Original article

Further insights of the variance component method for detecting QTL in livestock and aquacultural species: relaxing the assumption of additive effects

Victor MARTINEZ*

Faculty of Veterinary Sciences, Universidad de Chile, Avda Santa Rosa 11735, Santiago, Chile

Abstract – Complex traits may show some degree of dominance at the gene level that may influence the statistical power of simple models, *i.e.* assuming only additive effects to detect quantitative trait loci (QTL) using the variance component method. Little has been published on this topic even in species where relatively large family sizes can be obtained, such as poultry, pigs, and aquacultural species. This is important, when the idea is to select regions likely to be harbouring dominant QTL or in marker assisted selection. In this work, we investigated the empirical power and accuracy to both detect and localise dominant QTL with or without incorporating dominance effects explicitly in the model of analysis. For this purpose, populations with variable family sizes and constant population size and different values for dominance variance were simulated. The results show that when using only additive effects there was little loss in power to detect QTL and estimates of position, using or not using dominance, were empirically unbiased. Further, there was little gain in accuracy of positioning the QTL with most scenarios except when simulating an overdominant QTL.

QTL / additive effect / dominance / power / REML

1. INTRODUCTION

Quantitative trait loci (QTL) detection using mixed linear models is one of the preferred methods for estimating the contribution of a particular chromosomal segment to the observed variance in general pedigrees from outbred populations [2,19]. This method infers QTL segregation using as a covariance structure the number of alleles identical by descent (IBD) conditional on genetic markers in many positions of the genome [19,29].

It is customary that, when using crosses between outbred populations (the F_2 design), additive and dominance effects are fitted jointly in the regression

---

*Corresponding author: vmartine@uchile.cl

analysis [1]. Using the variance component method, it is usually assumed that only additive effects are of importance and therefore only IBD matrices conditional on marker data are fitted in the restricted maximum likelihood (REML) procedure (see [10,22] for traditional implementations in outbred pedigrees of pigs and sheep). Although this is indeed correct under the assumption of no dominance, it is not clear under the variance component framework what is the most powerful test of linkage and by what extent variance components can be biased if dominance is not accounted for in the model of analysis.

In light of recent results in cattle [17], where significant dominance effects have been estimated in the DGAT1 locus, this may be of importance, for example when the interest is to select genomic regions showing evidence of QTL at particular chromosomes, when predicting breeding values due to the QTL in order to select candidates in marker assisted selection programmes [8,13] or when performing confirmation studies within commercial populations. This may be important in cases where the original experiments from crosses between outbred lines show evidence of non-additive gene action at the QTL [6,7].

Under the assumption of genes with infinitesimal effects, modelling dominance is difficult since it is necessary to maximise the likelihood of the data, fitting extra parameters, such as dominance variance and the covariance between additive and dominance effects under inbreeding [5], and it is likely that the estimates of these variance components are subjected to large sampling correlations [23,24]. Also under the infinitesimal model it is difficult conceptually to deal with inbreeding depression, since it is doubtful that a genetic model of an infinite number of loci exists with directional dominance [5]. Nevertheless, at least in theory, the use of more complex models may help to improve accuracy of estimation, as well as help to exploit non-additive genetic variation within breeds [12]. However, in practice it is not easy to disentangle variation due to common environmental effects and dominance effects, since when using full-sib structures as in poultry or fish breeding both terms are completely confounded. Under mixed inheritance, non-additive genetic variance can be accommodated explicitly by extending the mixed inheritance model of the QTL. The covariance structure of dominance effects is proportional to the probability that two relatives share the same genotype at a locus [9]. Very little has been presented in the literature about this subject, although in practice it is an important issue for detecting QTL in outbred populations [21].

In the present paper, we investigated the behaviour of the mixed linear model when modelling dominance variance at the QTL in species where relatively large family sizes can be obtained, such as in pigs, poultry, and aquacultural species. Since *a priori*, in a given experiment where the actual genetic model is not known, we investigated a two-step approach in which additive effects

and genotypic effects at the QTL are first modelled for QTL detection in order to obtain the most likely position of the QTL. Testing for dominance effects was carried out conditionally at the most likely position of the QTL, previously obtained from both models using the required covariance structure in the mixed linear model. Using these methods, we calculated empirical power and accuracy of estimating variance components, using different livestock population structures likely to be encountered in practice.

2. MATERIALS AND METHODS

The outline of this paper is as follows. First, we present the mixed model including dominance at the QTL and show how covariances can be computed in full-sib structures; then we performed the testing regimes used first for detecting the presence of a segregating QTL (with or without using information of dominance) and then we made inferences about the mode of gene action at the QTL.

2.1. Genetic model

Let us assume a population of non-inbred full-sib families measured for a normally distributed trait (y_i). The model used to explain the phenotype of individual i is

$$y_i = \mu + a_i + d_i + p_i + e_i, \tag{1}$$

where μ is the contribution of fixed effects (such as the mean); a_i is the additive effect at the putative biallelic QTL of the individual, with values equal to a for QQ, 0 for Qq and qQ, and $-a$ for qq; d_i, is the dominance effect (d) expressed as the difference between the mid-homozygote value [9]; p_i is the polygenic component, explaining unlinked genes in the rest of the genome, and e_i is the residual. In scalar notation, the first and second moments of the vector of phenotypes are then equal to the following:

$$\begin{cases} \mathrm{E}(y_i) = \mu \\ \mathrm{Var}(y_i) = \sigma_a^2 + \sigma_d^2 + \sigma_p^2 + \sigma_e^2 \\ \mathrm{Cov}(y_i, y_j) = \Phi_{i,j}\sigma_a^2 + \Delta_{i,j}\sigma_d^2 + \pi_{i,j}\sigma_p^2 \end{cases}, \tag{2}$$

where $\Phi_{i,j}$ is the IBD proportion of alleles that individuals i and j share at the known QTL, Δ is a variable indicating whether individual i and j share the same genotype at the QTL, and $\pi_{i,j}$ is the additive relationship between individuals i and j. These variables are distributed according to the following distributions (by convention, the first allele is the paternal QTL allele inherited

from the father to individual i (q_i^f) and the second allele is the maternal QTL allele inherited to individual i (q_i^m)):

$$\Phi_{i,j} \begin{cases} 0 & \text{for } q_i^f \neq q_j^f \text{ and } q_i^m \neq q_j^m \\ 0.5 & \text{for } q_i^f = q_j^f \text{ and } q_i^m \neq q_j^m \text{ or } q_i^m = q_j^m \text{ and } q_i^f \neq q_j^f , \\ 1 & \text{for } q_i^f = q_j^f \text{ and } q_i^m = q_j^m \end{cases} \quad (3)$$

$$\Delta_{i,j} \begin{cases} 1 & \text{for } q_i^f q_i^m = q_j^f q_j^m \\ 0 & \text{otherwise} \end{cases} . \quad (4)$$

To set up the mixed model equations at the animal level, we need to obtain the inverse of the covariance structure due to polygenic effects and of additive and dominance effects at the QTL. We first explicitly constructed the actual matrices pertaining to each full-sib group, as detailed in the following section. These calculations were carried out using a computer program designed specifically for this purpose.

2.2. Computing covariance matrices given marker data for full-sib structures

Since the actual genotype of the QTL is not known, we inferred the expected IBD proportion between two full-sibs ($\phi_{i,j}$) using the marginal distribution of IBD proportions (0, 0.5, and 1) at the QTL conditional on marker information [28]. For the purposes of this analysis, completely informative markers with known ordered genotypes were assumed (see the description in [28]). First, consider what is the probability that sib i inherits QTL alleles from the maternal or paternal first homolog or the second homolog conditional on the marker haplotype. There are four possible QTL allelic classes conditional on flanking markers, each depending on the probability of recombination between the markers and the QTL, and between flanking markers in the male and female parent (see [28] for details). With random mating, the conditional probabilities of QTL genotypes can be calculated as the product between the corresponding probabilities of inherited gametes.

The expected IBD proportion ($\phi_{i,j}$) is then the product of the probabilities that both offspring receive, either the same or a different QTL allele from the mother or the father (therefore comparisons are made between paternally and maternally inherited alleles) multiplied each by the corresponding IBD

proportion (0, 0.5, and 1). In matrix notation, the value of ($\phi_{i,j}$) between any pair of full-sibs conditional on marker data is equal to the product of the vector of conditional genotype probabilities given marker data (vector $\mathbf{Q_i}$) and a matrix that represents all the possible alternatives of IBD proportion ($\mathbf{C}$) between both full-sibs considered (Equation (5)):

$$\phi_{i,j} = \mathbf{Q'_i C Q_j}, \tag{5}$$

where $\mathbf{C}$ is a 4×4 symmetric matrix with diagonal elements equal to 1 and off-diagonal elements equal to 0 or 0.5. The notation used in the vector $\mathbf{Q_i}$ stands for conditional probability of QTL genotype, given the ordered genotype at the flanking markers (where 1,i and 2,i represent the flanking markers (1 or 2)) and the allele inherited (say l) from the mother (m) or the father (f) (note that at most there are four different possibilities depending on whether the paternal or maternal allele (for the QTL or marker) inherited by the sib i came from the sire (fm, ff) or the dam (mf, mm) of each parent):

$$\mathbf{C} = \begin{bmatrix} 1.0 & 0.5 & 0.5 & 0.0 \\ 0.5 & 1.0 & 0.0 & 0.5 \\ 0.5 & 0.0 & 1.0 & 0.5 \\ 0.0 & 0.5 & 0.5 & 1.0 \end{bmatrix} \tag{6}$$

and

$$\mathbf{Q_i} = \begin{bmatrix} P\left(q_i^f q_i^m | \frac{M_{1,i}^m M_{2,i}^m}{M_{1,i}^f M_{2,i}^f}\right) & = & P\left(q_i^{ff} q_i^{fm} | \frac{M_{1,i}^m M_{2,i}^m}{M_{1,i}^f M_{2,i}^f}\right) \\ P\left(q_i^f q_i^m | \frac{M_{1,i}^m M_{2,i}^m}{M_{1,i}^f M_{2,i}^f}\right) & = & P\left(q_i^{ff} q_i^{mm} | \frac{M_{1,i}^m M_{2,i}^m}{M_{1,i}^f M_{2,i}^f}\right) \\ P\left(q_i^f q_i^m | \frac{M_{1,i}^m M_{2,i}^m}{M_{1,i}^f M_{2,i}^f}\right) & = & P\left(q_i^{mf} q_i^{fm} | \frac{M_{1,i}^m M_{2,i}^m}{M_{1,i}^f M_{2,i}^f}\right) \\ P\left(q_i^f q_i^m | \frac{M_{1,i}^m M_{2,i}^m}{M_{1,i}^f M_{2,i}^f}\right) & = & P\left(q_i^{mf} q_i^{mm} | \frac{M_{1,i}^m M_{2,i}^m}{M_{1,i}^f M_{2,i}^f}\right) \end{bmatrix}. \tag{7}$$

The value of ($\delta_{i,j}$) between two full-sibs can be obtained similarly, as the probability that both share the same genotype at the QTL [9]. Without using marker data, this value is equal to 1/4 for full-sib individuals. Using marker data, the expectation conditional on marker data can be calculated as the product of the vector $\mathbf{Q_i}$ transposed and the vector $\mathbf{Q_j}$:

$$\delta_{i,j} = \mathbf{Q'_i Q_j}. \tag{8}$$

2.3. Hypothesis testing of nested models

2.3.1. First stage: detecting QTL

The first question to be addressed when mapping a QTL is to test whether there is evidence of segregation of QTL in the linkage group under analysis. This is irrespective of whether there is an additive or a dominant QTL segregating. According to Table I, it is possible to compute two different tests for detecting QTL:

1. **ADDITIVE**: This test is calculated along the linkage group under analysis as minus twice the difference between the log-likelihood of a reduced model (only fitting the polygenic effects) and the log-likelihood of a model fitting additive effects at the QTL in addition to polygenic effects (Tab. I; $-2(L_I - L_{II})$). This is done adjusting the covariance structure due to additive effects at the QTL at every centiMorgan of the linkage group. This test assumes that an additive genetic model is the true underlying mode of gene action of the QTL.
2. **GENOTYPIC**: This test is calculated along the linkage group under analysis as minus twice the difference between the log-likelihood of a reduced model (only fitting the polygenic effects) and the log-likelihood of a model fitting additive and dominance effects at the QTL in addition to polygenic effects (Tab. I; $-2(L_I - L_{III})$). This is done by simultaneously adjusting the covariance structure due to additive and dominance effects at the QTL at every centiMorgan of the linkage group. This test assumes that an additive plus dominance genetic model is the true underlying mode of gene action of the QTL.

In both cases, the test statistic is computed along the linkage group and the highest value of the likelihood ratio (LR) test provides the most likely position of the QTL. Note that the location may differ between tests that were used to detect the QTL (**ADDITIVE** and **GENOTYPIC**) but on average both tests should give very similar locations if they are unbiased (see Sect. 4 below).

2.3.2. Second stage: inferences about the mode of gene action at the QTL

It is important to test the actual mode of gene action given the data independently of whether an additive or dominant model is assumed for detecting a QTL. Testing significance of dominance variance can be accomplished within the framework of the **ADDITIVE** test by fitting dominance effects at the most likely position as obtained from this test (say at position k in the linkage group). At this position, we include in the model the dominance effect with a covariance

Table I. Hypothesis testing framework under dominance. In every alternative, the test statistic is computed as twice the difference between the log-likelihood of the null model (L_0) and (L_1).

Test	Full model	Hypothesis tested (L_1)	Null hypothesis (L_0)
Polygenic (P)	$M_{\mathrm{I}} = Y = Xb + Zp + e$	$L_{\mathrm{I}} = l(Y\vert\sigma_e^2, \sigma_p^2)$	
Additive (A)	$M_{\mathrm{II}} = Y = Xb + Zp + Za + e$	$L_{\mathrm{II}} = l(Y\vert\sigma_e^2, \sigma_p^2, \sigma_a^2)$	L_{I}
Genotypic (A + D)	$M_{\mathrm{III}} = Y = Xb + Zp + Za + Zd + e$	$L_{\mathrm{III}} = l(Y\vert\sigma_e^2, \sigma_p^2, \sigma_a^2, \sigma_d^2)$	L_{I}
Dominance (at position k)	$M_{\mathrm{IV}} = Y = Xb_k + Zp_k + Za_k + Zd_k + e_k$	$L_{\mathrm{IV}} = l(Y\vert\sigma_{e,k}^2, \sigma_{p,k}^2, \sigma_{a,k}^2, \sigma_{d,k}^2)$	L_{II}

proportional to $\delta_{i,j}$ in addition to the additive effects. The LR test is equal to minus twice the difference between log-likelihoods of these two models at position k (Tab. I; $-2(L_{II} - L_{IV})$).

Using the **GENOTYPIC** test, testing for dominance variance at the QTL position detected can be carried out, computing the LR as minus twice the difference between the log-likelihood of the model in which no dominance effects were fitted (only additive effects of the QTL) and that of the model fitting dominance and additive effects, simultaneously (Tab. I; $-2(L_{II} - L_{III})$).

2.4. Simulations

We investigated different alternatives using the simulation of two generation outbred pedigrees structured as independent full-sib families with variable sizes (f_n; including parents and progeny). The scenarios simulated used typical values of f_n equal to 10, 20, 50, or 100, as observed in many livestock and aquacultural species. The total population size was constant ($n = 500$), thus the number of full-sib families is then equal to $n/(f_n)$. The analysis comprised a single linkage group of 50 cM with fully informative markers every 10 cM (six in total), with a QTL placed at position 25 cM. Phenotypes were simulated for parents and progeny with a broad sense heritability (including QTL and polygenic variance) equal to 0.4 and a constant additive genetic variance $\left(\sigma_a^2 = 150\right)$ due to a biallelic QTL. Allele frequencies at the QTL in the base population in Hardy-Weinberg equilibrium were equal to 0.5 and alleles of markers and QTL were uncorrelated (*i.e.* no linkage disequilibrium (LD) was assumed in the base population). Dominance variance due to the QTL was simulated using different ratios of the dominance (d) and additive (a) effects (Tab. II). For each case the residual genetic variance, due to variation under infinitesimal model assumptions (polygenic effects), was varied, such that it explained the remainder of the total genetic variance. The environmental variance was kept constant in all scenarios $\left(\sigma_e^2 = 600\right)$.

We calculated the significance thresholds for the different tests performed for detecting QTL (**ADDITIVE** and **GENOTYPIC**) for the null hypothesis of no segregating QTL (scenario 0; Tab. II). In this case, all the genetic variance was assumed to be due to polygenic effects, such that the heritability of the trait was equal to 0.40.

The unknown significance thresholds required for testing dominance variance were obtained under two different scenarios. In the first case, we simulated an additive segregating QTL that explained 25% of the total variance, giving a high probability of the design for detecting an additive QTL. The second case did not assume that a QTL is segregating; *i.e.* only polygenic effects were simulated.

Table II. Parametric settings for the different scenarios used in the simulations, where *a* is the additive effect at the QTL, *d* is the dominance value at the QTL, *p* are the polygenic effects, and *e* are the environmental effects.

SCENARIO	Genetic value of the QTL		d/a	σ_a^2	σ_d^2	σ_p^2	σ_e^2
	a	*d*					
0	17.32	0.00	0.00	0.00	0.00	400.00	600.00
I	17.32	0.00	0.00	150.00	0.00	250.00	600.00
II	17.32	12.25	0.71	150.00	37.50	212.50	600.00
III	17.32	17.32	1.00	150.00	75.00	175.00	600.00
IV	17.32	24.49	1.41	150.00	150.00	100.00	600.00

The heritability (including QTL and polygenic variance) was equal to 0.4. This second alternative reflects well a situation, where there is no prior knowledge of whether a QTL is segregating in the population. This may be of importance when empirical significance thresholds can be obtained while analysing real populations, by permuting genotypes and phenotypes within families due to the large family sizes especially in aquaculture [4]. In all these cases, 1000 replicates were simulated under the null hypothesis and each replicate comprised 50 full-sib families of size 10 (eight sibs and two parents), giving a population size equal to 500 individuals.

The variance components were estimated using REML with ASREML [11], using the defined matrices as explained in Section 2.2.

3. RESULTS

3.1. Distribution of the test statistics under the null hypotheses

#### 3.1.1. Test statistics used to detect QTL

The empirical distribution under the null hypothesis of no QTL for the different tests implemented is presented in Figure 1. Empirically, the **ADDITIVE** test is distributed as a χ^2 distribution of between 1 and 2 degrees of freedom (DF); this is seen here irrespective of the family sizes evaluated in this paper. From this result, it is clear that there is very good agreement between the empirical distribution obtained here and the distribution of the LR under the null hypothesis, as obtained previously by others [27].

For the **GENOTYPIC** test (including additive and dominance effects at the QTL in the model), there is no previous empirical evidence in the literature
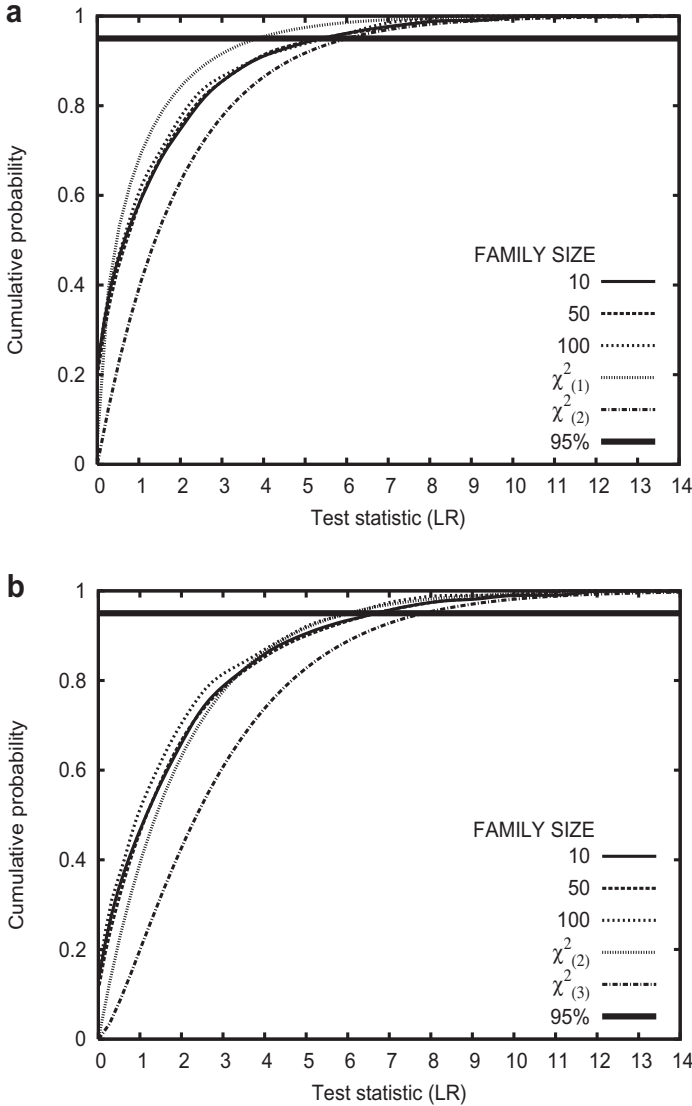
Figure 1. Distribution of the test statistic under the null hypothesis of no QTL. (a) Distribution of the test statistic under the null hypothesis of no QTL, obtained using the **ADDITIVE** test. (b) Distribution of the test statistic under the null hypothesis of no QTL, obtained using the **GENOTYPIC** test. The total number of individuals is equal to 500. The horizontal line is the 95% of the cumulative distribution.

regarding the distribution of this test under the null hypothesis. However, it would be expected to be distributed between $1/2\chi^2_{(2)}$ and $1/2\chi^2_{(3)}$ distributions, since, here, testing σ^2_a and σ^2_d is carried out in many positions along the linkage
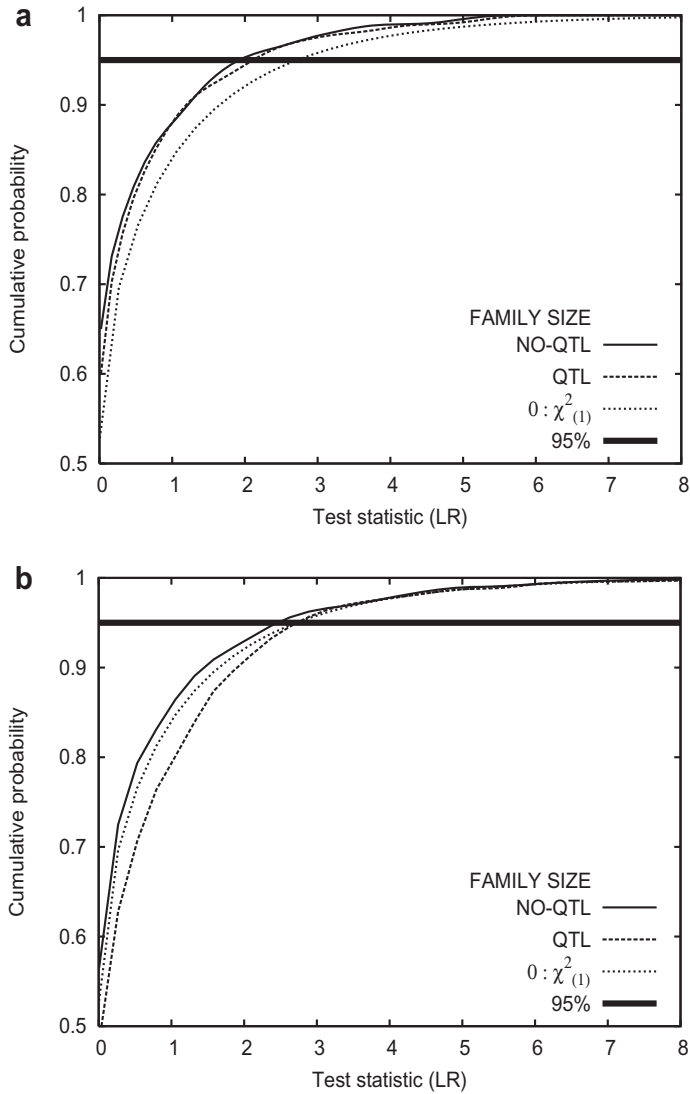
Figure 2. Distribution of the test statistic used for detecting dominance variance under the null hypothesis of no dominance. (a) Obtained at the best location from the **ADDITIVE** test. (b) Obtained as the best location from the **GENOTYPIC** test. The design was equal to 50 families of eight sibs plus the two parents each ($n = 500$). The horizontal line is the 95% of the cumulative distribution.

group [29]. The simulation results show that the significance thresholds tend to be more conservative, with a distribution very similar to a $\chi^2_{(2)}$ distribution (Fig. 1). Again the family size has little effect on the significance thresholds.

Table III. Power to detect QTL (first stage) using the **ADDITIVE** and **GENOTYPIC** model. In each case, 100 replicates were simulated according to the different scenarios simulated.

Model for QTL detection	Family size	SCENARIO			
		I	II	III	IV
ADDITIVE	10	0.42	0.59	0.80	0.89
	20	0.80	0.92	0.97	0.98
	50	0.96	0.93	0.96	0.98
	100	0.98	0.93	0.94	0.98
GENOTYPIC	10	0.40	0.56	0.83	0.91
	20	0.79	0.92	0.98	1.00
	50	0.94	0.95	0.98	0.99
	100	0.97	0.93	0.94	0.98

The distribution of the test statistic used to detect dominance variance conditional on the best location obtained when using the **ADDITIVE** test is unknown. As explained in Section 2.4 we performed two scenarios, with or without simulating additive effects at the QTL, to obtain significance thresholds under the null hypothesis of no dominance. As is seen in Figure 2, there is very little difference between the significance thresholds and the shape of the distributions of this test obtained with these two alternative scenarios. Also, note that the significance threshold tends to be more conservative than the one expected at a single position from a mixture distribution of 0 with a probability of 1/2 and a one degree of freedom ($\chi^2_{(1)}$) with a probability of 1/2 (for $\alpha = 0.05$; 2.71 [2,3]).

When using the **GENOTYPIC** test, the test of dominance variance was constructed as twice the difference between the log-likelihood of the **GENOTYPIC** model and the log-likelihood of the model in which only additive effects were fitted. This gave an empirical distribution very similar to a distribution from a mixture distribution of 0 with a probability of 1/2 and a one degree of freedom ($\chi^2_{(1)}$) with a probability of 1/2 (with a significance threshold similar to 2.71 for $\alpha = 0.05$; Fig. 2).

3.2. First stage: empirical power of QTL detection

The power estimates ($\alpha = 0.05$) over 100 replicates are presented in Table III. In this case, power was computed as the proportions of the 100 replicates with a LR greater than the average of the empirical significance thresholds for all the family sizes considered in the previous section. This was done due to the fact

that very little difference between the significance thresholds were obtained for the different family sizes simulated.

3.2.1. Power of the additive model

The **ADDITIVE** test was used to detect the QTL, irrespective of whether the true model involved dominance or not. The power calculations are presented in Table III for the different family sizes simulated. We found two main trends in terms of power to detect the QTL in these scenarios; power increased asymptotically when increasing the family size and secondly, power increased when a dominant QTL was simulated. This suggests that most of the information about linkage between markers and QTL can be captured in the random model through fitting only the additive relationships conditional on marker data.

3.2.2. Power of the genotypic model

The power of the **GENOTYPIC** test tends to be very similar to the test that only includes additive effects in the model. In fact, absolute differences in power to detect the QTL were only at the most extreme case about 3% and this holds for all degrees of dominance simulated in the present examples (such as when including overdominance, see Tab. III). Although not shown, there was a slight increase in the mean test statistic of this test when compared to the additive test alone (about 10%), however, this increase in test statistic was counterbalanced by the fact that this test had a higher significance threshold compared to the one obtained for the **ADDITIVE** test (see Fig. 1). Furthermore, the correlation between both the **ADDITIVE** and **GENOTYPIC** test statistics was very high and consistent with the ratio between them (ranged from 0.90 to 0.97).

3.3. Second stage: power of detecting dominance conditional on location

Testing dominance at the best position as obtained from the **ADDITIVE** model assumes that the most likely position obtained from the additive test is unbiased (which happens to be the case (see Sect. 3.4.1)). This test has the advantage in that there is no need to fit the covariance structure due to dominance at every position tested. For this reason, it requires only once to construct the dominance relationship matrix conditional on marker information. Therefore, this strategy should be computationally more efficient. The first trend observed was that power to detect dominance variance is relatively low even in cases where overdominance was simulated (Tab. IV). In these scenarios, dominance variance accounted for as much as 15% of the total variance (Tab. II). Secondly,

Table IV. Power of detecting dominance in the second stage, according to the conditional model (first estimating the most likely position of the QTL using the **ADDITIVE** and **GENOTYPIC** model and then fitting dominance effects at the most likely position detected). In each case, 100 replicates were simulated according to the different scenarios simulated.

Model for QTL detection	Family size	SCENARIO		
		II	III	IV
ADDITIVE	10	0.12	0.30	0.53
	20	0.17	0.35	0.81
	50	0.30	0.48	0.82
	100	0.29	0.51	0.77
GENOTYPIC	10	0.17	0.37	0.52
	20	0.15	0.36	0.77
	50	0.26	0.49	0.79
	100	0.27	0.54	0.78

the optimum family size in terms of power was around 50 individuals per family, which is very similar to the optimum observed for the tests of linkage (Tab. IV). In this case, power almost reached an asymptotic value when increasing the family size in excess of 50 individuals per family (Tab. IV). Finally, similar power was obtained whether or not the QTL was detected including or not including dominance in the first stage on QTL detection (Tab. IV).

3.4. Estimates of QTL parameters

3.4.1. Position

Estimates of position obtained from both methods (**ADDITIVE** and **GENO-TYPIC**) are presented in Table V. In general, estimates of position obtained from both methods were empirically unbiased. We note, however, that correlations between position estimates were only moderate-high (ranging from 0.70 to 0.80) for complete dominance and smaller for overdominance (0.50–0.60). This is because in some replicates, the maximum LR was obtained at opposite positions in the chromosome.

3.4.2. Variance components and bias due to model misspecification

Using the model that only incorporates the additive component gave biased estimates of variance components, if the true model incorporates dominance.

Table V. Position estimates according to the different testing procedures performed to detect the QTL at the first stage.

Family size	SCENARIO							
	I		II		III		IV	
	ADDITIVE	GENOTYPIC	ADDITIVE	GENOTYPIC	ADDITIVE	GENOTYPIC	ADDITIVE	GENOTYPIC
10	25.8 (10.5)	26.6 (10.9)	25.6 (10.8)	24.6 (10.7)	25.8 (6.7)	24.1 (6.5)	25.5 (6.6)	24.7 (5.6)
20	25.0 (7.0)	24.9 (6.7)	25.4 (6.3)	25.0 (5.7)	25.2 (5.6)	24.9 (5.8)	25.0 (3.9)	24.9 (3.0)
50	24.5 (5.9)	24.1 (6.0)	25.5 (5.0)	25.8 (4.8)	25.4 (3.9)	25.8 (4.5)	25.2 (3.5)	24.9 (2.3)
100	25.1 (4.7)	25.1 (4.8)	24.6 (6.0)	25.3 (5.6)	25.1 (6.5)	25.1 (5.3)	24.8 (4.2)	25.2 (2.7)
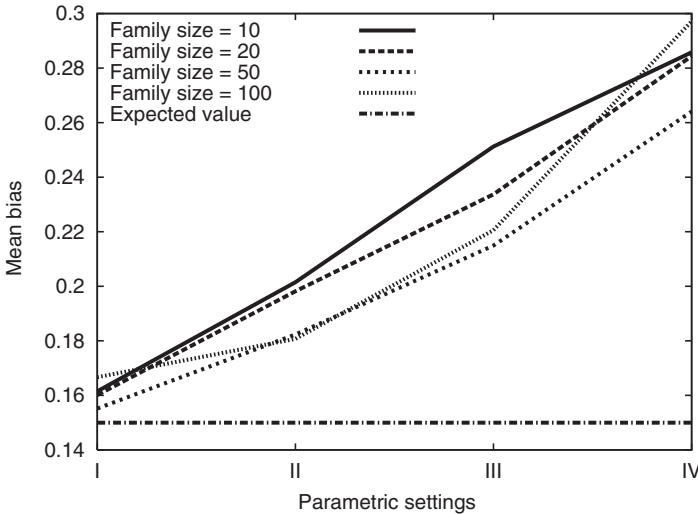
Figure 3. Empirical mean bias due to model misspecification of the heritability due to the QTL according to different levels of dominance and family size simulated. Roman numerals I, II, III, and IV correspond to σ_d^2 equal to 0, 37.5, 75, and 150, respectively.

This bias, due to model misspecification, was measured using the mean bias (MB) which was calculated as:

$$\text{MB} = \frac{\sum_{i=1}^{100}\left(\hat{\theta}_i - \theta_i\right)}{100}, \tag{9}$$

where θ_i is the simulated value of the heritability due to the QTL and ˆ denotes the estimate from each replicate simulated.

Not including dominance in the model gave upwardly biased estimates of the heritability of the QTL and the actual bias is clearly a function of the magnitude of the dominance variance explained by the QTL (Fig. 3). Due to the fact that the power of the variance component method is related to the magnitude of the variance component considered [14,25], this may explain why, under dominance, the power of the simple additive test is very similar to the test that includes dominance in the model. When explicitly including dominance effects in the model, estimates of the additive and dominance variance of the QTL tend to show only a slight (up/downward) bias, considering the complexity of the model fitted. Furthermore, standard deviations of variance components over replicates were relatively similar for the models fitting or not fitting dominance when detecting QTL in the first stage (results not shown).

4. DISCUSSION

In this paper, different models for detecting QTL under the variance component method were studied. It is customary to fit only additive effects to search for QTL, irrespectively, of whether the true underlying model involves dominance. In this investigation we obtained similar power for detecting the QTL using or not explicitly using dominance effects in the model of analysis when there is a non-additive genetic action influencing the trait under consideration. The most likely explanation of this finding is related to the amount of information required for estimating different parameters at the QTL. Recent results have shown the difficulties when trying to estimate additive and dominance effects using information from dense marker data using reasonable sample sizes [23]. This result may be explained largely by the high correlation between the IBD probabilities ($\phi_{i,j}$) and genotype IBD ($\delta_{i,j}$) between individuals. Thus, separating additive *versus* dominance effects within families can be difficult when estimating highly correlated parameters using the variance component method [20,23,24].

Position estimates may differ according to the models fitted to the data, but the results obtained here suggest that shifts in position are not due to a better fit when including dominance, *i.e.* unbiased estimates of position are obtained with or without including dominance. This has also been found in models involving dominance and inbreeding in which relatively large shifts in position have been observed when including dominance in the analysis, even when this effect only borders significance [21]. To investigate this finding further, we calculated the correlation between the difference of position estimates from both models (including or not including dominance when detecting the QTL) and we used the LR to test dominance at the most likely position from the **GENO-TYPIC** model, in order to see whether changes in position are associated with a better fit of this model. A significant association should be obtained if shifts in position estimates between both models are due to the incorporation of dominance. An example of such a trend is presented in Figure 4. This figure clearly shows that there is very little evidence suggesting that changes in position are due to the better fit given by using dominance in the model. Note that the example gave maximal power to detect the QTL, but the same trend was observed when the power to detect the QTL was small (data not shown). In spite of this finding, some increase in accuracy can still be obtained when including dominance effects in the model of analysis, but this was only evident when there was overdominance (see Tab. V).

So far, we completely assumed informative markers and linkage phase known. These assumptions made it possible to use a simple and very fast deterministic
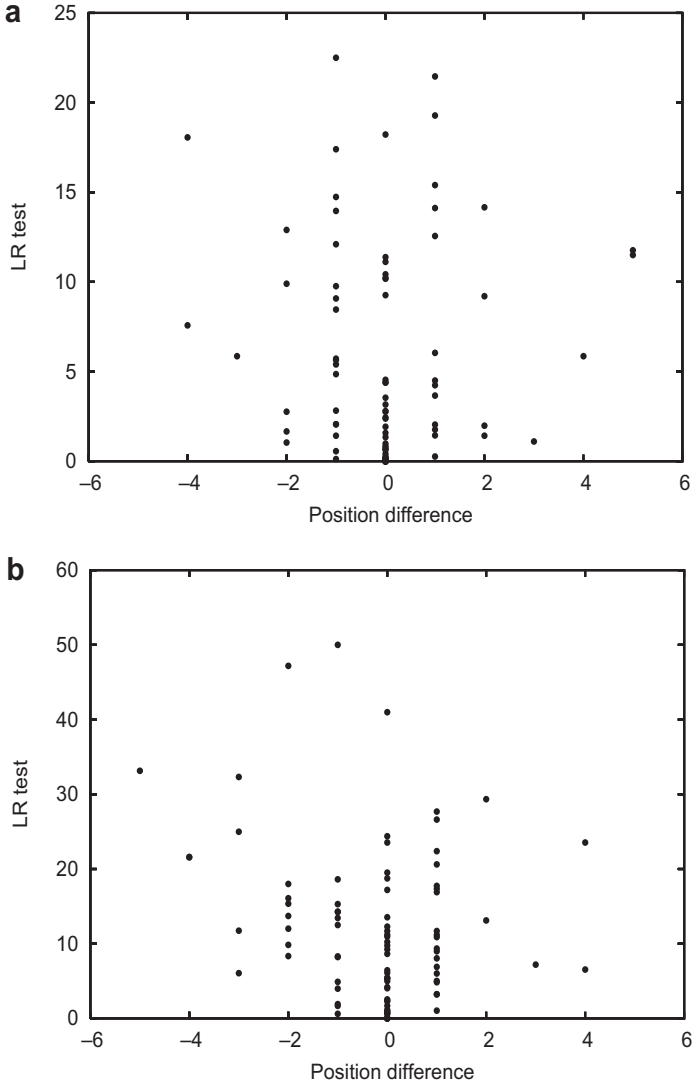
Figure 4. Relationship between the difference between position estimates obtained with or without including dominance to detect the QTL and the LR test for detecting dominance (using the **GENOTYPIC** test). The examples are given for complete dominance (a) and overdominance at the QTL (b). The family size was equal to 50 individuals.

procedure to obtain the matrices required for modelling the covariance structure at the QTL for additive and dominance effects, given the large family sizes simulated. This enabled us to obtain the distributions under the different null

hypothesis assumed. It would be expected that fully informative markers will not produce bias, but give the highest possible values of power calculations given the scenarios simulated in this paper. In practice, it would be expected that when markers are not fully informative, information from multiple markers along the chromosome can be used for inferring IBD probabilities using hidden Markov models and this improves the effects of reduced marker heterozygosity.

In a literature review [16] regarding QTL analysis in plants, it is shown that more than 50% of all the QTL surveyed gave evidence of a dominance and over-dominance mode of gene action as the most plausible models for the dominant QTL. These results, however, are likely due to the significance threshold imposed to declare that a QTL is real, which will cause very large bias in the dominance effects [16]. So the interpretation of a significant dominance, either a complete or an overdominant model, should be considered with caution. In practice, this can be particularly extreme in outbred populations where the family size is relatively low, markers may not be fully informative and the proportion of the parental crosses that is expected to be informative is small.

Significant quantities of genomic data are now available, with the potential for enhancing accuracy of position estimates of QTL effects. For instance, [26] reported a genetic variation map of the chicken genome containing 2.8 million single-nucleotide polymorphisms (SNP) and [15] found 2507 putative SNP in the salmon genome that could be valuable for this purpose. In order to capture the information from historical recombinations, calculation of IBD probabilities simultaneously require information from LD and linkage information [18] at specific targeted regions. Under these circumstances, greater accuracy of positioning the QTL is expected, however, this is associated with the extent of LD in the population. This is because, the values of $\delta_{i,j}$ only rely on LD but not on linkage information, which, for species with relatively small family sizes, would be a limitation when mapping dominant QTL.

REFERENCES

[1] Alfonso L., Haley C., Power of different F_2 schemes for QTL detection in livestock, Anim. Sci. 66 (1998) 1–8.

[2] Almasy L., Blangero J., Multipoint quantitative-trait linkage analysis in general pedigrees, Am. J. Hum. Genet. 62 (1998) 1198–1211.

[3] Amos C., de Andrade M., Genetic linkage methods for quantitative traits, Stat. Methods Med. Res. 10 (2001) 3–25.

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

[5] De Boer I., Hoeschele I., Genetic evaluation methods for populations with dominance and inbreeding, Theor. Appl. Genet. 86 (1993) 245–258.

[6] De Koning D.J., Janss L.L.G., Rattink A.P., Oers P.A.M.V., Vries B.J.D., Groenen M.A.M., der Poel J.J.V., Groot P.N.D., Brascamp E.W., van Arendonk J., Detection of quantitative trait loci for backfat thickness and intramuscular fat content in pigs (*Sus scrofa*), Genetics 152 (1999) 1679–1690.

[7] De Koning D., Nagamine Y., Evans G., Haley C., QTL detection in outbred populations; comparison of variance component and simple methods, in: 7th World Congress on Genetics Applied to Livestock Production, 19–23 August 2002, Montpellier, France, CD ROM 21-16.

[8] Dekkers J., Chakraborty R., Purebred selection for crossbred performance using QTL, in: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, 2002, CD-ROM Communication 09-07.

[9] Falconer D., Mackay T., Introduction to quantitative genetics, 4th edn., Longman Scientific and Technical, 1996.

[10] George A., Visscher P., Haley C., Mapping quantitative trait loci in complex pedigrees: a two step variance component approach, Genetics 156 (2000) 2081–2092.

[11] Gilmour A., Thompson R., Cullis B., Welham S., ASREML reference manual, New South Wales Agriculture, Orange, Australia, 2000.

[12] Goddard M., The validity of genetic models underlying quantitative traits, Liv. Prod. Sci. 72 (2001) 117–127.

[13] Goddard M., Meuwissen T., The use of linkage disequilibrium to map quantitative trait loci, Aust. J. Exp. Agri. 45 (2005) 837–845.

[14] Goring H., Terwilliger J., Blangero J., Large upward bias in estimation of locus-specific effects from genomewide scans, Am. J. Hum. Genet. 69 (2001) 1357–1369.

[15] Hayes B., Laerdahl J., Lien S., Moen T., Berg P., Hindar K., Davidson W., Koop B., Adzhubei A., Hoyheim B., An extensive resource of single nucleotide polymorphism markers associated with atlantic salmon (*Salmo salar*) expressed sequence, Aquaculture 265 (2007) 82–90.

[16] Kearsey M., Farquhar A., QTL analysis in plants; where are we now?, Heredity 80 (1998) 137–142.

[17] Kuehn C., Edel C., Weikard R., Thaller G., Dominance and parent-of-origin effects of coding and non-coding alleles at the acylcoa-diacylglycerol-acyltransferase (*DGAT1*) gene on milk production traits in German Holstein cows, BMC Genet. 8 (2007) 62.

[18] Lee S., van der Werf J., Using dominance relationship coefficients based on linkage disequilibrium and linkage with a general complex pedigree to increase mapping resolution, Genetics 174 (2006) 1009–1016.

[19] Lynch M., Walsh B., Genetics and analysis of quantitative traits, Sinauer Associates Inc., Sunderland, Massachusetts, 1997, Vol. I, 01375 USA.

[20] Meyer K., Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm, Genet. Sel. Evol. 21 (1989) 317–340.

[21] Perez-Enciso M., Fernando R.L., Bidanel J., LeRoy P., Quantitative trait locus analysis in crosses between outbred lines with dominance and inbreeding, Genetics 159 (2001) 413–422.

[22] Pong-Wong R., George A., Woolliams J., Haley C., A simple and rapid method for calculating identity-by-descent matrices using multiple markers, Genet. Sel. Evol. 33 (2001) 453–471.

[23] Visscher P., Medland S., Ferreira M., Morley K., Zhu G., Cornes B., Montgomery G., Martin N., Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between fullsibs, PLoS Genet. 2 (2006) e41.

[24] Visscher P., MacGregor S., Benyamin B., Zhu G., Gordon S., Medland S., Hill W., Hottenga J., Willemsen G., Boomsma D., Liu Y., Deng H., Montgomery G., Martin N., Genome partitioning of genetic variation for height from 11 214 sibling pairs, Am. J. Hum. Genet. 81 (2007) 1104–1110.

[25] Williams J., Blangero J., Power of variance component linkage analysis to detect quantitative trait loci, Ann. Hum. Genet. 63 (1999) 545–563.

[26] Wong G., Liu K., Wang J., Zhang Y., Yang X., Zhang Z., Meng Q., Zhou J., Li D., Zhang J., Ni P., Li S., Ran L., Li H., Zhang J., Li R., Li S., Zheng H., Lin W., Li G., Wang X., Zhao W., Li J., Ye C., Dai M., Ruan J., Zhou Y., Li Y., He X., Zhang Y., Wang J., Huang X., Tong W., Chen J., Ye J., Chen C., Wei N., Li G., Dong L., Lan F., Sun Y., Zhang Z., Yang Z., Yu Y., Huang Y., He D., Xi Y., Wei D., Qi Q., Li W., Shi J., Wang M., Xie F., Wang J., Zhang X., Wang P., Zhao Y., Li N., Yang N., Dong W., Hu S., Zeng C., Zheng W., Hao B., Hillier L., Yang S., Warren W., Wilson R., Brandstrom M., Ellegren H., Crooijmans R., van der Poel J., Bovenhuis H., Groenen M.A., Ovcharenko I., Gordon L., Stubbs L., Lucas S., Glavina T., Aerts J., Rothwell P.K.L., Young J., Rogers S., Walker B., van Hateren A., Kaufman J., Bumstead N., Lamont S., Zhou H., Hocking P., Morrice D., de Koning D., Law A., Bartley N., Burt D., Hunt H., Cheng H., Gunnarsson U., Wahlberg P., Andersson L., Kindlund E., Tammi M., Andersson B., Webber C., Ponting C.P., Overton I.M., Boardman P.E., Tang H., Hubbard S., Wilson S., Yu J., Wang J., Yang H., International chicken polymorphism map

consortium, a genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms, Nature 432 (2004) 717–722.

[27] Xie C., Gessler D., Xu S., Combining different line crosses for mapping quantitative trait loci using the identical by descent-based variance component method, Genetics 148 (1998) 1139–1146.

[28] Xu S., Computation of the full likelihood function for estimating variance at a quantitative trait locus, Genetics 144 (1996) 1951–1960.

[29] Xu S., Atchley W., A random model approach to interval mapping of quantitative trait loci, Genetics 141 (1995) 1189–1197.