Original article

Methods for the detection of multiple linked QTL applied to a mixture of full and half sib families

Hélène GILBERT[a*], Pascale LE ROY[b]

[a] Station de génétique quantitative et appliquée UR337, INRA, F-78352 Jouy-en-Josas, France
[b] INRA-Agrocampus, F-35040 Rennes, France

Abstract – A new multiple trait strategy based on discriminant analysis was studied for efficient detection of linked QTL in outbred sib families, in comparison with a multivariate likelihood technique. The discriminant analysis technique describes the segregation of a linear combination of the traits in a univariate likelihood. This combination is calculated for each pair of positions depending on the inheritance of the pairs of QTL haplotypes in the progeny. The gains in power and accuracy for position estimations of multiple trait methods in grid searches were evaluated in reference to single trait detections of linked QTL. The methods were applied to simulated designs with two correlated traits submitted to various effects from the linked QTL. Multiple trait strategies were generally more powerful and accurate than the single trait technique. Linked QTL were distinguished when they were separated enough to identify informative recombinations: at least two genetic markers and 25 cM between the QTL under the simulated conditions. Except in a particular case, discriminant analysis was at least as powerful as the multivariate technique and its implementation was five times faster. Combining the advantages from both methodologies, we finally propose a complete strategy for rapid and efficient systematic multivariate detections in outbred populations.

QTL detection / linked QTL / multiple trait / sib families / simulations

1. INTRODUCTION

Maximum likelihood methods [23], as well as linearised approximations [11, 15, 16] or variance component methods [1] have been widely used to detect individual QTL on traits of interest. Hints of joint influence from several chromosomal regions on some traits *e.g.* [2, 3] have been frequently pointed out. Multivariate developments have been explored [13, 32, 35, 36] to better describe these regions. First, maximum likelihood techniques were

---

* Corresponding author: helene.gilbert@jouy.inra.fr

set up to differenciate two linked QTL from one pleiotropic QTL in inbred designs [4, 11, 14, 17, 21, 24, 30, 31, 33]. Later, in outbred populations, similar techniques were applied to detect multiple but unlinked QTL [18] and variance component methods were conducted to distinguish linked QTL from a pleiotropic locus [22, 26]. However in practice, outbred populations with a mixture of half- (and full-) sib families have high computational requirements and low information contents compared to inbred populations: the QTL allele contrasts have to be described independently for each sire (and possibly dam, when full-sibs are used) family [5]. Thus, applying multivariate strategies can affect the power of detection due to increased numbers of parameters to be estimated jointly. Alternative single QTL techniques [6, 21, 27, 34] for multiple trait detections have been proposed using linear combinations of the traits. Thus, the likelihood results in a univariate function similar to single trait models, saving computational costs and increasing power. The computation of a pertinent linear combination for the putative QTL has been discussed. A first proposal [34] was derived from the phenotypic covariance matrix of the traits, calculated independently from the QTL influence: a high power of detection was reached only when the correlation between traits due to the QTL was similar to their phenotypic correlation. A second technique resulted from the maximisation of the ratio between the genetic covariance due to the putative QTL and the residual covariance of the model ([21] in inbred crosses, [6, 7] in outbred populations). In inbred populations, recombinant individuals in intervals between genetic markers were excluded and specific ratios were computed per interval based on grand-parental origins of the corresponding haplotype. In outbred populations, calculations were performed at each position considering the probability that the progeny received one or the other QTL haplotype from their sire. As a result of the discriminant analysis technique DA [29], the linear combination of traits that maximises the between group (QTL derived) and the within group (residual derived) covariance ratio best discriminates the groups of progeny. This alternative gives results identical to the multivariate likelihoods, when the model of the discriminant transformation is identical to the model of the multivariate likelihood [6, 27].

In the present paper, multiple trait techniques were applied to the detection of linked QTL in the framework of likelihood maximization techniques to get an efficient and more general alternative to variance component methods. Here, we describe the corresponding discriminant transformation in mixtures of full and half-sib families, and compare it to a multivariate likelihood technique for their computing costs, their power and their accuracy of parameter estimations, and their respective abilities to improve single trait detections.

Finally we propose a strategy to systematically implement the multivariate detections. The methods were programmed as options in the QTLMAP software [5, 6, 10, 25, 28] available upon request.

2. MATERIALS AND METHODS

2.1. Tests

To discriminate one QTL from two QTL segregating on a chromosome, the null hypothesis was set to "there is one QTL in the linkage group" and the alternative hypothesis was "there are two QTL at the tested positions". In notations, single trait hypotheses and multiple trait hypotheses, which imply pleiotropic QTL, were distinguished: the null hypotheses are denoted H1 (single trait) and H11 (multiple traits), and the alternative hypotheses are denoted H2 (single trait) and H22 (multiple traits).

Following [6], likelihoods at position x were noted Λ_l^x under H1 for trait l and Λ^x under H11 for traits $l = 1, \ldots, p$. Under the hypothesis of two QTL located at $\mathbf{x} = (x1; x2)$, $x1 < x2$, likelihoods were $\Lambda_l^{\mathbf{x}}$ for H2 for trait l, and $\Lambda^{\mathbf{x}}$ for H22. Maximum likelihood techniques were used to get parameter estimates, so likelihood ratio tests were applied at each $\mathbf{x}$. For single trait models, $LRT_l^{\mathbf{x}}$ was $-2\ln[max_x(\Lambda_l^x)/max_{\mathbf{x}}(\Lambda_l^{\mathbf{x}})]$, and for multiple trait tests $LRT^{\mathbf{x}} = -2\ln[max_x(\Lambda^x)/max_{\mathbf{x}}(\Lambda^{\mathbf{x}})]$. Since the tests compare the two QTL *vs.* the best single QTL hypothesis, it may be over conservative, due to putative ghost effects under the single QTL model.

2.2. Likelihood

Techniques for detecting two QTL were developed from single QTL methods proposed for mixtures of half- and full-sib families using partially linearised likelihoods [6, 25]. Main notations are summarised in Table I. Three techniques were addressed for two QTL tests: a single trait method (ST2), and two multiple trait strategies, MV2 using a multivariate likelihood, and DA2 based on a univariate likelihood applied to a linear combination of traits obtained from a discriminant analysis. Likelihood calculations under H1 and H11 were described in [6], we focussed this paper on the two-QTL models.

Grid search detections [9] were retained, supported by risks of "ghost QTL" detections [23] or missing QTL with single QTL models. Thus, the likelihoods described the joint distribution of traits due to each pair of QTL. In a first step, at positions $\mathbf{x} = \{x1; x2\}$, the probabilities that the progeny inherited a given

Table I. Notations for half- and full-sib family observations.

Notation	Signification
i for $i=1,\ldots,n$	sire
j for $j=1,\ldots,n_i$	dam (within sire)
k for $k=1,\ldots,n_{ij}$	progeny (within mate)
M_i	marker information for sire i family
yp_{ijkl} for $l=1,\ldots,p$	phenotype for quantitative trait l of progeny ijk
$\mathbf{yp}_{ijk}$	vector for p phenotypes for quantitative traits l of progeny ijk
$\widehat{hs}_i$	most probable sire i genotype for genetic markers
hd_{ij}	dam ij genotype having a probability greater than 0.1 for the genetic markers
xr for $r=1,2$	QTL$_r$ position
$\mathbf{x} = \{x1, x2\}$	vector of QTL positions
qr_s for $r=1,2$	sire haplotype transmitted for QTL$_r$
	$qr_s=1$ from the grand sire, $qr_s=2$ from the grand dam
$\mathbf{q_s} = \{q1_s, q2_s\}$	pair of haplotypes inherited from the sire at positions $\mathbf{x}$
qr_d for $r=1,2$	dam haplotype transmitted for QTL$_r$
	$qr_d=1$ from the grand sire, $qr_d=2$ from the grand dam
$\mathbf{q_d} = \{q1_d, q2_d\}$	pair of haplotypes inherited from the dam at positions $\mathbf{x}$
$d_{ijk}^{\mathbf{x}} = \{\mathbf{q_s}, \mathbf{q_d}\}$	pairs of haplotypes inherited by progeny ijk at positions $\mathbf{x}$
$\mu_{il}^{\mathbf{xq_s}}$	phenotypic mean of the progeny, which received $q1_s$ and $q2_s$ at $\mathbf{x} = (x1; x2)$ from the sire i for the quantitative trait l
$\mu_{ijl}^{\mathbf{xq_d}}$	phenotypic mean of the progeny which received $q1_d$ and $q2_d$ at $\mathbf{x} = (x1; x2)$ from the dam ij for the quantitative trait l
σ_{il}^2	variance of the sire i family for the quantitative trait l
$f(yp_{ijkl}; \mu_{il}^{\mathbf{xq_s}} + \mu_{ijl}^{\mathbf{xq_d}}, \sigma_{il}^2)$	normal penetrance function, conditional on the haplotypes $d_{ijk}^{\mathbf{x}}$ transmitted

pair $d_{ijk}^{\mathbf{x}} = (\mathbf{q_s}, \mathbf{q_d})$ of QTL haplotypes from its sire (haplotypes $\mathbf{q_s} = (q1_s, q2_s)$) and dam (haplotypes $\mathbf{q_d} = (q1_d, q2_d)$) were calculated (see Tab. I for notation and [5]). The probabilities that progeny inherited the QTL haplotypes at position $x1$ and $x2$, $p\left(d_{ijk}^{\mathbf{x}} = (qx_s, qx_d)/\widehat{hs}_i, hd_{ij}, M_i\right)$ [25] were multiplied, giving the joint probability $p\left(d_{ijk}^{\mathbf{x}}/\widehat{hs}_i, hd_{ij}, M_i\right)$: due to properties of interval mapping, the inheritance probabilities at positions $x1$ and $x2$ are conditionnally independent, given that there is at least one informative marker between the tested positions [23]. In a second step, given that only the most probable sire haplotype $\widehat{hs}_i$ was retained to compute the sire likelihood, the partial linearisation of the likelihood within full-sib families was obtained. Finally, the joint distribution of the traits depended on two QTL effects and a polygenic value for the genetic part [11, 12, 30]. Thus, Y_{ijkl} the linearised performance for the

individual ijk for trait l was the following:

$$Y_{ijkl} = yp_{ijkl} - \sum_{\mathbf{q_s}=(1,1)}^{(2,2)} \sum_{\mathbf{q_d}=(1,1)}^{(2,2)} p\left(d_{ijk}^{\mathbf{x}} = (\mathbf{q_s}, \mathbf{q_d})/\widehat{hs}_i, hd_{ij}, M_i\right)\left(\mu_{il}^{\mathbf{xq_s}} + \mu_{ijl}^{\mathbf{xq_d}}\right) \quad (1)$$

where $\mu_{il}^{\mathbf{xq_s}}$ (respectively $\mu_{ijl}^{\mathbf{xq_d}}$) is the phenotypic mean of the progeny which received $\mathbf{q_s} = (q1_s; q2_s)$ (respectively $\mathbf{q_d} = (q1_d; q2_d)$) at $\mathbf{x} = (x1; x2)$ from sire i (respectively dam ij) for quantitative trait l. These phenotypic means were

$$\mu_{il}^{\mathbf{xq_s}} = \mu_{il} - (-1)^{q1_s}\frac{\alpha_{il}^{x1}}{2} - (-1)^{q2_s}\frac{\alpha_{il}^{x2}}{2},$$

and $\mu_{ijl}^{\mathbf{xq_d}}$ was obtained similarly, where μ_{il}, μ_{ijl} are respectively the within half-sib and within full-sib phenotypic means, and α_{il}^{x1}, α_{il}^{x2}, α_{ijl}^{x1} and α_{ijl}^{x2} are respectively the within half-sib and within full-sib average effects of substitution following Fisher of the QTL1 ($x1$ location) and QTL2 ($x2$ location) for the quantitative trait l. Thus at $\mathbf{x}$, with ϕ_{ijk} the penetrance function for progeny ijk from sire i and dam ij, the partially linearised likelihood under any alternative hypothesis is the following:

$$\prod_{i=1}^{n} \prod_{ij=1}^{n_i} \sum_{hd_{ij}} p(hd_{ij}/\widehat{hs}_i, M_i) \prod_{k=1}^{n_{ij}} \phi_{ijk} \quad (2)$$

Single trait method (ST2): In the single trait model, the trait value had a gaussian distribution f, so the penetrance function was:

$$\phi_{ijk} = f(yp_{ijkl}/\widehat{hs}_i, hd_{ij}, M_i) = \frac{1}{\sqrt{2\pi}\sigma_{il}} \exp\left(-\frac{1}{2}\left(\frac{Y_{ijkl}}{\sigma_{il}}\right)^2\right).$$

Additional parameters were estimated compared to the likelihood under the null hypothesis H1: one QTL effect per sire $i = 1, \ldots, n$, one per dam $ij = 1, \ldots, n_i$, and one position, giving theoretically $n + \sum_{i=1}^{n} n_i + 1$ degrees of freedom for the comparison of H1 *vs.* H2.

Multivariate penetrance function (MV2): The first multiple trait model was based on a multivariate penetrance function. It was a multivariate extension of the single trait model (*cf.* previous paragraph), or equivalently a two-QTL extension of the multivariate MV method in [6] for single QTL detection. The multinormal penetrance function is given by:

$$\phi_{ijk} = f(yp_{ijk}/\widehat{hs}_i, hd_{ij}, M_i) = \sqrt{\frac{|\mathbf{VC}_i^{-1}|}{2\pi}} \exp\left(-\frac{1}{2}\mathbf{Y}'_{ijk}\mathbf{VC}_i^{-1}\mathbf{Y}_{ijk}\right),$$

where $\mathbf{Y}_{ijk} = \left[Y_{ijkl}; l = 1, \ldots, p \right]$, $\mathbf{VC}_i$ is the residual covariance matrix between traits for the sire family i, $\mathbf{Y}'_{ijk}$ is the transposed vector, $\mathbf{VC}_i^{-1}$ the inverse and $|\mathbf{VC}_i^{-1}|$ the determinant.

Comparing the multiple QTL (H22) *vs.* the single QTL (H11) multiple trait hypotheses, $(n + \sum_{i=1}^{n} n_i) * p + 1$ additional parameters were estimated under H22: one QTL effect for each sire and dam and each trait at the second position, plus the position.

Discriminant transformation (DA2): The discriminant analysis is an efficient alternative to the multivariate likelihood in multiple trait detections, based on a univariate analysis of a linear combination of the traits [6]. The major gain is expected for computing time, in comparison to the multivariate likelihoods in MV2.

For two-QTL models, linear combinations of traits were calculated at each $\mathbf{x}$, maximising the ratio of the genetic variability due to the QTL (variability between the four groups of progeny inheriting the different QTL haplotype combinations $\mathbf{q_s} = (q1_s, q2_s)$), and the residual variability due to any other factor (variability within those groups). In practice, QTL haplotypes inherited from the sire were known in probabilities, derived from the $p\left(d_{ijk}^{\mathbf{x}} = (\mathbf{q_s}, \mathbf{q_d})/\widehat{hs}_i, hd_{ij}, M_i\right)$ described in equation (1): each progeny thus belonged to every QTL haplotype group, weighted by this probability. At $\mathbf{x}$, a unique linear combination of traits $z_{ijk}^{\mathbf{x}} = \sum_{l=1}^{p} \beta_l^{\mathbf{x}} y_{ijkl}$ was obtained, and the Gaussian distribution f was used in equation (2) for ϕ_{ijk}, $f(z_{ijk}^{\mathbf{x}}/\widehat{hs}_i, hd_{ij}, M_i)$ resulting in a univariate likelihood. Thus now the mean, variance and QTL effect estimates from the likelihood maximisation refer to the linear combination.

2.3. Simulations

The three methods were compared for detection of two QTL using simulations. An intercross design was simulated, with 10 unrelated males from population P1 and 10 unrelated females from population P2 in a grand-parental generation (F0). Each mating produced one sire and two dams (parental generation F1), mated to obtain 10 half-sib families split into 20 balanced full-sib families to get 500 progeny (generation F2). The linkage group was sized 100 cM, with 9 equally spaced genetic markers. Each genetic marker had five alleles with a 0.2 frequency either in P1 and P2. Grand-parental haplotypes were randomly drawn and allelic transmissions followed Mendelian rules. Different QTL alleles were fixed in P1 and P2, so all F1 individuals were heterozygous for the biallelic QTL (this was not the case for genetic markers). Four sets of locations

Table II. Combinations of QTL simulated effects under the alternative hypothesis: QTL effects on the traits from one given haplotype.

Case	1		2		3		4		5		6	
Trait	1	2	1	2	1	2	1	2	1	2	1	2
QTL1	a	0	a	a	a	0	a	a	a	$-a$	a	$-a$
QTL2	a	0	a	a	$-a$	0	$-a$	$-a$	a	$-a$	$-a$	a

of a first QTL (QTL1) and second QTL (QTL2) were considered: (1) QTL1 at 31 cM and QTL2 at 81 cM (QTL physically independent), (2) QTL1 at 31 cM and QTL2 at 57 cM (two genetic markers between the QTL), (3) QTL1 at 31 cM and QTL2 at 43 cM (one genetic marker separates the QTL), (4) QTL1 at 29 cM and QTL2 at 34 cM (with no genetic marker separating them).

Two traits were simulated in the present study, using a completely additive model. For both traits, besides the QTL, the polygenic heritability was 0.2, the residual variance was 1 and the residual correlation was -0.4. We defined a panel of six combinations for the QTL effects (Tab. II), with respect to considerations on multiple trait or linked QTL detection in the literature:

– the number of traits influenced by the two QTL: in single QTL models, a significant improvement of QTL detection was achieved using multiple trait techniques if traits were correlated, either genetically or residually [6, 17, 21]. The effect of QTL_r on trait l was $|a_{rl}| = 0.5 = a$, $r = 1, 2$, $l = 1, 2$, except for cases 1 and 3, where the QTL were not pleiotropic ($a_{r2} = 0$, $\forall r$);

– the relationship between the residual and the genetic (due to the QTL) correlations: multiple trait detections proved to be more powerful and precise if the residual correlation had a sign opposite to the correlation due to the pleiotropic QTL allele segregation [14, 19, 27]. For two QTL models, the sum of the QTL effects from a given haplotype on each trait was considered to define a genetic correlation sign due to the QTL: cases 1, 2 and 4 corresponded to opposite genetic and residual correlations between the traits, whereas in cases 3, 5 and 6 they were in similar directions;

– the phase of the QTL alleles for each trait: a one-QTL analysis was demonstrated to be very powerful when the effects of two linked QTL were in phase (each haplotypic phase carries QTL alleles having the same sign effects, either positive or negative), a phenomena described as "ghost QTL" by Lander and Botstein [23], whereas the one-QTL model was very poor at detecting QTL when two QTL in repulsion phase were segregating (alleles of opposite sign effects are carried by a haplotypic phase). In two-QTL detection, impacts are not straightforward: with QTL in phase, a higher

maximum likelihood is expected under the null hypothesis, implying higher thresholds. We simulated cases 1, 2 and 5 with QTL alleles in phase, and cases 3, 4 and 6 with QTL alleles in repulsion.

2.4. Comparison of the methods

The three techniques for 2 QTL detections were successively applied to each simulated design. In each of the 24 combinations of effects (six cases) and QTL positions (four distances), 100 simulations were computed, and the maximum LRT of each technique, their positions and maximum likelihood estimates (MLE) under the null and alternative hypothesis were recorded. The smallest estimated position was arbitrarily attributed to QTL1. Finally, the computing time required to get the MLE with each technique under the two QTL model was stored. From these records, the statistical methods were compared for the following:

- computing time, averaged for each method over the 24×100 analysis of simulated designs;
- power of each method for the test of one QTL *vs.* two QTL, computed over the 100 simulations as the percentage of maximum LRT greater than an empirical threshold;
- accuracy of the position estimates, from average estimates of QTL1 and QTL2 positions, over the 100 simulations: Mean Squared Errors (MSE) were computed as the mean of the squared differences between the estimates in each simulation and the true value. This formally combines bias and estimation variance in one criterium;
- accuracy of the QTL effect estimates for ST2 and MV2, calculated over the 100 simulations using MSE for individual QTL effects a_{rl}. Since DA2 analysis estimates parameters for linear combinations of traits, no direct estimation of QTL effects on each trait was available.

2.5. Empirical threshold computation

Test statistic distributions under the null hypotheses were not available: in addition to the usual source of deviation from asymptotic conditions in single QTL tests, a grid search was applied in this study, involving an increased number of correlated tests compared to one dimension single QTL tests. To derive an asymptotic distribution of the test statistic under the null hypothesis would be too time consuming, thus 200 simulations were conducted under the null

hypothesis "there is one QTL in the linkage group" to get empirical 5% thresholds. For each combination of effects and QTL positions in two-QTL models (Tab. II), a particular set of position and effects from a single QTL had to be chosen. From the single QTL model analysis applied to each simulation under a two-QTL model, estimates for QTL position and effects were averaged (over the 100 simulations performed to calculate power). Given these estimates, a single QTL was simulated and various two-QTL tests were applied to obtain the corresponding thresholds. This proposition differed from [17], where a different threshold was empirically computed for each simulation. This approach was chosen because of the excessive computing times for the likelihood maximizations in models of mixture of full- and half-sib families. Some simulations with one QTL demonstrated a good approximation of the type I error, even if it tended to be slightly lower than 5%. In practice, the simulations would be conducted using position and effects estimated with a one-QTL analysis on the data set considered, which may be quite conservative.

3. RESULTS

3.1. Computing time

For multiple QTL tests, the single trait analysis lasted for 129.10 seconds (s), the discriminant analysis for 71.44 s and the multivariate likelihood for 345.36 s, on average. Compared to single QTL tests, computing times with univariate likelihoods were thus increased at least 30 times, due to both a larger number of individual tests in the grid search (21 likelihood maximisations *vs.* 221 with 100 cM and 5 cM-steps) and a higher number of estimated parameters. For the multivariate likelihood only, the number of parameters also depended on the number of traits. In two-trait models, the computing cost was 5 times greater with MV2 than with DA2, so multivariate two-QTL likelihoods can be computationally exhausting for systematic genome scans.

3.2. Thresholds

Estimates of the effects and position obtained under the single QTL hypothesis (Tab. III) confirmed risks of "ghost" QTL detections when QTL segregate with coupling phases [11, 30]: position estimations were close to the mean between the simulated QTL positions, and effect estimates were greater than the individual effect of each QTL on the trait. On the contrary, when QTL alleles were in repulsion, average position estimates were close to the middle of

Table III. Average parameter estimates under one pleiotropic QTL model, simulations with two traits l determined by two linked QTL_r, a_{lr} = QTL_r effect on trait l.

		Position estimates		Effect estimates	
a_{l1} a_{l2}		+ +	+ −	+ +	+ −
QTL1/QTL2	31/81	58.0	57.1	0.634	−0.025
(cM)	31/57	43.9	45.5	0.767	−0.003
	31/43	38.3	48.4	0.793	−0.010
	29/34	28.5	51.3	0.827	−0.010

Table IV. Thresholds for two-QTL *vs.* one-QTL tests.

Case	1	2	3 - 4 - 6	5	1	2	3 - 4 - 6	5
QTL1/QTL2		31/81 cM				31/57 cM		
ST2	72.0	63.6	57.6	67.8	69.8	67.5	63.7	67.1
MV2	111.3	109.0	99.1	109.5	110.4	112.7	99.3	111.8
DA2	91.7	123.4	103.6	98.5	102.1	133.3	111.7	109.3
QTL1/QTL2		31/43 cM				29/34 cM		
ST2	65.3	65.9	60.4	63.7	75.8	70.1	57.9	71.4
MV2	110.2	111.2	100.6	107.5	110.8	117.8	99.9	110.0
DA2	114.8	164.0	99.5	121.6	108.9	150.8	107.1	118.8

ST2: single trait method; MV2: multivariate method; DA2: discriminant analysis.

the linkage group and effect estimates were close to zero, so that very few QTL would be detected under the one-QTL model. Thus, to save computational time, one threshold common for cases 3, 4 and 6 was estimated, with the average effect equal to −0.01 for any distance simulated between the QTL.

When QTL alleles were in coupling phase (cases 1, 2 and 5), thresholds (Tab. IV) seemed to increase as QTL got closer to each other, in relation with increased simulated effects. On the contrary, when QTL alleles were in repulsion (case 3), thresholds were relatively independent of the QTL positions, due to similar simulated small effects. In these cases, thresholds were generally lower than in allele coupling cases. Differences related to the QTL genetic correlation and the residual correlation between the traits: as the effects were higher, and the correlations in the opposite direction, the thresholds increased (from case 1 to case 4 to case 2), especially for discriminant analyses. In fact, these are situations where single QTL detection are known to be more powerful. An accurate estimation of the QTL effects under the null hypothesis (one-QTL model) may thus be required to estimate pertinent empirical thresholds.

Table V. Power (%) of multiple QTL detections.

$a_{11}\ a_{21}$	$+0$	$++$	$+0$	$++$	$+-$	$+-$	$+0$	$++$	$+0$	$++$	$+-$	$+-$
$a_{12}\ a_{22}$	$+0$	$++$	-0	$--$	$+-$	$-+$	$+0$	$++$	-0	$--$	$+-$	$-+$
QTL1/QTL2			31/81 cM						31/57 cM			
ST2	42	65	90	91	55	78	6	10	30	23	7	27
MV2	50	100	86	100	84	90	8	33	48	97	11	59
DA2	62	99	20	100	84	37	11	47	5	28	26	4
QTL1/QTL2			31/43 cM						29/34 cM			
ST2	7	6	9	6	12	13	1	2	6	7	2	4
MV2	3	17	15	39	18	16	3	8	9	5	7	4
DA2	9	4	8	5	25	9	21	19	4	5	26	4

a_{lr} = QTL$_r$ effect on trait l, a = 0.5; ST2: single trait method; MV2: multivariate method; DA2: discriminant analysis.

3.3. Power

When comparing multiple trait strategies, two situations (Tab. V) arose: when the QTL alleles were in coupling phase, generally DA2 had greater power compared to MV2, but when QTL alleles were in repulsion, MV2 was often more powerful, and surprisingly DA2 barely reached the 5% of type I error. Actually, in the discriminant technique the haplotypic effects were nearly null in the groups of the parental haplotypes: QTL effects cancelled each other. However, compared to the groups of the recombinant haplotypes, these groups represent most of the progeny, thus most of the information content for the discrimination. In repulsion designs, the power of the discriminant analysis might thus have been lowered by construction. To confirm this hypothesis, new designs in repulsion phase were simulated with unbalanced effects, and DA2 showed at least equivalent power compared to MV2: as an example, when $a_{1l} = 0.5$ and $a_{2l} = -0.25$ for each trait l, 34% of detection was obtained with DA2, compared to 21% with MV2.

Power was higher using a multiple trait method than the single trait method even in cases 1 and 3 where only one of the correlated traits was influenced by the QTL. When the QTL affected both traits, higher power was achieved by multiple trait detections if the QTL generated a genetic correlation opposite to the residual correlation: higher power for case 2 than for case 5 (QTL alleles in coupling phase) and for case 4 than for case 6 (QTL alleles in repulsion). Moreover, both single and multiple trait detections were more powerful when the QTL alleles were in repulsion, related to lower threshold values.

Finally, power was very dependent on the distance and number of genetic markers separating the QTL. When QTL were 50 cM apart, two QTL were distinguished in more than 60% of the simulations, but only in more than 25%

when they were 26 cM apart, and from 10 to 40% when they were 12 cM apart. When they were 5 cM apart power reached the type I error of the tests.

3.4. Parameter estimates

Positions. Generally, biases of position estimates (not shown) were slightly higher using DA2 than MV2, but MSE (Tab. VI) were lower with DA2 in cases when the power of detection was high, which corresponds to lower estimation variances. Accuracy improvement using multiple trait strategies compared to ST2 was not systematic. Actually, very high accuracies were reached with ST2 in general, which is common when QTL effects are high enough.

Biases and MSE were higher for QTL2 than QTL1, especially when the QTL got closer, reaching more than 20 cM when the QTL were 5 cM apart, but few simulations were performed with the interval between the QTL centered on the linkage group (data not shown), showing equal bias on the two position estimates.

QTL effects. A higher accuracy was achieved with MV2 (Tab. VII) compared to ST2, except when QTL alleles were in repulsion (estimation variances were higher). No systematic improvement was observed when only one trait depended on the QTL, but in most favorable cases, MV2 reduced MSE from 50 to 85%. MSE were smaller when the QTL were 50 cM apart than when they got closer but no trend could describe the improvement obtained with MV2 compared to ST2 with closer QTL. Looking further, when the QTL got close to each other the MSE increased mostly because of increased biases. They could reach more than half of the simulated values, particularly for QTL2.

4. DISCUSSION

We studied the detections of two linked QTL in a mixture of half and full sib families using two multiple trait methodologies. Taking advantage of the correlation between the traits, improvements compared to the single trait detections for power of detection and accuracy of parameter estimates, were obtained in accordance with tendencies previously described for analysis of inbred QTL detection designs [17, 21]. Previous studies on outbred populations [26] generally proposed to test the hypothesis H1 "one pleiotropic QTL" *vs.* H2 "two non pleiotropic linked QTL". The test we envisaged in this paper is more general, due to the properties of the discriminant analysis. It should be considered as a first step in a genome scan, to identify pleiotropic regions, which would have to

Table VI. MSE of position estimates ($\times 10^2$).

$a_{11}\,a_{21}$	$+\,0$	$+\,+$	$+\,0$	$+\,+$	$+\,-$	$+\,-$	$+\,0$	$+\,+$	$+\,0$	$+\,+$	$+\,-$	$+\,-$
$a_{12}\,a_{22}$	$+\,0$	$+\,+$	$-\,0$	$-\,-$	$+\,-$	$-\,+$	$+\,0$	$+\,+$	$-\,0$	$-\,-$	$+\,-$	$-\,+$
QTL1/QTL2	31/81 cM						31/57 cM					
$ST2_1$	1.33	0.70	0.82	0.72	1.03	1.06	1.97	1.56	1.35	1.79	2.20	1.65
$ST2_2$	1.08	0.51	0.95	1.19	1.14	0.90	3.42	2.93	1.76	2.39	3.09	1.55
$MV2_1$	1.38	0.22	1.16	0.25	0.51	0.69	2.18	1.08	2.16	0.32	1.97	1.71
$MV2_2$	1.85	0.80	1.68	0.74	1.31	1.21	3.84	1.85	3.06	0.69	3.27	2.20
$DA2_1$	1.26	0.09	1.16	0.12	0.41	0.61	2.63	1.80	2.56	0.25	2.35	3.15
$DA2_2$	0.83	0.12	0.84	0.10	0.40	0.91	5.01	2.41	2.11	0.22	4.87	3.42
QTL1/QTL2	31/43 cM						29/34 cM					
$ST2_1$	1.83	1.98	3.82	5.16	1.80	3.81	1.13	1.38	7.06	8.38	1.48	5.83
$ST2_2$	6.21	6.41	9.48	9.46	5.01	6.36	8.47	8.99	17.23	17.91	8.59	17.61
$MV2_1$	1.24	1.48	4.49	2.56	1.79	3.28	1.42	1.20	6.50	6.53	1.18	6.97
$MV2_2$	6.98	4.29	8.60	4.87	5.36	5.89	9.22	6.97	16.56	18.69	10.87	19.00
$DA2_1$	1.94	1.58	5.78	1.67	2.41	5.03	1.23	0.78	8.38	7.12	0.71	7.43
$DA2_2$	12.57	9.16	10.49	3.41	9.78	9.84	18.79	18.81	16.90	17.10	18.85	17.50

a_{lr} = effect of the QTL_r alleles on the trait l, $a = 0.5$; $ST2_l$: single trait method, QTL_r; $MV2_r$: multivariate method, QTL_r; $DA2_r$: discriminant analysis, QTL_r.

Table VII. MSE of QTL effect estimates.

$a_{11}\,a_{21}$	+0	++	+0	++	+−	+−	+0	++	+0	++	+−	+−
$a_{12}\,a_{22}$	+0	++	−0	−−	+−	−+	+0	++	−0	−−	+−	−+
QTL1/QTL2	31/81 cM						31/57 cM					
$ST2^1_1$	3.76	3.07	3.95	3.56	3.14	7.17	8.33	6.28	27.78	16.71	14.18	13.81
$ST2^1_2$	4.40	3.13	3.85	4.31	3.13	7.08	8.81	6.66	27.84	15.97	8.71	14.16
$MV22^1_1$	3.41	2.75	6.16	2.77	3.11	3.25	8.11	5.17	34.70	14.28	8.37	13.73
$MV22^1_2$	3.39	2.77	6.30	2.97	2.97	3.25	8.09	5.35	33.45	13.88	7.39	14.25
$MV22^2_1$	4.48	2.70	5.09	3.02	2.95	3.35	7.80	5.71	22.82	11.15	6.68	18.00
$MV22^2_2$	4.13	2.83	5.22	3.19	3.10	3.24	7.66	5.63	22.60	13.08	6.39	19.11
QTL1/QTL2	31/43 cM						29/34 cM					
$ST2^1_1$	10.32	14.45	23.31	19.05	10.19	27.58	14.73	12.58	25.08	31.38	14.25	20.47
$ST2^1_2$	10.98	14.29	23.58	19.02	11.50	28.25	16.88	13.07	24.99	33.15	14.35	24.79
$MV22^1_1$	11.21	10.14	22.89	28.88	9.69	39.13	12.44	13.34	27.95	19.16	10.46	21.60
$MV22^1_2$	11.48	9.81	24.00	27.33	10.83	37.33	13.65	14.09	27.46	20.74	12.19	22.17
$MV22^2_1$	11.44	13.30	14.39	36.41	9.42	38.98	12.19	14.80	17.92	19.65	10.38	22.30
$MV22^2_2$	10.76	12.24	16.15	36.08	10.50	35.84	11.93	15.53	17.55	20.15	11.98	23.69

a_{lr} = effect of the QTL$_r$ alleles on the trait l, a = 0.5; ST2^l_r: single trait method, QTL$_r$ trait 1; MV2^l_r: multivariate method, QTL$_r$ trait l.

be further examined with more specific models – using multivariate functions for example or different linear transformations.

A first key to test the two-QTL hypothesis *vs.* the one-QTL hypothesis was the threshold estimations. The technique of Monte Carlo simulations under the null hypothesis was chosen, implying a hypothesis on the position and QTL effects to be simulated. Threshold calculations showed great robustness to the position assumptions, but undesired sensitivity to the QTL effect values input. Knott and Haley [17] observed similar trends and proposed alternatively non parametric bootstraps [24] to distinguish two linked QTL. The computational cost might be too high to be applied to mixtures of full and half sib family detections, but it could be envisaged in other designs.

A second key was related to the genetic information. Detections were powerful and parameter estimations accurate with at least two genetic markers or 26 cM between the QTL similarly to [17, 20, 31]. On the contrary, systematic biases were observed on both QTL location and effect estimates when the QTL got closer, due to limited information on recombinations given the population size and marker density. These biases should be related to the minimum and optimal map density of 20 cM in QTL designs described in [23]. Thus, efficient two-QTL detections depend not only on informative and dense genetic maps but also on a large number of recombinant progeny. In this paper, an approximate calculation of the probability of the joint transmission of the couples of haplotypes was chosen. It is exact only when the transmission probabilities are independent of each other, which occurs either when an informative marker separates the positions or when they are more than 50 cM apart. In practice, significant results in intervals where markers are not informative for all the families should be confirmed by additional genotyping. Similarly, the joint analysis of positions closer than 20 cM with no informative marker in between should be avoided, or results at least carefully questionned.

The third key for detection power was the coupling or repulsion states of the QTL alleles. Higher powers corresponded to repulsion phases, where usually no QTL are detected under one-QTL models. On the contrary, alleles in coupling phase often result in "ghost QTL" detection under one-QTL models, giving higher thresholds, so distinction using two-QTL tests are harder to achieve. An alternative would be to calculate thresholds from the highest likelihood ratio test at the two positions given by the two-QTL model: the tests would be less conservative, but they would also be less general at the chromosome level. Two-QTL detections require systematic searches, performed also in chromosomal regions where no QTL are mapped under one-QTL models. This might be more powerful than stepwise model selections based on significant one-QTL

detections [17, 31]: these may not detect linked QTL with alleles in repulsion, whereas practically, such QTL might be particularly useful to disentangle undesired genetic correlations between traits.

Finally, DA2 was at least as powerful and accurate for position estimations as MV2, except in the particular cases of QTL alleles with equal and opposite effects. Based on these results, a general strategy efficient and easily applicable in outbred populations was derived for multivariate QTL detections. It was originally aimed at saving computing time for analysis of mixtures of full- and half-sib families, but can evidently be applied to any design. When all sires are not heterozygous for the QTL, or allele contrasts differ from one sire to the other, Gilbert and Le Roy [7] showed that the discriminant transformation performed within families can improve the power of detection of a pleiotropic QTL if large sire families are available. A similar calculation could easily be set up to account for allele heterogeneity in the two-locus transformations.

The general approach has proved to be successful in performing a QTL detection for carcass composition traits on porcine chromosome 7 [8]. Considering a QTL detection design, a set of $k = K$ traits is defined depending on physiological functions (fat deposition) or productive items (carcass adiposity). Thus, the following strategy can be applied to every chromosomal region:

(1) on k traits, discriminant analysis: likelihood maximisations under no QTL, one-QTL and two-QTL models, p-value computations for the test 0/1 QTL ($T0/1_k$):

 – it is significant, p-value computation for the test 1/2 QTL ($T1/2_k$),
 – it is not significant, p-value computation for the test 0/2 QTL ($T0/2_k$).

The decision is:
(a) accept one-QTL model: if $T0/1_k$ is significant, and $T1/2_k$ is not,
(b) accept the two-QTL model: if $T0/1_k$ and $T1/2_k$ are significant or $T0/1_k$ is not significant and $T0/2_k$ is significant;

(2) if any test is significant at step 2, the trait with the smallest contribution to the linear combination at the maximum LRT is removed from the detection, and new searches are conducted with $k = K - 1$ traits:

(a) if one-QTL model was kept with $k + 1$ traits, $T0/1_k$ is conducted,
(b) if $T0/2_k + 1$ was significant, $T0/2_k$ is applied,
(c) if $T1/2_k + 1$ was significant, $T1/2_k$ is applied;

(3) if the k trait test is at least as significant as the $k + 1$ test, a new trait is removed from the detection and the analyses in step 2 are conducted again, with $k - 1$ traits. If the k trait test is not at least as significant as the $k + 1$ test, the $k + 1$ trait model is finally selected.

The estimates for QTL effects in the final set of traits should be obtained by maximising the multivariate likelihood of the current QTL model at the most likely location obtained in step 3. Then, some finer models, as with linked QTL influencing different traits, might be tested with multivariate likelihoods on limited regions.

Particular attention should be paid to the p-value calculation. Tolerant levels should be used in steps 1 and 2 for preliminary selections (5% chromosome wide for example) to balance the over-parameterisation of the tests. For the finest models, one should increase the rejection thresholds to be stringent enough for the final choice. In the selection procedure between steps 1 to 3, multivariate likelihoods can be used only for QTL effect estimations under the one-QTL model, if simulations or permutations are chosen for p-value computations. Thus DA strategies can save days of computation in preliminary steps of multivariate analysis at no cost concerning power and QTL location.

5. CONCLUSIONS

We propose an efficient strategy to start the fine mapping of pleiotropic chromosomal regions. The power of a discriminant analysis combined with the hugely reduced computational costs represents an original strategy and finally offers a rapid and efficient alternative to systematically define which chromosomal regions to focus on. In order to describe the finest models better, the opportunity to combine in bivariate likelihoods two linear combinations of the traits, related to two different positions for the putative QTL and describing potentially different trait covariances might be explored in further studies.

ACKNOWLEDGEMENTS

REFERENCES

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

[2] Andersson L., Haley C.S., Ellegren H., Knott S.A., Johansson M., Andersson K., Andersson-Eklund L., Edfors-Lilja I., Fredholm M., Hansson I., Håkanson J., Lundstróm K., Genetic mapping of quantitative trait loci for growth and fatness in pigs, Science 263 (1994) 1771–1774.

[3] Bidanel J.P., Milan D., Iannuccelli N., Amigues Y., Boscher M.Y., Bourgeois F., Caritez J.C., Gruand J., Le Roy P., Lagant H., Quintanilla R., Renard C., Gellin J., Ollivier L., Chevalet C., Detection of quantitative trait loci for growth and fatness in pigs, Genet. Sel. Evol. 33 (2001) 289–309.

[4] Carlborg O., Andersson L., Kinghorn B., The use of a genetic algorithm for simultaneous mapping of multiple interacting quantitative trait loci, Genetics 155 (2000) 2003–2010.

[5] Elsen J.M., Mangin B., Goffinet B., Boichard D., Le Roy P., Alternative models for QTL detection in livestock. I. General introduction, Genet. Sel. Evol. 31 (1999) 213–224.

[6] Gilbert H., Le Roy P., Comparison of three multitrait methods for QTL detection, Genet. Sel. Evol. 35 (2003) 281–304.

[7] Gilbert H., Le Roy P., Power of three multitrait methods for QTL detection in crossbred populations, Genet. Sel. Evol. 36 (2004) 347–361.

[8] Gilbert H., Le Roy P., Milan D., Bidanel J.P., Linked and pleiotropic QTL influencing carcass composition traits detected on the porcine chromosome 7, Genet. Res. (2007), to appear.

[9] Goffinet B., Mangin B., Comparing methods to detect more than one QTL on a chromosome, Theor. Appl. Genet. 96 (1998) 628–633.

[10] Goffinet B., Le Roy P., Boichard D., Elsen J.M., Mangin B., Alternative models for QTL detection in livestock. III. Heteroskedastic model and models corresponding to several distributions of the QTL effects, Genet. Sel. Evol. 31 (1999) 341–350.

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

[12] Haley C.S., Knott S.A., Elsen J.M., Mapping quantitative trait loci in crosses between outbred lines using least squares, Genetics 136 (1994) 1195–1207.

[13] Jansen R.C., Stam P., High resolution of quantitative traits into multiple loci via interval mapping, Genetics 136 (1994) 1447–1455.

[14] Jiang C., Zeng Z.B., Multiple trait analysis of genetic mapping for quantitative trait loci, Genetics 140 (1995) 1111–1127.

[15] Knapp S., Bridges W., Birked D., Mapping quantitative trait loci using molecular marker linkage maps, Theor. Appl. Gen. 79 (1990) 583–592.

[16] Knott S.A., Haley C.S., Maximum likelihood mapping of quantitative trait loci using full-sib families, Genetics 132 (1992) 1211–1222.

[17] Knott S.A., Haley C.S., Multiple least squares for quantitative trait loci detection, Genetics 156 (2000) 899–911.

[18] de Koning D.J., Schulmant N.F., Elo K., Moisio S., Kinos R., Vilkki I., Maki-Tanila A., Mapping of multiple quantitative trait loci by simple regression in half-sib designs, J. Anim. Sci. 79 (2001) 616–622.

[19] Korol A.B., Ronin Y.I., Kirzhner V.M., Interval mapping of quantitative trait loci employing correlated trait complexes, Genetics 140 (1995) 273–284.

[20] Korol A.B., Ronin Y.I., Nevo E., Hayes P.M., Multi-interval mapping of correlated trait complexes, Heredity 80 (1998) 273–284.

[21] Korol A.B., Ronin Y.I., Itskovich A.M., Peng J., Nevo E., Enhanced efficiency of quantitative trait loci mapping analysis based on multivariate complexes of quantitative traits, Genetics 175 (2001) 1789–1803.

[22] Kucerova J., Lund M.S., Sorensen P., Sahana G., Guldbrandtsen B., Nielsen V.H., Thomsen B., Bendixen C., Multitrait quantitative trait loci mapping for milk production traits in danish Holstein cattle, J. Dairy Sci. 89 (2006) 2245–2256.

[23] Lander E.S., Botstein D., Mapping mendelian factors underlying quantitative traits using RFLP linkage maps, Genetics 121 (1989) 185–199.

[24] Lebreton C.H., Visscher P.M., Haley C.S., Semikhodskii A., Quarrie S.A., A nonparametric bootstrap method for testing close linkage *vs.* pleiotropy of coincident quantitative trait loci, Genetics 150 (1998) 931–943.

[25] Le Roy P., Elsen J.M., Boichard D., Mangin M., Bidanel J.P., Goffinet B., An algorithm for QTL detection in mixture of full and half sib families, in: 6th World Congress of Genetic Applied to Livestock Production, 11–16 January 1998, Vol. 26, University of New England, Armidale, pp. 257–260.

[26] Lund M.S., Sorensen P., Guldbrandtsen B., Sorensen D.A., Multitrait fine mapping of quantitative trait loci using combined linkage disequilibria and linkage analysis, Genetics 136 (2003) 405–410.

[27] Mangin B., Thoquet P., Grimsley N.H., Pleiotropic QTL analysis, Biometrics 54 (1998) 88–99.

[28] Mangin B., Goffinet B., Le Roy P., Boichard D., Elsen J.M., Alternative models for QTL detection in livestock. II. Likelihood approximations and sire marker genotype estimations, Genet. Sel. Evol. 31 (1999) 225–237.

[29] Mardia K.V., Kent J.T., Bibby J.M., Discriminate Analysis, in: Multivariate Analysis, Academic Press, London, 1979, pp. 300–332.

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

[31] Nakamichi R., Ukai Y., Kishino H., Detection of closely linked multiple quantitative trait loci using a genetic algorithm, Genetics 158 (2001) 463–475.

[32] Ronin Y.I., Kirzhner V.M., Korol A.B., Linkage between loci of quantitative traits and marker loci: multitrait analysis with a single marker, Theor. Appl. Genet. 90 (1995) 776–786.

[33] Ronin Y.I., Korol A.B., Nevo E., Single- and multiple-trait mapping analysis of linked quantitative trait loci. Some asymptotic analytical approximations, Genetics 151 (1999) 387–396.

[34] Weller J.I., Wiggans G.R., VanRaden P.M., Ron M., Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multi-trait experiment, Theor. Appl. Genet. 92 (1996) 998–1002.

[35] Zeng Z.B., Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci, Proc. Natl. Acad. Sci. USA 90 (1993) 10972–10976.

[36] Zeng Z.B., Precision mapping of quantitative trait loci, Genetics 136 (1994) 1457–1468.