Theor Appl Genet (1996) 93: 1167–1174

G. Thaller · I. Hoeschele

A Monte Carlo method for Bayesian analysis of linkage between single markers and quantitative trait loci. II. A simulation study

Received: 30 May 1996 / Accepted: 14 June 1996

Abstract A Bayesian approach to the statistical mapping of Quantitative Trait Loci (QTLs) using single markers was implemented via Markov Chain Monte Carlo (MCMC) algorithms for parameter estimation and hypothesis testing. Parameters were estimated by marginal posterior means computed with a Gibbs sampler with data augmentation. Variables sampled included the augmented data (marker-QTL genotypes, polygenic effects), the event of linkage or nonlinkage, and the parameters (allele frequencies, QTL substitution effect, recombination rate, polygenic and residual variances). The analysis was evaluated empirically via application to simulated granddaughter designs consisting of 2000 sons, 20 related sires and their ancestors. Results obtained in this study and preliminary work on multiple linked markers and multiple QTLs support the usefulness of the Bayesian method for the statistical mapping of QTLs.

Key words Linkage analysis · Bayesian method · Markov chain Monte Carlo · Quantitative-trait loci · Simulation

Introduction

Bayesian methods have been suggested by several authors for the identification of major genes or linkages between marker loci and QTLa. Hoeschele and VanRaden (1993a,b) derived a Bayesian analysis of linkage between single genetic markers and quantitative trait loci. Hoeschele (1994) extended this method to multiple markers and described a Gibbs sampler for this analysis. Thomas and Cortessis (1992) developed a Bayesian method, implemented via Gibbs sampling, for a simple model of disease etiology, given a single marker. Janss et al. (1995) presented a Bayesian approach for the complex segregation analysis of a continuous trait.

In a companion paper, Thaller and Hoeschele (1996) implemented the method of Hoeschele and VanRaden (1993a,b) for Bayesian linkage analysis via MCMC algorithms for parameter estimation and testing for linkage. The parameter vector includes QTL gene frequency, substitution effect, marker-QTL recombination rate, polygenic and residual variances. Here, we apply the method of Thaller and Hoeschele (1996) to simulated granddaughter designs (GDD) or half-sib designs used in cattle for the mapping of QTLs, which are similar in size and structure to actual GDDs in the US Holstein population. Because a detailed derivation of the method is provided in the companion paper, here we will only describe the simulation designs and the results of the analyses.

Simulation

The simulated GDD consisted of 2000 sons (100 per sire) and 29 ancestors: two grandfathers of sires, seven fathers of sires, and 20 sires. This pedigree of the sires was chosen to mimic the relationship structure of the real GDD of the US public gene mapping project in dairy cattle, the Dairy Bull DNA Repository (Da et al. 1994). The simulated pedigree is depicted in Fig. 1.

Marker and QTL genotypes were assigned to base animals assuming Hardy-Weinberg equilibrium (HWE). Haplotypes were transmitted to offspring according to the linkage phase in the male parent, and haplotypes from the dams were generated under HWE. For the marker locus five al-

Communicated by E. J. Eisen

G. Thaller[1] · I. Hoeschele[2] (✉)
Department of Dairy Science,
Virginia Polytechnic Institute and State University,
Blacksburg, VA 24061-0315, USA

Present addresses:
[1] Lehrstuhl für Tierzucht, Technische Universität München,
D-85350 Freising-Weihenstephan, Germany
[2] Department of Dairy Science, 2160 Litton-Reaves Hall,
Virginia Tech, Blacksburg, VA 24061-0315, USA

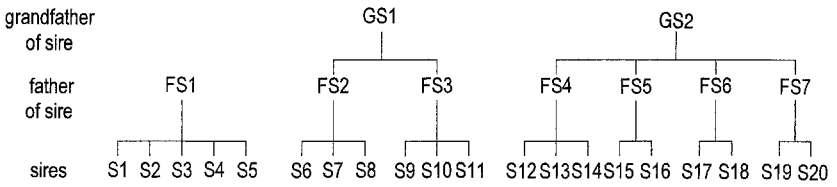

Fig. 1 Pedigree of granddaughter design

leles with equal frequencies, and for the QTL two alleles were assumed.

The trait simulated and analyzed was Daughter Yield Deviation (DYD) of the sons (Van Raden and Wiggans 1991) for fat yield. DYDs are unregressed averages of daughter records of sons, which are adjusted for systematic environmental effects and the genetic merits of the mates of the sons.

The model employed for the simulation of DYDs of sons was:

$$DYD_{son} = \frac{1}{n}\sum_{o=1}^{n} g_{o(son)} + 0.5u_{son} + \varepsilon_{son}; \quad (1)$$

$$Var(\varepsilon_{son}) = \frac{1}{n}(0.75\sigma_u^2 + \sigma_E^2)$$

where n was the number of daughters of the son, g was the QTL genotypic deviation equal to $-\alpha$, 0, or α under additive gene action, α was the QTL substitution effect, u was the polygenic effect, σ_u^2 was the polygenic variance, and σ_E^2 denotes a purely environmental component of variance. Note that the distribution of actual DYDs is more complex as these are adjusted for environmental effects and the merits of mates using animal model estimates.

The model of analysis for the DYDs of sons was:

$$DYD_{son} = \mu + 0.5g_{son} + 0.5u_{son} + e_{son}; \quad (2)$$

$$e_{son} \sim N(0, r_{son}\,\sigma_i^2),$$

where $i = 1$ if $G_{son} = QQ$ or qq and $i = 2$ if $G_{son} = Qq$ or qQ, μ was an overall mean, e was a residual, r_{son} = (1–Rel)/Rel = k/n was a weight, Rel was the reliability of the son's animal model evaluation contributed by daughters (Van-Raden and Wiggans 1991), and k was the variance ratio in a sire model without QTLs (note that the weight 1/n could be used in place of k/n). Two different residual variances were modeled because the variance among daughters of QTL-homozygous sons is $p(1-p)\alpha^2 + 0.75\sigma_u^2 + \sigma_E^2$, while the variance among daughters of heterozygous sons equals this amount plus $0.25\alpha^2$.

Several data sets were defined with common parameter values for reliability (Rel = 0.7) and heritability ($h^2 = 0.3$). The data sets differed in the QTL allele frequency (p), the marker-QTL recombination rate (r) and in the size of the QTL effect (α), expressed in additive genetic standard deviations. The data sets corresponding to the combinations of QTL parameter values investigated are listed in Table 1. Data sets III, VI, and VIII contained an unlinked QTL to check the methodology when the null hypothesis of no linkage was true.

Results

Starting values

Starting values were required for all parameters to initiate the Gibbs sampler. Starting values were not needed for the missing data (**MG**, **u**), because blocking was applied to sample genotypes and polygenic effects of sires and sons. MG and u effects were both sampled in the order sires – sons – ancestors, and in cycle 1 sires were treated as unrelated. The sampling probabilities for the QTL genotypes of selected individuals from the pedigree, evaluated at the true values of the parameters and missing data and the corresponding frequencies from a run of the Gibbs sampler, are given in Table 2. These numbers show that sufficient movement of genotypes could be expected from a Gibbs sampler of computationally acceptable length. Furthermore, for most individuals their true genotype was sampled most frequently.

The true values were used as starting values for the parameters, after trial runs resulted in close agreement (relative to Monte Carlo errors) between parameter estimates from samplers with very different sets of initial values.

Diagnostics from Gibbs output

The length of the Gibbs sampler and the burn-in period were determined based on the estimated autocorrelation structure of the samples for each parameter from trial runs of the Gibbs sampler for the designs in Table 1. Considering the least-favorable autocorrelation structure, subsequent analyses were based on a single Gibbs chain with 5000 burn-in cycles and a length of 750 000. Due to memory limitations, information from every other cycle was

Table 1 Granddaughter designs

Design	Gene frequency	Recombination rate	QTL Effect
I	0.5	0.1	1.0
II	0.5	0.2	1.0
III	0.5	0.5	1.0
IV	0.5	0.1	0.5
V	0.5	0.2	0.5
VI	0.5	0.5	0.5
VII	0.2	0.1	1.0
VIII	0.2	0.5	1.0

Table 2 Probabilities of the conditional sampling distribution of genotypes realized in Gibbs output and evaluated at true (in parentheses) QTL genotypes of relatives and true parameter values for design I and selected individuals in Fig. 1

Individual	True genotype	QTL genotype probabilities			
		1	2	3	4
GS1	3	0.010 (0.000)	0.014 (0.012)	0.970 (0.988)	0.006 (0.000)
FS3	2	0.946 (0.924)	0.047 (0.075)	0.007 (0.001)	0.000 (0.000)
S10	2	0.017 (0.000)	0.982 (10.00)	0.000 (0.000)	0.001 (0.000)
GS2	1	0.303 (0.932)	0.687 (0.068)	0.008 (0.000)	0.002 (0.000)
FS5	1	0.891 (0.917)	0.097 (0.083)	0.012 (0.000)	0.000 (0.000)
S16	2	0.101 (0.000)	0.895 (10.00)	0.000 (0.000)	0.004 (0.000)
FS1	1	0.957 (0.992)	0.005 (0.008)	0.038 (0.000)	0.000 (0.000)
S3	2	0.083 (0.000)	0.881 (10.00)	0.031 (0.000)	0.005 (0.000)

Table 3 Effective sample sizes (averages from five replicates) for allele frequency (p), recombination rate (r), QTL effect (α) and polygenic variance (σ_u^2)

Design[a]	p	r	α	σ_u^2
I	1113	1497	1645	320
II	869	2454	2629	680
III	248	2874	538	426
IV	325	1046	1422	390
V	262	1074	508	595
VI	276	2360	670	640
VII	407	2439	1421	503
VIII	206	3467	514	524

[a] Designs are defined in Table 1

Table 4 Ranges of autocorrelations across all designs and five replicates for allele frequency (p), recombination rate (r), QTLeffect (a) and polygenic variance (σ_u^2)

Lag	p	r	α	(σ_u^2)
1	0.963–0.993	0.866–0.979	0.879–0.974	0.969–0.992
10	0.763–0.948	0.414–0.757	0.514–0.829	0.930–0.963
50	0.315–0.806	0.151–0.509	0.164–0.521	0.725–0.876
100	0.161–0.668	0.087–0.390	0.079–0.382	0.538–0.745
500	0.055–0.482	0.025–0.213	0.028–0.166	0.298–0.416
1000	0.009–0.375	0.003–0.163	0.001–0.128	0.034–0.257
2000	0.0–0.246	0.004–0.010	0.002–0.079	0.046–0.346

stored. Formulae for estimating autocorrelations can be found in time series texts, in Geyer (1992), and in Sorensen et al. (1995), who also computed an effective sample size (ESS) or the number of independent samples containing as much information as a given number of dependent samples.

Table 3 contains the ESS for those parameters for which the smallest ESS numbers were found. Values are averages across five replicates. ESS numbers for the residual variances were in the order of twice the values obtained for polygenic variance. Designs with an unlinked QTL tended to yield a smaller ESS for p and α and a larger ESS for r. Ranges of autocorrelations across designs and replicates for different lags are given in Table 4. Autocorrelations for the residual variances were only slightly smaller than those for σ_u^2. Overall, the highest autocorrelations were found for the variance components and gene frequency p. Given the figures in Table 3, and the very close agreement between estimates from samplers with different initial values, the length of the sampler was considered as sufficient to provide reliable inferences.

Parameter estimates

True parameter values, average posterior mean estimates and their empirical standard errors across ten replicates are listed in Tables 5 to 7. In all analyses, a flat prior on [0, c], with c being a very large constant, was used for QTL effect α. For the QTL effect of one genetic SD and tight linkage (design I in Table 5), estimates of the QTL parameters were very accurate, although the QTL variance associated with the marker was slightly under-estimated. Polygenic variance was over-estimated and the residual variances were under-estimated, which can be explained with the strong, negative posterior correlations between these parameters in Table 8. For the lesser degree of linkage with true r = 0.2, QTL parameters p and r were still accurately estimated while QTL substitution effect α and QTL variance associated with the marker were under-estimated, with polygenic variance being over-estimated.

For the designs with a equal to half of the genetic SD (Table 6), estimates of r were biased upwards. As a consequence, the QTL variance associated with the marker was under-estimated while QTL substitution effect α and the residual variances were over-estimated. Gene frequency and polygenic variance were estimated quite accurately for both designs in Table 6.

Table 7 contains the parameter estimates for design III where a QTL of one genetic SD was present but not linked to the marker. Parameters p and r and the QTL variance associated with the marker were estimated accurately, while the substitution effect of the unlinked QTL was under-estimated, and the polygenic and residual variances were over-estimated. For design VII with a tightly linked QTL effect of one genetic SD and a frequency of the less-favorable allele of 0.8, the QTL parameters were estimated accurately, while polygenic variance was slightly over- and the residual variances were slightly under-estimated.

Ranges of the posterior correlations among parameters, estimated from Gibbs output, are provided in Table 8 for

Table 5 Average parameter estimates and empirical SE across ten replicates for designs I and II[a]

Parameter	True value Design I (II)	Design I Estimate (SE)	Design II Estimate (SE)
p	0.50	0.51 (0.02)	0.51 (0.03)
r	0.10 (0.20)	0.11 (0.01)	0.19 (0.03)
α	570.50	56.70 (1.35)	490.50 (20.03)
μ	0.00	−10.20 (10.21)	−0.88 (1.66)
σ^2_{e1}	793.36	7140.56 (750.23)	8020.57 (134.70)
σ^2_{e2}	860.41	767.90 (61.80)	830.80 (120.19)
σ^2_{u}	413.44	487.94 (360.28)	535.40 (49.75)
σ^2_{m}	1058.40 (595.35)	964.37 (60.74)	4590.29 (63.64)

[a] Designs are defined in Table 1

Table 6 Average parameter estimates and empirical SE across ten replicates for designs IV and V[a]

Parameter	True value Design I (II)	Design I Estimate (SE)	Design II Estimate (SE)
p	0.50	0.54 (0.03)	0.51 (0.04)
r	0.10 (0.20)	0.27 (0.03)	0.38 (0.03)
α	28.76	36.44 (1.91)	34.88 (1.78)
μ	0.00	0.18 (1.61)	−1.36 (1.75)
σ^2_{e1}	8180.51	8890.00 (1040.55)	847.92 (107.34)
σ^2_{e2}	8350.25	901.48 (1050.2)	985.39 (113.67)
σ^2_{u}	7230.52	702.18 (54.17)	737.63 (58.31)
σ^2_{m}	264.60 (148.84)	146.80 (31.77)	34.82 (7.62)

[a] Designs are defined in Table 1

Table 7 Average parameter estimates and empirical SE across ten replicates for designs III and VII[a]

Parameter	True value Design I (II)	Design I Estimate (SE)	Design II Estimate (SE)
p	0.50 (0.80)	0.51 (0.03)	0.78 (0.02)
r	0.50 (0.10)	0.49 (0.00)	0.10 (0.01)
α	570.51	380.02 (2.82)	590.03 (20.28)
μ	0.00 (−34.51)	−1.12 (1.83)	−51.88 (10.20)
σ^2_{e1}	793.36 (805.41)	941.65 (1440.55)	742.88 (83.71)
σ^2_{e2}	860.41 (872.47)	943.40 (1160.2)	795.89 (108.10)
σ^2_{u}	413.44 (5620.28)	623.66 (48.62)	6140.52 (46.35)
σ^2_{m}	0.00 (677.38)	20.25 (1.30)	713.80 (690.06)

[a] Designs are defined in Table 1

Table 8 Ranges (across ten replicates) of the posterior correlations among parameters evaluated from Gibbs output. Values above the diagonal are for design I, below for design V[a]

Item	p	r	α	σ^2_{e1}	σ^2_{e2}	σ^2_{u}
p		−0.22, 0.07	−0.28, 0.14	0.59, 0.73	−0.19, 0.21	−0.06, 0.26
r	−0.35, 0.17		0.48, 0.67	−0.16, 0.04	−0.45, −0.25	−0.16, 0.09
α	−0.30, 0.44	−0.11, 0.44		−0.31, 0.07	−0.56, −0.28	−0.29, 0.11
σ^2_{e1}	0.74, 0.83	−0.14, 0.36	−0.52, 0.66		0.77, 0.87	−0.91, −0.83
σ^2_{e2}	−0.21, 0.40	−0.13, 0.30	−0.23, 0.15	0.20, 0.84		−0.83, −0.60
σ^2_{u}	−0.25, 0.16	−0.14, 0.04	−0.28, −0.15	−0.93, −0.71	−0.85, −0.22	

[a] Designs are defined in Table 1

Table 9 Marginal posterior probabilities of linkage computed as Monte Carlo averages of conditional probabilities. Five replicates

Design[a]	Marginal posterior probability of linkage			Likelihood ratio	$n > LR_{MAX \mid Ho}$
	Average	MIN	MAX		
I	10.00	10.00	10.00	∞	5
II	0.886	0.535	10.00	1.941	5
III	0.052	0.019	0.095	0.014	–
IV	0.768	0.156	0.998	0.828	5
V	0.313	0.066	0.813	0.113	2
VI	0.079	0.047	0.140	0.020	–
VII	1.00	1.00	1.00	∞	5
VIII	0.046	0.014	0.093	0.193	–

[a] Designs are defined in Table 1

Table 10 Marginal posterior probabilities of linkage and marginal likelihood ratios (R) due to Newton and Raftery (1994) with d=0.1 and based on five replicates

Design	$\ln[f(\mathbf{y}, \mathbf{M} \mid r < 0.5)/f(\mathbf{y}, \mathbf{M} \mid r = 0.5)]$			$P(r < 0.5 \mid \mathbf{y}, \mathbf{M})$	R	$n > R_{MAX \mid Ho}$
	Average	MIN	MAX			
I	172.93	21.76	361.04	1.0	10^{75}	4
II	24.37	–26.59	159.56	1.0	10^{10}	1
III	–1.46	–15.89	30.57	0.058	0.232	–
IV	25.15	–32.72	90.18	1.0	10^{10}	4
V	–28.97	–135.25	19.72	0.0	10^{-13}	0
VI	12.68	–5.91	54.24	1.0	10^{5}	–
VII	39.04	–6.13	120.68	1.0	10^{17}	4
VIII	–8.65	–16.80	–4.81	0.0	10^{-4}	–

[a] Designs are defined in Table 1

designs I and V. Values for the other designs were similar. Most correlations were relatively weak. Strong correlations were found between the residual and polygenic variance components and between residual variance for the homozygous sons and QTL allele frequency. Correlations of intermediate strength were found between r and α and between residual variances and α.

Hypothesis testing

Table 9 contains averages, minima and maxima of the marginal posterior probabilities of linkage over five replicates, computed as $1-P(\mathcal{L} = 0 \mid \mathbf{y}, \mathbf{M})$ with $P(\mathcal{L} = 0 \mid \mathbf{y}, \mathbf{M})$ estimated as in equation (10) of Thaller and Hoeschele (1996). Corresponding marginal likelihood ratios (MLRs) were computed by rearranging equation (2) in Thaller and Hoeschele (1996) and replacing $P(\mathcal{L} = 0 \mid \mathbf{y}, \mathbf{M})$ by the average probability from the Gibbs sampler. These results were based on Gibbs runs where the prior probability of linkage was set to 0.2. Also listed is the number of replicates out of 5, in which the MLR estimated under the alternative hypothesis exceeded the maximum value obtained under the null hypothesis of nonlinkage. For designs I, II, and IV where the linkage hypothesis was true, posterior probabilities of linkage exceeded 0.5 and provided strong evidence for linkage, given a prior probability of 0.2. For design V with loose linkage (r = 0.2) and a smaller QTL effect (see Table 1), the average probability did not support linkage. The MLRs in favor of linkage exceeded 1 for designs I and II, but did not support linkage for all other designs. For the designs with an unlinked QTL (III, IV, VIII), the MLRs took on extremely small values leading to very small posterior probabilities of linkage and a clear rejection of the linkage hypothesis. MLRs less than one for the other designs exceeded those for III, IV and VIII by far.

Average, minima and maxima of the natural logarithms of the MLRs, calculated with the MC algorithm of Newton and Raftery (1994) as described in Thaller and Hoeschele (1996), are in Table 10. The imaginary portion of samples drawn from the prior distribution was set to 0.1 after an initial robustness study showed little effect in the range of 0.01 to 0.2. Based on average (across replicates) likelihood ratios, conclusions similar to those for Table 9 were reached, with the exception of design VI, where the null hypothesis was incorrectly rejected. Compared to the results in Table 9, the MLRs were very variable across replicates, and there was overlap in MLR values between designs with and without linkage. Comparing the MLRs for designs with a linked QTL to the maximum MLRs obtained for designs representing the nonlinkage hypothesis (last column of Table 10) revealed a loss of power in discriminating between the null and the alternative hypotheses. Even for design I, there was one outlier replicate indicating nonlinkage.

Table 11 Marginal posterior probabilities of linkage and marginal likelihood ratios (R) due to Meng and Wong (1993) with ln R calculated from mean $f_i(y,q_c)/f_j(y,q_c)$ and based on five replicates

Design	ln [f (**y**, **M** | r < 0.5)/f (**y**, **M** | r = 0.5)]			P (r < 0.5 | **y**, **M**)	R	$n > R_{MAX \mid Ho}$
	Average	MIN	MAX			
I	218.87	1650.54	2670.29	1.0	10^{96}	5
II	211.30	146.93	266.18	1.0	10^{92}	5
III	70.68	30.98	1110.07	1.0	10^{31}	–
IV	256.81	2180.04	292.36	1.0	10^{112}	5
V	2280.59	191.85	2660.02	1.0	10^{100}	5
VI	148.95	118.80	185.78	1.0	10^{65}	–
VII	183.16	1090.04	238.87	1.0	10^{80}	5
VIII	620.53	1.64	97.46	1.0	10^{27}	–

[a] Designs are defined in Table 1

Table 12 Marginal posterior probabilities of linkage and marginal likelihood ratios (R) due to Meng and Wong (1993) with ln R calculated from median $f_i(y,q_c)/f_j(y,q_c)$ and based on five replicates

Design	ln [f (**y**, **M** | r < 0.5)/f (**y**, **M** | r = 0.5)]			P (r < 0.5 | **y**, **M**)	R	$n > R_{MAX \mid Ho}$
	Average	MIN	MAX			
I	1040.07	530.55	165.47	1.0	10^{46}	5
II	420.08	3.36	118.75	1.0	10^{18}	5
III	–1.86	–2.19	–10.55	0.038	0.156	–
IV	45.70	–1.73	890.00	1.0	10^{20}	4
V	8.72	–1.75	500.20	1.0	10^{4}	1
VI	–1.72	–1.90	–1.60	0.043	0.179	–
VII	57.84	430.03	790.53	1.0	10^{25}	5
VIII	–10.56	–1.76	–1.41	0.050	0.209	–

[a] Designs are defined in Table 1

MLRs calculated via the method of Meng and Wong (1993) were first computed as in equation (19) of Thaller and Hoeschele (1996). Subsequently, instead of taking means in the numerator and denominator of (19), we calculated the medians of the square roots of the likelihood ratios in (19). The original method based on means (Table 11) led to very high MLRs both for designs with and without a linked QTL. These values appeared to be far too high and the comparison to the maximum achieved under H_0 (last column of Table 11) was not consistent with the behavior of the other methods (Tables 9 and 10). This finding appeared to be the result of outliers in a few cycles of the Gibbs sampler, affecting the mean strongly but not the median. The likelihoods in (19) depend on the frequencies of the MG genotypes realized in any Gibbs cycle [see equation (20) of Thaller and Hoeschele (1996)]. When analyzing data sets generated under H_0 (nonlinkage) by running a Gibbs sampler under the linkage assumption, most recombination rates were expectedly set to 0.5. When r values less than 0.5 were sampled, the Gibbs sampler quickly returned to 0.5 except for a few occasions where it remained at smaller values for several, say ten, cycles. Then, changes in the frequencies of the MG realizations led to very few unusually low ratios in the denominator of (19) which dominated the calculation of the MLR. Inspection of the distribution of the likelihood ratios across cycles depicted these few outliers. Therefore, means in the numerator and denominator of (19) were replaced by medians. MLRs based on medians are presented in Table 12 and are in much better agreement with our expectations and the results in Table 9, at least to the extent that the nonlinkage cases were now clearly identified.

To investigate the dependency of the outcome of the analysis on the assumed prior probability of linkage, designs I, III, and IV were analyzed twice (five replications each) by first assuming a prior probability of 0.2 and subsequently of 0.05. For design I, there was no difference between the two analyses, with the event of linkage being sampled in each of the 750 000 Gibbs cycles. For design IV, the outcome of the analysis depended on the assumed prior probability such that, on average, in 576 189 Gibbs cycles linkage was sampled given a prior probability of 0.2, while linkage was sampled in only 478 810 cycles, on average, for the prior of 0.05. For the prior of 0.2 the posterior probability of linkage was 0.77, and for the prior of 0.05 the posterior probability was 0.64. For design III (nonlinkage), linkage was sampled in 38 70 cycles, on average, for the prior of 0.2 and in 21 243 cycles, on average, for the prior of 0.05. The corresponding posterior probabilities of linkage were 0.052 and 0.028, respectively.

Marginal posterior densities of the QTL variance associated with the marker were estimated from Gibbs output as described in Thaller and Hoeschele (1996) under 'Hypothesis Testing', and representative density plots are depicted in Figs. 2 to 4. Under nonlinkage (Fig. 4, design III), the maximum value of the density was attained at zero, and

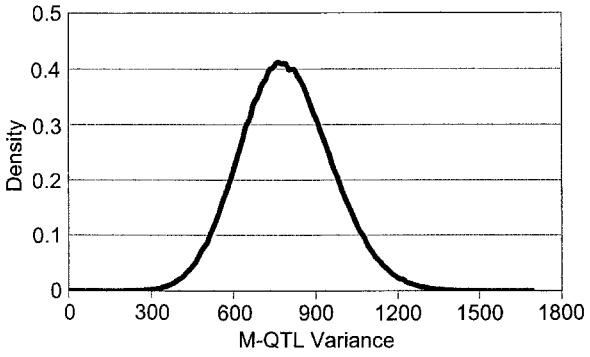

Fig. 2 Marginal posterior density of the QTL variance associated with the marker for design I

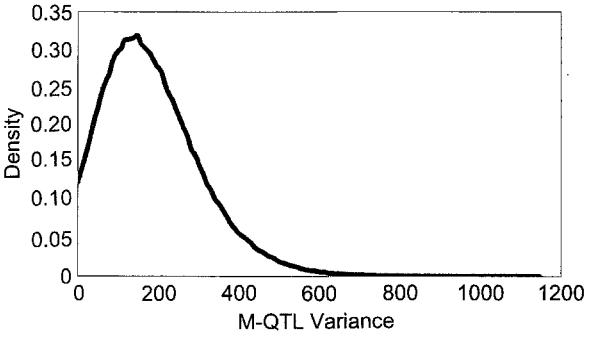

Fig. 3 Marginal posterior density of the QTL variance associated with the marker for design V

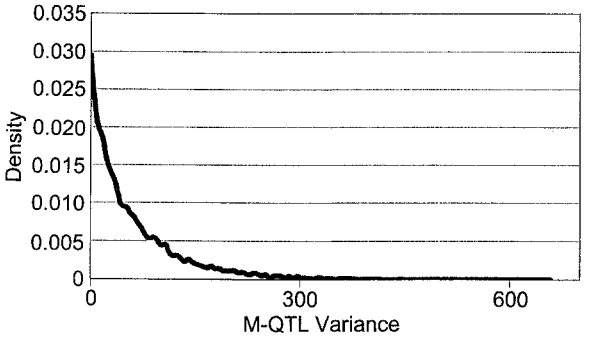

Fig. 4 Marginal posterior density of the QTL variance associated with the marker for design III

the density continuously decreased with increasing variance. For design I, the true variance was large (1058.40), and Fig. 2 shows very strong support for the presence of a QTL with the density approaching zero quickly for small values of the variance. For design V with a loosely linked QTL of smaller effect (true variance of 148.84), there was insufficient evidence for a linked QTL in Fig. 3, with the ratio of the maximum value of the marginal likelihood to its value at zero equal to only 2.36.

Conclusions

This study represents the first application of Bayesian QTL mapping to a very large granddaughter design, similar in size to real GDDs in the US Holstein population, implemented via MCMC algorithms and utilizing full pedigree information for the sires. Although the simulated GDDs only contained relationships among sires through male ancestors, those through females can be utilized if available.

The methodology can equally well be applied to daughter designs, where the phenotypes are records of individual daughters. An analysis of DYDs is approximate due to the assumption of normally distributed residuals in model (2). However, analysis of daughter phenotypes would require the inclusion of many individuals which have not been genotyped, but are contemporaries of the genotyped daughters, in order to adjust for systematic environmental effects present in field data. Performing such an analysis would most likely be difficult with any method of linkage analysis.

The autocorrelation structure found for the parameters of interest required a Gibbs sampler of length 750000 to be run. With substantial analysis of Gibbs output and information stored from every other cycle, one run of this sampler took about 15 CPU h on an IBM SP2 system, which could be somewhat reduced by optimization of the software for this system. This analysis is, however, computationally feasible, at least for the markers of special interest identified from an initial ad hoc linkage analysis.

The marginal posterior mean estimates of the parameters were fairly accurate, although some improvement might be expected from the utilization of multiple linked markers, which we are currently investigating. The deterioration in the accuracy of the parameter estimates and the increasing influence of the assumed prior probability of linkage on the outcome of the analysis when the QTL effect was reduced from 1 to 0.5 genetic SD, as well as results for 0.25 genetic SD not presented here (posterior probability of linkage always less than 0.5), indicate that QTL effects of 0.5 genetic SD represent the lower limit for successful statistical QTL mapping using single markers, at least for the designs and methodology investigated here. For a general interpretation of Bayesian parameter estimates, see the companion paper (Thaller and Hoeschele 1996).

Several tests for linkage, evaluated from Gibbs output, were investigated and, although not in close agreement in the numerical values, led to very similar decisions with respect to acceptance or rejection of linkage. The most promising test appears to be the one based on marginal posterior probabilities of linkage calculated from equations (10) or (11) in Thaller and Hoeschele (1996). This method generalizes straightforwardly to multiple linked markers.

The results presented here, our initial work on multiple linked markers (Uimari et al. 1996), and the study of Satagopan et al. (1996) fitting multiple QTLs in an MCMC algorithm, successfully indicate that this methodology will be very useful for obtaining parameter estimates and for

providing further evidence for QTLs located in promising regions of the genome identified by simpler methods which cannot provide parameter estimates nor utilize full pedigree information.

Acknowledgements The National Science Foundation provided generous support for this project (grant no. BIR-9596247). G. Thaller acknowledges financial support from the Deutsche Forschungsgemeinschaft in the form of a postdoctoral fellowship. I. Hoeschele acknowledges financial support from the European Human Capital and Mobility Fund while on research leave at Wageningen University, The Netherlands. This research was conducted using the resources of the Cornell Theory Center, which receives major funding from the National Science Foundation and New York State. Additional funding comes from the Advanced Research Projects Agency, the National Institutes of Health, IBM Corporation, and other members of the center's Corporate Research Institute.

References

Da Y, Ron M, Yanai A, Band M, Everts RE, Heyen DW, Weller JI, Wiggans GR, Lewin HA (1994) The dairy bull DNA repository: a resource for mapping quantitative trait loci. Proc 5th World Congr Genet Appl Livest Prod Sci 21:229–232

Geyer CJ (1992) Practical Markov chain Monte Carlo (with discussion). Stat Sci 7:467–511

Hoeschele I (1994) Bayesian QTL mapping via the Gibbs sampler. Proc 5th World Congr Genet Appl Livst Prod 21, Guelph, Canada, pp 241–244

Hoeschele I, VanRaden PM (1993a) Bayesian analysis of linkage between genetic markers and quantitative trait loci. I. Prior knowledge. Theor Appl Genet 85:953–960

Hoeschele I, VanRaden PM (1993b) Bayesian analysis of linkage between genetic markers and quantitative trait loci. II. Combining prior knowledge with experimental evidence. Theor Appl Genet 85:946–952

Janss LLG, Thompson R, van Arendonk JAM (1995) Application of Gibbs sampling in a mixed major gene – polygenic inheritance model in animal populations. Theor Appl Genet 91:1137–1147

Meng X-L, Wong WH (1993) Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. Technical Report No. 365, Department of Statistics, The University of Chicago

Newton MA, Raftery AE (1994) Approximate Bayesian inference with the weighted likelihood bootstrap. J Roy Stat Soc B 56:3–48

Satagopan JM, Yandell BS, Newton MA, Osborn TC (1996) Markov chain Monte Carlo approach to detect polygene loci for complex traits. Genetics (in press)

Scott WD (1992) Multivariate density estimation. Wiley and Sons, New York

Sorensen DA, Anderson S, Gianola D, Korsgaard I (1995) Bayesian inference in threshold models using Gibbs sampling. Genet Selec Evol 27:229–249

Thaller G, Hoeschele I (1996) A Monte Carlo method for Bayesian analysis of linkage between single markers and quantitative trait loci. I. Methodology. Theor Appl Genet (in press)

Thomas DC, Cortessis V (1992) A Gibbs sampling approach to linkage analysis. Hum Hered 42:63–76

Uimari P, Thaller G, Hoeschele I (1996) A Monte Carlo method for Bayesian analysis of linkage between multiple linked markers and a quantitative trait locus. Genetics 143:1831–1842

VanRaden PM, Wiggans GR (1991) Derivation, calculation and use of national animal model information. J Dairy Sci 74:2737–2746