Genet. Res., Camb. (1998), **72**, *pp.* 149–158. *With* 4 *figures.* Printed in the United Kingdom © 1998 Cambridge University Press

On the use of linear regression and maximum likelihood for QTL mapping in half-sib designs

P. V. BARET[1]*, S. A. KNOTT[1] AND P. M. VISSCHER[2]

[1] *University of Edinburgh, Institute of Cell, Animal and Population Biology, West Mains Road, Edinburgh EH9 3JT, Scotland*
[2] *University of Edinburgh, Institute of Ecology and Resources Management, West Mains Road, Edinburgh EH9 3JG, Scotland*

(*Received 2 September 1997 and in revised form 1 June 1998*)

Summary

Methods of identification of quantitative trait loci (QTL) using a half-sib design are generally based on least-squares or maximum likelihood approaches. These methods differ in the genetical model considered and in the information used. Despite these differences, the power of the two methods in a daughter design is very similar. Using an analogy with a one-way analysis of variance, we propose an equation connecting the two test-statistics (*F* ratio for regression and likelihood ratio test in the case of the maximum likelihood). The robustness of this relationship is tested by simulation for different single QTL models. In general, the correspondence between the two statistics is good under both the null hypothesis and the alternative hypothesis of a single QTL segregating. Practical implications are discussed with particular emphasis on the theoretical distribution of the likelihood ratio test.

1. Introduction

A powerful approach to the detection of quantitative trait loci (QTL) is based on crosses between inbred lines differing significantly for the trait of interest. In many cases, this kind of design is impossible to implement for economic, biological or ethical reasons and alternative designs have to be considered. One of these alternatives is the study of half-sib families where individuals have one parent in common. Methodologies for QTL identification in half-sib family design were recently developed (e.g. Weller *et al.*, 1990; Knott *et al.*, 1996). This approach is based on the phenotypic differences between the half-sib progeny that inherit one allele from the sire and those that inherit the other allele (Neimann-Sørensen & Robertson, 1961). Despite the fact that they are based on models more complex than those used in an inbred line approach, half-sib methods allow the use of the extant breeding structure of some livestock or plant populations. Most genetical analyses used to identify QTL are based on least-squares (LS) or maximum likelihood (ML) approaches. LS – or regression – methods are simpler and easier to implement. ML methods are more versatile but computationally demanding. In both approaches, significance thresholds might be inferred from theoretical distributions, simulations or permutation procedures. Theoretical approaches require a perfect knowledge of the distributions of the different components of the model. In a LS approach, the statistic is a ratio of mean squares (*F* ratio). In a ML approach, the statistic is a function of a ratio of maximum likelihoods called the likelihood ratio test (LRT).

The relationship between the two test-statistics is dependent on the structure of the population being analysed. In line crosses, *F* ratio and LRT are equivalent if the errors are independent and normally distributed at a marker (e.g. Haley & Knott, 1992; Doerge, 1995):

$$LRT = n \log_e \left(1 + \left(\frac{df1}{df2}\right) F\right), \quad (1)$$

with *F* being the standard *F* ratio for testing the model with $df1$ and $df2$ degrees of freedom.

In half-sib designs, the relationship between the *F* ratio and LRT is less straightforward. In the regression approach, phenotypic values are regressed on the

* Corresponding author. Present address: Université catholique de Louvain, Faculté des sciences agronomiques, Unité de Génétique, Place Croix du Sud 2 bte 14, B-1348 Louvain-la-Neuve, Belgium. Telephone: +32 10 47.37.23. Fax: +32 10 47.37.28. e-mail: baret@gena.ucl.ac.be.

putative genotype in an across-family analysis. In the ML case, a multiple-parameter likelihood equation is used. The number of parameters considered is dependent on the type of likelihood considered (e.g. Elsen *et al*., 1997). In practice, approximate likelihoods may be used to replace full likelihoods that are computationally not feasible. In this study, we will use the approximation proposed by Knott *et al*. (1996) (see Section 2).

In a ML approach the usual assumption is that, when the null hypothesis is true, the LRT follows a χ^2 distribution with degrees of freedom equal to the number of parameters tested (Wilks, 1938; Knott *et al*., 1996). For QTL identification in a half-sib design, when testing for the effect of a biallelic QTL and the proportion of sires that are heterozygous at the QTL, the expected number of degrees of freedom is 2. In a previous comparison of ML and LS methods for QTL detection in half-sib designs, both approaches gave very similar results in terms of power (Knott *et al*., 1996). Nevertheless, such a concordance does not imply a common underlying mechanism in the calculation of the two test-statistics.

Both approaches (LS and ML) are based on the same data: individual phenotypic values and marker genotypes. Nevertheless they differ in the assumptions made and the way this information is processed. In a regression approach, the number of alleles at the QTL locus is not defined. Conversely, in the ML approach, the QTL is assumed to be biallelic. Sires can be homozygous or heterozygous at the QTL locus. This information is processed differently according to the methods. In LS, the QTL effect is fitted within each family and can vary across the different families. In consequence, the QTL effect of homozygous sites is expected to be null. In ML, two parameters are estimated: the QTL effect, which is assumed to be identical in magnitude across all the families; and the expected proportion of heterozygous sires.

More generally, the regression is considered as a simple and robust approach and the ML as a more versatile and comprehensive procedure. ML and LS are expected to perform similarly when the interval between adjacent marker is small – say less than 30 cM (e.g. Knott *et al*., 1996). For larger intervals, ML is expected to perform better because it utilizes information both from linkage and from the distribution of the data. An extreme case is that of a single marker. Using LS, it is impossible to estimate both a QTL effect and location, since they are completely confounded, whereas for ML it is theoretically possible both to map the QTL relative to the single marker and to estimate the absolute value of its location. *A priori*, it is impossible to decide which method will perform better and whether their respective test-statistics are perfectly correlated. The aim of this study was to investigate the relationship between the statistical tests in the half-sib design for an approximate ML method and a linear regression method, in the presence and absence of a QTL.

2. Methods

(i) *Design*

All our simulations are based on a daughter design (Weller *et al*., 1990). Each sire is mated to the same number of dams and the trait is measured on a single offspring per mating (a daughter, if we refer to a dairy cattle situation). The number of sires ranges from 5 to 100. In our simulations, the number of daughters per sire is always equal to 200.

On a chromosome, a single marker and a single QTL share the same location. To mimic a fully informative situation, the number of alleles at the marker is equal to 16, occurring with equal frequency, and the dam allele is specifically coded to be always identifiable. For the QTL locus, the number of alleles is equal to 2 with equal frequency (0·5). As assumed by Weller *et al*. (1990), the polygenic heritability (excluding the QTL effect) is constant and equal to 0·2. The gene effect (GE) is given according to Weller's definition as α/SD where α is half the difference between the mean trait values for the two alternative homozygotes at the QTL and SD is the within-QTL genotype standard deviation for the quantitative trait. Conversion to the Falconer and Mackay's substitution effect (a) could be achieved using the following equation:

$$a^2 = \frac{\mathrm{GE}^2}{2pq\mathrm{GE}^2 + 1}, \tag{2}$$

where a is Falconer & Mackay's (1996) substitution effect, and p and q are the frequencies of alleles A_1 and A_2 of the QTL.

Using this equation and assuming that $p = 0{\cdot}5$, values of 0·1, 0·2 and 0·3 of Weller's gene effect are respectively equal to 0·0998, 0·1980 and 0·2935 in terms of Falconer and Mackay's substitution effect. Note that for a granddaughter design, the parameterization of the effect (GE) implies that the effect is a function of the polygenic heritability, because the within-grandsire within-QTL variation among sons depends on the heritability of the trait. This may lead to the paradoxical conclusion that the power of mapping a QTL in a granddaughter design is lower when the heritability of the background polygenes is higher (Weller *et al*., 1990).

(ii) *Simulation process*

The simulation process was based on an algorithm developed for the first workshop on QTL mapping applied to livestock (Bovenhuis *et al*., 1997).

https://doi.org/10.1017/S0016672398003450 Published online by Cambridge University Press

Each QTL has two alleles that are assigned at random to the sire haplotypes depending on their frequency. For each daughter, the QTL allele inherited from the sire is determined by a binomial distribution draw between the two sire haplotypes. The QTL allele inherited from the dam is drawn from a binomial distribution. Daughter phenotypes are simulated as detailed below.

Sire breeding value is equal to

$$A_S = A_{QTL_S} + A_{\inf_S}, \tag{3}$$

where A_{QTL_S} is the sum of the effects of QTL alleles allocated to the sire; and $A_{\inf_S}$ is the contribution of other chromosomes to the trait and is randomly simulated by a normal distribution assuming the infinitesimal model $N(0,[h^2\sigma_P^2])$.

Daughter phenotype is equal to

$$P_{daughter} = A_{QTLsire} + A_{QTLdam} + \tfrac{1}{2}A_{\inf_S} + I + E, \tag{4}$$

where $A_{QTLsire}$ is the effect of the QTL allele inherited by the daughter from the sire; A_{QTLdam} is the effect of the QTL allele inherited by the daughter from the dam; I is the remaining polygenic effect simulated by a normal distribution $N(0,[0{\cdot}75h^2\sigma_P^2])$ – this term includes dam infinitesimal effects; and E is the environmental noise simulated by a normal distribution $N(0,[(1-h^2)\sigma_P^2])$.

Reduction of the genetic variance due to negative gametic disequilibrium among loci (Bulmer, 1971) was not taken into consideration.

(iii) *Methods of analysis*

Two methods of analysis were used: least-squares (LS) and maximum likelihood using a simplex maximization routine (ML). Principles of LS and ML approaches applied to a half-sib design were extensively presented in Knott *et al.* (1996). The LS analysis is based on a regression of the value of the trait on the probabilities of inheriting a given gamete from the sire. The across-family regression is nested within sires to take into account both the difference in the linkage phase between the QTL alleles and the sire gamete and the fact that some of the sires may be homozygous at the QTL.

The ML approach is based on the approximate model proposed by Knott *et al.* (1996). Simplifications from the full maximum likelihood models are based on the following assumptions: (a) the effect of the QTL is relatively small so the within-marker distribution is not influenced significantly; (b) two QTL alleles are segregating; (c) between-family genetic variance heterogeneity due to dam contributions is ignored within sire within marker. Each half-sib record is adjusted for the mean of the HS group to take into account the between-family genetic variation. Under these assumptions, the ML at each position is entirely defined by three parameters: the proportion of sires homozygous at the QTL (h), the substitution effect of the QTL (a) and the within-marker within sire residual variance (σ_w^2). The likelihood equation is (Knott *et al.*, 1996):

$$\begin{aligned} L = \prod_{i=1}^{s} \Bigg\{ & h \prod_{j=1}^{n_j} \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left(\frac{-z_{ij}^2}{2\sigma_w^2}\right) \\ & + \frac{(1-h)}{2} \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma_w^2}} \Bigg[m_{ij} \exp\left(\frac{-(z_{ij}-a/2)^2}{2\sigma_w^2}\right) \\ & + (1-m_{ij}) \exp\left(\frac{-(z_{ij}+a/2)^2}{2\sigma_w^2}\right) \Bigg] \\ & + \frac{(1-h)}{2} \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma_w^2}} \Bigg[m_{ij} \exp\left(\frac{-(z_{ij}+a/2)^2}{2\sigma_w^2}\right) \\ & + (1-m_{ij}) \exp\left(\frac{-(z_{ij}-a/2)^2}{2\sigma_w^2}\right) \Bigg] \Bigg\}, \end{aligned} \tag{5}$$

where z_{ij} is the record for the jth half-sib offspring of the ith sire adjusted to remove the between-sire effect; m_{ij} is the conditional probability that offspring j inherits gamete 1 from sire i at the position being considered; and n_i is the number of offspring of sire i.

The simplex method used as a maximization routine in ML is based on a two-step approach. (1) Different values of the percentage of heterozygous sires at the QTL loci were tested using a one-dimensional grid search. The principle of the grid search is straightforward: the likelihood function was evaluated at different points evenly spaced along the total range of variation of the parameter. For the proportion of heterozygous sires, the range spans from 0 to 1. The number of runs of the simplex algorithm within this range is dependent on the chosen increment. The smaller the increment, the more computationally demanding is the method but the more precise is the ML estimation. For each of these grid points, the ML was estimated using the simplex method (Nelder & Mead, 1965) in a two-dimensional space (gene effect and residual variance). (2) From the grid point with the highest likelihood value, the simplex algorithm was repeated in a three-dimensional space (percentage of heterozygous sires, gene effect and residual variance) to refine the ML location. All the estimates presented are based on analyses with a grid search increment of 0·01.

(iv) *Significance thresholds*

Significant thresholds are a critical and sometimes controversial aspect of QTL mapping. Two factors have to be taken into account: (1) the chosen level of

https://doi.org/10.1017/S0016672398003450 Published online by Cambridge University Press

Table 1. *Significance thresholds determined either from an F distribution with degrees of freedom* $(s, s(n-2))$ *where s is the number of sires and n is the number of daughters* (*Theoretical*) *or from distributions of test statistics for 10000 simulations with no QTL effect* (*Empirical*)

	Significance level: 0·05			Significance level: 0·01		
	LS		ML	LS		ML
No. of sires	Theoretical	Empirical	Empirical	Theoretical	Empirical	Empirical
5	2·22	2·28	2·69	3·04	3·15	5·74
10	1·84	1·87	2·92	2·33	2·41	6·04
20	1·57	1·59	3·09	1·88	1·91	6·19
	ML theoretical (for all the designs)					
$\chi^2_{(1)}$		3·84			6·63	
$\chi^2_{(2)}$		5·99			9·21	

LS, least-squares; ML, maximum likelihood.

significance and (2) the scope of the threshold (comparison-wise, chromosome-wise or genome-wise). In terms of methodologies, thresholds can be determined either from a theoretical framework assuming a given distribution of parameters or from empirical methods as simulation or permutation procedures. In the specific case of half-sib designs and considering QTL and marker at the same location, theoretical thresholds can be used in a LS framework. However, due to the approximate nature of the likelihood model used in a half-sib context, the conditions required for the asymptotic results are not met. In consequence, no theoretical threshold can be applied in conjunction with the ML approach.

For LS, the statistic is the ratio of the between-marker alleles within-sire mean square to the residual mean square. For a single location, the test-statistic distribution for the LS approach is assumed to follow an F distribution with degrees of freedom proportional to the number of sires heterozygous at the marker, because in that case a simple marker contrast is fitted for each sire family. Theoretical F ratio thresholds were determined from an F distribution with degrees of freedom $(s, s(n-2))$, where s is the number of sires and n the number of daughters. For ML, the statistic is a likelihood ratio test, i.e. $-2\log(ML_{reduced}/ML_{QTL})$ where ML_{QTL} is the ML fitting a QTL and $ML_{reduced}$ is the ML when the QTL is omitted. Empirical F ratio thresholds and ML thresholds were determined by simulation (10000 replicates) (Table 1). Simulations were achieved by generating datasets using the process described in (3) and (4) with QTL effects equal to zero. Each simulated dataset was analysed using both LS and ML methods. The test-statistics were ranked to determine the significance thresholds. Difference between power estimations are considered as different if higher than $1{\cdot}96\sqrt{u(1-u)(1/n_1+1/n_2)}$, where u is the proportion of runs above the significance threshold pooled across methods; and n_1 and n_2 are the number of runs for each method.

As power is calculated in parallel on the same simulated data, the results of both methods of analysis (LS and ML) are positively correlated and this test is conservative.

(v) *Simulation of chromosome-wide QTL scans*

In the last part of this paper, we will discuss the application of our results to chromosome scans under H1 hypothesis (a single QTL segregating). Four situations differing by the QTL effect and the informativeness of the design are simulated. The daughter design comprises 10 sires and 100 daughters per sire. A single QTL is positioned at 35 cM on a 100 cM chromosome segment. The total heritability is equal to 25% and the simulated QTL effect is either small (explaining 2% of the phenotypic variance, i.e. 8% of the additive genetic variance) or large (explaining 10% of the phenotypic variance). In the highly informative design, 11 markers are evenly spaced (interval 10 cM) on a 100 cM chromosome segment and 16 alleles are segregating per marker. The allele inherited from the dams is always identified. In the poorly informative design, only 6 biallelic markers are evenly spaced on a 100 cM segment and there is no specific coding for the dam allele. For these chromosome scans, a grid search increment of 0·25 was used for the proportion of heterozygous sires for the grid search in the ML optimization (see Section 2).

3. Results

(i) *Power calculations*

We use Monte Carlo simulations (1000 replicates per design) to recalculate the power in some of the

https://doi.org/10.1017/S0016672398003450 Published online by Cambridge University Press

Table 2. *Power in % (1000 replicates)*

No. of sires[a]	GE[b]	Weller[c]	Significance thresholds used: Least squares Empirical	Least squares Theoretical	Maximum likelihood Empirical	Maximum likelihood $\chi^2_{(1)}$	Maximum likelihood $\chi^2_{(2)}$
			Significance level 5%				
5	0·1	11	10	12	11	7	3
	0·2	37	37	39	37	27	17
	0·3	68	69	71	68	60	46
10	0·1	15	15	17	15	10	5
	0·2	54	56	58	55	48	33
	0·3	88	87	87	86	83	74
20	0·1	20	20	21	20	14	6
	0·2	76	78	78	77	71	57
	0·3	99	98	98	98	97	95
			Significance level 1%				
5	0·1	3	3	4	3	2	1
	0·2	18	17	19	18	13	8
	0·3	50	49	51	47	42	32
10	0·1	5	6	6	5	5	1
	0·2	31	33	37	33	30	18
	0·3	76	75	77	74	72	61
20	0·1	7	6	7	5	4	2
	0·2	56	59	62	56	53	35
	0·3	95	96	96	95	94	89

[a] In all cases, the number of daughters was equal to 200.
[b] Gene effect following Weller *et al*. (1990).
[c] Weller refers to the power estimate obtained from deterministic simulation by Weller *et al*. (1990). Confidence interval (95%) spans from 1% (for extreme values close to 1% and 99%) to 3% (for median values around 50%).

daughter designs analysed by Weller *et al*. (1990). Both regression and ML approaches were used. The results of power calculations are given in Table 2. At the 5% significance level, the difference between the power predicted (using the approximation given by Weller *et al*., 1990) and Monte Carlo simulation never exceeds 4 percentage points. Power estimates obtained from ML analyses are closer to Weller's estimate than LS-based estimates, but the difference is small. At the 1% significance level, the difference can reach 6 points and the empirical regression performs slightly better than the ML. However, precise assessment of power at the 1% significance level would require a higher number of simulations.

Comparison of the power calculation based on both LS with a theoretically determined threshold (assuming degrees of freedom equal to $(s, s(n-2))$ and LS with an empirically determined threshold – based on simulations without any QTL effect – showed a good concordance (less than 4 points of difference). Using theoretical thresholds, the estimated power is slightly higher. This difference in power can be explained by the fact that, in these simulations, 6·25% ($\frac{1}{16}$) of sires are expected to be uninformative at the marker level as we simulated a 16 allele marker rather than making all sires heterozygous. The degrees of freedom of the theoretical model should be modified in consequence.

Generally, the LRT is assumed to follow a χ^2 distribution with degrees of freedom (d.f.) equal to the difference between the number of parameters in the full and reduced models. If this hypothesis is correct, a theoretical χ^2 distribution with 2 d.f. should be appropriate, with 1 d.f. for the effect and 1 for the proportion of heterozygous sires. In practice, if we compare power calculations based on a theoretical χ^2 distribution and on an empirical distribution there is a strong discrepancy and the use of χ^2 thresholds induces an underestimation of the power even if a χ^2 with 1 d.f. rather than 2 d.f. is used.

All the differences in power between the empirical LS approach and the empirical ML approach are non-significant.

(ii) *Properties of the ML test-statistic distribution*

The discrepancy between the observed power based on a significance threshold obtained empirically and from a χ^2 distribution is, in part, due to the particular distribution of the ML test-statistics under the null hypothesis. As already pointed out by Le Roy & Elsen (1995), the distribution is bimodal: a peak at zero and

https://doi.org/10.1017/S0016672398003450 Published online by Cambridge University Press

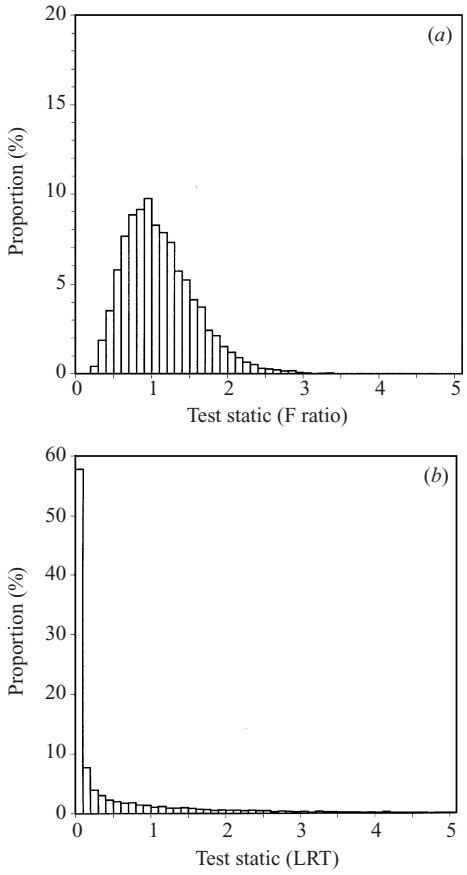

Fig. 1. Distribution of the test-statistics for a 10 sire daughter design under the null hypothesis (no QTL effect) (10000 simulations considering a single position). (*a*) Least-squares (LS) method. (*b*) Maximum likelihood (ML) method.

the rest following a χ^2 distribution (Fig. 1). A theoretical explanation is that, analogous to a random one-way ANOVA model where half the variance estimates are expected to be negative, the ML estimate is at the boundary of the permissible parameter space in half the cases and a zero LRT is obtained (e.g. Stram & Lee, 1994).

(iii) *Relationship between LS and ML test-statistics*

(*a*) *Theoretical approach*

The relationship between F and LRT can be predicted by drawing a parallel between the LS and ML approaches of a nested variance model used to estimate between- and within-level variances. This comparison is based on simplifying assumptions and will be tested using simulated data.

Consider a general one-way model:

$$MST = \frac{(a-1)\,MSB + a(b-1)\,MSW}{(ab-1)}, \tag{6}$$

where MST is the total mean square, MSB is the between-level mean square, MSW is the within-level mean square, and $(a-1)$ and $a(b-1)$ are the degrees of freedom of the between- and within-mean squares respectively. The likelihood equation for the full model is then

$$\begin{aligned} -2\log(L)\,|\,\sigma_b^2\,\sigma_w^2 = {} & \frac{(a-1)\,MSB}{(b\sigma_b^2+\sigma_w^2)} + \frac{a(b-1)\,MSW}{\sigma_w^2} \\ & + (a-1)\log(b\sigma_b^2+\sigma_w^2) \\ & + a(b-1)\log(\sigma_w^2), \end{aligned} \tag{7}$$

where σ_b^2 is the variance between levels and σ_w^2 is the variance within levels. And the likelihood equation for the reduced model is

$$-2\log(L)\,|\,\sigma_w^2 = \frac{(ab-1)\,MST}{\sigma_w^2} + (ab-1)\log(\sigma_w^2). \tag{8}$$

For $MSB < MSW$, the maximum likelihoods for the full and reduced models are equivalent, because the estimate of the between-level variance (σ_b^2) in the full model is zero. For $MSB > MSW$, it can be shown that the LRT is equal to

$$\begin{aligned} LRT = {} & (ab-1)\log\left[\frac{a(b-1)}{(ab-1)} + \frac{(a-1)}{(ab-1)}\frac{MSB}{MSW}\right] \\ & - (a-1)\log\left(\frac{MSB}{MSW}\right). \end{aligned} \tag{9}$$

As $F = MSB/MSW$,

$$\begin{aligned} LRT = {} & (ab-1)\log\left[\frac{a(b-1)}{(ab-1)} + \frac{(a-1)}{(ab-1)}F\right] \\ & - (a-1)\log(F). \end{aligned} \tag{10}$$

Equation (10) immediately implies that the relationship is valid only if $F > 1$ and that if $F = 1$, LRT = 0 for a one-way ANOVA model.

To apply this general equation to our daughter design, we modify the number of degrees of freedom according to the following equivalences between a one-way analysis of variance $(a-1, a(b-1))$ and the QTL model $(s, s(n-2))$:

$$\begin{cases} s = a-1 \\ s(n-2) = a(b-1) \end{cases} \tag{11}$$

and (10) becomes:

$$LRT = (s(n-1))\log\left[\frac{n-2}{n-1} + \frac{F}{n-1}\right] - s\log(F). \tag{12}$$

(*b*) *Empirical approach under H1*

As predicted by (12) the relationship between test-statistics obtained by LS and ML is clearly divided in two parts: (a) when the F ratio is equal to or smaller than 1, the LRT is equal to 0; (b) when the F ratio is higher than 1, F ratio and LRT are both positive and correlated according to a non-linear pattern. From

https://doi.org/10.1017/S0016672398003450 Published online by Cambridge University Press

Table 3. *Proportion of runs where the F ratio is below 1 and proportion of runs where likelihood ratio test (LRT) is equal to 0 for different daughter designs when no QTL was simulated (null hypothesis): 1000 replicates per design*

No. of sires[a]	*F* ratio < 1	LRT = 0
5	56·3	56·3
10	55·5	54·7
20	56·2	56·1
40	54·9	54·4
100	54·0	54·0

[a] In all cases, the number of daughters was equal to 200.

Fig. 2, it appears that the lower limit of this non-linear relationship is sharply defined, but the upper limit is less precise.

To confirm the relationship between *F* ratios smaller or equal to 1 and LRT equal to 0 in part (a) of the distribution, we ran 10 × 1000 simulations for a 10 sire design. The correlation is very high, the number of divergent results ranging from 5 to 16 out of 1000. When a divergent result was observed, the *F* ratio was smaller than 1 when the LRT was positive in 94% of the cases. In the remaining 6% the LRT was equal to 0 when the *F* ratio was higher than 1. This relationship is confirmed across a wide range of designs (Table 3).

(*c*) *Comparison of theoretical and empirical approaches under H0*

From the comparison of LRT predicted from the observed *F* using (12) and LRT obtained from ML simulations on the same dataset, it appears there is a high correlation between the two approaches (Fig. 2). The distribution of the simulated values along the theoretical curve is not symmetric. When there is a divergence, the value of LRT is always higher than the predicted value.

Studying the estimated parameters of the likelihood models after completion of the maximization process, we noted that the highest differences between the LRT and *F* ratio were observed when the estimated proportion of heterozygous sires differed markedly from 1. In the majority of simulations, the estimated proportion of heterozygous sires was close to 1 and in this case the predicted and observed LRT are very similar. In the other simulations, the estimated proportion of heterozygous sires was lower (in general, between 0 and 0·3) and the bias between the observed and predicted values was more important. To confirm this observation, we plotted the deviation from the theoretical model as a function of the test-statistic value and the estimated proportion of heterozygous sires. As expected, the highest deviations are observed when the test-statistics are higher (Fig. 3).

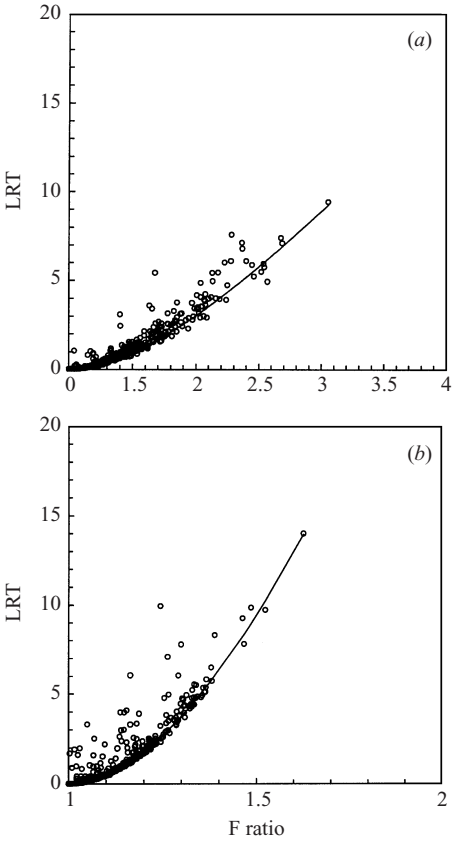

Fig. 2. The relationship between the *F* ratio and likelihood ratio test (LRT) observed with simulated data. The continuous line gives the predicted LRT calculated by transformation of the *F* ratio using equation (12). Scales of *x*-axes differ. (*a*) Ten sire design. (*b*) One hundred sire design.

(*d*) *Empirical approach under H1*

To test the robustness of the LRT prediction from an observed *F*, we compared the observed and predicted (using equation 12) LRT in four chromosome scan situations that differ by the simulated QTL effect and the informativeness of the design (see Section 2). In general, the concordance between the predicted and observed LRT is good, in both poorly and highly informative designs (Fig. 4). Generally the ML gives a slightly higher test-statistic value than the LS (predicted LRT), but ML significance thresholds obtained by simulation are also higher and power is very similar, and none of the power differences is significant (Table 4). When the quality of the design or the QTL effect increases, the difference decreases.

4. Discussion

To our knowledge, this study is the first analytical approach of the relationship between the *F* ratio and LRT in the context of QTL mapping in a half-sib design. The results not only confirm the similarity of the methods in terms of power and parameter

https://doi.org/10.1017/S0016672398003450 Published online by Cambridge University Press

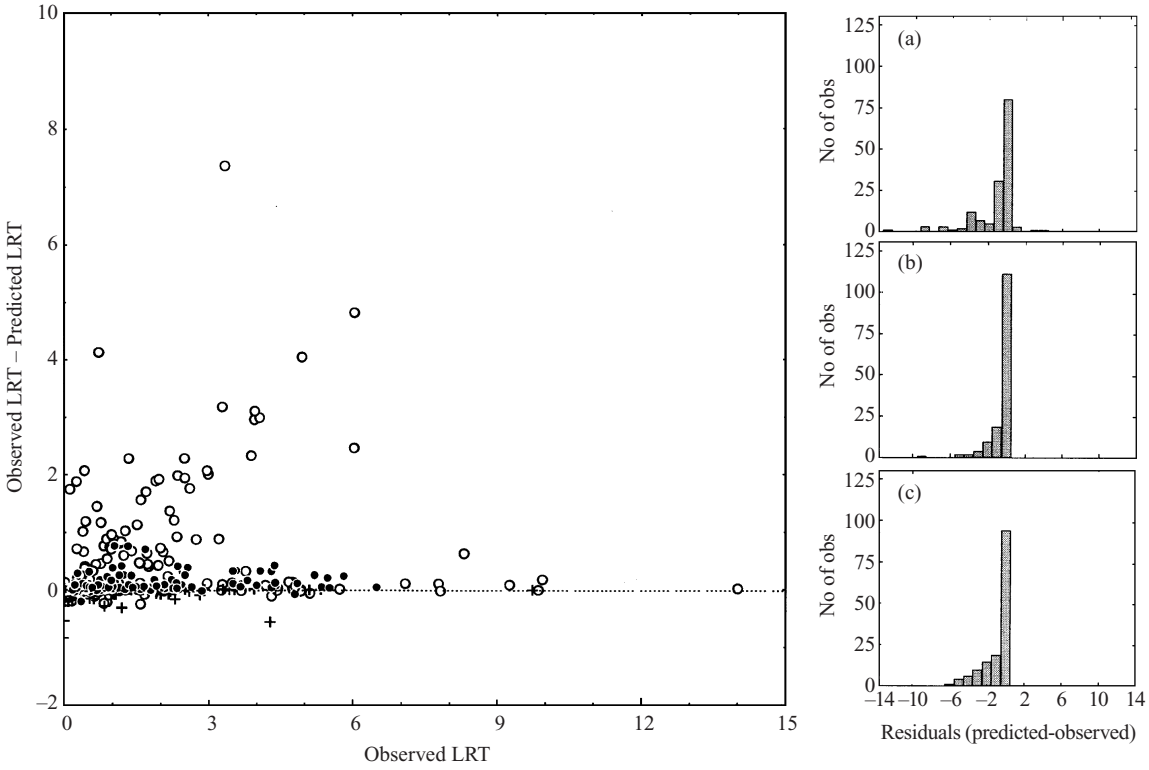

Fig. 3. Deviation of the observed LRT from the theoretical value as a function of the observed LRT and the proportion of heterozygous sires in a 100 sire design. Simulated data have been divided in three equal parts of 150 observations according to the proportion of heterozygous sires (open circles, 0·00–0·26; crosses, 0·26–0·97; filled circles, 0·98–1·00). Distribution of residuals (predicted LRT − observed LRT) according to the proportion of heterozygous sires (*a*: 0·00–0·26; *b*: 0·26–0·97; *c*: 0·98–1·00).

estimates, but provide a theoretical framework that aids in understanding this correspondence.

What are the implications of these observations? First, as demonstrated in the first part of the paper, assuming that the LRT follows a central χ^2 with 2 (or even 1) degrees of freedom under the null hypothesis will lead to a statistical test that is too conservative. Hence, power to detect a QTL will be considerably reduced. Asymptotically, a better distribution is a mixture of two χ^2 distributions of degrees of freedom 0 and 1, respectively. In this case, the significance thresholds are obtained by considering a $\chi^2_{(1)}$ and doubling the probability levels. For example, the 5% significance threshold is equal to a $\chi^2_{(1)}$ and $P = 0{\cdot}1$, i.e. 2·70, and the 1% significance threshold is equal to a $\chi^2_{(1)}$ and $P = 0{\cdot}02$, i.e. 5·1. These values are close to our empirical observations (see Table 1). For computationally demanding methods such as ML, it is tempting to take a significance threshold from standard statistical tables. But because there is not a complete agreement, our results reconfirm that an empirical method for setting significance thresholds, either by simulation or by permutation, is to be preferred.

Second, it is very tempting to extend our conclusions to practical applications, but this may not be appropriate. Despite the robustness of the proposed prediction across a wide range of designs, we have to keep in mind that the use of a one-way variance model as a bridge between LS and ML is based on a series of simplifying assumptions. In some situations, such as unbalanced designs, we cannot rule out the possibility that the discrepancy between methods might increase.

Moreover, such an extension is often irrelevant. In most of the practical applications, the choice of the method of analysis is more dependent on the characteristics of the designs and objectives of the study than on theoretical or statistical considerations. Regression is generally used in simple situations when the emphasis is on the location of the QTL. ML is usually considered in more complex designs, and when more parameters are estimated.

Nevertheless, as there is a quasi one-to-one correspondence between *F* ratios smaller than 1 and LRT equal to 0 under the null hypothesis, we can take advantage of this property in the calculation of the significance thresholds using a very simple algorithm:

(1) simulate data with a QTL effect equal to 0;
(2) calculate *F* ratio using a LS approach;
(3) if $F \leqslant 1$, LRT = 0;
(4) if $F > 1$, calculate LRT using a ML approach.

The implementation of this algorithm provides a gain in computing time of 43, 40 and 58% in significance

https://doi.org/10.1017/S0016672398003450 Published online by Cambridge University Press

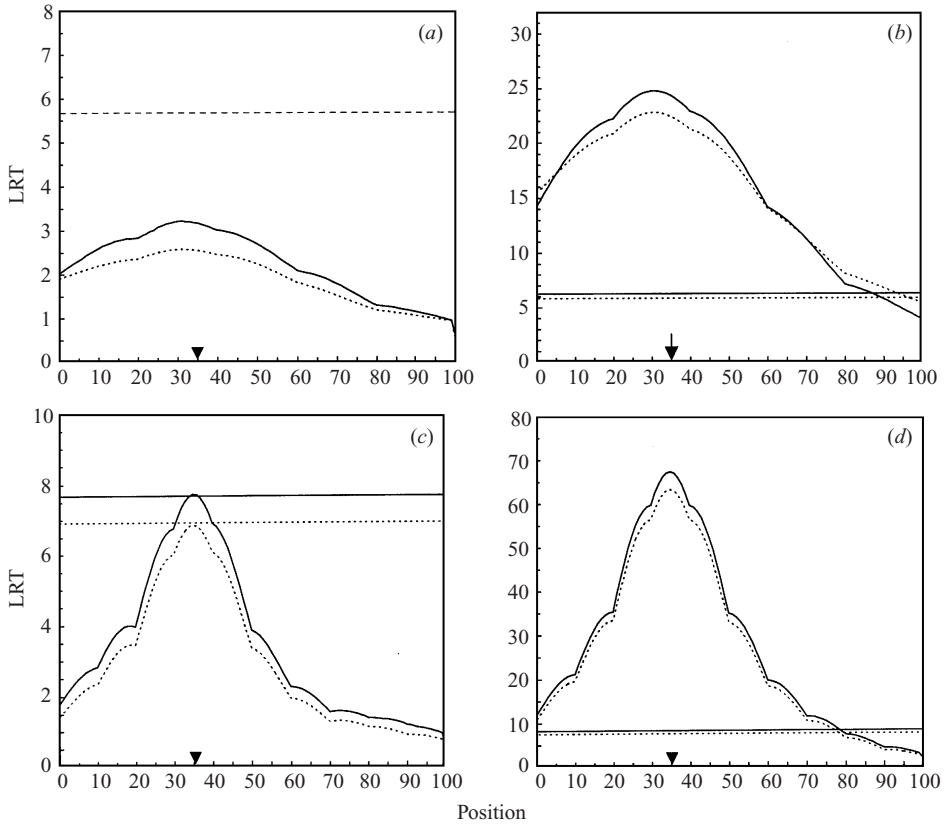

Fig. 4. Observed (continuous line) and predicted (dotted line) LRT curves. Predicted LRT is calculated by transformation of the F ratio using (12). Horizontal lines indicate 5% significance thresholds. Arrow indicates the simulated location of QTL. Scales of y-axes differ. (*a*) Small QTL, poor design; (*b*) small QTL, informative design; (*c*) large QTL, poor design; (*d*) large QTL, informative design.

Table 4. *Estimates of QTL location and power in chromosome scans (200 simulations). Significance thresholds were determined by simulation (800 replicates)*

		Mean location[a] and standard error		Power (5%)	
QTL effect	Design	LS	ML	LS	ML
Small	Poor[b]	46·5 ± 4·1	44·0 ± 3·8	30·0	30·5
	Informative[c]	38·3 ± 3·0	39·1 ± 3·1	55·0	52·0
Large	Poor[b]	33·0 ± 2·6	31·7 ± 2·4	95·0	95·5
	Informative[c]	34·8 ± 2·5	34·9 ± 2·5	100	100

[a] Simulated QTL location: 35·0.
[b] Marker positions: 0, 20, 40, 60, 80, 100. Two alleles per marker – dam alleles unidentified.
[c] Marker positions: 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100. Sixteen alleles per marker – dam alleles identified.

threshold calculations in a single-position model for 5, 10 and 20 sires respectively. The significance thresholds are not modified in any of the cases as most of the rejected values are in the bottom part of the distribution far from the tail which is relevant for the significance threshold calculations. The algorithm was also applied to significance threshold calculation in chromosome scans. For both poor and informative designs, the gain in computing time was 51%. In one case with the poor design the outcome of one run – out of 800 – was modified by the implementation of the algorithm and, in consequence, the overall significance threshold was slightly modified (5·98 vs 6·11). In consequence, the type I error (5·1%) was slightly higher than the nominal level (5·0%). As computing facilities are always a limiting factor for

https://doi.org/10.1017/S0016672398003450 Published online by Cambridge University Press

calculation of significance thresholds by bootstrap or simulation, this simplification could be used to allow an increase in the number of runs or the use of a more precise maximization algorithm.

We have demonstrated that a simple relationship between F and LRT exists when testing for a QTL in a balanced half-sib design. This relationship provides a theoretical framework to understand the similarities between the two approaches, especially in terms of power. Moreover it is demonstrated that, at a single location, the asymptotic distribution of LRT is a mixture of half $\chi^2_{(0)}$ and half $\chi^2_{(1)}$. Use of these properties in the algorithm used to calculate the significance thresholds allows for quicker simulations and, computing facilities being limited, leads to more accurate significance thresholds.

We are grateful to Ian White for helpful discussions and to Chris Haley, Bill Hill, Peter Keightley, Pierre Tilquin and Robin Thompson for their comments on an earlier version of the paper. We thank Bruce Walsh and another anonymous referee for their constructive comments. P. V. B. was funded by the BBSRC and S. A. K. was supported by the Royal Society.

References

Bovenhuis, H., van Arendonk, J. A. M., Davis, G., Elsen, J.-M., Haley, C. S., Hill, W. G., Baret, P. V., Hetzel, J. & Nicholas, F. W. (1997). Detection and mapping of quantitative trait loci in farm animals. *Livestock Production Science* **52**, 135–144.

Bulmer, M. G. (1971). The effect of selection on genetic variability. *American Naturalist* **105**, 201–211.

Doerge, R. W. (1995). Testing for linkage phase known/unknown. *Journal of Heredity* **86**, 61–62.

Elsen, J.-M., Knott, S. A., Le Roy, P. & Haley, C. S. (1997). Comparison between some approximate maximum likelihood methods for quantitative trait locus detection in progeny test designs. *Theoretical and Applied Genetics* **95**, 236–245.

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

Haley, C. S. & Knott, S. A. (1992). A simple method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315–324.

Knott, S. A., Elsen, J.-M. & Haley, C. S. (1996). Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *Theoretical and Applied Genetics* **93**, 71–80.

Le Roy, P. & Elsen, J.-M. (1995). Numerical comparison between powers of maximum likelihood and analysis of variance methods for QTL detection in progeny test designs: the case of monogenic inheritance. *Theoretical and Applied Genetics* **90**, 65–72.

Neimann-Sørensen, A. & Robertson, A. (1961). The association between blood groups and several production characteristics in three Danish cattle breeds. *Acta Agriculturae Scandinavica* **11**, 163–196.

Nelder, J. A. & Mead, R. (1965). A simplex method for function minimization. *Computer Journal* **7**, 147–151.

Stram, D. O. & Lee, J. W. (1994). Variance components testing in the longitudinal mixed model. *Biometrics* **50**, 1171–1177.

Weller, J. L., Kashi, Y. & Soller, M. (1990). Power of daughter and grand-daughter designs for determining linkage between marker loci and quantitative loci in dairy cattle. *Journal of Dairy Science* **73**, 2525–2537.

Wilks, S. S. (1938). The large sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematics and Statistics* **9**, 60–62.