QTL Mapping In Connected Porcine F_2-Crosses

C. Rückert[*] and J. Bennewitz[*]

Introduction

Mapping QTL in pigs often relied on an F_2-cross. These were either generated by mating an 'exotic' breed with a commercial breed or by mating two distinct commercial breeds (Rothschild *et al.* 2007). Usually, these experiments are analysed separately. However, it was shown that a combined analysis of several QTL experiments results in a higher statistical power to detect QTL and in a more precise localisation of QTL, which is due to the higher number of meioses used (e.g. Walling et al. 2000). A combined analysis might be especially beneficial if the crosses rely on same founder breeds, or even better, if same founder animals were used to generate the different cross.

The aim of this paper was to analyse three porcine F_2-crosses jointly. The crosses relied largely on same founder animals. For this purpose we applied a modified statistical method used in plant breeding for joint QTL analysis, which allowed the inclusion of imprinting. We compared the results of this joint analysis with those obtained from a separate analysis using a classical mapping approach.

Materials and Methods

F_2-crosses. Three crosses were included in the study. The first (MxW) was a cross obtained from Meishan (M) and European Wild Boar (W), the second (MxP) from M and Pietrain (P) and the third from W and P (WxP). The number of animals was around 320 for each cross with an equal sex ratio (males and females, no castrates). Notably, to some extent same founder animals were used in the crosses. The animals were genomewide genotyped to a large extent for same markers, mostly microsatellites. All F_2-animals were housed on one farm. In total 46 meat and carcass traits were recorded after slaughtering under standardises conditions. The experiment is described in more detail by Geldermann et al. (2003). For this study, the traits carcass weight and back fat depth were selected.

Linkage map and QTL-genotype probabilities. A common linkage map was calculated using Crimap (Green 1990). Because mostly same markers were used across the three designs a large number of co-informative meioses were observed and subsequently no problems occurred when building the map. It was assumed that two founder breeds (breed *i* and *j*) of a single cross are divergent homozygous at a QTL, i.e. showing only genotype Q_iQ_i and Q_jQ_j, respectively. Subsequently, for each F_2-individual of a certain cross four genotype probabilities $pr(Q_i^pQ_i^m)$, $pr(Q_j^pQ_i^m)$, $pr(Q_i^pQ_j^m)$ and $pr(Q_j^pQ_j^m)$ were calculated for each chromosomal position. The upper subscripts denotes for the parental origin of the alleles (i.e. paternal (*p*) or maternal (*m*) derived) and the lower subscript denotes

[*] *Universität Hohenheim, Institut für Tierhaltung und Tierzüchtung, D-70599 Stuttgart, Germany*

for the breed origin of the alleles. These probabilities were estimated using a modified version of Bigmap (Reinsch 1999).

Genetic and stastical model. The genetic model follows in general the multiallelic model of Liu and Zeng (2000), but was extended to account for imprinting. The genetic mean (μ) is defined as the mean of the breeds. The genetic model for an F_2 population generated from the two breeds *i* and *j* was

$$\begin{bmatrix} g_{ii}^{pm} \\ g_{ij}^{pm} \\ g_{ji}^{pm} \\ g_{jj}^{pm} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} a_i^p \\ a_i^m \\ a_j^p \\ a_j^m \\ d_{ij} \end{bmatrix} + \begin{bmatrix} \mu \\ \mu \\ \mu \\ \mu \end{bmatrix},$$

where the subscripts are used as defined above. This genetic model was used to set up the following statistical model, again following Liu and Zeng (2000),

$$y_{ijk} = \mathbf{x}_{ijk}\beta + (z_{ijk,i}^{p}\mathbf{w}_{ijk,i}^{p} + z_{ijk,i}^{m}\mathbf{w}_{ijk,i}^{m} + z_{ijk,j}^{p}\mathbf{w}_{ijk,j}^{p} + z_{ijk,j}^{m}\mathbf{w}_{ijk,j}^{m})\mathbf{a} + z_{ijk,ij}^{pm}\mathbf{w}_{ijk,ij}^{pm}\mathbf{d} + e_{ijk}$$

where y_{ijk} is the phenotypic observation of the k-th individual in the F_2 cross derived from line *i* and *j*. The term $\mathbf{x}_{ijk}$ denotes for the *ijk*-th row vector of a design matrix linking the phenotypic observation of the F_2-offsping to some fixed effects stored in β . The term e_{ijk} is a random residual with heterogeneous error variance, i.e. $e_{ijk} \sim N(0, \sigma_{ij}^2)$. Vector **a** contains the additive effects ($a_1^p, a_1^m, ...a_L^p, a_L^m$), with *L* being the number of founder breeds. Vector **d** the dominance effects ($d_{1,2}$, $d_{1,2}$, …, $d_{(L-1),L}$). The four **w** terms are row vectors of length 2*L with one element of one and the others are equal to zero. Each of them indicate one of the four possible additive effects in **a** that could be observed in the F_2-individual based on pedigree data. For example, $w_{ijk,i}^{p}$ denotes for the putative allele in offspring *ijk* (indicated by first lower subscript *ijk*) inherited paternally (indicated by upper subscript *p*) from line *i* (indicated by second lower subscript *i*). The four z terms are scalars and are either zero or one. They indicate if the offspring inherited the corresponding allele form corresponding parent. For each offspring these four terms sum up to two. Similarly, $w_{ijk,ij}^{pm}$ is a row vector of length *L*, indicating which dominance effect could be possible in the offspring based on pedigree data. The scalar $z_{ijk,ij}^{pm}$ is one if the offspring is heterozygous at the QTL and zero otherwise. The true *z* terms are unknown and have to be approximated using marker data. In this study the *z* terms were calculated from the estimated four QTL-genotype probabilities at each chromosomal position. For example, the term $z_{ijk,i}^{p}$ was set equal to $pr(Q_i^p Q_i^m) + pr(Q_i^p Q_j^m)$. The dominance $z_{ijk,ij}^{pm}$ term was the sum of the two heterozygous genotype probabilities. The final statistical model was a multiple linear regression. In order to avoid an over parameterisation, the regression was parameterised, taking the restrictions

$\sum_{i=1}^{L} a_i^p = 0$ and $\sum_{i=1}^{L} a_i^m = 0$ into account. Hence, 2*L-2 additive effects were estimated. The model was fitted every cM on the genome by adapting the z terms accordingly. When imprinting is not accounted for the model reduced to the proposed model of Liu and Zeng (2000). In this case L-1 additive effects were estimated. The test statistics from the multiple regression were converted into LOD-scores. Confidence intervals were calculated by the one-LOD drop off method. The QTL genome scan was repeated using previously identified QTL as cofactors in the model until no new QTL were detected.

The global null hypothesis was that every parameter in **a** and **d** are zero, the alternative hypothesis was that at least one was significantly different from zero. Chromosome-segment-wise critical threshold values (5 percent) for this global test were calculated using the quick method of Piepho (2001). A chromosome-segment was defined as a chromosomal region bounded by either the start or endpoint of a chromosome or by the bounds of a confidence interval of a second QTL on this chromosome. Once the global null hypothesis was rejected, a series of sub-hypothesis regarding the mode of inheritance (additive, dominant, imprinting QTL) and number of alleles were tested. Here the comparison wise error probability was used. In order to detect imprinting, it was tested whether the paternal inherited and maternal inherited QTL effects of a certain founder breed differed significantly. For comparison purpose, the crosses were analysed separately using the Haley-Knott-regression. The model included some fixed effects, additive, dominant and imprinting QTL effects and was parameterised as shown in Mantey et al. (2005).

Results and Discussion

The results revealed 16 QTL for carcass weight and 15 QTL back fat depth. When imprinting was not accounted for, these figures reduced to 9 and 8, respectively. This demonstrated the importance of modelling imprinting for these traits. Around one half of the identified QTL showed three alleles, i.e. were segregating between all three crosses.

When comparing the results with those obtained from separate analysis of the crosses (see Figure 1), it became obvious that the test statistics were much higher and much sharper, resulting in confidence intervals with a reduced width. Additionally, the number of chromosome-segments with a QTL was much higher, stressing the increased statistical power of the combined analysis. This power is a result of a around threefold increase of the number of meiosis and the decrease of the number of estimated QTL parameters from nine (three parameters in each cross) to seven. Strength of the applied method is also the speed of computation. Because the permutation test for significance threshold determination is replaced by the quick method, a genome scan was completed within several minutes. This makes it especially valuable to analyses the large number of traits that were collected for these animals.

Modelling heterogeneous error variances seemed to be important, because they really were heterogeneous. Reason for this might be a different number of segregating QTL between the crosses. Heterogeneity reduced if more QTL were included in the model (not shown).

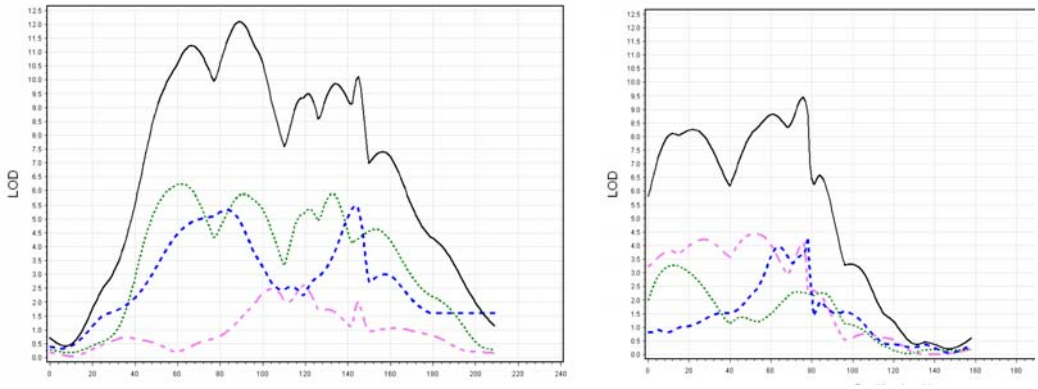

Figure 1: Test statistics profiles for carcass weight on chromosome 1 (left) and for back fat depth on chromosome 2 (right). The black line denotes for the results from the combined analysis. The violet (green, blue) line denotes for the results from the analysis of the MxP (WxP, WxM) analysis.

Conclusion

This study showed the benefit of analysing three connected F_2-crosses jointly. The advantage of this analysis is the larger number of meiosis and the reduced number of estimated parameters, compared to a separate analysis. Additionally, modelling imprinting was important for these traits. The models will be applied to the remaining set of traits and will be expanded to include interactions between loci.

References

Geldermann, H., Müller, E., Moser, et al. (2003). *J. Anim. Breed. Genet.*, 120: 363-393.

Green, P., Falls, K., Crooks, S. (1990). Washington University, School of Medicine, St. Louis.

Liu, Y., Zeng, Z.B. (2000). *Genet. Res.*, 75: 345-355.

Mantey, C., Brockmann, G.A., Kalm, E., Reinsch, N. (2005). *J. Heredity*, 96: 329-338.

Piepho, H.P. (2001). *Genetics*, 157: 425-432.

Reinsch, N. (1999). *J. Anim. Breed. Genet.*, 116: 425-435.

Rothschild, M.F., Hu, Z.L., Jiang, Z. (2005). *Int. J. Biol. Sci.*, 3: 192-197.

Walling, A.G., Visscher, P.M., Andersson, et al. (2000). *Genetics*, 155: 1369-1378.