Université catholique de Louvain
Faculté d'ingénierie biologique, agronomique et environnementale

Methodological aspects of the mapping of disease resistance loci in livestock

Pierre Tilquin

September 2003

Thèse présentée en vue de l'obtention
du grade de docteur en sciences agronomiques
et ingénierie biologique

Promoteurs : P. Baret et E. Le Boulengé

Part II

Methodological aspects

Chapter 7

Genetic resistance to nematodes parasites in sheep: use of repeated measurements in QTL mapping[6]

This third methodological chapter tackles the problem of using repeated measurements in the context of resistance to parasitic diseases.

The method used to simulate two correlated measurements is first presented [7.2.3 to 7.2.5]. Then different simple approaches are proposed to combine two faecal egg counts (FEC) in a QTL mapping analysis.

The power obtained when using either approach is then estimated by means of simulations [7.3].

The link between the heritability of a trait and the expected power of QTL detection is then discussed [7.4].

---

[6] Renaville B., Gruner L., Scala A., Baret P.V. & Tilquin P. (2003) Genetic resistance to nematode parasites in sheep: use of repeated measurements in QTL mapping. *In preparation*.

Summary

Most of methods for the detection of disease resistance loci (DRL) use only one resistance measurement, even if this resistance is often estimated by repeated measurements. Such a feature is especially encountered in the context of parasitic diseases for which multiple FEC (faecal egg counts) are carried out. To accommodate this constraint, we studied the interest of using aggregation methods – arithmetic or weighted means of two FEC – for the detection of QTL conferring resistance to parasites in sheep. Two correlated FEC influenced by the genotype were simulated. These two FEC were then combined into four different aggregates: (1) an arithmetic mean, (2) an arithmetic mean of the reduced FEC, (3) an environmentally weighted mean and (4) an heritability weighted mean. The aggregated FEC were analysed with two QTL detection methods: (1) a method based on least-squares, (2) a non-parametric test. Depending on experimental conditions, the gain of statistical power reaches 12 % when an aggregate FEC is used, compared to the use of a single FEC. The use of an aggregated trait instead of a single measurement would therefore increase the chance to detect the QTL and would also improve the accuracy of the estimated QTL position. This new technique only requires one additional FEC for which the cost is relatively low compared to the overall experimental cost (animal housing and genotyping).

7.1. Introduction

Nematode infection of sheep results in economic losses, either through a reduced production or through the cost of anthelminthics. Since differences are observed between individuals, in terms of infection level, it is of interest to select more resistant animals. Indeed, selected animals will show a lower production decrease and will need less anthelminthics. A criterion frequently used by parasitologists to quantify the resistance is to perform faecal egg counts (FEC).

As shown by numerous studies, differences in FEC values between animals are explained by a genetic component. Estimations of FEC heritability are ranging from 0.20 to 0.35 (Douch *et al.*, 1995; Bouix *et al.*, 1998; Eady *et al.*, 1998; Bishop & Stear, 1999). The classic selection procedure implies the selection of animals presenting low FEC values. However, this selection requires the infection of animals and the further assessment of FEC levels for each animal.

Another approach to select resistant animals is based on the identification of genes associated with the resistance to nematodes and on the selection at the molecular level of resistant animals (marker assisted selection, MAS). Up to now, five genome scans for quantitative trait loci (QTL) affecting resistance to nematodes were carried out (Table 7.1). Other studies (n = 8) focused on candidate genes (Blattman *et al.*, 1993; Hulme *et al.*, 1993; Schwaiger *et al.*, 1995; Stear *et al.*, 1996; Buitkamp *et al.*, 1996; Paterson *et al.*, 1998; Paterson *et al.*, 2001; Coltman *et al.*, 2001a). A special feature of all these studies is that, despite that FEC were measured repeatedly, the analysis was mainly run on a single FEC. In 4 out of 13 studies, the analysis was run on the arithmetic mean of multiple FEC from each challenge and in two other studies the arithmetic mean of all the FEC performed during the experiment was used.

Table 7.1: QTL scans for nematode resistance in sheep.

Reference	Design			QTL analysis			Results	
	Number of challenge	Number of FEC /challenge	Drenching	Variable[a]	MT[b]	Method[c]	Number of QTL	Chr.[d]
Beh *et al.* (2002)	2	2 / 3	yes	CM	CR	ML	6	1, 3, 6, 11, 12
Crawford *et al.* (1997)	2	3	no	CM	*log*	LS	1	1
Diez-Tascon *et al.* (2002)	2	1	no	S	*log*	LS	1	1
Janßen *et al.* (2002)	1	2	no	S	*log*	LS	1	20
Raadsma *et al.* (2002)	2	4	yes	CM, TM	CR	LS	1	--

[a] S = each FEC separately, CM = arithmetic mean of the FEC of one challenge, TM = the arithmetic mean of all the FEC from all challenges

[b] Mathematical transformation of phenotypic values : *log* = logarithmic, CR = cube root

[c] Interval mapping method: ML = maximum likelihood, LS = least squares

[d] Chromosomes on which the identified QTL are located

Even if QTL mapping methods for multiple trait are available (Jiang & Zeng, 1995; Weller *et al.*, 1996; Knott & Haley, 2000; Korol *et al.*, 2001), the application of these methods is not yet common practice. They are theoretically sophisticated and can present computational difficulties. Furthermore, to our knowledge, there is no straightforward way of handling repeated measurements of the same trait in a QTL mapping context. In other contexts, some studies analysed the different measurements separately (Moharramipour *et al.*, 1997). Other studies on cattle lactation used an index of test-day records (Ron *et al.*, 2001) or a model of the lactation curve (Rodriguez-Zas *et al.*, 2002). In the mouse, Jackson *et al.* (1999) specified a mixed model with an unstructured covariance matrix.

Instead, using single-trait QTL mapping methods that are currently available, we propose to combine multiple measurements into a unique parameter (an aggregate) summarising the information from the repeated measurements. Such an approach should increase the power of QTL detection as higher heritability estimates were obtained when using means of several FEC (Stear *et al.*, 1996; Bishop *et al.*, 1996; Baker, 1997; Stear *et al.*, 1999).

In this study, we compare four aggregates of two FEC: (1) the arithmetic mean of the raw data (M), (2) the arithmetic mean of the reduced data (RM) (3) an environmentally weighted mean of the reduced data (EWM) and finally, (4) an heritability weighted mean of the reduced data (HWM).

The aim is to quantify the advantage of using one or another aggregate of two FEC compared to using a single FEC. The comparison will be carried out in terms of statistical power and in terms of the accuracy of the estimated QTL location using two interval mapping methods (1) one based on least squares (LS) and, (2) another based an a non-parametric test (NP).

7.2. Materials and methods

7.2.1. Methodology

Simulations of FEC distributions are performed in order to mimic a real data set of FEC. As the theoretical distribution function of FEC is unknown,

normally distributed phenotypic values are first simulated and are then transformed into FEC using a transformation based on quantiles (*normal score back transformation*, see below).

In order to mimic the real conditions, two features of FEC are to be taken into account: firstly, the correlation between FEC measurements and secondly, the fact that FEC exhibit an increasing heritability during the grazing season. Two correlated FEC are simulated, with heritabilities of the first and the second FEC equal to 0.20 and 0.30 respectively.

7.2.2. Real data set

Distributions of FEC were obtained from a study performed on 960 ewes (unpublished results). Ewes were progenies from a backcross experiment between 10 rams F1 (Lacaune ram x Sarda ewe) and Sarda ewes. Ewes were born in the spring 1999 and were housed at the Institut de zootechnie (IZCS), Monastir, Sardinia. During the experiment, ewes lambed twice: at the beginning of the year 2000 and during a period from December 2000 to January 2001. The moments of FEC were chosen, on the one hand, in order to obtain sufficiently high values and, on the other hand, to be far apart from the parturient period. The FEC were done, with the McMaster technique, in a four to five days period in June and in October 2000.

Proportions of zero counts in FEC of June and October are 16 % and 3 % respectively. Since FEC distributions are right-skewed with a long tail, a mathematical transformation was applied:

$$\log_{10}\left[\left(FEC+7.5\right)/\log_{10}\left(FEC+7.5\right)\right] \qquad (7.1)$$

This transformation is an adaptation of the transformation proposed by Bosseray and Plommet (1976) where the factor 7.5 is half of the dilution factor of the McMaster technique used to measure FEC. As the heritability estimates could not be obtained from the data, the repeatability (0.30, data not shown) was used as a good estimate of heritability.

7.2.3. Simulation design

All the simulations are based on the same design: 30 families of 40 half-sibs. Each sire is randomly mated to 40 unrelated dams and the trait is measured twice on a single offspring per mating (FEC1, FEC2), which amounts to 1200 measured individuals and 2400 measurements. Both trait measurements are first simulated as normally distributed and are then transformed into FEC ($h_1^2 = 0.20$, $h_2^2 = 0.30$; see below). Simulations are repeated 1000 times.

For each offspring, a 100 cM chromosome segment carrying eleven markers evenly spaced (interval 10 cM) is simulated. A single QTL is simulated and positioned at 35 cM on the 100 cM chromosome segment (between the fourth and the fifth marker) (see Chapter 5). In all simulations, the number of alleles at the markers is equal to 16, occurring with equal frequency. Such a number of alleles was chosen to mimic a fully informative situation. Furthermore, the dam allele is specifically coded to be always identifiable. The number of alleles at the QTL is equal to 2 with equal frequency (0.5). The same QTL is acting on both measurements, but due to difference in heritability, QTL explains 8 % of the variance of FEC1 and 12 % of the variance of FEC2.

7.2.4. Simulation of the first measurement

The simulation process is based on an algorithm developed by Baret *et al.* (1998). Briefly, normally distributed phenotypic values for the first measurement (all terms in equations are labelled with a 1 as they belong to the first measurement) are simulated as the sum of a genetic component and an environmental component: $P_1 = G_1 + E_1$. In this equation, the genetic component (G_1) includes QTL and polygenic effects and the environmental noise (E_1) is simulated by a normal distribution $N\left(0,\left[\left(1-h_1^2\right)\sigma_p^2\right]\right)$ (see Chapter 5).

7.2.5. Simulation of a second correlated measurement

A second phenotype measurement is simulated as correlated to the first measurement. The simulation of a correlated distribution was based on the theory of bivariate normal distributions. The formula described in Ronin *et al.* (1995) was used.

The second phenotypic value (P_2) is also simulated as the sum of a genetic component and an environmental component $P_2 = G_2 + E_2$ and with the same total phenotypic variance as the first phenotypic value. These two components were simulated on the basis of the genetic and environment components of the first measurement.

The choice of the correlation coefficient between the two FEC was based on practice. As mentioned in the literature (Bishop *et al.*, 1996; Stear *et al.*, 1999; Woolaston & Windon, 2001), the genetic correlation between two repetitions of FEC is not significantly different from 1. This value was thus chosen for our simulations (Table 7.2). The genetic value at the second measurement is simulated as:

$$G_2 - m_{G2} = \frac{\sigma_{G2}}{S_{G1}}\left(G_1 - \overline{G}_1\right) \qquad (7.2)$$

where G_1 and G_2 are the genetic values of the first and the second measurement respectively, $\overline{G}_1$ and S_{G1} are the observed mean and standard deviation of the genetic values at the first measurement, m_{G2} and σ_{G2} are the theoretical mean (equal to 0) and standard deviation of the genetic values at the second measurement. This equation amounts to a scaling of the genetic effect of the first measurement to the scale of the second correlated measurement, since both measurements have different heritabilities. As the genetic effect is the sum of QTL and polygenic effects, this scaling is applied to both component. The heritability of the QTL at the second measurement increases thus in the same proportion as the heritability ($h^2_{QTL_2}$ = 12 %).

The environmental value for the second measurement is simulated as:

$$E_2 - m_{E2} = R_E \frac{\sigma_{E2}}{S_{E1}}\left(E_1 - \overline{E_1}\right) + \sqrt{1 - R_E^2}\,\sigma_{E2} z_E \qquad (7.3)$$

where E_1 and E_2 are the environmental values at the first and second measurements respectively, $\overline{E}_1$ and S_{E1} are the observed mean and standard deviation of the environmental values at the first measurement, m_{E2} and σ_{E2} are the theoretical mean (equal to 0) and standard deviation (equal to $\sigma_{E2} = \sqrt{(1 - h_2^2)\sigma_p^2}$) of the environmental values at the second measurement, R_E is the correlation coefficient between the two environmental values (E_1 and E_2). Four values of correlation coefficients were chosen 0.0, 0.3, 0.6 and 0.9 – these values correspond to phenotypic correlation of 0.24, 0.47, 0.69 and 0.92 respectively – and z_E is a random number drawn in a normal distribution $N(0,1)$.

7.2.6. Transformation of normal phenotypes in FEC

Since FEC distributions cannot be simulated with a known function, an approach used in geostatistics and referred to as *normal score back transformation* is applied. The rationale of this transformation is explained in Tilquin *et al.* (2001) (a thorough explanation over the *normal score back transformation* is given in Appendix B). It can be seen as a correspondence table between the quantiles of two different distributions, a simulated normally distributed phenotype and an observed FEC distribution. The simulation of two repeated and correlated FEC is made by applying this transformation to all the simulated values of the two normal correlated phenotypes. In order to avoid interpretation problems, only one distribution (October, 3 % of zeros) was used to mimic FEC.

7.2.7. Aggregates of two FEC

Four different aggregates of the two FEC were built, the arithmetic mean of the raw data (M), the arithmetic mean of the reduced data (RM), i.e. the arithmetic mean of the two FEC standardised so that they have a mean of

zero and a standard deviation of one, and two weighted means of the reduced data $\omega_1 FEC_1 + \omega_2 FEC_2$, where FEC_1 and FEC_2 are the value of FEC after mathematical transformation and standardisation, and ω_1 and ω_2 are the weights attributed to FEC_1 and FEC_2 respectively. For the first type of weighted mean (EWM), the weights are attributed in order to minimise the environmental variance of the mean, these weights are:

$$\omega_1 = \hat{\sigma}^2_{E_2} \Big/ \left[\hat{\sigma}^2_{E_1} + \hat{\sigma}^2_{E_2} \right] \tag{7.4}$$

$$\omega_2 = \hat{\sigma}^2_{E_1} \Big/ \left[\hat{\sigma}^2_{E_1} + \hat{\sigma}^2_{E_2} \right] \tag{7.5}$$

Environmental variance can be estimated and the weights are estimated as:

$$\omega_1 = \left(1 - \hat{h}^2_2\right) \Big/ \left[\left(1 - \hat{h}^2_1\right) + \left(1 - \hat{h}^2_2\right) \right] \tag{7.6}$$

$$\omega_2 = \left(1 - \hat{h}^2_1\right) \Big/ \left[\left(1 - \hat{h}^2_1\right) + \left(1 - \hat{h}^2_2\right) \right] \tag{7.7}$$

where $\hat{h}^2_1$ and $\hat{h}^2_2$ are heritability estimates obtained for each simulation on mathematically transformed FEC (see Methods of analysis).

For the second weighted mean (HWM), the weights are attributed in order to maximise the genetic variation of the mean, these weights are:

$$\omega_1 = \hat{h}^2_1 \Big/ \left(\hat{h}^2_1 + \hat{h}^2_2 \right) \tag{7.8}$$

$$\omega_2 = \hat{h}^2_2 \Big/ \left(\hat{h}^2_1 + \hat{h}^2_2 \right) \tag{7.9}$$

Simulation parameters are summarised in Table 7.2.

Table 7.2: Parameters of simulation.

	FEC1	FEC2
QTL frequency ($p = q$)	0.5	0.5
QTL substitution effect (a)	0.4	0.5
Heritability (h^2)	0.2	0.3
QTL heritability (h^2_{QTL})	0.08	0.12
Genetic correlation (R_G)	1	
Environmental correlation (R_E)	0.0, 0.3, 0.6, 0.9	

7.2.8. Methods of analysis

Heritability estimates are obtained using the classical method based on least squares as four times the sire component in a nested analysis of variance of half-sib families (Falconer & Mackay, 1996).

Two QTL mapping methods are used: regression interval mapping (Knott *et al.*, 1996) and non-parametric interval mapping using midranks for ties (Coppieters *et al.*, 1998; Tilquin *et al.*, 2001; Tilquin *et al.*, 2003). Both are applied on mathematically transformed individual FEC or on the different aggregates of the two FEC.

7.2.9. Comparison of the different aggregates

The four different aggregates of the two FEC will be compared to the single FEC approach in terms of: (1) statistical power, (2) accuracy of the estimated QTL position.

The significance threshold is obtained by use of within family permutations (1000 permutations of each simulation) (Churchill & Doerge, 1994). The proportion of significant simulations is used as a power estimate. Power estimates are considered as different if difference is higher than $z_{1-\alpha/2}\sqrt{u(1-u)(1/n_1+1/n_2)}$, where $z_{1-\alpha/2}$ is the 1-α/2 quantile of the standard

$N(0,1)$, u is the proportion of significant runs pooled across methods; and n_1 and n_2 are the number of runs for each method. The significance level of this test was adjusted for multiple comparisons using the Bonferroni method.

The accuracy of the estimated QTL position is assessed by the estimated bias (the difference between the estimated QTL position and the simulated QTL location, i.e. 35).

It doesn't escape to our notice that, in the case of two normally distributed measurements (with $\sigma^2_{P1} = \sigma^2_{P2} = \sigma^2_P$), the heritability of their mean can be determined theoretically according to the properties of the variance (see Appendix F for demonstrations).

This heritability depends on the heritability of the two measurements and on their genetic and environmental correlations (equations 7.10, 7.11, 7.12).

$$h^2_{M=RM} = \frac{h_1^2 + h_2^2 + 2\sqrt{h_1^2 h_2^2}R_G}{2 + 2\sqrt{h_1^2}\sqrt{h_2^2}R_G + 2\sqrt{1-h_1^2}\sqrt{1-h_2^2}R_E} \quad (7.10)$$

$$h^2_{EWM} = \frac{\left(1-h_2^2\right)^2 h_1^2 + \left(1-h_1^2\right)^2 h_2^2 + 2\left(1-h_2^2\right)\left(1-h_1^2\right)\sqrt{h_1^2 h_2^2}R_G}{\left(1-h_2^2\right)^2 + \left(1-h_1^2\right)^2 + 2\left(1-h_2^2\right)\left(1-h_1^2\right)\sqrt{\left(h_1^2\right)\left(h_2^2\right)}R_G + 2\sqrt{\left(1-h_2^2\right)^3\left(1-h_1^2\right)^3}R_E} \quad (7.11)$$

$$h^2_{HWM} = \frac{\left(h_1^2\right)^3 + \left(h_2^2\right)^3 + 2\sqrt{\left(h_1^2\right)^3\left(h_2^2\right)^3}R_G}{2\sqrt{\left(h_1^2\right)^3}\sqrt{\left(h_2^2\right)^3}R_G + \left(h_1^2\right)^2 + \left(h_2^2\right)^2 + 2\sqrt{\left(h_1^2\right)^2\left(1-h_1^2\right)}\sqrt{\left(h_2^2\right)^2\left(1-h_2^2\right)}R_E} \quad (7.12)$$

By replacing in these equations, the heritability values of the first ($h_1^2 = 0.20$) and second FEC ($h_2^2 = 0.30$) and the values of the genetic correlation ($R_G = 1$) by the values used in the simulations (Table 7.2), we plotted the heritability of the aggregate phenotype as a function of the environmental correlation (Figure 7.1).

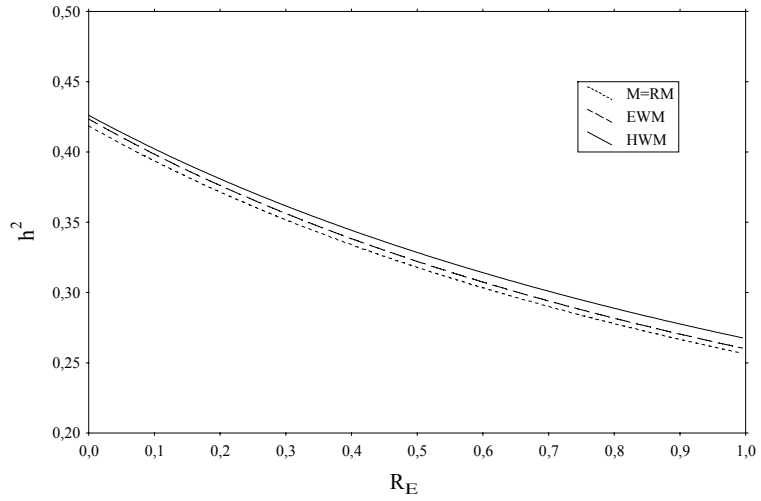

Figure 7.1: Theoretical relationship between the heritability (h^2) and the environmental correlation coefficient (R_E), for the arithmetic mean of both measurements (M) and for the heritability weighted mean (HWM).

As the statistical power depends on the heritability of the trait, it is possible to determine, prior to analysis, if it is more interesting to use an aggregated trait or only the most heritable measurement. For example, with heritability values of 0.20 and 0.30 for first and second measurements respectively, an aggregate is justified for correlation lower than, approximately, 0.70. However, all this theoretical development is based on the variance properties of the normal distribution. In the case of the disease resistance to nematode parasites, the FEC do not show a normal distribution. As the variance properties of such distribution are unknown, it isn't possible to theoretically compute the heritability of aggregated FEC. In order to evaluate the interest of aggregating multiple FEC, simulations were done.

7.3. Results

Average percentages of zeros (mean ± s.d., over the 1000 replicates) were 3.12 ± 0.59 % and 3.02 ± 0.48 % for FEC1 and FEC2, which is close to the

simulation parameter of 3 %. In all simulations, phenotypic correlation coefficients were in accordance with expected values (data not shown).

Table 7.3: Mean heritability estimates (± s.d.) of both FEC and their aggregates as a function of the environmental correlation coefficient R_E (based on 1000 replicates of each alternative).

	R_E			
Phenotype [a]	0.0	0.3	0.6	0.9
FEC1	0.20 ± 0.07	0.20 ± 0.07	0.20 ± 0.07	0.20 ± 0.07
FEC2	0.30 ± 0.08	0.30 ± 0.08	0.30 ± 0.08	0.30 ± 0.08
M	0.38 ± 0.10	0.33 ± 0.09	0.29 ± 0.08	0.26 ± 0.08
M(N)[b]	*0.42*	*0.35*	*0.30*	*0.27*
RM	0.41 ± 0.10	0.34 ± 0.09	0.29 ± 0.09	0.26 ± 0.08
RM(N) [b]	*0.42*	*0.35*	*0.30*	*0.27*
EWM	0.36 ± 0.12	0.31 ± 0.10	0.28 ± 0.09	0.25 ± 0.08
EWM(N) [b]	*0.42*	*0.36*	*0.31*	*0.27*
HWM	0.41 ± 0.10	0.35 ± 0.09	0.31 ± 0.08	0.27 ± 0.08
HWM(N) [b]	*0.43*	*0.36*	*0.31*	*0.28*

The simulated values of heritabilities were set to 0.20 and 0.30 for the first and the second FEC respectively.

[a] FEC1 and FEC2 are the first and the second simulated FEC after mathematical transformation and normal standardisation, M is the arithmetic mean of the raw data, RM is the arithmetic mean of the reduced data, EWM is the environmentally weighted mean of the reduced data and finally, HWM is the heritability weighted mean of the reduced data.

[b] Theoretical heritability values for aggregates of normally distributed phenotypes (see Figure 7.1)

Heritability estimates for the single FEC were in accordance with the simulated values, i.e. 0.20 and 0.30 (Table 7.3). For an environmental correlation of 0.0, heritability estimates of all aggregates were higher than the heritability estimates of any single FEC. Heritability estimates of these aggregates decrease as the environmental correlation coefficient (R_E) between both FEC increases. If environmental correlation is equal to 0.60 or

0.90, the heritability of aggregates is smaller than that of the second single FEC.

The greater heritability for the aggregates than for the single FEC is due to the genetic correlation between the two FEC. Indeed, the variance of a sum is greater than the sum of variances, if the terms of the sum are positively correlated. As the heritability is the ratio between genetic and phenotypic variances, an increased genetic variance contributed to increased heritability. For the same reasons, an increased environmental correlation leads to an increased phenotypic variance and may contribute to a decreased heritability (see equations 7.10, 7.11, 7.12).

Table 7.4: Power (%) at the 5 % level for the least-squares method (LS) and the non-parametric test (NP) as a function of the environmental correlation (R_E) and when analysing both FEC separately (FEC1 and FEC2) or their aggregates(M, RM, EWM and HWM).

	LS				NP			
R_E	0.0	0.3	0.6	0.9	0.0	0.3	0.6	0.9
FEC1	59 [a]	59 [a]	59 [a]	59 [a]	56 [a]	56 [a]	56 [a]	56 [a]
FEC2	84 [b]	84 [b]	84 [b]	84 [b]	80 [b]	81 [b]	81 [b]	82 [b]
M	94 [c]	89 [c]	84 [b]	76 [b]	92 [c]	87 [c]	80 [b]	71 [b]
RM	96 [c]	90 [c]	84 [b]	75 [b]	94 [c]	88 [c]	80 [b]	72 [b]
EWM	93 [c]	87 [bc]	81 [b]	72 [b]	91 [c]	85 [bc]	78 [b]	69 [b]
HWM	96 [c]	90 [c]	85 [b]	78 [b]	94 [c]	88 [c]	82 [b]	75 [b]

Power estimates based on 1000 replicates of each alternative, using 1000 permutations to obtain the significance level of each replicate. Heritability of FEC1 and FEC2 were respectively set to 0.20 and 0.30. Heritability due to the QTL was set to 8 % and 12 % of the total phenotypic variance for FEC1 and FEC2 respectively.

[a] Comparison of statistical power between measurements (single and aggregates) within columns (*i.e.* within the same method and with the same environmental correlation coefficient); values with the same superscript are not significantly different. The significance level of this test was adjusted for multiple comparisons using the Bonferroni method.

Power estimates at the 5 % level (and at the 1 % level; data not shown) of all aggregates of both FEC are significantly higher than the power observed when analysing the first FEC (Table 7.4). If the correlation between both FEC is small, power estimates of all aggregates are also significantly higher than the power observed when analysing the second FEC. This gain of power is progressively lost when the correlation coefficient increases. This phenomenon is observed either with the least-squares method (LS) or with the non-parametric test (NP). Power of NP is slightly smaller than the power of LS.

For all QTL mapping methods and for all phenotypes, the estimated QTL location tends to be biased towards the centre of the chromosome[7] (Table 7.5). However, the bias of the estimated QTL location is smaller for every aggregates of FEC than for any single FEC.

For example, when there is no environmental correlation between both FEC, the average bias amounts to the value of 0.6 cM for all aggregates compared to 2.6 cM and 1.4 cM for FEC1 and FEC2 respectively; this difference is only significant when compared to FEC1 (see Table 7.5). Bias is not significantly different between methods of aggregation.

For all four aggregates, the mean value of test statistics along the chromosome was higher than using one of the two single FEC, especially at the position of the simulated QTL (Figure 7.2). Between aggregates, the heritability weighted mean gave the highest peak.

[7] See Appendix C for an explanation about the apparent contradiction between the mean bias of the estimated QTL location (Table 7.5) and mean test statistic curves (Figure 7.2).

Table 7.5: Mean bias (± s.e.) of the estimated QTL location with the least-squares method (LS) or the non-parametric test (NP) for both single FEC and their aggregates.

	LS				NP			
R_E	0.0	0.3	0.6	0.9	0.0	0.3	0.6	0.9
FEC1	2.60 ± 0.59 [a]	2.60 ± 0.59 [a]	2.60 ± 0.59 [a]	2.60 ± 0.59 [a]	3.52 ± 0.61 [a]	3.52 ± 0.61 [a]	3.52 ± 0.61 [a]	3.52 ± 0.61 [a]
FEC2	1.36 ± 0.41 [ab]	1.49 ± 0.42 [a,b]	1.72 ± 0.42 [a]	1.39 ± 0.42 [a]	1.48 ± 0.42 [b]	1.18 ± 0.43 [b]	1.53 ± 0.43 [a,b]	1.63 ± 0.44 [a]
M	0.58 ± 0.29 [b]	0.94 ± 0.36 [a,b]	1.82 ± 0.44 [a]	2.17 ± 0.48 [a]	1.06 ± 0.33 [b]	1.27 ± 0.39 [b]	1.79 ± 0.44 [a,b]	2.13 ± 0.49 [a]
RM	0.59 ± 0.28 [b]	0.73 ± 0.34 [a,b]	1.53 ± 0.43 [a]	2.00 ± 0.47 [a]	0.58 ± 0.27 [b]	0.85 ± 0.35 [b]	1.95 ± 0.45 [a,b]	2.30 ± 0.50 [a]
EWM	0.63 ± 0.30 [b]	0.66 ± 0.36 [b]	1.74 ± 0.45 [a]	2.13 ± 0.51 [a]	0.96 ± 0.32 [b]	1.05 ± 0.38 [b]	2.12 ± 0.46 [b]	2.28 ± 0.52 [a]
HWM	0.66 ± 0.28 [b]	1.05 ± 0.35 [b]	1.43 ± 0.41 [a]	1.69 ± 0.46 [a]	0.73 ± 0.29 [b]	0.77 ± 0.35 [b]	1.42 ± 0.42 [b]	2.28 ± 0.47 [a]

See Table 7.4

[a] Comparison of mean estimates of QTL location between phenotype within columns, *i.e.* with the same environmental correlation coefficient (Bonferroni *t* tests); values with the same superscript are not significantly different; $P > 0.05$.

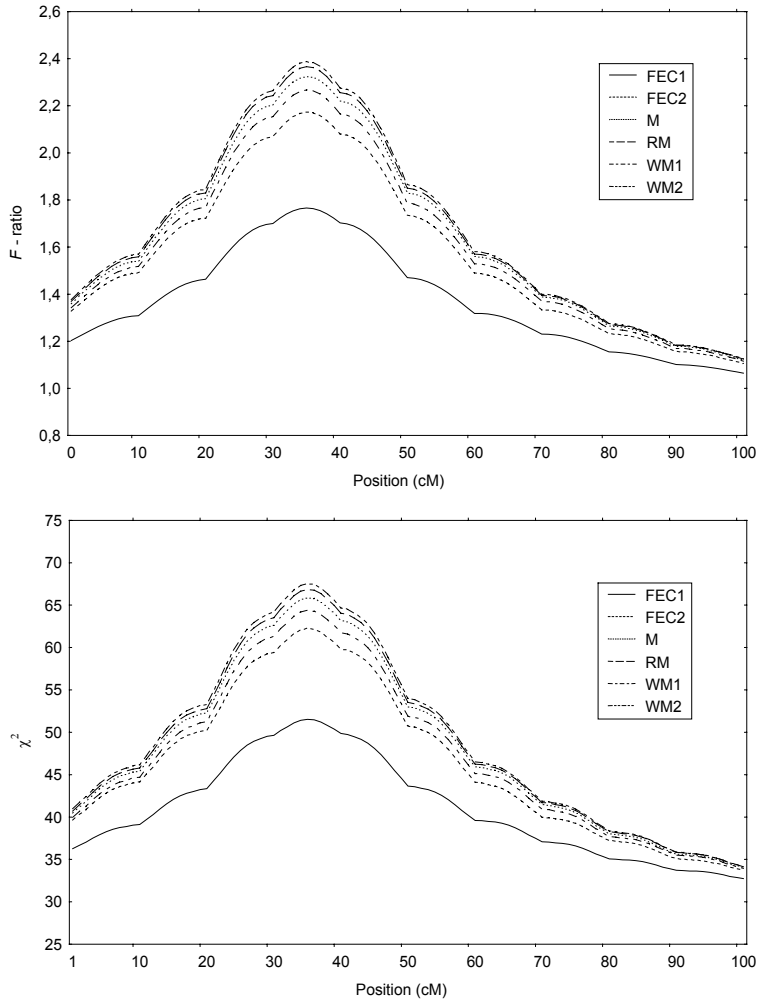

Figure 7.2: Mean test statistic values (1000 replicates) along the chromosome for the least-squares (LS) and the non-parametric (NP) QTL mapping method when single FEC and their aggregates are analysed (the environmental correlation coefficient between both FEC was 0.3).

7.4. Discussion

This study focused on the way to perform QTL analysis when repeated measurements of disease resistance phenotypes, such as FEC, are carried out. As compared to QTL mapping studies on production traits in which multiple traits are measured, e.g. Georges *et al.* (1995), QTL studies on the resistance of livestock animals to nematodes parasites use repeated measurements of the same trait, i.e. faecal egg counts (Crawford *et al.*, 1997; Beh *et al.*, 2002). This indicator trait is repeatedly measured to reduce the environmental noise and have a more accurate estimate of the disease resistance genetic value of animals (Gasbarre *et al.*, 1996).

Although there are QTL mapping methods designed for the combined analysis of multiple traits (Jiang & Zeng, 1995; Knott & Haley, 2000; Korol *et al.*, 2001), the way to analyse repeated measurements of the same trait is not straightforward. One can either assume that the same QTL controls identically the phenotypic expression of each measurement, or assume that different QTL are acting on the different measurements. The latter assumption would justify the use of multiple traits methods. However, in the context of FEC, the former assumption is supported by the fact that genetic correlation between FEC in lambs older than 3 months were not significantly lower than 1.0, indicating that faecal egg counts at different ages are expressions of the same trait (Bishop *et al.*, 1996) and can therefore be assumed to be governed by the same genes.

We therefore proposed to use simple aggregates of multiple FEC, such as the arithmetic mean or an heritability weighted mean and to use single-trait QTL mapping methods that are currently available to simplify the analysis. The principle of this approach is similar to using breeding values as dependent variable in QTL scans, and is somewhat analogue to the principle of canonical transformations which reduce the trait space (Weller *et al.*, 1996; Mangin *et al.*, 1998). Evidently, by averaging repeated measurements, the residual (environmental) variance is reduced and the heritability of the resulting trait is increased.

QTL effect influences both heritability of the trait and the power of QTL detection. In consequence, heritability and power of QTL detection are positively correlated, as observed in numerous simulation studies, e.g. Korol *et al.* (1995) and Nagamine & Haley (2001). Based on a normality assumption, we have shown that the heritability of the aggregates proposed in this study is easily obtained from simple formulas (see Appendix F) that are function of the environmental correlation between both measurements (equation 7.10, 7.11, 7.12; Table 7.3; Figure 7.1). If the heritability of aggregates is higher than the one of single measurements, the power of QTL detection is expected to increase. However, the link between heritability and power is not obvious to describe (as shown in Figure 7.3). In a single-marker test context, deterministic formulas were proposed in the case of a normally distributed trait (Weller *et al.*, 1990).

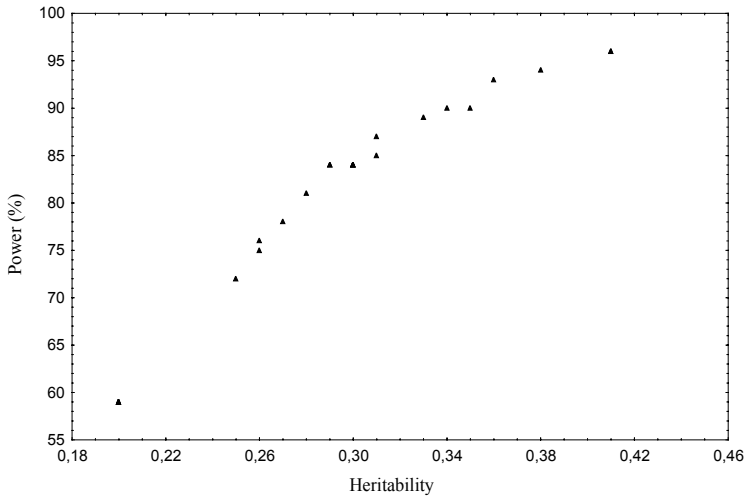

Figure 7.3: Scatter plot of heritability and power estimates obtained under the 16 alternatives of aggregates (R_E x aggregate) and for the two single FEC when using the least squares-based interval mapping method.

The originality of the present study resides in the use of repeated FEC measurements that are non-normally distributed. A simulation-based approach was therefore necessary to evaluate the power of using either a

single FEC or an aggregate of two FEC in the QTL scan. As heritability formulas derived for aggregates hold only for normally distributed phenotypes (equations 7.10, 7.11, 7.12), the heritability of aggregate traits was estimated.

The heritability of every aggregate of both FEC is higher than the one of the single FEC if the environmental correlation is small. This higher heritability of an arithmetic mean of FEC is also mentioned in the literature. For example, Stear *et al.* (1997a) obtained, for two FEC, heritability estimates of 0.12 and 0.22, and a value of 0.33 for the mean of both FEC. In an experiment including six FEC, Gruner *et al.* (1998) obtained heritability estimates of 0.33 for single FEC and of 0.55 for the mean of these FEC. As for normal measurements (see Figure 7.2), the heritability of aggregate FEC decreases as the environmental correlation increases (Table 7.3).

An extra information from the study is the comparison between different methods of aggregation (M, RM, EWM an HWM). All aggregates give a higher power of QTL detection compared to the power obtained when using a single FEC. All other things being equal, the aggregate offering the greatest increase in power is the heritability weighted mean (HWM).

This study has shown that the statistical power obtained when analysing resistance to nematode parasites is higher if aggregates of FEC are used. Furthermore, the power of QTL detection increases as the environmental correlation between FEC decreases. If the correlation is higher or equal to approximately 0.70 (under our conditions, $h_1^2 = 0.20$, $h_2^2 = 0.30$), a higher power would be obtained by analysing the most heritable FEC. However, the values of environmental correlation mentioned in the literature are often smaller than 0.70 (Shaw *et al.*, 1999). Moreover, the environmental correlation is reduced if FEC are spaced in time and if separated by a drenching (e.g. Gruner *et al.*, 1992; Bouix *et al.*, 1998).

As far as the model of infection allows for it, *i.e.* if the immune system of animals does not memorise previous infections, in QTL mapping experiments aimed at locating genes for resistance to nematode parasites, practitioners should be encouraged to use repeated FEC and aggregate them

to increase the power of QTL detection. To guarantee a small environmental correlation between FEC, a drenching should be favoured between each FEC, as far as FEC are performed sufficiently far away from this treatment. As different QTL might affect the disease resistance of lambs and ewes, aggregating repeated FEC should be restricted to FEC coming from animals at the same developmental stage.

7.5. Conclusion

The identification of QTL involved in helminth resistance would allow the substitution of a sporadic approach based on medicines by a long-term solution founded on selection of animals with enhanced resistance. Our simple approach, by increasing the opportunities to detect a QTL, should provide a tool to practitioners to facilitate the decomposition of traits of interest into their individual genetic factors. Our conclusions could be extended to other disease resistance phenotypes for which multiple measurements are usual and if the genetic basis of each measurement is the same.

7.6. Acknowledgements

We are grateful to Patrick Bogaert for helpful suggestions. Financial support is acknowledged from the Université catholique de Louvain and from the European Union (FAIR CT-98-4311 and QLRT-99-30656).