Genet. Res., Camb. (2002), **80**, *pp.* 245–256. *With* 1 *figure.*
DOI: 10.1017/S0016672302005906 Printed in the United Kingdom

On flexible finite polygenic models for multiple-trait evaluation

MARCO C. A. M. BINK[1,2]*
[1] *Biometris – Wageningen University and Research Centre, PO Box 100, 6700 AC Wageningen, The Netherlands*
[2] *Institute for Animal Science and Health, Lelystad, The Netherlands*

(*Received 7 January 2002 and in revised form 8 April and 25 July 2002*)

Summary

Finite polygenic models (FPM) might be an alternative to the infinitesimal model (TIM) for the genetic evaluation of pedigreed multiple-generation populations for multiple quantitative traits. I present a general flexible Bayesian method that includes the number of genes in the FPM as an additional random variable. Markov-chain Monte Carlo techniques such as Gibbs sampling and the reversible jump sampler are used for implementation. Sampling of genotypes of all genes in the FPM is done via the use of segregation indicators. A broad range of FPM models, some combined with TIM, are empirically tested for the estimation of variance components and the number of genes in the FPM. Four simulation scenarios were studied, including genetic models with 5 or 50 additive independent diallelic genes affecting the traits, and random selection or selection on one of the traits was performed. The results in this study were based on ten replicates per simulation scenario. In the case of random selection, uniform priors on additive gene effects led to posterior mean estimates of genetic variance that were positively correlated with the number of genes fitted in the FPM. In the case of trait selection, assuming normal priors on gene effects also led to genetic variance estimates for the selected trait that were negatively correlated with the number of genes in the FPM. This negative correlation was not observed for the unselected trait. Treating the number of genes in the FPM as random revealed a positive correlation between prior and posterior mean estimates of this number, but the prior hardly affected the posterior estimates of genetic variance. Posterior inferences about the number of genes should be considered to be indicative where trait selection seems to improve the power of distinguishing between TIM and FPM. Based on the results of this study, I suggest not replacing TIM by the FPM, but combining TIM and FPM with the number of genes treated as random, to facilitate a highly flexible and thereby robust method for variance component estimation in pedigreed populations. Further study is required to explore the full potential of these models under different genetic model assumptions.

1. Introduction

Accurate, unbiased estimation of genetic variance components plays a key role in breeding programs. Fisher (1918) introduced the infinitesimal model (TIM), in which it is assumed that traits are determined by an infinite number of unlinked, additive genetic loci, each with an infinitesimal contribution to the trait. TIM has been successfully applied in animal breeding and proved to be robust when analysing long-term selection responses (e.g. Martinez *et al.*, 2000). Thompson & Skolnick (1977) proposed a finite polygenic model (FPM) to estimate genetic variance components by fitting a finite number of unlinked polygenic loci that describe the genetic covariance among pedigree members. Fernando *et al.* (1994) proposed the FPM for complex segregation analysis. The FPM allows the inclusion of non-additive genetic effects, such as dominance and epistasis, whereas this inclusion is hard in the infinitesimal model for theoretical and computational reasons (e.g. DeBoer & Hoeschele, 1993).

However, the FPM has not yet been widely applied for estimation of genetic parameters. A major problem in implementing an FPM using a standard likelihood approach is the calculation of the genotype probability

* Tel: +31 317 47 70 00. Fax: +31 317 41 80 94.
e-mail: m.c.a.m.bink@plant.wag-ur.nl

for all the loci. The large number of possible genotype combinations quickly causes practical computation problems as the number of individuals increases, and is likely to explode with complex pedigree structures involving loops. Markov-chain Monte Carlo (MCMC) approaches based upon Gibbs sampling algorithms, which were previously suggested for segregation studies of single genes (Guo & Thompson, 1992; Janss *et al*., 1995), have recently been extended to the multiple unlinked genes in the FPM (Pong-Wong *et al*., 1999). Instead of sampling the genotypes of individuals, the segregation of genes can also be fully specified by meiosis indicators (e.g. Thompson, 1999). In the computation of probabilities of gene descent in pedigrees, these meiosis indicators have also been called inheritance vectors (Lander & Green, 1987), descent graphs (Sobel & Lange, 1996) and segregation indicators (Thompson, 1994). Implementation of descent graphs to sample genotypes jointly for all individuals has been explored by Tier & Henshall (2001) and Du & Hoeschele (2000); Du and Hoeschele reported that the joint sampling procedure did not improve the parameter estimation compared with simpler implementation of descent graphs.

The estimates for genetic parameters can be heavily affected by the number of genes that are fitted in the FPM and the assumed prior distribution for gene effects (Pong-Wong *et al*., 1999; Du *et al*., 1999). Assuming uniform priors on gene effects, Pong-Wong *et al*. (1999) and Du *et al*. (1999) found that the number of loci in the model positively biased the estimates for genetic variance. This bias was not observed when normal priors on gene effects were assumed (Pong-Wong *et al*., 1999). Du & Hoeschele (2000) postulated that the positive bias found by Pong-Wong *et al*. (1999) might have been due to lack of data. Du & Hoeschele (2000) simulated larger data sets and found much smaller biases owing to the number of genes in the model, especially when assuming normal priors. However, they suggested further research to determine the optimal number of genes in the FPM. Also, most of these results were obtained for situations of random selection. It has not been well studied whether different priors on gene effects behave similarly in situations of trait selection, as in breeding programs.

In this study, I describe a flexible FPM for the estimation of genetic parameters. First, I include the number of genes in the FPM as an additional unknown variable via a reversible jump Metropolis–Hastings algorithm (Green, 1995). Second, I sample genotypes for genes using segregation indicators (Thompson, 1994). Furthermore, I combine the FPM with TIM, making the joint procedure as flexible as possible. The method considers multiple-trait evaluation, which is a logical extension and has not been explored before. I use Monte Carlo simulation to study the performance of many FPM models for scenarios with few and many genes simulated and for situations of random selection or trait selection. In the simulation, I focus on additive genetic models, whereas the method is described for additive and dominance genetic models.

2. Methods

The method is described for two quantitative traits, but there is no theoretical or practical restriction for more than two traits. Also, for reasons of clarity and focus, I omit the description of TIM (e.g. Wang *et al*., 1993; Sorensen, 1996) in this section, although I do consider combinations of the FPM and TIM in the analysis of simulated data.

(i) *Finite polygenic model*

Under this model, the two quantitative traits are assumed to be genetically controlled by L unlinked biallelic loci, so the genotypes can be denoted as AA, AB and BB. Following the same notation as Falconer (1989), each locus l for trait t has an additive (a_{lt}) and a dominance (d_{lt}) effect with a frequency of alleles A and B in the base population of p_l and q_l, respectively. Here, I assume that the base population is produced by random mating; that is, it is in Hardy–Weinberg equilibrium and there is linkage equilibrium among loci. If m_{lt} denotes the mean for locus l for trait t, then

$$m_{lt} = (p_l)^2 \times (-a_{lt}) + (2p_lq_l)d_{lt} + q_l^2 a_{lt}.$$

The total genetic (co)variance explained by locus l for traits t and t' is thus

$$\begin{aligned}\sigma_{ltt'} = {} & p_l^2(m_{lt} - (-a_{lt}))(m_{lt'} - (-a_{lt'})) \\ & + q_lp_l(m_{lt} - d_{lt})(m_{lt'} - d_{lt'}) \\ & + q_l^2(m_{lt} - a_{lt})(m_{lt'} - a_{lt'}).\end{aligned}$$

If $t' = t$ then $\sigma_{ltt'} = \sigma_{lt}^2$ (the genetic variance); otherwise, it equals the genetic covariance. Because the loci are assumed to be unlinked and in linkage equilibrium, the total genetic (co)variances are the sums over all the loci. As in the studies by Pong-Wong *et al*. (1999), Du *et al*. (1999) and Du & Hoeschele (2000), I assume that the allele frequencies p_l and q_l are fixed at 0·5 in the founders.

Let $\mathbf{y}_t$ and $\mathbf{y}_{t'}$ denote the vectors of dimension n_y of data for the two traits t and t'. Let α_{lt} denote the two-dimensional vector $[a_{lt}\ d_{lt}]^{\mathrm{T}}$ for the lth locus. Now the following model can be assumed:

$$\begin{aligned}\begin{bmatrix}\mathbf{y}_t \\ \mathbf{y}_{t'}\end{bmatrix} = {} & \begin{bmatrix}\mathbf{X}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{t'}\end{bmatrix} \times \begin{bmatrix}\mathbf{b}_t \\ \mathbf{b}_{t'}\end{bmatrix} \\ & + \sum_{l-1}^{L} \left(\begin{bmatrix}\mathbf{Z}_{lt} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{lt'}\end{bmatrix} \times \begin{bmatrix}\boldsymbol{\alpha}_{lt} \\ \boldsymbol{\alpha}_{lt'}\end{bmatrix} \right) \\ & + \begin{bmatrix}\mathbf{I}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{t'}\end{bmatrix} \times \begin{bmatrix}\mathbf{e}_t \\ \mathbf{e}_{t'}\end{bmatrix},\end{aligned}$$

where $\mathbf{b}_t$ ($\mathbf{b}_{t'}$) is the vector of dimension k_t ($k_{t'}$) of 'nuisance effects', and $\mathbf{e}_t$ ($\mathbf{e}_{t'}$) is the vector of dimension n_t ($n_{t'}$) of residuals of trait t and t'. Incidence matrices $\mathbf{X}_t$ and $\mathbf{X}_{t'}$ relate the trait data to the nuisance effects and are identical to each other when the same nuisance parameters are included for all traits. For simplicity, I assume only an overall mean effect μ for each trait (μ_t and $\mu_{t'}$). The incidence matrices $\mathbf{Z}_{lt}$ and $\mathbf{Z}_{lt'}$ are typically unknown because the genotypes of individuals are unobserved. The conditional distribution of the data (ordered individuals within traits) is:

$$\begin{pmatrix} \mathbf{y}_t \\ \mathbf{y}_{t'} \end{pmatrix} \Bigg| \mathbf{b}_t,\ \mathbf{b}_{t'}, \boldsymbol{\alpha}_t, \boldsymbol{\alpha}_{t'}, \mathbf{R}_e \Bigg) \sim N\left(\begin{matrix} \mathbf{X}_t\mathbf{b}_t + \sum_{l-1}^{L} Z_{lt}\boldsymbol{\alpha}_{lt} \\ \mathbf{X}_{t'}\mathbf{b}_{t'} + \sum_{l-1}^{L} \mathbf{Z}_{lt'}\boldsymbol{\alpha}_{lt'} \end{matrix}, \mathbf{R}_e \otimes \mathbf{I}_n \right),$$

where $\mathbf{R}_e = \begin{bmatrix} \sigma^2_{e,t} & \sigma_{e,tt'} \\ \sigma_{e,tt'} & \sigma^2_{e,t'} \end{bmatrix}$.

(ii) *Prior distributions*

For the nuisance effects, I assume proper uniform distributions (omitting the bounds $\mathbf{b}_{t(\min)}$, $\mathbf{b}_{t(\max)}$, $\mathbf{b}_{t'(\min)}$, and $\mathbf{b}_{t'(\max)}$),

$p(\mathbf{b}_t) \propto$ constant and $p(\mathbf{b}_{t'}) \propto$ constant.

Based on the results of Pong-Wong *et al.* (1999), I consider two possible prior distributions to model the gene effects: uniform and normal. Pong-Wong *et al.* (1999) chose a folded-over normal because a_t was, by their definition, the effect of the favourable homozygous genotype. In a multiple-trait situation, this folded-over normal is not applicable because a genotype might be favourable for one trait and unfavourable for another trait; that is, it might act in opposite directions for different traits. Inverted Wishart (IW) distributions are used as prior distributions for the (co)variance components, mainly for simplicity. The Wishart density describes the distribution of sums of squares and cross-products of standard normal random variables (see also Van Tassell & Van Vleck, 1996).

$$p(\mathbf{R}_e|\nu_e, \mathbf{V}_e) \propto |\mathbf{R}_e|^{-\frac{1}{2}(\nu_e+n_t+1)} \exp\left[-\tfrac{1}{2} tr(\mathbf{R}_e^{-1}\mathbf{V}_e^{-1})\right],$$

where ν_e and $\mathbf{V}_e$ are hyperparameters of the distributions, which I assume are known, and n_t is the number of traits (and the dimension of $\mathbf{R}_e$). In the univariate case, the IW distribution reduces to an inverted ψ-square distribution. This IW distribution reduces to a uniform distribution if we set $\nu = -(n_t+1)$ and $\mathbf{V}=0$ (see also Sorensen, 1996). In the case of normal priors for the random variables a_{lt} and d_{lt}, IW distributions are also used as priors for the hyperparameters Λ_a and Λ_d, respectively.

The number of unlinked genes (L) was assumed to be known and kept fixed in previous studies by Pong-Wong *et al.* (1999), Du *et al.* (1999) and Du & Hoeschele (2000). Here, I consider L as a random variable and want to infer its distribution from the data. Similar to several quantitative-trait-locus (QTL) mapping studies (e.g. Sillanpää & Arjas, 1998; Stephens & Fish, 1998; Lee & Thomas, 2000), I use a Poisson distribution with hyperparameter λ as the prior distribution for the number of genes in the model.

(iii) *Joint posterior distribution*

Combining the conditional distribution of the complete data and the priors on model parameters lead to the required joint posterior density of all unknowns (including the $L \times N$ genotypes in $\mathbf{G}$),

$$\begin{aligned} p(\mathbf{b}, \mathbf{a}, \mathbf{d}, \mathbf{p}, \mathbf{R}_e, \mathbf{G}, L \,|\, \mathbf{y}) \propto & \prod_{i=1}^{N} p(\mathbf{y}_i \,|\, \mathbf{b}, \mathbf{a}, \mathbf{d}, \mathbf{R}_e, \mathbf{G}_i) \\ & \times p(\mathbf{b}) \times p(\mathbf{a}) \times p(\mathbf{d}) \times p(\mathbf{R}_e) \\ & \times \prod_{i=1}^{N_b} \prod_{j=1}^{L} p(g_{ij}) \\ & \times \prod_{i=N_b+1}^{N_d} \prod_{j=1}^{L} p(g_{ij} \,|\, g_{f_i j}, g_{m_i j}), \end{aligned}$$

where N is the number of individuals (assuming all individuals have phenotypes), N_b and N_d are the number of base individuals (founders) and the number of descendants, respectively ($N = N_b + N_d$), $\mathbf{G}_i$ is row i of $\mathbf{G}$ and $\mathbf{G}_i = \{g_{ij}\}$, where g_{ij} is the genotype of individual i at locus j, and $p(g_{ij})$ and $p(g_{ij}|g_{f_i j}, g_{m_i j})$ are the population frequency and transition probability, respectively, of genotype g_{ij}.

The genotypes for a locus are determined together by the alleles of founder individuals and by segregation indicators of non-founder individuals (Lange & Matthysse, 1989; Thompson, 1994; Sobel & Lange, 1996). The segregation indicators uniquely describe the flow of genes through a pedigree and the implementation was similar to those described (for QTL mapping) by Uimari & Sillanpää (2001) and Bink *et al.* (2002). Similar to Du & Hoeschele (2000), we modified the block-sampling scheme of Janss *et al.* (1995), such that both male and female parents are sampled unconditionally on the genotypes of their final offspring (offspring not having progeny themselves), i.e. integration of the final offspring.

The number of affecting loci (L) was updated through reversible jump sampling (Green, 1995; Waagepetersen & Sorensen, 2001). I have adopted the implementation proposed by Sillanpää & Arjas (1998). As a basic strategy, only single step moves

were allowed (i.e. only one locus may be added or deleted during an updating cycle). In the locus addition proposal (i.e. the birth step), new values for α were generated from their priors (where $\alpha=[a, d]^{T}$ (the gene effects)). To increase the probability of acceptance for the addition step, we used a scaling factor w that is smaller than unity (e.g. 0·05) (see also Uimari & Sillanpää, 2001) on the prior distribution of α. New genotypes and segregation indicators for founders and non-founders, respectively, were also proposed from the priors. Given a truncated Poisson prior distribution on L with parameter λ, the acceptance ratio of the birth step reduces to (Sillanpää & Arjas, 1998)

$$A=\frac{P(\mathbf{y}\mid G', \alpha', \textit{others})}{P(\mathbf{y}\mid G, \alpha, \textit{others})}\times\frac{\lambda}{(L)^{2}}\times w,$$

where $'$ indicates the old values (of existing loci) plus proposed values for the new locus, and 'others' refer to all other random variables, which are constants in both numerator and denominator. If a deletion (death step) was proposed, the locus to be deleted was chosen randomly. The acceptance ratio for a deletion step is $1-A$. Refer to Uimari & Sillanpää (2001) and Bink *et al.* (2002) for more details on the sampling steps of the MCMC simulation.

(iv) *Simulated population*

Four scenarios are considered, differing in genetic model (5 or 50 genes underlying the quantitative traits) and in selection (random or trait selection). These scenarios are denoted as S05, S05s, S50 and S50s, where the number indicates the number of genes and "s" indicates trait selection. Each scenario will be studied by analysing ten replicates of simulated data sets.

(a) *Genetic parameters*

The environmental and genetic variances for two quantitative traits were 1·0. The genetic variance was due to 5 or 50 independent genes that each had the same contribution (Table 1). All genes were assumed to be diallelic and to act additively (i.e. no dominance) on both traits. The direction (or correlation) differed between genes (Table 1). For the base population, the allele frequency was 0·5 for each gene and I assumed Hardy–Weinberg equilibrium within genes and linkage equilibrium between genes.

(b) *Population structure*

The structure of the simulated population consisted of a base population of 80 unrelated individuals (40 males and 40 females) plus 5 discrete generations each of 400 individuals. At each generation 40 males and 40 females were selected, where selection was either random or based on phenotypic superiority for trait 1. A single mating resulted in 4 full sibs, 2 males and 2 females, where the parents were a random sample of available candidates with no restriction on the number of offspring per parent nor on the kinship between parents.

Table 1. *Scenarios of simulated data sets*

Scenario	Number of genes	Correlation gene effects on traits	Selection
S05	5	2 similar, 3 opposite	Random
S05s	5	2 similar, 3 opposite	Trait 1
S50	50	25 similar, 25 opposite	Random
S50s	50	25 similar, 25 opposite	Trait 1

(v) *Models of analysis*

As previously stated, two possible priors for gene effects are considered: uniform priors and normal priors. In the case of normal priors, the hyperparameters of the normal distribution (Λ) can be treated as unknown and must be estimated from the data. However, when the number of unlinked genes is small, the sampling of random variable Λ can become problematic owing to lack of degrees of freedom. Therefore, I restrict the use of normal priors for gene effects to cases where there are enough genes in the model (at least five genes). In the case of L being a random variable, this might occasionally happen in the Markov chain. The number of unlinked genes (L) is *a priori* either known (fixed) or unknown (random) (Table 2). In the case where the number of genes is treated as unknown, its prior distribution was assumed to be Poisson distributed with a mean of 2 or 20 (Table 2). For reasons of comparison, I also include models combining TIM and FPM, where the number of genes in the FPM are again treated as known or unknown and only uniform priors on gene effects are applied (Table 2).

In the following, I refer to all models with a fixed number of genes as F* models and those with a random number of genes as R* models. Furthermore, models including TIM are referred to as T* models.

3. Results

The results presented below are the averages of the posterior mean estimates from ten replicates of simulating the four scenarios given in Table 1, analysed by the 15 models given in Table 2. For each combination of scenario + model, we obtained 1000 realizations from a Markov chain, storing every 25th or 50th sample for the F* and R* models, respectively. This sampling protocol provided low autocorrelation between consecutive realizations; that is, the Monte Carlo standard errors (Geyer, 1992; Sorensen *et al.*,

Table 2. *Models of analysis in the finite polygenic model* (*FPM*)

	Number of FPM genes						
	Fixed					Random	
P(*a*)*	0	5	10	20	50	2	20
Uniform		F05u	F10u	F20u	F50u	R02u	R20u
Normal		F05n	F10n	F20n	F50n	R02n	R20n
TIM‡	TIM					TR02u	TR20u

* Prior distribution of (additive) gene effects.
‡ The infinitesimal model.

1995) were below 0·5% for (ratios of) variance components. The required computation time on a Pentium III (1 GHz) was almost 18 h for the models fitting 50 genes, whereas less than 1·5 h were needed for models fitting five genes.

(i) *Distribution of number of genes*

Table 3 presents the posterior probabilities and posterior mean estimates for the number of genes for all R* models and TR* models for the four different scenarios. The genetic models R02u and R02n resulted in highly similar posterior distributions across all scenarios, except for a small upwards shift of R02n in scenario S50s. The posterior distributions for these two models were both shifted upwards (with the mean in the direction of the simulated values) under trait selection versus random selection. The posterior mean estimates were close to five in scenario S5s and close to eight in scenario S50s.

The genetic models R20u and R20n resulted in posterior distributions that were always higher than those for the genetic models R02u and R02n, with differences in posterior mean estimates ranging from 1·1 (R20u versus R02u in S05s) to 3·2 (R20n versus R02n in S50 and S50s). The genetic model R20n consistently gave an up-shifted posterior distribution compared with model R20u. Like models R02u and R02n, trait selection caused the posterior mean estimates of models R20u and R20n to be closer to the simulated values. This shift was downwards in S05s, because the posterior distribution was higher than the simulated value in S05.

In general, a normal prior distribution on gene effects gave higher estimates for the number of genes than a uniform prior distribution, especially when many genes were expected and when selection was applied. These higher numbers might result from the expected regression towards zero for gene effects in case of normal priors. Also, trait selection provides substantial information for estimating the number of genes (Table 3).

Across all scenarios and the two different priors, the T* models always resulted in posterior mean estimates (and distributions) that were lower than those from the R* models. The differences in posterior mean estimates ranged from 0·9 (TR20u versus R20u in S05s) to 7·1 (TR20u versus R20n in S50s). For model TR02u, the effect of trait selection versus random selection resulted in differences in posterior distribution only between scenario S05 and scenario S05s, where trait selection gave higher posterior estimates (closer to the simulated value). For model TR20u, the effect of trait selection versus random selection was only observed in the posterior distributions between scenarios S50 and S50s, where trait selection gave lower posterior estimates.

(ii) *Variance components*

(a) *Scenario S05: five genes and random selection*

Table 4 presents the averaged posterior mean estimates of residual and genetic variance components for traits 1 and 2, for the 15 models given in Table 2, across ten replicates of simulation of scenario S05. Taking uniform priors on gene effects, a strong positive correlation was observed between the number of genes in the FPM and the estimated genetic variance (e.g. for trait 1, the posterior estimates were 1·07 and 1·40 for models F05u and F50u, respectively). A similar negative (but smaller) correlation was observed between the estimates for residual variance and the number of genes in the FPM model (Table 4). For models R02u and R20u, the posterior mean estimates for variance components were close to those for model F05u, which agrees well with the posterior mean estimates for the number of genes, which were 4·4 and 6·2 for R02u and R20u, respectively (Table 3). Taking normal priors on gene effects, all F* models and R* models resulted in very similar posterior mean estimates, ranging from 1·03 to 1·06 (Table 4). These results confirmed the conclusions of Pong-Wong *et al.* (1999) and Du *et al.* (1999) for random selection.

For the T* models, the posterior mean estimates for total genetic variance were slightly higher than those for the FPM models with normal priors. When combining TIM and FPM, the estimated standard deviation of the posterior densities for the two genetic variance components were much larger than fitting only one component (either TIM or FPM). This indicates that, for random selection, the phenotypic data have relatively little power to distinguish between TIM and FPM as genetic models underlying the quantitative traits.

The estimated genetic and residual correlations between the two traits were very similar in all studied models (Table 4) and agreed well with the simulated values. Those for model TR02u varied considerably, with the genetic correlation owing to the FPM being

Table 3. *Prior and posterior probabilities and posterior mean estimates of number of genes, calculated as averages over ten replicates per simulation scenario. Scenarios S05, S05s, S50, and S50s are defined in Table 1, genetic models are defined in Table 2. Prior and posterior probabilities >0·005 are shown*

	Number of genes (*L*)																
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean
Prior																	
P02*	0·14	0·27	0·27	0·18	0·09	0·04	0·01										
S05																	
R02u				0·04	0·56	0·33	0·07										4·4
R02n				0·04	0·52	0·37	0·07										4·5
TR02u		0·06	0·36	0·43	0·13	0·02											2·7
S05s																	
R02u					0·35	0·54	0·10	0·01									4·8
R02n					0·38	0·51	0·10	0·01									4·7
TR02u			0·16	0·54	0·27	0·04											3·2
S50																	
R02u				0·03	0·29	0·41	0·22	0·05									5·0
R02n				0·02	0·29	0·37	0·24	0·07	0·01								5·1
TR02u	0·05	0·40	0·36	0·15	0·04	0·01											1·7
S50s																	
R02u						0·05	0·09	0·34	0·38	0·13	0·01						7·5
R02n							0·05	0·29	0·37	0·18	0·10	0·01					8·0
TR02u	0·08	0·43	0·33	0·13	0·03												1·6
Prior																	
P20*										0·01	0·01	0·02	0·03	0·05	0·07	0·09	
S05																	
R20u					0·07	0·19	0·35	0·26	0·11	0·03							6·2
R20n					0·02	0·11	0·24	0·26	0·20	0·10	0·04	0·02					7·1
TR20u			0·01	0·07	0·24	0·32	0·22	0·11	0·03								5·2
S05s																	
R20u					0·04	0·29	0·42	0·21	0·04								5·9
R20n					0·02	0·18	0·34	0·28	0·13	0·04	0·01						6·5
TR20u				0·07	0·26	0·38	0·22	0·06	0·01								5·0
S50																	
R20u						0·09	0·20	0·27	0·27	0·13	0·04						7·3
R20n						0·02	0·09	0·25	0·25	0·20	0·12	0·05	0·02				8·3
TR20u			0·03	0·11	0·22	0·28	0·21	0·11	0·03	0·01							5·0
S50s																	
R20u							0·01	0·03	0·11	0·27	0·30	0·16	0·10	0·02			9·8
R20n								0·01	0·03	0·09	0·18	0·27	0·23	0·12	0·05	0·02	11·2
TR20u		0·03	0·10	0·24	0·26	0·22	0·11	0·04	0·01								4·1

* Priors P02 and P20 denote a Poisson prior with expected mean of 2 and 20, respectively. For P20. the probabilities for $n>15$ are not given in this table; their joint probability is $\sim 0{\cdot}72$.

slightly positive and that owing to the TIM being strongly negative. However, accounting for the TIM proportion of the total genetic variance, the overall genetic correlation for the TR02u (and TR20u) models agreed well with those from the other models ($\sim -0{\cdot}23$).

(b) *Scenario S05s: five genes and trait selection*

When selection was based on trait 1, the posterior mean estimates for genetic variances for trait 1 were clearly affected by the number of genes in the FPM models (Table 5), irrespective of which prior was taken for gene effects. Taking uniform priors, however, the trend was very different from the one observed in scenario S05 (random selection): the posterior estimates of genetic variance decreased when the number of genes increased in the F* models. This same negative trend was also observed taking a normal prior on gene effects; estimates were 1·09 and 0·84 for models F05n and F50n, respectively. In the case of normal priors, the trends of trait 1 were also observed for trait 2 but, in

Table 4. *Posterior mean estimates for error variance* (σ^2_e), *total genetic variance* (σ^2_g), *proportion of the genetic variance owing to TIM* ($Prop_{TIM}$) *for traits 1 and 2, and error correlation* (r_e) *and genetic correlation in FPM* ($r_{g,FPM}$) *and in TIM* ($r_{g,TIM}$) *for scenario S05. Calculated as the average of ten replicates. Standard errors across the ten replicates ranged between 0·04 and 0·07 for* σ^2_e, *and between 0·07 and 0·11 for* σ^2_g (*similar for both traits*)

	F05u	F10u	F20u	F50u	R02u	R20u	F05n	F10n	F20n	F50n	R02n	R20n	TIM	TR02u	TR20u
Trait 1															
σ^2_e	0·98	0·96	0·93	0·85	0·99	0·98	0·99	0·98	0·98	0·98	1·00	0·99	0·98	0·98	0·97
σ^2_g	1·07	1·10	1·18	1·40	1·05	1·07	1·03	1·05	1·06	1·06	1·03	1·04	1·09	1·08	1·09
$Prop_{TIM}$													1·00	0·40	0·21
Trait 2															
σ^2_e	0·98	0·95	0·93	0·84	0·98	0·97	0·98	0·97	0·97	0·98	0·99	0·98	0·97	0·96	0·95
σ^2_g	1·07	1·10	1·17	1·38	1·06	1·08	1·06	1·06	1·06	1·05	1·05	1·06	1·10	1·11	1·12
$Prop_{TIM}$													1·00	0·49	0·35
Correlation															
r_e	0·02	0·02	0·03	0·03	0·02	0·02	0·03	0·03	0·03	0·03	0·02	0·03	0·03	0·04	0·03
$r_{g,FPM}$	−0·22	−0·22	−0·21	−0·18	−0·23	−0·22	−0·23	−0·23	−0·23	−0·24	−0·23	−0·23		0·05	−0·14
$r_{g,TIM}$													−0·25	−0·61	−0·46

Table 5. *Posterior mean estimates for error variance* (σ^2_e), *total genetic variance* (σ^2_g), *proportion of the genetic variance owing to TIM* ($Prop_{TIM}$) *for traits 1 and 2, and error correlation* (r_e) *and genetic correlation in FPM* ($r_{g,FPM}$) *and in TIM* ($r_{g,TIM}$) *for scenario S05s. Calculated as the average of ten replicates. Standard errors across the ten replicates ranged between 0·02 and 0·04* (*trait 1*) *and 0·06 and 0·09* (*trait 2*) *for* σ^2_e, *and between 0·02 and 0·06* (*trait 1*) *and 0·08 and 0·14* (*trait 2*) *for* σ^2_g

	F05u	F10u	F20u	F50u	R02u	R20u	F05n	F10n	F20n	F50n	R02n	R20n	TIM	TR02u	TR20u
Trait 1															
σ^2_e	1·02	1·01	0·99	0·97	1·02	1·02	1·02	1·01	1·00	0·99	1·02	1·02	0·99	1·00	1·01
σ^2_g	1·11	1·04	1·02	1·01	1·12	1·08	1·09	1·01	0·95	0·84	1·11	1·06	0·80	1·05	1·06
$Prop_{TIM}$													1·00	0·23	0·13
Trait 2															
σ^2_e	0·97	0·95	0·91	0·81	0·98	0·97	0·98	0·96	0·95	0·94	0·99	0·97	0·95	0·97	0·96
σ^2_g	1·04	1·08	1·12	1·24	1·02	1·04	1·01	1·01	0·98	0·86	1·00	1·01	0·83	1·03	1·06
$Prop_{TIM}$													1·00	0·18	0·12
Correlation															
r_e	0·01	0·01	0·01	0·01	0·01	0·01	0·01	0·01	0·02	0·02	0·01	0·01	0·02	0·01	0·02
$r_{g,FPM}$	−0·21	−0·20	−0·19	−0·19	−0·21	−0·21	−0·21	−0·21	−0·22	−0·24	−0·21	−0·21		−0·15	−0·14
$r_{g,TIM}$													−0·23	−0·42	−0·67

the case of uniform priors, the trend of trait 2 was similar to those observed in the scenario S05 (random selection). The posterior mean estimates for the R* models were all very close to those obtained by the F05u and F05n models, because the posterior mean for the number of genes in these models was close to 5 (Table 3). TIM gave posterior mean estimates that were very similar to those in F50n – relatively very low estimates for genetic variance. The estimates for (total) genetic variance for the TR* models were close to the simulated values, where the proportions of genetic variance explained by TIM were only about 0·2 and 0·1, respectively, for the two traits in models TR02u and TR20u. The posterior mean estimates for residual and (overall) genetic correlation were consistent across all models studied and close to the simulated values (Table 5).

(c) *Scenario S50: 50 genes and random selection*

In this scenario the trends of results for the different models of analysis were similar to those in scenario S05. Taking a uniform prior on gene effects, the posterior mean estimates for genetic and residual variances increased and decreased, respectively, with the number of genes in the F* models (Table 6). Taking a normal prior on gene effects, the number of genes in the model did not affect the estimates for variance components – the points estimates were very similar across all models. It is also noteworthy that, in this scenario,

Table 6. *Posterior mean estimates for error variance* (σ^2_e), *total genetic variance* (σ^2_g), *proportion of the genetic variance owing to TIM* ($Prop_{TIM}$) *for traits 1 and 2, and error correlation* (r_e) *and genetic correlation in FPM* ($r_{g,FPM}$) *and in TIM* ($r_{g,TIM}$) *for scenario S50. Calculated as the average of ten replicates. Standard errors across the ten replicates ranged between 0·05 and 0·08 for* σ^2_e, *and between 0·07 and 0·12 for* σ^2_g (*similar for both traits*)

	F05u	F10u	F20u	F50u	R02u	R20u	F05n	F10n	F20n	F50n	R02n	R20n	TIM	TR02u	TR20u
Trait 1															
σ^2_e	1·01	0·97	0·94	0·86	1·01	0·99	1·01	0·99	0·99	0·98	1·02	1·00	0·97	0·98	0·98
σ^2_g	1·01	1·06	1·13	1·33	1·00	1·03	0·99	1·00	1·01	1·02	0·98	1·00	1·06	1·05	1·05
$Prop_{TIM}$													1·00	0·77	0·41
Trait 2															
σ^2_e	1·06	1·02	0·98	0·89	1·06	1·04	1·06	1·04	1·04	1·03	1·06	1·05	1·02	1·02	1·02
σ^2_g	0·94	0·99	1·08	1·31	0·94	0·97	0·93	0·96	0·96	0·97	0·93	0·94	1·02	1·00	1·02
$Prop_{TIM}$													1·00	0·76	0·53
Correlation															
r_e	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·01	0·00	0·00	0·01	0·00	0·00
$r_{g,FPM}$	0·03	0·02	0·02	0·02	0·03	0·02	0·03	0·02	0·02	0·01	0·03	0·02		−0·01	−0·06
$r_{g,TIM}$													0·01	0·03	0·13

Table 7. *Posterior mean estimates for error variance* (σ^2_e), *total genetic variance* (σ^2_g), *proportion of the genetic variance owing to TIM* ($Prop_{TIM}$) *for traits 1 and 2, and error correlation* (r_e) *and genetic correlation in FPM* ($r_{g,FPM}$) *and in TIM* ($r_{g,TIM}$) *for scenario S50s. Calculated as the average of ten replicates. Standard errors across the ten replicates ranged between 0·02 and 0·04* (*trait 1*) *and 0·06 and 0·09* (*trait 2*) *for* σ^2_e, *and between 0·06 and 0·08* (*trait 1*) *and 0·09 and 0·13* (*trait 2*) *for* σ^2_g

	F05u	F10u	F20u	F50u	R02u	R20u	F05n	F10n	F20n	F50n	R02n	R20n	TIM	TR02u	TR20u
Trait 1															
σ^2_e	1·07	1·04	1·03	0·99	1·05	1·05	1·07	1·05	1·04	1·03	1·05	1·05	1·02	1·02	1·03
σ^2_g	1·43	1·25	1·18	1·18	1·31	1·25	1·41	1·23	1·12	1·03	1·28	1·20	0·98	1·00	1·03
$Prop_{TIM}$													1·00	0·88	0·72
Trait 2															
σ^2_e	1·05	0·99	0·95	0·86	1·01	0·99	1·05	1·01	1·00	0·98	1·02	1·00	0·98	0·98	0·98
σ^2_g	1·05	1·13	1·19	1·39	1·10	1·13	1·04	1·08	1·07	1·06	1·07	1·08	1·07	1·08	1·10
$Prop_{TIM}$													1·00	0·74	0·53
Correlation															
r_e	0·01	0·01	0·01	0·01	0·01	0·01	0·01	0·01	0·01	0·01	0·01	0·01	0·01	0·01	0·01
$r_{g,FPM}$	0·04	0·03	0·02	0·01	0·03	0·03	0·04	0·03	0·02	0·01	0·03	0·03		−0·09	−0·05
$r_{g,TIM}$													0·01	0·04	0·05

with 50 genes simulated, model F50u leads to a severe overestimation of the genetic variance, whereas model F05u seems to give consistent estimates, and the number of genes fitted is much lower than simulated.

The posterior mean estimates for genetic variance for the T* models were slightly higher than those for the models with normal priors. The proportion of genetic variance explained by the TIM decreased with the number of genes fitted in the FPM: 0·77 and 0·41 in models TR02u and TR20u, respectively. In TR02u and TR20u, the posterior mean estimates for the number of genes were 1·7 and 5·0, respectively (Table 3). The posterior mean estimates for the residual and genetic correlation agreed very well with the simulated value of 0·0.

(d) *Scenario S50s: 50 genes and trait selection*

Table 7 shows that the posterior mean estimates for residual and genetic variances have very similar trends to those in scenario S05s. Taking uniform priors on gene effects, posterior mean estimates for residual variances decreased for both traits when the number of genes in the FPM increased. For the genetic variance, the posterior means for trait 1 and trait 2 decreased and increased when the number of genes increased. Similar to scenario S50, the genetic variance for trait 2 seemed to be severely overestimated by model F50u. Accounting for the number of genes fitted in models R02u and R20u (Table 3), these models gave consistent estimates with the F* models with uniform priors. Taking

normal priors, the posterior mean estimates for genetic variance decreased with the number of genes in the model, whereas those for residual variance also decreased slightly. The trend observed for genetic variance of trait 1 was not observed for trait 2, whereas the trend for residual variance of trait 1 was clearer in trait 2. The models TR02u and TR20u gave very similar estimates to TIM and these estimates were the most consistent with the simulated values. In these models, the proportion of genetic variance explained by TIM was always higher than 0·5, differing between the two models and the two traits (Table 7). There was a small decrease in genetic correlation when the number of genes in the F* models increased, regardless of which prior was taken. For all TR02u and TR20u models, the genetic correlation for the FPM and TIM were slightly negative and positive, respectively, although the posterior densities were very broad (results not shown). Their overall genetic correlation was again very consistent with those in the F* and R* models.

4. Discussion

(i) *Method*

A Bayesian method was presented to study the genetic architecture of multiple quantitative traits by exploring a range of FPMs, some combined with TIM. Within the FPM, the number of genes is modelled as an additional random variable by using a reversible jump algorithm. In the proposed FPM, the genes can act as additive or dominant but, in the simulation study, we concentrated on additive genes only. The multiple-traits implementation facilitates the study of the behaviour of individual genes at correlated traits. This would allow the identification of genes that act favourably on all traits of interest, either positively or negatively correlated traits. I realise that the scenarios studied here are limited and more simulation studies are required to evaluate the potential of our method to pick up interesting (major) genes that act on multiple traits in different directions and different magnitudes. For example, simulating many small genes and one major gene might do this where the major gene acts upon one or more traits. However, this was outside the scope of this paper.

(ii) *Simulation results*

The analysis with uniform priors on gene effects in our simulations confirmed the previous findings of Pong-Wong *et al.* (1999) and Du *et al.* (1999) that the number of genes in the model can greatly affect the estimates of genetic variance in the FPM. That is, for more genes, the estimates of genetic and residual variances increased and decreased, respectively. In the case of random selection, this dependency was not observed taking normal priors on gene effects, irrespective of the genetic model that was used in the simulation. However, in the case of trait selection, there were dependencies between parameter estimates and the number of genes, irrespective of the choice of priors on gene effects. For the selected trait, the number of genes was negatively correlated with the estimated genetic variance. This resulted in an underestimation by the F* models with more genes in scenario S05s, and in an overestimation by the F* models with fewer genes in scenario S50s. Treating the number of genes as a random variable did make the FPM more robust to trait selection, but the posterior mean estimates from these models were still somewhat biased when many genes was simulated. The prior expected number of genes in models R02n and R20n was much lower than the number simulated in scenario S50s, although models such as R50n were not studied to see whether the estimates would improve or not. Across the two scenarios of trait selection, models TR02u and TR20u proved to be very robust; that is, they yielded consistent estimates for variance components. In scenario S05s, the random FPM explains most of the variance and, in scenario S50s, TIM explains most of the variance (see also Fig. 1). Treating the number of genes as random primarily facilitates this desired flexibility. However, one should treat the posterior distribution of the number of genes in the FPM as indicative, because this number cannot easily be inferred from phenotypic data alone, unless the number of genes is very small and they have very different contributions to multiple traits.

(iii) *Mixed inheritance models*

For the TIM (TF00u), the estimated genetic variance under trait selection was much lower than the estimates from the F* and R* models. The difference was largest in the S05s scenario but also still present in the S50s scenario. These differences in estimates might be explained by the different model assumptions – allowing for changes in allele frequencies that interfere with selection acting on one of the traits. Under assumption of normality, all models performed similarly in the case of random selection but very differently in the case of trait selection. Tables 4–7 indicated that the TR* models with random number of genes performed well under very different models of simulation. Therefore, when one is interested in robust estimation of variance components (and not, for example, in detecting major genes), I propose the use of model TR02 (a combination of TIM and FPM with a random number of genes). In this study, I only considered uniform priors on gene effects for this model but normal priors could be used as well, although the scale parameters of this normal distribution can not then be inferred from the

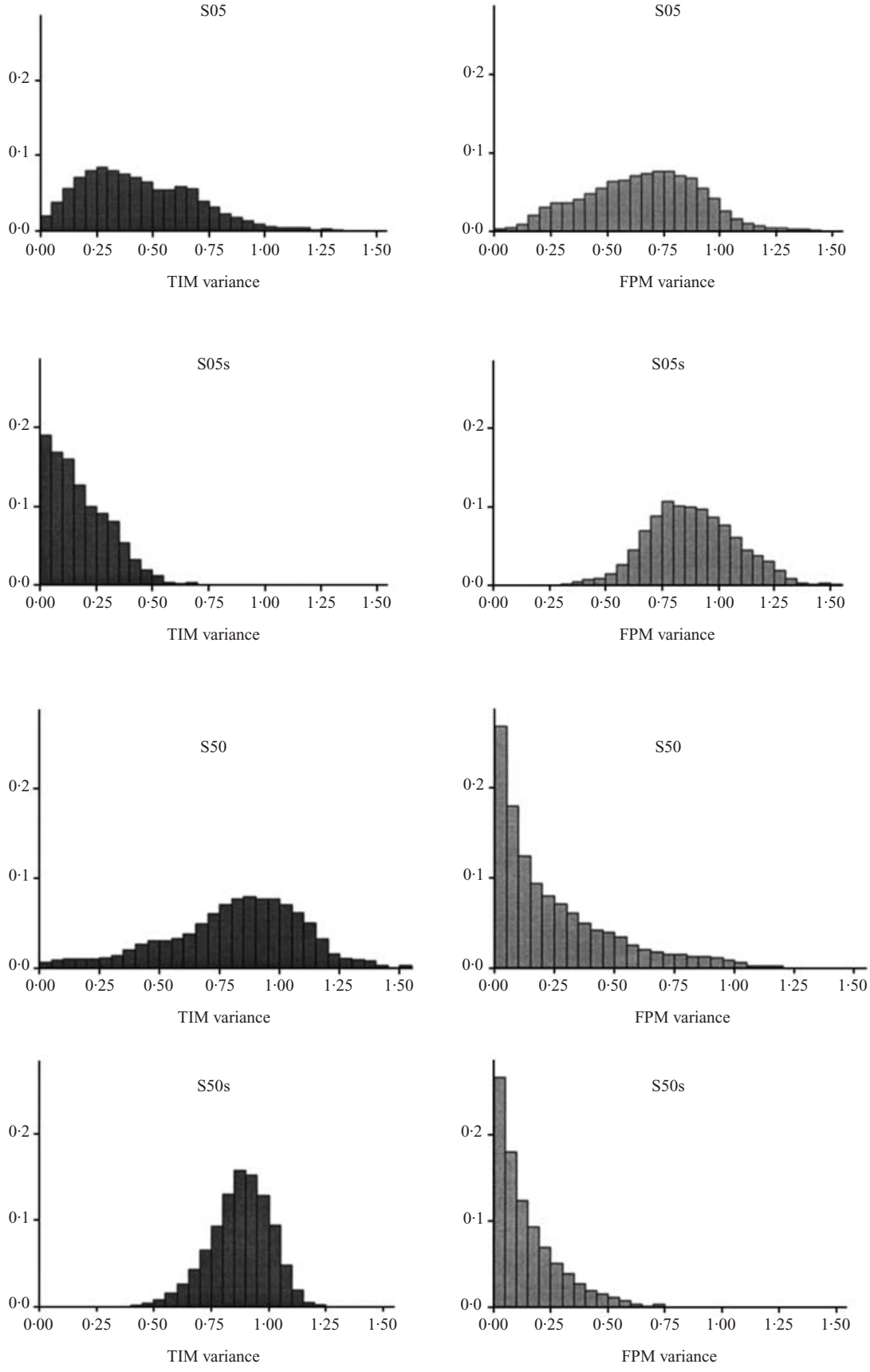

Fig. 1. Marginal posterior distributions of genetic variance owing to TIM and FPM in model TR02u for four different scenarios (as averages over ten replicates).

data. That is, there might be too few genes present, leading to too few degrees of freedom for the IW distribution. Taking the scale parameters as fixed solves this problem but one should then have some prior indication of the size of these scale parameters. The effects of taking fixed scale parameters needs to be studied more precisely.

(iv) *Major-gene detection*

The results also point to the need for some caution with respect to the detection of major genes. For example, the FPM part of the TR* models explained a significant fraction of the genetic variance in the S50 and S50s simulation scenarios, although no major genes were simulated, a result that was also reported by Tier & Henshall (2001). In these cases, one should base inferences on the full posterior density of the genetic variance (or heritability) of the FPM instead of only the posterior mean. Figure 1 clearly shows that, in scenarios S50 and (especially) S50s, there is a relatively high posterior probability of having zero or very little variance owing to the FPM. Only when there is very little or no mass near zero might a major gene be postulated. So, for scenarios S50 and S50s, I would have rejected the presence of major genes based upon the posterior distribution of the genetic variance or heritability. Also, one may examine the effects of individual genes, although it should be realized that their identity cannot be established in cases of multiple genes. Notice also that I simulated genes of equal size, whereas unequally sized genes might be more realistic and the performance of a FPM with random number of genes remains unknown. A natural way to reduce the false discovery of major genes is by imposing a very stringent prior on the number of genes by imposing a high prior probability mass at or near zero. By the same reasoning, the TIM model would have been excluded or rejected in scenario S05s based on the high posterior probability near zero (Fig. 1).

(v) *Sampling genotypes*

Du *et al*. (1999) suggested sampling the genotypes of parents unconditionally on their final offspring, which is an extension of the sampling scheme of Janss *et al*. (1995). Du & Hoeschele (2000) explored a genotype-sampling scheme based on descent graphs (Thompson, 1994, Sobel & Lange, 1996) to sample the genotypes at all loci jointly. However, Du & Hoeschele (2000) reported that descent-graph sampling of genotypes only modestly improved variance-component estimation compared with Gibbs sampling. I sampled genotypes for genes using segregation indicators (Lange & Matthysse, 1989; Thompson, 1994) and did not observe unsatisfactory mixing. However, more research is needed to evaluate the mixing properties of segregation indicators for large and deep pedigrees, because these might exist in practical livestock breeding programs. Then, the mixing of segregation indicators or founder alleles for ancient parents might be hampered by the large amount of information coming from all their descendants.

(vi) *Extensions*

Estimation of non-additive genetic variances can be a major incentive to implement FPM; this was not explored in this study but has already been studied by Du *et al*. (1999) and Du & Hoeschele (2000). The next step in my method is to incorporate molecular marker data to map QTLs to chromosomes. This novel Bayesian approach then allows three sources of genetic variance (QTL, FPM, and TIM). This method of mapping QTLs has already been applied to data from humans (Uimari & Sillanpää, 2001) and plants (Bink *et al*. 2002), but no allowance has yet been made for unlinked genes. In practice, mapping experiments are likely to provide a limited number of marker loci and these markers only partially cover the genome. For example, I have been analysing data from pig selection lines, in which marker data are available for a single chromosome, and used the flexibility of the FPM to account for putative QTLs on other (unmarked) chromosomes. Furthermore, combinations of QTL and FPM models will also be used to analyse data on complex plant pedigrees, for which the presence of high rates of inbreeding hampers the use of TIM. In this way, experimental and non-experimental data on animal and plant breeding populations will be used more efficiently to explore the genetic variation underlying quantitative traits.

The contributions of P. Uimari to the software are gratefully acknowledged. Suggestions from T. Meuwissen, R. Jansen and an anonymous reviewer on earlier versions of the manuscript are acknowledged.

References

Bink, M. C. A. M., Uimari, P., Sillanpää, M. J., Janss, L. L. G. & Jansen, R. C. (2002). Multiple QTL mapping in related plant populations via a pedigree-analysis approach. *Theoretical and Applied Genetics* **104**, 751–762.

DeBoer, I. J. M. & Hoeschele, I. (1993). Genetic evaluation methods for populations with dominance and inbreeding. *Theoretical and Applied Genetics* **86**, 245–258.

Du, F. X., Hoeschele, I. & Gage-Lahti, K. M. (1999). Estimation of additive and dominance variance components in finite polygenic models and complex pedigrees. *Genetical Research* **74**, 179–187.

Du, F. X. & Hoeschele, I. (2000). Estimation of additive, dominance, and epistatic variance components using finite locus models implemented with a single-site Gibbs and a descent graph sampler. *Genetical Research* **76**, 187–198.

Falconer, D. S. (1989). *Introduction to Quantitative Genetics*, 3rd edn. Harlow, UK: Longman.

Fernando, R. L., Stricker, C. & Elston, R. C. (1994). The finite polygenic mixed model: an alternative formulation for the mixed model of inheritance. *Theoretical and Applied Genetics* **88**, 573–580.

Fisher, R. A. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Trans. Royal Society Edinburgh* **52**, 399–433.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statistical Science* **7**, 467–511.

Green, P. J. (1995). Reversible jumping Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732.

Guo, S. W. & Thompson, E. A. (1992). A Monte Carlo method for combined segregation and linkage analysis. *American Journal of Human Genetics* **51**, 1111–1126.

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.

Lander, E. S. & Green, P. (1987). Construction of multilocus genetic linkage maps in humans. *Proceedings of the National Academy of Sciences of the USA* **84**, 2363–2367.

Lange, K. & Matthysse, S. (1989). Simulation of pedigree genotypes by random walks. *American Journal of Human Genetics* **45**, 959–970.

Lee, J. K. & Thomas, D. C. (2000). Performance of Markov chain–Monte Carlo approaches for mapping genes in oligogenic models with an unknown number of loci. *American Journal of Human Genetics* **67**, 1232–1250.

Martinez, V., Bunger, L. & Hill, W. G. (2000). Analysis of response to 20 generations of selection for body composition in mice: fit to infinitesimal model assumptions. *Genetics Selection Evolution* **32**, 3–21.

Pong-Wong, R., Haley, C. S. & Woolliams, J. A. (1999). Behaviour of the additive finite locus model. *Genetics Selection Evolution* **31**, 193–211.

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.

Sorensen, D. A., Andersen, S., Gianola, D. & Korsgaard, I. (1995). Bayesian inference in threshold models using Gibbs sampling. *Genetics Selection Evolution* **27**, 229–249.

Sorensen, D. A. (1996). *Gibbs Sampling in Quantitative Genetics*. Danish Institute of Agricultural Sciences. DIAS Internal report no. 82, Foulum, Denmark.

Stephens, D. A. & Fisch, R. D. (1998). Bayesian analysis of quantitative trait locus data using reversible jump Markov chain Monte Carlo. *Biometrics* **54**, 1334–1347.

Thompson, E. A. & Skolnick, M. H. (1977). Likelihoods on complex pedigrees for quantitative traits. *Proceedings of the International Conference on Quantitative Genetics* (eds E. Pollack, O. Kempthorne & T. B. Bailey Jr), pp. 815–818. Ames: Iowa State University Press.

Thompson, E. A. (1994). Monte Carlo likelihood in genetic mapping. *Statistical Science* **9**, 355–366.

Thompson, E. A. (1999). *Statistical Inference From Genetic Data on Pedigrees*. Institute of Mathematical Statistics Books. Beachwood, OH: USA.

Tier, B. & Henshall, J. (2001). A sampling algorithm for segregation analysis. *Genetics Selection Evolution* **33**, 587–603.

Uimari, P. & Sillanpää, M. (2001). A Bayesian MCMC linkage analysis with segregation indicators for complex pedigrees. *Genetic Epidemiology* **21**, 224–242.

Van Tassel, C. P. & Van Vleck, L. D. (1996). Multiple-trait Gibbs sampler for animal models: flexible programs for Bayesian and likelihood-based (co)variance component inference. *Animal Science* **74**, 2586–2597.

Waagepetersen, R. & Sorensen, D. (2001). A tutorial on reversible jump MCMC with a view toward applications in QTL mapping. *Review of the International Statistical Institute* **69**, 49–62.

Wang, C. S., Rutledge, J. J. & Gianola, D. (1993). Marginal inferences about variance components in a mixed linear model using Gibbs sampling. *Genetics Selection Evolution* **25**, 41–62.