Alternative models for QTL detection in livestock. II. Likelihood approximations and sire marker genotype estimations

B. Mangin, B. Goffinet, P. Le Roy, D. Boichard, J.M. Elsen

Genet. Sel. Evol. 31 (1999) 225–237

Original article

Alternative models for QTL detection in livestock. II. Likelihood approximations and sire marker genotype estimations

Brigitte Mangin[a], Bruno Goffinet[a], Pascale Le Roy[b], Didier Boichard[b], Jean-Michel Elsen[c*]

[a] Biométrie et intelligence artificielle, Institut national de la recherche agronomique, BP27, 31326 Castanet-Tolosan, France
[b] Station de génétique quantitative et appliquée, Institut national de la recherche agronomique, 78352 Jouy-en-Josas, France
[c] Station d'amélioration génétique des animaux, Institut national de la recherche agronomique, BP27, 31326 Castanet-Tolosan, France

(Received 20 November 1998; accepted 7 April 1999)

Abstract – In this paper, we compare four different methods of dealing with the unknown linkage phase of sire markers which occurs in the detection of quantitative trait loci (QTL) in a half-sib family structure when no information is available on grandparents. The methods are compared by considering a Gaussian approximation of the progeny likelihood instead of the mixture likelihood. In the first simulation study, the properties of the Gaussian model and of the mixture model were investigated, using the simplest method for sire gamete reconstruction. Both models lead to comparable results as regards the test power but the mean square error of sib QTL effect estimates was larger for the Gaussian likelihood than for the mixture likelihood, especially for maps with widely spaced markers. The second simulation study revealed that the simplest method for sire marker genotype estimation was as powerful as complicated methods and that the method including all the possible sire marker genotypes was never the most powerful. © Inra/Elsevier, Paris

half-sib family / QTL detection / unknown linkage phase / Gaussian approximation / log-likelihood ratio test

Résumé – Modèles alternatifs pour la détection de QTL dans les populations animales. II. Approximations de la vraisemblance et estimations du génotype des mâles aux marqueurs. Dans ce papier, nous comparons quatre méthodes, qui permettent de résoudre le problème relatif à la phase inconnue des mâles

* Correspondence and reprints
E-mail: elsen@toulouse.inra.fr

aux marqueurs dans des familles de demi-germains, lorsque aucune information sur les grands-parents n'est disponible. Ces méthodes sont comparées, en utilisant l'approximation gaussienne de la vraisemblance à l'intérieur de chaque descendance à la place de la vraisemblance du mélange de distribution. Dans la première étude par simulation, les propriétés respectives du modèle gaussien et du modèle de mélange sont étudiées pour la méthode la plus simple de reconstruction des gamètes des mâles. Les deux modèles conduisent à des tests comparables au regard de leur puissance mais l'erreur quadratique moyenne d'estimation de l'effet de substitution du QTL intra-famille est plus grande pour le modèle gaussien que pour le modèle de mélange, en particulier pour les cartes génétiques très peu denses. La deuxième étude par simulation montre que la plus simple méthode d'estimation du génotype des mâles aux marqueurs est aussi puissante que les méthodes plus sophistiquées et que la méthode qui consiste à prendre en compte dans la vraisemblance tous les génotypes possibles d'un mâle aux marqueurs n'est jamais la plus puissante.

famille de demi-frères / détection de QTL / phase de linkage inconnue / approximation gaussienne / test du rapport de vraisemblance

1. INTRODUCTION

The present paper deals with the detection of one QTL in half-sib families when no information is available on grandparents.

A general form of the likelihood of detecting QTL in simple pedigree structures such as half-sib or full-sib families when marker information is available on progeny, parents and grandparents was presented by Elsen et al. [2]. This likelihood is a two-level mixture distribution with different possible sire marker genotypes given marker information, and different possible progeny QTL genotypes given sire marker genotype and offspring marker information. This paper describes simulations carried out to compare simplified likelihoods.

As an alternative to the mixture approach, we suggest simplifying the likelihood by considering only one sire marker genotype. Three solutions were explored: the first one, close to the Knott et al. proposal [7], is the likelihood of quantitative phenotypes conditional on the most probable sire marker genotype given marker information, while in the others, the sire marker genotype is treated as a fixed effect, estimating the likelihood of the quantitative trait observation conditionally or jointly with the sire marker genotype.

These comparisons were performed on a simplified form of the likelihood with regard to the mixture of the progeny QTL genotypes. This simplified likelihood is the one used in interval mapping by linear regression [5, 8] but instead of least squares tests as in the above papers, maximum log-likelihood ratio tests were used. The properties of this simplification are described in the first part of the paper, using the likelihood of the quantitative phenotypes conditional on the most probable sire marker genotype given marker information.

2. COMPARISON OF LIKELIHOOD AND SIMPLIFIED LIKELIHOOD

Most hypotheses and notations are given in Elsen et al. [2]. Notations related to this paper are summarized in *table I.*

Table I. Notations for half-sib family observations.

Notation	Signification
i for $i = 1, \ldots, n$	sire
ij for $j = 1, \ldots, n_i$	progeny (one progeny per mate)
l for $l = 1, \ldots, L$	marker locus
ms_i^{l1}, ms_i^{l2}	marker alleles for sire i at locus l
$mp_{ij}^{l1}, mp_{ij}^{l2}$	marker alleles for descendant ij at locus l
$ms_i = \{ms_i^{l1}, ms_i^{l2}\}_{l=1,\ldots,L}$	$L \times 2$ matrix of marker phenotypes for sire i
$mp_{ij} = \{mp_{ij}^{l1}, mp_{ij}^{l2}\}_{l=1,\ldots,L}$	$L \times 2$ matrix of marker phenotypes for progeny ij
M_i	Marker information for sire i family
yp_{ij}	trait data for descendant ij
hs_i^1	chromosome transmitted by the grand sire to the sire i
hs_i^2	chromosome transmitted by the grand dam to the sire i
$hs_i = \{hs_i^1, hs_i^2\}$	$L \times 2$ matrix of marker genotypes for sire i
x	QTL position
$d_{ij}^x = 1$ or 2	sire allele received by progeny ij at location x
μ_i^{x1} or μ_i^{x2}	means of trait distribution at location x
$\phi(yp_{ij}; \mu_i^{xq}, \sigma_e^2)$	normal penetrance function, conditional to the q ($q = 1$ or 2) chromosome segment transmitted by the sire

Let hs, μ^{x1}, μ^{x2} denote the vectors of sire marker genotypes hs_i and of phenotypic means of trait distribution μ_i^{x1}, μ_i^{x2}. Let Λ_0 be the likelihood under the null hypothesis that no QTL is segregating in the pedigree

$$\Lambda_0 = \prod_{i=1}^{n} \prod_{j=1}^{n_i} \phi(yp_{ij}; \mu_i, \sigma_e^2)$$

where μ_i is the phenotypic mean of sire i offspring. Let μ be the vector of μ_i.

2.1. Test statistics

The general form of the likelihood presented by Elsen et al. [2] is

$$\Lambda^x = \prod_{i=1}^{n} \Lambda_i^x = \prod_{i=1}^{n} \sum_{hs_i} p(hs_i/M_i) \Lambda_i^{x,hs_i}$$

$$= \prod_{i=1}^{n} \sum_{hs_i} p(hs_i/M_i) \prod_{j=1}^{n_i} \sum_{q=1}^{2} p(d_{ij}^x = q/hs_i, M_i) f(yp_{ij}/d_{ij}^x = q)$$

That leads to the maximum log-likelihood ratio test

$$T = 2\left(\sup_{x,\mu^{x1},\mu^{x2},\sigma_e^2} \sum_{i=1}^{n} \log\left(\sum_{hs_i} p(hs_i/M_i)\Lambda^{x,hs_i}\right) - \sup_{\mu,\sigma_e^2} \log(\Lambda_0)\right)$$

Full maximum likelihood for this type of likelihood requires a lot of computation because the number of possible sire marker genotypes hs_i, in the first summation, grows exponentially with the number of informative markers per sire. *Table II* presents for T and the other tests proposed in this paper, the CPU time needed for one simulation. Although our program could certainly be optimized, these results show that computing T test is possible for one data set but cannot reasonably be considered for simulations; simulations that are generally needed to obtain significant thresholds.

Table II. CPU time (in seconds) for one simulation for all the log-likelihood ratio tests.

Number of descendants	Number of markers	T	T^1	T^2	T^3	T^4	T^5
50	11	800	280	4.5	6	6	11.5
20	3	675	165	1	2	2.5	10

For T and T^1 a mixture of distributions in progeny is used. For T^2, T^3, T^4 and T^5 a Gaussian distribution in progeny is used.

A natural way of dealing with this difficulty is to work in two steps: in the first step a probable marker genotype for each sire is estimated and in the second step the part of the likelihood corresponding only to these probable marker genotypes is maximized.

A possible estimate for the sire marker genotypes, very close to the sire gamete reconstruction proposed by Knott et al. [7] may be based on

$$\widehat{hs_i} = \text{argmax}_{hs_i} p(hs_i/M_i)$$

Let $\widehat{hs}$ be the vector of estimated sire marker genotypes. For the second step, the likelihood is reduced to

$$\Lambda^{x,\widehat{hs}} = \prod_{i=1}^{n} \Lambda_i^{x,\widehat{hs_i}} = \prod_{i=1}^{n} \prod_{j=1}^{n_i} \sum_{q=1}^{2} p(d_{ij}^x = q/\widehat{hs_i}, M_i) f(yp_{ij}/d_{ij}^x = q)$$

In order to simplify the maximization step, the mixture of distributions in progeny can be approximated by a normal distribution with expectation equal to the expectation of the mixture. Then a linear model is obtained at each position x along the chromosome. Let $\widetilde{\Lambda}^{x,\widehat{hs}}$ denote this simplified likelihood equal to

$$\widetilde{\Lambda}^{x,\widehat{hs}} = \prod_{i=1}^{n} \widetilde{\Lambda}^{x,\widehat{hs_i}} = \prod_{i=1}^{n} \prod_{j=1}^{n_i} \phi(yp_{ij}; \sum_{q=1}^{2} p(d_{ij}^x = q/\widehat{hs_i}, M_i)\mu_i^{xq}, \sigma_e^2)$$

A simulation study was carried out to compare the power of QTL detection, using maximum log-likelihood ratio tests, T^1 and T^2 where

$$T^1 = 2\,(\sup_{x,\mu^{x1},\mu^{x2},\sigma_e^2} \log(\Lambda^{x,\widehat{hs}}) - \sup_{\mu,\sigma_e^2} \log(\Lambda_0))$$

$$T^2 = 2\,(\sup_{x,\mu^{x1},\mu^{x2},\sigma_e^2} \log(\widetilde{\Lambda}^{x,\widehat{hs}}) - \sup_{\mu,\sigma_e^2} \log(\Lambda_0))$$

2.2. Simulation results

Sire designs with 20 sire families of 50 or 20 descendants per sire were simulated. The linkage group comprised three or eleven equally spaced markers, each with two alleles segregating at equal frequency in the population. Polygenic heritability was fixed at 0.2 and residual variability at 1. The power studies were based on a QTL with two alleles at equal frequency, located either at 5 or 35 cM from one end of the linkage group with additive effect equal either to 0.5 or to 1 and no dominance.

2.2.2. Threshold and power

The null distributions of the test statistics were estimated simulating data sets with polygenic effects corresponding to the heritability value used in the simulation model. Significant thresholds for T^1 and T^2 are shown in *table III*. The largest difference between the test powers, shown in *table IV*, was observed for a 20 half-sib progeny design, an 11 marker map and a QTL located at 35 cM with an additive effect equal to 1. In this situation, a gain of about 10 % was obtained with the mixture likelihood as compared to the Gaussian likelihood. However, other cases did not show large differences and either the first or the second test may be the most powerful depending on the case studied.

In the back-cross design, these tests have been proven to be asymptotically equivalent when the QTL effect is small [9]. In order to limit computing time the Gaussian approximation only will be considered in the second part of this paper and in its companion paper [4]. Methods and simulation results given with the Gaussian approximation may be extended to include a mixture of distributions.

2.2.2. Parameter estimates

Despite power results that were quite similar for both methods, it is worthwhile comparing parameter estimates for the QTL location and sib QTL effect.

Mean estimates of position and of empirical standard deviation of the position estimate are shown in *table V*. Obviously, due to the fact that the position estimate is constrained in order to belong to the chromosome, its bias was found to be more important for a QTL located at the beginning of the chromosome than for a QTL located near the middle of the chromosome, but both methods gave similar bias. Standard deviations of the position estimates

Table III. Empirical 5 % significant thresholds for T^1, T^2, over 1 000 replications.

Number of descendants	Number of markers	T^1	T^2
50	11	39.40	38.22
20	11	43.51	40.14
50	3	40.56	35.94
20	3	45.56	37.49

Table IV. Percentage of replicates significant at the empirical 0.05 significant threshold for T^1, T^2, over 500 replications.

Number of descendants	Number of markers	QTL position	QTL additive effect	T^1	T^2
50	11	0.05	0.5	51.6	51.0
50	11	0.35	0.5	59.0	62.8
50	11	0.05	1.0	98.2	97.4
50	11	0.35	1.0	99.4	99.4
20	11	0.05	0.5	14.2	16.2
20	11	0.35	0.5	16.4	18.0
20	11	0.05	1.0	56.0	57.6
20	11	0.35	1.0	75.6	68.8
50	3	0.05	0.5	24.8	25.4
50	3	0.35	0.5	20.2	21.0
50	3	0.05	1.0	78.8	74.6
50	3	0.35	1.0	72.4	70.6
20	3	0.05	0.5	8.8	11.2
20	3	0.35	0.5	9.6	9.6
20	3	0.05	1.0	27.2	27.8
20	3	0.35	1.0	26.0	25.8

were slightly larger for a Gaussian likelihood than for a mixture likelihood for the more widely spaced marker map but they were comparable for the other map studied.

Mean square errors of the within half-sib QTL substitution effect are shown in *table VI*.

As the bias of $\widehat{\alpha_i^x}$ is small (data not shown), the mean square error is closely related to

$$\sum_{i=1}^{n} \mathrm{Var}(\widehat{\alpha_i^x})/n$$

Results for the Gaussian likelihood in the 11 equally spaced marker maps may be explained by considering the idealized case where the QTL position is known and located on a marker and for which all sires are heterozygous for this marker. The variance of $\widehat{\alpha_i}$ depends only on the number of informative descendants per sire. For a marker with two alleles at equal frequency, the

Table V. Mean estimates of the QTL position (and their empirical standard deviation) for T^1, T^2, over 500 replications.

Number of descendants	Number of markers	QTL position	QTL additive effect	T^1		T^2	
50	11	0.05	0.5	0.206	(0.269)	0.206	(0.257)
50	11	0.35	0.5	0.377	(0.216)	0.390	(0.220)
50	11	0.05	1.0	0.079	(0.116)	0.075	(0.096)
50	11	0.35	1.0	0.348	(0.082)	0.348	(0.083)
20	11	0.05	0.5	0.363	(0.348)	0.354	(0.333)
20	11	0.35	0.5	0.423	(0.304)	0.413	(0.287)
20	11	0.05	1.0	0.155	(0.230)	0.182	(0.253)
20	11	0.35	1.0	0.353	(0.186)	0.352	(0.184)
50	3	0.05	0.5	0.349	(0.330)	0.311	(0.367)
50	3	0.35	0.5	0.480	(0.306)	0.459	(0.366)
50	3	0.05	1.0	0.125	(0.146)	0.137	(0.228)
50	3	0.35	1.0	0.424	(0.224)	0.405	(0.280)
20	3	0.05	0.5	0.431	(0.336)	0.424	(0.401)
20	3	0.35	0.5	0.487	(0.310)	0.475	(0.377)
20	3	0.05	1.0	0.286	(0.292)	0.293	(0.360)
20	3	0.35	1.0	0.447	(0.268)	0.445	(0.336)

Table VI. Mean square error of the within half-sib QTL substitution effect for T^1, T^2, over 500 replications.

Number of descendants	Number of markers	QTL position	QTL additive effect	T^1	T^2
50	11	0.05	0.5	0.212	0.233
50	11	0.35	0.5	0.186	0.193
50	11	0.05	1.0	0.234	0.278
50	11	0.35	1.0	0.188	0.208
20	11	0.05	0.5	0.594	0.625
20	11	0.35	0.5	0.564	0.574
20	11	0.05	1.0	0.667	0.768
20	11	0.35	1.0	0.619	0.670
50	3	0.05	0.5	0.523	1.605
50	3	0.35	0.5	0.531	1.337
50	3	0.05	1.0	0.660	1.973
50	3	0.35	1.0	0.672	1.364
20	3	0.05	0.5	1.140	4.076
20	3	0.35	0.5	1.160	3.406
20	3	0.05	1.0	1.424	5.372
20	3	0.35	1.0	1.475	3.835

number of informative descendants is roughly $n_i/2$ and the variance of $\widehat{\alpha_i^x}$ is then $8/n_i$ times the residual variance. For 50 (respectively 20) descendants per sire and a residual variance equal to 1, a 0.16 (respectively 0.4) mean square error is expected in the idealized case. The unknown QTL position, the distance between the QTL position and heterozygous markers for sire, the unknown sire marker genotypes and the overestimation of the residual variance when the additive QTL effect is great [10] explain the increase in the mean square error.

Results for the Gaussian likelihood in the three equally spaced marker maps may be explained considering a second idealized case where the QTL is known to be located at the beginning of the chromosome. As only sires heterozygous at least at one marker are considered, three cases of sires (c_1, c_2, c_3) exist with different variance of $\widehat{\alpha_i^x}$. c_1 contains sires that are heterozygous for the first marker, c_2 those that are homozygous for the first marker and heterozygous for the second one, and c_3 those that are heterozygous only for the last marker. The proportion of sires in the three classes are about 4/7, 2/7 and 1/7. The variance of $\widehat{\alpha_i^x}$ for sires in the class c_i is about

$$\frac{8\sigma_e^2}{n_i(1-2r_{c_i})^2}$$

where r_{c_i} denotes the recombination rate between the first marker heterozygous in the class c_i and the QTL located at the beginning of the chromosome. With 50 descendants per sire (respectively 20) and a residual variance equal to 1, a 1.7 (respectively 4.2) mean square error is expected. A more favourable location of the QTL (near the middle of the chromosome) decreases the mean square error.

The estimation of the within half-sib QTL substitution effect with the mixture likelihood does not only use the mean difference between informative descendants carrying allele A at a marker and those carrying allele B, but takes advantage of information from higher moments of the mixture distribution. Even if this information becomes negligible when the number of descendants per sire is large, in a finite population and especially for a widely spaced maker map, it leads to a significant reduction of the mean square error.

3. OTHER METHODS TO DEAL WITH UNKNOWN SIRE MARKER GENOTYPES

Errors in sire gamete reconstruction can decrease the power of both methods. Knott et al. [7] found that in their worst situation only 6 % of informative sires were incorrectly reconstructed, but they had studied large half-sib families with 100 descendants per sire.

Table VII shows, for one male, the empirical probability of correct reconstruction based on $\widehat{hs_i}$ over 1 000 replications. We confirm a 6 % maximum error in large families but found up to 30 % errors in smaller families, which led us to study alternative methods.

The rationale of the following alternative methods is that their aim is not to improve the quality of sire gamete reconstructions but to increase the power

Table VII. Empirical probability (in %) of correct sire gamete reconstruction over 1 000 replications.

Number of markers (distance between markers)	Number of alleles per marker	Number of descendants per sire 10	20	50	100
3	2	70.7	80.0	91.3	95.5
(50 cM)	4	74.7	87.6	95.5	98.4
5	2	77.0	89.8	96.7	98.7
(25 cM)	4	84.9	94.7	99.2	99.6
11	2	88.2	96.4	99.1	99.8
(10 cM)	4	96.2	99.5	100.0	100.0

of QTL detection. It is not necessary to work in two steps and the hs_i marker genotypes can be treated as nuisance parameters.

3.1. Estimations of sire marker genotypes based on conditional likelihood of quantitative phenotypes

The first alternative method is to treat the hs_i parameters as fixed parameters in the likelihood of quantitative phenotypes given the marker information, $\prod_i \Lambda_i^{x,hs_i}$. The full maximum is obtained after a search on a continuous space for the QTL location and effect, within sire mean and variance parameters and on a discrete space for the sire marker genotype parameters. This leads, with the Gaussian approximation of the mixture in progeny, to estimating the sire marker genotypes by

$$\widehat{\widetilde{hs}} = \text{argmax}_{hs}\Big(\sup_{x,\mu^{x1},\mu^{x2},\sigma_e^2} \sum_{i=1}^{n} \log(\widetilde{\Lambda}_i^{x,hs_i})\Big)$$

The maximum log-likelihood ratio test then gives

$$T^3 = 2\,\Big(\sup_{x,hs,\mu^{x1},\mu^{x2},\sigma_e^2} \sum_{i=1}^{n} \log(\widetilde{\Lambda}^{x,hs_i}) - \sup_{\mu,\sigma_e^2} \log(\Lambda_0)\Big)$$

3.2. Estimations of sire marker genotypes on weighted conditional likelihood

Estimating the sire marker genotypes by using only the previous likelihood function means neglecting information contained in $p(hs_i|M_i)$. Alternatively, the within sire conditional likelihood could be weighted by $p(hs_i|M_i)$ giving the weighted conditional likelihood to be maximized $\prod_i p(hs_i|M_i)\Lambda_i^{x,hs_i}$.

This leads, with the Gaussian approximation of the mixture in progeny, to estimating sire marker genotypes by

$$\widehat{\widehat{\widetilde{hs}}} = \text{argmax}_{hs}\Big(\sup_{x,\mu^{x1},\mu^{x2},\sigma_e^2} \sum_{i=1}^{n} \log(p(hs_i/M_i)\ \widetilde{\Lambda}_i^{x,hs_i})\Big)$$

The maximum log-likelihood ratio test is then equal to

$$T^4 = 2\,(\sup_{x,hs,\mu^{x1},\mu^{x2},\sigma_e^2} \sum_{i=1}^{n} \log(p(hs_i/M_i)\widetilde{\Lambda}^{x,hs_i}) - \sup_{\mu,\sigma_e^2} \log(\Lambda_0))$$

3.3. No estimation of sire marker genotypes

The last method is based on the likelihood function Λ^x proposed by Elsen et al. [2], using the Gaussian approximation of the mixture in progeny. The maximum log-likelihood ratio test is equal to

$$T^5 = 2\,(\sup_{x,\mu^{x1},\mu^{x2},\sigma_e^2} \sum_{i=1}^{n} \log(\sum_{hs_i} p(hs_i/M_i)\widetilde{\Lambda}^{x,hs_i}) - \sup_{\mu,\sigma_e^2} \log(\Lambda_0))$$

In practice, the three tests proposed should be slightly modified to take into account that the sire marker genotype space is growing exponentially with the number of informative markers per sire. This sire marker genotype space could be limited to genotypes that satisfy $p(hs_i|M_i)$ greater than a given value, fixed in the simulation study to 0.01.

3.4. Simulation results

Significant thresholds and powers for T^2, T^3, T^4 and T^5 are shown in *tables VIII* and *IX*. On the whole the compared tests gave very similar power for all of the situations studied, suggesting that the simplest method can be used, to avoid unnecessary computation. This similarity between tests may be attributed to the high percentage of correct sire gamete reconstruction. Only when markers were widely spaced and when family size was limited, did estimating sire marker genotypes on the weighted likelihood given the marker information lead to a slightly more powerful test.

Table VIII. Empirical 5 % significant thresholds for T^2, T^3, T^4 and T^5 over 5 000 replications.

Number of descendants	Number of markers	T^2	T^3	T^4	T^5
50	11	38.19	39.17	38.28	36.30
20	11	39.80	46.38	41.22	32.17
50	3	35.44	40.18	36.58	29.21
20	3	37.56	44.73	39.62	28.97

T^2: most probable sire marker genotype on marker information. T^3: most probable sire marker genotype on conditional likelihood. T^4: most probable sire marker genotype on weighted conditional likelihood. T^5: all sire marker genotypes.

Table IX. Percentage of replicates significant at the empirical 0.05 significant threshold for T^2, T^3, T^4 and T^5, over 1 000 replications.

Number of descendants	Number of markers	QTL position	QTL additive effect	T^2	T^3	T^4	T^5
50	11	0.05	0.5	49.8	49.0	49.6	48.6
50	11	0.35	0.5	60.8	60.0	60.9	58.6
50	11	0.05	1.0	96.6	96.4	96.7	96.6
50	11	0.35	1.0	98.7	98.7	98.8	98.6
20	11	0.05	0.5	18.1	16.6	16.6	16.5
20	11	0.35	0.5	19.4	18.7	19.0	17.9
20	11	0.05	1.0	58.2	53.2	57.9	54.4
20	11	0.35	1.0	70.2	67.4	70.1	67.2
50	3	0.05	0.5	26.1	26.0	25.4	24.1
50	3	0.35	0.5	24.3	24.4	23.7	22.0
50	3	0.05	1.0	78.2	75.4	78.2	77.9
50	3	0.35	1.0	71.1	69.2	71.5	70.5
20	3	0.05	0.5	11.9	11.9	12.2	11.8
20	3	0.35	0.5	10.4	11.3	11.9	11.2
20	3	0.05	1.0	29.7	28.7	30.6	30.4
20	3	0.35	1.0	25.9	26.0	27.3	25.7

T^2: most probable sire marker genotype on marker information. T^3: most probable sire marker genotype on conditional likelihood. T^4: most probable sire marker genotype on weighted conditional likelihood. T^5: all sire marker genotypes.

4. DISCUSSION AND CONCLUSION

When the marker map is known, Λ^x is proportional to the joint likelihood of quantitative and marker observations considering the hs_i as random parameters with uniform prior distribution, so joint or conditional likelihoods give the same results in terms of QTL parameter estimates and detection test. The joint likelihood was first used by Georges et al. [3] to map QTL in dairy cattle by considering only sire-by-sire analyses. Then Jansen et al. [6] computed the conditional likelihood and considered pooled sire analysis. As mentioned by Georges et al. [3], the problem with this likelihood is due to the fact that only the $|\mu^{x1} - \mu^{x2}|$ can be estimated when there is no information on grandparents. Indeed, using the alternative parametrization $\mu_i^{x1} = \mu_i + \alpha_i^x/2$, $\mu_i^{x2} = \mu_i - \alpha_i^x/2$ it has been proved (in the Appendix) that the sign of α_i^x cannot be estimated. This is not important for an objective of QTL detection but shows the limit of this method, if an objective is to pursue QTL effect estimation simultaneously.

For all the methods studied, the empirical significance thresholds were obtained by simulations of sire and progeny marker genotypes and of the quantitative performances. In practice a permutation test [1] or a Monte-Carlo simulation taking account of the correct marker structure should be used. This could lead to slightly different threshold values. However, because very high correlations between tests were observed, we can guess that the conclusions

concerning the different tests should not depend on the chosen significance threshold.

Modelling progeny quantitative observations with mixture distributions is a more computationally demanding approach than methods using Gaussian distributions. Previous studies [5, 10] have compared in a single family the estimates obtained using mixture or Gaussian models. They concluded that the estimation accuracy is similar in both models, except for the residual variance when a QTL of large effect is mapped in a widely spaced marker map. Our study on multiple families showed that the accuracy of within half-sib QTL substitution effect estimates decreased significantly for the Gaussian model compared to the mixture model, especially in a more widely spaced marker map and even if the QTL effect was not large, although the test power and the accuracy of the QTL position remained comparable.

Our comparison of alternative methods for handling the problem of unknown sire marker linkage phases showed clearly that a simple method of reconstructing the sire genotype is almost as powerful as more complex methods, especially the one that takes into account all the possible sire marker genotypes since the T^5 test was never the most powerful test.

REFERENCES

[1] Churchill G.A., Doerge R.W., Empirical threshold values for quantitative trait mapping, Genetics 138 (1994) 963–971.

[2] Elsen J.M., Mangin B., Goffinet B., Le Roy P., Boichard D., Alternative models for QTL detection in livestock. I. General introduction, Genet. Sel. Evol. (1999) 213–224.

[3] Georges M., Nielsen D., Mackinnon M., Mishra A., Okimoto R., Pasquino A.T., Sargeant L.S., Sorensen A., Steele M.R., Zhao X., Womack J.E., Hoeschele I., Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing, Genetics 139 (1995) 907–920.

[4] Goffinet B., Le Roy P., Boichard D., Elsen J.M., Mangin B., Alternative models for QTL detection in livestock. III. Assumption about the model, Genet. Sel. Evol. (1999) in press.

[5] Haley C.S., Knott S.A., A simple regression method for mapping quantitative trait loci in lines crosses using flanking markers, Heredity 69 (1992) 315–324.

[6] Jansen R.C., David L.J., Van Arendonk J.A.M., A mixture model approach to the mapping of quantitative trait loci in complex populations with an application to multiple cattle families, Genetics 148 (1998) 391–400.

[7] Knott S.A., Elsen J.M., Haley C.S., Methods for multiple-marker mapping of quantitative trait loci in half-sib populations, Theor. Appl. Genet. 93 (1996) 71–80.

[8] Martinez O., Curnow R.N., Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers, Theor. Appl. Genet. 85 (1992) 480–488.

[9] Rebaï A., Goffinet B., Mangin B., Comparing power of different methods for QTL detection, Biometrics 51 (1995) 87–99.

[10] Xu S., A comment on simple regression method for interval mapping, Genetics 141 (1995) 1657–1659.

APPENDIX: Proof of the nonestimability of the sign of the within half-sib QTL substitution effect

The likelihood Λ^x is a function of parameters μ_i^{x1} and μ_i^{x2}. Using the alternative parametrization $\mu_i^{x1} = \mu_i + \alpha_i^x/2$, $\mu_i^{x2} = \mu_i - \alpha_i^x/2$, Λ^x is a function of α_i^x and we denote by $\Lambda_{ij}^{x,hs_i}(\alpha_i^x)$ the terms corresponding to descendants

$$\Lambda_{ij}^{x,hs_i}(\alpha_i^x) = \sum_{q=1}^{2} p(d_{ij}^x = q/hs_i, M_i) f(yp_{ij}/d_{ij}^x = q)$$

Consider a genotype $hs_i = \{hs_i^1, hs_i^2\}$ and its symmetrical $hs_i' = \{hs_i^2, hs_i^1\}$. When there is no ancestry information, it is obvious that

$$p(hs_i'|M_i) = p(hs_i|M_i)$$

Using the relation $p(d_{ij}^x = q/hs_i, M_i) = 1 - p(d_{ij}^x = q/hs_i', M_i)$ it can easily be proven that

$$\Lambda_{ij}^{x,hs_i'}(\alpha_i^x) = \Lambda_{ij}^{x,hs_i}(-\alpha_i^x)$$

In Λ^x, terms corresponding to hs_i and hs_i' can be grouped to obtain

$$\begin{aligned}\Lambda^x &= \prod_{i=1}^{n} \sum \left(p(hs_i/M_i) \prod_{j=1}^{n_i} \Lambda_{ij}^{x,hs_i}(\alpha_i^x) + p(hs_i'/M_i) \prod_{j=1}^{n_i} \Lambda_{ij}^{x,hs_i'}(\alpha_i^x) \right) \\ &= \prod_{i=1}^{n} \sum \left(p(hs_i/M_i) \left(\prod_{j=1}^{n_i} \Lambda_{ij}^{x,hs_i}(\alpha_i^x) + \prod_{j=1}^{n_i} \Lambda_{ij}^{x,hs_i}(-\alpha_i^x) \right) \right)\end{aligned}$$

Because Λ^x is a symmetrical function of the α_i^x parameters, their sign cannot be estimated.