Theor Appl Genet (1995) 90:65–72

P. Le Roy · J. M. Elsen

Numerical comparison between powers of maximum likelihood and analysis of variance methods for QTL detection in progeny test designs: the case of monogenic inheritance

Received: 2 February 1994 / Accepted: 17 May 1994

Abstract Simulations are used to compare four statistics for the detection of a quantitative trait locus (QTL) in daughter and grand-daughter designs as defined by Soller and Genizi (1978) and Weller et al. (1990): (1) the Fisher test of a linear model including a marker effect within sire or grand-sire effect; (2) the likelihood ratio test of a segregation analysis without the information given by the marker; (3) the likelihood ratio test of a segregation analysis considering the information from the marker; and (4) the lod score which is the likelihood ratio test of absence of linkage between the marker and the QTL. In all cases the two segregation analyses are more powerful for QTL detection than are either the linear method or the lod score. The differences in power are generally limited but may be significant (in a ratio of 1 to 3 or 4) when the QTL has a small effect (0.2 standard deviations) and is not closely linked to the marker (recombination rate of 20% or more).

Key words QTL · Genetic marker · Likelihood ratio test · Segregation analysis

Introduction

The DNA polymorphisms discovered in the seventies enable systematic production of genetic makers (Soller 1990). Large-scale gene-mapping projects have been started recently for domestic animals, and in particular for pigs and cattle. The identification of quantitative trait loci (QTLs) is most often advanced as the main practical interest for these future maps. When a QTL is linked to a marker gene, the quantitative trait distribution in offspring from a double heterozygous parent will differ depending on the marker allele they received from this parent. This is the basic principle used in all QTL-detection designs in plant as well as in animal breeding.

Depending on the biological characteristics of the species and on available experimental facilities, these designs are more or less controlled. In cattle a suggestion has been made to use existing, large half-sib families produced by artificial insemination, an idea which was formulated as early as 1961 by Neimann-Sorensen and Robertson.

More recently, Soller and Genizi (1978) and Weller et al. (1990) evaluated the power of "daughter" and "grand-daughter" designs. The test statistic proposed was the ratio of variances between and within marker class within family. This type of criterion exploits only mean differences between classes of offspring from double heteroygous sires. The maximum likelihood technique is an appealing alternative to this linear approach. It has been described and evaluated in simple genetic situations such as a backcross or an F_2 between pure lines by Weller (1986), Luo and Kearsey (1989, 1991), Simpson (1989, 1992), Lander and Botstein (1989) and Haley (1991).

In the case of outbred populations, the segregation-analysis models designed by MacLean et al. (1984) or by Risch (1984) for nuclear human families must be adapted to a livestock structure. Thus, Knott and Haley (1992) applied maximum likelihood for mapping of QTLs in full-sib families. A first evaluation of these likelihood techniques was published by Le Roy and Elsen (1991) for half-sib families. These were found to be very computer intensive and thus efficient approximations would be welcome (Dentine and Cowan 1990).

In the present paper, we give new comparisons between linear and maximum-likelihood (ML) techniques when applied to daughter and grand-daughter designs. The relative interest of ML is briefly described in algebraic terms, and simulation results are given in simple

Communicated by E. J. Eisen

P. Le Roy (✉)
Institut National de la Recherche Agronomique, Centre de recherches de Jouy en Josas, Station de génétique quantitative et appliquée, 78352 Jouy en Josas Cedex, France

J. M. Elsen
Institut National de la Recherche Agronomique, Centre de recherches de Toulouse, Station d'amélioration génétique des animaux, BP 27, Auzeville 31326 Castanet-Tolosan Cedex, France

cases using models without approximations. In particular, the simulations were limited to the test of monogenic inheritance vs no heredity; no polygenic effect was added to the generated data, nor was it included in the likelihood function. Nevertheless, the general expression of the test statistics is given in order to show its complexity.

Models and test statistics

We consider the case of a single biallelic QTL and a single marker.

The s sires in the daughter design, as well as the t grand-sires in the grand-daughter design, were all assumed to be heterozygous MN at the marker locus, while their genotype at the quantitative locus might have been AA, AB or BB. In the population of these males, the four possible genotypes (MA/NA, MA/NB, MB/NA and MB/NB) were at frequencies p_1, p_2, p_3 and p_4 respectively, with $\sum_i p_i = 1$. In the population of dams, the frequency of the A allele was q. Within a QTL genotype g, the trait was assumed to be normally distributed with a mean μ_g.

Sire design

Amongst the three types of sire families studied by Soller and Genizi (1978), we considered only the half-sib one. Each sire i ($i = 1, \ldots, s$) was given a family of daughters j ($j = 1, \ldots, d$), all measured at the marker locus. $\mathscr{M}_i$, $\mathscr{N}_i$ and $\mathscr{U}_i$ were the subsets of sire i daughters depending on the marker alleles received from i (M, N or unknown).

Linear statistics of Soller and Genizi (1978). Soller and Genizi (1978) considered only informative daughters, i.e., belonging to $\mathscr{M}_i$ or $\mathscr{N}_i$ subsets. Their model may be written as:

$$y_{ij} = \begin{cases} u_i + \frac{1}{2}m_i + e_{ij} & \text{if } j \in \mathscr{M}_i \\ u_i - \frac{1}{2}m_i + e_{ij} & \text{if } j \in \mathscr{N}_i \end{cases}$$

where y_{ij} is the performance of the jth daughter of the sire i, u_i is the sire i effect and m_i the within-sire marker effect, i.e., the effect of the marker received by a daughter from its sire. The residuals, e_{ij}, are assumed to be normally distributed, $\mathscr{N}(0, \sigma_e^2)$.

With this model, a QTL linked to the marker locus is detected through the significance of the m_i effect. The corresponding test statistics was written as SG.

Maximum likelihood. Let G_i be the additive polygenic value of the ith sire, that is twice the sire effect u_i. G_i was assumed to be normally distributed, $\mathscr{N}(0, \sigma_G^2)$. We then considered the following density functions:

the density of y_{ij} given the daughter genotype g [$g \in$(AA, AB, BB)] and the polygenic sire value G_i:

$$f_g(y_{ij}|G_i) = \frac{1}{\sqrt{2\pi}\sigma_e} \exp\left[-\frac{1}{2}\left(\frac{y_{ij} - \mu_g - G_i/2}{\sigma_e}\right)^2\right],$$

the density of the polygenic sire value G_i:

$$f(G_i) = \frac{1}{\sqrt{2\pi}\sigma_G} \exp\left[-\frac{1}{2}\left(\frac{G_i}{\sigma_G}\right)^2\right],$$

the density of y_{ij} given the QTL allele (A or B) received from the sire and G_i:

$$\phi_{Aij} = p f_{AA}(y_{ij}|G_i) + (1-p) f_{AB}(y_{ij}|G_i)$$

$$\phi_{Bij} = p f_{AB}(y_{ij}|G_i) + (1-p) f_{BB}(y_{ij}|G_i).$$

Then, the likelihood of the observations, given the marker information is:

$$\begin{aligned} V_2 = \prod_i \int_{G_i} f(G_i) \Big\{ & p_1 \prod_j \phi_{Aij} \\ & + p_2 \prod_{j \in \mathscr{M}_i} [(1-r)\phi_{Aij} + r\phi_{Bij}] \prod_{j \in \mathscr{N}_i} [(1-r)\phi_{Bij} + r\phi_{Aij}] \\ & \times \prod_{j \in \mathscr{U}_i} [1/2\phi_{Aij} + 1/2\phi_{Bij}] + p_3 \prod_{j \in \mathscr{M}_i} [(1-r)\phi_{Bij} + r\phi_{Aij}] \\ & \times \prod_{j \in \mathscr{N}_i} [(1-r)\phi_{Aij} + r\phi_{Bij}] \prod_{j \in \mathscr{U}_i} [1/2\phi_{Aij} + 1/2\phi_{Bij}] \\ & + p_4 \prod_j \phi_{Bij} \Big\} dG_i \end{aligned}$$

where r is the recombination rate between the QTL and the marker locus. In this expression, the elements of the sum correspond to the genotypes MA/NA, MA/NB, MB/NA and MB/NB for sire i. Note that, as in Elston (1980), V_2 is the likelihood of the observations y_{ij} obtained after integrating out the sire effects G_i, and not the joint likelihood of y_{ij} (observable) and G_i (non-observable).

The likelihood V_1 of the observations, ignoring the marker information, is simply obtained from V_2 by letting $r = 1/2$. It must be emphasized that V_1 is the likelihood of the observations under the hypothesis that a QTL exists and is not linked to the marker locus.

The likelihood V_0 of the observations under the hypothesis of the absence of a QTL segregating in the population is obtained with $\mu_{AA} = \mu_{AB} = \mu_{BB} = \mu$.

Three test statisfics were computed:

SA $= -2\ln(V_0/V_1)$, which is the classical statistic used in segregation analysis (Elston and Steward 1971);

SAM $= -2\ln(V_0/V_2)$, which is the criteria used by McLean et al. (1984) and Risch (1984) when marker information is added in the segregation analysis;

LS $= -2\ln(V_1/V_2)$, which is similar to the lod score (Morton 1955). It tests the absence of linkage between two loci, in this case the marker locus and the QTL, assuming that a QTL is segregating. Thus it should not be used as a test for H_0. However, it was included in our analysis because, as with the ANOVA methods, a segregating QTL will not be detected by LS if it is unliked to the studied marker.

Grand-daughter design

Each of the t grand-sires ($i = 1, \ldots, t$) had s sires ($j = 1, \ldots, s$) with d daughters per son ($k = 1, \ldots, d$). The marker genotypes of the grand-sires, the sires and their dams were known. $\mathscr{M}_i$, $\mathscr{N}_i$ and $\mathscr{U}_i$ were the subsets of grand-sire i sons depending on the marker allele received from i (M, N or unknown).

Linear statistics of Weller et al. (1990). As in the case of Soller and Genizi (1978) for the daughter design, Weller et al. (1990) considered only informative sons and the model is:

$$y_{ijk} = \begin{cases} v_i + \frac{1}{2}m_i + u_{ij} + e_{ijk} & \text{if } j \in \mathscr{M}_i \\ v_i - \frac{1}{2}m_i + u_{ij} + e_{ijk} & \text{if } j \in \mathscr{N}_i \end{cases}$$

where y_{ijk} is the performance of the kth grand-daughter of the jth son of the grand-sire i, v_i is the grand-sire i effect, m_i the within grand-sire marker effect and u_{ij} the within-grand-sire, within-marker, sire ij effect. The residuals e_{ijk} are assumed to be normally distributed, $\mathscr{N}(0, \sigma_e^2)$.

With this model, a QTL linked to the marker locus is detected through the significance of the m_i effect. The corresponding test statistic was written as WKS.

Maximum likelihood. Let G_i and G_{ij} be the grand-sire and sire polygenic values, assumed to be normally distributed, $\mathscr{N}(0, \sigma_G^2)$ and

$\mathcal{N}(0, \sigma'^2_G)$. As in the daughter design, the following density functions were defined:

the density of y_{ijk} given the grand-daughter genotype g and the polygenic sire value G_{ij}:

$$f_g(y_{ijk}|G_{ij}) = \frac{1}{\sqrt{2\pi}\sigma_e} \exp\left[-\frac{1}{2}\left(\frac{y_{ijk} - \mu_g - G_{ij}/2}{\sigma_e}\right)^2\right],$$

the density of the polygenic grand-sire value G_i:

$$f(G_i) = \frac{1}{\sqrt{2\pi}\sigma_G} \exp\left[-\frac{1}{2}\left(\frac{G_i}{\sigma_G}\right)^2\right],$$

the density of the polygenic sire value G_{ij} given the polygenic grand-sire value:

$$f(G_{ij}|G_i) = \frac{1}{\sqrt{2\pi}\sigma'_G} \exp\left[-\frac{1}{2}\left(\frac{G_{ij} - G_i/2}{\sigma'_G}\right)^2\right]$$

with $\sigma'^2_G = 3/4\sigma^2_G$,

the density of y_{ijk} given the QTL allele (A or B) received by daughter ijk from the sire ij and G_{ij}:

$$\phi_{Aijk} = pf_{AA}(y_{ijk}|G_{ij}) + (1-p)f_{AB}(y_{ijk}|G_{ij})$$

$$\phi_{Bijk} = pf_{AB}(y_{ijk}|G_{ij}) + (1-p)f_{BB}(y_{ijk}|G_{ij}).$$

Two new density functions must be considered:

$$\Phi_{Aij} = \int_{G_{ij}} f(G_{ij}|G_i)\left\{p\prod_k \phi_{Aijk} + (1-p)\prod_k (\phi_{Aijk} + \phi_{Bijk})/2\right\} dG_{ij}$$

$$\Phi_{Bij} = \int_{G_{ij}} f(G_{ij}|G_i)\left\{p\prod_k (\phi_{Aijk} + \phi_{Bijk})/2 + (1-p)\prod_k \phi_{Bijk}\right\} dG_{ij}$$

which are the likelihoods of sire ij offspring given the QTL allele (A or B) received by the sire ij from the grand-sire i and G_i.

The algebraic forms of the likelihoods of the observations, given the marker information W_2, and ignoring the marker information W_1 or assuming the absence of QTL W_0, are similar to V_2, V_1 and V_0 replacing ϕ_{Aij} and ϕ_{Bij} by Φ_{Aij} and Φ_{Bij}.

The test statistics were again written as SA, SAM and LS.

What are the advantages of the ML methods?

As explained in the introduction, the principle of QTL detection is the detection of differences in the trait distribution between offspring of double heterozygous parents classified on the marker allele they received. Compared to the test statistics used by Soller and Genizi (1978) and by Weller et al. (1990), the ML techniques should make use of two extra sources of information: the differences between distributions (beyond the difference between means) and the segregation of alleles at the QTL.

Differences between distributions

Tables 1 and 2 give the differences in mean and in variance between offspring classes defined by the allele received from the sire for both designs respectively, where $a = (\mu_{AA} - \mu_{BB})/2$ and $d = (\mu_{AA} - \mu_{AB})/(\mu_{AA} - \mu_{BB})$ (as defined in Morton and McLean 1974).

As previously explained by Soller and Genizi (1978) and Weller et al. (1990), the mean difference is zero for homozygous sires, is half in the grand-daughter design as compared to the daughter design, and decreases to zero when the recombination rate increases to 50%. It must also be noted that this difference is independent of the A allele frequency, p, if the QTL is additive ($d = 1/2$), but varies between 0 and $2a$ (a for the grand-daughter design) for a fully-dominant locus ($d = 0$ or 1).

Table 1 Differences between offspring receiving M, rather than N, from their sire in the daughter design [a]

Sire genes	Mean differences	Variance differences
MA/NA	0	0
MA/NB	$2a(1-2r)[pd + (1-p) \times (1-d)]$	$4a^2(1-2r)p(1-p) \times (2d-1)$
MB/NA	$-2a(1-2r)[pd + (1-p) \times (1-d)]$	$-4a^2(1-2r)p(1-p) \times (2d-1)$
MB/NB	0	0

[a] $a = (\mu_{AA} - \mu_{BB})/2$ and $d = (\mu_{AA} - \mu_{AB})/(\mu_{AA} - \mu_{BB})$, p is the allele A frequency, r is the recombination rate

Table 2 Differences between offspring receiving M, rather than N, from their sire in the grand-daughter design [a]

Grand-Sire genes	Mean differences	Variance differences
MA/NA	0	0
MA/NB	$a(1-2r)[pd + (1-p) \times (1-d)]$	$2a^2(1-2r)[p(1/2 + p - 2p^2) \times d^2 + (1-p)(1/2 - 3p + 2p^2)(1-d)^2 + \frac{1}{2}p \times (1-p)(1-2p)]$
MB/NA	$-a(1-2r)[pd + (1-p) \times (1-d)]$	$-2a^2(1-2r)[p(1/2 + p - 2p^2) \times d^2 + (1-p) \times (1/2 - 3p + 2p^2) \times (1-d)^2 + \frac{1}{2}p \times (1-p)(1-2p)]$
MB/NB	0	0

[a] $a = (\mu_{AA} - \mu_{BB})/2$ and $d = (\mu_{AA} - \mu_{AB})/(\mu_{AA} - \mu_{BB})$, p is the allele A frequency, r is recombination rate

Tables 1 and 2 show that there are differences in variances as well as in means between offspring classes. This result may be generalized to higher moments of the distributions. Figure 1 reports relevant values for these variance differences. For an additive QTL, there is no variance difference for the daughter design and the differences of variance are maximum at fixation in the grand-daughter design. For a dominant QTL, these differences are very sensitive to the A-allele frequency, with a local maximum at $P = 0.5$.

Segregation at the QTL

Due to the segregation of alleles at the QTL, some sires (or grand-sires) are homozygous AA or BB. The presence of such sires decreases the constrasts studied by Soller and Genizi (1978) and Weller et al. (1990), directly affecting the power of their tests. Practically, the apparent differences between genotypes at the QTL are proportionally smaller when the proportion of homozygous sires in greater.

Using ML techniques, we then do a segregation analysis which, like any mixture analysis, estimates the characteristics (mean, variance, proportion) of unknown, sub-distribution components of an observable global distribution. For instance, even if all the sires are known to be AA, estimates of μ_{AA} and μ_{AB} are still obtained from the

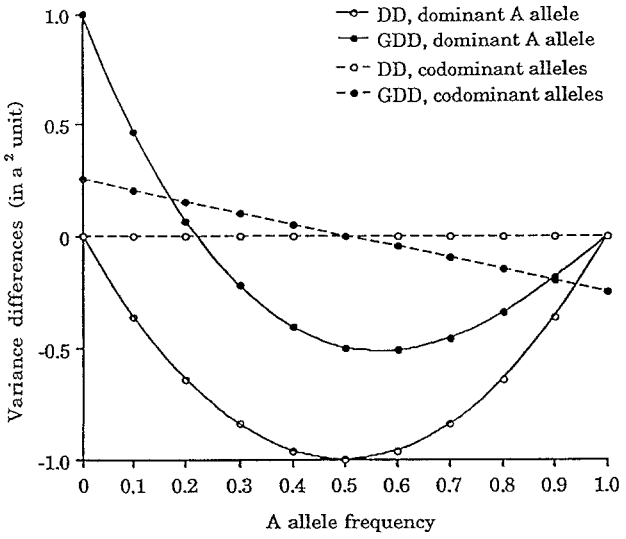

Fig. 1 Variance differences between offspring receiving M, rather than N, from a MA/NB sire. (GDD = grand-daughter design, DD = daughter design)

data. Moreover, posterior probabilities of individual (grand-sires, sires or daughters) genotypes are implicitely used in the parameter estimates, avoiding the bias in within-genotype means estimations.

Similarly, the ML method gives an estimate of the recombination rate, r, when, in the linear approach, separate estimates of the effect and the position of QTL are impossible, the power of the tests decreasing when r increases.

Simulations

Restricted hypotheses

Due to the huge amount of computations needed (about 200 h of CPU on a 3090 IBM computer for the whole study), some restictions were imposed to perform simulations and in the formulation of the likelihood function. First, as stated previously, zero polygenic variance was assumed in the likelihoods, avoiding the quadratures in G_i and G_{ij}, and the heritability of the trait was set to zero in the generated data. Secondly, the likelihoods were expressed assuming Hardy-Weinberg equilibrium at the QTL and linkage equilibrium between the marker locus and the QTL. Thus, the proportions p_1, p_2, p_3 and p_4 have been replaced respectively by p^2, $p(1-p)$, $(1-p)p$ and $(1-p)^2$, removing two parameters to be optimized.

Populations studied

The main comparisons concerned large populations of ten sires and 500 daughters/sire for the daughter design, five grand-sires, 100 sires/grand-sire and ten or 100 daughters/sire, or five grand-sires, 20 sires/grand-sire and 500 daughters/sire for the grand-daughter design. The QTL was defined by its effect ($a = 0.1$ or 0.3 phenotypic standard deviation) and its type (dominant, $d = 1$, or additive, $d = 1/2$). The frequency of p was fixed at 1/2. Four recombination rates (r) were studied, 0, 0.1, 0.2, 0.5. Extra simulations were done for some small populations (ten sires and ten or 20 daughters) in the daughter design in order to evaluate its ability to a detect a major QTL ($a = 1$ and $d = 0$ or 1/2) with a recombination rate of 0, 0.1 and 0.2.

In all cases, we considered a highly-polymorphic marker and assumed that the allele (M or N) received by the daughter or a son from its sire may be identified without ambiguity.

Computing techniques

The simulations were done in Fortran using appropriate NAG routines (G05CCF, G05DDF and G05CAF). The data were generated following the different hypotheses described above, concerning both the normality of within-genotype distribution and the values of the parameters. As suggested by Weller et al. (1990), as linear test statistics for the daughter and the grand-daughter designs we used the F score given by the Type-3 sum of squares in SAS-GLM. Specific Fortran routines were written for computing the likelihoods. These likelihoods were maximised with a quasi-Newton algorithm from the NAG library (E04JBF). Three hundred replications were simulated under each H_0, and 100 under each H_1 situation described above. The maximisation of the likelihood under H_1 was done from four starting points, the best result being retained. Under H_2, two starting points derived from the solution obtained under H_1 were compared. The 5% quantiles under H_0 were obtained with the Harrel and Davis (1982) estimator. The estimated powers under H_1 were simply the proportions of test-statistics values greater than these quantiles in the sample of replicates.

Results

Distribution under H_0

Tables 3 and 4 give the main characteristics of the test-statistics distributions under the H_0 hypothesis. As expected, the SG and WKS criteria follow central Fisher distributions with s and $2s(d-1)$ degrees of freedom for SG, and t and $ts(d-1)$ degrees of freedom for WKS. By constrast, the distributions of the maximum-likelihood

Table 3 Distributions of test statistics under H_0 in large populations. Daughter design

Method[a]	Numbers of		Percentage of 0	Without 0[b]		With 0[b]		Quantiles ± SD (with 0[b])	
	Sires	Daughters/sire		Mean	SD	Mean	SD	At 5%	At 1%
SG	10	500	0	1.02	0.45	1.02	0.45	1.79 ± 0.05	2.35 ± 0.15
SA	10	500	23.6	1.90	2.05	1.45	1.97	5.66 ± 0.57	8.91 ± 0.99
SAM	10	500	13.0	2.62	2.36	2.28	2.37	7.08 ± 0.48	9.91 ± 0.62
LS	10	500	58.7	1.99	1.87	0.82	1.55	4.59 ± 0.59	6.70 ± 0.47

[a] SG = Soller and Genizi (1978) test statistic; SA = Segregation analysis test; SAM = Segregation analysis including information from a marker; LS = Lod Score test of the linkage between QTL and marker

[b] Means, standard deviations and quantiles estimated without or with the 0 values obtained for the likelihood ratio test in the simulations

Table 4 Distributions of test statistics under H_0 in large populations. Grand-daughter design (five grand-sires)

Method[a]	Numbers of		Percentage of 0	Without 0[b]		With 0[b]		Quantiles ± SD (with 0[b])	
	Sons/grand-sire	Daughters/sire		Mean	SD	Mean	SD	At 5%	At 1%
WKS	100	10	0	0.98	0.64	0.98	0.64	2.24 ± 0.14	3.09 ± 0.14
	100	100	0	0.98	0.66	0.98	0.66	2.23 ± 0.10	3.24 ± 0.23
	20	500	0	0.98	0.63	0.98	0.63	2.20 ± 0.08	2.98 ± 0.14
SA	100	10	7.0	2.04	2.13	1.90	2.12	6.33 ± 0.37	9.52 ± 0.92
	100	100	9.7	1.93	2.13	1.74	2.10	6.17 ± 0.33	8.60 ± 0.72
	20	500	19.7	1.85	2.15	1.48	2.06	5.60 ± 0.39	9.46 ± 0.73
SAM	100	10	4.0	2.54	2.37	2.44	2.37	7.20 ± 0.47	10.70 ± 0.90
	100	100	13.7	2.49	2.40	2.19	2.39	6.84 ± 0.46	10.60 ± 0.87
	20	500	20.7	2.43	2.45	1.93	2.39	7.18 ± 0.55	10.04 ± 0.42
LS	100	10	20.7	1.89	2.15	1.50	2.06	5.69 ± 0.57	9.67 ± 1.38
	100	100	19.7	1.97	2.14	1.58	2.07	5.69 ± 0.28	9.35 ± 0.94
	20	500	33.7	1.76	5.00	1.19	4.18	4.94 ± 0.77	13.54 ± 0.40

[a] WKS = Weller et al. (1990) test statistic; SA = Segregation analysis test; SAM = Segregation analysis including information from a marker; LS = Lod score test of the linkage between QTL and marker

[b] Means, standard deviations and quantiles estimated without or with the 0 values obtained for the likelihood ratio test in the simulations

Table 5 Power of the test statistics in large populations at the 5% level. Daughter design

Numbers of		Mean values[a]			Recombination rate	SG[b]	SA[c]	SAM[d]	LS[e]
Sires	Daughters/sire	AA	AB	BB					
10	500	−0.1	0	0.1	0	31	22	33	20
					0.1	22	21	25	16
					0.2	11	23	27	9
					0.5	3	24	24	2
10	500	−0.3	0	0.3	0	98	97	91	99
					0.1	98	99	100	97
					0.2	86	97	100	86
					0.5	4	98	94	2
10	500	0	0	0.2	0	20	20	19	25
					0.1	19	24	24	10
					0.2	8	18	21	9
					0.5	1	14	13	2
10	500	0	0	0.6	0	99	100	96	99
					0.1	97	97	96	99
					0.2	78	100	100	81
					0.5	4	94	94	4

[a] In within genotype standard deviation
[b] SG = Soller and Genizi (1978) test statistic
[c] SA = Segregation analysis test
[d] SAM = Segregation analysis including information from a marker
[e] LS = Lod score test of the linkage between QTL and marker

tests are not the χ^2 distributions expected for large samples (Wilks 1938). In particular, a number of simulations resulted in test statistics close to zero. For SA and SAM, this deviations from the classical asymptotic results described by Wilks (1938) come from the breakdown in regularity conditions generally encountered in mixture problems (e.g., McLachlan and Basford 1987). For LS, it comes from the definition of H_1 as the null hypothesis of the test when simulations were performed under H_0.

Power

The powers of the four test statistics in large populations for the daughter and grand-daughter designs are given in Tables 5 and 6. Concerning our main objective of the comparison between test statistics, the segregation-analysis method SA is often, and SAM is generally, more powerful than the linear statistics SG and WKS (excluding some inversions due to our limited number of simulations). As expected, the extra power is larger when the recombination rate is higher. It is more often quite limited when the QTL and marker locus are totally linked, in particular for larger QTLs. The power may increase by more than 300% with 20% recombination and a small additive QTL. The lod score and linear statistics reach a very similar power, at least for recombination rates different from 0.5. However, the information given by the marker in the segregation analysis is limited. A maximum of 20% extra power is gained when

Table 6 Power of the test statistics in large populations at the 5% level. Grand-daughter design

Numbers of			Mean values [a]			Recombination rate	WKS[b]	SA[c]	SAM[d]	LS[e]
Grand-sires	Sons/grand-sire	Daughters/sire	AA	AB	BB					
5	100	10	−0.1	0	0.1	0	13	9	13	5
						0.1	7	2	5	3
						0.2	8	6	9	2
						0.5	5	8	8	1
5	100	10	−0.3	0	0.3	0	75	65	90	53
						0.1	48	51	75	39
						0.2	40	55	66	10
						0.5	5	54	52	2
5	100	100	−0.1	0	0.1	0	75	63	95	63
						0.1	61	64	85	40
						0.2	37	64	77	18
						0.5	3	72	74	4
5	100	100	−0.3	0	0.3	0	98	100	100	97
						0.1	98	100	100	95
						0.2	90	100	100	88
						0.5	2	100	100	1
5	100	100	0	0	0.2	0	75	66	95	68
						0.1	61	73	91	41
						0.2	33	67	87	31
						0.5	4	66	63	2
5	100	100	0	0	0.6	0	95	100	100	96
						0.1	96	100	100	96
						0.2	94	100	100	88
						0.5	5	100	100	4
5	20	500	−0.1	0	0.1	0	70	90	100	43
						0.1	52	88	98	28
						0.2	25	92	95	9
						0.5	5	94	95	0
5	20	500	−0.3	0	0.3	0	94	100	100	99
						0.1	86	100	100	68
						0.2	55	100	100	42
						0.5	5	100	100	0
5	20	500	0	0	0.2	0	67	91	98	47
						0.1	47	90	95	23
						0.2	33	90	96	13
						0.5	2	93	95	0
5	20	500	0	0	0.6	0	93	100	100	96
						0.1	78	100	100	71
						0.2	62	100	100	40
						0.5	6	100	100	0

[a] In within genotype standard deviation
[b] WKS = Weller et al. (1990) test statistic
[c] SA = Segregation analysis test
[d] SAM = Segregation analysis including information from a marker
[e] LS = Lod score test of the linkage between QTL and marker

the loci are totally linked and when the QTL effect is small.

On the other hand, the result with ML tests are almost unaffected by the type of QTL (additive or dominant) and the population structure (100 sires and 100 grand-daughters or 20 sires and 500 grand-daughters). This result is particularly encouraging since the major limitation in the grand-daughter design will be the number of marker genotype identifications, value $t \times s$, rather than the number of phenotypic observations. Note that the superiority of the ML test over WSK is more important for the last design (20 sires and 500 daughters/sire) where differences in within-sire distribution are more precisely estimated.

Finally, the case of a small population with a large QTL emphasizes the preceeding results (Table 7): the

Table 7 Power of the test statistics in large populations at the 5% level. Daughter design

Numbers of		Mean values[a]			Recombination rate	SG[b]	SA[c]	SAM[d]	LS[e]
Sires	Daughters/sire	AA	AB	BB					
10	10	0	0	2	0	19	41	47	26
		0	0	2	0.1	10	37	43	15
		0	0	2	0.2	10	39	36	6
10	10	0	1	2	0	15	27	28	18
		0	1	2	0.1	18	21	27	20
		0	1	2	0.2	6	19	19	7
10	20	0	0	2	0	56	85	96	71
		0	0	2	0.1	26	85	93	47
10	20	0	0	2	0	56	65	87	65
		0	1	2	0.1	37	56	73	39
		0	1	2	0.2	20	54	59	19

[a] In within genotype standard deviation
[b] SG = Soller and Genizi (1978) test statistic
[c] SA = Segregation analysis test
[d] SAM = Segregation analysis including information from a marker
[e] LS = Lod score test of the linkage between QTL and marker

superiority of maximum-likelihood techniques over SG statistics increases with the recombination rate, reaching nearly 400% in very small populations when $r = 0.2$. The extra gain due to the marker information in segregation analyses in limited; the SG and lod score have very similar behaviour.

Discussion

Our numerical results were obtained under the following genetic hypotheses: a biallelic QTL, Hardy-Weinberg equilibrium, and linkage equilibrium between the marker and the QTL. The described ANOVA methods are still valid under deviations to these hypotheses, for the marker effects are estimated within sire or grandsire. In contrast, the likelihood equations should be modified to account for more complex genetic situations. In general we should expect that increasing the number of parameters to be estimated may decrease the power of these likelihood-based methods, to an unpredictable level. However, for a multiallelic QTL the within-sire differences between offspring classes defined by the sire marker allele will be lowered, increasing the relative interest of likelihood methods.

On the whole, our results, obtained on a limited set of parameters values, show a higher power for segregation-analysis methods as compared with simpler tests based on the marker effect on trait mean. This gain comes from the use of more information than with the analysis of variance tests: the marker effect on the whole-trait distribution, and Mendelian segregation of the trait in daughters, even if the sire or the grand-sire are homozygous at the QTL.

However, the power of the lod score, often used in this kind of study, is not higher than the power of ANOVA tests. This relatively-low performance is due to the definition of the tested hypotheses: the lod score compares H_1 (a QTL exists, which is not linked to the marker) and H_2 (a QTL exists, which is linked to the marker), while the segregation analysis compares H_2 and H_0 (no QTL). The lod score should only be used in cases when a QTL is known, it is a position test and not a detection test. The case of the ANOVA methods is similar since the absence of marker effect may be understood either as the absence of any QTL or as the absence of linkage between an existing QTL and the marker. This ambiguity about the objective of ANOVA tests must be emphasized. It may be described either as a position or as a detection test. Only the latter case was considered in this study. In the former case, we could assume as a very general situation the segregation of genes at an unknown, probably large, number of QTLs. Thus the precise definition of the null hypothesis, including possible linkage between QTLs, is probably intractable.

It must also be emphasized that the extra information provided by the marker may be small in segregation analysis, which is nearly optimal without the marker, at least for the population sizes and structures studied, in particular for larger QTLs and closer linkage. A positive argument for the use of genetic markers is the very good power of the grand-daughter design with only five grand-sires and 20 sons per grand-sire, since the costs of marker identifications may be very limited in such populations. Moreover, detecting linkage between markers and the QTL gives an efficient tool for genetic improvement, in particular for the early and fast selection of favourable allele carriers. Finally, the future multiloci approaches using marker maps, will allow for the breakdown analysis of oligogenic inheritance, which is nearly impossible with classical segregation analysis. One may also expect the costs of typing to decrease rapidly.

The interest of ANOVA methods is their computing simplicity. They reach good power for strong linkage (this will be the usual case when dense maps are available) and must be used as a first approach for QTL detection. Moreover, they may be more robust to deviation from normality, a quality which should be evaluated. Nevertheless, segregation analysis is not only a test but also an estimation method and gives information about the size and position of detected QTLs. Some

effort should now be directed to finding efficient, simplified, segregation-analysis methods for detecting QTLs in livestock.

Indeed, these results were obtained assuming that there was only one QTL segregating. The general situation probably involves the segregation of several QTLs, i.e., an oligogenic (few number of QTLs), a polygenic (large number of small QTLs), or a mixed (one large QTL + a large number of small ones) inheritance. Thus, the conclusions given here must be considered as a first indication and the test statistics actually used must include this extra source of genetic variation. Nevertheless, due to computational difficulties, the likelihood must be approximated using, for instance, the propositions of Hasstedt (1982), Demenais et al. (1990), or Le Roy et al. (1989).

References

Demenais FM, Murigane C, Bonney GE (1990) Search for faster methods of fitting the regressive models to quantitative traits. Genet Epidemiol 7:319–334

Dentine MR, Crown CM (1990) An analytical model for the estimation of chromosome substitution effects in the offspring of individuals heterozygous at a segregating marker locus. Theor Appl Genet 79:775–780

Elston RC (1980) A general model for the genetic analysis of pedigree data. Hum Hered 21:523–542

Elston RC, Stewart J (1971) Segregation analysis. Curr Dev Anthropol Genet 1:327–354

Haley CS (1991) Use of DNA fingerprints for the detection of major genes for quantitative traits in domestic species. Anim Genet 22:259–277

Harrel FE, Davis CE (1982) A new distribution-free quantile estimator. Biometrika 69:635–640

Hasstedt SJ (19982) A mixed-model likelihood approximation on large pedigrees. Comput Biomed Res 15:295–307

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

Le Roy P, Elsen JM (1991) Détection de gènes majeurs chez les animaux: apport des marqueurs. Séminaire de Méribel 2, "Marquage moléculaire et sélection". INRA, Département de génétique et d'amélioration des plantes, pp 79–84

Le Roy P, Elsen JM, Knott SA (1989) Comparison of four statistical methods for detection of a major gene in progeny-test design. Genet Sel Evol 21:341–357

Luo ZW, Kearsey MJ (1989) Maximum-likelihood estimation of linkage between a marker gene and a quantitative locus. Heredity 63:401–408

Luo ZW, Kearsey MJ (1991) Maximum-likelihood estimation of linkage between a marker gene and a quantitative locus. II. Application to backcross and doubled haploid populations. Heredity 66:117–124

MacLachlan GJ, Basford KE (1987) Mixture models: inference and applications to clustering. Marcel Dekker, New York

MacLean CJ, Morton NE, Yee S (1984) Combined analysis of genetic segregation and linkage under and oligogenic model. Computers and Biomed Res 17:471–480

Morton NE (1955) Sequential tests for the detection of linkage. Am J Hum Genet 7:277–318

Morton NE, MacLean CJ (1974) Analysis of family resemblance. III. Complex segregation of quantitative traits. Am J Hum Genet 26:499–503

Niemann-Sorensen A, Robertson A (1961) The association between blood groups and several production characteristics in three Danish cattle breeds. Acta Agric Scand 11:163–196

Risch N (1984) Segregation analysis incorporating linkage markers. I. Single-locus models with an application to type-I diabetes. Am J Hum Genet 36:363–386

Simpson SP (1989) Detection of linkage between quantitative trait loci and restriction fragment length polymorphisms using inbred lines. Theor Appl Genet 77:815–819

Simpson SP (1992) Correction: detection of linkage between quantitative trait loci and restriction fragment length polymorphisms using inbred lines. Theor Appl Genet 85:110–111

Soller M (1990) Genetic mapping of the bovine genome using deoxyribonucleic acid-level markers to identify loci affecting quantitative traits of economic importance. J Dairy Sci 73:2628–2646

Soller M, Genizi A (1978) The efficiency of experimental designs for the detection of linkage between a marker locus and locus affecting a quantitative trait in segregating populations. Biometrics 34:47–55

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

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

Wilks SS (1938) The large sample distribution of the likelihood ratio for testing composite hypotheses. Ann Math Stat 9:60–62