Original article

Mapping quantitative trait loci in outcross populations via residual maximum likelihood. II. A simulation study

FE Grignola[1], I Hoeschele[1*], Q Zhang[1], G Thaller[2]

[1] *Department of Dairy Science, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0315, USA;*
[2] *Lehrstuhl fuer Tierzucht, Technical University of Munich at Weihenstephan, 85350 Munich, Germany*

Summary – Position and variance contribution of a single QTL together with additive polygenic and residual variance components were estimated using a residual maximum likelihood method and a derivative-free algorithm. The variance–covariance matrix of QTL effects and its inverse were computed conditional on incomplete information from multiple-linked markers. Simulation was employed to investigate the accuracy of parameter estimates and likelihood ratio tests. The design was a granddaughter design with 2 000 sons, 20 sires of sons and 9 ancestors of sires. Designs with 1 000 and 600 sons were also investigated. Data were simulated under three different genetic models for the QTL, a biallelic model, a multiallelic model with ten alleles at equal frequencies, and a model with normally and independently distributed QTL allelic effects for base individuals. The trait analyzed was daughter yield deviation or the daughter average adjusted for environmental effects and merits of mates of the sons. Genotypes for five markers situated on the same chromosome were generated for all sons and their ancestors. Data were analyzed with and without relationships among sires. Parameters were estimated with good accuracy under all three simulation models. The REML method was fairly robust to the number of alleles at the QTL for the designs studied.

quantitative trait loci / residual maximum likelihood / mapping / simulation / granddaughter design

Résumé – La cartographie de locus de caractère quantitatif dans des populations en ségrégation à l'aide du maximum de vraisemblance résiduelle. II. Étude de simulation. *La position et la contribution à la variance d'un locus de caractère quantitatif (QTL) ont été estimées avec les variances polygéniques additives et résiduelles par une méthode de maximum de vraisemblance résiduelle et un algorithme sans dérivation. La matrice de*

---

* Correspondence and reprints

variance–covariance des effets de QTL et son inverse ont été calculées conditionnellement à l'information incomplète relative à des ensembles de marqueurs liés. Une simulation a été réalisée pour déterminer la précision des paramètres estimés et les tests de rapport de vraisemblance. Le plan d'expérience était un schéma petite-fille, avec 2 000 fils, 20 pères et 9 ancêtres de pères. Des schémas avec 1 000 et 600 fils ont également été étudiés. Les données ont été simulées sous trois modèles génétiques différents pour le QTL, un modèle biallélique, un modèle à 10 allèles d'égale fréquence, et un modèle avec des effets alléliques distribués normalement et d'une manière indépendante chez les individus de base. Le caractère analysé était la production des petites-filles en écart à la moyenne, ou leur moyenne ajustée pour les effets de milieu et les valeurs génétiques des conjointes des fils. Les génotypes pour cinq marqueurs situés sur un même chromosome ont été générés pour tous les fils et leurs ancêtres, et les données ont été analysées avec ou sans relation de parenté entre les pères. Les paramètres étaient estimés avec une bonne précision dans les trois modèles simulés. La méthode était d'une robustesse satisfaisante relativement au nombre d'allèles dans les schémas étudiés.

locus de caractère quantitatif / maximum de vraisemblance résiduelle / cartographie / simulation / schéma petite-fille

INTRODUCTION

In a companion paper (Grignola et al, 1996), a residual maximum likelihood (REML) method was derived for estimating position and variance contribution of a single QTL together with additive polygenic and residual variance components. The REML analysis was implemented with a derivative-free algorithm. The method overcomes the shortcomings of the traditional methods of linear regression (eg, Haley et al, 1994; Zeng, 1994) and maximum likelihood (ML) interval mapping (eg, Weller, 1986; Lander and Botstein, 1989; Knott and Haley, 1992). ML and regression methods cannot fully account for the more complex data structures of outcross populations, eg, data on several families with relationships across families, unknown linkage phases in parents, unknown number of QTL alleles in the population, and varying amounts of data information on different QTLs or in different families.

The REML method is based on a mixed linear model including random polygenic effects and random QTL effects. Polygenic effects represent the sum of the additive effects at all loci not linked to the markers. The QTL allelic effects are assumed to have a prior normal distribution with variance–covariance matrix conditional on information from multiple linked markers. Because the true nature of the QTLs is unknown, ie, the number of alleles at a QTL in the population studied is unknown, the robustness of the REML analysis to the number of alleles at the QTL must be evaluated.

One of the main experimental designs for QTL mapping in livestock is the half-sib design, used in cattle in the form of daughter or granddaughter designs (Weller, 1990). In this paper, we evaluate the accuracy of the REML analysis in QTL mapping using granddaughter designs. The simulated designs resemble actual designs for the US Holstein population. Data are simulated under several genetic models differing in the number of QTL alleles. The analysis is carried out with and without consideration of relationships among sires. Here, we only present simulation results. The analysis is described in detail in the companion paper (Grignola et al, 1996).

SIMULATION

Design

The most frequently used design for mapping QTL in dairy cattle is the granddaughter design (GDD), where marker genotypes are collected on sons and phenotypes on daughters of the sons. A GDD was simulated with a pedigree structure resembling the real GDD of the US public gene mapping project for dairy cattle based on the Dairy Bull DNA Repository (Da et al, 1994). The simulated GDD consisted of 2 000 sons, 20 sires, and 9 ancestors of the sires (fig 1), and is identical to the design used in the Bayesian linkage analyses of Thaller and Hoeschele (1996a,b) and Uimari et al (1996a). While this design had 100 sons per sire, designs with only 50 or 30 sons per sire were also simulated.
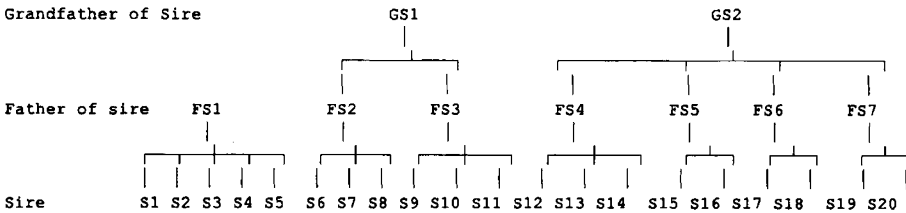

Fig 1. Pedigree of granddaughter design.

The phenotype simulated was daughter yield deviation (DYD) of sons (Van-Raden and Wiggans, 1991). DYD is an average of the phenotypes of the daughters adjusted for systematic environmental effects and genetic values of the daughter's dams. Total variance of DYD equals $\text{Var}(\text{DYD}) = 0.25\sigma_a^2/R$ and can be factored into

$$\text{Var}(\text{DYD}) = 0.25\sigma_a^2 + \left\{[(1-R)/R] \cdot 0.25\sigma_a^2\right\}$$

where R is reliability or squared accuracy of the son's estimated additive genetic effect or breeding value, and σ_a^2 is the additive genetic variance. In this factorization, the first component is the variance among the sons' transmitting abilities or half of their additive genetic values, and the second term is 'residual variance' or the variance of the average dam and Mendelian genetic effect of the daughters and the average environmental effect. Variance of DYD can be rewritten as

$$\text{Var}(\text{DYD}) = 0.25\sigma_a^2 + w \cdot \sigma_e^2$$

where $w = (1-R)/R$ and $\sigma_e^2 = 0.25\sigma_a^2$. Therefore, when analyzing DYD with the weight w, DYD has an expected heritability of 0.5, because the expected value of the estimate of σ_a^2 is $0.25\sigma_a^2$. Hence, there is the option of treating heritability as known in the REML analysis when phenotypic data are DYDs.

Marker and QTL genotypes were simulated according to Hardy–Weinberg frequencies and the map positions of all loci. One linkage group was considered which consisted of five marker loci and one QTL. Each marker locus had five alleles at

equal frequencies, with the exception of one design where each marker had only three alleles at equal frequencies. The markers were spaced 20 cM apart and, for the results presented here, the QTL was located in interval 3 at 5 cM from the left marker (other QTL positions were simulated to verify that the analysis was working properly).

Polygenic and QTL effects were simulated according to the pedigree in figure 1. Data were analyzed (i) using full pedigree information and (ii) assuming that the 20 sires in figure 1 were unrelated, as is common practice. The QTL contribution to the DYDs of sons was generated by sampling individual QTL allelic effects of daughters under each of the genetic models as described below. This sampling of QTL effects assures that DYD of a heterozygous son or of a son with substantial difference in the additive effect of its two QTL alleles has larger variance among daughters due to the QTL than a homozygous son or a son with similar QTL allelic effects.

Genetic models

Three different genetic models were used to simulate data. Common to all models were the parameters narrow sense heritability of individual phenotypes $h^2 = 0.3$, phenotypic SD $\sigma_p = 100$, and the order of and recombination rates among all loci. Under all three models, phenotypes were simulated as

$$\mathrm{DYD}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} g_{ij} + u_i + e_i; \quad \mathrm{Var}(e_i) = \frac{1}{n_i}(0.75\sigma_u^2 + \sigma_e^2)$$

where n_i was the number of daughters of son i, g was the sum of the v effects in daughter j of son i, u was a normally distributed polygenic effect, e was a normally distributed residual, polygenic variance (σ_u^2) was equal to the difference between additive genetic variance (σ_a^2) and the variance explained by the QTL $(2\sigma_v^2)$, and σ_e^2 was environmental variance. Number of daughters per son was set to 50, corresponding to a reliability (VanRaden and Wiggans, 1991) of near 0.8. The ratio of the QTL allelic variance (σ_v^2) to the additive genetic variance (σ_a^2) is denoted by v^2 below.

Model 1. Normal-effects model

For each individual with one or both parents unknown, one or both QTL effects, respectively, were drawn from $N(0, \sigma_v^2)$. For the pedigree in figure 1, there were 32 distinct base alleles, and the QTL was treated as a locus with 32 distinct alleles in passing on alleles to descendants. The parameter σ_v^2 was set to $0.25\sigma_a^2$ or $0.0625\sigma_a^2$, ie, the simulated QTL accounted for 50% $(2v^2 = 0.5)$ or 12.5% $(2v^2 = 0.125)$ of the total additive genetic variance, respectively. In an additional simulation, v^2 was set to zero to obtain the empirical distribution of the test statistic under the null hypothesis.

Model 2. Multiallelic model

The QTL had ten alleles with equal frequencies. For the biallelic QTL with $2v^2 = 0.5$ (see below), the difference among homozygotes was $2\alpha = 2\sigma_a$. For the multiallelic

QTL, means of the ten homozygous genotypes ranged from $-\sigma_a$ to σ_a at equal intervals. Means of heterozygotes were calculated assuming additive gene action. Given these means (μ), additive effects of alleles were determined

$$\alpha_i = p \sum_{j=1}^{10} \mu_{ij}$$

as where $p_1 = p_2 = \ldots = p_{10} = p = 0.1$. The variance at the QTL was

$$\sigma_{QTL}^2 = 2v^2\sigma_a^2 = 2p \sum_{i=1}^{10} \alpha_i^2$$

which yielded a value of $2v^2 = 0.204$.

Model 3. Biallelic model

The QTL was biallelic with allele frequency of 0.5. The variance at the QTL was

$$2\sigma_v^2 = 2v^2\sigma_a^2 = 2p(1-p)\alpha^2$$

where for $p = 0.5$ and $2v^2 = 0.5$ or $2v^2 = 0.125$, QTL substitution effect α was determined and used to compute the additive effects of the two QTL alleles as $-p\alpha$ and $(1-p)\alpha$.

RESULTS

The REML analysis for single QTL mapping using all markers on a chromosome is decribed in detail in the companion paper (Grignola et al, 1996). Analyses were performed with and without considering relationships among sires and with the marker linkage phases of sires and ancestors known or unknown. For the designs considered here, the most probable (more than 90%) linkage phase of the sires was always the true phase, but phases probabilities calculated for the ancestors indicated more uncertainty about their true phases. When phases were treated as unknown, the analysis employed equation [11] of Grignola et al (1996) to calculate the inverse of the variance–covariance matrix of the QTL effects of ancestors and sires. Contributions from sons were calculated assuming that the most likely sires phases equalled the true phases.

Analysis of a single data set took around 8 mins of computing time on an IBM SP2 system with RS6000 390 and 590 nodes. In a preliminary investigation, the REML analysis, using a derivative-free Simplex algorithm was started from very different initial values for the parameters to verify convergence to the same estimates. As an example, a particular data set simulated under the biallelic QTL model ($2v^2 = 0.5$) was analyzed using the two starting value sets [0.9, 0.45, 0.6, 200] and [0.1, 0.05, 0.4, 1500] for $[h^2, v^2, d_Q, \sigma_e^2]$ with the parameters defined in table I. Parameter estimates and likelihood for the first starting value set were 0.6380, 0.2790, 0.4434, 542.6, and $-4\,331.3$. Corresponding figures for the second starting value set were 0.6380, 0.2780, 0.4440, 541.6, and $-4\,331.3$.

Table I. Definition of model parameters and likelihood ratio.

Symbol	Parameter
h^2	Narrow sense heritability
v^2	Ratio of QTL to additive genetic variance
d_Q	QTL location (cM)
σ_e^2	Residual variance
LR	Likelihood ratio test statistic $= -2 \log[L(H_0: v^2 = 0)/L(H_A: v^2 \neq 0)]$

Figure 2 contains cumulative distribution functions with 90th percentiles of 2.706 and 4.605 for chi-square distributions with one and two degrees of freedom, respectively. The figure also contains an empirical distribution function with 90th percentile of 3.746 for the likelihood ratio statistic from 1 000 replicates generated without a QTL. Xu and Atchley (1995) found the distribution to be close to the 2 *df* chi-square, while here the empirical distribution was in between the 1 *df* and 2 *df* chi-square distributions and slightly closer to the 2 *df* distribution.
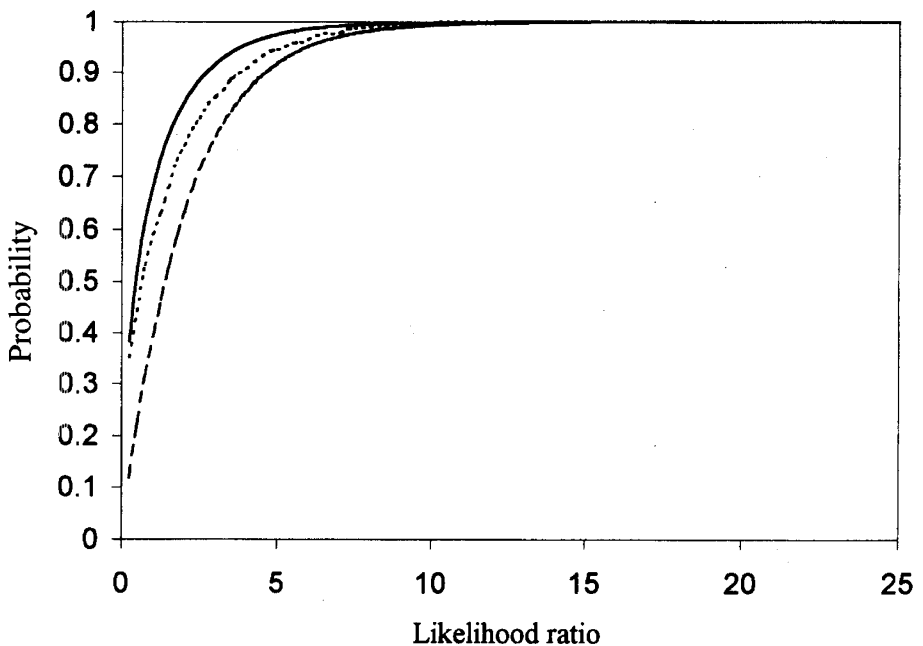

Fig 2. Cumulative distribution functions with 90th percentiles for chi-square distributions with one and two degrees of freedom, respectively, and empirical cumulative distribution function with 90th percentile of the likelihood ratio statistic from 1 000 replicates generated under the null hypothesis of no QTL. Chi-square 1 *df* (———); chi-square 2 *df* (– – –); empirical distribution (- - - -).

Parameter estimates and likelihood ratio statistics are shown in tables II–VI. All results are based on 50 replicates. In table II, results for the normal-effects QTL model with $2v^2 = 0.5$ are presented. Several different analyses, described in table II, were conducted. First, sires were simulated as unrelated and analyzed without relationships. Then, relationships among sires were simulated according to figure 1. In analyses of these data, sires were treated as unrelated, treated as related with known marker linkage phases (ie, the true phases were used for all sires and ancestors), treated as related with the most likely linkage phases used in place of the true phases, or treated as related with linkage phases considered as unknown (by using equation [11] in Grignola et al, 1996).

Table II. Means and their standard errors (SE) for the estimates of h^2, v^2, d_Q, and σ_e^2 and log likelihood ratio for the normal-effects QTL model explaining 50% of the additive genetic variance (50 replicates).

Parameter [a]	True value	Analysis I [b] Analysis II [c]	Analysis III [d] Analysis IV [e]	Analysis V [f]	Analysis VI [g] Analysis VII [h]
h^2	0.5	0.558 (0.037) 0.537 (0.039)	0.558 (0.039) 0.544 (0.038)	0.541 (0.037)	0.494 (0.034) 0.5 (fixed)
v^2	0.25	0.243 (0.015) 0.279 (0.018)	0.280 (0.015) 0.279 (0.015)	0.280 (0.015)	0.288 (0.016) 0.262 (0.013)
d_Q	0.4527	0.461 (0.004) 0.451 (0.005)	0.450 (0.005) 0.450 (0.005)	0.449 (0.005)	0.444 (0.005) 0.449 (0.005)
σ_e^2	750	748 (75) 828 (86)	767.4 (79) 798.1 (79)	802.3 (79)	903.8 (79) 759.1 (5.1)
LR	–	125.7 (6.8) 125.7 (6.8)	151.9 (9.6) 151.9 (9.6)	151.7 (9.6)	127.8 (7.8) 151.5 (9.6)

[a] Parameters and likelihood ratio are defined in table I. [b] Analysis I: no relationships simulated, no relationships in analysis. [c] Analysis II: relationships simulated, no relationships in analysis. [d] Analysis III: relationships simulated and used in analysis with known phases. [e] Analysis IV: relationships simulated and used in analysis with most likely phases. [f] Analysis V: relationships simulated and used in analysis with unknown phases. [g] Analysis VI: relationships used in analysis and phases unknown; three alleles at markers. [h] Analysis VII: relationships used in analysis and phases unknown; h^2 fixed at 0.5.

Accuracy of parameter estimates was slightly higher when sires were simulated and analyzed as being unrelated (analysis I), compared to the case where sires were simulated and analyzed as related (analyses III–V). When sires were simulated related and analyzed unrelated, the estimate of heritability was slightly lower and the likelihood ratio statistic was lower than the corresponding values obtained with sires treated as related (analyses II and IV). When sires were analyzed using relationships, treating the most likely marker linkage phases of sires and ancestors as the true phases produced almost identical parameter estimates and likelihood

Table III. Means and their standard errors (SE) for the estimates of h^2, v^2, d_Q, and σ_e^2 and log likelihood ratio for the normal-effects QTL model explaining 12.5% of additive genetic variance (50 replicates).

Parameter [a]	True value	Analysis I [b] Analysis II [c]	Analysis III [d] Analysis IV [e]	Analysis V [f] Analysis VI [g]
h^2	0.5	0.617 (0.042) 0.559 (0.042)	0.602 (0.042) 0.472 (0.040)	0.538 (0.041) 0.467 (0.041)
v^2	0.0625	0.058 (0.005) 0.063 (0.006)	0.069 (0.006) 0.095 (0.011)	0.083 (0.008) 0.112 (0.013)
d_Q	0.4527	0.473 (0.016) 0.434 (0.017)	0.435 (0.016) 0.467 (0.020)	0.507 (0.030) 0.526 (0.030)
$\sigma^2 3_e$	750	685.7 (97) 810.2 (98)	694.5 (83) 994.6 (88)	852.0 (95) 1033.8 (99)
LR	–	20.46 (1.9) 18.94 (1.7)	10.42 (1) 9.61 (0.89)	6.56 (0.74) 6.00 (0.74)

[a] Parameters and likelihood ratio are defined in table I. [b] Analysis I: relationships simulated, 100 sons/sire, relationships in analysis. [c] Analysis II: relationships simulated, 100 sons/sire, no relationships in analysis. [d] Analysis III: relationships simulated, 50 sons/sire, relationships in analysis. [e] Analysis IV: relationships simulated, 50 sons/sire, no relationships in analysis. [f] Analysis V: relationships simulated, 30 sons/sire, relationships in analysis. [g] Analysis VI: relationships simulated, 30 sons/sire, no relationships in analysis.

ratio statistics as considering linkage phases as unknown (analyses III–V). Only small differences between analyses with known, most likely, or unknown linkage phases are to be expected for a granddaughter design with marker information on 100 sons per sire.

A further analysis (analysis VI in table II) was conducted on the same GDD except that markers had three alleles at equal frequencies rather than the five alleles simulated for all other design variations. Parameter estimates were not noticeably affected by the decline in marker polymorphism, but the average likelihood ratio statistic was reduced.

In the last analysis of table II (VII), heritability was fixed at 0.5. Accuracy of estimates of the QTL parameters and of residual variance was improved but the value of the likelihood ratio statistic was almost unchanged.

Table III contains results for data sets generated with $2v^2 = 0.125$ and with relationships among sires, and analyzed first by ignoring relationships among sires and secondly by accounting for relationships with linkage phases treated as unknown (equation [11] in Grignola et al, 1996). Analyses I and II in table III were conducted for the GDD with 100 sons per sire. The estimate of heritability and the likelihood ratio statistic were again lower when relationships among sires were ignored. Expectedly, position of this smaller QTL was less accurately estimated, and

Table IV. Means and their standard errors (SE) for the estimates of h^2, v^2, d_Q, and σ_e^2 and log likelihood ratio for the multiallelic QTL model explaining 20.2% of additive genetic variance (50 replicates).

Parameter [a]	True value	Estimate	SE
h^2	0.5	0.5718	0.0443
		0.5861	0.0420
v^2	0.102	0.1123	0.0103
		0.1086	0.0083
d_Q	0.4527	0.4404	0.0080
		0.4444	0.0076
σ_e^2	750	793.4	97.9
		741.9	91
LR	–	18.94	1.638
		20.47	1.934

Relationships among sires simulated, no relationships considered in analysis or relationships used with unknown phases. [a] Parameters and likelihood ratio are defined in table I.

Table V. Means and their standard errors (SE) for the estimates of h^2, v^2, d_Q, and σ_e^2 and log likelihood ratio for the biallelic QTL model explaining 50% of additive genetic variance (50 replicates).

Parameter [a]	True value	Estimate	SE
h^2	0.5	0.4688	0.0360
		0.5387	0.0319
v^2	0.25	0.3042	0.0177
		0.2565	0.0124
d_Q	0.4527	0.4498	0.0031
		0.4478	0.0032
σ_e^2	750	981.1	82.46
		770.4	68.39
LR	–	145.3	5.918
		150.6	6.072

Relationships among sires simulated, no relationships considered in analysis or relationships used with unknown linkage phases. [a] Parameters and likelihood ratio are defined in table I.

the likelihood ratios were considerably lower than those in table II. In the analyses of the granddaughter designs with only 50 (analyses III and IV) or 30 (analyses V and VI) sons, heritability estimates were also lower when sires were treated as unrelated, and QTL variance contribution was overestimated. The differences

Table VI. Means and their standard errors (SE) for the estimates of h^2, v^2, dQ, and σ_e^2 and log likelihood ratio for the biallelic QTL model explaining 12.5% of additive genetic variance (50 replicates).

Parameter [a]	True value	Estimate	SE
h^2	0.5	0.4789	0.0378
		0.5069	0.0393
v^2	0.0625	0.0742	0.0053
		0.0713	0.0049
d_Q	0.4527	0.4304	0.0146
		0.4284	0.0152
σ_e^2	750	960	86.57
		896.3	85.27
LR	–	22.33	1.754
		24.20	1.817

[a] Parameters and likelihood ratio are defined in table I. Relationships among sires simulated, no relationships considered in analysis or relationships used with unknown linkage phases.

between analyses with and without relationships in the likelihood ratio statistics were rather small for the designs with 30 and 50 sons, where the likelihood ratio statistics were near the threshold values of 5.99 (0.05 type-I error level) and 9.21 (0.01 type-I error level) when assuming a chi-square distribution with two degrees of freedom.

Figures 3 and 4 depict residual likelihood profiles for single replicates generated with $2v^2 = 0.5$ and $2v^2 = 0.125$, respectively, and obtained from analyses with relationships among sires and with linkage phases of sires and ancestors treated as unknown. Both figures display the marker positions, the most likely QTL location, and a confidence interval (CI) for the QTL position calculated by the LOD drop-off method of Lander and Botstein (1989). The limits of this CI were found by determining the map position at either side of the most likely position, where the LOD score had fallen by one unit (this calculation required converting natural logarithms to base 10 logarithms). As pointed out earlier, information from all markers was utilized, leading to smoother profiles (Knott and Haley, 1992; Georges et al, 1995) than the original interval mapping method of Lander and Botstein (1989).

Results for the multiallelic model ($2v^2 = 0.204$) are presented in table IV. Again, data sets were analyzed by first treating sires as unrelated and then by utilizing relationships among sires with linkage phases of sires and ancestors treated as unknown. Parameters were estimated quite accurately, and likelihood ratios were significant and similar to those for the normal-effects QTL model with $2v^2 = 0.125$.

Results for the biallelic models with half of the homozygote difference equal to one additive genetic SD ($2v^2 = 0.5$) or to one half of it ($2v^2 = 0.125$) are presented in tables V and VI, respectively. For both tables, sires were related
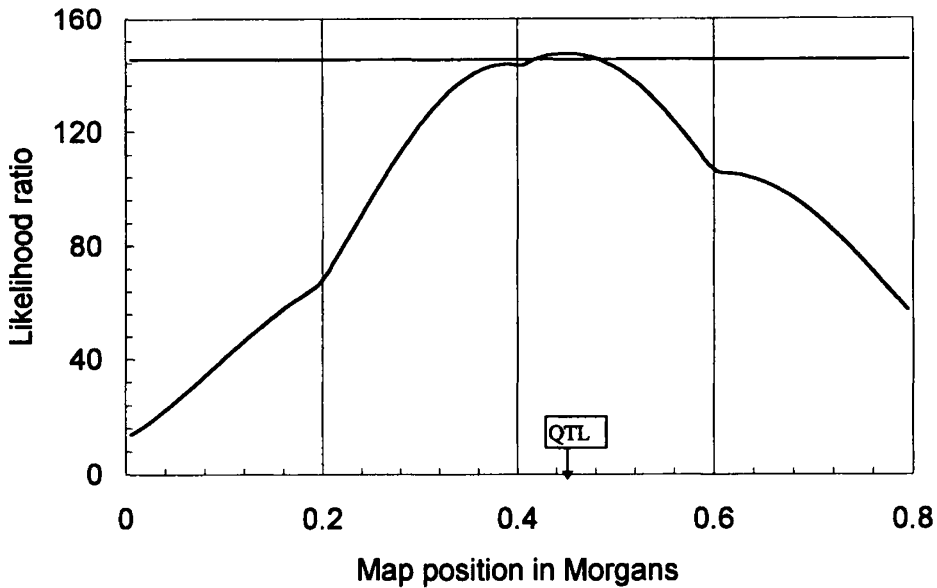
Fig 3. Likelihood profile for a single data set generated with $2v^2 = 0.5$ (see table I). Markers are located at 0, 20, 40, 60 and 80 cM. The QTL is located at 45.27 cM. Significance threshold for log likelihood ratio is 5.991 from $\chi^2(2$ *df*, $\alpha = 0.05)$.

in the simulation, and data sets were analyzed by ignoring relationships and by accounting for relationships with linkage phases unknown (equation [11] in Grignola et al, 1996). Ignoring relationships among sires again decreased the estimate of heritability but increased the estimate of v^2. The least accurate estimate of QTL position was obtained under the biallelic model with $2v^2 = 0.125$. Overall, and at least when relationships among sires were considered in the analyses, parameter estimates were not noticeably inferior to those obtained under the correponding normal-effects QTL models (table II), and likelihood ratios for the biallelic and normal-effects models were similar.

CONCLUSIONS

REML analysis based on a mixed linear model with random QTL allelic effects, having a prior normal distribution, provided quite accurate estimates of QTL location and of the variance components, in particular of the QTL variance contribution. Data were generated under three different genetic models for the QTL, the biallelic, the multiallelic (ten alleles) and the normal-effects (two effects per founder drawn from independent and identical normal distributions) model. While the normal-effects model is very similar to the analysis model, the biallelic model with gene frequency p and substitution effect α deviates the most from the model of analysis. In the biallelic model, 50% of the individuals are expected to be
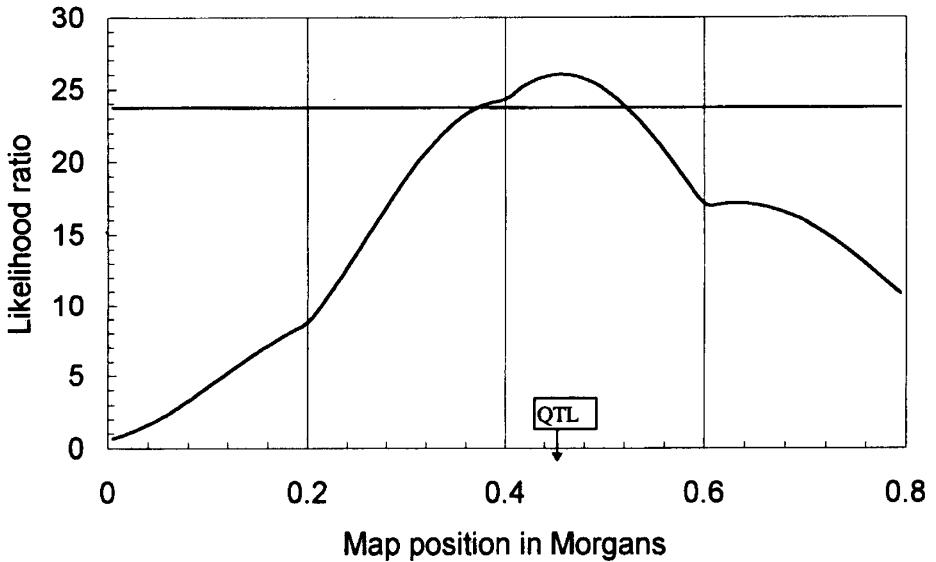
Fig 4. Likelihood profile for a single data set generated with $2v^2 = 0.125$ (see table I). Markers are located at 0, 20, 40, 60 and 80 cM. The QTL is located at 45.27 cM. Significance threshold for log likelihood ratio is 5.991 from $\chi^2(2df, \alpha = 0.05)$.

homozygous for a gene frequency of $p = 0.5$, and variance among daughters of a homozygous son is equal to

$$p(1 - p)\alpha^2 + 0.75\sigma_u^2 + \sigma_e^2$$

while variance among daughters of heterozygous sons is higher by $0.25\alpha^2$. Therefore, Thaller and Hoeschele (1996a,b) and Uimari et al (1996a) fitted two different residual variances of DYD when performing Bayesian analysis of linkage of a biallelic QTL.

Despite the discrepancies between the biallelic model and the model of analysis, the REML analysis was quite robust to the number of alleles at the QTL, a result which is in agreement with findings of Xu and Atchley (1995); ie, polymorphism at the QTL did not strongly affect parameter estimates or hypothesis tests. This finding confirms the usefulness of the REML analysis as an alternative method of analysis which, although not nonparametric, requires fewer parametric assumptions (number of QTL alleles and their frequencies) than maximum likelihood and Bayesian analyses based on biallelic QTL models. Furthermore, REML is in general known to be quite robust to deviations from normality.

The REML analysis can be considered as an approximation to the Bayesian analysis of Hoeschele et al (1996) fitting a normal-effects QTL model. The Bayesian analysis has the advantage of also being able to fit a biallelic model, but for the designs considered here, it does not give more accurate parameter estimates than

the REML analysis and requires several hours of computing time. Although the REML analysis fitting a single QTL requires only around 8 mi of CPU time, this requirement still prohibits the calculation of genome-wide significance thresholds for this method using data permutation (Churchill and Doerge, 1994), unless a number of less stringent significance levels are chosen to obtain threshold values and these are used to extrapolate to the desired significance threshold (Uimari et al, 1996b). Only the least-squares method allows direct computation of genome-wide significance thresholds from a large number of permutations (eg, 10 000 to 100 000).

The REML analysis has been implemented in the Fortran program SMREML, which is available from the authors upon request. The program is currently being extended to fit two linked QTLs per chromosome, rather than using the approach of Xu and Atchley (1995) fitting the variances associated with the next-to-flanking markers to account for additional, linked QTLs. Their approach is only approximate as effects associated with marker alleles identified within founders erode over generations. Furthermore it requires many additional parameters when the marker polymorphism is limited, causing the flanking and next-to-flanking markers to differ among families. In the near future, the program will be extended to other designs (eg, full-sibships within half-sibships), where the current computation of the inverse of the variance–covariance matrix becomes approximate due to uncertain linkage phases in parents of final offspring, and other ways of computing this inverse exactly (eg, Van Arendonk et al, 1994) will be implemented.

ACKNOWLEDGMENTS

REFERENCES

Churchill G, Doerge R (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963-971

Da Y, Ron M, Yanai A, Band M, Everts RE, Heyen DW, Weller JI, Wiggans GR, Lewin HA (1994) The Dairy Bull DNA Repository: a resource for mapping quantitative trait loci. *Proc 5th World Congr Genet. Appl Livest Prod* 21, 229-232

Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao X, Womack JE, Hoeschele I (1995) Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* 139, 907-920

Grignola FE, Hoeschele I, Tier B (1996) Mapping quantitative trait loci in outcross populations via residual maximum likelihood. I. Methodology. *Genet Sel Evol* 28, 479-490

Haley CS, Knott SA, Elsen JM (1994) Mapping quantitative trait loci in crosses between outbred lines using least-squares. *Genetics* 136, 1195-1207

Hoeschele I, Uimari P, Grignola FE, Zhang Q, Gage KM (1996) Statistical mapping of polygene loci in livestock. *Proc Int Biometric Soc* (in press)

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

Thaller G, Hoeschele I (1996a) A Monte-Carlo method for Bayesian analysis of linkage between single markers and quantitative trait loci: I. Methodology. *Theor Appl Genet* 92 (in press)

Thaller G, Hoeschele I (1996b) A Monte-Carlo method for Bayesian analysis of linkage between single markers and quantitative trait loci: I. A simulation study. *Theor Appl Genet* 92 (in press)

Uimari P, Thaller G, Hoeschele I (1996a) The use of multiple markers in a Bayesian method for mapping quantitative trait loci. *Genetics* 143, 1831-1842

Uimari P, Zhang Q, Grignola FE, Hoeschele I, Thaller G (1996b) Analysis of QTL workshop I granddaughter design data using least-squares, residual maximum likelihood and Bayesian methods. *J Quantitative Trait Loci* (in press)

van Arendonk JAM, Tier B, Kinghorn BP (1994) Use of multiple genetic markers in prediction of breeding values. *Genetics* 137, 319-329

VanRaden PM, Wiggans GR (1991) Derivation, calculation, and use of national animal model information. *J Dairy Sci* 74, 2737-2746

Weller JI (1986) Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. *Biometrics* 42, 627-640

Weller JI, Kashi Y, Soller M (1990) Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J Dairy Sci* 73, 2525-2537

Xu S, Atchley WR (1995) A random model approach to interval mapping of quantitative trait loci. *Genetics* 141, 1189-1197

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