

Random regression models for detection of gene by environment interaction

Marie LILLEHAMMER*, Jørgen ØDEGÅRD, Theo H.E. MEUWISSEN

Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences,
1432 Ås, Norway

(Received 25 January 2006; accepted 6 November 2006)

Abstract – Two random regression models, where the effect of a putative QTL was regressed on an environmental gradient, are described. The first model estimates the correlation between intercept and slope of the random regression, while the other model restricts this correlation to 1 or -1 , which is expected under a bi-allelic QTL model. The random regression models were compared to a model assuming no gene by environment interactions. The comparison was done with regards to the models ability to detect QTL, to position them accurately and to detect possible QTL by environment interactions. A simulation study based on a granddaughter design was conducted, and QTL were assumed, either by assigning an effect independent of the environment or as a linear function of a simulated environmental gradient. It was concluded that the random regression models were suitable for detection of QTL effects, in the presence and absence of interactions with environmental gradients. Fixing the correlation between intercept and slope of the random regression had a positive effect on power when the QTL effects re-ranked between environments.

gene by environment interaction / QTL detection / random regression / reaction norms

1. INTRODUCTION

Several studies have detected genotype by environment interactions both in dairy cattle *e.g.* [5–7, 13, 22] and in other species *e.g.* [1, 16, 19]. It has also been reported that single genes can show different effects when expressed in different environments [11]. Genotype by environment interactions can be caused either by alleles being expressed only in specific environments or by changes in gene regulation depending on the environment [20]. To be able to quantify differences in gene effects between environments, reaction norms have been used [13], which describe the effect of a genotype as a function of the environment. Usually, a linear or quadratic function is used, which can be estimated

* Corresponding author: marie.lillehammer@umb.no

from data by random regression models [4, 13]. When genotype \times environment interactions are ignored in QTL detection, genes having a substantial average effect across environments might still be detected. However, some genes, having a significant effect in some environments might have a small average effect across environments [14]. Globalisation of selective breeding, such as in dairy cattle, makes knowledge about genotype by environment interactions increasingly important [2]. Overall, selection for a trait with an interaction effect will be less efficient if the interaction effect is not taken into account [5].

The granddaughter design is most commonly used to search for QTL in dairy cattle. The data consists of marker information on grandsires and their sons and phenotypic information on their granddaughters only. The use of “daughter yield deviations” (DYD) of sires provides useful information about each sire’s average performance over the environments, in which his daughters can be found. A daughter yield deviation averages records over all environments and represents the average sire value. Information about the different environments of the daughters, needed to search for environmental interactions, cannot be directly implemented in the DYD-approach.

One way to include the environmental information is to handle the gene effects as different traits when measured in different environments. A multiple trait method can then be used to estimate specific genetic effects for each environment. This method has been used to search for genotype \times environment interaction *e.g.* [5, 7, 24] and QTL \times environment interaction [21, 23]. A drawback of using this method is the loss of information about the environmental differences within each environmental group. Furthermore, multiple trait analysis with many traits can be computationally demanding.

Reaction norm models fitted by random regression are especially useful when the environment is measured on a continuous scale and changes gradually [12], and where the QTL-effect is assumed to be a function of the environmental gradient. Furthermore, such a random regression model will have much less parameters to be estimated compared with a multi-trait model analysing groups of environments. Random regression models have also been used to detect QTL for body weight in pigs at different ages [3]. Furthermore, Lund *et al.* [14] showed that random regression could be used to detect a simulated QTL where the effect changes gradually over time during lactation. In their study, DYDs were calculated for the sires at each test-day, which made a set of 55 DYDs for each sire, during lactation. This was made possible by testing the simulated cows at 55 specific days in milk only. However, in real data,

environments are highly variable and the observations cannot be transformed to a limited set of DYDs without loss of information. To avoid loss of environmental information by averaging over environments, using single daughter records directly in the analysis is a possibility. Ideally, in an analysis of single daughter records, the daughters' QTL alleles should be fitted in the model, but these are often unknown since only sires and grandsires are genotyped. However, each daughter inherits one of her father's alleles, and there is an equal probability of inheriting the one allele compared to the other. Hence, the sire's alleles can both be fitted in the model, modelling the effect on the daughter's production of having a father with a specific allele.

If we search for QTL using reaction norm/random regression models, we are using a variance component QTL mapping approach. In variance component QTL mapping, no assumptions are made with respect to the numbers of alleles in each locus [10]. In the case of a linear reaction norm, each assumed QTL allele has an effect on the slope and on the intercept of reaction norm, and the relations between those effects are modelled through their correlation [14]. However, if the QTL has actually only two alleles, as is often presumed, the alleles fall into two categories, say the Q and q alleles. Let a_I and a_S denote the allelic effects of the intercept and the slope, respectively, if a_I and a_S of the Q-alleles are both positive, and those of the q-alleles are both negative, the correlation between the alleles in the infinite alleles model is expected to be 1 (Fig. 1a). Alternatively, if a_I and a_S have opposite signs, the correlation is expected to be -1 (Fig. 1b). Hence, Goddard [9] suggested restricting this correlation to 1 or -1 in variance component QTL mapping in order to reduce the number of parameter estimates and to increase the precision of the QTL mapping. Whether the true correlation is 1 or -1 depends only on where the y-axis is placed relative to the crossing point between the reaction norms of the two alleles (Fig. 1). A special case occurs if the y-axis is placed exactly in the crossing point. Then, the true variance in intercept is 0, and the correlation between intercept and slope is thus not defined.

The main objective of this study was to investigate the ability of random regression models to detect gene by environment interactions. Two different random regression models were used to analyse simulated data. One of the random regression models estimated the correlation between the level and the slope of the random regression, while the other fixed this correlation to 1 or -1 . The results were compared to results from analyses done with or without interaction between QTL and environment. Different scenarios were simulated to investigate the ability of the models to detect different possible gene by environment interactions.

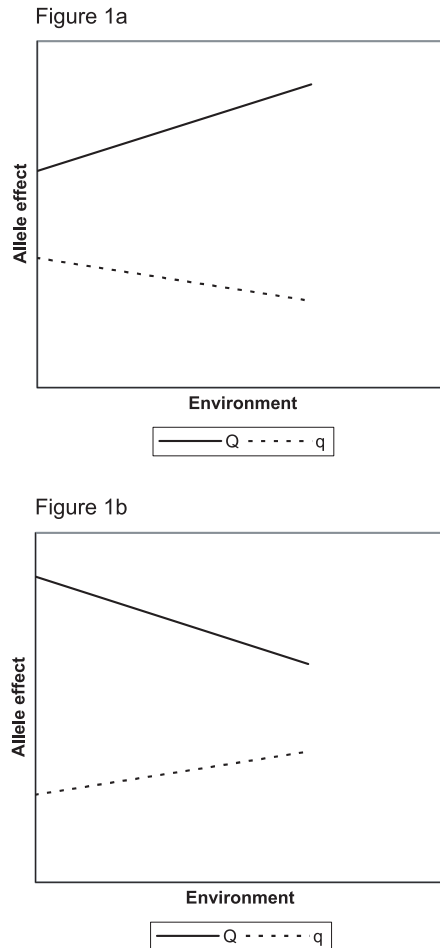


Figure 1. The allelic effects of two alleles with a pleiotropic effect on slope and intercept. The y-axes cross the x-axes in the intercepts. In Figure 1a allele Q has a positive effect on both intercept and slope, and the correlation between intercept and slope is 1. In Figure 1b, allele Q has a positive effect on intercept but a negative effect on slope. The correlation between slope and intercept is -1 . The figures show how shifting the environment where the intercept is located can change the sign of the correlation between slope and intercept.

2. MATERIALS AND METHODS

2.1. Simulations

A population was simulated for 100 generations at a constant population size of 100 (see Meuwissen and Goddard [17] for details). This created linkage

disequilibrium between the markers and the QTL. Generation 101 consisted of 100 males and 100 females, made by random mating of the animals in generation 100. Generation 102 consisted of 500 males and 50 000 females. The females were created by random mating of all animals in generation 101, while the males were created by random mating of ten males in generation 101 (referred to as selected grandsires) with all of the females from generation 101. Generation 103, consisting of 50 000 females, was created by mating each male from generation 102 with 100 females from the same generation. This gives a granddaughter design with 10 grandsires (generation 101), each grand-sire with 50 sons (generation 102) and each son with a half-sib group of 100 daughters (generation 103). The population structure was chosen to reflect a livestock population with many individuals but a low effective population size. The pedigree used for subsequent analyses included all the animals in generation 100 and 101 and the males of generation 102.

Additive genetic effects in the base population were sampled from a normal distribution with a mean of zero. In subsequent generations, additive genetic effects of each individual were calculated as the sum of the mean additive genetic effect of the parents and a random Mendelian sampling term ($N(0, 1/2\sigma_{BV}^2)$). In the simulation, additive genetic values were assumed to have an interaction with the environment. Genetic intercept and slope values were sampled for all individuals. The genetic variances in intercept and slope values were both set to 0.1, and the correlation between the two values was set to 0.5. The positive correlation gave a situation with mostly scaling effects of sire effects within the environmental range when the environment was defined to be positive [13].

The linkage map consisted of 20 multiallelic marker loci. It was assumed that the region between position 40 cM and 50 cM had been identified in an earlier QTL mapping experiment and more markers had been positioned in this region. The markers were positioned at 0, 10, 20, 30, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 90 and 100 cM. The biallelic QTL was positioned in the midpoint between marker 9 and 10, 44.5 cM from the first marker. The allele frequency of QTL allele 1 was about 0.25, but varied between replicates [17].

Each daughter (generation 103) was assigned to a simulated environmental value (Env). These values were assumed to be independently and normally distributed. The average environmental value was set to $Env = 2.50$, and the environmental standard deviation was set to 1.0. Only positive environmental values were simulated.

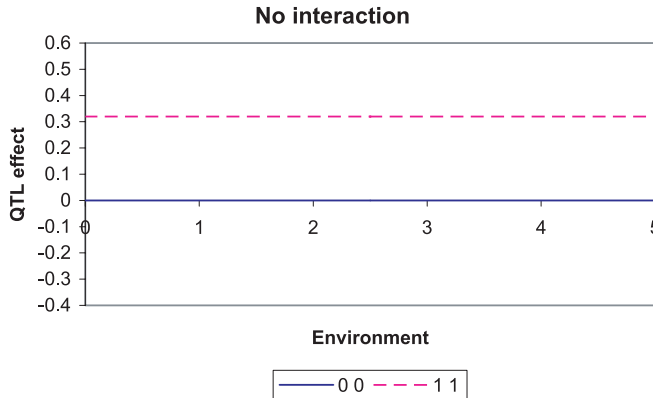


Figure 2. The effect of genotype 00 and genotype 11 of the simulated QTL is shown. In scenario “No interaction”, the QTL effect is kept constant all over the environmental range. The substitution effect is constantly 0.16, leading to a variance due to the QTL of 0.0096. The QTL contributes only to the variance in intercept.

Observations were simulated for each daughter. Each observation was the sum of the additive genetic effect (as a function of the environment), the additive QTL effect and a random deviation. The random deviations were normally distributed with mean 0 and variance 0.5. Three different kinds of QTL effects were simulated.

In the first scenario, further referred to as scenario “No interaction”, the QTL had a general effect on the trait, but did not interact with the environment (Fig. 2). The substitution effect of the QTL was 0.16, giving a variance due to the QTL of 0.01 when the QTL frequencies were 0.25 for allele 1 and 0.75 for allele 0.

In the second scenario, “Scaling”, the QTL effect was zero in the lowest environment ($Env = 0$) and increased linearly with the environmental value (Fig. 3). The average variance due to the QTL was 0.0132, given allele frequencies of 0.25 and 0.75. When splitting the variance in an intercept variance and a slope variance, the intercept variance was 0 and the slope variance was 0.0021.

In the third scenario, “Reranking”, the QTL effect was zero in the average environment ($Env = 2.50$), and the positive allele in lower environments was the negative allele in higher environments (Fig. 4). The slope values for the different genotypes were the same as in the scaling effect scenario, but the crossing point was shifted from the lowest environment to the mean environment. The average effect of the QTL is 0, because the effects on each side of the mean environment cancel each other out. When splitting the QTL variance

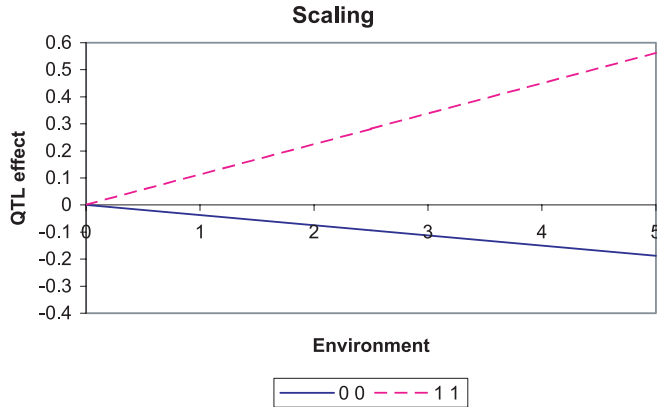


Figure 3. The effect of genotype 00 and genotype 11 of the simulated QTL is shown. In scenario “Scaling”, the QTL effect is 0 when $Env = 0$, and it increases linearly with an increase in environmental value. When $Env = 2.5$, the average environment, the variance due to the QTL is 0.0132. The variance in slope due to this QTL is 0.0021, while the variance in intercept is 0 when intercept is defined as $Env = 0$ and 0.0132 when the intercept is defined as $Env = 2.50$.

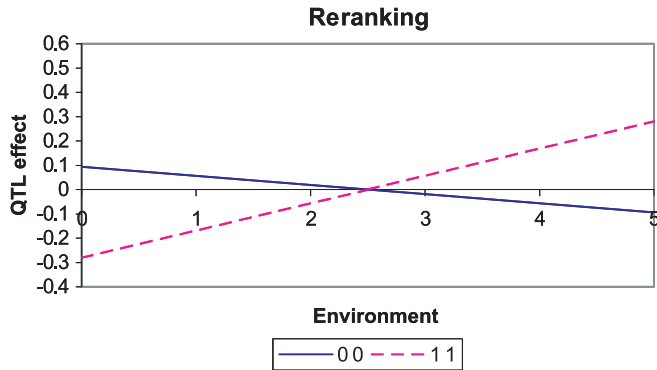


Figure 4. The effect of genotype 00 and genotype 11 of the simulated QTL is shown. In scenario “Reranking”, the QTL effect is 0 when $Env = 2.50$. Allele 1 is the negative allele when $Env < 2.50$ and the positive allele when $Env > 2.50$. The slope variance due to the QTL is 0.0021 and the intercept variance due to the QTL is 0.0132, given that the intercept is defined as $Env = 0$.

into slope and intercept variances, the slope variance is 0.0021, just as in the scaling effect scenario, while the intercept variance is 0.0132, the same as the average variance in the scaling effect scenario.

For all scenarios, a hundred replicates were simulated.

2.2. Analyses of simulated data

For the analyses, the marker-genotypes were assumed known for the sires and for the ten selected grandsires, but the linkage phases were assumed unknown. Observations and environmental information were available from the daughters only.

Combined linkage disequilibrium and linkage analysis [18] were used to join the marker and pedigree information into renumbered haplotypes and IBD-matrices for the haplotypes. The midpoint of each marker bracket was regarded as a putative position for a QTL.

The covariance associated to the IBD matrices was estimated by variance component analysis. The variance component analyses were performed using three different models. All models analysed single daughter records instead of daughter means. Furthermore, the QTL-effects were assumed to be either not affected by the environment (RR-Q) or a linear function of the environment (RR-QEe and RR-QEf).

All models had the following general characteristics:

$$y_{ij} = \mu + \sum_{k=0}^1 s_{ki} Env_j + \sum_{m=0}^p \sum_{n=1}^2 h_{mni} Env_j + e_{ij}$$

where

y_{ij} = the observation from daughter j of sire i ,

μ = the overall mean,

s_{ki} = the k^{th} order random regression coefficient of sire i on the environment,

Env_j = the environmental value of daughter j ,

h_{mni} is the m^{th} order random regression factor of the n^{th} haplotype of sire i on the environment of daughter j , and

e_{ij} is the random residual for the record y_{ij} .

Furthermore, $\mathbf{s} = \begin{bmatrix} \mathbf{s}_0 \\ \mathbf{s}_1 \end{bmatrix} \sim N(\mathbf{0}, \mathbf{A} \otimes \mathbf{S})$, where \mathbf{A} is the additive relationship matrix and \mathbf{S} is the variance-covariance matrix of sire random regression coefficients, and $\mathbf{h} \sim N(\mathbf{0}, \mathbf{G} \otimes \mathbf{H})$, where \mathbf{G} is the IBD-matrix between the haplotypes and \mathbf{H} is the variance-covariance matrix of the haplotype coefficients. Even though each daughter inherits only one of the sire's alleles (either h_{1i} or h_{2i}), both the alleles were included in the model and assumed to be different classes within the same random effect.

In the model RR-Q, the parameter p was set to 0, assuming genotype by environment interactions on the sire effects but no gene by environment interactions on the QTL effects. By setting $p = 1$, this model was extended to allow for gene by environment interactions on the QTL effects (Models RR-QEe and RR-QEf).

The correlation between the intercept and the first order regression coefficient of haplotype is estimated with RR-QEe, while when RR-QEf is applied, the correlation between the slope and intercept of the haplotype effects is fixed to be 0.99 (further referred to as RR-QEf+) or -0.99 (further referred to as RR-QEf-). Although the models assume that an infinite number of haplotypes are segregating in the population, the true (simulated) number of alleles is 2. Assuming a biallelic QTL where each allelic effect is a linear function of the environment, these linear functions will cross each other at one single point. This point might be far outside the environments the animals are exposed to, but by extrapolating the curves, such a point will always exist for two non-parallel lines. By using RR-QEf, a linear function is still estimated for each haplotype, but all the lines are forced to cross each other at the same point.

The (co)variance components estimates from random regressions will be affected by shifting the environmental scale. This was done when analysing the scenario "Scaling". In the scaling scenario, both the alleles have a value of zero at $Env = 0$. With the intercept defined in this point, the intercept variance due to the QTL is zero. Furthermore, since there is no variance in intercept, the true correlation between intercept and slope will not be defined. To obtain a variance in intercept, and a true correlation of 1 between intercept and slope, 2.50 was subtracted from every environmental value before applying model RR-QEf+ to the data simulated with a scaling effect. The environment defined as intercept was thus shifted from $Env = 0$ to $Env = 2.50$.

For all the analyses, likelihood ratio test statistics [15] were used to test the significance of the detected genes. The likelihood ratio was calculated as twice the difference between the log likelihood obtained from the different models (RR-Q, RR-QEe or RR-QEf) and the log likelihood obtained from a model omitting the QTL effect. When using RR-QEe and RR-QEf, a second likelihood ratio test was applied with RR-Q as the reduced model, to test for interaction. Threshold values were obtained from the χ^2 -table, and a nominal significance level of 1% was used. The ASREML package [8] was used for the variance component analyses.

For each scenario, the number of detected QTL was reported. The number of detected QTL was split into the number detected in the exact correct position ("Detected correct pos"), the number detected in a different bracket, but within

the dense part of the linkage group (“Det. Dense part”) and the number of QTL detected outside the dense part of the linkage group (“Det. Sparse part”). The proportion of QTL detected in or close to the correct position was used as a measurement of precision of QTL mapping.

2.3. Handling convergence problems

Convergence problems (looping) were detected in some of the random regression analyses. The results from these positions were deleted before the results were analysed. If less than 15 positions converged in one particular replicate with one of the models, the QTL was reported as not detected due to convergence problems (for this specific model and replicate).

3. RESULTS

3.1. Without interaction between QTL effect and environment

When the QTL was simulated with no interaction, it was detected in 67% of the replicates with model RR-Q, 51% of the replicates with model RR-QEe and 33% of the replicates with model RR-QEf- (Fig. 5). Applying a random regression model to data where there was no QTL by environment interaction thus resulted in lower power of detection. RR-QEe falsely detected interaction in three replicates, while RR-QEf- falsely detected interaction in two replicates. When applying model RR-Q, 64 of the detections were done between 40 and 50 cM, and 23 of the detections were done in the exact correct position (between markers 9 and 10).

3.2. QTL with scaling effect

When the QTL was simulated with a scaling effect, model RR-Q detected 19% of the QTL, while the random regression models detected 36% (RR-QEe) and 39% (RR-QEf+) (Fig. 6). When applying RR-QEf+, 20 replicates were deleted because of less than 15 marker brackets converging. RR-QEe detected the QTL in four of the replicates deleted from the RR-QEf+ analyses. There were no significant differences in the number of detected QTL or in the fraction of QTL detected in the correct position between the two random regression models. RR-QEe detected the interaction term in 31 replicates, while RR-QEf+ detected the interaction term in 35 replicates.

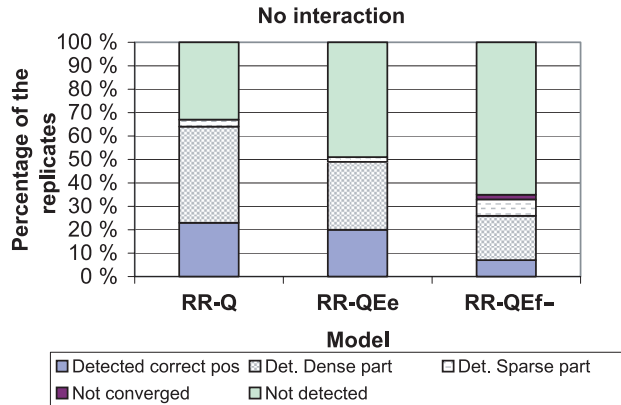


Figure 5. The power of each model is represented by the total height of the three bars denoting detection. Precision is higher when the “Det. Sparse part” is smaller and the “Detected correct pos” is higher. RR-Q had the highest power of the models in this scenario, and precision was generally high, although a little lower for RR-QEf–, which also had some convergence problems.

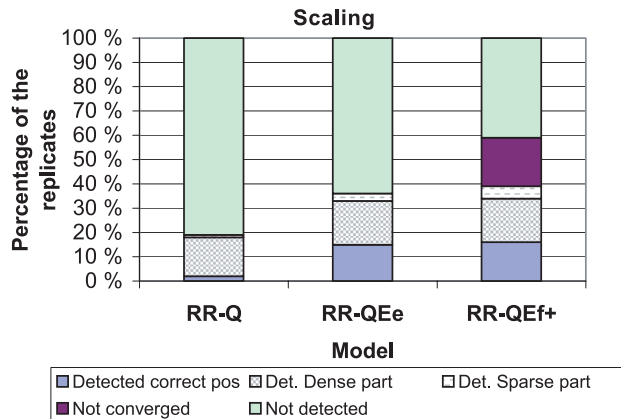


Figure 6. In the scaling scenario, the random regression models had higher power than RR-Q, and all models had quite high precision. The random regression models differed less from each other with regard to power or precision, but RR-QEf+ had a substantial proportion of brackets with convergence problems, while RR-QEe converged in all replicates.

3.3. Reranking of QTL effects between environments

In the scenario where the QTL alleles were simulated to have reranking of effects between different environments RR-Q detected 6% of the QTL, while RR-QEe and RR-QEf– detected 49% and 60% respectively (Fig. 7). Applying

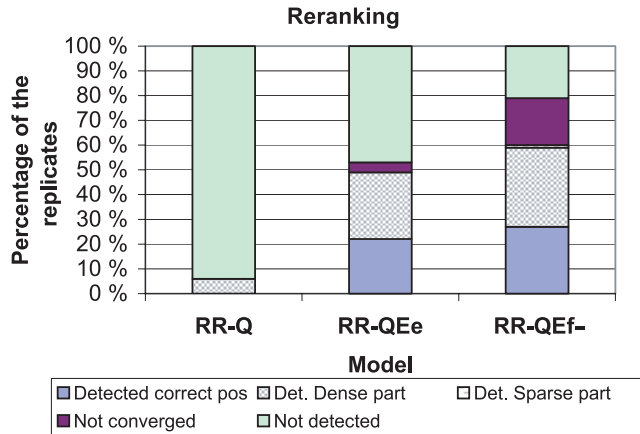


Figure 7. As seen from the total height of the bars denoting detection, the random regression models had a much higher power than RR-Q, and there were differences between the random regression models as well. RR-QEf- had a higher power and more convergence problems, while both random regression models had high precision.

RR-QEe gave four deleted replicates due to less than 15 brackets converging, while RR-QEf- gave 19 such deleted replicates. In each replicate where RR-QEe did not converge, RR-QEf- did not converge either. In every replicate where RR-QEf- gave less than 15 converged brackets, RR-QEe did not detect the QTL. There were still no differences between the two random regression models in the proportion of detected QTL in the correct bracket. RR-QEe detected the interaction term in 48 replicates, while RR-QEf- detected the interaction term in 58 replicates.

4. DISCUSSION

To be able to detect gene by environment interaction, there is a need for methods with a high power of QTL detection, but without detecting a lot of false QTL. Furthermore, the method should provide precise estimates of QTL positions, and be able to detect not only the gene, but also its interactions with the environment. All these issues are discussed below.

4.1. Power of QTL detection

The power of a test is defined to be the probability that the null hypothesis is rejected when it is indeed false [15]. Considering the null-hypothesis to be that

no QTL exists on the linkage group, power is the fraction of the 100 datasets in which a QTL is detected.

When the QTL was simulated without environmental interaction, RR-Q had the highest power (Fig. 5). This was expected, since the addition of another parameter in the model increases the threshold value, without improving the fit of the model.

When the QTL is simulated to have an interaction effect with the environment, both RR-QEe and RR-QEf give a higher power of QTL detection than RR-Q (Figs. 6 and 7). This shows that using random regression is not only a way to detect the interaction term of a gene, which is impossible with the RR-Q-model; it also increases the power of detecting such genes. The relative benefit of using random regression models, compared to RR-Q, is higher in re-ranking scenarios than in scaling-scenarios (Figs. 6 and 7). This is both due to random regression models detecting more Reranking-QTL than Scaling-QTL and due to RR-Q detecting more Scaling-QTL than Reranking-QTL.

RR-Q is in theory not able to detect the simulated Reranking-QTL. The only ways RR-Q can detect Reranking-QTL is through an uneven distribution of the haplotypes over the environmental scale, so that some haplotypes get more information from one side of the mean than from the other side, or through random differences in residual values leading to a “false discovery”. RR-QEe might detect Reranking-QTL easier than Scaling-QTL because the Reranking-QTL have both intercept and slope effect, while Scaling-QTL contribute only to the slope-variance. In addition, it might be harder to distinguish between Scaling-QTL and the background genetics than between Reranking-QTL and the background genetics, since the background genetics were scaling effects as well.

The power is affected by a combination of the model and the QTL effect. When the QTL is fitted as a single general effect, like in RR-Q, the modelled effect will be the average effect over all environments. The power of detection should then depend on the QTL effect in the average environment. If so, RR-Q should have higher power in the scaling scenario (which has the highest average QTL effect) than in the no interaction scenario, and the lowest power in the reranking scenario (which has an average QTL effect of 0). However, RR-Q detects more than 3 times as many replicates in the no interaction scenario than in the scaling scenario (Figs. 5 and 6). Again, the QTL in the scaling scenario acts quite similar to the general genetic effect, which is also a scaling effect, and that might be the reason why it is so hard to detect. When model RR-Q was applied to the simulated scaling QTL, higher variances for error and sire-effects were estimated than when RR-Q was applied to the simulated no

interaction scenario. The sire term seems to be more capable of modelling the QTL effect than the QTL term in the model, since the sire-term includes environmental interaction. QTL with interaction are therefore hard to detect with a model assuming interaction for the sire term but not for the QTL.

Even though using random regression increases the power of detecting QTL from the scaling scenario, this scenario still gave the lowest power of detection. Since the additive sire effect in real data is a sum of the effect of all the genes affecting the trait, many genes may have an interaction pattern similar to the sire by environment interaction pattern. If so, a majority of genes will be hard to detect due to their pattern similarity with each other, while genes with a more extreme interaction pattern, different from the majority, will be easier to detect. These genes might be very interesting selection candidates in order to change the environmental sensitivity in the population. If a trait has a positive genetic correlation between production in the average environment and environmental sensitivity, selecting for higher production will also give increased environmental sensitivity. A QTL with an unusual interaction pattern gives the opportunity to select for higher production without getting the usual response in environmental sensitivity.

Fixing the correlation between slope and intercept of the random regression of QTL effects resulted in convergence problems. However, even though 19% of the brackets had to be deleted due to convergence problems in the reranking scenario, RR-QEf- had a power of 60%, which is higher than any other model in that scenario (Fig. 7). Thus fixing the correlation between slope and intercept increased the power of detection with more than 20% in the reranking scenario. Fixing the correlation between slope and intercept had only a minor effect on the power in the scaling scenario (Fig. 6).

4.2. Precision of QTL mapping

Precision can be measured by the fraction of QTL detected in the correct marker bracket and the fraction of QTL detected within the dense part of the linkage group (Figs. 5, 6 and 7). Goddard [9] proposed to fix the correlation between slope and intercept to 1 or -1 in order to obtain higher precision through steeper likelihood ratio curves. This should lead to a higher proportion of detected QTL to be in the correct bracket. Figures 6 and 7 do not support that theory, since in both scenarios, the fraction of detections done in the dense part or in the correct bracket is unaffected by the choice of random regression model. In general, very few detections were done in the sparse part. With a marker map as dense as a marker every 1 cM around the QTL position, very

few detections missed the correct position with more than 5 cM. Both the random regression models gave a high precision of QTL mapping.

4.3. Test for gene by environment interaction

Model RR-QEe falsely detected gene by environment interaction in three replicates of the no interaction scenario, while RR-QEf $_{-}$ falsely detected interaction in two replicates. Using a 0.01 threshold value, 1 false positive out of 100 replicates was expected. However, even though the interaction test was performed for only one position per chromosome, the test was done for the position where RR-QEe or RR-QEf $_{-}$ gave the highest likelihood value. The likelihood value can be higher than expected due to an overestimation of the intercept-variance or the slope-variance. Choosing the positions with the highest likelihood value for testing the interaction term increases the risk of testing the positions with falsely positive interaction terms. The number of false positives might then be higher than if a random position was tested for interaction per replicate. Assuming that one of a hundred positions shows a false interaction term, this should happen in 20 random positions from the hundred replicates, and the interaction test could falsely detect gene by environment interaction in 20% of the replicates. The dependency between the tested positions within a replicate lowers the number of false detections, since it increases the probability that more of the positions with a false interaction term fall into the same replicate. The low number of false detections makes the likelihood ratio test of RR-QEe or RR-QEf $_{-}$ vs. RR-Q well suitable for rejecting a non-existing gene by environment interaction.

The use of a random regression model detected the interaction term in most cases where the gene was detected in both scenarios “Scaling” and “Reranking”. Random regression models seem to have a high power to detect an existing interaction without detecting many false interactions. Fixing the correlation between slope and intercept to 1 or -1 did not affect the fraction of the detected QTL that also got a detected interaction term.

4.4. More than two alleles

If the QTL is multi allelic, the true correlation between slope and intercept of the allelic effects is no longer necessarily 1 or -1 , making the RR-QEf model less appropriate, depending on the allele frequencies. If the frequency of the two most common alleles adds up to almost 1, the true correlation between level and slope might be so close to one that fixing the correlation between slope and intercept has a positive impact on the power of detection.

5. CONCLUSIONS

The power of detecting a QTL depends not only on its average effect but also on its interaction with the environment, also when the model does not include an interaction term for the QTL. A QTL with an environmental interaction can be hard to detect even though it has a large average effect.

Using random regression to detect gene by environment interaction increases the power to detect QTL that interact with the environment, both when the gene has a scaling effect or when it has reranking of allele effect between environments. Fixing the correlation between the slope and the intercept of the random regression to ± 1 might in some cases increase the power compared to a random regression model that estimates this correlation, but may also lead to more convergence problems.

The random regression models had high precision in QTL mapping, and the precision was not affected by whether the correlation between slope and intercept of the random regression was fixed to ± 1 or whether it was estimated in the statistical analysis.

REFERENCES

- [1] Bidanel J.P., Ducos A., Genetic correlations between test station and on-farm performance traits in Large White and French Landrace pig breeds, *Livest. Prod. Sci.* 45 (1996) 55–62.
- [2] Bryant J., López-Villalobos N., Holmes C., Pryce J., Simulation modelling of dairy cattle performance based on knowledge of genotype, environment and genotype by environment interactions: current status, *Agr. Syst.* 86 (2005) 121–143.
- [3] Buitenhuis A.J., Lund M.S., Labouriau R., Petersen A.H., Madsen L.B., Panitz F., Bendixen C., Comparing body weight measured as single time point versus modelled as a longitudinal trait for QTL detection in pigs, EAAP 55th Annual Meeting, 5–9 September 2004, Bled, Slovenia.
- [4] Calus M.P.L., Veerkamp R.F., Estimation of environmental sensitivity of genetic merit for milk production traits using a random regression model, *J. Dairy Sci.* 86 (2003) 3756–3764.
- [5] Calus M.P.L., Groen A.F., de Jong G., Genotype \times environment interaction for protein yield in Dutch Dairy Cattle as quantified by different models, *J. Dairy Sci.* 85 (2002) 3115–3123.
- [6] Castillo-Juarez H., Oltenacu P.A., Blake R.W., McCulloch C.E., Cienfuegos-Rivas E.G., Effect of herd environment on the genetic and phenotypic relationships among milk yield, conception rate, and somatic cell score in Holstein Cattle, *J. Dairy Sci.* 83 (2000) 807–814.
- [7] Fikse W.F., Rekaya R., Weigel K.A., Genotype \times environment interaction for milk production in Guernsey Cattle, *J. Dairy Sci.* 86 (2003) 1821–1827.

- [8] Gilmour A.R., Cullis B.R., Welham S.J., Thompson R., ASREML reference manual (2001).
- [9] Goddard M.E., The validity of genetic models underlying quantitative traits, *Livest. Prod. Sci.* 72 (2001) 117–127.
- [10] Grignola F.E., Zhang Q., Hoeschele I., Mapping linked quantitative trait loci via residual maximum likelihood, *Genet. Sel. Evol.* 29 (1997) 529–544.
- [11] Jiang C., Zeng Z.-B., Multiple trait analysis of genetic mapping for quantitative trait loci, *Genetics* 140 (1995) 1111–1127.
- [12] de Jong G., Phenotypic plasticity as a product of selection in a variable environment, *Am. Nat.* 145 (1995) 493–512.
- [13] Kolmodin R., Strandberg E., Madsen P., Jensen J., Jorjani H., Genotype by environment interaction in Nordic Dairy Cattle studied using reaction norms, *Acta Agric. Scand., Section A Anim. Sci.* 52 (2002) 11–24.
- [14] Lund M.S., Sørensen P., Madsen P., Linkage analysis in longitudinal data using random regression, in: *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, Montpellier, 19–23 August 2002, France.
- [15] Lynch M., Walsh B., *Genetics and analysis of quantitative traits*, Sinauer Associates, Sunderland, 1998.
- [16] Macaranas J.M., Mather P.B., Lal S.N., Vereivalu T., Lagibalavu M., Capra M.F., Genotype and environment: A comparative evaluation of four tilapia stocks in Fiji, *Aquaculture* 150 (1997) 11–24.
- [17] Meuwissen T.H.E., Goddard M.E., Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci, *Genetics* 155 (2000) 421–430.
- [18] Meuwissen T.H.E., Karlsen A., Lien S., Olsaker I., Goddard M.E., Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping, *Genetics* 161 (2002) 373–379.
- [19] Pollott G.E., Greeff J.C., Genotype \times environment interactions and genetic parameters for fecal egg count and production traits of Merino sheep, *J. Anim. Sci.* 82 (2004) 2840–2851.
- [20] Via S., Gomulkiewicz R., de Jong G., Scheiner S.M., Schlichting C.D., van Tienderen P.H., Adaptive phenotypic plasticity: consensus and controversy, *Trends Ecol. Evol.* 10 (1995) 212–217.
- [21] Wang D.L., Zhu J., Li Z.K., Paterson A.H., Mapping QTLs with epistatic effects and QTL \times environment interactions by mixed linear model approaches, *Theor. Appl. Genet.* 99 (1999) 1255–1264.
- [22] Windig J.J., Calus M.P.L., Veerkamp R.F., Influence of herd environment on health and fertility and their relationship with milk production, *J. Dairy Sci.* 88 (2005) 335–347.
- [23] Yang J., Zhu J., Methods for predicting superior genotypes under multiple environments based on QTL effects, *Theor. Appl. Genet.* 110 (2005) 1268–1274.
- [24] Zwald N.R., Weigel K.A., Fikse W.F., Rekaya R., Identification of factors that cause genotype by environment interaction between herds of Holstein Cattle in seventeen countries, *J. Dairy Sci.* 86 (2003) 1009–1018.