Theor Appl Genet (1997) 95: 236–245

J. M. Elsen · Sara Knott
P. Le Roy · C. S. Haley

Comparison between some approximate maximum-likelihood methods for quantitative trait locus detection in progeny test designs

Received: 3 December 1996 / Accepted: 3 January 1997

Abstract The power and efficiency of parameter estimation of four approximate maximum-likelihood segregation-analysis methods for QTL detection were numerically compared using Monte Carlo simulation. The approximations were designed to avoid the long computation required by exact maximum-likelihood segregation analysis for populations composed of large, independent half-sib families, as found in forest-tree and animal-breeding programs. The methods were compared both when information from a marker closely linked to the QTL was available and when it was not. Three of the approximations were from the literature: the Modal-Estimation method initially developed by Le Roy et al., an approximate Regressive Model from Demenais and Bonney, and the Within-Sire method used by Boichard et al. The fourth method was derived from this Within-Sire method by ignoring between-male-parent information and segregation within families due to the alleles inherited from the female parents. The relative advantages of the criteria are compared for various hypotheses concerning the characteristics of the QTL and the size of the population. No one method was clearly superior over all situations studied. The fourth, and simplest, method, however, performed sufficiently well when marker data were available, particularly in terms of power, for it to provide a tool for rapid preliminary screening of data from QTL mapping studies.

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

Communicated by C. Smith

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

S. Knott
Institute of Cell, Animal and Population Biology, University of Edinburgh, EH9 3JT, Scotland

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

C. S. Haley
Roslin Institute (Edinburgh), Roslin, Midlothian, EH25 9PS Scotland, UK

Introduction

A large number of major genes influencing disease sensitivity in humans have been found using phenotypic and pedigree information. In crop plants, as well as in forest trees, many studies have also reported the existence of such major genes (e.g. Paterson et al. 1991; Bradshaw et al. 1994). To a lesser extent, major genes controlling quantitative traits in farm animals have been discovered by looking at the distribution of phenotypes in controlled populations; for example: double muscling in pigs (Ollivier 1980), low technological yield of ham in pigs (Le Roy et al. 1990), and prolificacy in sheep (Piper and Bindon 1982).

Evidence for these major genes was obtained from observations of performance recorded in populations in which the genes were segregating. The best statistical method to help proving their existence is surely segregation analysis, initially proposed by Elston and Stewart (1971), and its generalisation, complex pedigree analysis, proposed by Morton and McLean (1974). This method is basically a maximum-likelihood approach: given the phenotypes in the pedigree, different models of inheritance are compared on the basis of their likelihood. The most widely tested hypotheses are the polygenic (an infinite number of small independent genes) *versus* the mixed inheritance (a major gene plus an infinite number of small independent genes).

Even when considering independent full- or half-sib families, applying segregation-analysis methods to forest trees or livestock (Plomion and Durel 1996) populations requires a great deal of computing due to the very large number of progeny as compared with human pedigree structures. Simplified statistical criteria have been proposed specifically for these structures (Le Roy et al. 1989; Boichard et al. 1990).

The discovery of DNA polymorphisms and the development of molecular biology tools, such as the PCR, has boosted gene-mapping projects. Marker maps with about one marker locus every 20 cM are becoming available for tree species (pine: Devey et al. 1994; Plomion et al. 1995; poplar: Bradshaw et al. 1994; eucalyptus: Grattapaglia and Sederoff 1994; Byrne et al. 1995) and the major animal species (pigs: Rohrer et al. 1994; cattle: Barendse et al. 1994; Bishop et al. 1994; sheep: Crawford et al. 1994). These maps allow easier and better identification of quantitative trait loci (QTLs). The principle is the identification, in the offspring of an individual, of those which received one or the other of the two chromosomal fragments surrounding the studied marker. If a quantitative locus is located on this fragment, and if the parent is heterozygous at both the marker and the QTL, then a systematic difference will be observed between the two groups of offspring. The idea is old (Sax 1923), and its application to outbred populations, where large families are recorded in the classical context of selection schemes, has been studied a number of times recently (Niemann–Sorensen and Robertson 1961; Soller and Genizi 1978; Weller, Kashi and Soller 1990; O'Malley and Mckeand 1994; Bradshaw and Stettler 1995).

The maximum-likelihood methods described above may be generalised to situations where markers are available. The first extensions to include markers concerned human populations (MacLean et al. 1984; Risch 1984). Further extensions have been made for livestock situations, including tests for monogenic transmission in half sibs (Bovenhuis and Weller 1994; Le Roy and Elsen 1994), and for mixed transmission in full sibs (Knott and Haley 1992) and single half-sib families (Georges et al. 1995). The numerical difficulties remain, and specific simplifications when considering mixed inheritance in large families are needed.

A least-squares approach applied to QTL detection in crosses between lines seems to be an efficient alternative to maximum likelihood (Knapp et al. 1990; Haley and Knott 1992; Haley et al. 1994). Considering monogenic inheritance in outbred populations, Le Roy and Elsen (1994) found that the superiority of maximum likelihood over least squares increases when the distance between the QTL and marker locus increases or when the QTL effect decreases when considering linkage to a single marker.

In the present paper, we compare different simplified segregation-analysis methods using quantitative trait measurements only or quantitative trait measurements plus information from a linked marker. Their numerical properties based on simulations (power and precision of parameter estimations) are described. The use of one of these simplified methods in the case of multiple markers, and its comparison with least squares is given elsewhere (Knott et al. 1996).

Methods

Models

General hypotheses and notation

The population studied was a set of s half-sib families each with a common male parent and with d progeny per family. The s male and sd female parents were assumed to be unrelated. y_{ij} was the performance of the j^{th} progeny ($j = 1, \ldots, d$) of the male i ($i = 1, \ldots, s$).

Under the mixed-inheritance hypothesis H_1, a major gene was segregating with two alleles A and B. The major genotype g of an individual was AA, AB or BB. These three genotypes g_i had probabilities p_1, p_2 and p_3 in the male-parent population and the A allele frequency in the female-parent population was q. The major locus genotype of the j^{th} progeny of the i^{th} male parent was g_{ij}, with $\gamma_i = (g_{i1}, g_{i2}, \ldots, g_{id})$.

Under H_1, within the major genotype, g, the trait was assumed to be normally distributed, $\mathcal{N}(\mu_g, \sigma^2)$. Under the polygenic-inheritance hypothesis, H_0, the trait was assumed to be normally distributed, $\mathcal{N}(0, \sigma^2)$, in the global population.

The male-parent effect u_i of the i^{th} male parent, that is half the additive polygenic value G_i, was assumed to be normally distributed, $\mathcal{N}(0, \sigma_u^2)$. Given the major genotype (under H_1) and the male-parent effect u_i, the residuals of progeny phenotype were distributed as normal, $\mathcal{N}(0, \sigma_e^2)$.

The comparison of methods with a marker locus available was based on the following assumptions: all the male parents were heterozygous MN at the marker locus; the marker male-parent allele was identifiable in all progeny; and the QTL and marker locus were completely linked. $\mathcal{M}_i$ and $\mathcal{N}_i$ were the subsets of the progeny of male parent i depending on the marker allele received from i (M, N). Assuming linkage equilibrium in the male-parent population, the four possible genotypes h_i for these male parents (MA/NA, MA/NB, MB/NA and MB/NB) were at frequencies $p_1, p_2/2$, $p_2/2$ and p_3 respectively.

The marker allele received by the j^{th} progeny from its male parent i was v_{ij}, with $v_i = (v_{i1}, v_{i2}, \ldots, v_{id})$.

The density of y normally distributed, $\mathcal{N}(\mu, \sigma^2)$, was written as $f(y|\mu, \sigma)$ and, in the multidimensional situation, the density of the observation vector $\mathbf{y}$ with mean μ and variance-covariance matrix $\mathbf{V}$, was written as $f(\mathbf{y}|\mu, \mathbf{V})$.

In the following we describe only the H_1 hypothesis since the algebra under H_0 can be directly deduced from this general situation. The exact likelihood will first be derived, followed by four approximations. In each case, the corresponding test statistics are approximated likelihood ratio tests, i.e. $-2 \times \log$ of the ratio of the maximum of the approximate likelihoods under H_0 and H_1.

The exact likelihood

Without any approximation, the likelihood L without marker information of the (y_{ij}, $i = 1, \ldots, s$, $j = 1, \ldots, d$) may be written as:

$$L = \prod_{i=1}^{s} \sum_{g_i} p(g_i) \sum_{\gamma_i} p(\gamma_i / g_i) f(\mathbf{y}_i | (\mu_{g_{ij}})_{\gamma_i}, \mathbf{V}),$$

where the vector $(\mu_{g_{ij}})_{\gamma_i}$ is the vector $(\mu_{g_{i1}}, \mu_{g_{i2}}, \ldots, \mu_{g_{id}})$ corresponding to the γ_i vector of major genotypes, and where $\mathbf{V}$, the variance-covariance matrix of one of the male-parent families, is given by:

$$\mathbf{V} = \mathbf{I}_d \sigma_e^2 + \mathbf{J}_d \sigma_u^2,$$

$\mathbf{I}_d$ and $\mathbf{J}_d$ being, respectively, the identity matrix and a matrix of 1s.

The conditional probabilities $p(g_{ij}/g_i)$, elements of $p(\gamma_i/g_i)$, are a direct function of q, the A allele frequency of the female-parent population.

The summation in γ_i comprises 2^d or 3^d terms, depending on g_i. An alternative formulation of L avoids this difficulty:

$$L = \prod_{i=1}^{s} \int_{u_i} f(u_i|0,\sigma_u) \sum_{g_i} p(g_i) \prod_{j=1}^{d} \sum_{g_{ij}} p(g_{ij}|g_i) f(y_{ij}|\mu_{g_{ij}} + u_i, \sigma_e) du_i.$$

Nevertheless, with this alternative formulation, new numerical difficulties arise from the integration in u_i.

When a marker is available, the likelihood L is modified to consider the additional information:

$$L = \prod_{i=1}^{s} \sum_{h_i} p(h_i) \sum_{\gamma_i} p(\gamma_i/h_i, v_i) f(\mathbf{y}_i | \{\mu_{g_{ij}}\}_{\gamma_i}, \mathbf{V}).$$

The alternative formulation of the likelihood is now given by:

$$L = \prod_{i=1}^{s} \int_{u_i} f(u_i|0,\sigma_u) \sum_{h_i} p(h_i) \times \left[\prod_{j=1}^{d} \sum_{g_{ij}} p(g_{ij}|h_i, v_{ij}) f(y_{ij}|\mu_{g_{ij}} + u_i, \sigma_e) \right] du_i.$$

First approximation: the modal estimation method ME

Le Roy et al. (1989) and Hoeschele (1988) proposed to approximate the exact likelihood by the joint likelihood of the observations and a "modal estimation", u_i, of the male-parent effects u_i. With this approximation, the quadrature in male-parent effect, u_i, is replaced by a maximisation of the likelihood in u_i, an iterative algorithm based on the gradient method being proposed. An extension of this method was proposed by Knott et al. (1991a, b) and by Elsen and Le Roy (1990), where the male-parent effects were estimated within the major genotype, g_i. The method proved to be powerful and less computationally demanding than the true segregation analysis. The approximate likelihood L_{ME} is given by:

$$L_{ME} = \prod_{i=1}^{s} \frac{1}{\sqrt{|\mathbf{V}|}} \sum_{g_i} p(g_i) f(u_{ig_i}|0,\sigma_u) \times \left[\prod_{j=1}^{d} \sum_{g_{ij}} p(g_{ij}|g_i) f(y_{ij}|\mu_{g_{ij}} + u_{ig_i}, \sigma_e) \right].$$

When a marker is available, this formula is modified accordingly:

$$L_{ME} = \prod_{i=1}^{s} \frac{1}{\sqrt{|\mathbf{V}|}} \sum_{h_i} p(h_i) f(u_{ih_i}|0,\sigma_u) \times \left[\prod_{j=1}^{d} \sum_{g_{ij}} p(g_{ij}|h_i, v_{ij}) f(y_{ij}|\mu_{g_{ij}} + u_{ih_i}, \sigma_e) \right]$$

Second approximation: the regressive model RE

The regressive model was proposed by Bonney (1984, 1986). Its main advantage is its flexibility for considering different patterns of statistical dependencies. Demenais and Bonney (1989) proved the equivalence of the mixed (major gene + polygenic inheritance) and "class-D" regressive models, where equal sib-sib correlations are considered. The numerical difficulties of the exact mixed-segregation analysis remains in the regressive model. Demenais et al. (1990) compared approximations of the class-D model. Some of them proved to be very efficient in terms of power and parameter estimates. Their 4th approximation, similar to the approximation used by Hasstedt (1982) in her PAP software for pedigree analysis, was included in our comparison. Applying Bonney's regressive model to the half-sib structure considered here without phenotypic information on parents, the true likelihood may be written as:

$$L = \prod_{i=1}^{s} \sum_{g_i} p(g_i) \left[\sum_{g_{i_1}} p(g_{i1}|g_i) f(y_{i1}/g_i, g_{i1}) \times \left(\sum_{g_{i_2}} p(g_{i2}|g_i) f(y_{i2}/g_i, g_{i1}, g_{i2}, y_{i1}) \cdots \left(\sum_{g_{id}} p(g_{id}|g_i) f(y_{id}/g_i, g_{i1}, g_{i2}, \ldots, g_{id-1}, y_{i1} \cdots y_{id-1}) \right) \cdots \right) \right]$$

with $f(y_{ij}/g_i, g_{i1}, \ldots, g_{ij-1}, y_{i1} \cdots y_{ij-1}) = f(y_{ij}|m_{ij}, v_j)$, where

$$m_{ij} = \mu_{g_{ij}} + [cov(y_{ij}, y_{i1}), \ldots, cov(y_{ij}, y_{ij-1})] [\mathbf{V_j}]^{-1} \times \begin{bmatrix} y_{i1} - \mu_{g_{i1}} \\ \vdots \\ y_{ij-1} - \mu_{g_{ij-1}} \end{bmatrix}$$

and

$$v_j = var(y_{ij}) - [cov(y_{ij}, y_{i1}), \ldots, cov(y_{ij}, y_{ij-1})] [\mathbf{V_j}]^{-1} \times \begin{bmatrix} cov(y_{ij}, y_{i1}) \\ \vdots \\ cov(y_{ij}, y_{ij-1}) \end{bmatrix},$$

$\mathbf{V_j}$ being the variance-covariance matrix of the $y_{i1}, \ldots, y_{ij-1}$.

The numerical diffculties result from the presence of the $\mu_{g_{il}}$ $(l = 1, \ldots, j-1)$ in the m_{ij}, i.e. before the summation in g_{il}. In the approximation proposed by Demenais et al. (1990), the mean value $\mu_{g_{il}}$ is replaced by an averaged value $\tilde{\mu}_{il}$ given by:

$$\tilde{\mu}_{il} = \sum_{g_{il}} \mu_{g_{il}} \frac{p(g_{il}/g_i) \exp\left[-\frac{1}{2} \left(\frac{y_{il} - \mu_{g_{il}}}{\sigma} \right)^2 \right]}{\sum_g p(g/g_i) \exp\left[-\frac{1}{2} \left(\frac{y_{il} - \mu_g}{\sigma} \right)^2 \right]},$$

with $\sigma^2 = \sigma_e^2 + \sigma_u^2$.

This simplification gives the following approximate likelihood:

$$L_{RE} = \prod_{i=1}^{s} \sum_{g_i} p(g_i) \prod_{j=1}^{d} \sum_{g_{ij}} p(g_{ij}|g_i) f\left(y_{ij} | \mu_{g_{ij}} + \beta_j \sum_{l=1}^{j-1} (y_{il} - \tilde{\mu}_{il}), v_j \right)$$

with $\beta_j = \dfrac{\sigma_u^2}{\sigma_e^2 + (j-1)\sigma_u^2}$, $\beta_1 = 0$ and $v_j = \sigma_e^2 \dfrac{\sigma_e^2 + j\sigma_u^2}{\sigma_e^2 + (j-1)\sigma_u^2}$.

When a marker is available the previous approximate likelihood must be changed as follows:

$$L_{RE} = \prod_{i=1}^{s} \sum_{h_i} p(h_i) \prod_{j=1}^{d} \sum_{g_{ij}} p(g_{ij}|h_i, v_{ij}) \times f\left(y_{ij} | \mu_{g_{ij}} + \beta_j \sum_{l=1}^{j-1} (y_{il} - \hat{\mu}_{il}), v_j \right),$$

with

$$\hat{\mu}_{il} = \sum_{g_{il}} \mu_{g_{il}} \frac{p(g_{il}/h_i) \exp\left[-\frac{1}{2} \left(\frac{y_{il} - \mu_{g_{il}}}{\sigma} \right)^2 \right]}{\sum_g p(g/h_i) \exp\left[-\frac{1}{2} \left(\frac{y_{il} - \mu_g}{\sigma} \right)^2 \right]}.$$

Third approximation: the within-sire method WS

Boichard et al. (1988) proposed another simplification based on the distributional properties of the deviation of offspring phenotypes

from the within-family mean, $y_{ij} - \overline{y_i}$. The initial data $\mathbf{y_i}$ were linearly transformed in $\mathbf{z_i} = \mathbf{T}\mathbf{y_i}$, using the $\mathbf{T}$ matrix:

$$\mathbf{T} = \begin{pmatrix} 1-\frac{1}{d} & -\frac{1}{d} & \cdots & -\frac{1}{d} & -\frac{1}{d} \\ -\frac{1}{d} & 1-\frac{1}{d} & \cdots & -\frac{1}{d} & -\frac{1}{d} \\ \vdots & \vdots & \ddots & \vdots & \\ -\frac{1}{d} & -\frac{1}{d} & \cdots & 1-\frac{1}{d} & -\frac{1}{d} \\ \frac{1}{d} & \frac{1}{d} & \cdots & \frac{1}{d} & \frac{1}{d} \end{pmatrix}$$

giving $\mathbf{z_i'} = (y_{i1} - \overline{y_i}, y_{i2} - \overline{y_i}, \ldots, y_{id-1} - \overline{y_i}, \overline{y_i})$. Thus the density of $\mathbf{z_i}$ conditional on v_i was proportional to the corresponding density of $\mathbf{y_i}$, with $E(\mathbf{z_i}/\gamma_i) = \mathbf{T}.E(\mathbf{y_i}/\gamma_i)$ and $var(\mathbf{z_i}/\gamma_i) = \mathbf{T.V.T}'$.

The principle of the Boichard et al. (1988) method was to replace the $\mathbf{z_i}$ density by an asymptotic approximation. After some algebra, it is found that:

$$var(\mathbf{z_i}/\gamma_i) = \begin{pmatrix} \sigma_e^2\left(\mathbf{I}_{d-1} - \frac{\mathbf{1}}{d-1}\mathbf{J}_{d-1}\right) & 0 \\ 0 & \sigma_u^2 + \frac{1}{d}\sigma_e^2 \end{pmatrix} = \mathbf{W}.$$

As the number d of progeny increases, the variance matrix $\mathbf{W}$ tends to

$$\begin{pmatrix} \sigma_e^2\mathbf{I}_{d-1} & 0 \\ 0 & \sigma_u^2 \end{pmatrix}.$$

Conditional on the genotypes g_i and γ_i, the expectation of the elements of $\mathbf{z_i}$ are $\mu_{g_{ij}} - \overline{\mu_{\gamma_i}}$ for $j \in [1, d-1]$ and $\overline{\mu_{\gamma_i}} = \sum_{j=1}^{d} \mu_{g_{ij}}/d$ for $j = d$. Conditional on g_i alone, the quantity $\overline{\mu_{\gamma_i}}$ is a random variable the distribution of which depends on the A allele frequency q in the female parents. The expectation of this random variable depends on g_i:

g_i	$= AA$	AB	BB
$E(\overline{\mu_{\gamma_i}}/g_i) =$	$q\mu_1 + (1-q)\mu_2$	$\frac{1}{2}[q\mu_1 + \mu_2 + (1-q)\mu_3]$	$q\mu_2 + (1-q)\mu_3$

and will be noted μ_{g_i}.

Asymptotically, $E(\mathbf{z_i}/\gamma_i)$ tends towards its expectation $\mu_{g_{i_1}} - \mu_{g_i}, \ldots, \mu_{g_{id-1}} - \mu_{g_i}, \mu_{g_i}$. Finally, Boichard et al. (1988) used, as an approximate likelihood:

$$L_{WS} = \prod_{i=1}^{s} \sum_{g_i} p(g_i) f(z_{id}|\mu_{g_i}, \sigma_u) \left[\prod_{j=1}^{d-1} \sum_{g_{ij}} p(g_{ij}|g_i) f(z_{ij}|\mu_{g_{ij}} - \mu_{g_i}, \sigma_e) \right].$$

This formula may be modified to consider the case where a marker is available:

$$L_{WS} = \prod_{i=1}^{s} \sum_{h_i} p(h_i) f(z_{id}|\mu_{h_i}, \sigma_u) \left[\prod_{j=1}^{d-1} \sum_{g_{ij}} p(g_{ij}|h_i, v_{ij}) f(z_{ij}|\mu_{g_{ij}} - \mu_{g_i}, \sigma_e) \right].$$

Fourth approximation: *only the within-family deviation method WO*

This is a direct simplification of the previous method. It is assumed: (1) that information from the variability among male-parent means is limited as compared to the information from the variability among within male-parent variances, (2) that the non-normality of within-male-parent distribution is negligible in homozygous (AA or BB) male-parent families as compared to heterozygous AB male-parent families. With the first assumption, the $f(z_{id} = \overline{y_i}|\mu_{gi}, \sigma_u)$ term is neglected. With the second assumption, the within AA (or BB) male-parent distribution is assumed to be normal with a 0 expectation $[0 = qE(z_{ij}/g_i = AA,\ g_{ij} = AA) + (1-q)E(z_{ij}/g_i = AA, g_{ij} = AB)]$ and variance σ_w^2; and the within-AB male-parent distribution is a 1/2-1/2 mixture of a normal $\mathcal{N}(\frac{\alpha}{2}, \sigma_w^2)$ and a normal $\mathcal{N}(-\frac{\alpha}{2}, \sigma_w^2)$, with α, the substitution effect (Falconer 1989): $\frac{\alpha}{2} = qE(z_{ij}/g_i = AB, g_{ij} = AA) + (1-q)E(z_{ij}/g_i = AB,\ g_{ij} = AB) = -[qE(z_{ij}/g_i = AB, g_{ij} = AB) + (1-q)E(z_{ij}/g_i = AB, g_{ij} = BB)]$. The following resulting simplified likelihood is obtained:

$$L_{WO} = \prod_{i=1}^{s} \left\{ p\left[\prod_{j=1}^{d} f(z_{ij}|0, \sigma_w) \right] + (1-p)\left[\prod_{i=1}^{d} \tfrac{1}{2} f(z_{ij}|\tfrac{\alpha}{2}, \sigma_w) + \tfrac{1}{2} f(z_{ij}|-\tfrac{\alpha}{2}, \sigma_w) \right] \right\}$$

with $p = p(g_i = AA \text{ or } g_i = BB)$.

This likelihood should be modified when a marker is available:

$$L_{WO} = \prod_{i=1}^{s} \left\{ p\left[\prod_{j=1}^{d} f(z_{ij}|0, \sigma_w) \right] + \frac{1-p}{2}\left[\prod_{j\in\mathcal{M}} f(z_{ij}|\tfrac{\alpha}{2}, \sigma_w) \right]\left[\prod_{j\in\mathcal{N}} f(z_{ij}|-\tfrac{\alpha}{2}, \sigma_w) \right] + \frac{1-p}{2}\left[\prod_{j\in\mathcal{M}} f(z_{ij}|-\tfrac{\alpha}{2}, \sigma_w) \right]\left[\prod_{j\in\mathcal{N}} f(z_{ij}|\tfrac{\alpha}{2}, \sigma_w) \right] \right\}.$$

Comparison of the methods

A numerical evaluation of these methods was performed using Monte Carlo simulation. Two population structures (10 or 20 male parents with 100 progeny each) and six types of major genes (with 2, 1 or 0.5 phenotypic standard deviations between QTL genotypes AA and BB, with complete dominance or additivity) were considered. All the data were generated with a within-major-genotype variance of 1 and an heritability of 0.20, giving $\sigma_u = 0.224$ and $\sigma_e = 0.975$. The A allele frequency was 0.5 in the female parents and the male parents (with the assumption of Hardy Weinberg equilibrium).

The simulations were written in Fortran using appropriate NAG routines (G05CCF, G05DDF and G05CAF; Numerical Algorithms Group 1990). Specific Fortran routines were written for computing the likelihoods. These likelihoods were maximized with a quasi-Newton algorithm from the NAG library (E04JBF). Two-thousand replicate data sets were simulated under each H_0 and 200 under each H_1 situation described above. The maximisation of the likelihood under H_0 was obtained algebraically and under H_1 was carried out from three different starting points based around the parameters used to simulate the data, the best result being retained. The 10% 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-statistic values greater than these quantiles in the sample of replicates. Eight parameters were estimated under H_1 for the first three methods: the means μ_1, μ_2, μ_3, the variances σ_u^2, σ_e^2, the genotype frequencies in the male-parent population p_1, p_2, and the female-parent allele A frequency q. Three parameters were estimated for the fourth method: the substitution effect α, the within-family variance σ_w^2, and the frequency of heterozygous male parents $(1-p)$.

From these parameters, we computed, for the first three methods:

a the additive effect of the gene $[(\mu_3 - \mu_1)/2]$
d the dominance effect $[\mu_2 - (\mu_1 + \mu_3)/2]$ and
α the substitution effect as defined in the 4th model
Note that, $\alpha = a + d(1-2q)$, giving $a = \alpha$ when $q = 0.5$.

The quality of parameter estimates was characterised by the mean and standard deviation of their empirical distribution.

Results

Computational efficiency

The CPU time needed on a 3090 IBM computer to reach convergence was estimated on ten replications in the situation of an additive QTL with two standard-deviation effect, considering a population of 20 families. Without markers, 35, 41, 36 and 3 s were respectively needed for ME, RE, WS and WO methods to reach the solution. With markers, the times needed were 30, 49, 22 and 1 s. These figures show clearly the large superiority of the simplest method, the three others not being very different.

Table 1 Test statistic distribution under H_0

Population	Marker	Test	Mean	Variance	% 0[a]	10% Quantile
10 males	No	ME	3.06	3.33	28	7.82
		RE	4.96	3.20	3	9.16
		WS	5.76	3.82	1	10.82
		WO	0.63	1.38	62	2.22
	Yes	ME	1.34	2.77	57	4.97
		RE	4.70	3.31	4	9.11
		WS	4.81	3.70	4	9.71
		WO	0.51	1.17	54	1.77
20 males	No	ME	2.52	3.33	38	7.23
		RE	5.56	3.56	4	10.38
		WS	5.38	3.57	1	10.14
		WO	0.78	1.56	58	2.74
	Yes	ME	1.06	2.43	54	4.07
		RE	4.93	3.28	4	9.69
		WS	4.79	3.46	4	9.30
		WO	0.55	1.26	52	1.78

[a] Percentage of zero test statistics

Distribution under H_0

The results are summarised in Table 1. Many simulations gave a zero value for the Modal Estimation and Within-sire-Only test statistics. This is a common feature of these types of method. (e.g. McLachlan and Basford 1987). Nevertheless, the results for the situations with 10 and 20 male-parent families are relatively consistent when comparing the distribution characteristics.

Power

Table 2 gives the empirical power at the 10% level. As expected, the power decreases with the effect of the gene and with the number of male parents, with a few exceptions which are probably due to the limited number of simulations.

Without marker information, a dominant gene with an effect of two standard deviations between homozygotes was detected by all methods. For the dominant gene of intermediate effect, RE performed best and WS worse, although for the gene of smallest effect this was reversed. When the simulated gene was additive in effect all approaches gave a lower power than that observed for a dominant gene with the same additive effect. Again, at intermediate powers (i.e. for the large additive gene) RE performed better than the others.

The inclusion of marker information in the analysis dramatically increased power, especially for the QTL with an additive effect. QTLs with an additive effect of 0.5 or 1 (i.e. one or two standard deviations between homozygotes) are detected in nearly all replicates

Table 2 Percentage of replicates with a test statistic greater than 10% quantile of the distribution of the test statistic under $H0$

Population structure	Type of gene (μ_1, μ_2, μ_3)	Power							
		Without marker				With marker			
		ME	RE	WS	WO	ME	RE	WS	WO
10 males	0 0 2	100	100	100	99	100	100	100	100
	0 0 1	28	33	19	32	97	91	99	97
	0 0 0.5	14	13	17	16	57	31	30	49
	0 1 2	41	46	35	39	100	100	100	98
	0 0.5 1	17	17	12	19	98	89	91	96
	0 0.25 0.5	10	15	11	14	57	27	33	51
20 males	0 0 2	100	100	100	100	99	100	100	100
	0 0 1	47	49	41	43	100	100	100	99
	0 0 0.5	15	7	17	14	78	40	49	74
	0 1 2	40	63	53	52	97	100	100	100
	0 0.5 1	10	22	14	14	100	98	98	100
	0 0.25 0.5	10	15	14	9	72	48	47	72

whether their effect is additive or dominant. For both the additive and dominant QTLs with smallest effect WO and ME have higher power, especially with the higher number of male parents.

Parameter estimates

Table 3 gives the means and standard deviations of the estimates for the additive and dominance effects of the QTL, the variances and the female-parent A allele frequency for the first three methods (ME, RE and WS). The results are shown for the population of 20-male-parent families, the 10-male-parent-families population behaving quite similarly. For the largest gene the estimates for the effect of the QTLs are unbiased with all three methods giving similar precision. With smaller effect genes, the additive effect is overestimated, with an increasing bias as the gene effect diminishes. The inclusion of the marker information decreases the bias for the smaller effect genes, such that the estimates for the intermediate gene now look correct, although the smallest gene effect is still overestimated. Considering the dominance effect, WS is, on average, better when additive QTLs are simulated but the standard errors are high. The estimates from ME are more precise for the smaller effect genes.

The male-parent variance, σ_u^2, is underestimated except in situations with the large QTLs analysed using RE or WS. WS gives the best results with highest precision and ME the worst. The estimates are less biased when marker information is included and when more male parents (20 compared with 10 male parents) are incorporated in the analysis (data not shown).

Table 3 Mean parameter estimates for the first three approximations with the empirical standard deviation over the replicates in parentheses. (20 male parent families)

Type of gene[a] (μ_1, μ_2, μ_3)	Parameter (true value)[b]	Test statistics[c]					
		Without marker			With marker		
		ME	RE	WS	ME	RE	WS
0 0 2	a(1)	1.03 (1.03)	1.01 (0.15)	1.01 (0.14)	1.00 (0.07)	0.99 (0.09)	0.99 (0.09)
	d(−1)	−0.95 (0.19)	−1.00 (0.23)	−0.99 (0.22)	−1.00 (0.10)	−1.01 (0.13)	−1.01 (0.12)
	σ_e (0.975)	0.97 (0.03)	0.96 (0.03)	0.96 (0.03)	0.97 (0.02)	0.97 (0.02)	0.97 (0.02)
	σ_u(0.224)	0.17 (0.07)	0.24 (0.08)	0.24 (0.05)	0.20 (0.04)	0.25 (0.06)	0.25 (0.04)
	q(0.5)	0.50 (0.06)	0.50 (0.06)	0.51 (0.06)	0.51 (0.11)	0.50 (0.04)	0.50 (0.03)
0 1 2	a(1)	0.89 (0.21)	0.95 (0.49)	1.00 (0.21)	0.99 (0.18)	0.99 (0.32)	1.01 (0.20)
	d(0)	−0.07 (0.27)	−0.22 (0.73)	−0.02 (0.27)	−0.07 (0.13)	−0.03 (0.44)	0.00 (0.21)
	σ_e (0.975)	1.02 (0.07)	1.00 (0.06)	0.97 (0.09)	0.99 (0.03)	1.00 (0.05)	0.98 (0.04)
	σ_u(0.224)	0.19 (0.12)	0.22 (0.12)	0.21 (0.06)	0.21 (0.05)	0.24 (0.07)	0.24 (0.04)
	q(0.5)	0.49 (0.14)	0.43 (0.25)	0.50 (0.20)	0.51 (0.12)	0.45 (0.26)	0.50 (0.18)
0 0 1	a(0.5)	0.69 (0.16)	0.70 (0.27)	0.57 (0.26)	0.52 (0.11)	0.53 (0.19)	0.48 (0.17)
	d(−0.5)	−0.33 (0.27)	−0.47 (0.41)	−0.62 (0.33)	−0.48 (0.17)	−0.58 (0.26)	−0.61 (0.26)
	σ_e (0.975)	0.96 (0.05)	0.94 (0.05)	0.94 (0.05)	0.97 (0.03)	0.96 (0.03)	0.95 (0.04)
	σ_u(0.224)	0.14 (0.09)	0.17 (0.10)	0.20 (0.07)	0.19 (0.06)	0.23 (0.08)	0.23 (0.05)
	q(0.5)	0.53 (0.14)	0.56 (0.22)	0.54 (0.20)	0.51 (0.08)	0.53 (0.15)	0.50 (0.14)
0 0.5 1	a(0.5)	0.61 (0.17)	0.67 (0.39)	0.65 (0.34)	0.50 (0.11)	0.54 (0.39)	0.55 (0.32)
	d(0)	−0.19 (0.33)	−0.16 (0.65)	−0.02 (0.52)	−0.02 (0.19)	−0.06 (0.59)	−0.05 (0.50)
	σ_e (0.975)	0.96 (0.06)	0.94 (0.05)	0.94 (0.05)	0.97 (0.02)	0.96 (0.03)	0.96 (0.03)
	σ_u(0.224)	0.18 (0.10)	0.16 (0.11)	0.18 (0.07)	0.21 (0.07)	0.22 (0.08)	0.23 (0.06)
	q(0.5)	0.50 (0.12)	0.44 (0.28)	0.50 (0.29)	0.49 (0.07)	0.44 (0.29)	0.49 (0.27)
0 0.5	a(0.25)	0.59 (0.16)	0.50 (0.65)	0.52 (0.34)	0.40 (0.13)	0.46 (0.39)	0.45 (0.30)
	d(−0.25)	−0.26 (0.32)	−0.40 (0.75)	−0.44 (0.54)	−0.13 (0.28)	−0.36 (0.58)	−0.29 (0.59)
	σ_e (0.975)	0.95 (0.06)	0.93 (0.05)	0.91 (0.06)	0.96 (0.03)	0.94 (0.04)	0.93 (0.05)
	σ_u(0.224)	0.16 (0.09)	0.16 (0.09)	0.17 (0.06)	0.16 (0.07)	0.19 (0.08)	0.21 (0.05)
	q(0.5)	0.50 (0.14)	0.52 (0.28)	0.49 (0.29)	0.51 (0.07)	0.55 (0.27)	0.51 (0.26)
0 0.25 0.5	a(0.25)	0.60 (0.20)	0.55 (0.38)	0.58 (0.44)	0.38 (0.12)	0.44 (0.37)	0.39 (0.27)
	d(0)	−0.22 (0.38)	−0.25 (0.67)	−0.03 (0.73)	−0.03 (0.27)	−0.19 (0.72)	−0.01 (0.65)
	σ_e (0.975)	0.94 (0.06)	0.91 (0.05)	0.91 (0.05)	0.96 (0.02)	0.92 (0.05)	0.92 (0.05)
	σ_u(0.224)	0.17 (0.09)	0.13 (0.09)	0.16 (0.06)	0.16 (0.07)	0.18 (0.09)	0.21 (0.06)
	q(0.5)	0.49 (0.13)	0.52 (0.28)	0.53 (0.29)	0.50 (0.06)	0.51 (0.26)	0.49 (0.24)

[a] The simulated effect of the QTL genotypes in within-QTL-genotype standard deviations

[b] a is the additive effect at the QTL, d is the dominance effect, σ_e is the residual standard deviation, σ_u is the between male-parent standard deviation and q is the female-parent allele frequency (see text for more detailed descriptions)

[c] ME = modal estimation, RE = regressive and WS = within-sire method

The estimate of the female-parent allele frequency is unbiased with the three methods but not very precise. In general, the empirical standard deviation of the estimate decreases as the simulated effect of the QTL increases. The standard deviations from ME are the lowest and improve further when markers are incorporated in the analysis.

The comparison of the first three methods with WO are given in Table 4. The substitution effect, α, and the frequency of heterozygous male parents are considered As expected, given the results above, when marker information is not available, the substitution effect is overestimated for the smaller effect QTL with the first three methods. For the largest QTL WO gives a more biased estimate than the other methods, especially when more male parents are included in the analysis (data not shown). With marker information the performance of WO improves, giving unbiased estimates of the substitution effect for all QTLs. The estimates are also more precise than with the other approximations. The male-parent heterozygote frequency is not well estimated except when the effect of the QTL is large. The inclusion of marker information reduces the bias. Using WO the estimate increases as the QTL effect decreases when marker information is not available, and decreases with QTL effect when a marker is used. This is in contrast with the other approaches which tend to give increasing underestimates with decline in QTL effect both with and without markers.

Discussion

Four approximations to the likelihood for QTL detection in a progeny test design have been presented. The first three give a more precise description of the situation, with eight parameters describing the within-genotype means and variances and the QTL genotype frequency in the male parents and the allele frequency in the female parents. On the other hand, the fourth approximation (WO), using additional assumptions concerning the distribution, involves only three parameters, mixing the distribution components described in the others.

In this paper, we focused on maximum likelihood partly because it is the more widely used technique in classical segregation analysis and in order to measure directly the usefulness of marker information for QTL

Table 4 Mean parameter estimates for the all four approximations with the empirical standard deviation over the replicates in parentheses. (20 male parent families)

Type of gene[a] (μ_1, μ_2, μ_3)	Parameter[b] (true value)	Test statistics			
		ME[c]	RE	WS	WO
		Without marker			
0 0 2	α(1)	0.99 (0.21)	1.02 (0.22)	1.02 (0.20)	1.74 (0.12)
	p_2(0.5)	0.51 (0.12)	0.48 (0.14)	0.47 (0.14)	0.28 (0.10)
0 1 2	α(1)	0.88 (0.24)	0.80 (0.70)	0.98 (0.36)	1.16 (0.26)
	p_2(0.5)	0.34 (0.24)	0.36 (0.21)	0.46 (0.20)	0.33 (0.33)
0 0 1	α(0.5)	0.70 (0.26)	0.76 (0.58)	0.60 (0.46)	0.96 (0.32)
	p_2(0.5)	0.42 (0.28)	0.37 (0.28)	0.34 (0.30)	0.37 (0.31)
0 0.5 1	α(0.5)	0.62 (0.24)	0.64 (0.80)	0.68 (0.68)	0.82 (0.32)
	p_2(0.5)	0.30 (0.32)	0.27 (0.28)	0.29 (0.28)	0.41 (0.41)
0 0 0.5	α(0.25)	0.62 (0.24)	0.50 (1.26)	0.52 (0.68)	0.72 (0.38)
	p_2(0.5)	0.31 (0.33)	0.27 (0.33)	0.32 (0.34)	0.41 (0.39)
0 0.25 0.5	α(0.25)	0.64 (0.36)	0.62 (0.76)	0.68 (0.86)	0.64 (0.36)
	p_2(0.5)	0.33 (0.35)	0.27 (0.30)	0.23 (0.30)	0.53 (0.41)
		With marker			
0 0 2	α(1)	1.02 (0.11)	0.98 (0.14)	0.98 (0.14)	0.98 (0.10)
	p_2(0.5)	0.51 (0.11)	0.49 (0.11)	0.49 (0.11)	0.48 (0.12)
0 1 2	α(1)	0.90 (0.68)	0.96 (0.60)	1.02 (0.38)	0.98 (0.08)
	p_2(0.5)	0.49 (0.20)	0.49 (0.11)	0.51 (0.11)	0.50 (0.11)
0 0 1	α(0.5)	0.52 (0.18)	0.54 (0.38)	0.44 (0.34)	0.48 (0.12)
	p_2(0.5)	0.51 (0.17)	0.51 (0.16)	0.51 (0.17)	0.44 (0.23)
0 0.5 1	α(0.5)	0.50 (0.16)	0.52 (0.76)	0.58 (0.64)	0.48 (0.12)
	p_2(0.5)	0.54 (0.18)	0.52 (0.19)	0.47 (0.18)	0.44 (0.21)
0 0 0.5	α(0.25)	0.42 (0.18)	0.54 (0.76)	0.52 (0.60)	0.26 (0.14)
	p_2(0.5)	0.34 (0.29)	0.38 (0.34)	0.35 (0.32)	0.31 (0.37)
0 0.25 0.5	α(0.25)	0.40 (0.16)	0.52 (0.80)	0.42 (0.52)	0.26 (0.14)
	p_2(0.5)	0.34 (0.31)	0.29 (0.28)	0.36 (0.33)	0.39 (0.37)

[a] The simulated effect of the QTL genotypes in within-QTL-genotye standard deviations
[b] α is the substitution effect at the QTL, p_2 the male-parent heterozygote frequency (see text for more detailed descriptions)
[c] ME = modal estimation, RE = regressive, WS = within-sire ad WO = only-within-family deviation method

detection. However, other approaches are possible including the least-squares method (Haley and Knott 1992) and non-parametric tests (Lander and Kruglyak 1995).

Due to the very large amount of computation required, this study has been based on a limited number of replications (200 under each H_1), thus limiting the precision of the comparisons. Nevertheless, consistent patterns have been observed, which enables general conclusions to be drawn. In particular, ME and WO methods appear to be more powerful in linkage analysis, even though all the methods yield reasonable parameter estimations. In the without-marker situation, the comparison of power and parameter estimation seemed to reveal a constant pattern, with the RE method generally performing better in terms of power.

Thus, despite the additional simplification, the power of the fourth approach is comparable with the more-complicated approximations. The parameter estimates give a less-complete description of the QTL and its effect, but when a marker is available the estimate for the substitution effect is unbiased and precise; the frequency of heterozygous male parents, however, is not well estimated. The reduction in the number of parameters enables much faster computation and hence, this fourth approach should be used as a first a step, similar to other simple test statistics which have been proposed for preliminary data exploration (see Le Roy and Elsen 1992 for a review) or for a rapid screening of the genome when markers are available. Subsequent analysis could then be performed in the area of interest, thus allowing a more complete description of the QTL.

From a comparison of the other approximations it is not clear that any one method is always consistently preferable to the others. When marker information is not available, RE seems to perform better in terms of power. Considering parameter estimates, however, none is consistently better than the others. With marker information ME gives higher power than the other two approximations and parameter estimates are reasonable.

These approaches were compared here in extreme situations where either no marker or a unique totally linked marker were available. In practice, recorded individuals may have sets of genetic markers put on their genome, and the amount of marker information will vary from individual to individual and from place to place on the chromosomes. In these circumstances, information may be fully exploited by multilocus approaches as described for instance by Haley et al. (1994). In a companion paper (Knott et al. 1996), we propose an extension of this approach to the case of half-sib families in outbred populations, using either the least squares method or our "only within-family deviation method".

Robustness was not evaluated in this paper. All four approximations compared were based on an identical parametric model assuming normality of within-QTL genotype performances and are probably similar in their behaviour to non-respect of this hypothesis.

Any complicating factors such as fixed effects have been omitted. Large half-sib families are required and, hence, data will be obtained from existing populations that will be distributed over a number of plots or years, for example. All of the likelihoods presented can be extended to take account of these effects. Nonetheless, there may be difficulties due to the increased computation required in order to calculate the likelihood and estimate the parameters of interest.

It has been implicitly assumed that possible familial relations between male and female parents should not change the classification between methods. A more complete treatment of QTL detection in a progeny test design should consider these genetic relationships. While much more complicated in their algebraic and computational developments, such methods are already available using peeling (Hasstedt 1982) or Gibbs sampling (Guo and Thompson 1994). The extra power and robustness given by a correct consideration of these relations are still to be evaluated in the large-sized progeny test design often used in plant and animals.

Our H_0 hypothesis considered a polygenic inheritance described by an additive genetic effect with a 0.2 heritability. Thus, implicitly, H_0 dealt with unlinked QTLs. In theory, it might be possible to extend models and associated likelihoods to situations with a major unlinked QTL. More generally, an oligogenic inheritance with a finite number of QTLs could be considered. However, computation would become tedious, if not impossible, in practice. Moreover, there is no obvious reason why extra differences between the compared methods should be found in these situations.

A major difficulty when using the maximum-likelihood method is the behaviour of the test statistic when $H0$ holds, as the distribution of the test cannot be predicted at present. In this paper, the rejection thresholds were obtained by simulation, with 2000 replicates. Real situations will be much more complicated, in particular in multilocus approaches. Simulations can be used or a permutation test applied (Churchill and Doerge 1994) if the test statistics can be rapidly computed. However, since the thresholds vary between traits, permutation tests have to be performed for all traits successively. Approximations were recently proposed for the computation of thresholds (Rebai et al. 1994) for simple situations found in plant breeding; an effort should be made to extend these to outbred population structures.

Acknowledgements This work was partly supported by a Human Capital and Mobility project of the European commission (CT940508). We are very grateful for the help from C. Plomion and referees reading the manuscript.

References

Barendse W, Armitage SM, Kossarek LM, Shalom A, Kirkpatrick BW, Ryan AM, Clayton D, Li L, Neibergs HL, Zhang N, Grosse WM, Weiss J, Creighton P, McCarthy F, Ron M, Teale AJ, Fries R, McGraw RA, Moore SS, Georges M, Soller M, Womack JE, Hetzel DJS (1994) A genetic linkage map of the bovine genome. Nature Genet 6:227–235

Bishop MD, Kappes SM, Keele JW, Stone RT, Sunden SLF, Kawkins GA, Toldo SS, Fries R, Grosz MD, Yoo J, Beatttie CW (1994) A genetic linkage map in cattle. Genetics 136:619–639

Boichard D, Elsen JM, Le Roy P, Bonaiti B (1990) Segregation analysis of fat content data in Holstein × European Freisian crossbred cattle. In: Hill WG, Thompson R, Woolliams JA (eds). 4th World Congress on Genetics Applied to Livestock Production, Edinburgh, 23–27 July 1990, vol IV, pp 167–170

Bonney GE (1984) On the statistical determination of major gene mechanisms in continuous human traits: regressive models. Am J Med Genet 18:731–749

Bonney GE (1986) Regressive logistic models for familial disease and other binary traits. Biometrics 42:611–625

Bovenhuis H, Weller JL (1994) Mapping and analysis of dairy cattle quantitative trait loci by maximum-likelihood methodology using milk protein genes as genetic markers. Genetics 137:267–280

Bradshaw HD, Stettler RF (1995) Molecular genetics of growth and development of *Populus* IV. Mapping QTLs with large effects on growth, form and phenology traits in a forest tree. Genetics 139:963–973

Bradshaw HD, Villar M, Watson BD, Otto KG, Stewart S, Stettler RF (1994) Molecular genetics of growth and development in *Populus*. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS, and RAPD markers. Theor Appl Genet 89:167–178

Byrne M, Murrell JC, Allen B, Moran GF (1995) An integrated genetic linkage map for eucalypts using RFLP, RAPD and isozyme markers. Theor Appl Genet 91:869–875

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Crawford AM, Dodds KG, Ede AJ, Pierson CA, Montgomery GW, Garmonsway HG, Beattie AE, Davies K, Maddox JF, Kappes SW, Stone RT, Nguyen TC, Penty JM, Lord EA, Broom JE, Buitkamp J, Schwaiger W, Epplen JT, Matthew P, Matthews ME, Hulme DJ, Beh KJ, McGraw RA, Beattie CW (1994) An autosomal linkage map of the sheep genome. Genetics 140:703–724

Demenais FM, Bonney GE (1989) Equivalence of the mixed and regressive models for genetic analysis. I. Continuous traits. Genet Epidemiol 6:597–617

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

Devey ME, Fiddler TA, Liu BH, Knapp SJ, Neale BD (1994) An RFLP linkage map for loblolly pine based on a three-generation outbred pedigree. Theor Appl Genet 88:273–278

Elsen JM, Le Roy P (1990) Detection of major genes and determination of genotypes. Application to discrete variables. In: Hill WG, Thompson R, Wolliams JA (eds). 4th World Congress on Genetics Applied to Livestock Production, Edinburgh 23–27 July 1990, vol XV, pp 37–49

Elston RC, Stewart J (1971) A general model for the genetic analysis of pedigree data. Hum Hered 21:523–542

Falconer DS (1989) Introduction to quantitative genetics, 3rd edn. Longman, UK

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

Grattapaglia D, Sederoff R (1994) Genetic linkage maps for *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. Genetics 137:1121–1137

Guo SW, Thompson EA (1994) Monte Carlo estimation of mixed models for large complex pedigrees. Biometrics 50:417–432

Haley CS, Knott SA (1992) A simple regression model for interval mapping in line crosses. Heredity 69:315–324

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

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

Hasstedt SJ (1982) A mixed-model likelihood approximation on large pedigrees. Comput Biomed Res 15:295–307

Knapp SJ, Bridges WC, Birkes D (1990) Mapping quantitative trait loci using molecular marker linkage maps. Theor Appl Genet 79:583–592

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

Knott SA, Haley CS, Thompson R (1991a) Methods of segregation analysis for animal breeding data: a comparison of power. Heredity 68:299–311

Knott SA, Haley CS, Thompson R (1991b) Methods of segregation analysis for animal breeding data: parameter estimates. Heredity 68:313–320

Knott SA, Elsen JM, Haley CS (1996) Methods for multiple marker mapping of quantitative trait loci in half-sib populations. Theor Appl Genet 93:71–80

Lander ES, Kruglyak (1995) Genetic direction of complex traits: guidelines for interpreting and reporting linkage results. Nature Genetics 11:241–247

Le Roy P, Elsen JM (1992) Simple test statistics for major gene detection: a numerical comparison Theor Appl Genet 83:635–644

Le Roy P, Elsen JM (1994) Numerical comparison between powers of maximum likelihood and analysis of variance methods for QTL detection in progeny test designs. The case of monogenic inheritance. Theor Appl Genet 90:65–72

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

Le Roy P, Naveau J, Elsen JM, Sellier P (1990) Evidence for a new major gene influencing meat quality in pigs. Genet Res 55:33–40

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 an oligogenic model. Comput Biomed Res 17:471–480

Morton NE, McLean CJ (1974) Analysis of family resemblance. III. Complex segregation of quantitative traits. Am J Hum Genet 26:489–503

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

Numerical Algorithms Group (1990) The NAG Fortran Library Manual – Mark 14. NAG Ltd., Oxford

Ollivier L (1980) Le déterminisme génétique de l'hypertropie musculaire chez le Proc. Ann Génét Sél Anim 12, 383–394

O'Malley DM, McKeand SE (1994) Marker-assisted selection for breeding value in forest trees. For Genet 1:207–218

Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE, Lander ES, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations and environments. Genetics 127:181–197

Piper LR, Bindon BM (1982) The Booroola Merino and the performance of medium non-peppin crosses at Armidale. In: Piper LR, Bindon BM, Nethery RD (eds). The Booroola Merino, CSIRO, Melbourne, Australia, pp 9–20

Plomion C, Bahrman N, Durel CE, O'Malley DM (1995) Genomic mapping in *Pinus pinaster* (Maritime pine) using RAPD and protein markers. Heredity 74 : 661–668

Plomion C, Durel C-E (1996) Estimation of the average effects of specific alleles detected by the pseudo-testcross QTL-mapping strategy. Gen Sel Evol 28 (3) : 217–320

Rebai A, Goffinet B, Mangin B (1994) Approximate thresholds of interval mapping tests for QTL detection. Genetics 138 : 235–240

Ricordeau G (1984) Major genes in sheep and goats. In: 2nd World Congress on Genetics Applied to Livestock Production, vol VI, Madrid, Spain, pp 454–461

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

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

Sax K (1923) The association of size differences with seed coat pattern and pigmentation in *Phaesolus vulgarus*. Genetics 8 : 552–560

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

Weller JL, 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