

The power of two experimental designs for detecting linkage between a marker locus and a locus affecting a quantitative character in a segregating population

ZW Luo *

*University of Edinburgh, Institute of Cell Animal and Population
Biology, King's Buildings, Edinburgh EH 9 3JT, UK*

(Received 14 November 1991; accepted 9 February 1993)

Summary – The statistical power of 2 experimental designs (backcrossing and intercrossing) for detecting linkage between a marker gene and a quantitative trait locus (QTL) in families derived from a segregating population is investigated. Formulae which relate power to the recombination frequency (r) between the genes, the genetical properties of the quantitative trait controlled by the QTL and the design parameters are developed. The reliability of some simplifying assumptions was confirmed by computer simulations. Application of these formulae has shown that the power of the 2 designs with population size of 1 000 was $< 20\%$ when r was 0.3 for all heritabilities of single gene considered, few large families are better than many small families, and backcrossing is generally more efficient than intercrossing. The allele frequencies and dominance properties of the QTLs have important interactions in their effects on power.

statistical power / marker – QTL linkage / backcross / intercross

Résumé – Puissance de 2 plans d'expérience pour détecter une liaison génétique entre un locus marqueur et un locus influençant un caractère quantitatif dans une population en ségrégation. Cet article étudie la puissance statistique de 2 plans d'expérience (rétrocroisement et intercroisement de F_1) pour détecter une liaison génétique entre un gène marqueur et un locus de caractère quantitatif (QTL) dans des familles dérivées d'une population en ségrégation. Des formules sont établies pour exprimer la puissance en fonction du taux de recombinaison (r) entre les gènes, des propriétés génétiques du caractère quantitatif contrôlé par le QTL et des paramètres du plan d'expérience. La fiabilité de

* Correspondence and reprints: Institute of Animal Physiology and Genetics Research, Roslin, Edinburgh EH25 9 PS, UK.

quelques hypothèses simplificatrices a été confirmée par des simulations sur ordinateur. L'application de ces formules montre que la puissance des 2 plans, pour une taille de population de 1 000, est inférieure à 20% quand r est supérieur à 0,3 pour toutes les héritabilités du gène considéré, qu'un nombre limité de familles de grande taille vaut mieux qu'un grand nombre de petites familles, et que le rétrocroisement est généralement plus efficace que l'intercroisement. Les fréquences alléliques et la dominance au locus du caractère quantitatif interagissent fortement dans leurs effets sur la puissance.

puissance statistique / liaison marqueur-QTL / rétrocroisement / intercroisement

INTRODUCTION

With the rapid development of molecular techniques in the last decade, their application to the investigation of the genetical basis of quantitative characters has become a subject of considerable activity (Botstein *et al*, 1980; Beckmann and Soller, 1986; Lander and Botstein, 1989). The central idea of these new investigations was to use the newly-discovered molecular markers (for example, RFLPs) at defined map positions for tracing linked quantitative trait loci (QTLs). Methodologically, this can be accomplished by detecting linkage between a genetic marker(s) and a QTL(s) through various appropriate experimental designs (Breese and Mather, 1957, 1960; Thoday, 1961; Jayakar, 1970; Hill, 1975; Weller, 1986; Luo, 1989; Luo and Kearsey, 1989; Lander and Botstein, 1989).

Hill (1975) demonstrated the use of analysis of variance for detecting linkage between a marker gene and a QTL by means of a nested backcrossing or intercrossing experiment and attempted to work out the power of these designs. However, because of the varying sizes of each of the nested groups, the numerator of the final test statistic used in the analysis of variance to detect the marker-QTL linkage cannot be expressed as a constant times a random χ^2 variable. Therefore, she was unable to work out analytical expression for the power of the experimental designs. Soller *et al* (1976, 1978) suggested excluding the offspring with heterozygous marker genotypes in the power analyses of the intercross design in order to increase the power of the designs. This has also avoided the complexity caused by the unequal sample sizes among the different marker genotypes and allowed use of the normal procedure of hierarchical analysis of variance so as to set up an *F*-distributed test statistic. Obviously, this results in the loss of useful information and artificially inflates the expected variance between offspring marker classes.

The present paper will focus on exploring a statistical approach to work out the experimental power of the designs suggested by Jayakar (1970) and Hill (1975) and relate the power directly to genetic parameters of the marker gene and the QTL and the relevant design parameters. This will allow factors affecting the power to be investigated comprehensively.

THEORY

Basic assumptions and experimental design

The method involves analysing progeny from natural or controlled matings in a population. Consider 2 autosomal loci, one affects a quantitative character (QTL) while the other is a codominant marker. The 2 loci are linked with a recombination fraction of r ($r' = 1 - r$). Let the frequency of allele Q_1 at the QTL be denoted p ($p = 1 - q$) and the phenotypic distributions of the 3 genotypes at the QTL, *ie* Q_1Q_1 , Q_1Q_2 and Q_2Q_2 are assumed to be $N(\mu+a, \sigma^2)$, $N(\mu+d, \sigma^2)$ and $N(\mu-a, \sigma^2)$ respectively, where a and d represent the additive and dominant effect at the QTL (Falconer, 1989). With just one QTL, σ^2 will be the environmental variance alone, but with other unlinked QTLs, it will also include genetic variance at these loci. The phenotypes of the 3 marker genotypes, *viz* M_1M_1 , M_1M_2 and M_2M_2 are distinguishable, *ie* the marker locus is codominant and we assume that the QTL and the marker gene are in linkage equilibrium in the population. One can score the progeny of these families where parents are $M_1M_1 \times M_1M_2$ or $M_1M_2 \times M_1M_2$ (*ie* backcrossing or intercrossing) and record the quantitative phenotype and marker genotype. If, for example, we consider an experiment consisting of s sibships, within each of which there are m marker classes ($m = 2$ and 3 for backcrossing and intercrossing designs, respectively). Let n_{ij} represent the number of sibs within the j th marker class within the i th sibship, then the variation for the quantitative trait can be partitioned into that between and within sibships, while that of within sibships can be further partitioned into variation within and between marker genotypes. For such unbalanced 2-way nested classification data, variance components have been worked out by Searle (1971, p 475-477). If it is further assumed that each sibship has a constant size of n then the total experimental size is $s \times n$ and analysis of variance for both backcrossing and intercrossing designs is illustrated in table I, in which:

$$n_0 = \frac{1}{m - 1} \left[n - \frac{1}{n} \sum_{j=1}^m n_{ij}^2 \right] \tag{1}$$

following Searle (1961) and Snedecor and Cochran (1968, p 189-191).

Table I. Analysis of variance of the 2 nested designs.

<i>Items</i>	<i>DF</i>	<i>EMS</i>
Between sibships	$s - 1$	-
Between markers within sibships	$s(m - 1)$	$\sigma_w^2 + n_0\sigma_B^2$
Within marker within sibships	$\sum_{i=1}^s \sum_{j=1}^m (n_{ij} - 1)$	σ_w^2

Statistical model

In the analysis of variance described in table I, the linear model for phenotypic record of the quantitative trait measured on the k th sib ($k = 1, 2, \dots, n_{ij}$) with the j th marker genotype ($j = 1, 2, \dots, m$) within the i th sibship ($i = 1, 2, \dots, s$) can be written as:

$$Y_{ijk} = \mu + \alpha_i + \beta_{ij} + e_{ijk} \quad [2]$$

where μ is an overall population mean while α_i , β_{ij} and e_{ijk} are contributions from the sibship, from the marker genotype within sibship and residual error respectively. They are assumed to be independently and normally distributed with zero means and variances σ_α^2 , σ_β^2 and σ_e^2 respectively. The frequency distribution of the QTL genotypes, the expected means and variances of the progenies within the i th marker genotypes and within all possible sibships were obtained by Hill (1975), and these were carefully rederived by Luo (1989). It was found that the expected variance between marker genotypes within sibships (σ_β^2) is:

$$\sigma_{M(I)}^2 = \frac{1}{2}(r - r')^2 pq [a^2 + (1 - 8pqr'r')d^2 + 2(q - p)ad] \quad [3.1]$$

and the expected variance within marker genotypes within sibships (σ_e^2) is:

$$\sigma_{W(I)}^2 = \frac{1}{2}pq\{(1 + 4rr')a^2 + [4rr'(1 - 4rr'pq) + 8rr'(1 - 2rr')pq + p^2 + q^2]d^2 + 2(1 + 4rr')(q - p)ad\} + \sigma^2 \quad [3.2]$$

for the intercross design; while the corresponding variances for the backcross design are:

$$\sigma_{M(B)}^2 = \frac{1}{2}pq(r - r')^2[a^2 + (1 - 2pq)d^2 + 2(q - p)ad] \quad [4.1]$$

and:

$$\sigma_{W(B)}^2 = \frac{1}{2}pq\{(1 + 4rr')a^2 + [1 + 4rr'(1 - 2pq)]d^2 + 2(1 + 4rr')(q - p)ad\} + \sigma^2 \quad [4.2]$$

It is easily seen from equations [3.1] and [4.1] that the expected variance between marker genotypes within sibship ($\sigma_{M(I)}^2$ or $\sigma_{M(B)}^2$) for either the intercross or backcross design will be statistically zero if the marker gene is not linked with the QTL, i.e. $r = 0.5$. The expected variance could also be zero if one of alleles at the QTL is fixed, i.e. $p = 0$ or 1 , but these situations are trivial. As pointed out by Jayakar (1970), under the null hypothesis $H_0 : r = 0.5$, the following ratio of mean squares:

$$F = \frac{MS_m}{MS_w} \quad [5]$$

is distributed as a central F -variable with expected value of 1. However, the ratio will be a noncentral F -variable when r is less than 0.5.

The denominator of the right side of [5] is distributed as $\chi^2_{(s(n-m))}$. However, when the cell sizes (n_{ij}) are not constant over the marker genotypes, the numerator of the F -ratio, cannot be expressed as a linear combination of chi-square variables. Therefore it is difficult to determine the power of the test directly, contrary to a traditional F -test when the null hypothesis is false.

However, under the assumption of constant size of sibships, the following approximation:

$$n_{i1} : n_{i2} : n_{i3} \approx 1 : 2 : 1 \quad [6.1]$$

can be incorporated into equation [1] for the intercrossing design and [1] can thus be rewritten as:

$$n_0 \approx \frac{1}{2}n \left(1 - \frac{3}{8}\right) \quad [6.2]$$

Similarly, the following approximation holds between sizes of 2 subsibships for the backcrossing design:

$$n_{i1} : n_{i2} \approx 1 : 1 \quad [7.1]$$

which directly results in:

$$n_0 \approx \frac{1}{2}n \quad [7.2]$$

therefore, the expectation of MS_m in equation [5] can be approximated by a general form:

$$E(MS_m) \approx \sigma_M^2 + n_0\sigma_W^2 \quad [8]$$

where σ_M^2 and σ_W^2 are respectively defined by [3.1] and [3.2] for intercrossing design or by [4.1] and [4.2] for backcrossing design. If the marker genotypes $[\beta_{ij}]$ in model [2] are considered to be fixed effects in analysis of variance described in table I, then the statistic for testing the presence of linkage between the marker gene and QTL is:

$$E(F) \approx 1 + n_0 \frac{\sigma_M^2}{\sigma_W^2} \quad [9]$$

where F is a noncentral F -variable with degrees of freedom described in table I and the noncentrality parameter:

$$\delta = (m \times s - 1)n_0 \frac{\sigma_{M(I)}^2}{\sigma_{W(I)}^2} \quad [10]$$

whose definition is the same as that in Kendall *et al* (1983, p 37) and in Johnson and Kotz (1970, p 191).

By definition, the power function of the 2 designs for detecting the linkage can be written in the following general form:

$$\beta = P_r\{F_{v_1, v_2; \delta} > F_{\alpha, v_1, v_2}\} \quad [11]$$

where $F_{v_1, v_2; \delta}$ represents a noncentral F -variable with degrees of freedom v_1 and v_2 and noncentral parameter δ while $F_{\alpha; v_1, v_2}$ stands for the upper α point of a central F -variable with degrees of freedom v_1 and v_2 .

Power calculation

So far, the power for detecting the linkage by use of these designs has been shown to be a function of the recombination fraction (r) and the basic genetic parameters at the QTL, namely the allelic frequency p ($q = 1 - p$), the additive and dominant effects at the QTL (a and d), the residual variance (σ^2) as well as the experimental design parameters s (ie the number of sibships) and n (ie the size of the sibships).

For a given broad heritability (h_b^2) and dominance ratio ($f = \frac{d}{a}$) at the QTL, the genetic variance associated with the QTL in an F_2 population is:

$$V_{gF_2} = 2pq[a^2 + (1 - 2pq)d^2 - 2(p - q)ad] \quad [12]$$

For convenience, let the phenotypic variance of the quantitative trait in the F_2 population be 100, the additive and dominant effect (a and d) can be solved as:

$$V_{gF_2} = 100h_b^2 \quad [13.1]$$

$$\sigma^2 = 100 - V_{gF_2} \quad [13.2]$$

and the additive and dominance effects at the QTL are obtained from:

$$a = \sqrt{\frac{V_{gF_2}}{2pq[1 + (1 - 2pq)f^2 - 2(p - q)f]}} \quad [14.3]$$

$$d = f \times a \quad [14.4]$$

Once the design parameters (s and n) and the genetic parameters at the QTL (p , f and h_b^2) are given together with the recombination frequency between the marker and QTL (r), the value of the noncentral F -variable can be calculated by using equation [9]. For a given significance level α of the test, the power of detecting the linkage can thus be worked out through equation [11] directly by using the relevant statistical tables such as that by Tang (1938) or Tiku (1967). Although these tables are available to provide the power of an F -test they are restricted to a limited number of degrees of freedom and to a limited range of values of the noncentral parameter. However, several procedures are available to approximate the power of the F -test (Patnaik, 1949; Laubscher, 1960; Tiku, 1965, 1967). For its higher accuracy, Tiku's 3-moment common approximation by using Laguerre series was programmed in *Mathematica* (Wolfram, 1991) to evaluate the experimental power in the present paper.

Power evaluation from simulations

Since approximations [6.2] and [7.2] were made in deriving the power function, the reliability of these approximations was checked by comparing the theoretical prediction of the power to the powers which were calculated from simulation experiments.

A Fortran-77 computer programme was designed for: i) simulating the inheritance of the marker-QTL linkage in the 2 nested experiments as described above for any combinations of experimental design and genetic parameters (Luo, 1989); ii) computing F -value from analysis of variance using the simulation data following the algorithm described by Searle (1971); and iii) calculating the frequency of significant F -values in replicated simulation trials as in Carbonell *et al.*, (1992), which gives the empirical power.

RESULTS

Although the power of the 2 designs can be easily investigated at any combinations of experimental design and genetic parameters, a total experimental size of 1 000 was only considered here. The powers of the 2 designs were evaluated by both theoretical prediction and computer simulation for all possible combinations of 2 design structures (10 (sibships) \times 100 (sibs) and 20 \times 50), heritability $h^2 = 0.01, 0.05$ and 0.10 , allelic frequency $p = 0.25, 0.5$ and 0.75 , dominance ratio $f = 0.0, 0.5$ and 1.0 as well as recombination frequency between the marker gene and QTL $r = 0.0, 0.1$ and 0.3 . The powers were evaluated at a significant level (α) equal to 0.05 . For simplicity, only part of the results were listed in table II for demonstrating an agreement between powers evaluated from theoretical prediction and simulation based on 500 replicates (in parentheses).

The powers of the 2 designs were also computed analytically for the experimental size of 1 000 but realistically smaller size of sibships and were tabulated in table III.

It could be interesting to compare the present power predictor to that of Soller and Genizi (1978). Table III in Soller and Genizi (1978) listed the number of sibships and the total experimental sizes required for achieving a power of 90% when the allelic frequency (p), dominance ratio (f) and contrast at the QTL were $0.5, 0.0$ and 0.01 (equivalent to 1% heritability in the present study) respectively, and the recombination frequency between the marker and QTL was zero. The powers with these population structures and the same genetic parameters were evaluated by use of the present method. The difference of the evaluated powers to 90% has been summarised in table IV.

Effects of recombination frequency between the marker and QTL (r), allelic frequency (p) and dominance ratio (f) at the QTL on the power of both backcrossing and intercrossing designs have been illustrated in figure 1 for a given heritability of 0.1 .

DISCUSSION

Derivations in the present paper have shown that the power of the 2 kinds of designs for detecting linkage between a marker gene and a QTL can be expressed as 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 consideration of a wide range of situations (table II), suggesting reliability of the theoretical analysis.

Recombination frequency between the marker and QTL displayed a pronounced effect on the power when $h^2 > 0.05$ (tables II, III). In this case, both designs

Table II. Comparison of powers between 2 design structures (10×100) and (20×50) of a given total experimental size of 1 000 for backcrossing and intercrossing designs where DP, h^2 , f , p and r represent design parameter, heritability, dominance ratio, allelic frequency at the QTL and recombination frequency between the marker gene and QTL. The powers were evaluated from theoretical approximation and simulation (in parentheses).

DP	h^2	f	p	Backcrossing			Intercrossing		
				$r = 0.0$	$r = 0.1$	$r = 0.3$	$r = 0.0$	$r = 0.1$	$r = 0.3$
10(100)	0.05	0.0	0.25	0.65(0.73)	0.42(0.50)	0.12(0.12)	0.47(0.52)	0.29(0.35)	0.09(0.11)
10(100)	0.05	0.0	0.50	0.65(0.66)	0.42(0.43)	0.12(0.11)	0.47(0.51)	0.29(0.32)	0.09(0.10)
10(100)	0.05	0.0	0.75	0.65(0.63)	0.42(0.44)	0.12(0.12)	0.47(0.46)	0.29(0.35)	0.09(0.11)
10(100)	0.05	0.5	0.25	0.74(0.66)	0.50(0.55)	0.13(0.22)	0.59(0.54)	0.36(0.38)	0.10(0.13)
10(100)	0.05	0.5	0.50	0.65(0.63)	0.42(0.43)	0.23(0.24)	0.53(0.50)	0.31(0.34)	0.09(0.08)
10(100)	0.05	0.5	0.75	0.48(0.47)	0.30(0.39)	0.10(0.10)	0.39(0.36)	0.23(0.24)	0.08(0.06)
10(100)	0.05	1.0	0.25	0.76(0.72)	0.52(0.44)	0.14(0.12)	0.66(0.74)	0.40(0.41)	0.10(0.10)
10(100)	0.05	1.0	0.50	0.65(0.59)	0.42(0.51)	0.12(0.11)	0.63(0.64)	0.37(0.40)	0.09(0.11)
10(100)	0.05	1.0	0.75	0.37(0.32)	0.23(0.28)	0.09(0.04)	0.42(0.46)	0.22(0.32)	0.07(0.08)
10(100)	0.10	0.0	0.25	0.96(0.93)	0.80(0.78)	0.21(0.20)	0.87(0.83)	0.62(0.62)	0.15(0.16)
10(100)	0.10	0.0	0.50	0.96(0.92)	0.80(0.80)	0.21(0.21)	0.87(0.80)	0.62(0.62)	0.15(0.14)
10(100)	0.10	0.0	0.75	0.96(0.95)	0.80(0.76)	0.21(0.20)	0.87(0.71)	0.62(0.64)	0.15(0.19)
10(100)	0.10	0.5	0.25	0.98(0.90)	0.87(0.85)	0.25(0.23)	0.94(0.91)	0.73(0.79)	0.17(0.21)
10(100)	0.10	0.5	0.50	0.96(0.94)	0.80(0.78)	0.21(0.24)	0.90(0.83)	0.66(0.63)	0.15(0.17)
10(100)	0.10	0.5	0.75	0.86(0.79)	0.63(0.56)	0.16(0.21)	0.78(0.71)	0.63(0.56)	0.16(0.21)
10(100)	0.10	1.0	0.25	0.98(0.96)	0.88(0.79)	0.26(0.28)	0.97(0.97)	0.79(0.84)	0.18(0.18)
10(100)	0.10	1.0	0.50	0.96(0.95)	0.79(0.67)	0.21(0.24)	0.96(0.95)	0.73(0.69)	0.16(0.13)
10(100)	0.10	1.0	0.75	0.73(0.61)	0.49(0.49)	0.13(0.19)	0.81(0.68)	0.48(0.54)	0.11(0.17)
20(50)	0.05	0.0	0.25	0.51(0.53)	0.31(0.39)	0.10(0.06)	0.33(0.41)	0.20(0.21)	0.08(0.10)
20(50)	0.05	0.0	0.50	0.51(0.51)	0.31(0.34)	0.10(0.09)	0.33(0.38)	0.20(0.22)	0.08(0.08)
20(50)	0.05	0.0	0.75	0.51(0.55)	0.31(0.31)	0.10(0.19)	0.33(0.31)	0.20(0.22)	0.08(0.14)
20(50)	0.05	0.5	0.25	0.60(0.58)	0.37(0.38)	0.11(0.12)	0.42(0.50)	0.25(0.19)	0.08(0.08)
20(50)	0.05	0.5	0.50	0.51(0.55)	0.31(0.36)	0.10(0.06)	0.34(0.44)	0.22(0.23)	0.08(0.06)
20(50)	0.05	0.5	0.75	0.36(0.40)	0.22(0.23)	0.08(0.09)	0.27(0.36)	0.16(0.16)	0.07(0.07)
20(50)	0.05	1.0	0.25	0.62(0.64)	0.39(0.43)	0.11(0.11)	0.49(0.52)	0.28(0.24)	0.08(0.10)
20(50)	0.05	1.0	0.50	0.51(0.50)	0.31(0.35)	0.10(0.17)	0.46(0.46)	0.25(0.23)	0.08(0.09)
20(50)	0.05	1.0	0.75	0.27(0.22)	0.17(0.24)	0.07(0.08)	0.29(0.30)	0.16(0.16)	0.06(0.03)
20(50)	0.10	0.0	0.25	0.90(0.90)	0.66(0.60)	0.16(0.20)	0.72(0.72)	0.46(0.45)	0.11(0.11)
20(50)	0.10	0.0	0.50	0.90(0.86)	0.66(0.66)	0.16(0.16)	0.72(0.72)	0.46(0.51)	0.11(0.12)
20(50)	0.10	0.0	0.75	0.90(0.87)	0.66(0.71)	0.16(0.16)	0.72(0.73)	0.46(0.50)	0.11(0.16)
20(50)	0.10	0.5	0.25	0.95(0.88)	0.75(0.67)	0.18(0.18)	0.84(0.83)	0.56(0.64)	0.13(0.14)
20(50)	0.10	0.5	0.50	0.90(0.83)	0.66(0.61)	0.16(0.20)	0.78(0.77)	0.49(0.55)	0.11(0.12)
20(50)	0.10	0.5	0.75	0.74(0.67)	0.49(0.46)	0.12(0.10)	0.61(0.65)	0.35(0.40)	0.09(0.07)
20(50)	0.10	1.0	0.25	0.96(0.90)	0.78(0.74)	0.19(0.20)	0.90(0.82)	0.62(0.62)	0.13(0.13)
20(50)	0.10	1.0	0.50	0.89(0.90)	0.66(0.64)	0.16(0.13)	0.88(0.87)	0.56(0.58)	0.12(0.14)
20(50)	0.10	1.0	0.75	0.59(0.54)	0.37(0.36)	0.11(0.14)	0.66(0.58)	0.34(0.30)	0.08(0.07)

Table III. Theoretical prediction of powers of 2 designs of a given experimental size of 1 000 where NS , h^2 and r represent the number of sibships, heritability of the QTL and recombination frequency between the marker gene and QTL. The powers listed here were evaluated at $p = 0.5$ and $f = 0.0$.

NS	r	<i>Backcrossing</i>			<i>Intercrossing</i>		
		$h^2 = 0.01$	$h^2 = 0.10$	$h^2 = 0.20$	$h^2 = 0.01$	$h^2 = 0.10$	$h^2 = 0.20$
50	0.0	0.08	0.71	0.99	0.07	0.46	0.92
50	0.1	0.07	0.44	0.88	0.06	0.27	0.65
50	0.2	0.06	0.23	0.54	0.06	0.15	0.33
50	0.3	0.06	0.11	0.21	0.05	0.09	0.14
100	0.0	0.07	0.50	0.94	0.06	0.28	0.70
100	0.1	0.06	0.29	0.69	0.06	0.17	0.41
100	0.2	0.06	0.16	0.36	0.05	0.11	0.21
100	0.3	0.05	0.09	0.15	0.05	0.07	0.10
200	0.0	0.06	0.30	0.73	0.06	0.15	0.36
200	0.1	0.06	0.18	0.44	0.05	0.05	0.21
200	0.2	0.05	0.11	0.22	0.05	0.08	0.12
200	0.3	0.05	0.07	0.10	0.05	0.06	0.08

Table IV. Difference between the power predicted in the present study and that estimated in Soller and Genizi (1978). The predictions were made when the allelic frequency (p), dominance ratio (f) and heritability (h^2) at the QTL were 0.5, 0.0 and 0.01 (equivalent to contrast of 0.01 in Soller and Genizi, 1978) respectively. The recombination frequency between the marker and QTL was zero.

NS	<i>Backcrossing</i>		<i>Intercrossing</i>	
	<i>Sibship size</i>	<i>Diff in power</i>	<i>Sibship size</i>	<i>Diff in power</i>
4	2 578	0.08	2 804	0.06
6	1 580	0.06	1 736	0.03
8	1 190	0.05	1 288	- 0.01
10	978	0.04	1 044	- 0.04
20	568	0.02	592	- 0.10
30	430	0.02	440	- 0.14
40	355	0.01	362	- 0.15
50	307	0.01	316	- 0.16
60	281	0.02	281	- 0.17
100	204	0.02	204	- 0.19
1 000	56	0.00	56	- 0.20

lost $\approx 70\%$ of their power with an increase of r from 0.1 to 0.3. Moreover, the linkage would be unlikely to be detected (power $< 20\%$) when the QTL would be linked to the marker with a recombination frequency > 0.3 when $h^2 \leq 0.1$. It has been pointed out by Risch (1991) and Collins and Morton (1991) that power is dramatically reduced when the recombination frequency is > 0.3 . Recently, Luo and

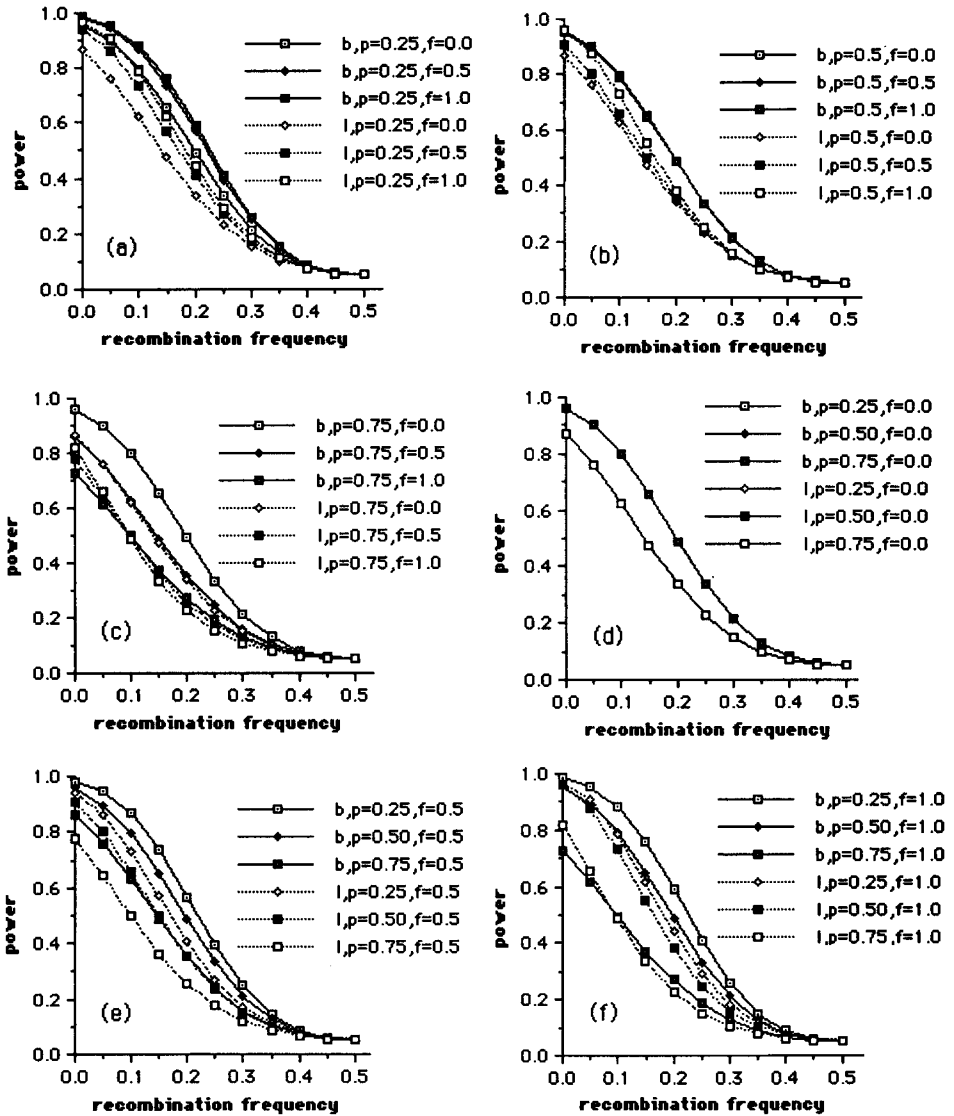


Fig 1. Power curves of backcrossing (b) and intercrossing (i) designs of 10 sibships with size of 100 for a fixed heritability of 0.1 but different allelic frequency (p) and dominance ratio (f) at QTL.

Woolliams (1992) studied the effect of recombination frequency between marker and QTL on accuracy of estimation of genetic parameters of the QTL with heritability of 0.1 and found that maximum likelihood estimates of these parameters is usually biased once the recombination frequency reaches 0.3.

The power of both designs increased with increasing dominance ratio at low allelic frequency ($p = 0.25$) (fig 1a), but decreased with increasing dominance ratio at high allelic frequency ($p = 0.75$) (fig 1c). However, there was little effect of dominance on the power of backcrossing at the allelic frequency of 0.5. At the same allelic frequency the power of intercrossing still increased with increasing dominance ratio (fig 1b).

There was no evidence of effect of allelic frequency on the power of both designs when gene effect at the QTL was purely additive ($f = 0.0$) (fig 1d). However, the power decreased with increasing allelic frequency when the allele displayed dominance (fig 1e, f).

Soller and Genizi (1978) published the first comprehensive theoretical study of the same designs as addressed in the present paper but through investigating significance of contrast between means of marker genotypes of interest in quantitative trait. They concluded that the effects of gene frequency and dominance level would be important when the number of families was small. Because when the number of families is small, the probability that the contrast in each of the families be zero is so large that the power requirement will not be met for any size of family. They suggested that the probability of zero contrast would be 0.90 for backcrosses and 0.94 for intercrosses when $a = d$ and $2pq = 0.3$. Therefore, at least 22 and 34 families for the 2 designs respectively must be sampled in order that on average non zero contrasts can be expected in 2 of these families. However, if the power of these designs is calculated in the way developed in the present paper, loss in the power due to probability of sampling families with zero contrast can be avoided, since it is likely that those families with zero contrast will nevertheless contribute to significance of the variance between marker types within families. In fact, in the case of $h^2 = 0.01$, $f = 1.0$, $2pq = 0.3$ and experimental size of 5 000 (10×500), a power of 0.76 (0.70) for the backcrossing design or 0.64 (0.62) for the intercrossing design was obtained from the theoretical prediction (simulation) in the present study.

Comparison in table IV was made between the present power predictor and that in Soller and Genizi (1978). It can be seen that the powers of the backcrossing designs were slightly higher in the present paper than in Soller and Genizi (1978). While the present method yielded higher power of the intercrossing designs with small number of sibships and large sibship size than in Soller and Genizi (1978), in which offspring class with heterozygous genotype at the marker locus was excluded. However, with large number of sibships and small sibship size, the power of the intercrossing designs was lower in the present paper than in Soller and Genizi (1978).

Several researchers (Hill, 1975; Soller and Genizi, 1978) have found that for a given total experimental size, design with fewer sibships but larger sibship size was more powerful than that with more sibships but smaller sibship size. This was confirmed by the present study. Moreover, it was found that decline in power due to smaller sibship size was more severe for intercrossing design than for backcrossing design (tables II, III). The effect of the population structure on the power is parallel to its effect on degrees of freedom of the residual expected mean square (table I).

For most animal species, realistic full sibship size is very small, eg 5 to 20, but half-sibship size might be very large. Weller *et al* (1990) investigated daughter and granddaughter designs and their powers for detecting the marker and QTL linkage

in dairy cattle populations in which one sire might have several hundred daughters or granddaughters if breeding was by AI. By organising experiment of such animal species into half-sibship population structure, one might expect more power since the residual expected mean square would have more degrees of freedom.

Comparison of power of the 2 designs revealed that backcrossing was generally more powerful than intercrossing. This agrees with the conclusion of Soller and Genizi (1978).

The present studies have not directly provided the total experimental size required for a given power under a particular genetic and design situation. The theoretical calculation in the present paper can, however, be easily used in the procedure suggested by Fox (1956) so as to obtain the size of experiment for a given power in a specific situation.

ACKNOWLEDGMENTS

I thank MJ Kearsey, R Thompson and C Haley for their useful comments and helpful discussions in this paper and 3 anonymous reviewers for their many constructive suggestions and criticism.

REFERENCES

- Beckmann JS, Soller M (1986) Restriction fragment length polymorphisms and genetic improvement of agricultural species. *Euphytica* 35, 111-124
- Botstein D, White RL, Skolink M, Davies RW (1980) Construction of a genetic linkage map using restriction fragment length polymorphisms. *Am J Hum Genet* 27, 314-331
- Breese EL, Mather K (1957) The organisation of polygenic activity within a chromosome in *Drosophila*: I. Hair characters. *Heredity* 11, 373-395
- Breese EL, Mather K (1960) The organisation of polygenic activity within a chromosome in *Drosophila*: II. Viability. *Heredity* 14, 375-399
- Carbonell EA, Gerig TM, Balansard E, Asins MJ (1992) Interval mapping in the analysis of nonadditive quantitative trait loci. *Biometrics* 48, 305-315
- Collins A, Morton NE (1991) Significance of maximal lod. *Ann Hum Genet* 55, 39-41
- Falconer DS (1989) *Introduction to Quantitative Genetics*. Longman Scientific and Technical, 3rd edn
- Fox M (1956) Charts of the power of the *F*-test. *Ann Math Stat* 27, 484-497
- Hill AP (1975) Quantitative linkage: a statistical procedure for its detection and estimation. *Ann Hum Genet (Lond)* 38, 439-449
- Jayakar SD (1970) On the detection and estimation of linkage between a locus influencing a quantitative character and a marker locus. *Biometrics* 26, 451-464
- Jonhson NL, Kotz S (1970) *Distributions in Statistics: Continuous Univariate Distributions*. Houghton Mifflin, Boston, MA
- Kendall MG, Stuart A, Ord JK (1983) *The Advanced Theory of Statistics, Vol 3: Design and Analysis, and Times-Series*. Charles Griffin Ltd, 4th edn

- Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185-199
- Laubscher NH (1960) Normalizing the noncentral t and F distributions. *Ann Math Stat* 31, 1105-1112
- Luo ZW (1989) Polygene location and selection for heterotic traits in *Drosophila melanogaster*. Ph D thesis, Univ Birmingham, UK
- Luo ZW, Kearsley MJ (1989) Maximum likelihood estimation of linkage between a marker gene and a quantitative locus. *Heredity* 63, 401-408
- Luo ZW, Woolliams J (1992) Estimation of genetic parameters by using linkage between a locus underlying a quantitative trait and a marker locus. *Heredity* 70, 245-253
- Patnaitk PB (1949) The noncentral χ^2 - and F -distributions and their applications. *Biometrika* 36, 202-232
- Risch N (1991) A note on multiple testing procedures in linkage analysis. *Am J Hum Genet* 48, 1058-1064
- Searle SR (1961) Variance components in the unbalanced 2-way nested classification. *Ann Math Stat* 32, 1161-1166
- Searle SR (1971) *Linear Models*. John Wiley and Sons, London
- Snedecor GW, Cochran WG (1968) *Statistical Methods*. The Iowa State Univ Press, Ames, IA
- Soller M, Brody T (1976) On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theor Appl Genet* 47, 35-39
- Soller M, Genizi A (1978) The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. *Biometrics* 34, 47-55
- Tang CP (1938) The power function of the analysis of variance test with tables and illustrations of their use. *Stat Res Mem* 2, 126-149
- Thoday ML (1961) Location of polygenes. *Nature (Lond)* vol 191, 368-370
- Tiku ML (1965) Laguerre series forms of noncentral χ^2 - and F -distributions. *Biometrika* 52, 415-427
- Tiku ML (1967) Tables of the power of the F -test. *J Am Stat Assoc* 62, 525-539
- Weller JI (1986) Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. *Biometrics* 42, 627-640
- Weller JI, Kasi Y, Soller M (1990) Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J, Dairy Sci* 73, 2525-2537
- Wolfram S (1991) *Mathematica: A System for Doing Mathematics by Computer*. Addison-Wesley Publ Co Inc, 2nd edn