

**Comparing body weight measured as single time point versus modelled as a longitudinal trait
for QTL detection in pigs**

A.J. Buitenhuis*, M.S. Lund, R. Labouriau, A.H. Petersen, L.B. Madsen, F. Panitz, C. Bendixen

Danish Institute of Agricultural Sciences, Department of Animal Breeding and Genetics, P.O. Box 50, 8830 Tjele, Denmark.

*Corresponding author: Bart.Buitenhuis@agrsci.dk

Abstract. Genes involved in the regulation of these traits may be differentially expressed over time. Many traits studied have a repeated measurement over time, e.g. growth, but are often analysed as single time points. Taking into account the (co)variance structure between these point may improve the detection of QTL. In this paper a single trait QTL model, a multi trait QTL model and a random regression method are used to detect QTL for body weight (BW) in a large pig population on pig chromosome 2. The single trait QTL mapping method revealed a QTL for BW1, while the QTL in BW2-BW4 were not significant. The multi trait QTL analysis gave the same profile likelihood as the single trait analysis, but did not reach significance. This is due to the many parameters to be estimated in the model. The most likely model suggests a QTL affecting all four traits (BW1, BW2, BW3, BW4). The random regression model reveals a significant QTL with an increasing effect over time, however, reveal convergence problems as well. In conclusion: the longitudinal model (random regression) has more power and reveals more information about the effect over time, however, it is not robust. Therefore, simpler models should be fitted as well.

Keywords: Quantitative Trait Loci; Random regression; body weight; pigs

INTRODUCTION

The ability to study the genetics underlying quantitative variation has enormously enhanced by the development of genetic maps and the development of statistical methods to identify loci influencing quantitative traits (Lander and Botstein, 1989; Haley et al., 1994; Meuwissen and Goddard, 2000). Quantitative traits are controlled by a large number of loci (QTL) and are complex in nature. Genes involved in the regulation of these traits may be differentially expressed over time. Many traits studied have a repeated measurement over time, e.g. growth, but are often analysed as single time points, assuming the QTL equally expressed over time. The average effect of a differentially expressed QTL over time has likely a small average effect and will be difficult to detect. Allowing the analysis for time dependent QTL effects, by using longitudinal models may increase the statistical power to detect QTL (Lund et al., 2002).

Identification of time dependent QTL effect is of help in understanding the biology of the trait studied. One can imagine that QTL alleles effecting persistency in milk production in cattle or egg production in layers might be more interesting then the QTL alleles influencing the beginning of the production period. Insight in the expression profile of a QTL over time may help to improve selection decisions based on QTL information.

When a trait was measured over time one can consider each single time point as a separate trait. Using a single trait QTL analysis to analyse the subsequent time points individually is a valid way of analysing the data. However, the power to detect QTL may be different due to different in sample size between time points and the QTL may not be expressed equally in each time point. To improve power of QTL detection one can take the (co)variance structure into account between the traits by using a multi-trait analysis. Taking into account the correlation between traits can increase the power to detect QTL not detected in the single trait analysis. In addition, the multi trait analysis reveals information on the correlation between the traits and facilitates detection of pleiotropic

QTL. However, the number of parameters to be estimated in the multi trait analysis increases as well compared to the single trait analysis, which influences the significance threshold (Piepho, 2001).

Modelling the trait over time using random regression methods is another way of dealing with the (co)variance structure between the time points i.e., modelling the trait as a longitudinal trait. The advantages of using random regression is that (i) less parameters need to be estimated which increase the power of QTL detection and (ii) differential expression within time periods is taken into account. In this model the QTL effect over time as well as the change in variance and covariance over time is taken into account. The (co)variances between QTL allelic effects are modelled using the random QTL model (Fernando and Grossman, 1989). The random allelic effects are regressed on the curve parameters similar as the additive polygenic effect. In this way the shape of the curve depends on the QTL genotypes. This approach has been successfully applied on simulated DYD data (Lund et al., 2002) and real data (unpublished).

The aim of this paper is: i) to detect QTL for body weight in pigs for pig chromosome (SSC) 2 using a single trait QTL analysis, a multi-trait QTL analysis and a random regression QTL analysis ii) to compare the different methods iii) to validate the supposed advantage of the use of a longitudinal QTL analysis over a single trait or multi trait QTL analysis.

MATERIALS AND METHODS

Population

Twelve pure breed Duroc boars coming from a commercial population were mated to 709 Landrace x Large White dams, resulting in 1,126 litters. Litters with eight or more piglets were included in the experiment. Directly after birth piglets were ear tagged with an individual number. In total there were about 12,000 offspring obtained. The piglets were born in the period from February 2000 to March 2001 divided over three herds in Danmark.

Phenotyping

Body weight of all the piglets was recorded at four different time points during the rearing period. At birth (BW1) (0-4 d), at weaning (BW2) (18-40 d), start at the first growth period (BW3) (58-115 d) and at slaughter (BW4) (140-201 d) body weight was measured.

Genotyping and map construction

Sequences of ESTs were available from the Danish-Chinese pig genome sequencing project (<http://www.genome.kvl.dk>) and sequences from an exon-resequencing project in which exons of different genes were sequenced in a panel containing the boars used in this QTL experiment. Selected sequences containing polymorphisms were evaluated (Barbujani et al., 1997) and typed on the all animals in the pedigree. In total 150 SNPs were typed and used to calculate distances between each linked marker pair using CRI-MAP (Green et al., 1990). The comparative human-pig map is used to assign the linkage groups to the corresponding pig chromosomes, based on the homology between human and pig genes. To verify QTL segregating in the population for BW, a valid method is to start with chromosomal regions containing potentially a QTL (Evans et al., 2003). In this paper we only focus on pig chromosome (SSC) 2 as a start.

Statistical analyses

Single trait single QTL analysis. Each of the different periods were analysed separately using linkage analysis. The full model can be expressed as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{q} + \mathbf{e}, \quad (1)$$

where \mathbf{y} is a vector of n observations, \mathbf{X} is a known design matrix, $\boldsymbol{\beta}$ is a vector of unknown fixed effects, \mathbf{W} is a known matrix relating each individual record to its unknown additive QTL effect, \mathbf{q} is a vector of unknown additive QTL effects of individuals and \mathbf{e} is a vector of residuals. The random variables \mathbf{u} , \mathbf{q} and \mathbf{e} are assumed to be multivariate normally distributed and mutually independent. BW was corrected for the effects of SEX ($i=2$), HERD ($i=3$), AGE, where AGE has 4 levels for BW1, 22 levels for BW2, 57 levels for BW3, and 48 levels for BW4.

Multi trait single QTL analysis. Model (1) can be extended to a multi trait single QTL model where \mathbf{y} is an $n * t$ vector of n observations on t traits (Sørensen et al., 2003).

Random regression mixed model. The time points were analysed together in a random regression model including QTL effects. The full model can be expressed as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{p} + \mathbf{Z}_2\mathbf{u} + \sum_{i=1}^{nqtl} \mathbf{W}_{qi} + \mathbf{e}, \quad (2)$$

where \mathbf{y} is a vector of n observations taken at different time points, $\boldsymbol{\beta}$ is a vector of effects describing the fixed curve over time, \mathbf{X} is a design matrix relating fixed effects to records, $\mathbf{Z}_1\mathbf{p}$, $\mathbf{Z}_2\mathbf{u}$, and \mathbf{W}_{qi} are curves of random deviations from the fixed curve due to permanent environmental effects, polygenic effects, and allelic effects of the i^{th} QTL, respectively. \mathbf{p} , \mathbf{u} , \mathbf{q}_i entail the random regression coefficients, \mathbf{Z}_1 , \mathbf{Z}_2 , and \mathbf{W} are design matrixes with covariates of the curve. The random regression coefficients are assumed to be multivariate normal distributed (Lund et al., 2002).

IBD matrix. First the gametic relationship matrix (Fernando and Grossman, 1989) was calculated and then using the linear relationship between the gametic relationship matrix and the IBD matrix, the IBD matrix was designed (George et al., 2000). The covariance structure among the random QTL allelic effects of all animals in de the pedigree, are described in the gametic

relationship matrix. The information of the transmission of linked markers is used to calculate the IBD probabilities at the position of a putative QTL position (Sørensen et al., 2003).

Random regression applied to Body Weight. The analysis was performed within families. In the fixed part of the model ($\mathbf{X}\beta$) the following effects were fitted SEX (I=2), HERD (I=3), AGE (I=18). SEX*AGE (i= 36) and HERD*AGE (i=54) are the interaction terms in the fixed part. In the random part of the model the permanent the environment effect (Ep), polygenic effect (PG) and a QTL effect was fitted. The Legendre Polynomial up to the 3rd order (n=3) was used on the Ep and the QTL effect. The residuals were weighted with one over the mean weight in each age class to account for the increasing variance over time. Using this model severe convergence problems were encountered due to high correlation (~1) between the parameters to be estimated in the model. Therefore, a reparameterization was applied to the LP in the form of (1+kLP1). The optimum k value was tested for the model by making a profile likelihood with k as random variable. The reparameterization was applied to the Ep as well as to the QTL effect. Using the reparameterized LP facilitates convergence of the model.

Significance thresholds were calculated using a quick method to compute approximate threshold levels that control the genome-wise type I error (Piepho 2001). A significance level of 5% chromosome wise was considered to be significant.

RESULTS

Linkage map. SSC2 contains 6 SNP markers and has a length of 125.6 cM.

Single trait single QTL analysis. The QTL analysis was performed within single families for the four periods for BW (BW1-BW4). Out of the 14 sire families, only one family (family 5) showed a significant QTL for BW1 in the end of the chromosome (111 cM). For BW2-BW4 there is a peak in the same region on SSC2, however, QTL for BW2-BW4 were not significant (Table I).

Multi trait single QTL analysis. A multi trait single QTL analysis was performed using sire family 5 to test whether the same QTL was influencing all four periods. The phenotypic correlation between body weight in the different periods are presented in Table II. Running the full model (QTL influencing all four traits) reveal a loglikelihood of 17.3. Although this is fairly close to the significance threshold this was not significant. Reducing the model step by step (n=15 trait/QTL combinations) did not reveal in a significant detection of a QTL. Figure I shows the most important changes given in the last point of the chromosome. The QTL variances obtained in the multi trait analysis are shown in Table III. The correlation between the QTL effects in the different periods suggests a pleiotropic QTL affecting all four traits (Table IV).

Random regression QTL analysis. Using the reparameterization for the Legendre Polynomials (1+kLP1) resulted in an optimum k value of 0.8 for the Ep and an optimum k value of 0.815 for the QTL effect (data not shown). When applying this to the data of sire 5, the peak in the end of the chromosome (position 10 to 12) is in the same position as the single trait analysis and the multi trait analysis (Figure II). The peak in the first half of the chromosome was not detected with the single trait or multi trait analysis in this family. The variance of the QTL was higher than the variance estimated in the same period using the multi trait analysis. A plot of the variance explained by the QTL over time is presented in Figure III.

DISCUSSION

Data obtained in an experiment can be analysed in different ways. Especially when data is recorded over time it would be of interest to take the (co)variance structure between time points into account as well. In this study we used body weight measures in a large pig population as a trait to study the advantages or disadvantages of using different methods for QTL detection.

The random regression method applied to the body weight data encountered some convergence problems due to highly correlated parameters. These problems of correlated parameters were not encountered in the simulation study of milk production data (Lund et al., 2002) and in real milk production data (unpublished). Although reparameterization of the Legendre polynomials reduced the problem of correlated parameters, this indicates that using random regression methods to model the growth curve is not straightforward. An advantage of the random regression would be the higher power to detect QTL compared to a multi trait single QTL analysis due to the fact that in the random regression the number of parameters to be estimated in the model are less compared to the multi trait analysis. In this study one can see that the QTL was detected using the random regression method, while this was not the case using the multi trait analysis. However, in the random regression there was a peak in the beginning of the chromosome, not detected neither in the single trait or in the multi trait analysis. Although this peak was not significant it remains unclear whether it is a true peak because the random regression is more sensitive than the single trait and multi trait model, or a peak caused by the method. The analysis with the single trait model and multi trait model give a similar log likelihood profile. Indicating that there is a QTL segregating in this part of the chromosome. The advantage of using a multi trait analysis over a single trait analysis is that the correlation between the traits is taken into account in the multi trait analysis. This can facilitate QTL detection when the right model is applied. Comparing the variances between the single and multi trait model show that there is a difference but they are similar and have the same trend. The advantage of using multi-trait models is underlined by much smaller standard errors on the variance estimates. However, comparing the variance explained by the QTL over time in the random regression method show a higher variance in the BW4 period than both the single or multi trait analysis. Whether this is a feature of the model used is in this stage still unclear. One problem in the present data is probably that there are too few observations per animal to estimate the within animal covariance structure over time properly. To validate the

behaviour of the random regression method with regard to sensitiveness compared to the single and multi trait analysis and estimating the variance over time, a simulation study has to be performed using the data structure as the one of this pig population.

This pig population is a very powerful population for QTL detection for production traits. The power estimated for QTL detection for body weight using the method of Van der Beek et al. (1995) is around 90% depending on the final SNP map density. It is therefore anticipated that in case there is a QTL segregating in the population it is likely to be detected using the standard QTL mapping methods. Selecting potential chromosomes, which possibly contain a QTL is an efficient method to screen the population for QTL for production traits (Evans et al., 2003). Although there is ample need in addition in increasing the marker density on SSC2, to draw a more precise conclusion on the QTL, it may be suggested that the actual QTL is outside the scanned region of the chromosome.

In conclusion, the random regression method identified a significant QTL, which has an increasing effect over time. This QTL was only detected for BW1 using the single trait QTL model and although the multi trait model showed the same profile likelihood this was not significant. Using the random regression method there is potentially more power to detect QTL and reveals extra information about effects over time. However, it is not clear whether the method is robust in the present data.

REFERENCES

- Barbujani G., Magagni A., Minch E., Cavalli-Sforza, L.L., An apportionment of human DNA diversity, *Proc. Natl. Acad. Sci.* 94 (1997) 4516-4519.
- Evans G.J., Giuffra E., Sanchez A., Kerje S., Davalos G., Vidal O., Illán S., Noguera J.L., Varona L., Velander I., Southwood O.I., De Koning D.-J., Haley C.S., Plastow G.S., Andersson L., Identification of quantitative trait loci for production traits in commercial pig populations, *Genetics* 164 (2003) 621-627.
- Fernando R.L., Grossman M., Marker-assisted selection using best linear unbiased prediction, *Genet. Sel. Evol.* 21 (1989) 476-477.
- George A.W., Visscher P.M., Haley C.S., Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach, *Genetics* 156 (2000) 2081-2092.
- Green P., Falls K., Crook S., *Documentation for CRI-MAP*, Version 2.4 Washington University School of Medicine, St. Louis, USA (1990).
- 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.
- Lander E.S., Botstein D., Mapping mendelian factors underlying quantitative traits using RFLP linkage maps, *Genetics* 121 (1989) 185-199.
- Lund M. S., Sørensen P., Madsen P., Linkage analysis in longitudinal data using random regression, In: 7th World Congress on Genetics Applied to Livestock Production, 19-23 August 2002, Montpellier, CD ROM communication 19-23, Inra, Castanet-Tolosan.
- Meuwissen T.H., Goddard M.E., Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci, *Genetics* 155 (2000) 421-430.
- Piepho H.P., A quick method for computing approximate thresholds for quantitative trait loci detection, *Genetics* 157 (2001) 425-432.

Sørensen P., Lund M.S., Guldbrandtsen B., Jensen J., Sørensen D., A comparison of bivariate and univariate QTL mapping in livestock populations, *Genet. Sel. Evol.* 35 (2003) 605-622.

Van der Beek S., Van Arendonk J.A.M., Groen A.F., Power of two- and three-generation QTL mapping experiments in an outbred population containing full-sib and half-sib families. *Theor. Appl. Genet.* 91 (1995) 1115-1124.

Table I. QTL variance with the standard error (SE) given for the last position of SSC2 for body weight measured in four different periods detected in family 5 using single trait analysis.

Trait ¹	Position (Morgan)	σ^2_{QTL}	SE	$\sigma^2_{\text{residual}}$	SE	Loglikelihood
BW1	1.11	0.006	0.008	0.09	0.01	8.27 [*]
BW2	1.11	0.049	0.085	2.38	0.16	2.01
BW3	1.11	3.02	6.73	33.34	6.73	6.27 [†]
BW4	1.11	35.33	33.29	67.22	42.88	4.82

¹BW = body weight measured at birth (BW1), weaning (BW2), start 2nd growth period (BW3) and slaughter (BW4).

* 5% chromosome wide significance level; [†] suggestive linkage

Table II. Phenotypic correlation between the body weight traits measured

Trait ¹	BW1	BW2	BW3	BW4
BW1	1.0			
BW2	0.58	1.0		
BW3	0.52	0.69	1.0	
BW4	0.44	0.61	0.79	1.0

¹BW = body weight measured at birth (BW1), weaning (BW2), start 2nd growth period (BW3) and slaughter (BW4).

Table III. QTL variance with the standard error (SE) given for the last position of SSC2 for body weight measured in four different periods detected using multi trait analysis.

Trait ¹	Position (Morgan)	σ^2_{QTL}	SE	$\sigma^2_{\text{residual}}$	SE	Loglikelihood
BW1	1.11	0.007	0.0002	0.097	0.003	17.29
BW2	1.11	0.12	0.003	2.32	0.079	17.29
BW3	1.11	11.58	0.27	23.33	0.79	17.29
BW4	1.11	18.0	0.43	93.83	3.19	17.29

Table IV. Correlation between the effects of QTL detected for body weight measured at the four different periods using the multi trait analysis.

	QTL effect	QTL effect	QTL effect	QTL effect
	BW1	BW2	BW3	BW4
QTL effect BW1	1.0			
QTL effect BW2	0.9989	1.0		
QTL effect BW3	0.9993	0.9997	1.0	
QTL effect BW4	0.9989	0.9993	0.9998	1.0

Figure I. Profile likelihood of the multi trait single QTL analysis when the QTL is affecting all four (BW1, BW2, BW3, BW4) traits (Full model). For the reduced models the test statistics for the last point is given. On the X-axis the length of the chromosome (SSC2) is given in Morgan. On the Y-axis the log likelihood is given.

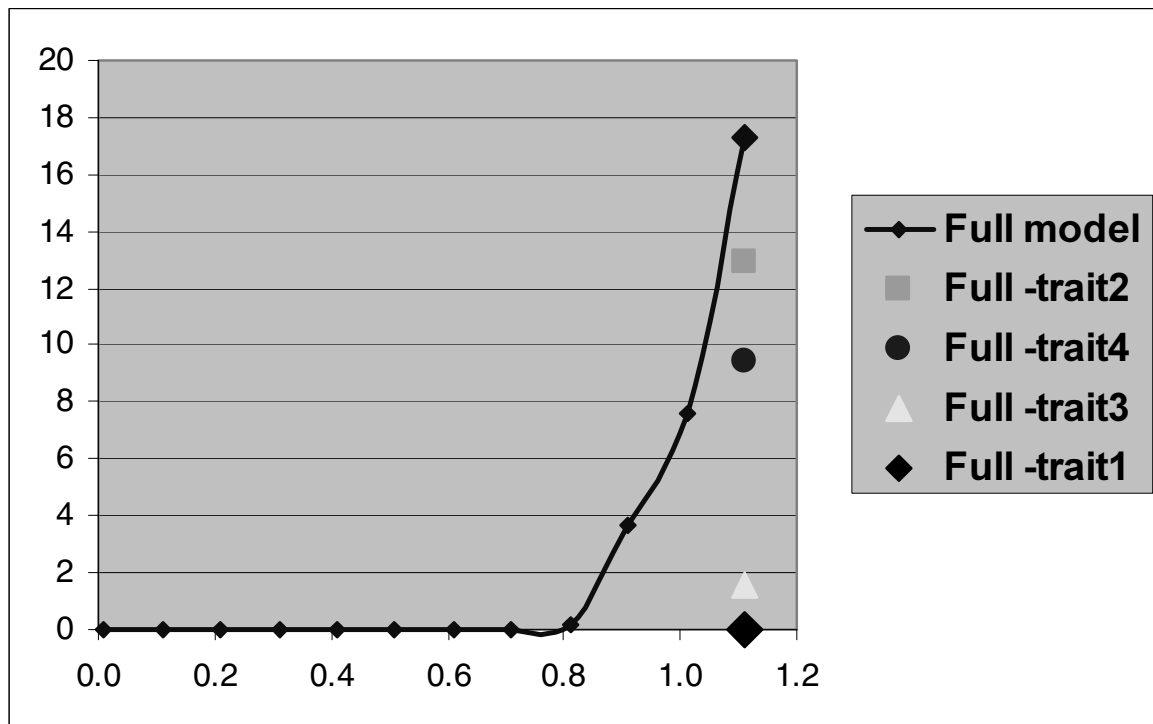


Figure II. Profile likelihood of the random regression QTL mapping method (solid line). The significance threshold is given as a dashed line. On the X-axis the length of the chromosome (SSC2) is given in Morgan. On the Y-axis the log likelihood is given.

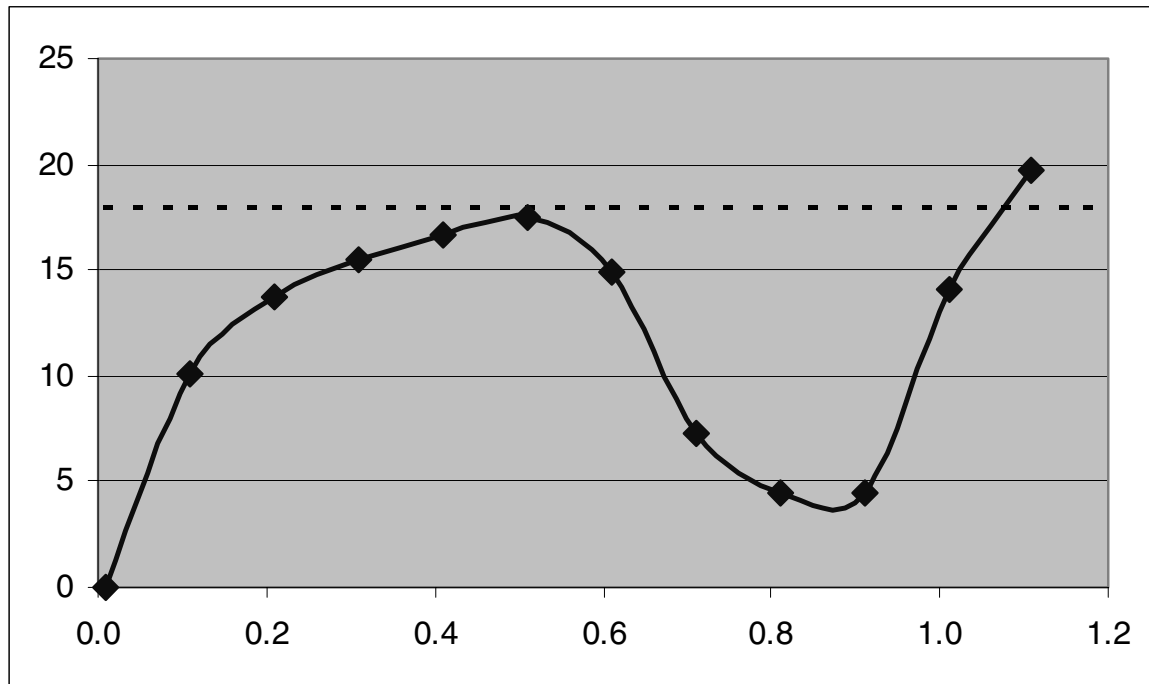


Figure III. Variance explained by the QTL over time.

