QTL DETECTION IN OUTBRED POPULATIONS; COMPARISON OF VARIANCE COMPONENT AND SIMPLE METHODS

D.J. de Koning[1], Y. Nagamine[1], G. Evans[2] and C.S. Haley[1]

[1] Roslin Institute, Roslin, Midlothian, EH25 9PS, UK
[2] PIC International, Cambridge, UK

INTRODUCTION

In poultry and pigs, many QTL have been successfully mapped in experimental populations (Andersson, 2001). These experimental populations are mostly analysed by regression methods under a line-cross model (Haley *et al*., 1994) or a half-sib model (Knott *et al.*, 1996). However, the regression methods do not take all relationships in a complex pedigree into account, do not generally include polygenic components and provide no genotypic value for the individual animals. Methods to jointly estimate QTL and polygenic effects have been suggested by Fernando and Grossman (1989). George *et al.* (2000) present a two-step approach for a variance component analysis of QTL. In the first step, the identity by descent (IBD) proportions at pre-defined genome locations are estimated between all related individuals, while in the second step the QTL and the polygenic effects are estimated by Residual Maximum Likelihood (REML). The IBD proportions are calculated using Monte Carlo Markov Chain (MCMC) methods, which require a large number of iterations (Heath, 1997). Pong-Wong *et al.* (2001) present a deterministic alternative for the estimation of IBD proportions, which increases the computation speed dramatically. When genotyping is only performed on two generations, and there are few additional relationships between families, the estimation of IBD proportions can be restricted to those within half-sib and full-sib families. Nagamine *et al.* (submitted) have developed a "simple" deterministic IBD approach that estimates within family IBD proportions. In this paper we compare the performance of the simple deterministic method (SMD) to that of the MCMC based method and the half-sib regression method, using data from commercial pig breeds.

MATERIAL AND METHODS

IBD proportions using MCMC. The IBD proportions are estimated using an adapted version of the QTL mapping software LOKI (Heath, 1997). This program uses a Gibbs sampler to obtain IBD proportions in arbitrary pedigrees with missing marker data and unknown haplotypes. A detailed description of the method can be found in Thompson and Heath (1999). We used at least 7,000 iterations and verified convergence by repeating some of the analyses with >20,000 iterations.

IBD proportions using deterministic approach. Knott *et al.* (1996) introduced a method to estimate the conditional probability of inheriting a specific paternal haplotype. These methods can also be implemented to estimate allelic IBD proportions between two paternal half-sibs. Let pj and pk be the probability of inheriting the 'first' paternal haplotype for animals j and k at a given location. The paternal allelic IBD probability between animals j and k can be described by:

(pj x pk) +[(1-pj) x (1-pk)]

Likewise, the maternal allelic IBD probability can be estimated when full sib structures exist. The genotypic IBD proportions are estimated by averaging the maternal and paternal allelic IBD probabilities.

Variance Component Analysis. The animal model for the quantitative trait, including a random QTL effect is:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Zu} + \mathbf{Zv} + \mathbf{e}$$

where **y** is a vector of phenotypes, **X** is a design matrix, **β** is a vector of fixed effects, **Z** is an incidence matrix relating animals to phenotypes, **u** is a vector of polygenic effects, **v** is a vector of additive genotypic QTL effects and **e** is a residual vector. The random genetic effects **u** and **v** are assumed to be distributed as multivariate normal densities with mean zero and variances $\mathbf{A}\sigma^2_u$ and $\mathbf{G}\sigma^2_v$, respectively. **A** is the standard additive genetic relationship matrix and **G** is the covariance matrix for the additive QTL effects, represented by the proportion of alleles IBD. ASREML (Gilmour *et al.,* 1998) was used to perform the variance component analysis. A test statistic for a given location was obtained by running an animal model without a QTL effect and subsequently calculating a likelihood ratio (LR).

Implementation for commercial pig data. Data was used from the PigQtech EU demonstration project in which ten different pig populations were targeted for regions where QTL have been identified in experimental populations. This study used data from two different commercial lines based on Hampshire and Large White breeds from PIC. Both lines consisted of ten paternal half-sib families, sub-divided into 71 full-sib families with a total of 511 and 542 offspring for the Hampshire and Large White derived lines, respectively. Genotypes for 29 microsatellite markers in ten candidate regions were obtained on all animals. Three phenotypes, that were available for the offspring, were analysed: daily gain during test (TDG), daily gain during life (LDG) and P2, a combined measure of back fat thickness. Exploratory analyses were performed under a half-sib model (Knott *et al.,* 1996) using the QTL express software at: http://qtl.cap.ed.ac.uk/. Thirty-two candidate positions were defined, equally distributed across the ten chromosome regions, and analysed using the two-step variance component analyses using both the MCMC and the SMD method. For significance testing, we imposed a 5% threshold assuming the LR to be distributed as a mixture between zero and a χ^2 with 1 d.f.

RESULTS AND DISCUSSION

The exploratory least squares (LS) analyses showed evidence for QTL affecting LDG and TDG on SSC3 and SSC7 in the Hampshire line. In the Large White line, QTL for LDG and TDG were detected on SSC1 and SSC6, and a QTL for P2 on SSC4.

Testing 32 locations in two breeds for linkage with any of the three traits, using either SMD or MCMC IBD proportions, resulted in a total of 284 ASREML analyses. From these, the LR of the best location for every chromosome segment and trait is represented in Figure1. Figure 1a shows that there is good agreement between the half-sib least squares results and the SMD variance component analyses. The variance component analyses did not confirm the two QTL on SSC1 for LDG and TDG (the two left most points above the F-ratio threshold in Figure 1a). However, two additional QTL were detected on SSC10 for LDG and SSC6 for P2. Figure 1b shows the relationship between the two IBD methods. Although there is generally good

agreement between the two methods, the two additional QTL on SSC6 and SSC10 were not confirmed using the MCMC IBD approach. Figure 2 illustrates for an example set of data the differences in IBD proportions when using the SMD or the MCMC method for a single position flanked by two markers. The correlation between the estimates under the two methods is 0.78 while fitting a regression through the points gives a slope of 0.99. The difference between the MCMC and SMD method is most distinctive around the SMD values of 0.25 and 0.50. Here, some full- or half-sib pairs are uninformative under the SMD method, while proportions from the MCMC approach cover the full range of possible IBD proportions within such sibs. The clustering of points from the MCMC method at values of 0.0 and 0.5 where the SMD method infers values of 0.25 suggests that the MCMC approach is using extra information, whereas variation within these clusters probably represents in part sampling variation. One point of concern is the stability of ASREML for the relatively small number of animals used here. This is reflected by the observations that the estimate of a variance component could sometimes decrease tenfold, without much change in the likelihood. In one other case, the polygenic variance became negligible, when the QTL was significant. However, the good agreement with the regression analysis suggests that this is evidence of difficulty in estimating the polygenic component in such a small data set rather than suggesting there is likely to be a problem with the spurious detection of QTL.

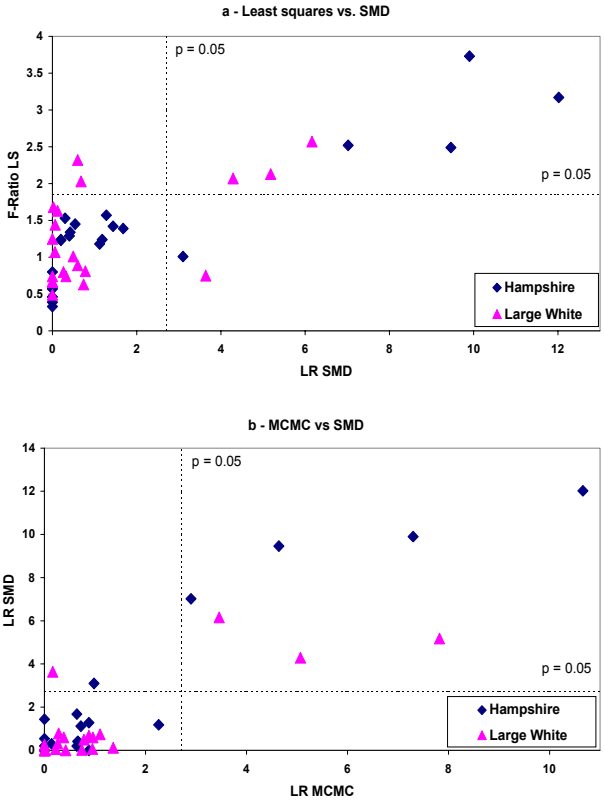

Figure 1. Comparisons between the highest test statistic for ten QTL regions and three traits

CONCLUSION

This research demonstrates that the two-step variance component methods can be used to confirm QTL in commercial lines, with comparable power for the SMD and the MCMC IBD methods. Although this data structure included full-sib families, the least-squares half-sib analysis provided a robust QTL detection method, while the potential extra power from using variance component methods, arising from considering the maternal QTL genotypes, was not evident in the results presented in this paper. Variance component analyses might prove more powerful when parents also have phenotypes and full-sib family sizes are large.

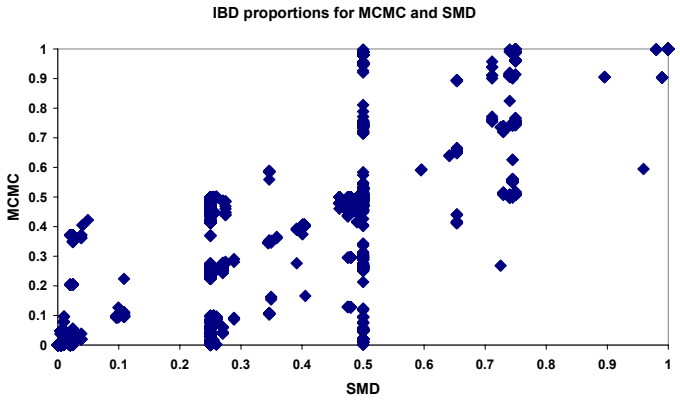

Figure 2. Scatter plot for the IBD proportions, estimated by MCMC or SMD method

ACKNOWLEDGEMENTS

This research was supported by the European Union through the PigQTech project (contract no. BIO4-CT97-2243; coordinated by Leif Andersson). DJdK acknowledges support from DEFRA through Link project LK0625. YN acknowledges support from the National Agricultural Research Organization of Japan. CSH acknowledges support from the BBSRC.

REFERENCES

Andersson, L. (2001) *Nat. Rev. Gene.t* **2** : 131-138.

Fernando, R.L. and Grossman, M. (1989) *Genet. Sel. Evol.* **21** : 467-477.

George, A.W., Visscher, P.M. and Haley, C.S. (2000) *Genetics* **156** : 2081-2092.

Gilmour, A.R., Cullis, B.R., Welham, S.J. and Thompson, R. (1998) "ASREML: user manual". Orange Agricultural Institute, NSW, Australia.

Haley, C.S., Knott, S.A. and Elsen, J.M. (1994) *Genetics* **136** : 1195-1207.

Heath, S.C. (1997) *Am. J. Hum. Genet.* **61** : 748-760.

Knott, S.A., Elsen, J.M. and Haley, C.S. (1996) *Theor. Appl. Genet.* **93** : 71-80.

Pong-Wong, R., George, A.W., Woolliams, J.A. and Haley, C.S. (2001) *Genet. Sel. Evol.* **33** : 453-471.

Thompson, E.A. and Heath, S.C. (1999) "Statistics in Molecular Biology", p 498-506, Springer-Verlag IMS lecture note series.