$
$$+ \sum_{t=1}^{10} f_{st}\sum_{j=1}^{3}\frac{T_j^2}{6}\left[\sum_{k=1} f_{stjk}(1-f_{stjk})\gamma_{stjk}^2\right.$$
$$\left. -2\sum_{k<=k'} f_{stjk}f_{stjk'}\gamma_{stjk}\gamma_{stjk'}\right] + \sigma_e^2$$

The noncentral parameter of the denominator statistic is

$$\delta = mn\sum_{s=1}^{G}\sum_{t=1}^{10} f_s f_{st}\frac{C_1}{C_2} \tag{16}$$

where

$$C_1 = \omega_{st}^2 - \frac{1}{mf_s}\left[\sum_{t=1}^{10} f_{st}(1-f_{st})\omega_{st}^2 - 2\sum_{t<t'} f_{st}f_{st'}\omega_{st}\omega_{st'}\right]$$
$$+ \sum_{j=1}^{3}\frac{T_j^2}{n}\left[\sum_{k=1} f_{stjk}(1-f_{stjk})\gamma_{stjk}^2 - 2\sum_{k<k'} f_{stjk}f_{stjk'}\gamma_{stjk}\gamma_{stjk'}\right]$$
$$- \sum_{j=1}^{3}\frac{T_j^2}{mnf_s}\left[\sum_{t=1}^{10}\sum_{k=1} f_{st}f_{stjk}(1-f_{st}f_{stjk})\gamma_{stjk}^2\right.$$
$$\left. -2\sum_{t<t'}\sum_{k<k'} f_{st}f_{stjk}f_{st'}f_{st'jk'}\gamma_{stjk}\gamma_{st'jk'}\right]$$

and

$$C_2 = \sum_{j=1}^{3}\frac{T_j^2}{6}\left[\sum_{k=1} f_{stjk}(1-f_{stjk})\gamma_{stjk}^2\right.$$
$$\left. -2\sum_{k<k'} f_{stjk}f_{stjk'}\gamma_{stjk}\gamma_{stjk'}\right] + \sigma_e^2$$

The power of function (11) can be evaluated using the cumulative distribution of the doubly noncentral F distribution, which was approximated by an infinite Poisson-weighted series of multiple of incomplete beta function (Bulgren, 1971).

It is noted that $F(v_1, v_2, \lambda', \delta', 1-\alpha)$ stands for the upper α-point ($\alpha = 0.05$) of the doubly noncentral F distribution with the same degree of freedom but the noncentral parameters of numerator λ' and denominator δ' are

calculated under the null hypotheses in equations (12)–(16). For any given parameters of the doubly noncentral F distribution $v_1, v_2, \lambda', \delta'$ and a given significance level α, the threshold F_c was calculated from numerically solving equation $\int_{F_c}^{\infty} f(v_1, v_2, \lambda', \delta', x)dx = \alpha$. In the present studies, α was set to be 5%.

In the above analysis, we presented the formulation for only one of all possible combinations of the marker-QTL linkage orders. Following the same principles, we formulated the analysis for other combinations but do not present them here to simplify the paper.

However, when two markers are considered, m families of L_1, L_2 and L_3 populations are divided into nine categories. One of the complications is that the two genetic markers are so close to each other that some expected frequency of some marker genotypes is too small to be observed in practice. In the simulated analysis, the actual observed marker genotypes may be used for analysis of variance. When the expected number of genotypes is less than two in the theoretical analysis, we classify them as missing data. In this situation, some statistics, such as the degree of freedom and the expected sum of squares between- and within-marker genotypes, must be correspondingly adjusted and approximated in both theoretical and simulated analyses.

Simulation study and numerical analysis

Simulation study

To validate the analytical predictions aforementioned, we carried out simulations that mimic a mating experiment with TTC design (100 (families) × 20 (progenies per family)). In our simulations, we varied genetic parameters and sample sizes. In particular, genetic crossover events between genes at linked marker loci and/or QTL were simulated according to the random walk 'algo-rithm' described elsewhere (Luo and Kearsey, 1992) and recombination interference was ignored. Each individual phenotype was generated as its genotypic value from equation (1) plus a random number sampled from a standard normal distribution. The simulation for each of simulated parameter configurations was repeated 1000 times. For each set of simulation parameters, the simulation was also carried out under the null hypothesis ($i_{aa} = i_{ad} = i_{da} = i_{dd} = 0$). The 95 percentile of the F values derived from 1000 simulations under the null hypothesis was used as 5% threshold to test for significance of the corresponding alternative hypothesis. The proportion of the significant tests in 1000 simulations was defined as the empirical power, which was used to compare with the theoretical prediction.

Results

Tabulated in Tables 2–4 are the expected sum of squares between families, the expected mean square within full-sib families and F statistic, together with their corresponding standard deviations, over 1000 replicates of simulations and those predicted from calculations based on the theoretical analyses developed in the present study. The theoretical predictions are in good agreement with the simulated observations, validating the theoretical model presented here. In Tables 2–4, simulated observations of the powers and the thresholds of statistically testing for epistasis are shown together with the theoretical predictions for all the simulated populations. The theoretical calculations of the power provided adequate predictions to the corresponding simulated values.

Table 2 shows that some kinds of epistasis are more likely to be detected than others. TTC design indicated higher power for testing additive-by-additive (i_{11}) variance component than that for detecting additive-by-

Table 2 Comparison of powers for detecting epistatic components with $a_1 = a_2 = 0.5$, $d_1 = d_2 = 0.25$ and the genetic distance between the QTL of 45 cM

Genetic parameter				Simulated					Predicted				
i_{aa}	i_{ad}	i_{da}	i_{dd}	$\hat{F}_c$	ESS_β	EMS_w	$\hat{F}$	$\hat{\beta}$	F_c	ESS_β	EMS_w	F	β
1.0	0	0	0	1.2478	1.6290 ± 0.2280	1.3312 ± 0.0344	1.2393 ± 0.1750	44.1	1.2509	1.6242	1.3307	1.2205	42.98
0	1.0	0	0	1.2483	1.3893 ± 0.1906	1.2465 ± 0.0300	1.1279 ± 0.1641	22.7	1.2509	1.3926	1.2459	1.1177	21.21
0	0.0	1.0	0	1.2439	1.3893 ± 0.1921	1.2454 ± 0.0310	1.1337 ± 0.1593	23.6	1.2509	1.3926	1.2459	1.1177	21.21
0	0	0	1.0	1.2499	1.3653 ± 0.1851	1.2724 ± 0.0306	1.0625 ± 0.1614	13.2	1.2509	1.3721	1.2725	1.0782	14.63

Expected sum of squares between families (ESS_β), the expected mean square within full-sib families (EMS_w) and F statistic, together with their corresponding standard deviations, and predicted from theoretical calculation, as well as the observed threshold (F_c) and powers (β) and their corresponding theoretical predictions in analysis of variance are given. The same is true for Tables 3 and 4.

Table 3 The effect of linkage and linkage phases on epistatic detection with $a_1 = a_2 = 0.5$, $d_1 = d_2 = 0.25$ and $i_{aa} = i_{ad} = i_{da} = i_{dd} = 0.5$

Distance (cM)	Linkage phase	Simulated					Predicted				
		$\hat{F}_c$	ESS_β	EMS_w	$\hat{F}$	$\hat{\beta}$	F_c	ESS_β	EMS_w	F	β
15	Coupling	1.2557	1.2543 ± 0.1803	1.1850 ± 0.0291	1.0592 ± 0.1551	10.1	1.2533	1.2503	1.1855	1.0546	10.55
	Repulsion	1.2317	1.2567 ± 0.1800	1.2562 ± 0.0289	1.0002 ± 0.1455	6.9	1.2494	1.2638	1.2559	1.0063	5.77
30	Coupling	1.2603	1.3992 ± 0.1909	1.2328 ± 0.0299	1.1357 ± 0.1581	21.3	1.2526	1.3946	1.2325	1.1315	23.67
	Repulsion	1.2670	1.3330 ± 0.1924	1.2807 ± 0.0292	1.0393 ± 0.1524	7.8	1.2501	1.3235	1.2801	1.0339	8.47
45	Coupling	1.2422	1.4911 ± 0.2141	1.2536 ± 0.0303	1.1901 ± 0.1732	36.8	1.2523	1.4968	1.2534	1.1941	37.55
	Repulsion	1.2614	1.3899 ± 0.1976	1.2902 ± 0.0310	1.0778 ± 0.1552	12.1	1.2506	1.3922	1.2923	1.0773	14.23

Table 4 Comparison of powers of additive effects and dominance effects for detecting epistasis with $i_{aa} = i_{ad} = i_{da} = i_{dd} = 0.5$ and the genetic distance of 30 cM between the QTL

Genetic parameter				Simulated					Predicted				
a_1	d_1	a_2	d_2	$\hat{F}_c$	ESS_β	EMS_w	$\hat{F}$	$\hat{\beta}$	F_c	ESS_β	EMS_w	F	β
0.50	0.25	0.50	0.25	1.2397	1.3869 ± 0.1969	1.2323 ± 0.0298	1.1262 ± 0.1616	24.8	1.2526	1.3946	1.2325	1.1315	23.67
1.00	0.25	1.00	0.25	1.2467	1.9047 ± 0.2726	1.7507 ± 0.0490	1.0889 ± 0.1581	15.1	1.2591	1.9241	1.7520	1.0982	17.18
1.50	0.25	1.50	0.25	1.2536	2.8623 ± 0.4157	2.6909 ± 0.0892	1.0646 ± 0.1567	13.1	1.2602	2.8545	2.6922	1.0602	13.44
0.50	0.50	0.50	0.50	1.2499	1.4401 ± 0.2093	1.2754 ± 0.0324	1.1303 ± 0.1678	24.9	1.2579	1.4379	1.2758	1.1270	21.16
0.50	1.00	0.50	1.00	1.2514	1.7288 ± 0.2435	1.5575 ± 0.0384	1.1108 ± 0.1591	18.8	1.2586	1.7293	1.5572	1.1105	18.98
0.00	0.00	0.00	0.00	1.2365	1.3206 ± 0.1861	1.1548 ± 0.0295	1.1442 ± 0.1724	28.8	1.2448	1.3176	1.1555	1.1402	25.49

dominance (i_{12} or i_{21}) variance, whereas the design showed lowest power for detecting dominance-by-dominance (i_{22}) variance component, conditional on the same genetical background. Table 3 shows that there is a trend toward increase in the power as the genetical distance between the two QTL increases, which implies that epistasis may be more difficult to be detected when the QTL are tightly linked. The test tends to be more efficient when the QTL are linked in coupling than that in repulsion, which was verified by Pooni and Jinks (1976) in a theoretical study of the same design. The effects of heritability and dominance ratio on epistatic detection are summarized in Table 4. The decline of heritability (narrow sense heritability) and dominance increases the power for detecting epistasis and the power reaches the highest value when the additive and dominance effects approach zero on the condition that the epistatic effects are constant.

Having demonstrated the superiority of the theoretical predictions without marker information, we now implement the theoretical analyses when marker information(s) exists. In the same genetic background (i.e., we set $a_1 = a_2 = 0.5$, $d_1 = d_2 = 0.25$ and $i_{aa} = i_{ad} = i_{da} = i_{dd} = 0.5$ in all the simulated populations), 10 populations were simulated for 10 different sets of parameters as summarized in Table 5. To allow fair comparisons, the genetic distance between two QTL was set to 45 cM in all simulations.

The TTC has greater power to detect epistasis when marker information exists. Comparison among populations 1, 7, and 10 shows that the ranking of epistatic detection is $AM_1M_2B > AM_1B > AB$, where loci A and B are QTL and M_1 and M_2 are marker loci, and the power is determined by the marker that is closest to the QTL when two markers locate at the same side of QTL (comparison between populations 6 and 9). The relative position of QTL and markers may affect the power of epistatic detection. There is a trend toward decrease in the power of the epistatic detection as the number of marker loci between the two QTL decreases (comparison among populations 1, 3, 5 and 6; see also comparison between populations 7 and 9).

Comparing the power of epistatic detection with one-marker, two-marker and without marker information, it will be seen from Table 5 that the standard deviations of estimated expected mean square with one marker indicated higher value than that with two markers, whereas the corresponding value without marker information showed the lowest value.

Using marker information, tabulated in Table 5 are expected mean squares of between-marker genotype,

within-marker genotype and the F ratio estimated from simulation, together with their corresponding standard errors predicted from theoretical calculations, as well as the observed powers and their corresponding theoretical predictions. It can be seen from Table 5 that theoretical predictions of the powers using equation (11) also provide an adequate approximation for the simulated values in all 10 populations.

Discussion

The TTC, which was originally proposed by Kearsey and Jinks (1968), provides not only a direct test for epistatic variance component but also unbiased estimates of additive and dominant components whenever epistasis among polygenes is absent. It has been shown that it is the most advanced design so far to investigate the genetic architecture of both experimental and natural populations (Kearsey and Jinks, 1968). The analysis of designs with or without marker information, presented here, provides useful predictions of statistical power. Our comments are confined to two-locus epistatic effects but these modifications indicate the general issues.

To calculate the power we need the distribution of the test statistics under both the null and the alternative hypothesis. Under the assumption of the fixed model, as both between- and within-marker genotype mean squares follow noncentral χ^2 distribution, the F statistic for significance test of the between-marker variances (or between families) given follows a doubly noncentral F distribution under both alternative and null hypothesis. The power for a given degree of freedom, significance level and the noncentral parameters of numerator and denominator have been calculated (Bulgren, 1971; Johnson et al., 1995). Derivations in the present paper have shown that the power for detecting epistasis can be expressed as a function of design parameters and parameters describing genetic properties of the marker and QTL. The powers from theoretical evaluation agree very well with those from stochastic simulation under a wide range of situations, suggesting reliability of the theoretical analysis.

Assuming no interference (in recombination), we demonstrate that there is an interaction between epistasis and linkage information that is responsible in part for the improvement in sensitivity of detection under the finite locus model, which is very helpful in understanding where the increase in power comes from using markers.

However, in real experimental organisms, genetic interference will affect crossovers, as has been well known since Haldane (1919). Interference will have the

Table 5 Numerical results of analysis of variance

Population	Order	D_{12}	D_{23}	D_{34}	Simulated					Predicted				
					$\hat{F}_c$	ESS_β	EMS_w	$\hat{F}$	$\hat{\beta}$	F_c	ESS_β	EMS_w	F	β
1	AM_1M_2B	5	35	5	2.0882	2.7470±1.1967	1.2723±0.1971	2.2310±1.0319	48.4	2.0526	2.9315	1.2660	2.3155	52.46
2	AM_1M_2B	10	25	10	2.0507	2.5194±1.1859	1.2853±0.1879	2.0045±1.0083	41.5	2.0975	2.5264	1.2922	1.9551	42.13
3	M_1AM_2B	5	40	5	1.9880	2.5742±1.1401	1.2792±0.1897	2.0395±0.9556	44.9	2.1459	2.6245	1.2779	2.0538	47.98
4	M_1AM_2B	5	5	40	2.1799	2.4914±1.2458	1.2957±0.1953	1.9733±1.0559	36.0	2.3974	2.4128	1.3050	1.8489	34.27
5	M_1ABM_2	15	45	5	1.9814	2.5432±1.1142	1.2847±0.1929	2.0228±0.9555	42.5	2.1312	2.4606	1.2928	1.9033	36.38
6	M_2M_1AB		5	45	2.0717	2.1840±1.1493	1.3086±0.1875	1.7052±0.9407	28.3	2.4066	2.1959	1.3120	1.6737	28.91
7	AM_1B	5		40	2.8003	4.6784±3.2846	1.6429±0.3191	2.6763±1.4869	43.7	2.7699	4.7521	1.6162	2.9402	46.89
8	AM_1B	20		25	2.7113	4.2191±2.9968	1.6048±0.2907	2.4852±1.4495	40.8	2.7677	4.1820	1.6948	2.4675	39.43
9	M_1AB	5		45	2.7245	3.1938±2.5119	1.5466±0.2705	1.9550±1.3128	27.7	2.8808	3.2206	1.5827	2.0349	28.14
10	AB		45		1.2504	1.3770±0.1934	1.2510±0.0308	1.1015±0.1575	17.7	1.2511	1.3991	1.2534	1.1162	19.42

Expected mean squares of between-marker genotype (EMS_W), within-marker genotype (EMS_W) and F statistic (F) estimated from simulation, together with their corresponding standard errors and predicted from theoretical calculations, as well as the observed threshold ($\hat{F}_c$) and powers (β) and their corresponding theoretical predictions with $a_1=a_2=0.5$, $d_1=d_2=0.25$ and $i_{aa}=i_{ad}=i_{da}=i_{dd}=0.5$.

effect of decreasing the apparent genetic distances between the loci along the linkage group (McPeek and Speed, 1995), which will result in a decrease in the power for QTL detection (Xu *et al.*, 2005). The simulation experiments in Table 3 show that there is a trend in the decrease of the power of the epistatic detection as the genetic distance between two QTL deceases. In view of its importance, the problem of interference deserves further investigation.

Wade (2002) identified additive-by-additive epistasis as one of the most important kinds of epistasis because it contributed most heavily to the generation of new additive variance. Table 2 shows that the TTC design indicated higher power for testing additive-by-additive (i_{aa}) variance component than that for detecting additive-by-dominance (i_{ad} or i_{da}) and dominance-by-dominance (i_{dd}) variance component, conditional on the same genetical background. Because additive-by-additive epistasis are relatively easier to detect than other types of epistasis, it seems that more examples of additive-by-additive epistasis are available from QTL studies, for instance, in the mouse, additive-by-additive epistasis has been shown to characterize genes affecting lung tumors (Fijneman *et al.*, 1996).

Pooni and Jinks (1976) demonstrated the greater power of the TTC for detecting epistasis, compared to other alternatives without marker information. Our investigation, presented here, differs from that of Pooni and Jinks in several aspects. Firstly, we develop a quantitative genetics model and method for detecting epistasis by making use of the TTC experiments with marker information and exploring the statistical power of an experimental design with or without incorporating marker information. Secondly, Pooni and Jinks (1976) focus on exploring statistical power for detecting complementary and duplicate epistasis under the infinitesimal model, whereas the statistical power for detecting arbitrary types of epistasis of the mating design was assessed under a finite locus model. In addition, it could be interesting to compare the statistical power for detecting epistasis to that of Pooni and Jinks's. Table 4 shows that the decline in heritability (narrow heritability) and dominance results in an increase in the power for detecting epistasis. In contrast, Pooni and Jinks found that there was a trend toward an increase in the power as the heritability and dominance ratio increase. The reason for this contradiction is that the epistatic effects are expressed as a linear function of additive and dominance effects in Pooni and Jinks (1976), for example, $i^2_{aa}=a^2/10$, $i^2_{ad}=i^2_{da}=(a^2+d^2)/10$ and $i^2_{dd}=d^2/10$. It is obvious that the power will be improved in this formulation, because the epistatic effects will increase as heritability and dominance ratio increase. Their linear function (above) is hard to justify; consequently, their results may be misleading.

The TTC may have increased power for epistatic detection when marker information exists. More importantly, the availability of the molecular markers offers the opportunity for detecting pairwise interactions between QTL. The simulation experiments show that there is a trend toward decrease in the power for detecting epistasis as the number of marker loci between the two QTL decreases. Therefore, one of the optimal choices for increasing the power of epistatic detection is to explore more molecular markers between two linked epistatic QTL.

Although prediction of power in more general models, with recombination interference, multiple marker alleles, multiple markers, natural population, etc., requires tedious algebra, it is relatively simple to implement the analysis of variance with both real and simulated data. Hence analysis of variance provides a useful tool to enable quick screening of the genetic architecture of a population preliminary to the use of computationally demanding methods such as maximum likelihood or a Bayesian approach. The maximum likelihood or Bayesian approach may, however, provide more power as well as a better framework for the estimation of epistatic effects, which has been well developed in QTL mapping (Kao *et al.*, 1999; Carlborg and Haley, 2004; Yi *et al.*, 2005).

Acknowledgements

References

Boer MP, Ter Braak CJ, Jansen RC (2002). A penalized likelihood method for mapping epistatic quantitative trait loci with one-dimensional genome searches. *Genetics* **162**: 951–960.

Bulgren WG (1971). On representations of the doubly non-central F distribution. *J Am Stat Assoc* **66**: 184–186.

Carlborg O, Haley C (2004). Epistasis: too often neglected in complex trait studies? *Nat Rev Genet* **5**: 618–625.

Doerge RW (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nat Rev Genet* **3**: 43–52.

Falconer DS, Mackay TFC (1996). *Introduction to Quantitative Genetics*, 4th edn. Longman Science and Technology: Essex, UK.

Flint J, Mott R (2001). Finding the molecular basis of quantitative traits: successes and pitfalls. *Nat Rev Genet* **2**: 437–445.

Fijneman RJA, de Vries SS, Jensen RC, Demant P (1996). Complex interaction of new quantitative trait loci Sluc1, Sluc2, Sluc3 and Sluc4 that influence the susceptibility to lung cancer in the mouse. *Nat Genet* **14**: 465–467.

Goldringer I, Brabant P, Gallais A (1997). Estimation of additive and epistatic genetic variances for agronomic traits in a population of doubled-haploid lines of wheat. *Heredity* **79**: 60–71.

Haldane JBS (1919). The combination of linkage values and the calculation of distances between the loci of linked factors. *J Genet* **8**: 229–309.

Haley CS, Knott SA (1992). A simple regression method for mapping quantitative trait loci of linked factors. *Genetics* **8**: 299–309.

Holland JB (1998). EPISTACY: a SAS program for detecting two-locus epistatic interactions using genetic marker information. *J Hered* **89**: 374–375.

Hua JP, Xing YZ, Wu WR, Xu CG, Sun XL, Yu SB *et al.* (2003). Single-locus heterotic effects and dominance-by-dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. *Proc Natl Acad Sci USA* **100**: 2574–2579.

Jansen RC (2003). Studying complex biological systems using multifactorial perturbation. *Nat Rev Genet* **4**: 145–151.

Jinks JL, Perkins JM, Breese EL (1969). A general method of detecting additive, dominance and epistatic variation for metrical traits II. Application to inbred lines. *Heredity* **24**: 45–57.

Jinks JL, Perkins JM (1970). A general method of detecting additive, dominance and epistatic variation for metrical traits III F_2 and backcross populations. *Heredity* **25**: 419–429.

Johnson NL, Kotz S, Balakrishnan N (1995). *Continuous Univariate Distributions*, 2nd edn. John Wiley & Sons: New York.

Kao CH, Zeng ZB, Teasdale RD (1999). Multiple interval mapping for quantitative trait loci. *Genetics* **152**: 1203–1216.

Kao CH, Zeng ZB (2002). Modeling epistasis of quantitative trait loci using Cockerham's model. *Genetics* **160**: 1243–1261.

Kearsey MJ, Jinks JL (1968). A general method of detecting additive, dominance and epistatic variation for metrical traits I theory. *Heredity* **23**: 403–409.

Knott SA (1994). Prediction of the power of detection of marker-quantitative trait locus linkage using analysis of variance. *Theor Appl Genet* **89**: 318–322.

Lander ES, Botstein D (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.

Li ZK, Luo LJ, Mei HW, Wang DL, Shu QY, Tabien R *et al.* (2001). Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. *Genetics* **158**: 1737–1753.

Luo ZW, Kearsey MJ (1992). Interval mapping of quantitative trait loci in an F_2 population. *Heredity* **69**: 236–242.

Luo ZW (1993). The power of two experimental designs for detecting linkage between a marker locus and a locus affecting a quantitative character in a segregating population. *Genet Select Evol* **25**: 249–261.

Luo ZW (1998). Detecting linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations. *Heredity* **80**: 198–208.

Mackay TFC (2001). The genetic architecture of quantitative traits. *Nat Rev Genet* **35**: 303–339.

McPeek MS, Speed T (1995). Modeling interference in genetic recombination. *Genetics* **139**: 1031–1044.

McMullen MD, Byrne PF, Snook ME, Wiseman BR, Lee EA, Widstrom NW *et al.* (1998). Quantitative trait loci and metabolic pathways. *Proc Natl Acad Sci USA* **95**: 1996–2000.

Moore JH (2003). The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Hum Hered* **56**: 73–82.

Pooni HS, Jinks JL (1976). The efficiency and optimal size of the triple test cross design for detecting epistatic variation. *Heredity* **36**: 215–227.

Pooni HS, Jinks JL, Pooni GS (1980). A general method for the detection and estimation of additive, dominance and epistatic variance for metric traits. IV. Triple test cross analysis for normal families and their selfs. *Heredity* **44**: 177–192.

Satagopan JM, Yandell BS (1996). Markov chain Monte Carlo approach to detect Polygene loci for complex traits. *Genetics* **144**: 805–816.

Searle S (1987). *Linear Models for Unbalanced Data*. John Wiley & Sons Inc.: New York.

Sen S, Churchill G (2001). A statistical framework for quantitative trait mapping. *Genetics* **159**: 371–387.

Snedecor GW, Cochran WG (1989). *Statistical Methods*, 8th edn. Iowa State University Press, AMES, IA.

Wade MJ (2002). A gene's eye view of epistasis, selection and speciation. *J Evol Biol* **15**: 337–346.

Xu ZL, Zou F, Vision TJ (2005). Improving quantitative trait loci mapping resolution in experimental crosses by the use of genotypically selected samples. *Genetics* **170**: 401–408.

Yi NJ, Xu SZ (2003). Bayesian model choice and search strategies for mapping interacting quantitative trait loci. *Genetics* **79**: 185–198.

Yi NJ, Yandell BS, Churchill GA, Allison DA, Eisen EJ, Pomp D (2005). Bayesian model selection for genome-wide epistatic quantitative trait loci analysis. *Genetics* **170**: 1333–1344.

Yu SB, Li JX, Xu CG, Tan YF, Gao YJ, Zhang Q *et al.* (1997). Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc Natl Acad Sci USA* **94**: 9226–9231.

Zeng ZB (1994). Precision mapping of quantitative trait loci. *Genetics* **136**: 1457–1468.

Zeng ZB (1999). Estimating the genetic architecture of quantitative traits. *Genet Res* **74**: 279–289.