Genet. Res., Camb. (2002), **80**, *pp.* 237–243. *With* 1 *figure.*
DOI: 10.1017/S0016672302005918 Printed in the United Kingdom

Simple deterministic identity-by-descent coefficients and estimation of QTL allelic effects in full and half sibs

Y. NAGAMINE[1]*, S. A. KNOTT[2], P. M. VISSCHER[2] AND C. S. HALEY[1]
[1] *Roslin Institute* (*Edinburgh*), *Midlothian, EH25 9PS, UK*
[2] *Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh, EH9 3JT, UK*

(*Received 11 February 2002 and in revised form 28 June and 12 August 2002*)

Summary

Accurate and rapid methods for the detection of quantitative trait loci (QTLs) and evaluation of consequent allelic effects are required to implement marker-assisted selection in outbred populations. In this study, we present a simple deterministic method for estimating identity-by-descent (IBD) coefficients in full- and half-sib families that can be used for the detection of QTLs via a variance-component approach. In a simulated dataset, IBD coefficients among sibs estimated by the simple deterministic and Markov chain Monte Carlo (MCMC) methods with three or four alleles at each marker locus exhibited a correlation of greater than 0·99. This high correlation was also found in QTL analyses of data from an outbred pig population. Variance component analysis used both the simple deterministic and MCMC methods to estimate IBD coefficients. Both procedures detected a QTL at the same position and gave similar test statistics and heritabilities. The MCMC method, however, required much longer computation than the simple method. The conversion of estimated QTL genotypic effects into allelic effects for use in marker-assisted selection is also demonstrated.

1. Introduction

The detection and ultimate identification of quantitative trait loci (QTLs) in livestock will aid understanding of the causes of quantitative genetic variation and will provide tools to enhance further the success of breeding programmes. Many of the initial QTL mapping studies have focused on experimental crosses between genetically diverse populations (e.g. Andersson *et al*., 1994). Recently, more effort is being focused on studies within outbred populations, where QTL segregation is immediately relevant to within breed improvement. Detection of QTLs in such populations is more challenging than it is in data from designed experiments and several alternative methods of analysis have been proposed. One approach that has been applied for analysis of data from simple population structures is that based on least squares (Knott *et al*., 1996; Hoeschele *et al*., 1997; Knott *et al*., 1998; de Koning *et al*., 1999). The least-squares (LS) method is especially useful when performing a genome scan or implementing permutation or bootstrap analyses (Visscher *et al*., 1996), because it does not require large computational resources. However, the LS approach assumes that all effects are fixed and therefore cannot be used to estimate the polygenic (i.e. non-QTL) effect through the relationship matrix. Also, this method is used to estimate the differences between two allelic effects of QTLs within a parent but not the actual effect of each allele.

The variance component method with a random QTL effect has some advantages for the estimation of QTL genotypic or allelic effects (Xu & Atchley, 1995; Grignola *et al*., 1996*a*, 1996*b*), because this method does not require specification of the number of alleles, and polygenic effects can be estimated through the relationship matrix separately from QTL effects. Grignola *et al*. (1996*a*) introduced restricted or residual maximum likelihood (REML) to estimate QTL variance components. A two-step process was applied by George *et al*. (2000) to estimate first identity-by-descent (IBD) coefficients and subsequently the variance component caused by a QTL using available

* Corresponding author. Department of Animal Breeding and Reproduction, National Institute of Livestock and Grassland Science, Tsukuba Norin Kenkyu-danchi, 305-0901, Japan. e-mail: nagamine@affrc.go.jp
*Permanent address: National Institute of Livestock and Grassland Science, Tsukuba 305-0901, Japan.

computer packages (i.e. LOKI (Heath, 1997) and ASReml (Gilmour *et al.*, 1999), respectively, for the two steps). This is a flexible approach because, in principle, LOKI can handle pedigrees of any structure and can cope with missing marker information through a Markov chain Monte Carlo (MCMC) method. The ASReml program is similarly flexible. One problem with this approach is the computational time required for each step, especially the time taken during the MCMC procedure to estimate IBD coefficients at test positions near the markers.

For the second step, estimation of the variance due to the QTL, it is possible to fit the QTL as either genotypic or allelic effects in a mixed model (Grignola *et al.*, 1996*a*; Almasy & Blangero, 1998). Although the genotypic and allelic effect models are equivalent when it comes to QTL detection in a simple additive model, the advantage of the QTL allelic model is that it provides direct information for marker-assisted selection (Fernando & Grossman, 1989). By contrast, the allelic or gametic IBD matrix is four times as large as the genotypic IBD matrix, thus increasing the computational problem, especially in the context of a genome scan examining multiple positions.

In this study we have extended the simple deterministic gametic probability estimation from the half-sib LS method (Knott *et al.*, 1996) to estimate IBD coefficients for full- and half-sib populations. Furthermore, these IBD coefficients are compared to those obtained with the MCMC method for both simulated and real datasets. In the second step of our model, genotypic effects are used instead of allelic effects, because it is considerably faster to invert the genotypic IBD matrix in Henderson's mixed-model equations (Henderson, 1984). After estimation of the QTL position, we present a method to convert genotypic effects into allelic effects without inversion of the allelic IBD matrix.

2. Methods

(i) *Simple deterministic* (*SMD*) *estimation of IBD coefficients for pairs of sibs*

Knott *et al.* (1996) introduced a method for calculating the conditional probability of inheriting alleles at given locations from a parental gamete given parent and offspring genotypes at linked markers. This method is extended to estimate IBD coefficients for full- and half-sib populations. We assumed that the half-sib and full-sib populations had no inbreeding (i.e., no relationship between parents) for simplicity in demonstrating the analysis. The two IBD matrices were designated as the 'allelic IBD matrix' and the 'genotypic IBD matrix', signifying the allelic and genotypic relationships, respectively, between individuals at a particular test position.

To demonstrate the simple deterministic method to estimate IBD coefficients we follow the example of Knott *et al.* (1996), in which a sire has the following genotype for six loci A–F spaced at 20 cM intervals: Aa BB Cc dd EE Ff. Only markers that are heterozygous in the sire are relevant for the reconstruction of the sire's haplotypes. These haplotypes are determined using the number of offspring inheriting the four alternative gametes for adjacent pairs of loci. There are two alternative haplotype reconstructions for a pair of loci for which a sire is heterozygous. Each of the four possible progeny gamete would represent a parental gamete under one reconstruction or a recombinant gamete under the alternative reconstruction. The haplotype reconstruction that minimizes the proportion of recombinant progeny gametes for a pair of adjacent loci is selected. If the two reconstructions are equally likely, one is selected at random. For the following calculations, we assume that two haplotypes were determined from the sire (Knott *et al.*, 1996):

Haplotype 1:	A	c	f
Haplotype 2:	a	C	F

Three offspring were genotyped as follows,

HS1:	AA	cc	ff
HS2:	aa	Cc	ff
HS3:	aa	00	FF

where 00 indicates an unknown genotype. Having reconstructed the haplotypes of the sire, the procedures below were used to estimate the IBD coefficient between sibs. For illustration the coefficient at 30 cM from marker A (i.e. between marker loci B and C) was used (the QTL test position).

(a) *Estimation of the recombination rate*

The recombination rate (r) between markers and that between markers and the QTL test position were estimated from the known marker map distances (d cM), using Haldane's mapping function (Haldane, 1919).

$$r = (1 - e^{-0{\cdot}02d})/2.$$

In our example, the recombination rate, r_{AQ}, between marker A and the QTL test position (Q) is $(1 - e^{-0{\cdot}02 \times 30})/2 = 0{\cdot}225$. The other recombination rates involved are $r_{QC} = 0{\cdot}091$, $r_{AC} = 0{\cdot}275$, $r_{QF} = 0{\cdot}377$ and $r_{AF} = 0{\cdot}432$.

(b) *Estimation of conditional probabilities*

The second step involves estimation of the conditional probability of inheriting the allele for the QTL from one of the sire haplotypes (e.g. sire haplotype 1). Probabilities are calculated by using $1 - r$ when no crossovers occurred (i.e. when adjacent alleles have

https://doi.org/10.1017/S0016672302005918 Published online by Cambridge University Press

been inherited from the same sire haplotype) and r when a crossover must have occurred. In this example, the probabilities of inheriting the QTL allele from haplotype 1 of the sire are: $(1-r_{AQ})(1-r_{QC})/(1-r_{AC})=0{\cdot}972$ for HS1, $r_{AQ}(1-r_{QF})/r_{AF}=0{\cdot}325$ for HS2 and $r_{AQ}r_{QF}/(1-r_{AF})=0{\cdot}149$ for HS3. The marker F is used instead of marker C for HS2 and HS3 because, for HS2, it cannot be determined which allele at marker C has been inherited from the sire and HS3 has no marker genotype for C.

(c) *Estimation of the paternal IBD (PA-IBD) coefficient*

If two half sibs have the conditional probabilities $pr1$ and $pr2$ of inheriting the QTL allele from sire haplotype 1, the paternal IBD (PA-IBD) coefficient between them is determined as

$$\text{PA-IBD coefficient}=[\text{pr1 pr2}+(1-\text{pr1})(1-\text{pr2})]/2.$$

In our example data, the PA-IBD coefficient between HS1 and HS2 was estimated as $[0{\cdot}972\times0{\cdot}325+(1-0{\cdot}972)\times(1-0{\cdot}325)]/2=0{\cdot}167$. The PA-IBD coefficient between HS1 and HS3 was estimated as 0·084, and that between HS2 and HS3 as 0·311.

When an offspring has only one informative marker in a linkage group, the recombination rate between this marker and the QTL test position is used.

If dams are genotyped, the maternal allelic IBD (MA-IBD) coefficients are calculated in the same manner. The sum of the PA-IBD and the MA-IBD coefficients is taken as the genotypic IBD coefficient between full sibs. In the absence of informative markers, the PA-IBD or MA-IBD coefficient was set to 0·25.

(ii) *Comparison of genotypic IBD estimation methods in simulated data*

Genotypic IBD coefficients estimated with simple deterministic (SMD) and MCMC methods were compared using simulated data. For the first dataset, 300 progeny were generated from ten sires and 300 dams. This is a simple half-sib population, with each sire having 30 progeny and each dam only one progeny. In the second population, 300 progeny were generated from ten sires and 60 dams, whereby one sire was mated with six dams. This resulted in a half-sib and full-sib population, with each sire producing six half-sib families and each dam having five progeny as a full-sib family. In both types of dataset, a 40 cM chromosome was simulated, with five marker loci located at positions 0, 10, 20, 30 and 40 cM with genotypes available on sires, dams and offspring. In three different sets of simulations, each marker was assumed to have two, three or four alleles, and the frequencies of all marker alleles were equal in the parent generation. Both SMD and MCMC methods were applied to estimate IBD matrices. Additionally, a fully informative IBD (FIN-IBD) matrix was obtained because, in a simulated data set, the alleles inherited by all progeny were known for all markers. The correlations between SMD and FIN-IBD coefficients, and between MCMC and FIN-IBD coefficients were used as criteria to indicate the precision of the estimated IBD coefficients from the two methods. The program LOKI (Heath, 1997) was used to estimate IBD coefficients with the MCMC method. IBD coefficients of all combinations of full sibs and half sibs (4350 pairs) were estimated at the 5 cM and 25 cM positions in the linkage group.

(iii) *Conversion of QTL genotypic effect into allelic effects*

There are two equivalent animal models, one using QTL genotypic effects (**w**) and the other QTL allelic effects (**v**), with both including fixed and polygenic effects (Grignola *et al.*, 1996*a*; Almasy & Blangero, 1998).

The QTL genotypic effects model can be written as

$$\mathbf{y}=\mathbf{Xf}+\mathbf{Zu}+\mathbf{Zw}+\mathbf{e},$$

and the QTL allelic effects model can be written as

$$\mathbf{y}=\mathbf{Xf}+\mathbf{Zu}+\mathbf{ZTv}+\mathbf{e},$$

where **y** (size: $n\times1$) is a vector of phenotypic records for n animals. Vectors **w** ($n\times1$) and **v** ($2n\times1$) are estimated values for QTL genotypic and QTL allelic effects, respectively. All random values, **w**, **v** and the residual (**e**) are assumed to be normally distributed (Xu & Atchley, 1995; Fernando & Grossman, 1989). **X**, **Z** and **ZT** are incident matrixes for fixed (**f**), genotypic (**w**) and allelic effects (**v**), respectively.

T is an incident matrix relating each animal to its two allelic effects. If all animals have records, **T** is, using the Kronecker product (∗)

$$\mathbf{I}_n*[1\ 1]=\begin{pmatrix}1\ 1 & 0\ 0 & 0\ 0 & \cdots\\ 0\ 0 & 1\ 1 & 0\ 0 & \cdots\\ 0\ 0 & 0\ 0 & 1\ 1 & \cdots\\ \cdots & \cdots & \cdots & \cdots\end{pmatrix}.$$

Estimation of the parameters at the putative QTL position was performed using Henderson's mixed-model equations (Henderson, 1984), which require the inverse of the genotypic or allelic IBD matrix. **G** represents the allelic IBD matrix, and the genotypic IBD matrix (represented by **Q**) is $0{\cdot}5\mathbf{TGT}'$ (Van Arendonk *et al.*, 1994). The allelic effects model requires the inverse of **G**, which is a $2n\times2n$ matrix. The genotypic effects model, however, requires only the inverse of **Q** ($=0{\cdot}5\mathbf{TGT}'$), which is an $n\times n$ matrix.

https://doi.org/10.1017/S0016672302005918 Published online by Cambridge University Press

If there is no allelic interaction (i.e. under an additive model), vector **w** can simply be obtained from **v** as follows

$$\mathbf{w} = \mathbf{T}\mathbf{v}.$$

This equation means that the sum of two allelic values from an animal (e.g. $\mathbf{v}_{11}$ and $\mathbf{v}_{12}$) compose its genotypic value $\mathbf{w}_1 (= \mathbf{v}_{11} + \mathbf{v}_{12})$.

Of interest here, however, is the conversion from **w** to **v**, which is less straightforward:

$$\mathbf{T}\mathbf{v} = \mathbf{T}\mathbf{G}\mathbf{T}'(\mathbf{T}\mathbf{G}\mathbf{T}')^{-1}\mathbf{w}$$

$$\mathbf{T}\mathbf{v} = 0{\cdot}5\mathbf{T}\mathbf{G}\mathbf{T}'\mathbf{Q}^{-1}\mathbf{w}.$$

Therefore,

$$\mathbf{v} = 0{\cdot}5\mathbf{G}\mathbf{T}'\mathbf{Q}^{-1}\mathbf{w}.$$

Hence, **v** can be calculated without the inverse of **G** ($2n \times 2n$) and requires only the inverse of **Q** ($n \times n$). In fact, it is not necessary to calculate $\mathbf{Q}^{-1}$ for this conversion because it has already been obtained for use in the mixed-model equations to estimate **w**.

In this paper, we have used the genotypic effects model. After estimation of the QTL position, the allelic effects were obtained from the genotypic effects using the equation above.

(iv) *Example: QTL detection using data from a pig population*

(a) *Animal records*

Ten sires were mated with 146 dams to produce 391 offspring. Because each dam was mated with only one sire, 146 full-sib families and ten half-sib families were produced. Parents were assumed to be unrelated in all analyses. There were 544 average daily gain (ADG, measured in grams per day) records from these 547 animals. These animals represent approximately the best 20% and worst 20% progeny with respect to their daily gain record within each sire family. Three animals, one dam and two progeny, had no phenotypic records. The average start and end ages of the test were 86 and 134 days, respectively. Five markers on chromosome 7 (SW1354, SWR1078, TNFB, SW2019 and S0102) were used. Map distances between markers were estimated using CRI-MAP (Green *et al.*, 1990).

(b) *Correlation of estimated genotypic IBD coefficients from two methods*

A point approximately at the midpoint of the linkage group, 20 cM from marker 1, was selected as the test position. The estimated IBD values of all combinations of full sibs and half sibs (8376 pairs) were estimated from SMD and MCMC. The correlation between coefficients from two methods was calculated. The obligate values of 0 (unrelated), 0·5 (parent and offspring) and 1 (itself) were omitted in this calculation.

(c) *Statistical models to detect QTL position*

Polygenic and QTL genotypic effects were fitted as random effects and sex as a fixed effect in the REML analysis models. To test for the presence of a QTL against no QTL at a particular position, the likelihood ratio (LR) test statistic, $\log \mathrm{LR} = -2 \ln (L_0 - L_1)$, was calculated, where L_0 and L_1 are the respective likelihood values with no presence (H_0) or presence (H_1) of a QTL. The two-step approach (George *et al.*, 2000) was applied to detect the QTL position. Variance components were evaluated by REML, using the ASReml program (Gilmour *et al.*, 1999). The χ^2 distribution with one (χ^2_1) and two degrees (χ^2_2) of freedom was used to provide threshold values against which to judge the different methods (Xu & Atchley, 1995). Genotypic effects at the peak of test statistic were estimated with their standard error (Gilmour *et al.*, 1999). The two allelic effects of each animal were obtained from its genotypic effect.

3. Results

The estimated map locations for the five markers were (1) SW1354, 0 cM; (2) SWR1078, 8·9 cM; (3) TNFB, 27·5 cM; (4) SW2019, 29·3 cM; (5) S0102, 39·3 cM. These values correspond to distances from the first marker, SW1354, and are consistent with other published results (e.g. http://www.thearkdb.org/).

(i) *Comparison of IBD values from the two methods in simulated and real data sets*

The correlations of estimated IBD coefficients from SMD and MCMC methods (Table 1) were estimated as nearly one (0·99 to 1·00) when three or four alleles per marker were simulated for the first (one progeny per dam) and second (five progeny per dam) datasets. These correlations were slightly lower (0·92 to 0·95) with only two alleles per marker. IBD values obtained using both methods were highly correlated (0·94 to 0·98) with IBD values from FIN-IBD when there were three or four alleles per marker. The lowest correlations were between the IBD values from the SMD procedure and the FIN-IBD values when there were only two alleles per marker (0·76 to 0·85). The IBD values from the MCMC and the FIN-IBD in this situation had slightly higher correlations (0·81 to 0·91).

In the real data, the number of alleles at a marker varied from 5 to 13 and averaged 6·8 alleles per marker. The IBD values of all combinations of full sibs and half sibs at 20 cM were estimated. The average of the IBD coefficients obtained from each method (SMD and MCMC) was 0·259. Despite 4% of animals not

https://doi.org/10.1017/S0016672302005918 Published online by Cambridge University Press

Table 1. *Correlations among sibs' IBD coefficients in two populations. Population 1 (Pop1) contained 10 sires, 300 dams and 300 progeny, and Pop2 contained 10 sires, 60 dams and 300 progeny*

Alleles/marker	2 alleles				3 alleles				4 alleles			
Population	Pop1		Pop2		Pop1		Pop2		Pop1		Pop2	
Test position	5 cM	25 cM	5 cM	25 cM	5 cM	25 cM	5 cM	25 cM	5 cM	25 cM	5 cM	25 cM
SMD* & MCMC*	0·95	0·95	0·92	0·94	0·99	1·00	0·99	0·99	0·99	1·00	0·99	1·00
SMD & FIN*	0·76	0·80	0·82	0·85	0·94	0·98	0·94	0·97	0·95	0·97	0·94	0·98
MCMC & FIN	0·81	0·83	0·90	0·91	0·94	0·98	0·94	0·98	0·95	0·97	0·95	0·98

* SMD, MCMC and FIN represent IBD coefficients from the simple deterministic method, the Markov-chain Monte Carlo method and the fully informative assumption, respectively.

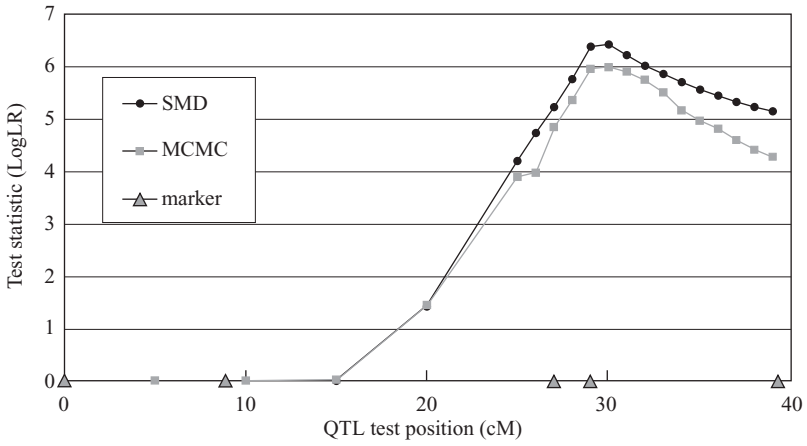

Fig. 1. Quantitative trait locus (QTL) position using identity-by-descent coefficients from the simple deterministic (SMD) and Markov-chain Monte Carlo (MCMC) methods for average daily weight gain from pig data.

being fully genotyped, the correlation between the IBD coefficients from the two methods was greater than 0·99.

(ii) *Pig data analyses*

The log LR statistic testing for the presence of QTL variance is plotted against the QTL test position in Fig. 1. The peak is observed at 30 cM for both the SMD and the MCMC methods. The peak values of the test statistic (6·44 from SMD and 6·00 from MCMC) were significant at the nominal 5% level in both χ^2 distributions ($\chi^2_1 = 3{\cdot}84, \chi^2_2 = 5{\cdot}99$). Many iterations were required in the MCMC method. Initially, 5000 iterations were applied to test every 5 cM (i.e. 5, 10, 15, 20, 25, 30 and 35 cM) from marker 1. After narrowing the candidate range to between 26 and 39 cM, 10 000 iterations were applied for every 1 cM. However, some test positions near a marker, 27 and 28 cM near marker 3 (27·5 cM) and 29 and 30 cM close to marker 4 (29·3 cM), did not yield a log LR value, because the genotypic IBD matrix was non-invertible. For these four positions, 20 000 iterations were needed to overcome this problem. The correlation between IBD values from SMD and MCMC was greater than 0·99 at 30 cM.

Estimates of the heritabilities for the polygenic and QTL genotypic effect changed depending on the test position (Table 2). With the MCMC method, we estimated a polygenic heritability of 23·2% and a QTL heritability of 0% at 5 cM. By contrast, we estimated a polygenic heritability of 1·7% and a QTL heritability of 16·2% at 30 cM. The SMD method revealed a polygenic heritability of 23·2% and a 0% heritability from the QTL at 5 cM, in contrast to values of 0% and 18·2%, respectively, at 30 cM. Evidently, in both procedures, nearly all genetic variance explained by polygenes at 5 cM was accounted for by a QTL at 30 cM. However, the QTL genotypic variance component from both methods was not precisely estimated. For example, the relative size of standard error of QTL variance component at 30 cM was 63% from both methods (Henderson, 1984; Gilmour *et al.*, 1999).

(iii) *QTL genotypic and allelic effects from the pig data*

The mean ADG phenotypic value was 1030 g day^{-1} and ranged between 688 g day^{-1} and 1455 g day^{-1}. QTL genotypic and allelic effects of all 547 animals were estimated at the 30 cM test position. QTL genotypic effects ranged from −82·8 g day^{-1} to

https://doi.org/10.1017/S0016672302005918 Published online by Cambridge University Press

Table 2. *Variance components* (σ^2) *and heritabilities* (h^2) *at 5 cM and 30 cM. The heritabilities for polygene and QTL genotypic effect are* $h_p^2 = \sigma_p^2/(\sigma_p^2+\sigma_q^2+\sigma_e^2)$ *and* $h_q^2 = \sigma_q^2/(\sigma_p^2+\sigma_q^2+\sigma_e^2)$, *respectively*

	5 cM		30 cM	
	SMD*	MCMC*	SMD	MCMC
Polygene (σ_p^2)	2795	2806	1	195
h_p^2	23·2%	23·2%	0%	1·7%
QTL (σ_q^2)	0	0	2171	1930
h_p^2	0%	0%	18·2%	16·2%
Error (σ_e^2)	9232	9225	9742	9759

* SMD and MCMC represent IBD coefficients from the simple deterministic method and the Markov-chain Monte Carlo method, respectively.

90·4 g day^{-1}, whereas QTL allelic effects varied from $-58{\cdot}7$ g day^{-1} to 53·9 g day^{-1}. Some sires showed large differences between the estimated effects of their two haplotypes at this location. For example, the QTL allelic effects of sire 1 were $-58{\cdot}7$ g day^{-1} and 17·3 g day^{-1}, whereas the genotypic effect was $-41{\cdot}4$ g day^{-1} ($=-58{\cdot}7+17{\cdot}3$ g day^{-1}) with a standard error of 28·9 g day^{-1}. The genotypic effect (=allelic effects 1+2) of sire 4 was $-10{\cdot}4$ g day^{-1} ($=-33{\cdot}3+22{\cdot}9$ g day^{-1}) with a standard error of 29·0 g day^{-1}, and that for sire 10 was 62·5 g day^{-1} ($=53{\cdot}9+8{\cdot}6$ g day^{-1}) with a standard error of 34·2 g day^{-1}.

4. Discussion

The correlation between estimated genotypic IBD coefficients obtained from the MCMC and SMD procedures was extremely high in the pig data and in the simulated data. However, the matrices of IBD coefficients estimated by the two methods had different properties. The genotypic matrix using IBD values estimated by MCMC was often non-invertible near a marker position, even with 10 000 sampling iterations, in contrast to that from the SMD method, which did not exhibit this problem. Because we used Henderson's mixed-model equations (Henderson, 1984), which require the inverse of the genotypic IBD matrix, this inability to invert the IBD matrix is a potential problem for MCMC analyses. If a pair of full sibs inherited the same marker genes from both parents, their IBD value near the marker should be very close to one. Because double recombination was considered in the deterministic method, estimated IBD values among sibs were always less than one other than at marker positions. However, with the MCMC method, estimated IBD values between sibs could equal one simply because all sampled realizations for a particular sib pair gave this value. For example, an IBD matrix using 10 000 iterations was singular at 29 cM, which was only 0·3 cM from the fourth marker (29·3 cM). This was due to the fact that a pair of full sibs that inherited the same alleles at flanking markers from their parents had, by chance, an estimated IBD coefficient between them of one. Hence, these two individuals will have the same estimated IBD coefficients with all other sibs and the two lines for these sibs in the matrix would be linearly dependent, giving a singular matrix. Increasing the number of iterations reduced the chance of two such sibs having an IBD coefficient of one when not at a marker position. The estimated IBD coefficient between sibs from SMD was very close to, but less than, one and the matrix was positive-definite. Both SMD and MCMC methods, however, will have the singular problem at the marker position and other computational strategies that do not require the inverse of the genotypic IBD matrix can be used (Visscher *et al.*, 1999).

The other problems of the MCMC method involve irreducibility and criteria for convergence. Jansen & Sheehan (1998) showed that, in a locus with more than two alleles, the underlying MCMC was not guaranteed to be irreducible. Finding convergence criteria was also difficult in practice. We used a maximum of 20 000 sampling iterations because of the large computation times required. However, even this was not always enough to obtain fully converged values. With runs starting from different random seeds, the IBD matrix might be invertible with some and non-invertible with others; even when the matrix is not singular, slight variation in the test statistic was observed. It was difficult to estimate how many iterations are sufficient to ensure convergence.

Because the parental haplotypes were estimated using only offspring informative at adjacent pairs of markers, the accuracy of these haplotypes is not expected to be high with low numbers of offspring. In the pig data, however, the average number of offspring per dam was only 2·7, and 22 dams (15%) had only one progeny. Even with a small number of offspring per dam, SMD provided fairly accurate estimated IBD values for both real and simulated datasets. This suggests that SMD is still applicable to animals that produce a limited number of progeny per dam, such as cattle and sheep.

The number of alleles at a marker has an important effect on the accuracy of estimated IBD coefficients. The IBD coefficients estimated using SMD and MCMC with two alleles per marker had a slightly lower correlation with FIN-IBD values in the simulated data than the situations with more alleles at each marker. Increasing the number of alleles increases the heterozygosity in the parents and reduces the probability that parents and offspring will be heterozygous for the same alleles, hence it is more frequently possible to determine which allele has been inherited by the progeny.

Some sires had a large difference between the effects of their QTL alleles. For example, the QTL allelic effects of sire 10 were 53·9 g day^{-1} and 8·6 g day^{-1}. This is useful information for marker-assisted selection between offspring of the same sire. If we consider only the polygenic gene effect, all full sibs without their own phenotypic records have the same expected breeding values. However, the QTL allelic effect will provide information about the different expected genetic values among sibs. This permits very early selection on traits generally measured later in life, such as fat thickness and ADG, by using their marker information.

Because only selected progeny groups within the sire family were genotyped, there is a possibility of overestimating genetic parameters. The parameters also had large standard errors as they were estimated from a small dataset. The use of unselected and larger datasets is necessary to obtain accurate parameters for marker-assisted selection. The SMD method has been shown to work for simple pedigree structures, but we need to extend it and to investigate its performance for more complicated pedigree structures.

In summary, we have presented a simple deterministic method to estimate IBD coefficients between full and half sibs, which can be used to detect QTL positions and evaluate QTL allelic effects without requiring the inverse of the allelic IBD matrix. The deterministic method is rapid and accurate when applied to a relatively simple population. Large differences between two QTL allelic effects are observed in some sires and this method is a useful tool for marker-assisted selection among sibs in commercial breeding programmes.

We acknowledge R. Pong-Wong for his comments and A. W. George for allowing the use of his computer software. We are grateful to the BBSRC and the Royal Society for funding, and for support from the QTL Transfer Consortium of the UK pig-breeding industry.

References

Almasy, L. & Blangero, J. (1998). Multipoint quantitative-traits linkage analysis in general pedigrees. *American Journal of Human Genetics* **62**, 1198–1211.

Andersson, L., Haley, C. S., Ellegren, H., Knott, S. A., Johansson, M., Andersson, K., Andersson-Eklund, L., Edfors-Lilja, I., Fredholm, M., Hansson, I., Hakansson, J. & Lundström, K. (1994). Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science* **263**, 1771–1774.

de Koning, D. J., Janss, L. L. G., Rattink, A. P., van Oers, P. A. M., de Vries, B. J., Groenen, M. A. M., van der Poel, J. J., de Groot, P. N., Brascamp, E. W. & van Arendonk, J. A. M. (1999). Detection of quantitative trait loci for backfat thickness and intramuscular fat content in pigs. *Genetics* **152**, 1679–1690.

Fernando, R. L. & Grossman, M. (1989). Marker assisted selection using best linear unbiased prediction. *Genetics Selection Evolution* **21**, 467–477.

George, A. W., Visscher, P. M. & Haley, C. S. (2000). Mapping quantitative trait loci in complex pedigrees. *Genetics* **156**, 2081–2092.

Gilmour, A. R., Cullis, B. R., Welham, S. J. & Thompson, R. (1999). *ASREML Program User Manual*. New South Wales: Orange Agricultural Institute.

Green, P., Falls, K. & Crooks, S. (1990). *Documentation for CRI-MAP* (version 2.4). St Louis: Washington University School of Medicine.

Grignola, F. E., Hoeschele, I. & Tier, B. (1996*a*). Mapping quantitative trait loci in outcross populations via residual maximum likelihood. I. Methodology. *Genetics Selection Evolution* **28**, 479–490.

Grignola, F. E., Hoeschele, I., Zhang, Q. & Thaller, G. (1996*b*). Mapping quantitative trait loci in outcross populations via residual maximum likelihood. II. A simulation study. *Genetics Selection Evolution* **28**, 491–504.

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**, 299–309.

Heath, S. C. (1997). Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *American Journal of Human Genetics* **61**, 748–760.

Henderson, C. R. (1984). *Applications of linear models in animal breeding*. Guelph: University of Guelph Press.

Hoeschele, I., Uimari, P., Grignola, F. E., Zhang, Q. & Gage, K. M. (1997). Advances in statistical methods to map quantitative trait in outbred populations. *Genetics* **147**, 1445–1457.

Jansen, C. S. & Sheehan, H. (1998). Problems with determination of noncommunicating classes for Monte Carlo Markov chain applications in pedigree analysis. *Biometrics* **54**, 416–425.

Knott, S. A., Elsen, J. M. & Haley, C. S. (1996). Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *Theoretical and Applied Genetics* **93**, 71–80.

Knott, S. A., Marklund, L., Haley, C. S., Anderson, K., Davies, W., Ellegren, H., Fredholm, M., Hansson, I., Hoyheim, B., Lundstrom, K., Moller, M. & Anderson, L. (1998). Multiple marker mapping of quantitative trait loci in a cross between outbred wild boar and Large White pigs. *Genetics* **149**, 1069–1080.

Van Arendonk, J. A. M., Tier, B. & Kinghorn, B. P. (1994). Use of multiple genetic markers in prediction of breeding values. *Genetics* **137**, 319–329.

Visscher, P. M., Thompsom, R. & Haley, C. S. (1996). Confidence intervals in QTL mapping by bootstrapping. *Genetics* **143**, 1013–1020.

Visscher, P. M., Haley, C. S., Heath, S. C., Muir, W. J. & Blackwood, D. H. R. (1999). Detecting QTLs for uni and bipolar disorder using a variance component method. *Psychiatric Genetics* **9**, 75–84.

Xu, S. & Atchley, W. R. (1995). A random model approach to interval mapping of quantitative trait loci. *Genetics* **141**, 1189–1197.