# COMPARING VARIOUS MODELS FOR DETECTING LITTER-SIZE QTL

**S. A. Kayis[1], N. J. Maqbool[2], P. C. Thomson[1], C. Moran[2], and F. W. Nicholas[2]**

[1]Biometry Unit, Department of Crop Sciences, University of Sydney, NSW 2006
[2]Department of Animal Science, University of Sydney, NSW 2006

## SUMMARY

A genome scan for QTL for mouse litter size was performed using Mapmaker/QTL and a novel generalised linear model. Both models produced similar results in terms of detecting and locating QTL. Mapmaker/QTL identified a QTL affecting litter size on chromosome 3 from 4$^{th}$ parity data of backcross 2. Analyses via generalised linear model (all parities and backcross 1 and 2 simultaneously) showed evidence for QTL located on chromosome 3 and 4. Generalised linear models and generalised linear mixed models produced similar results.

**Keywords:** Mapmaker/QTL, GLM, QTL, maximum likelihood, discrete data.

## INTRODUCTION

Many methods have been developed for quantitative trait locus (QTL) mapping. Some programs, like Mapmaker/QTL (Lincoln *et al.* 1993), have become widely used. This program is based on maximum likelihood (ML) and assumes a normal distribution of phenotype. It can not handle additional fixed effects such as parity or repeated records or data from more than one cross.

In this study, Mapmaker/QTL was compared with a generalised linear model (GLM) approach that allows additional fixed effects and more than one backcross (developed by Kayis *et al.* 1998). In addition, a generalised linear mixed model (GLMM) developed by Thomson (1998) to incorporate random animal effects was also evaluated. Unlike the MapMaker/QTL model, the GLM and GLMM approaches have been developed based in the Poisson distribution for modelling count data, rather than the normal distribution. In the GLM approach, within-animal correlation occurs as a result of a common (unobserved) QTL effect across all parities, while the GLMM has the additional common random animal effect across all parities..

## MATERIAL AND METHODS

**Data.** Two highly inbred mice strains, namely C57BL/6 ($P_1$) and IQ5 ($P_2$) were crossed to produce an $F_1$ and backcross (BC) progeny were performance tested. The details of mating conditions and rearing have been given by Silva (1994), who also kindly provided the data. There were 53 $BC_1$ (=$F_1$ female × $P_1$ male), and 49 $BC_2$ (=$F_1$ female × $P_2$ male) females, each with four completed litter records. Animals were genotyped by Silva (1994) and Maqbool (1998) for 65 genetic markers on chromosomes 1, 2, 3, 4, 5, 7, 8, 9, 10, 12, 13, and 14.

**Statistical models to analyse data.** Mapmaker/QTL can search only for QTL that occur between flanking markers; it can not search for QTL terminal to the first and last marker on a chromosome. It produces a LOD score for each putative QTL location on the chromosome between genetic markers. A LOD score exceeding a threshold value indicates significant evidence for QTL in that region. It

also is able to analyse data from an $F_2$ intercross design and it can handle missing data such as marker genotypes in a given analysis.

The GLM model was developed for discrete data such as litter size and assumes a Poisson distribution. It includes fixed effect parameters for parity ($\beta_i$, i = 1,...,4), QTL effects ($\gamma_k$, $k$ = $QQ$, $Qq$, $qq$, $qQ$), and uses information from $BC_1$ and $BC_2$ simultaneously to increase power to detect a QTL. The model is fitted using a ML method with a separate fit at each possible QTL position. It uses the flanking-marker method between markers and a single-marker method for the rightmost and leftmost marker on the chromosome. The log-likelihood value (log L) is obtained at each putative QTL location on the chromosome. A 95% confidence interval (95% CI), via the likelihood ratio test, for QTL location is obtained.

The GLM model can be extended to a GLMM by allowing for the inclusion of random animal effects. An approach outlined by Thomson (1998) is a semi-parametric model, in that it does not require a strictly Poisson distribution of litter size, and this method has been used in the current application. A generalised estimating equation approach has been used to fit these models, as ML is not possible unless the model is fully specified parametrically.

**Comparisons.** Litter size data were first analysed using Mapmaker/QTL and GLM (ie. assuming distribution of phenotype normal versus Poisson). To enable a fairer comparison of these methods, the GLM method was modified to analyse single parity and single BC data. The GLM method can not handle missing data, so genetic markers with missing genotypes were removed. Comparison were made of the likelihood profiles for Mapmaker/QTL and GLM. In addition, the GLM and GLMM (fixed effect versus mixed) models were compared, analysing all parities and backcrosses simultaneously.

**RESULTS**

In general, Mapmaker/QTL and GLM produced similar likelihood profiles with similarly located peaks for all analyses (4 parities __2 BCs __12 chromosomes = 96 analyses). For illustration, interval maps for the Mapmaker/QTL and GLM models are shown in Figure 1, for the only case where each produced significant evidence for a QTL (chromosome 3: parity 4, BC2). This analysis indicates that there is a QTL for litter size located between the genetic markers D3Mit24 and D3Mit12. There was no evidence for QTL on any other chromosomes in single parity / single BC analyses.

Simultaneous analysis of litter size from all parities for $BC_1$ and $BC_2$ via the GLM method showed some evidence for apparently the same QTL on chromosomes 3 and for another on chromosome 4. For chromosome 3, the estimated position is the same as for the single parity analysis. For chromosome 4, the QTL is most likely located between the genetic markers D4Mit37 and D4Mit204. The interval maps for chromosome 3 and 4 based on all parities and both backcrosses are shown in Figures 2 and 3 respectively. For comparison, the GLMM procedure was fitted to the same data, and in general revealed the same locations of QTL as the GLM procedure.

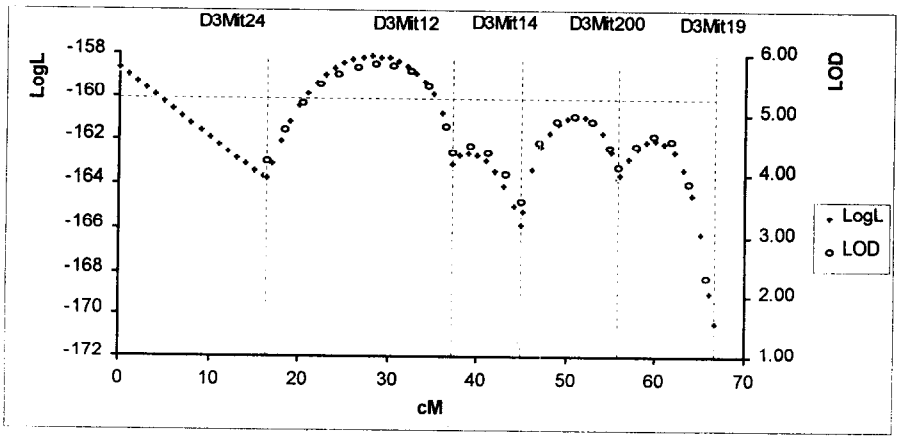


**Figure 1. Comparison of interval maps for chromosome 3 for data from parity 4 of $BC_2$. Dots and circles show LogL from GLM (assuming litter size has a Poisson distribution) and LOD from Mapmaker/QTL (assuming litter size has a normal distribution) respectively. Vertical dashed lines show genetic marker locations and the horizontal dashed line is the 95% CI for QTL location via the GLM method.**

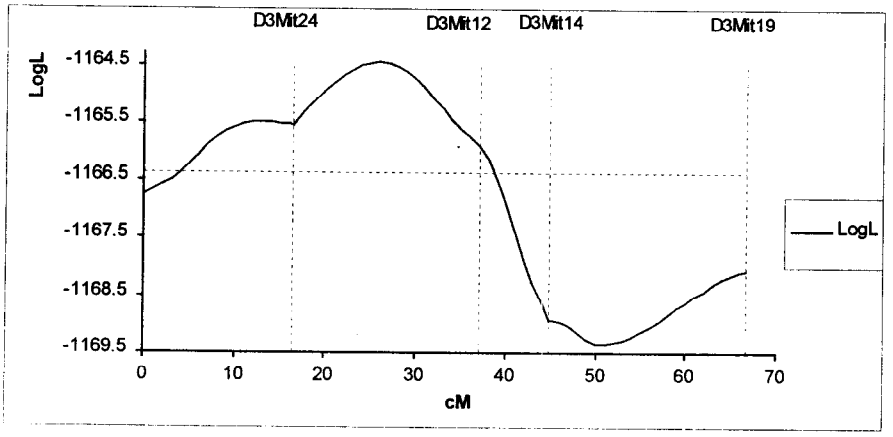


**Figure 2. Graph of interval map for chromosome 3, using all parities and both backcrosses together, via the GLM method. Vertical dashed lines show genetic marker locations and horizontal dashed line is the 95% CI for QTL location.**

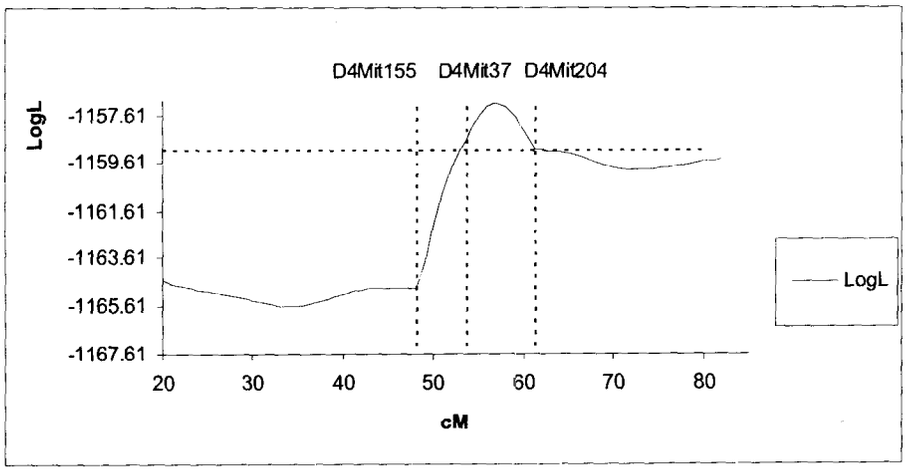


**Figure 3. Graph of interval map for chromosome 4 using all parities and both backcrosses together, via the GLM method. Vertical dashed lines show genetic marker locations and horizontal dashed line is the 95% CI for QTL location.**

## DISCUSSION

Despite the contravention of the assumption of normality, Mapmaker/QTL was apparently as efficient as the GLM model in detecting litter size QTL based on single parity data from a single backcross. However, the advantage of the GLM model lies in its capacity to simultaneously analyse multiple parities and both backcrosses, which enabled an additional QTL on chromosome 4 to be found. The GLMM method has produced similar results to the GLM procedure. However, simulation studies have revealed that GLMM procedures provide less biased estimates of QTL effects compared with GLM methods, even though the estimated location of the QTL is very similar. While the GLMM method is computationally intense, a fully parametric approach is currently being developed, which should be able to be implemented using standard GLMM or even LMM (linear mixed model) software.

## REFERENCES

Kayis, S.A., Thomson, P.C., and Nicholas, F.W. (1998) *45th Annual Conference Genetics Society of Australia*, Sydney, Australia. p. 67

Lincoln S.E., Daly M.J., and Lander E.S (1993) Constructing genetic maps with MAPMAKER/QTL 1.1. Whitehead Institute Technical Report 3rd edition

Maqbool N.J. (1998) Ph.D Thesis (submitted) Faculty of Agriculture, The University of Sydney

Silva, L. P. (1994) Ph.D Thesis, Faculty of Agriculture, The University of Sydney

Thomson, P.C. (1998) *Proc.6th World Congr. Genetics Appl. Lives. Prod.* **26**: 233