Original article

Use of sib-pair linkage methods for the estimation of the genetic variance at a quantitative trait locus

H Hamann, KU Götz [*]

*Bayerische Landesanstalt für Tierzucht, Prof Dürrwaechter-Platz 1,
85586 Grub, Germany*

Summary – Until recently, the sib-pair linkage method of Haseman and Elston could only be used for the detection of linkage between a quantitative trait locus (QTL) and a marker locus. It was not possible to estimate the amount of genetic variance contributed by the QTL or its recombination fraction with the marker locus. With the advent of dense marker maps for nearly every domestic species, every QTL should be located between 2 flanking markers. In this situation, the Haseman–Elston test can be modified to estimate the variance of a putative QTL as well as its recombination fractions with the 2 flanking markers. In the present paper, we derive 2 different estimation methods for the QTL variance based on the squared performance of full sibs: in one only the QTL variance is estimated, while in the other both the QTL variance and the recombination fractions are estimated. The method that estimates only the QTL variance turns out to be more powerful than the other. With respect to the estimation of QTL variance both methods give results close to the true values. However, the estimation of recombination fractions resulted in an overall underestimation of the true parameters.

sib-pair linkage / quantitative trait locus / genetic marker / genetic variance / recombination fraction

Résumé – Emploi des méthodes d'évaluation des liaisons génétiques par les couples de germains pour estimer la variance génétique à un locus de caractère quantitatif. *Jusqu'à une période récente, le test de liaison génétique de Haseman et Elston, basé sur les couples de germains, ne pouvait être utilisé que pour la mise en évidence de liaisons entre un locus à effet quantitatif (QTL) et un locus marqueur. Il n'était pas possible d'estimer la part de la variance génétique totale liée au QTL, ni le taux de recombinaison avec le locus marqueur. Suite au développement de cartes denses dans la plupart des espèces d'élevage domestiques, chaque QTL est susceptible d'être localisé entre 2 locus marqueurs flanquants. Dans cette situation, le test de Haseman-Elston peut être modifié pour estimer*

[*] Correspondence and reprints

à la fois la variance du QTL et les taux de recombinaison avec chacun des locus marqueurs flanquants. Dans le présent article, 2 méthodes d'estimation de la variance du QTL basées sur les différences quadratiques des performances des germains sont développées : l'une n'estime que la variance du QTL, en revanche l'autre estime la variance du QTL et les 2 taux de recombinaison. Une étude de simulation permettant d'apprécier la puissance et la qualité des 2 méthodes d'estimation est présentée. La méthode permettant d'estimer la variance du QTL uniquement apparaît plus puissante que la seconde. Chaque méthode donne des résultats assez proches des vraies valeurs en ce qui concerne la variance du QTL. Les taux de recombinaison sont en revanche globalement sous-estimés.

liaison génétique par couples de germains / locus de caractère quantitatif / marqueur génétique / variance génétique / taux de recombinaison

INTRODUCTION

Haseman and Elston (1972) developed the idea of detecting linkage between a genetic marker and a quantitative trait locus (QTL) by examining the squared difference of the performance of full-sibs. Other studies (Blackwelder and Elston, 1982; Amos and Elston, 1989) showed that the method is robust against a variety of distributions of the trait examined and that it can also make use of multivariate data (Amos *et al*, 1989). Götz and Ollivier (1992) found that in animal populations, especially in pigs, the power of the method is at least comparable to that of methods based on the analysis of variance. However, the Haseman–Elston method in its original form could only detect linkage between a marker and a QTL, but could not estimate whether this was due to a QTL with large effect at a large distance, or to a QTL with small effect that is closely linked to the marker.

Studies for the establishment of a complete linkage map of the pig genome are under way (Anderson *et al*, 1993; Rohrer *et al*, 1994). This will lead to the situation that in the near future every QTL of economic importance will be in the vicinity of 2 flanking markers. This article will show how sib-pair linkage tests can be applied for the estimation of the variance caused by a QTL located between 2 flanking markers. A simulation study will be presented to examine the power and properties of the method.

MATERIALS AND METHODS

Theory

Haseman and Elston's test (1972) is based on the idea that the difference in the performance of full-sibs becomes smaller if the sibs share a larger proportion of alleles identical by descent (*ibd*) at a QTL with large effect. Elston (1990) gives a general description of the method that will only briefly be outlined here for the simplified case of a QTL with no dominance. The basic variable of the Haseman–Elston test is the squared difference (Y_j) between 2 sibs (1 and 2) within a family j:

$$Y_j = (x_{1j} - x_{2j})^2$$

Given the proportion of genes *ibd* at the QTL (π_{jt}), Elston (1990) shows that the expectation of Y_j is:

$$E(Y_j|\pi_{jt}) = \sigma_e^2 + 2\sigma_q^2 - 2\sigma_q^2\pi_{jt}$$

where σ_q^2 is the additive genetic variance due to the QTL and σ_e^2 the variance of the difference of all other genetic environmental components. Since the proportion of genes *ibd* at the QTL cannot be observed, the proportion of genes *ibd* at the linked marker locus (π_{jm}) must be used to estimate π_{jt}. The expectation of Y_j given π_{jm} is:

$$\begin{aligned}
E(Y_j|\pi_{jm}) &= \sum_{\pi_{jt}} E(Y_j|\pi_{jt})p(\pi_{jt}|\pi_{jm}) \\
&= \sigma_e^2 + 2[1 - 2\theta(1-\theta)]\sigma_q^2 - 2(1-2\theta)^2 \cdot \sigma_q^2\pi_{jm} \quad\quad [1]
\end{aligned}$$

where θ is the recombination frequency between QTL and marker locus. This is a general linear regression equation and can be written as:

$$E(Y_j|\pi_{jm}) = \alpha + \beta\pi_{jm}$$

The expectation of the regression coefficient is:

$$E(b) = -2(1-2\theta)^2\sigma_q^2 \quad\quad [2]$$

where b is an estimator of β. This expectation is zero if either σ_q^2 is zero or θ is equal to 0.5. Blackwelder and Elston (1982) showed that the distribution of the estimated regression coefficients is asymptotically normal. Thus, a simple one-sided *t*-test can be applied to test whether the regression coefficient is significantly negative. However, it can also be seen from the expectation of b that a significantly negative estimate can result from a large θ together with a large QTL effect or from a small QTL effect and tight linkage.

To estimate θ and σ_q^2, we suppose that there are 2 markers flanking the QTL. This assumption seems valid in the case where a complete marker map exists. The number of parameters to be estimated increases to 3: 2 recombination frequencies, which will be designated θ_1 and θ_2, and the QTL variance σ_q^2. The total recombination frequency between the 2 markers (θ_t) can be supposed to be known from a mapping experiment or can be estimated directly from the data.

Method I. Estimation using 2 separate tests of linkage

Two different approaches can be taken to estimate σ_q^2 in the case of 2 markers. The first approach arises in a situation where separate test linkage for 2 markers lead to significant results. If the marker loci are known to be linked, all 3 parameters can be estimated using the expectations of the 2 regression coefficients (b_1 and b_2). As a side condition, the relationship between the 2 recombination frequencies and θ_t is needed. This relationship can be assumed to be known, if assumptions about the mode of interference are made. Throughout the rest of the paper we will assume

no interference between θ_1 and θ_2.

$$E(b_1) = -2(1 - 2\theta_1)^2\sigma_q^2 \tag{3}$$

$$E(b_2) = -2(1 - 2\theta_2)^2\sigma_q^2 \tag{4}$$

$$\theta_t = \theta_1 + \theta_2 - 2\theta_1\theta_2 \tag{5}$$

Since θ_2 can be inferred from θ_1 and θ_t *via* equation [5], solutions for the 3 unknowns can be found. However, because the range of possible values for the 2 regression coefficients is theoretically between plus and minus infinity, there is not always a solution in the range of real numbers.

Method II. Estimation using the combined information of 2 markers

The second approach starts out from the fact that from equation [2] the estimator of the regression coefficient divided by -2 is already a biased estimator of σ_q^2. In the single marker case, the bias increases rapidly even for small recombination frequencies, rendering the estimator practically useless. If, however, the data is restricted to sib-pairs with the same proportion of genes *ibd* at both marker-loci ($\pi_{jm_1} = \pi_{jm_2}$), then in the majority of cases the proportion of genes *ibd* at the QTL (π_{jt}) is equal to the proportion of genes *ibd* at the 2 marker loci. This will not occur in 2 rare situations: i) in case of double recombination and the 2 recombinations take place on either side of the QTL; and ii) if 2 separate recombination events in 2 sibs take place on different sides of the QTL. Consequently, the proportion of alleles *ibd* at both marker loci is a reliable estimator of π_{jt}. The price to be paid for this is that the proportion of usable sib-pairs is reduced by a factor that can be expressed as:

$$\sum_{\pi_{m_1}} p(\pi_{m_1} = \pi_{m_2}) = \Psi_t^2 + \frac{(1 - \Psi_t)^2}{2}$$

where $\Psi_t = (1 - 2\theta_t + 2\theta_t^2)$. In the case of a 20 cM marker map and informative matings, 55% of the sib-pairs would be selected, and if the markers were at distances of 4 cM that fraction would increase to 86%.

The expectation of Y_j given a certain proportion (x) of alleles *ibd* at the marker loci is:

$$E(Y_j|\pi_{jm_1} = \pi_{jm_2} = x) = \sum_{\pi_{jt}} E(Y_j|\pi_{jt})p(\pi_{jt}|\pi_{jm_1} = \pi_{jm_2} = x)$$

which can again be written as a linear function of π_{jm_1}:

$$E(Y_j|\pi_{jm_1} = \pi_{jm_2}) = \sigma_e^2 + \sigma_q^2\left(\frac{2\Psi_1\Psi_2}{\Psi_t}\right) - 2\sigma_q^2\left(\frac{2\Psi_1\Psi_2}{\Psi_t} - 1\right)\pi_{jm_1} \tag{6}$$

where $\Psi_1 = (1 - 2\theta_1 + 2\theta_1^2)$ and $\Psi_2 = (1 - 2\theta_2 + 2\theta_2^2)$.

From this it follows that the expectation of the regression coefficient (b_0) is:

$$E(b_0) = -2\left(\frac{2\Psi_1\Psi_2}{\Psi_t} - 1\right)\sigma_q^2 \tag{7}$$

Again, $\widehat{\sigma}_q^2 = -b_0/2$ is a biased estimator of the QTL variance. Whether the bias is acceptable or not, it depends on the size of the biasing factor

$$k = \left(\frac{2\Psi_1 \Psi_2}{\Psi_t} - 1 \right)$$

in the range of realistic values for θ_t.

Figure 1 shows the value of k as a function of θ_1 for 4 different values of θ_t. The maximum bias always occurs if $\theta_1 = \theta_2$. The maximum is not equal to $\theta_t/2$ because with no interference, the 2 recombination rates do not act additively. For large values of θ_t, k can take values down to 0.93, while for smaller values the bias is negligible. Figure 2 shows that range of possible values and the expectation of k depending on θ_t. It can be seen that the expectation of k results in a bias of less than 5% over the whole range considered.
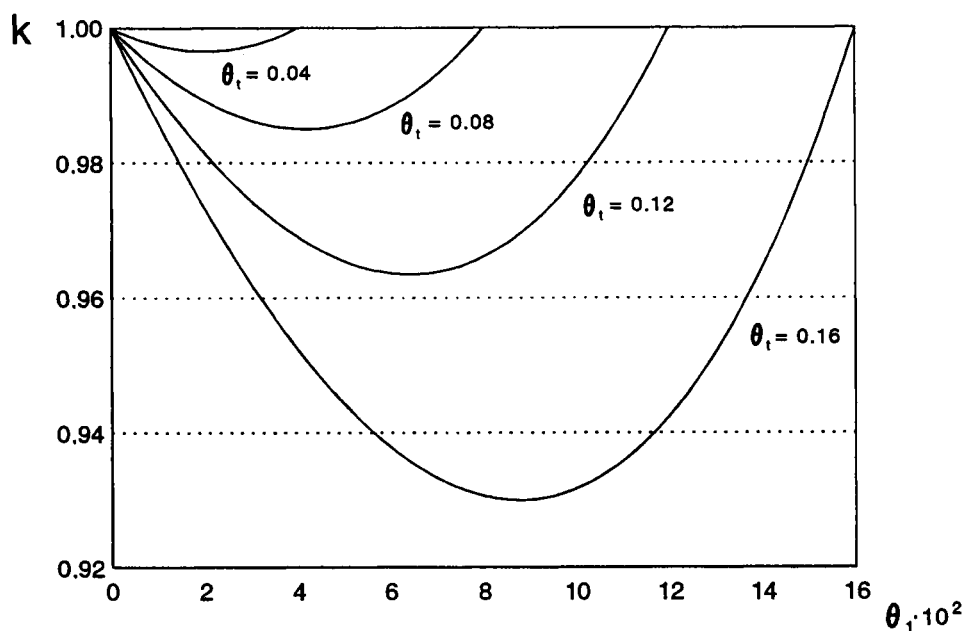
Fig 1. Value of the biasing factor k as a function of total recombination rate between the 2 markers (θ_t) and the recombination rate between the first marker and the QTL (θ_1).

The expectation of k for a given θ_t can be easily calculated. Since k is always between 0 and 1, a second estimator for the QTL variance can be derived by dividing the initial estimator by the expected value of k:

$$\widehat{\sigma}_q^{2*} = -\frac{b_0}{2E(k)} = \frac{\widehat{\sigma}_g^2}{E(k)}$$
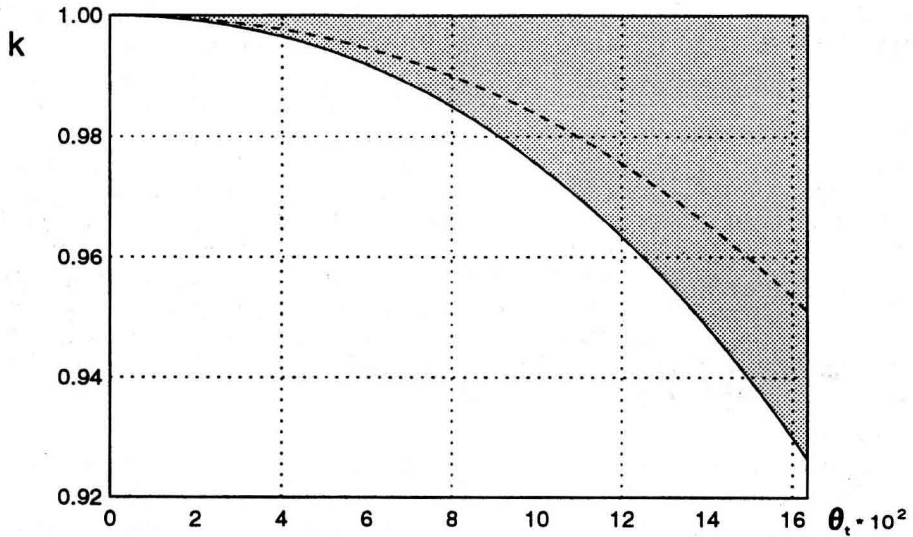
Fig 2. Range of possible values for k (shaded area) and expected value of k ($E(k)$) (dotted line) as a function of total recombination rate between the 2 markers (θ_t).

where $E(k)$ is given by:

$$E(k) = \frac{1}{\theta_t} \int_{\theta_1=0}^{\theta_t} k(\theta_1, \theta_t) d\theta_1 = 1 - \frac{4\theta_t^2}{3\Psi_t}$$

Simulation

A simulation study was conducted in order to examine the power of the 2 methods and the goodness of estimation. Data were simulated according to the following model:

$$x_{ij} = \mu + q_{ij} + \frac{1}{2}(bv_{s_j} + bv_{d_j}) + \phi_{ij} + ce_j + e_{ij}$$

where:

x_{ij} = phenotypic value of animal i in family j

μ = overall mean

q_{ij} = effect of the QTL genotype of animal i

bv_{s_j} = sire's contribution to polygenic breeding value (without QTL genotype)

bv_{d_j} = dam's contribution to polygenic breeding value (without QTL genotype)

ϕ_{ij} = Mendelian sampling effect

ce_j = effect of common litter environment

e_{ij} = residual error

For the constant parameters in the simulation, the following values were used: total phenotypic variance was set to 1 000, the heritability of the trait was 0.3

(including the QTL effect) and common environmental variance was 0.2. The population structure simulated was that of a typical pig-breeding situation with 25 sires, 10 dams per sire and 8 progeny per sire-dam pair. Thus, a total of 2 000 progeny were simulated in each replication. For a discussion of the effects of the mating structure and common environment, see Götz and Ollivier (1992).

Götz and Ollivier (1992) found that the use of fully informative matings can increase the power of the Haseman–Elston test for a given number of genotypings. Consequently, only these matings were used in the calculations. This has no consequence for the validity of the results, but it should be borne in mind that the number of genotyped individuals in practice would be slightly higher than 2 275. Within any family, all possible differences between full-sibs were used for the calculation of Y_js as proposed by Blackwelder and Elston (1982). This resulted in 28 comparisons per family and 7 000 comparisons per round of simulation.

Variable parameters in the simulation were:
(i)   the distance between the 2 markers (θ_t)
(ii) the position of the QTL between the 2 markers as expressed by θ_1 and θ_2
(iii) the size of the QTL effect

The distance between the 2 markers was varied approximately between 0.04 and 0.154, assuming no interference. The combinations of θ_1 and θ_2 that were simulated are given in table I.

Table I. Combinations of θ_1 and θ_2 used in the simulations.

θ_2	θ_1			
	0.02	*0.04*	*0.06*	*0.08*
0.02	x			
0.04		x		
0.06	x		x	
0.08		x		x
0.10	x		x	
0.12		x		
0.14	x			

Two codominant alleles with equal frequencies were assumed at the QTL. For both marker loci 10 alleles with equal frequencies were simulated and for the QTL effect genetic variances of 40, 80 and 120 were assumed. This resulted in a total of 30 different variants, each of them being simulated with 1 000 replications.

Analysis of simulation results

The power of the methods was defined as the percentage of replications where the null hypothesis was rejected at the 5% level. For *Method II* this approach is unambiguous while this is not the case for *Method I*. For the first method there are 2 null hypotheses of which 1 or both can be rejected at $\alpha = 0.05$. Since both tests rely on the same values for Y_j, they are not independent so that the

nominal type I errors for a global error of 5% can only be determined by simulation under the null hypothesis. However, these type I errors still depend on the 2 recombination frequencies so that the true state of nature must be known for an exact determination. Therefore, it was decided that a replication was significant for *Method I* if both null hypotheses were rejected at a 5% level. For the interpretation of the results it should be borne in mind that *Method I* has a slight disadvantage.

For the estimation with *Method II* only sib-pairs with the same proportion of alleles *ibd* at both marker loci were selected from the same data that were used for the estimation with *Method I*. As was explained previously, this results in a reduction of the number of effective sib-pairs of between 15% (for $\theta_1/\theta_2 = 0.02/0.02$) and 42% (for $\theta_1/\theta_2 = 0.02/0.14$).

To assess the goodness of the estimation, all replications of a certain variant (significant and non-significant) must be averaged. As can be seen from equations [3] and [4], the first method requires the square-root of the ratio of b_1 and b_2 for the estimation of σ_q^2. In practical applications this is not likely to cause problems, since significant regression coefficients always have a negative sign. In a simulation with a low value for the QTL effect, however, this causes problems because the estimated regression coefficients are normally distributed and a certain fraction can be expected with positive values. For these replicates a value for σ_q^2 cannot be estimated. Because regression coefficients at positive values are all non-significant, the missing QTL variances cause an overestimation of this parameter.

RESULTS

Power of the 2 methods

Table II shows the power of the 2 methods of estimation for all simulated variants. For a QTL effect of 40, the power is low for both methods and all variants. However, it can be observed that *Method II* has higher power in all variants and that the decrease in power with increasing θ_t is less for the second method. For a QTL effect of 80, the superiority of the second method is evident. The superiority is more pronounced if the values of θ_1 and θ_2 are unequal, which is caused by an increasing proportion of replicates where only one of the 2 tests in *Method I* gives a significant result. If the QTL effect is 120, both methods have high power with differences occurring only if the 2 recombination rates were of very different size.

Estimation of θ_1, θ_2 and σ_q^2 using Method I

The average estimated values for σ_q^2 are given in table III for the 2 methods. For low values of σ_q^2 an overestimation occurs, which is caused by the fact that a certain number of replications could not be calculated for reasons mentioned above. This could be as much as 24% of the replicates. With σ_q^2 equal to 80 and 120, the percentage of replicates without result decreased to 10 and 3%, respectively. In accordance with these numbers, the overestimation is less with increasing QTL variance and decreasing recombination fractions within QTL variance. For a QTL variance of 120 some slight underestimations occur with higher recombination

Table II. Power (in percent) of the 2 methods for different QTL effects and combinations of θ_1 and θ_2.

θ_1/θ_2	QTL effect					
	40		80		120	
	Method I	Method II	Method I	Method II	Method I	Method II
0.02/0.02	24	31	68	75	94	95
0.02/0.06	18	28	56	67	88	93
0.02/0.10	15	27	43	61	77	89
0.02/0.14	12	25	36	56	59	84
0.04/0.04	21	31	57	66	89	93
0.04/0.08	14	26	51	64	80	90
0.04/0.12	9	22	42	60	67	84
0.06/0.06	14	26	49	63	82	88
0.06/0.10	12	24	39	55	72	82
0.08/0.08	11	25	39	56	74	83

fractions. In accordance with the fact that most of the dropouts occur if the 2 recombination fractions are of very different sizes, the worst estimates are achieved if the QTL is located close to 1 of the 2 flanking markers.

The estimated values for the recombination fractions equally suffer from the problem of replications without solution. In contrast to the estimation of QTL variance, this leads to an underestimation of recombination fractions for small values of σ_q^2. Table IV shows that for a QTL variance of 40 the estimators are heavily biased downwards. This improves with increasing values for σ_q^2. A remarkable decrease of the standard deviation of the estimates can also be observed. However, none of the estimates are very precise, mainly due to the low expected numbers of double recombinants within the 2 000 progeny.

Estimation of σ_q^2 using Method II

The estimates for σ_q^2 using the second method are also presented in table III. From theory, *Method II* is expected to underestimate the true value of the parameter. This expectation is confirmed by the results with a single exception.

Especially for QTL variance of 40 the estimates are clearly superior to those of *Method I*. For larger values of θ_t the underestimation gets larger but stays within the range that can be explained by the decreasing value of k.

The results for the second estimator ($\widehat{\sigma}_q^{2*}$) are also given in table III as *Method IIc*. On average, this manipulation reduces the underestimation of σ_q^2 from about 3% to less than 1%.

H Hamann, KU Götz

Table III. Estimated values for the QTL variance using *Methods I*, *II* and *Method II* with correction for $E(k)$ (*Method IIc*) (averages of 1 000 replicates).

θ_1/θ_2	QTL variance								
	40			80			120		
	Method I	Method II	Method IIc	Method I	Method II	Method IIc	Method I	Method II	Method IIc
0.02/0.02	41.9	39.1	39.2	79.9	79.0	79.2	120.1	120.0	120.3
0.02/0.06	42.8	38.1	38.4	79.7	78.0	78.8	118.7	118.0	119.1
0.02/0.10	45.5	40.0	40.9	79.0	77.3	79.0	119.8	118.3	120.9
0.02/0.14	47.4	40.0	41.7	83.3	77.2	80.5	117.2	114.8	119.7
0.04/0.04	46.1	40.9	41.2	80.5	77.5	78.1	119.3	118.4	119.5
0.04/0.08	45.1	37.8	38.6	82.1	78.3	80.0	120.1	118.5	121.0
0.04/0.12	44.9	37.4	38.9	85.1	79.9	83.1	117.7	115.4	120.0
0.06/0.06	44.4	37.1	37.8	80.9	76.4	78.0	117.9	114.7	117.0
0.06/0.10	46.1	39.4	40.9	82.0	75.7	78.6	118.8	112.5	116.9
0.08/0.08	45.1	38.6	40.1	80.1	74.4	77.3	116.2	110.9	115.3
SD of estimates	32.6	34.0	34.0	34.4	36.4	36.5	39.0	40.2	40.3
(min–max)	−38.7	−42.6	−44.2	−40.8	−45.4	−47.2	−45.2	−47.4	−49.2

Table IV. Estimated values for recombination rates (θ_1/θ_2) (*Method I*, averages of 1 000 replicates) and their standard deviations (SDs).

| θ_1/θ_2 | QTL variance | | | | | |
| | 40 | | 80 | | 120 | |
	$\widehat{\theta}_1/\widehat{\theta}_2$	SDs	$\widehat{\theta}_1/\widehat{\theta}_2$	SDs	$\widehat{\theta}_1/\widehat{\theta}_2$	SDs
0.02/0.02	0.014/0.009	0.10/0.10	0.019/0.015	0.05/0.05	0.019/0.019	0.02/0.04
0.02/0.06	0.013/0.039	0.12/0.13	0.006/0.060	0.09/0.07	0.015/0.060	0.04/0.04
0.02/0.10	0.025/0.068	0.13/0.12	0.003/0.096	0.13/0.09	0.007/0.101	0.09/0.06
0.02/0.14	0.016/0.107	0.18/0.12	0.003/0.132	0.14/0.07	0.000/0.143	0.10/0.05
0.04/0.04	0.028/0.025	0.14/0.11	0.032/0.036	0.08/0.09	0.037/0.038	0.04/0.04
0.04/0.08	0.032/0.056	0.13/0.13	0.027/0.076	0.10/0.08	0.034/0.078	0.07/0.05
0.04/0.12	0.043/0.075	0.15/0.15	0.024/0.115	0.11/0.07	0.023/0.122	0.10/0.05
0.06/0.06	0.045/0.046	0.12/0.11	0.052/0.053	0.08/0.08	0.057/0.054	0.06/0.05
0.06/0.10	0.051/0.071	0.14/0.14	0.050/0.091	0.09/0.09	0.053/0.095	0.07/0.06
0.08/0.08	0.061/0.061	0.13/0.16	0.072/0.068	0.09/0.12	0.074/0.077	0.06/0.06

DISCUSSION

The present study has shown that with 2 flanking markers for a given QTL, the principle of sib-pair linkage methods can be applied to obtain estimates of the QTL variance and to locate the QTL in the interval. The simulation study made use of the results of Götz and Ollivier (1992), who showed that in animal breeding the preselection of fully informative matings is an appropriate way to improve the power of QTL-detection for a given number of genotypings. If many markers are to be examined, the parents will only be informative for a fraction of the markers. Since the major costs in the given design arise from the typing of progeny, these should only be typed for the markers where their parents are informative. With *Method II* it should also be possible to use multiple markers as proposed by Haley *et al* (1994). However, without modification this only seems feasible if the linkage phases in the parents are known, recombinant sibs are excluded from the analysis and the markers are not so far apart that double recombinants become important.

The results on power for the detection of a given QTL suggest that a QTL contributing 8% of the phenotypic variance can be detected with a power between 55 and 75% for the given design. In comparison with the results of Götz and Ollivier (1992) it must be taken into account that in the present study the QTL effect was included in the total genetic variance. In a comparable situation the power of both methods presented here is less than in Götz and Ollivier (1992). The reasons are that the type I error for *Method I* is not comparable to the 5% level in the previous study and that *Method II* uses fewer sib-pairs. The power of *Method II* is superior to that of *Method I* in all of the simulated variants. The reason is evident, since *Method I* does not use the prior information that the 2 marker loci are linked with a known recombination fraction for the test of linkage. This information is only used in the estimation step, given that linkage of both loci to the QTL was detected.

Future research should be directed towards a way of incorporating the information of linkage between the markers in the detection of linkage with a QTL.

One way to do this has recently been presented by Fulker and Cardon (1994). They used the information of 2 markers to estimate π_t and regressed Y_j on this estimated value. Since the estimation only works if θ_1 and θ_2 are known, they use an approach similar to interval mapping (Lander and Botstein, 1989) to plot the t-statistics against the putative QTL position. The authors also encountered problems in trying to determine the correct t-value for a certain type I error rate, even for the single interval case, which they solved by simulation under the null hypothesis. However, in every realistic scenario this problem will occur since usually many markers will be examined at a time. In this situation, none of the test statistics has a simple distribution and one would always have to escape to simulation studies in order to examine the distribution of the test statistic.

A comparison of the results of Fulker and Cardon (1994) at $h^2 = 0.125$ with our results (*Method II*) for a QTL variance of 120 shows little difference. This indicates that sibs with different percentages of alleles *ibd* at the 2 marker loci contribute little or nothing to the estimation of π_t.

In the estimation of the QTL variance, *Method I* is characterized by the overestimation of σ_q^2 if the true value is small. In practice, this is not likely to be a problem, since significant regression coefficients are always negative, but it makes it difficult to prove the unbiasedness of the estimator. However, for a QTL variance of 120 no overestimations occurred and underestimations were in all cases less than 3%. *Method II* is *a priori* a biased estimator. The results show that the bias is small if the QTL is located close to one of the markers and that the estimation of σ_q^2 can be improved by dividing the initial estimator by $E(k)$. The maximum bias can also be quantified if the recombination fraction between the 2 markers is known. However, since *Method I* gives similar estimates and information about the location of the QTL, one could use *Method II* to detect simultaneous linkage of 2 markers with a QTL and then use *Method I* for the estimation. Unfortunately, Fulker and Cardon (1994) give little information about the quality of their estimator of the QTL variance. The only result they give indicates that their algorithm leads to an overestimation of σ_q^2.

The estimation of recombination frequencies between the QTL and the markers leads to unsatisfactory results. The majority of recombination fractions were underestimated, although for higher QTL effects the estimators came close to the true values. Non-estimable replicates certainly influenced these results as well. The same observation can be made from the results of Fulker and Cardon (1994) which show relatively flat curves in the vicinity of the true QTL location and a tendency to place the QTL in the middle of the interval. In comparison, our *Method I* tends to locate the QTL closer to the marker with the smaller recombination fraction.

Knott and Haley (1992) examined the application of maximum likelihood (ML) in outbreeding populations with a full-sib structure. The advantage of ML is the fact that it is possible to estimate the gene effects at the QTL as well as the gene frequencies. However, the computational effort is much higher for ML than for the approach in the present paper. The authors conclude that for the treatment of realistic population structures and the inclusion of fixed effects, numerical approximations are needed to render practical data tractable by ML.

Haley and Knott (1992) presented a method for the mapping of QTLs by regression in crosses between inbred lines. This design is not tractable with our methods because of the complete linkage disequilibrium in an F_2 derived from inbred lines. However, those authors found that there 'seemed little advantage to be gained from resort to maximum likelihood methods for the analysis of these types of data'. The method is similar to that of Fulker and Cardon (1994) since both methods are based on the idea of interval mapping (Lander and Botstein, 1989).

CONCLUSIONS

The extension of Haseman and Elston's (1972) method of sib-pair linkage presented here allows for the estimation of QTL variance and recombination fractions if a relatively dense and informative marker map is available. Since the method uses only intra-family comparisons it does not need to take fixed effect into account so long as they affect all sibs in a family in the same way.

However, there are some limitations of the method that shall be mentioned here. The first is that the results presented rely heavily on the availability of highly polymorphic markers. If one or both of the markers is not very polymorphic, the number of parents to be typed increases dramatically. For a discussion of this topic see Götz and Ollivier (1992). The second limitation is the dependency of sib-pair linkage tests on the magnitude of the residual variance. For traits with low heritability the power of the method is low, while high heritability and common environmental effects are favourable. In addition, large family sizes increase the power of the method (Götz and Ollivier, 1992).

In outbreeding populations the detection of segregating QTLs is generally more difficult than in crosses between inbred lines (Knott and Haley, 1992). However, the detected QTLs are known to be segregating while with QTLs detected in crossing experiments a large fraction of favourable QTLs will already be fixed in the superior line. It is doubtful whether preselection of fully informative matings will still work if many markers are to be examined, since parents will be informative only for a fraction of markers. Nevertheless, it would be possible to type the progeny only for those markers where their parents are informative.

ACKNOWLEDGMENTS

REFERENCES

Amos CI, Elston RC (1989) Robust methods for the detection of genetic linkage for quantitative data from pedigrees. *Genet Epidem* 6, 349-360

Amos CI, Elston RC, Wilson AF, Bailey-Wilson JE (1989) A more powerful robust sib-pair test of linkage for quantitative traits. *Genet Epidemiol* 6, 435-449

Anderson L, Archibald AL, Gellin J, Schook LB (1993) 1st pig gene mapping workshop (PGM1), August 7, 1992, Interlaken, Switzerland. *Anim Genet* 24, 205-216

Blackwelder WC, Elston RC (1982) Power and robustness of sib-pair linkage tests and extension to larger sibships. *Commun Statist-Theor Methods* 11, 449-484

Elston RC (1990) A general linkage method for detection of major genes. *In: Advances in Statistical Methods Applied to Livestock Production* (K Hammond, D Gianola, eds), Springer, Berlin, Germany

Fulker DW, Cardon LR (1994) A sib-pair approach to interval mapping of quantitative trait loci. *Am J Hum Genet* 54, 1092-1103

Götz KU, Ollivier L (1992) Theoretical aspects of applying sib-pair linkage tests to livestock species. *Genet Sel Evol* 24, 29-42

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69, 315-324

Haley CS, Knott SA, Elsen JM (1994) Multi-marker approaches to mapping quantitative trait loci in livestock. Proc 45th EAAP-Meeting, Edinburgh, UK, 5-8 September 1994

Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 1, 3-19

Knott SA, Haley CS (1992) Maximum likelihood mapping of quantitative trait loci using full-sib families. *Genetics* 132, 1211-1222

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

Rohrer GA, Alexander LJ, Keele JW, Smith TP, Beattie CW (1994) A microsatellite linkage map of the porcine genome. *Genetics* 136, 231-245