

# Evaluating the Performance of a Block Updating McMC Sampler in a Simple Genetic Application

N.A. SHEEHAN \*      B. GULDBRANDTSEN<sup>1</sup>  
D.A. SORENSEN<sup>1</sup>

Markov chain Monte Carlo (McMC) methods have provided an enormous breakthrough in the analysis of large complex problems such as those which frequently arise in genetic applications. The richness of the inference and the flexibility of an McMC Bayesian approach in terms of design, data structure that can be analysed, and models that can be posed, is indisputable. However, despite the strengths of the Bayesian approach, it is important to acknowledge that there are other, often easier, ways of tackling a problem. This is so, especially when simpler, qualitative answers are sought, such as presence or absence of a quantitative trait locus. We critically evaluate the behaviour of a Bayesian McMC block sampler for the detection of a quantitative trait locus by linkage with marker data, and compare it with a traditional least squares method. Some practical issues are illustrated by discussing the pros and cons of a Bayesian block updating sampling scheme versus the least squares method in the context of a simple genetic mapping problem. Depending on the focus of analysis, we show that the McMC sampler does not always outperform the simpler approach from a frequentist perspective, and, more to the point, may not always perform appropriately in any particular replication.

**Key Words:** Markov chain Monte Carlo methods, Bayesian analysis, least squares, block updating, mixing, genetic mapping

1

## 1 INTRODUCTION

It is well known that algorithms for performing exact probability and likelihood computations, such as the statistical genetics *peeling* method for pedigree

---

\*Address for correspondence: Departments of Health Sciences and Genetics, University of Leicester, 2nd Floor Adrian Building, University Road, Leicester LE1 7RH, UK

<sup>1</sup>Department of Animal Breeding and Genetics, Danish Institute of Agricultural Sciences, PB 50, DK-8820 Tjele, Denmark

applications (Elston & Stewart 1971, Cannings, Thompson & Skolnick 1978, Lander & Green 1987), and the more general local computation algorithms (Lauritzen & Spiegelhalter 1988, Cowell, Dawid, Lauritzen & Spiegelhalter 1999) developed by the expert systems community for calculations on Bayesian networks, all break down in practice when the underlying graph becomes too complex. Such complexity arises when the graph has too many interconnecting undirected cycles or *loops* forcing large cutsets in a peeling sequence, or large cliques in the corresponding triangulated graph, both of which can lead to insupportable storage requirements (Thompson 2000, Lauritzen & Sheehan 2003). In genetic applications, exact calculations are computationally intensive, either when the pedigree graph itself is complex or when the graphical model induced by both the structure and genetic model under consideration is complex. When exact methods fail, probabilities and likelihoods of interest must be approximated. Sometimes, pedigree information is sacrificed completely and exact analyses carried out on a simple structure extracted from the original, such as the half-sib design we will discuss below. Sometimes likelihoods can be approximated by numerical integration (Knott & Haley 1992) or *iterative* peeling methods (Janss, Van Arendonk & Van der Werf 1995). Sometimes the original structure is simplified less drastically by cutting a sufficient number of loops to facilitate computation (Wang, Fernando, Stricker & Elston 1996). Alternatively, the complexity of the problem is preserved and Markov chain Monte Carlo (McMC) methods (Hastings 1970, Metropolis, Rosenbluth, Rosenbluth & Teller 1953) employed to provide estimates of the required quantities. (See Thompson (2001) for an overview.)

Although McMC methods are not generally restricted by the complexity of the underlying graph, the estimates they yield are often unreliable. In particular, the Markov chain defined by the popular single-site Gibbs sampler (Geman & Geman 1984) may be reducible in discrete genetics applications (Sheehan 1992, Sheehan & Thomas 1993) and some relaxation of the probability model is then required which may considerably enlarge the state space of the sampler (Gilks, Clayton, Spiegelhalter, Best, McNeil, Sharples & Kirby 1993). This, in turn, can lead to inefficiency due to large amounts of sampling time being spent in illegal states. Even when the chain is theoretically irreducible, some parts of the search space may be separated by configurations of extremely low probability and thus communication between these states, or *mixing*, can be poor. Single-site updating schemes can perform particularly badly in such cases. Indeed, fear of using heated or relaxed samplers has led to some confusion in the literature over whether a particular application defines an irreducible chain for the single-site Gibbs sampler or not (Cannings & Sheehan 2002). The main problem, however, is that these sampling methods have not been sufficiently well tested on large complex problems and are thus viewed with some suspicion in practice (Hoeschele, Uimari, Grignola, Zhang & Gage 1997). Despite the abundance of McMC convergence diagnostics (Cowles & Carlin 1996) there is no certainty as to when the equilibrium distribution has been reached and how long the chain must be in order to get a representative sample.

The main thrust of this paper is to critically evaluate the behaviour of a Bayesian McMC block sampling scheme for the detection of a quantitative trait locus by linkage with marker data when a few specific aspects of the problem change. McMC applications do not have to be Bayesian (see Thompson (1994) for example) but they frequently are in animal genetics applications (Hoeschele & VanRaden 1993, Thaller & Hoeschele 1996, Sillanpää & Arjas 1998, George, Mengersen & Davis 2000, Yi & Xu 2000). We have restricted our investigations to the simple pedigree structure of the half-sib design, and to a trivial mapping scenario in order to properly assess sampler behaviour. Nonetheless, the problem is sufficiently complicated that exact methods no longer apply and hence correct answers on a real data set are unavailable for comparison. Simulated data are consequently required in order to properly assess the efficiency of the sampler and we have carried out a study on a suite of simulated datasets. One question that is rarely asked in these complex applications but which should always be pertinent is whether a full Bayesian McMC analysis is sufficiently better than a simpler approximation as to make it worth the extra computational effort required. When, for example, is the approximation afforded by a traditional least squares method such as that described in Haley & Knott (1992), almost as good as the full McMC analysis and when is it considerably worse? There is currently a tendency to use the heaviest computational tools available for all applications, regardless of whether they are really necessary or not. The underlying assumption is that the complex analysis will always be at least as good as the simpler one. This is known not to be the case in many other areas of applied statistics where an overly complex analysis can sometimes be itself misleading. Our findings serve as a warning that the same may be true for McMC methods—and not just for genetic applications. In particular, depending on the focus of the analysis, we show that our seemingly “perfect” sampler does not always outperform the simpler approach from a frequentist perspective and, more to the point, may not always perform appropriately in any particular situation.

In the interests of readability, Section 2 will provide a brief review of the basic genetic terminology required for this paper (a more rigorous treatment can be found in Thompson (2000) or Sham (1998), for example) and the mapping problem. Section 3 will outline the Bayesian block updating sampling scheme described in fuller detail in Sheehan, Guldbrandtsen, Lund & Sorensen (2002).

## 2 THE GENETIC MAPPING PROBLEM

The motivating example for this paper is the problem of detecting the location of an autosomal quantitative trait locus by linkage with marker data. We will follow Falconer & Mackay (1996) and refer to *quantitative* traits as traits that are determined by the combined effects of genes at many loci (*polygenic variation*) superimposed with some additional non-genetic continuous variation. We define a *quantitative trait locus* (QTL) to be a segment of chromosome which affects a quantitative trait but not on such a scale as to cause an observable discontinuity and which is hence not detectable using methods for Mendelian traits. By

*phenotype* we will mean any observable characteristic, be it the actual genotype itself, a disease state (e.g. affected or normal) or some value for a continuous quantity. We will also use the common partitioning of a phenotype,  $y$ , into a genetic and a non-genetic component,  $y = g + e$ , where  $g$  is the effect due to the actual genotype of the individual at the locus (or loci) affecting the trait (the *genotypic value*) and  $e$  is an environmental deviation from this effect (Falconer & Mackay 1996, Lynch & Walsh 1998). It is assumed that  $E(e) = 0$  implying that the expected phenotype (for a given genotype) is the relevant genotypic value:  $E(y|g) = g$ .

For a single diallelic locus with alleles  $A_1$  and  $A_2$ , suppose that the homozygote genotype  $A_1A_1$  confers a genotypic value  $\mu + a$ , the other homozygote genotype  $A_2A_2$  confers  $\mu - a$  and the heterozygote genotype  $A_1A_2$  confers a genotypic value  $\mu + d$ , where  $\mu$  is an intercept parameter (see Figure 1). The dominance effect  $d$  is the deviation between the average genotypic value of the homozygote genotypes, and statistically, measures interaction between alleles within the locus. If  $A_1$  and  $A_2$  act in a completely additive fashion,  $d = 0$  and the heterozygote genotype is equal to  $\mu$ . Allele  $A_1$  is dominant if  $d > 0$  (in which case,  $A_2$  is the recessive allele), while  $d < 0$  implies that  $A_2$  is dominant over  $A_1$  (and  $A_1$  is the recessive allele). We say that we have complete dominance if  $d = -a$  or  $d = a$ , over-dominance if  $d > a$  and under-dominance if  $d < -a$ .

FIGURE 1 ABOUT HERE

The focus of a linkage analysis is the *recombination fraction*,  $r$ , which is the probability of a recombination event. There is a one-to-one correspondence (Ott 1999) between the recombination fraction and the *genetic map distance*, defined as the expected number of crossovers to occur between two loci in a gamete and measured in units called Morgans, or centiMorgans (Haldane 1919). We will use Haldane’s mapping function to relate the two. While it is natural to work with recombination rates when taking account of linkage, map distances are often more straightforward measures. We will refer to map distances (as differences between map positions) throughout this paper but note that the terms are interchangeable.

## 2.1 THE DESIGN, MODEL AND NOTATION

FIGURE 2 ABOUT HERE

In animal breeding applications, data are frequently collected on simple structures despite the common availability of more detailed pedigree information. Although animal breeders have had decades of experience with fairly complex pedigrees, existing methods cannot cope with analyses of complex genetics models, such as we are considering here, on the large complex pedigrees which often feature in animal populations. Even though current MCMC methods are very flexible with regard to the structure of the problem, the computational intensity and perceived unreliability of these methods are obstacles to their widespread acceptance as standard methods for general pedigrees. In practice, in human, animal and plant applications, analyses are still routinely carried out on simple structures such as nuclear families, sib-pairs and inbred line cross designs. For

this reason, although our methods are entirely general, we will follow Sheehan et al. (2002) and concentrate on the half-sib design as it enables comparison with the commonly used regression method of Section 4 and is simple enough to permit investigation into the behaviour of our McMC sampling scheme.

In many managed populations, especially those in which artificial insemination is practised, one male individual can have an enormous number of offspring with potentially as many different females. The population then contains large families of these half-siblings who share the same father, or sire. In a typical half-sib design, there are 10 to 15 sires, each with up to 100 offspring. For a QTL analysis, we would expect to have marker data on the sire and his offspring (although in practice, some of this could be incomplete) and records for the trait, deemed to be partly determined by the QTL, on offspring only. Detecting a QTL for milk yield in dairy cattle would be an example of such an analysis. Note that detection of a recombination event between the two markers requires that the parent individual be heterozygous at both loci. In this case, we say that the individual is informative for linkage at these loci. Figure 2 shows a half-sib design for a case with one sire and four offspring. As indicated, the mothers, or dams, of the offspring are all assumed to be unobserved and are represented as founders of the pedigree. This is in accordance with the assumptions of the half-sib design: the dams are all distinct, unrelated and assumed to be uninformative for the trait. Genetic inference is therefore centred on segregation from the common sire.

FIGURE 3 ABOUT HERE

For our purposes, we will assume that in addition to an effect from the polygenes, there is a single QTL coding for our continuous trait and our interest is on whether or not this QTL is located between two particular markers. For greater transparency, the markers are held to be diallelic with alleles  $M, m$  at the “M-locus” and  $N, n$  at the “N-locus” with corresponding population allele frequencies  $p_M, p_m = 1 - p_M$  and  $p_N, p_n = 1 - p_N$ , respectively. We note that this assumption is not required for either of our methods and actually makes the mapping problem more difficult (Knott & Haley 1992) as marker heterozygosity, and hence informativeness, increases with the number of alleles in the system. Furthermore, as is consistent with a model where only one allele is believed to influence the trait, we will assume the QTL to be diallelic with alleles  $Q$  and  $q$  and allele frequencies,  $p_Q$  and  $p_q = 1 - p_Q$ . For the half-sib design, it is common to presume that the population from which sires have been sampled is in both Hardy-Weinberg and linkage equilibrium at the three loci. By the first, we mean that given two alleles at a locus with frequencies  $p_1$  and  $p_2$ , the three genotype frequencies are  $p_1^2, 2p_1p_2$  and  $p_2^2$ . By the second, we mean that the haplotype frequencies are a product of the relevant locus allele frequencies e.g. for the  $Mqn$  haplotype in our example,

$$P(Mqn) = p_M p_q p_n.$$

These are strong assumptions for what is in reality a non-random mating population with strong selection but they are commonly made for any analysis on such a design. The map locations of the marker loci,  $\lambda_M$  and  $\lambda_N$  (with

$\lambda_M < \lambda_Q < \lambda_N$ ) are known so we know the genetic distance and thus the recombination fraction between them. The QTL location,  $\lambda_Q$ , is unknown and has to be estimated or, equivalently, the recombination fractions between the QTL and each marker have to be estimated. If recombinations in non-overlapping intervals are held to be independent (i.e. there is no *genetic interference*), only one of the two is required to parameterise the problem. This simple mapping scenario is illustrated in Figure 3.

We will use the notation of Sheehan et al. (2002) for fitting a normal linear mixed model (Gelfand, Hills, Racine-Poon & Smith 1990) to this QTL-mapping problem on a half-sib design and incorporating linkage with marker data. We denote the phenotypic record for the trait on offspring  $j$  of sire  $i$  by  $y_{ij}$  where  $i = 1, 2, \dots, k$ ,  $j = 1, 2, \dots, n_i$  and the full phenotype record vector by  $\mathbf{y}$  of dimension  $n = n_1 + n_2 + \dots + n_k$ . Let  $\mathcal{M}_i$  denote the (known) marker information at both marker loci for sire  $i$ ,  $i = 1, \dots, k$ , while  $\mathcal{M}_{ij}$  gives the marker information for his  $j^{th}$  offspring,  $j = 1, \dots, n_i$ . We let  $\mathbf{Q}_i$  and  $\mathbf{Q}_{ij}$  denote the (unobserved) QTL genotypes for the  $i^{th}$  sire and  $ij^{th}$  offspring, respectively where these can be any of the three distinct unordered types  $\{\mathbf{Q}\mathbf{Q}, \mathbf{Q}\mathbf{q}, \mathbf{q}\mathbf{q}\}$ . The normal linear mixed model will be written as

$$y_{ij} = s_i + q_{ij} + e_{ij} \quad (1)$$

with

$$q_{ij} = \begin{cases} \mu_1 \equiv \mu + a & \text{if } \mathbf{Q}_{ij} = \mathbf{Q}\mathbf{Q} \\ \mu_2 \equiv \mu + d & \text{if } \mathbf{Q}_{ij} = \mathbf{Q}\mathbf{q} \\ \mu_3 \equiv \mu - a & \text{if } \mathbf{Q}_{ij} = \mathbf{q}\mathbf{q} \end{cases}$$

acting as the “*fixed* effects” describing the effect of the QTL genotype on the trait phenotype as consistent with Figure 1. In addition to the QTL effect, we must account for the effect due to the polygenes which are also segregating from the sire. These are the *random* effects, or *sire* effects,  $\mathbf{s} = (s_1, \dots, s_k)'$ , where  $s_i$  denotes the average additive genetic effect of the  $i^{th}$  sire on the phenotypes of his offspring and which cannot be explained by the QTL. It is assumed that these are normally and independently distributed with common variance,  $\sigma_s^2$ , the *sire variance component*:

$$\mathbf{s} | \sigma_s^2 \sim N(\mathbf{0}, \mathbf{I}\sigma_s^2), \sigma_s^2 \in \mathbb{R}_+.$$

The polygenic variance,  $\sigma_u^2$ , is defined as the total additive genetic variance unexplained by the QTL. We have that  $\sigma_s^2 = \frac{1}{4}\sigma_u^2$  since half the genes of an offspring are shared with its sire. Finally,

$$e_{ij} \sim N(0, \sigma_{res}^2)$$

where  $\sigma_{res}^2 \in \mathbb{R}_+$  is the residual variance component and accounts for all the variation in the data which cannot be explained either by the sire effect or the QTL. In terms of the above,  $\sigma_{res}^2 = \frac{3}{4}\sigma_u^2 + \sigma_e^2$  where  $\sigma_e^2$  is the environmental variance component. We note that the above model (1) is the same as that for full sibs described by Knott & Haley (1992).

A fully Bayesian approach to modelling the above problem requires that we assign prior distributions on the allele frequencies, QTL map position, QTL

genotype effects, sire effects and sire and residual variance components. These are as follows. Allele frequencies are all taken to derive from a Beta distribution with known parameters. Thus, for example

$$\pi(p_Q) \sim \text{Beta}(p_Q|a, b) \text{ where } a \text{ and } b \text{ are known.}$$

The map location for the QTL,  $\lambda_Q$ , is assumed to have a uniform distribution over the interval  $(\lambda_N - \lambda_M)$  while the “fixed” effects  $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)'$  are taken to have an improper prior distribution

$$\pi(\boldsymbol{\mu}) \propto \text{constant.}$$

As a prior distribution on the sire variance component  $\sigma_s^2$  we assume a scaled inverted chi-square distribution with  $\nu_s$  degrees of freedom and scale factor  $S_s$ , both known:

$$\pi(\sigma_s^2) \propto (\sigma_s^2)^{-(\frac{\nu_s}{2}+1)} \exp\left(\frac{-\nu_s S_s}{2\sigma_s^2}\right). \quad (2)$$

The residual variance component  $\sigma_{res}^2$  is also assumed to have a scaled inverted chi-square prior distribution with known degrees of freedom  $\nu_{res}$  and scale factor  $S_{res}$ . (See Sheehan et al. (2002) for further details.)

### 3 AN MCMC BLOCK UPDATING SAMPLING SCHEME

Even without the Bayesian interpretation and despite the simplicity of both the pedigree and the mapping problem, the model described above is too complicated for an exact likelihood approach to estimating the parameters of interest. In addition to having to sum over all possible QTL genotypes and marker phases, an industrial-sized half-sib design induces correlations amongst up to 100 offspring of a sire via the shared random effect  $s_i$  thus creating *loops* in the underlying graphical model representing the problem, even though the corresponding pedigree is unlooped (Sheehan et al. 2002). Local computation and peeling algorithms all fail when there are too many loops in the relevant graph. Thus, we are already in a situation where alternative methods for estimating the quantities of interest, such as MCMC methods, must be entertained. With the full Bayesian approach, such alternatives are even more essential. It is well known that the single site Gibbs sampler (Geman & Geman 1984) popularly used in many Bayesian applications can mix very slowly in these complex models involving both discrete and continuous nodes, even when the sampler is theoretically irreducible (Janss, Thompson & Van Arendonk 1995, Heath 1997, Jensen & Kong 1999, Lund & Jensen 1999). Some kind of blocking or joint updating of variables is hence required in order to sample more efficiently.

Jensen, Kjærulff & Kong (1995) were the first to call up a peeling calculation on a set of discrete genotypes on a large complex pedigree from within a Gibbs

sampling framework. Their sampler repeatedly updates large, overlapping pre-defined blocks of genotypes conditionally on the current values of the usually small number of remaining genotypes. Mixing has been shown to be fast but the sampler is not necessarily irreducible. Other blocking samplers include the “L-sampler” of Heath (1997) which permits multipoint linkage analysis by updating all genotypes at each locus conditionally on the current values at all other loci. This requires that the pedigree can be peeled at a single locus. The “M-sampler” of Thompson & Heath (2000), on the other hand, peels along the chromosome rather than the pedigree and updates all the meioses in a single individual in one block. The L-sampler is restricted by pedigree complexity whereas the M-sampler is restricted by the number of loci involved. Various methods for alternating between the two have been shown to greatly improve mixing (Thompson 2000, Thomas, Gutin, Abkevich & Bansal 2000).

FIGURE 4 ABOUT HERE

We now briefly review the blocking sampler of Sheehan et al. (2002) but first introduce some further notation. Information on phase is required for linkage so it is convenient to use *meiosis* or *segregation* indicators (Thompson 1994, Sobel & Lange 1996, Thompson 2001) to indicate which of the two parental gametes is transmitted to the offspring. The segregation indicators are binary variables taking the value 0 if a copy of the maternal gamete is inherited from the parent and 1 if the paternal gamete is copied. Note that this labelling is random for the sires as the parental origins of the sire gametes are unspecified. Thus, for example,  $S_{ij}^M$  describes the transmission of gametes from sire  $i$  to his  $j$ th offspring at the “M-locus” and in the absence of any other information, we assume that it takes the value 0 or 1 with probability 0.5 i.e. inheritance is Mendelian at the first locus. Let  $M_{(ij,1)}$  denote the actual gamete inherited by individual  $ij$  from its sire  $i$ . Its allelic value depends on both alleles in  $i$  and on the value of the segregation indicator determining which one is inherited. The corresponding maternal allele,  $M_{(ij,0)}$  is drawn randomly from the population according to our assumptions about allele frequencies. The observed marker genotype,  $M_{ij}$ , is determined by the values of these two. Gametes and segregation indicators for the other loci are defined similarly for individuals  $i$  and  $ij$  with the one difference being that subsequent segregation indicator values depend on the preceding segregation indicator because of linkage. This dependency is a function of the recombination fraction between the two loci, or equivalently, of the QTL map location  $\lambda_Q$ .

The graph describing the full Bayesian analysis on the simple mapping problem of Section 2 is displayed in Figure 4 for a single sire and two of his offspring. The computational complexity is clearly illustrated indicating why linkage calculations are challenging. Although there are no loops in the pedigree of Figure 2 there are many loops in the corresponding Bayesian network for this three-locus mapping problem. The block updating MCMC sampling scheme proceeds as follows. All the discrete nodes in the graph comprising all genes, unobserved genotypes and segregation indicators (i.e. the grey nodes in Figure 4) are jointly sampled, conditionally on the current values at the remaining nodes. QTL genotype effects,  $\mu_i$ ,  $i = 1, 2, 3$  and sire effects  $s_i$ ,  $i = 1, \dots, k$  are block sampled from their full joint



conditional distribution, given values elsewhere. The variance components,  $\sigma_s^2$  and  $\sigma_{res}^2$ , and allele frequencies  $p_M, p_q$  and  $p_N$  are each sampled from their respective full conditional distributions using a single-site Gibbs sampler. The QTL map location parameter,  $\lambda_Q$ , is updated using a Metropolis-Hastings algorithm as its full conditional distribution does not have a standard form. Further details including full derivation of all the relevant conditional distributions and sampling steps are given in Sheehan et al. (2002).

We should emphasise at this point that, as MCMC samplers go, this is a very sophisticated one. It jointly carries out the “L” and “M” steps (i.e. locus and meiosis peeling), as described above, automatically deals with variation between families and with its elegant handling of the problem of unknown phase, it is essentially “perfect” in its exact treatment of the discrete part of the model. Of course, this is only possible because of the simplicity of the problem (i.e. simple pedigree *and* small number of loci) but we emphasise that, as a result, mixing problems with respect to the genotypic configuration should not be an issue in this application. The actual implementation of the sampler was via a graphical model representation and invoked the *random propagate* algorithm (Andersen, Olesen, Jensen & Jensen 1989) of the HUGIN package (<http://www.hugin.com>) to jointly update all the discrete variables. We note that the particular choice of representation is independent of our discussion here.

## 4 A REGRESSION APPROACH

A much simpler approach to the mapping problem of Section 2 is provided by a method based on the least squares principle (Haley & Knott 1992, Martinez & Curnow 1992, Haley, Knott & Elsen 1994). It only applies to simple designs like the half-sib design, does not use all the information in the data, does not deal with random polygenic effects and does not provide estimates of many of the parameters of interest other than the QTL position. Nonetheless, such methods are frequently used in practice as they are computationally straightforward and quite robust to violations of the assumptions (Knott 2005). If all that is required is whether or not the data provide evidence for a QTL affecting a particular trait, an approximate regression analysis may be just as efficient as an MCMC approach in certain circumstances. It would certainly be a lot quicker. It is therefore important to ask the question: when can we really expect to gain from the full MCMC analysis? Although least squares methods are standard in the animal breeding literature, the details of the method pertaining to this particular design are not easily found and so we will derive them here.

Consider our QTL with alleles  $Q$  and  $q$  and corresponding allele frequencies  $p_Q$  and  $p_q = 1 - p_Q$ . The three genotypes are  $QQ$ ,  $Qq$  and  $qq$  with genotypic effects  $\mu + a, \mu + d$  and  $\mu - a$  as given above in Equation 1. As before, we let  $s_i$  denote the expected additive polygenic contribution of the  $i^{th}$  sire to his offspring *excluding* the effect of the QTL. Ignoring the marker data and the polygenes for now, consider the expected genotypic effect of offspring which inherit a  $Q$  allele from their sire. These offspring can be of only two possible genotypes,  $QQ$  and

$Qq$ , and have inherited the other allele from the dam. By the design assumptions, inheritance from the dams is equivalent to drawing an allele at random from the general population. Hence, we have

$$\begin{aligned}
E(y_{ij}|Q \text{ from sire } i) &= E(y_{ij}|Q_{ij} = QQ) \Pr(Q_{ij} = QQ|Q \text{ from sire } i) \\
&+ E(y_{ij}|Q_{ij} = Qq) \Pr(Q_{ij} = Qq|Q \text{ from sire } i) \\
&= (\mu + a)p_Q + (\mu + d)p_q \\
&= \mu + (ap_Q + dp_q).
\end{aligned}$$

Likewise, the expected effect in an individual receiving a  $q$  allele from its sire is:

$$\begin{aligned}
E(y_{ij}|q \text{ from sire } i) &= (\mu + d)p_Q + (\mu - a)p_q \\
&= \mu + (dp_Q - ap_q).
\end{aligned}$$

In the absence of marker data and conditional on the sire's polygenic effect and QTL genotype, the expected value of an offspring of sire  $i$  is the sum of the expected additive polygenic contribution from sire  $i$  and the average of the expected genotypic effects due to inheritance of the different possible alleles at the QTL. Specifically,

$$E(y_{ij}|Q_i, s_i) = \begin{cases} s_i + \mu + (ap_Q + dp_q) & \text{if } Q_i = QQ \\ s_i + \mu + \frac{1}{2}(a(p_Q - p_q) + d) & \text{if } Q_i = Qq \\ s_i + \mu + (dp_Q - ap_q) & \text{if } Q_i = qq \end{cases} . \quad (3)$$

Now, without loss of generality, we arbitrarily label the two QTL alleles carried by the  $i^{th}$  sire as  $Q_1$  and  $Q_2$ , each of which can be either  $Q$  or  $q$ . Given the marker information on sire  $i$  and on his  $j^{th}$  offspring,  $(\mathcal{M}_i, \mathcal{M}_{ij})$ , we can calculate the probability that the  $Q_1$  allele is transmitted to offspring  $ij$ . This probability is a function of the marker gene frequencies, of the marker phase in the sire, of the recombination rate between the marker loci and of the recombination rate between the marker locus and the putative QTL. (See Haley et al. (1994), for example, for a full derivation.) We denote this probability  $\pi_{ij} \equiv P(Q_1 \text{ transmitted} | Q_i, \mathcal{M}_i, \mathcal{M}_{ij})$ . The probability that the  $Q_2$  allele is transmitted to offspring  $ij$  is  $1 - \pi_{ij}$ . If the sire is homozygous at the QTL such that  $Q_1 = Q_2$  then the offspring receives the same version of the QTL regardless of whether  $Q_1$  or  $Q_2$  is transmitted. Therefore the conditional expectation of  $y_{ij}$  does not depend on this transmission. Marker information only affects the expectation of the offspring's phenotype by providing information about transmission of  $Q_1$  and  $Q_2$ . Hence, if the sire is homozygous at the QTL, markers carry no information about the expectation of the offspring's phenotype. In contrast, for heterozygous sires (i.e.  $Q_1 \neq Q_2$ ) the expectation of  $y_{ij}$  *does* depend on the segregation at the QTL and hence on the marker information. Specifically, when  $Q_1 = Q$  and

$Q_2 = q$ , we deduce from (3) that

$$\begin{aligned}
E(y_{ij}|s_i, Q_i = \mathbf{Qq}, \mathcal{M}_i, \mathcal{M}_{ij}) &= s_i + \pi_{ij}E(y_{ij}|Q \text{ from sire } i) \\
&+ (1 - \pi_{ij})E(y_{ij}|q \text{ from sire } i) \\
&= s_i + \pi_{ij}(\mu + (ap_Q + dp_q)) \\
&+ (1 - \pi_{ij})(\mu + (dp_Q - ap_q)) \\
&= s_i + \mu + (dp_Q - ap_q) + \pi_{ij}(a + d(p_q - p_Q)) \\
&- E(y_{ij}|Q_i = \mathbf{Qq}, s_i) + E(y_{ij}|Q_i = \mathbf{Qq}, s_i) \\
&= \alpha_i + (2\pi_{ij} - 1)\beta_i
\end{aligned} \tag{4}$$

where  $\alpha_i = s_i + \mu + \frac{1}{2}(a(p_Q - p_q) + d)$  and  $\beta_i = \frac{1}{2}(a + d(p_q - p_Q))$ . For the reverse labelling with  $Q_1 = q$  and  $Q_2 = Q$ :

$$\begin{aligned}
E(y_{ij}|s_i, Q_i = \mathbf{qQ}, \mathcal{M}_i, \mathcal{M}_{ij}) &= s_i + \mu + (p_Q a + p_q d) - \pi_{ij}(a + d(p_Q - p_q)) \\
&= \alpha_i + (2\pi_{ij} - 1)(-\beta_i)
\end{aligned} \tag{5}$$

Note that the difference in sign between (4) and (5) above is entirely due to the arbitrary labelling of the QTL alleles. Furthermore, the distinction between the genotypes  $Qq$  and  $qQ$  in the conditioning is only to highlight this labelling as genotypes are, by definition, unordered.

A regression approach for detecting a QTL thought to affect a continuous trait from marker data on a half-sib design therefore involves regressing the observed offspring phenotypes for the trait onto the conditional probabilities,  $\pi_{ij}$ . As mentioned above, this is only informative when the sires are heterozygous at the QTL. If sire  $i$  is homozygous at both marker loci, then the markers carry no information about whether  $Q_1$  or  $Q_2$  is transmitted to the offspring. Hence,  $\pi_{ij}$  is equal to  $1/2$  for all offspring of sire  $i$ . When only one of the markers is heterozygous it can be shown that QTL position and QTL effect are confounded, such that it is not possible to distinguish a nearby QTL of small effect from a QTL of larger effect at a position further away.

For estimation purposes, this approach only yields estimates for specific functions of the parameters of interest, and not the parameters themselves. In the fully informative case, for example, we obtain a least squares estimate for the expected offspring genotypic value  $s_i + \mu + \frac{1}{2}(a(2p_Q - 1) + d)$  from the intercept term and  $\pm \frac{1}{2}(a + d(1 - 2p_Q))$  from the slope term (which is peculiar to each sire). In the absence of dominance (i.e. when  $d = 0$ ), the absolute value of the slope provides an unbiased estimate of the average effect of substituting a  $Q$  for a  $q$  allele for heterozygous sires, and thus estimates the increase in genotypic value due to the  $Q$  allele. For homozygous sires, the slope is an estimate of zero. Thus, a standard test for significance of a regression slope coefficient provides a test for a segregating sire. The sire polygenic effect is always confounded with the mean genotypic value in the intercept term.

## 5 SIMULATION STUDY RESULTS

A simulation study was conducted in which we attempted to compare the performance of the McMC block-updating analysis with the performance of the regression-type analysis described in Section 4. A number of different datasets were generated, all comprising ten sires each with one hundred offspring. Marker data at two diallelic marker loci 10cM apart were recorded for each sire and each offspring. The QTL was placed in the centre of the interval. Without loss of generality, we can assume that the “first” marker is at the beginning of the chromosome, so  $\lambda_M = 0$  cM,  $\lambda_Q = 5$  cM and  $\lambda_N = 10$  cM. To produce the marker and QTL genotypes, chromosomes were sampled from a population in linkage equilibrium. Phenotypic values were generated for each of the half-sib offspring assuming the model described by equation (1) of Section 2.1 so we have one thousand phenotype records for each dataset. The following parameter values were used to simulate the data:  $\sigma_s^2 = 7.5$ ,  $\sigma_{res}^2 = 92.5$ ,  $p_Q = 0.3$ ,  $p_M = p_N = 0.5$  and  $\mu = 0$ .

Five different scenarios were investigated and the genetic models underlying these are shown in Table 1. The models reflect no QTL ( $a = 0.0$ ), additive QTL of small ( $a = 2.5$ ), intermediate ( $a = 5.0$ ) and large ( $a = 10.0$ ) effects as well as dominant gene-action and intermediate QTL effect ( $a = d = 5.0$ ). As a measure of the relative size of the QTL effect, we define the genetic variance explained by the QTL to be  $\sigma_{qtl}^2$  and express this as a proportion of the overall *phenotypic* or marginal variance,  $\sigma^2$ , which is the sum of the sire variance, variance due to the QTL and the residual variance. In the notation of Section 2,  $\sigma^2 = \sigma_s^2 + \sigma_{qtl}^2 + \sigma_{res}^2$ . In terms of the fixed effects and the allele frequencies, it can be shown (Falconer & Mackay 1996) that

$$\sigma_{qtl}^2 = 2p_Q p_q (a + d(p_q - p_Q))^2 + (2p_Q p_q d)^2.$$

One hundred replicate data sets were simulated for each scenario in Table 1. All datasets were analysed with both methods.

TABLE 1 ABOUT HERE

### 5.1 REGRESSION ANALYSIS

The first step of the regression analysis involves estimation of marker allele frequencies. Allele frequencies are used to compute the conditional probabilities  $\pi_{ij}$  and to reconstruct the most probable linkage phase in the sires. The second step of the analysis is the estimation of the slope and the intercept of the regression equation for each sire, conditional on the estimated allele frequencies and phase. At the family level, a pseudo F-statistic is calculated for a sire family and the significance level of the regression slope assessed via a permutation test (Churchill & Doerge 1994) by repeatedly permuting offspring phenotypes within sire families. This is the test that was used to generate the results in Table 2 below. All sires corresponding to significant slope coefficients at the 5% level are classified as

heterozygous at the QTL locus and hence deemed to be segregating for a QTL. The remaining sires are classified as either QQ or qq homozygotes. Note that the regression method ignores uninformative sires (i.e. those homozygous at both markers) by giving them zero-weighting in the analysis. In practice, inferences are usually made at the dataset level (i.e. across all sire families) whereby a pseudo F-statistic is obtained by computing the ratio of the model mean square to the residual mean square. Significance is assessed via an appropriate permutation test. The test for the slope for each sire reveals whether the sire is segregating for the QTL. The overall test reveals whether there is a QTL segregating in the population. Results from this dataset level test are displayed in Table 5 in Section 5.2.

#### TABLE 2 ABOUT HERE

Table 2 shows the test-based classification rates (i.e. homozygous versus heterozygous) alongside the “true” simulated types. The individual sire acceptance and rejection probabilities have been averaged across families in each dataset and then across datasets. The probability of correctly rejecting the null hypothesis that the true slope coefficient is zero and thus detecting a segregating sire, increases on average with increasing genotypic effect,  $a$ , at the QTL from 3.2% to 64.4%. As we would expect, in the absence of any QTL effect, or when that effect is small, detection probabilities are very low. The power of detection for the dominance scenario **5-5** is slightly better (42.9%) than the power for the scenario **5-0** with same additive effect (32.7%).

We have seen in Section 4 that a point estimate of the increase in QTL genotypic value  $a$  due to the  $Q$  allele can be obtained from the slope term. This estimate is unbiased for the additive model (i.e. when  $d = 0$ ) but is confounded with the dominance effect and allele frequencies otherwise. Because of the difference in signs caused by the arbitrary labelling of the QTL alleles (equations 4 and 5), we have recorded the absolute value of this coefficient for each sire and taken averages and standard deviations of these values across families and datasets. The results are shown in Table 3.

#### TABLE 3 ABOUT HERE

A distinction has been made between sires which were classified as homozygotes by the permutation test (i.e. where the null hypothesis could not be rejected) and those which were classified as heterozygous. If the classification were perfect, the expected values should be close to zero for all the sires for which  $H_0$  was not rejected since homozygotes provide no information about this effect. Values should be close to 0, 2.5, 5 and 10, respectively, in the additive scenarios for those classified as heterozygous at the QTL. (Note that these expected values are not *exactly* 0, 2.5, 5 and 10 due to the fact that we are looking at absolute values.) As one might expect, the estimates for the sires classified as homozygotes are not really meaningful and are similar across the different scenarios. The estimates for the presumed heterozygous sires tend to be too high. This is because of the selection bias introduced by the testing criterion: only sires with very strong evidence for heterozygosity were categorised as such.

In summary the regression approach provides a qualitatively satisfactory

analysis for the presence/absence of a QTL. However, QTL effects are not well estimated and in the presence of dominance, QTL effects are confounded with gene frequencies (see Section 4).

## 5.2 MCMC ANALYSIS

The degrees of freedom and scale parameters of the scaled inverted chi squared prior distributions on the sire and residual variance components were arbitrarily set to  $\nu_s = \nu_{res} = 5$  and  $S_s = 7.5$  and  $S_{res} = 92.5$ , respectively. These settings give modal values for the residual and sire variances,  $\sigma_{res}^2$  and  $\sigma_s^2$ , of 66.1 and 5.36, respectively. (Recall from the beginning of Section 5 that the actual values for  $\sigma_{res}^2$  and  $\sigma_s^2$  chosen for the simulated datasets were 92.5 and 7.5). Prior distributions for allele frequencies were uniform on the interval  $[0, 1]$ , and improper uniform priors were assigned to the QTL effects,  $\mu_1, \mu_2$  and  $\mu_3$  (Section 2.1). As we are really interested in the magnitude of the additive effect  $a$  and the dominance effect  $d$  (Figure 1), rather than the individual combined genotype effects  $\mu_1, \mu_2, \mu_3$  themselves, we define the following functions of  $\mu_i$ ,  $\Delta_a = \frac{1}{2} |\mu_1 - \mu_3| \equiv |a|$  and  $\Delta_d = \mu_2 - (\mu_1 + \mu_3)/2 \equiv d$  (Section 2.1) and monitor these quantities instead. For each simulated dataset, the sampler was implemented with a single long chain of length 100,200. After a nominal burn-in period of 200, every tenth value was stored for each chosen parameter in the model giving a chain length of 10,000 for the resulting estimates. A little preliminary experimentation confirmed that this size of chain resulted in Monte Carlo coefficients of variation of estimates of posterior means of the chosen parameters ranging from 0.7 to 1.4%.

We report results in two sections. In section 5.2.1, the presentation is in the form of analyses explicitly constructed to produce output directly comparable to that of the regression analysis. In the following section (5.2.2), results are not directly comparable. In both sections, we report frequentist behaviour of chosen parameters (i.e. averaging across simulated datasets) and also various features of posterior distributions for a specific dataset.

### 5.2.1 QTL GENOTYPE CLASSIFICATION AND EFFECT SIZE VIA MCMC

Table 4 shows the posterior probabilities of the sire QTL genotypes corresponding to the underlying simulated values. For each scenario and for each of the three possible simulated sire types, an average posterior probability distribution of QTL genotypes is calculated as the proportion of samples assigned to each possibility, averaged over sires within a dataset and then across datasets. Also shown are the probabilities that each QTL genotype is classified as homozygote (QQ/qq). For instance in the **5-5** scenario, sires simulated as qq had on average a posterior probability of .297 of being qq and a posterior probability of .597 of being homozygous at the QTL.

TABLE 4 ABOUT HERE

In the **0-0** scenario where the QTL has no effect on the phenotype, the posterior probabilities of qq, Qq and QQ genotypes are very close to  $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ , independently of the simulated value, as would be expected. Things are not very

different in the **2.5-0** scenario where the QTL effect is small but, with increasing effect size, the posterior probability of correctly identifying a QTL heterozygote individual as such increases from .517 to .851. However, the posterior probability of the correct QTL genotype was low if the simulated value was homozygous. This remained true even in the **10-0** scenario where the QTL has a large effect on the phenotype explaining almost 30% of the variability (Table 1). On the other hand the posterior probability of identifying a sire as homozygote (i.e. either qq or QQ) did in fact increase with increasing effect size from 0.530 to 0.770 for qq and from 0.523 to 0.823 for QQ.

Although not directly comparable, these findings are broadly similar to those from the regression classification in Table 2: a large effect size is required in order to get reasonable classification rates. This is reflected in the posterior distributions of QTL genotype for each of the ten sires in a single dataset simulated for each of the three scenarios with a non-zero additive QTL effect. The general pattern in Figure 5 is very clear: as the size of the QTL effect increases, the marginal posterior distributions of QTL effects of the individual sires become more distinguishable. However, we also noted that genotypes can sometimes be poorly inferred, even in cases where the QTL effect is large as can be seen in subfigure (c) of Figure 5 where the marginal posterior distributions for some sires are relatively flat and uninformative. In this case, poor inference seemed to be caused by the fact that the sire effects happened to be extreme for the individuals in question and so the sampler is correctly reflecting the fact that information about the genotype effect has been dampened.

FIGURE 5 ABOUT HERE

Although the Bayesian approach is not directly comparable with the least squares analysis of Section 5.1, it can be viewed as so when the focus is on the same parameter and features across replicated datasets are of interest. In particular, for determination of sire QTL genotypes and effect sizes it is reasonable, and indeed practical, to ask if it is worth the extra computational effort. This question cannot be answered without making some kind of formal comparative statement and so we propose the following. As noted above, due to the non-identifiability of the sign of the difference between the effects of the homozygous QTL genotypes, we chose to study the absolute value of the difference,  $\Delta_a = \frac{1}{2} |\mu_1 - \mu_3|$ . To detect a QTL, the posterior distribution of  $\Delta_a$  would have to be well separated from 0. A hypothesis test for such a separation would give us results that are more comparable with those obtained from the pseudo dataset level F-test referred to in Section 5.1. From Sheehan et al. (2002), it can be seen that the marginal posterior distribution of genotype effects for this model is multivariate normal. Hence we expect the effect difference  $a$  to be univariate normal and its absolute value to look like a normal mixture random variable. In particular, if

$$a \sim N(\mu_a, \sigma_a^2), \quad (6)$$

then  $\Delta_a$  should have the density

$$\frac{1}{\sqrt{2\pi}\sigma_a} \left\{ \exp\left(-\frac{(\Delta_a - \mu_a)^2}{2\sigma_a^2}\right) + \exp\left(-\frac{(-\Delta_a - \mu_a)^2}{2\sigma_a^2}\right) \right\}. \quad (7)$$

From the output of a single McMC analysis on any particular dataset, we fitted the distribution in (7) to the 10,000  $\Delta_a$  samples using an EM-algorithm (Dempster, Laird & Rubin 1977) and estimated  $\mu_a$  and  $\sigma_a^2$ . The posterior distribution for a simulated data set was judged to be well separated from 0 (at the 5% level) if

$$z \equiv \frac{\widehat{\mu}_a}{\sqrt{\widehat{\sigma}_a^2}} \geq 1.96.$$

The parameter and test statistic values were then averaged across the 100 simulated datasets and the proportion of tests that rejected the null hypothesis of  $\Delta_a = 0$  was recorded for each scenario. These are displayed in Table 5 alongside the proportion of datasets in which the analogous test on the slope coefficient was rejected by the regression method of Section 5.1. We note that the similarity of the rejection rates from the two approaches is remarkable.

TABLE 5 ABOUT HERE

## 5.2.2 BAYESIAN ANALYSIS TREATED MORE APPROPRIATELY

TABLE 6 ABOUT HERE

A Bayesian analysis produces posterior distributions of *all* model parameters, including those that cannot be estimated from simpler approaches such as the regression method. However, due to paucity of information, inferences about specific parameters may be associated with high uncertainty. Here we present averages of relevant quantities over the 100 simulated datasets and we also look at the behaviour of the McMC approach on particular datasets. Table 6 gives various summary statistics of the Monte Carlo output for eight model parameter posterior distributions. 2.5% and 97.5% percentile values were obtained from the McMC sampler on each dataset along with the mean and variance for each parameter. These quantities were then averaged over the 100 simulated datasets and sample standard deviations calculated. We can see at once that the QTL allele frequency is poorly estimated in all cases with an over-estimated posterior mean. The ambiguity in the labelling of the respective alleles makes these quantities unidentifiable and the uniform prior on the interval  $[0, 1]$  takes over. The sire variance component tended to have a positively skewed posterior distribution (not shown). Estimates of both variance components display high variability. Estimation of the dominance effect in the **5-5** scenario is poor and the presence of dominance appears to affect the estimate of the additive effect by comparison with the **5-0** scenario where the signal appears to be stronger and less variable.

FIGURE 6 ABOUT HERE

Poor mixing behaviour can usually be detected from trace plots of the sampled values for each parameter separately. Figure 6 shows trace plots for  $\Delta_a$ ,  $\Delta_d$ , the two variance components and the QTL map position for one particular dataset simulated under the **5-0** scenario. The 5 parameters are all shown to mix well, as we expected, and the sampled values are spread around the true values used in the simulations. It looks though as if there could be some serial correlation for the map position of the QTL,  $\lambda_Q$ , and the sampled values are spread over



the whole range of the allowed interval. This is not surprising as linkage analysis of complex diseases can only identify large chromosomal regions of typically tens of cMs (Roberts, MacLean, Neale, Eaves & Kendler 1999). There is not enough information in the design to pinpoint the QTL location within a 10 cM interval so statistics on  $\lambda_Q$  are meaningless and should ideally not be produced at all. We placed the QTL in the centre of the interval for this reason but the trace plot of Figure 6 looks similar for other placings of the QTL.

FIGURE 7 ABOUT HERE

FIGURE 8 ABOUT HERE

As might be expected, our sampler showed very variable behaviour on different datasets generated from the same parameter values. For instance, Figure 7 shows the posterior distributions of the dominance effect, measured by  $\Delta_d$ , for two different datasets simulated for the scenario **5-0** (i.e. a scenario without dominance). In one case, the correct value of 0 for the dominance effect is fairly certain whereas the second case is vague about this value. Both posterior distributions are centred around the true value of 0 but there is considerable difference in the posterior variance of the distributions indicating large differences in the reliability of the estimated effect. Due to sampling variation, inferences can vary substantially between data sets and MCMC methods are just as influenced by such variation as any other method. In this case, it would seem that the sampler is correctly reflecting the lack of information about the dominance effect in the second data set.

FIGURE 9 ABOUT HERE

Figure 8 shows plots of the dominance effect,  $\Delta_d$ , versus  $(\mu_1 - \mu_3)/2$ , the signed QTL genotype effect size, from the MCMC output for sampler runs on each of 9 distinct datasets (those numbered 31 – 39) simulated under the **5-0** scenario. Ideally most of the sampled effect sizes  $(\mu_1 - \mu_3)/2$  should lie around  $\pm 5.0$  and the sampled dominance effects  $\mu_2 - (\mu_1 + \mu_3)/2$  should centre around 0.0. This behaviour is clearly demonstrated in subfigures (b), (d) and (e) with the difference between (b), on the one hand, and (d) and (e) on the other being that the sampler manages to flip between  $+5.0$  and  $-5$  in (b) but tends to settle for one of these in (d) and (e). However, we note that the (absolute) effect size is heavily over-estimated in (d) with very little uncertainty. In subfigures (a), (c), (f), (g) and (h), the sampler puts most of the weight on models exhibiting varying degrees of dominance i.e. where  $(\mu_2 - (\mu_1 + \mu_3)/2)$  is non-zero and is in the interval  $\pm(\mu_1 - \mu_3)/2$ . This is most pronounced for the dataset featured in subfigure (h) where almost complete dominance seems to be favoured—again with very little uncertainty. Subfigure (i) presents the interesting situation where the sampler spends most of the time moving between dominant and over-dominant models and is fairly sure about a zero effect size.

Figure 9 shows the posterior distributions of  $\Delta_a$  and  $\Delta_d$  for the datasets corresponding to subfigures (b) and (i) in Figure 8. These histograms clearly demonstrate the difficulty the sampler has in distinguishing between different genetic models. For dataset (b), the sampler gives high posterior probabilities to values of  $\Delta_a$  around 5 (the simulated value) and puts significant probability on

values of  $\Delta_d = 0$  (the simulated value) with a posterior mode slightly above 0. Hence, for dataset (b) the sampler detects the correct simulated genetic model. The corresponding plots for dataset (i) however, give high posterior probability to a zero effect size and give a strong indication of a non-zero dominance effect. The plots in Figure 8 indicate that the sampler is also capable of giving high posterior probabilities to the *wrong* values on an individual dataset. Inspection of the simulated three-locus genotypes for the ten sires in each dataset did not reveal an obvious explanation as to why this occurred. For instance, subfigures (d) and (e) both exhibit equivalent certainty about the size of the genotypic effect but (e) is more or less the correct value whereas (d), as noted above, is too high. Each dataset happened to have 7 informative sires. Of these, (d) had three fully informative (i.e. heterozygous also at *both* markers) whereas (e) had four. Likewise, the datasets corresponding to subfigures (h) and (i) were similarly uninformative in terms of segregating sires. One of the ten sires in (h) was partially informative being heterozygous at the QTL and one marker while one sire in (i) was fully informative. With this lack of information, we would expect both datasets to exhibit high uncertainty about parameter estimates via flat posterior distributions. It is somewhat disquieting to observe such certainty about the wrong values.

A similar range of behaviour patterns was observed when we looked at ten different datasets simulated under the dominance **5-5** scenario in that varying degrees of dominance were detected in some cases whereas additive models of gene action were fixed upon in others. (Results not shown.) QTL effects are either poorly inferred or not identifiable in the case of a least squares approach. In principle, this is not the case for the Bayesian model. However, the McMC sampler can lead to misleading inferences regarding these parameters in particular replications of data simulated under the same model.

### 5.3 GENERAL CONCLUSIONS

If the question of interest is whether there is a QTL segregating in a given region or not, the regression approach is fast, efficient and surprisingly reliable. It requires a large QTL effect for reasonable detection rates. However QTL effect sizes are not well estimated even in the absence of dominance and are generally confounded with allele and recombination rates. The Bayesian approach also provides a qualitatively satisfactory analysis for the presence/absence of a QTL and QTL effect sizes can, in principle, be disentangled even in the presence of dominance. However, large QTL effects are still required for sound inference. The Bayesian model uses the information contained in the variation within families allowing inferences to be drawn even in families where the sire is homozygous at the markers. The regression method cannot infer anything in this case but as such sires are truly uninformative for segregation at the QTL, these Bayesian estimates should be regarded critically. Clearly, a better quantitative analysis is possible using the Bayesian McMC approach. In particular, estimates of posterior distributions can be obtained for any model parameter, and plots

such as those presented in Figures 5, 6, 7 and 9 should all be inspected before drawing conclusions. The power of the Bayesian approach is in providing posterior probabilistic statements about parameters rather than point estimates. We have seen that such inferences can be vague when correctly reflecting sampling variability and uncertainty. However, on a specific dataset, it is possible for these posterior distributions to be peaked at the wrong parameter values leading to misleading inferences. In general, however, the unified approach of the Bayesian analysis accounting for all levels of uncertainty is undeniably attractive. Linkage phase, in particular, is inferred jointly with all the other parameters of the model whereas the regression method requires that this must be estimated *before* estimating the other parameters, conditionally on the inferred phase, and no account is taken of the uncertainty associated with the initial estimate. However, the advantages of the Bayesian approach come at a substantial computational cost. For this example, a Bayesian analysis takes approximately 30 cpu hours. In contrast, the corresponding regression analysis, including 1000 permutations for the permutation test, takes approximately 15 cpu minutes.

## 6 DISCUSSION

In the current climate of rapid development in genetics research, routine systematic analyses are required on increasingly complicated problems. Exact probability and likelihood calculations are frequently intractable and approximate methods must be used. Many of these perform standard calculations on a simplified version of the problem. For example, the least squares approach of Section 4, extracts a half-sib structure from a more complicated pedigree, ignores all the other relationships between these individuals and imposes the usual regression assumptions on the trait phenotypes. McMC methods are computationally intensive but provide a powerful alternative whereby much of the complexity of the problem can be preserved. They also have the advantage of being able to deal with missing marker data, highly polymorphic loci and non-informative sires in our example. The accuracy of the resulting estimates will depend heavily on the efficiency of the sampling scheme and the informativeness of the data. Point estimates for any parameters of interest will always be provided, but these should always be considered alongside their posterior distributions. Poor inferences can arise in the form of the usual uncertainty due to random sampling variation. More disturbingly, we found that posterior distributions occasionally peaked around the wrong values in particular cases, even when the data appeared to be completely uninformative, leading to misleading conclusions. Also, McMC methods can be easily misused to extract more information from the data than is actually available. Because of the pressure on investigators to deliver believable results quickly and the perceived unreliability of these methods, there is a general reluctance in practice to resort to McMC analysis on big complex genetics problems. In contrast, on the methodological side, proponents of McMC methods tend to advocate this approach to any problem, regardless of the practicalities involved. The aim of this paper was to highlight the need for

compromise between the theoretical and applied communities. We clearly need McMC methods but we also need to think more carefully about *when* we need them and *how* to determine if they are behaving properly.

We chose to focus on the “toy” QTL mapping problem of Sheehan et al. (2002) in order to illustrate our point, and we used simulated data to assess the sampler’s performance. Human or animal genetic mapping problems do not get any simpler than this one. A real problem would involve multiple markers with multiple alleles, multiple QTL and a complex pedigree structure. Our blocking scheme was perfectly designed to avoid the poor mixing problems encountered with single-site updating approaches. While we would not expect it to compete with a simple custom-built approach in terms of time, it is not unreasonable to demand that it deal with this problem easily and consistently. Otherwise, we can have very little confidence in recommending it for a real problem to which the true solution is unknown. Our results speak for themselves. In no situation did it seem that McMC had an advantage over the Least Squares approach in detecting the presence of a QTL besides the neater inferences clearly afforded by permitting probabilistic statements. When smaller, less informative sibships of size 10 were considered, both methods were adversely affected by the drop in information with the regression method faring slightly worse. For example, the hypothesis test results analogous to those in Table 5 indicated that the power of detection for the additive model with a large QTL effect (the **10-0** scenario) was reduced from 98% to 26% for the Bayesian method and from 95% to 12% for the regression method. There was no evidence, however, of an increase in false positives on the smaller sibships with regression. Overall, the regression approach is surprisingly good at what it purports to do and if a simple qualitative analysis is all that is required (i.e. is there a QTL with reasonable effect in this area or not?), it is difficult to sell the idea of a full McMC approach. The drawback is that regression is designed for and restricted to simple designs, requires large datasets and ignores information from homozygous sires.

Of course, we often do want more than a qualitative analysis. Estimates of sire variance components, or QTL effects in the presence of dominance, cannot be obtained from simple approaches. The regression method is not a competitor to the fully Bayesian McMC sampler on these terms. Nonetheless, we argue that it is important to a practitioner to be able to say whether McMC is “better”, or at least “necessary” for a particular application. Despite the perceived strengths of the Bayesian approach, it is important to acknowledge that there are other, often easier, ways of tackling the problem, and a formal comparison is essential to address such issues. We justify our frequentist interpretation of the Bayesian output in Section 5.2.1 on these grounds and argue that more comparisons of this type should be made. Haley & Knott (1992) compare regression with maximum likelihood methods which are also computationally intensive and conclude, as we do, that regression is fast and surprisingly good at doing what it claims to do. There are not very many instances in the literature where McMC methods are compared with simple alternative approaches to a problem.

Our example was motivated by applications in animal breeding, but our

message is far more general. McMC methods are frequently used for complex high-dimensional problems on which their performance cannot be properly assessed. Comparison with other methods is rarely, if ever, carried out. The flexibility of the McMC approach in terms of design, models and inference is indisputable. However, care should be taken in interpreting the results of such an analysis as estimates can be unreliable, especially if the sampler is mixing badly and the data are poor. In our case, for example, previous results from a single run of this particular blocking sampler yielded reasonable estimates of both variance components (Sheehan et al. 2002) and implied that dominant models of gene action could be identified (Guldbrandtsen, Sheehan & Sorensen 2002), but as demonstrated above, these findings were far from consistent when the sampler was tested over many datasets. At the risk of stating the obvious, McMC methods cannot compensate for lack of information in the data or poor study design, but they will always produce estimates of anything of interest. Further investigations of the kind we have embarked on here are essential before these methods can be exploited to their full potential in complex applications.

## 7 ACKNOWLEDGEMENTS

The authors acknowledge funding for research visits from Wellcome Trust Biomedical Research Collaboration Grant 056266/Z/98/Z. In addition, Daniel Sorensen acknowledges partial financial support from the Danish Agricultural and Veterinary Research Council grant 53-00-0332, Bernt Guldbrandtsen acknowledges support from the Danish Ministry of Agriculture, Fisheries and Food grant FREM98-DJF-1 and Nuala Sheehan acknowledges support from a Value-in-People award funded by the Wellcome Trust.

## REFERENCES

- Andersen, S. K., Olesen, K. G., Jensen, F. V. & Jensen, F. (1989), Hugin—a shell for building Bayesian belief universes for expert systems, *in* Proceedings of the 11th International Joint Conference on Artificial Intelligence, Morgan Kaufmann, San Mateo, pp. 1080–1085.
- Cannings, C. & Sheehan, N. A. (2002), On a misconception about irreducibility of the single-site gibbs sampler in a pedigree application, *Genetics* **162**, 993–996.
- Cannings, C., Thompson, E. A. & Skolnick, M. H. (1978), Probability functions on complex pedigrees, *Advances in Applied Probability* **10**, 26–61.
- Churchill, G. A. & Doerge, R. W. (1994), Empirical threshold values for quantitative trait mapping, *Genetics* **138**, 963–971.
- Cowell, R. G., Dawid, A. P., Lauritzen, S. L. & Spiegelhalter, D. J. (1999), *Probabilistic Networks and Expert Systems*, Statistics for Engineering and Information Science, Springer-Verlag, New York, Inc.

- Cowles, M. K. & Carlin, B. P. (1996), Markov chain Monte Carlo convergence diagnostics: a comparative review, *Journal of the American Statistical Society* **91**, 883–904.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977), Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society, Series B* **39**, 1–38.
- Elston, R. C. & Stewart, J. (1971), A general model for the genetic analysis of pedigree data, *Human Heredity* **21**, 523–542.
- Falconer, D. S. & Mackay, T. F. C. (1996), *Introduction to Quantitative Genetics*, fourth edn, Longman Group Ltd.
- Gelfand, A. E., Hills, S. E., Racine-Poon, A. & Smith, A. F. M. (1990), Illustration of Bayesian inference in normal data models using Gibbs sampling, *Journal of the American Statistical Society* **85**, 972–985.
- Geman, S. & Geman, D. (1984), Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on pattern analysis and machine intelligence* **45**, 721–741.
- George, A. W., Mengersen, K. L. & Davis, G. P. (2000), Localization of a quantitative trait locus via a Bayesian approach, *Biometrics* **56**, 40–51.
- Gilks, W. R., Clayton, D., Spiegelhalter, D. J., Best, N. G., McNeil, A. J., Sharples, L. D. & Kirby, A. J. (1993), Modelling complexity: applications of gibbs sampling to medicine, *Journal of the Royal Statistical Society, Series B* **55**, 39–52.
- Guldbrandtsen, B., Sheehan, N. A. & Sorensen, D. A. (2002), Detecting quantitative trait locus overdominance using graphical models based MCMC method, in *Proceedings of the Seventh World Congress on Genetics Applied to Livestock Production*. Montpellier, France.
- Haldane, J. B. S. (1919), The combination of linkage values and the calculation of distances between the loci of linked factors, *Journal of Genetics* **8**, 229–309.
- Haley, C. S. & Knott, S. (1992), A simple regression method for mapping quantitative trait loci in line crosses using flanking markers, *Heredity* **69**, 315–324.
- Haley, C. S., Knott, S. A. & Elsen, J. M. (1994), Mapping quantitative trait loci in crosses between outbred lines using least squares, *Genetics* **136**, 1195–1207.
- Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**(1), 97–109.
- Heath, S. C. (1997), Markov chain Monte Carlo segregation and linkage analysis for oliogenic models, *American Journal of Human Genetics* **61**, 748–760.

- Hoeschele, I. & VanRaden, P. M. (1993), Bayesian analysis of linkage between genetic markers and quantitative trait loci. ii. combining prior knowledge with experimental evidence, *Theoretical and Applied Genetics* **85**, 946–952.
- Hoeschele, I., Uimari, P., Grignola, F., Zhang, Q. & Gage, K. (1997), Advances in statistical methods to map quantitative trait loci in outbred populations, *Genetics* **147**, 1445–1457.
- Janss, L. L. G., Thompson, R. & Van Arendonk, J. A. M. (1995), Application of Gibbs sampling for inference in a mixed major gene-polygenic inheritance model in animal populations, *Theoretical and Applied Genetics* **91**, 1137–1147.
- Janss, L. L. G., Van Arendonk, J. A. M. & Van der Werf, J. H. J. (1995), Computing approximate monogenic model likelihoods in large pedigrees with loops, *Genetics, Selection, Evolution* **27**, 567–579.
- Jensen, C. S. & Kong, A. (1999), Blocking Gibbs sampling for linkage analysis in large pedigrees with many loops, *American Journal of Human Genetics* **65**, 885–901.
- Jensen, C. S., Kjærulff, U. & Kong, A. (1995), Blocking Gibbs sampling in very large probabilistic expert systems, *International Journal of Human-Computer Studies* **42**, 647–666.
- Knott, S. A. (2005), Regression-based quantitative trait loci mapping: robust, efficient and effective, *Philosophical Transactions of the Royal Society B* **360**, 1435–1442.
- Knott, S. A. & Haley, C. S. (1992), Maximum likelihood mapping of quantitative trait loci using full-sib families, *Genetics* **132**, 1211–1222.
- Lander, E. S. & Green, P. (1987), Construction of multilocus genetic linkage maps in humans, *Proceedings of the National Academy of Sciences (USA)* **84**, 2363–2367.
- Lauritzen, S. L. & Sheehan, N. (2003), Graphical models for genetic analyses, *Statistical Science* **18**, 489–514.
- Lauritzen, S. L. & Spiegelhalter, D. J. (1988), Local computations with probabilities on graphical structures and their applications to expert systems, *Journal of the Royal Statistical Society, Series B* **50**, 157–224.
- Lund, M. S. & Jensen, C. S. (1999), Blocking Gibbs sampling in the mixed inheritance model using graph theory, *Genetics, Selection, Evolution* **31**, 3–24.
- Lynch, M. & Walsh, B. (1998), *Genetics and Analysis of Quantitative Traits*, Sinauer Associates Inc., USA.

- Martinez, O. & Curnow, R. (1992), Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers, *Theoretical and Applied Genetics* **85**, 480–488.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N. & Teller, A. H. (1953), Equations of state calculations by fast computing machines, *Journal of Chemistry and Physics* **21**, 1087–1091.
- Ott, J. (1999), *Analysis of Human Genetic Linkage*, third edn, The Johns Hopkins University Press, Baltimore.
- Roberts, S. B., MacLean, C. J., Neale, M. C., Eaves, L. J. & Kendler, K. S. (1999), Replication of linkage studies of complex traits: An examination of variation in location estimates, *American Journal of Human Genetics* **65**, 876–884.
- Sham, P. (1998), *Statistics in Human Genetics*, Applications of Statistics Series, Arnold, London.
- Sheehan, N. (1992), Sampling genotypes on complex pedigrees with phenotypic constraints: the origin of the B allele among the Polar Eskimos, *IMA Journal of Mathematics Applied in Medicine and Biology* **9**, 1–18.
- Sheehan, N. & Thomas, A. (1993), On the irreducibility of a Markov chain defined on a space of genotype configurations by a sampling scheme, *Biometrics* **49**, 163–175.
- Sheehan, N. A., Guldbrandtsen, B., Lund, M. S. & Sorensen, D. A. (2002), Bayesian MCMC mapping of quantitative trait loci in a half-sib design: a graphical model perspective, *International Statistical Review* **70**, 241–267.
- Sillanpää, M. J. & Arjas, E. (1998), Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data, *Genetics* **148**, 1373–1388.
- Sobel, E. & Lange, K. (1996), Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics, *American Journal of Human Genetics* **58**, 1323–1337.
- Thaller, G. & Hoeschele, I. (1996), A Monte Carlo method for Bayesian analysis of linkage between single markers and quantitative trait loci I. Methodology, *Theoretical and Applied Genetics* **93**, 1161–1166.
- Thomas, A., Gutin, A., Abkevich, V. & Bansal, A. (2000), Multilocus linkage analysis by blocked Gibbs sampling, *Statistics and Computing* **10**, 259–269.
- Thompson, E. A. (1994), Monte Carlo likelihood in genetic mapping, *Statistical Science* **9**(3), 355–366.



- Thompson, E. A. (2000), *Statistical Inference from Genetic Data on Pedigrees*, Vol. 6 of *NSF-CBMS regional Conference Series in Probability and Statistics*, Institute of Mathematical Statistics and the American Statistical Association, Beachwood, Ohio, USA.
- Thompson, E. A. (2001), Monte Carlo methods on genetic structures, *in* O. E. Barndorff-Nielsen, D. R. Cox & C. Kluppelberg, eds, *Complex Stochastic Systems*, Chapman & Hall, chapter 4, pp. 176–218.
- Thompson, E. A. & Heath, S. C. (2000), Estimation of conditional multilocus gene identity among relatives, *in* F. Seiller-Moiseiwitsch, ed., *Statistics in molecular biology and genetics*, IMS Lecture Notes. Institute of Mathematical Statistics. American Mathematical Society, pp. 95–113.
- Wang, T., Fernando, R. L., Stricker, C. & Elston, R. C. (1996), An approximation to the likelihood for a pedigree with loops, *Theoretical and Applied Genetics* **93**, 1299–1309.
- Yi, N. & Xu, S. (2000), Bayesian mapping of quantitative trait loci under the identity-by-descent-based variance component model, *Genetics* **156**, 411–422.

## 8 TABLES

Table 1: Scenarios for which data sets were simulated. The last column shows the expected percentage of the total phenotypic variance,  $\sigma^2$ , that is explained by variation at the QTL.

Name	$a$	$d$	$\lambda_Q$	$\frac{\sigma_{qtl}^2}{\sigma^2}$
<b>0-0</b>	0.0	0.0	5.0 $cM$	0.0%
<b>2.5-0</b>	2.5	0.0	5.0 $cM$	2.6%
<b>5-0</b>	5.0	0.0	5.0 $cM$	9.5%
<b>10-0</b>	10.0	0.0	5.0 $cM$	29.6%
<b>5-5</b>	5.0	5.0	5.0 $cM$	20.0%

Table 2: Test-based classification rates for individual sires alongside the “true” simulated classifications for each of the five scenarios.

Scenario	True QTL	Regression Classification	
		qq or QQ	Qq
<b>0-0</b>	qq, QQ	96.3%	3.7%
	Qq	96.8%	3.2%
<b>2.5-0</b>	qq, QQ	96.8%	3.2%
	Qq	88.5%	11.5%
<b>5-0</b>	qq, QQ	96.4%	3.6%
	Qq	67.3%	32.7%
<b>10-0</b>	qq, QQ	95.3%	4.7%
	Qq	35.6%	64.4%
<b>5-5</b>	qq, QQ	98.0%	2.0%
	Qq	57.1%	42.9%

Table 3: Regression estimates of QTL effects. Figures give the mean  $\pm$  the standard deviation of the absolute values of the regression estimates of QTL effect conditional on whether or not the null hypothesis of no QTL segregating in the interval was rejected in the family.

<b>Scenario</b>	<b><math>H_0</math> not rejected</b>	<b><math>H_0</math> rejected</b>
<b>0-0</b>	1.59 $\pm$ 0.76	6.57 $\pm$ 0.57
<b>2.5-0</b>	1.67 $\pm$ 0.83	6.79 $\pm$ 0.75
<b>5-0</b>	1.80 $\pm$ 0.94	7.80 $\pm$ 0.91
<b>10-0</b>	1.68 $\pm$ 0.97	10.70 $\pm$ 1.68
<b>5-5</b>	1.98 $\pm$ 1.01	8.96 $\pm$ 1.18

Table 4: Bayesian classification probabilities of sire QTL genotypes.

True QTL	Average QTL genotype posterior probabilities			
	Scenario 0-0			
	qq	Qq	QQ	Homozygote
qq	<b>0.261</b>	0.479	0.260	0.521
Qq	0.258	<b>0.486</b>	0.257	0.515
QQ	0.259	0.483	<b>0.257</b>	0.516
	Scenario 2.5-0			
	qq	Qq	QQ	Homozygote
qq	<b>0.266</b>	0.470	0.264	0.530
Qq	0.243	<b>0.517</b>	0.239	0.482
QQ	0.261	0.478	<b>0.262</b>	0.523
	Scenario 5-0			
	qq	Qq	QQ	Homozygote
qq	<b>0.277</b>	0.432	0.291	0.568
Qq	0.185	<b>0.628</b>	0.187	0.372
QQ	0.278	0.430	<b>0.293</b>	0.571
	Scenario 10-0			
	qq	Qq	QQ	Homozygote
qq	<b>0.389</b>	0.230	0.381	0.770
Qq	0.073	<b>0.851</b>	0.076	0.149
QQ	0.381	0.172	<b>0.447</b>	0.823
	Scenario 5-5			
	qq	Qq	QQ	Homozygote
qq	<b>0.297</b>	0.403	0.300	0.597
Qq	0.151	<b>0.695</b>	0.154	0.305
QQ	0.382	0.293	<b>0.326</b>	0.706

Table 5: Bayesian detection of additive QTL effects via hypothesis testing. For each scenario the estimated mean and variance of the posterior distribution of  $\Delta_a$  is shown along with the average  $z$ -test statistic as well as the proportion of  $H_0$ 's rejected at  $\alpha = 5\%$  within 100 replicate data sets generated for each scenario. Also shown is the power of detection using the regression method.

Scenario	0-0	2.5-0	5-0	10-0	5-5
$\bar{\mu}_a$	1.69	2.09	3.70	8.65	4.85
$\sigma_a^2$	2.76	2.91	3.24	2.91	3.63
$\bar{z}$	1.05	1.26	2.28	5.83	2.95
Prop. Rej. $H_0$	7.0%	15.0%	53.0%	98.0%	70.0%
Regression	6.0%	17.0%	51.0%	95.0%	69.0%

Table 6: Monte Carlo output showing sample mean $\pm$ standard deviation (across replicate data sets), correct to three decimal places, of the 2.5% and 97.5% percentiles, mean and variance of the posterior distributions for 8 parameters in the model. We omit the scenario with small additive effect (2.5 – 0) and give the “true” values used in the simulations alongside the different parameters.

		2.5%	Mean	Variance	97.5%
<b>Scenario 0-0</b>					
$\sigma_s^2$	7.5	2.877 $\pm$ 0.311	8.484 $\pm$ 0.621	19.688 $\pm$ 1.271	20.022 $\pm$ 0.721
$\sigma_{res}^2$	92.5	59.842 $\pm$ 1.551	81.812 $\pm$ 0.793	109.795 $\pm$ 13.662	100.565 $\pm$ 1.843
$P_M$	0.5	0.455 $\pm$ 0.002	0.503 $\pm$ 0.003	0.001 $\pm$ 0.000	0.557 $\pm$ 0.007
$P_Q$	0.3	0.457 $\pm$ 0.002	0.499 $\pm$ 0.001	0.000 $\pm$ 0.000	0.541 $\pm$ 0.001
$P_N$	0.5	0.441 $\pm$ 0.002	0.499 $\pm$ 0.001	0.001 $\pm$ 0.000	0.557 $\pm$ 0.001
$\lambda_Q$	0.05	0.004 $\pm$ 0.000	0.050 $\pm$ 0.001	0.001 $\pm$ 0.000	0.096 $\pm$ 0.000
$\Delta_a$	0	0.083 $\pm$ 0.007	2.130 $\pm$ 0.168	2.622 $\pm$ 0.371	6.004 $\pm$ 0.419
$\Delta_d$	0	-10.885 $\pm$ 0.426	-0.308 $\pm$ 0.216	36.008 $\pm$ 3.602	10.616 $\pm$ 0.496
<b>Scenario 5-0</b>					
$\sigma_s^2$	7.5	3.137 $\pm$ 0.400	9.640 $\pm$ 0.686	25.575 $\pm$ 1.970	22.480 $\pm$ 0.772
$\sigma_{res}^2$	92.5	60.893 $\pm$ 1.468	85.262 $\pm$ 0.876	118.556 $\pm$ 12.916	104.387 $\pm$ 1.869
$P_M$	0.5	0.456 $\pm$ 0.005	0.504 $\pm$ 0.003	0.001 $\pm$ 0.000	0.551 $\pm$ 0.002
$P_Q$	0.3	0.458 $\pm$ 0.001	0.499 $\pm$ 0.001	0.000 $\pm$ 0.000	0.540 $\pm$ 0.001
$P_N$	0.5	0.442 $\pm$ 0.001	0.498 $\pm$ 0.001	0.001 $\pm$ 0.000	0.555 $\pm$ 0.001
$\lambda_Q$	0.05	0.004 $\pm$ 0.000	0.049 $\pm$ 0.001	0.001 $\pm$ 0.000	0.096 $\pm$ 0.000
$\Delta_a$	5	0.232 $\pm$ 0.031	3.843 $\pm$ 0.307	4.945 $\pm$ 0.628	8.552 $\pm$ 0.664
$\Delta_d$	0	-11.204 $\pm$ 0.329	-1.017 $\pm$ 0.486	30.676 $\pm$ 2.345	10.714 $\pm$ 0.293
<b>Scenario 10-0</b>					
$\sigma_s^2$	7.5	2.928 $\pm$ 0.061	9.826 $\pm$ 0.620	34.591 $\pm$ 5.253	25.399 $\pm$ 1.901
$\sigma_{res}^2$	92.5	69.793 $\pm$ 0.696	93.419 $\pm$ 1.556	156.506 $\pm$ 12.127	117.883 $\pm$ 2.254
$P_M$	0.5	0.455 $\pm$ 0.003	0.502 $\pm$ 0.004	0.001 $\pm$ 0.000	0.552 $\pm$ 0.005
$P_Q$	0.3	0.460 $\pm$ 0.002	0.499 $\pm$ 0.001	0.000 $\pm$ 0.000	0.539 $\pm$ 0.001
$P_N$	0.5	0.445 $\pm$ 0.002	0.499 $\pm$ 0.001	0.001 $\pm$ 0.000	0.552 $\pm$ 0.001
$\lambda_Q$	0.05	0.003 $\pm$ 0.000	0.046 $\pm$ 0.001	0.001 $\pm$ 0.000	0.095 $\pm$ 0.001
$\Delta_a$	10	2.477 $\pm$ 1.331	8.466 $\pm$ 0.544	6.540 $\pm$ 1.243	12.450 $\pm$ 0.217
$\Delta_d$	0	-7.810 $\pm$ 1.155	-0.762 $\pm$ 0.289	9.917 $\pm$ 1.820	5.988 $\pm$ 1.102
<b>Scenario 5-5</b>					
$\sigma_s^2$	7.5	2.922 $\pm$ 0.026	8.938 $\pm$ 0.246	24.211 $\pm$ 1.915	21.867 $\pm$ 0.746
$\sigma_{res}^2$	92.5	59.228 $\pm$ 0.949	85.370 $\pm$ 1.748	162.887 $\pm$ 14.654	110.283 $\pm$ 3.499
$P_M$	0.5	0.451 $\pm$ 0.001	0.500 $\pm$ 0.000	0.001 $\pm$ 0.000	0.552 $\pm$ 0.001
$P_Q$	0.3	0.459 $\pm$ 0.000	0.500 $\pm$ 0.000	0.000 $\pm$ 0.000	0.541 $\pm$ 0.000
$P_N$	0.5	0.443 $\pm$ 0.001	0.500 $\pm$ 0.000	0.001 $\pm$ 0.000	0.557 $\pm$ 0.000
$\lambda_Q$	0.05	0.003 $\pm$ 0.000	0.050 $\pm$ 0.000	0.001 $\pm$ 0.000	0.096 $\pm$ 0.000
$\Delta_a$	5	0.105 $\pm$ 0.015	2.881 $\pm$ 0.396	5.103 $\pm$ 1.062	8.267 $\pm$ 0.715
$\Delta_d$	5	-11.204 $\pm$ 0.061	0.229 $\pm$ 0.252	37.383 $\pm$ 1.155	11.220 $\pm$ 0.113

## 9 FIGURES



Figure 1: A representation of genotypic value for a diallelic locus adapted from Falconer & Mackay (1996).

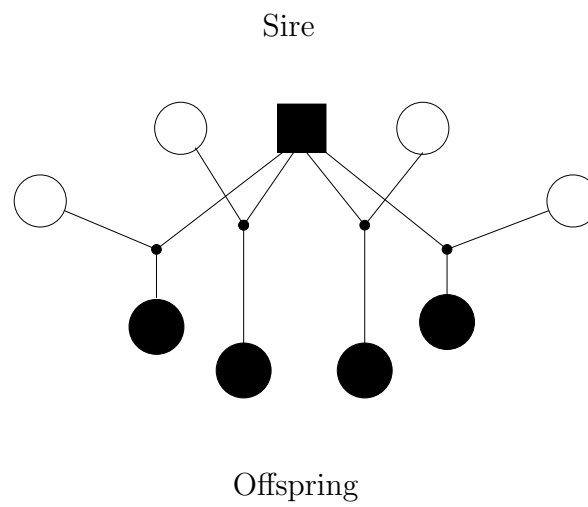


Figure 2: A *marriage node* graph of a half-sib design comprising a sire with 4 offspring. Individual nodes are represented by squares for males and circles for females while marriages (i.e. fertile matings) are represented by small black dots. The nodes corresponding to observed individuals are depicted in black.

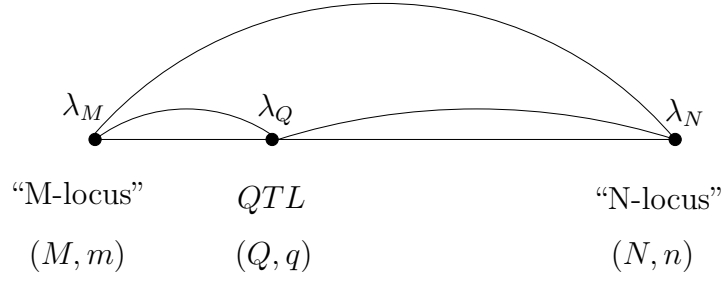


Figure 3: A very simple mapping problem in which the hypothesis of interest is that there is a single diallelic QTL coding for our continuous trait somewhere between two diallelic markers. The marker positions  $\lambda_M$  and  $\lambda_N$  are known so only the QTL location  $\lambda_Q$  has to be estimated.

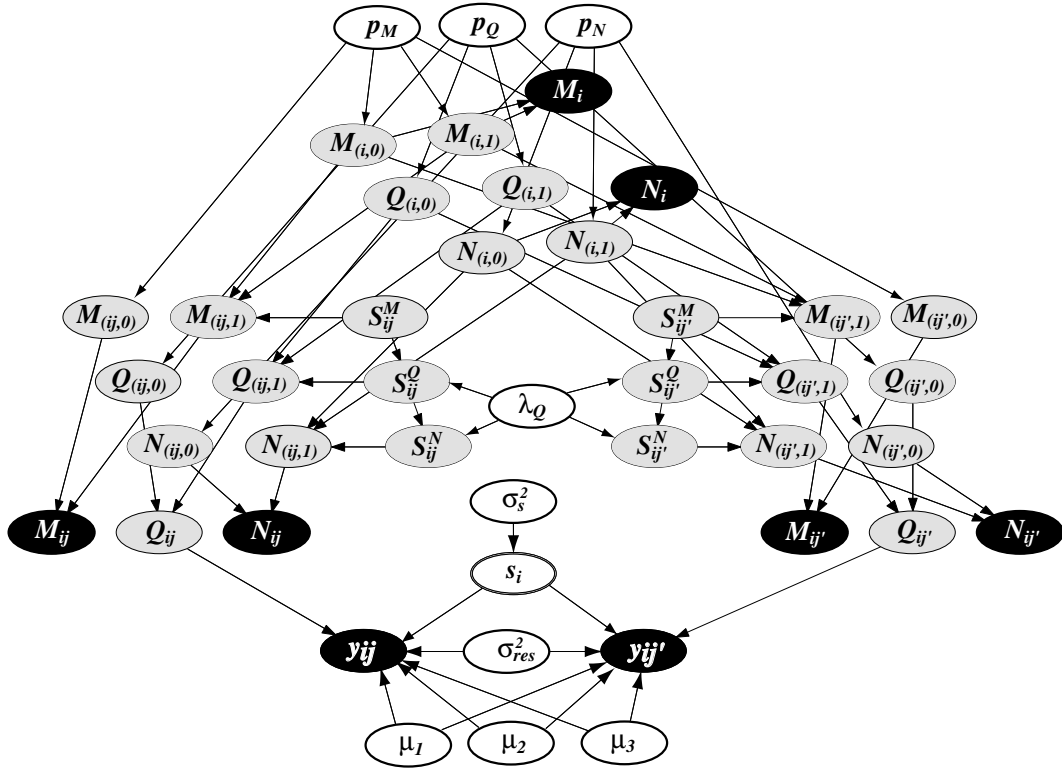


Figure 4: The graphical model taken from Sheehan et al. (2002) for the full Bayesian analysis on a half-sib design with one sire  $i$  and two daughters,  $ij$  and  $ij'$ . The black nodes are observed. Arrows connecting the nodes representing the segregation indicators  $S_{ij}^M$  and  $S_{ij}^Q$ , for example, denote the linkage between the loci, and dependency on the map location,  $\lambda_Q$ , is reflected by further arrows.



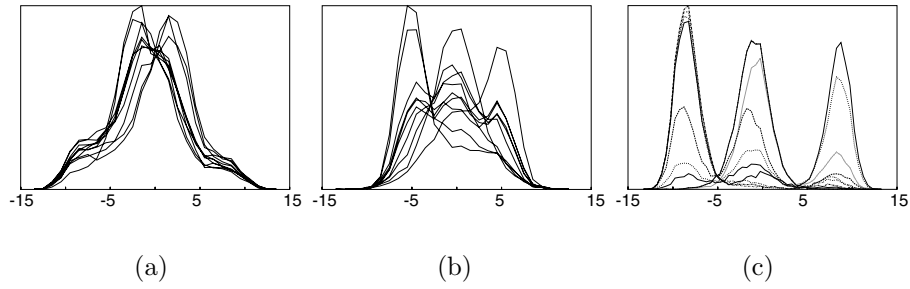


Figure 5: Distributions of sire QTL genotype effects for three scenarios. Each line corresponds to the posterior distribution of the effects of the QTL in one of the 10 sires in the simulation. (a):  $a = 2.50$ ,  $d = 0$  (b):  $a = 5.0$ ,  $d = 0$  (c):  $a = 10.0$ ,  $d = 0$ .

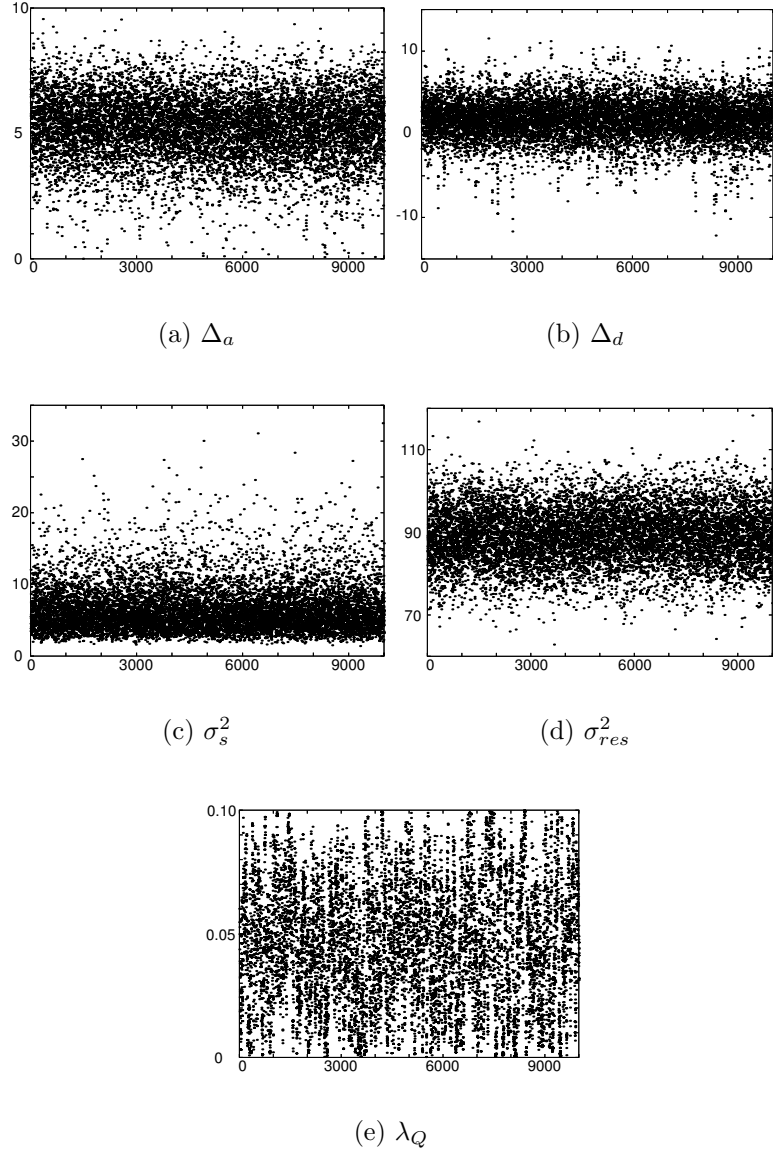


Figure 6: Trace plots for 5 parameters for one dataset simulated under scenario **5-0**. The dataset shown is the same as shown in dataset (b) in figure 8.

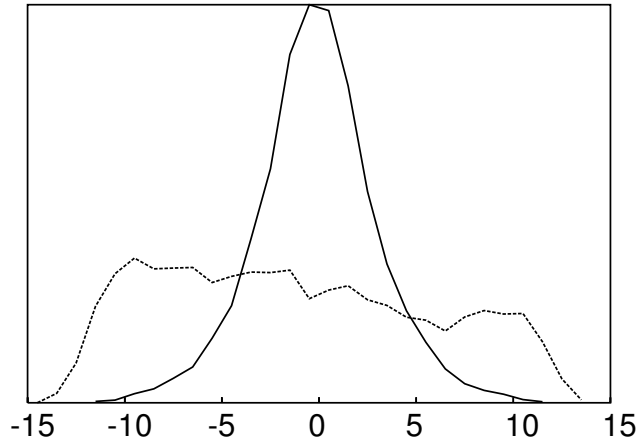


Figure 7: Comparison of distributions of  $\Delta_d$  for two different data sets generated with the same parameters ( $a = 5.0$  and  $d = 0.0$ ).

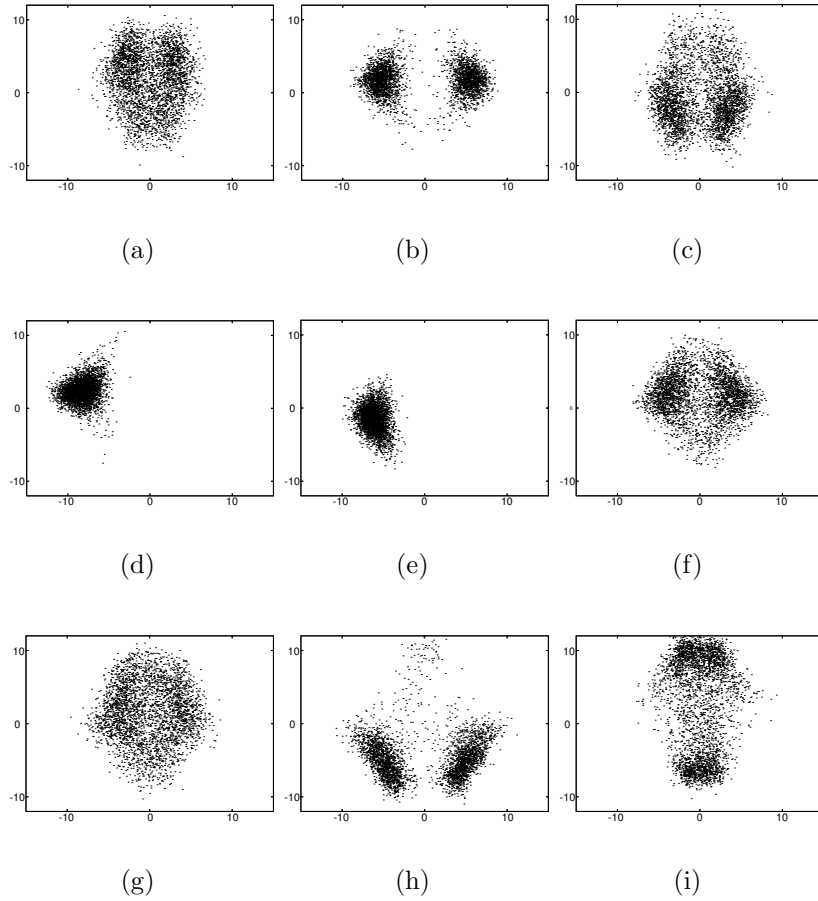


Figure 8: Nine examples of the behaviour of the sampler with respect to  $(\mu_1 - \mu_3)/2$  on the x-axis vs.  $\Delta_d = \mu_2 - (\mu_1 + \mu_3)/2$  on the y-axis. Each subplot corresponds to the output of the Gibbs sampler for a distinct dataset simulated under the scenario **5-0**. All subfigures are on the same scale, the x-axis runs from -15 to +15 while the y-axis runs from -12 to +12.

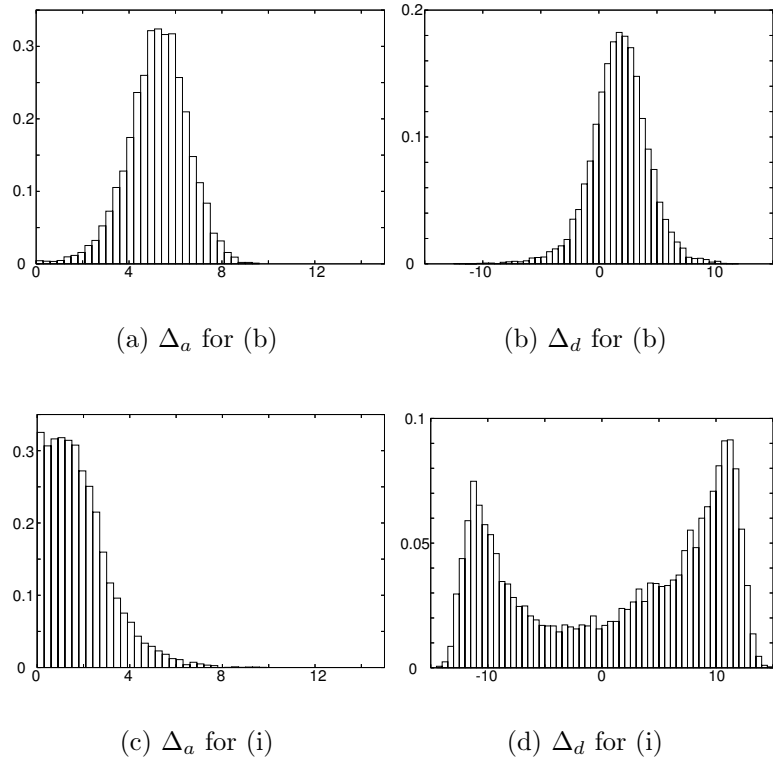


Figure 9: Plots of the posterior densities for  $\Delta_a$  and  $\Delta_d$  for two selected datasets simulated under the scenario **5-0**. The top two figures correspond to subfigure (b) in figure 8, while the bottom two correspond to subfigure (i) in figure 8.