

Performance of Markov Chain–Monte Carlo Approaches for Mapping Genes in Oligogenic Models with an Unknown Number of Loci

Jae K. Lee¹ and Duncan C. Thomas²

¹Department of Health Evaluation Sciences, University of Virginia School of Medicine, Charlottesville, and ²Department of Preventive Medicine, University of Southern California, Los Angeles

Markov chain–Monte Carlo (MCMC) techniques for multipoint mapping of quantitative trait loci have been developed on nuclear-family and extended-pedigree data. These methods are based on repeated sampling—peeling and gene dropping of genotype vectors and random sampling of each of the model parameters from their full conditional distributions, given phenotypes, markers, and other model parameters. We further refine such approaches by improving the efficiency of the marker haplotype-updating algorithm and by adopting a new proposal for adding loci. Incorporating these refinements, we have performed an extensive simulation study on simulated nuclear-family data, varying the number of trait loci, family size, displacement, and other segregation parameters. Our simulation studies show that our MCMC algorithm identifies the locations of the true trait loci and estimates their segregation parameters well—provided that the total number of sibship pairs in the pedigree data is reasonably large, heritability of each individual trait locus is not too low, and the loci are not too close together. Our MCMC algorithm was shown to be significantly more efficient than LOKI (Heath 1997) in our simulation study using nuclear-family data.

Introduction

Quantitative trait-locus (QTL) mapping methods have been well established in experimental genetics (see, e.g., Lander and Botstein 1989; Kruglyak and Lander 1995). Recently, there has been a surge of interest in Bayesian approaches to mapping QTLs. Although the origins of these methods are in the plant- and animal-breeding fields (Jansen and Stam 1994; Jansen 1996) and are based on experimental crosses between highly inbred lines, similar approaches have recently been explored in human genetics. In experimental genetics, the samples in the offspring generation are treated as independent observations using a relatively simple regression of the phenotype on a putative genotype, whose probability distribution is related to the observed marker data. Stephens and Smith (1993), Satagopan et al. (1996), and Uimari et al. (1996) describe Markov chain–Monte Carlo (MCMC) implementations of a fully Bayesian treatment of the problem of estimating both the location and segregation parameters for multiple QTLs, where the number of QTLs is fixed. However, the number of QTLs is in fact unknown, and hence the dimensionality

of the parameter space is also unknown, a problem that has not been solvable in a Bayesian framework until the introduction of the reversible-jump MCMC method by Green (1995). Sillanpää and Arjas (1998, 1999) and Stevens and Fisch (1998) applied this approach to the case of an unknown number of QTLs in line crosses, and George et al. (2000) applied similar methods to the problem of ordering marker loci. Although these methods, and the related methods in human genetics, were originally developed for mapping genes involved in quantitative traits, they are easily extended to binary, censored age at onset, or multivariate traits; nevertheless, we retain the term “QTL” in this discussion because of its historical context, without wishing to imply any such restriction.

The application of these ideas to human genetics is relatively new. Additional complications in human genetics derive from the absence of simple experimental-cross designs, leading to the need to consider all possible haplotypes corresponding to the observed genotypes, missing marker data on some individuals, more-complex pedigree structures, and uncertainty about the form of the disease model. At the 10th Genetic Analysis Workshop (GAW), Heath et al. (1997) and Thomas et al. (1997) independently introduced similar methods based on reversible-jump MCMC, where the genotypes at the trait loci were sampled using peeling, conditional on the current assignments of locations and segregation parameters. Given these sampled genotypes, the updating of these parameters and the number of trait loci then followed methods similar to those used in breed-

Received May 15, 2000; accepted for publication August 21, 2000; electronically published October 13, 2000.

Address for correspondence and reprints: Dr. Jae K. Lee, Department of Health Evaluation Sciences, University of Virginia School of Medicine, P.O. Box 800717, Charlottesville, VA 22908. E-mail: jaeklee@virginia.edu

© 2000 by The American Society of Human Genetics. All rights reserved. 0002-9297/2000/6705-0021\$02.00

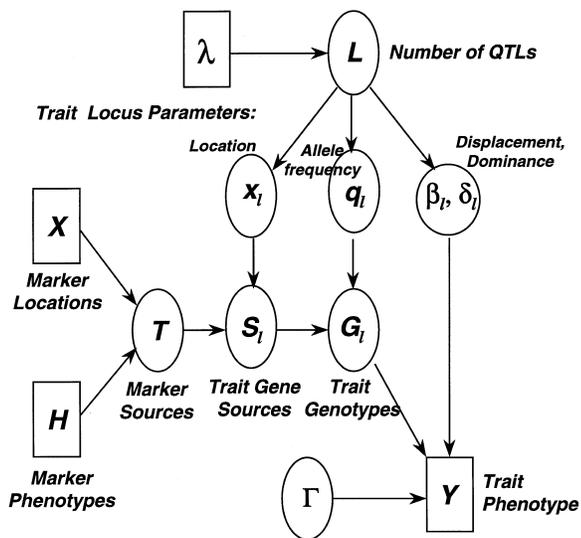


Figure 1 Directed acyclic graph for the QTL mapping model. Squares indicate observed data ($X, H,$ and Y) or fixed hyperparameters λ . Circles indicate unknown parameters ($G_l, S_l,$ and T) or latent variables ($x_l, q_l, \beta_l, \delta_l,$ and Γ). Not shown are the hyperpriors ϕ for q_l and ω for β_l .

ing and other human genetics applications. Heath’s method, which is applicable to extended pedigrees, has been described fully in a subsequent paper (Heath 1997) and implemented in a program called LOKI. Several applications of the method to real pedigree data on alcoholism (Daw et al. 1999b) and to simulated nuclear-family data (Hinrichs et al. 1999; Uimari et al. 1999) were presented at the 11th GAW. Daw et al. (1999a) applied LOKI to pedigree data on the age at onset of Alzheimer disease, whereas Yuan et al. (2000) applied it to data on familial hypobetalipoproteinemia. To date, however, there has been no systematic simulation study of the performance of these methods. Here, we describe some further refinements of our algorithm and investigate its performance on simulated nuclear-family data. In particular, our approach differs from that of Heath by a new algorithm for adding trait loci that leads to better acceptance rates.

Methods

Notation

Let $i = 1, \dots, I$ index the sibships and $j = 1, \dots, J_i$ its members, and let Y_{ij} denote their observed phenotypes. Let $p = m, f$ index the parents of sibship i , and let Y_{ip} denote their phenotypes. (As in the GAW applications, the phenotype is assumed here to be continuous with normally distributed errors, but the methods are easily extended to other types of data, including multivariate

phenotypes.) We postulate that there are an unknown number L of diallelic trait loci that influence this phenotype. Let $G_{ij\ell}$ denote the unobserved trait genotype at locus ℓ for subject ij , and let $S_{ij\ell p}$ be the grandparental source for the allele inherited from parent p . Let q_ℓ denote the unknown population frequency of allele A at locus ℓ . Finally, let Z_{ij} denote a vector of observed covariates.

Now suppose we have $m = 1, \dots, M_c$ marker loci on chromosomes $c = 1, \dots, C$. Let $H_{ijcm} = (H_{ijcm1}, H_{ijcm2})$ denote the marker phenotypes at locus cm , and let T_{ijcmp} be the grandparental source of the marker allele inherited from parent p . Let C_ℓ denote the chromosome on which trait locus ℓ is located and P_ℓ the position of locus ℓ on that chromosome relative to the markers ($0 =$ left of marker 1; $m =$ right of marker m). Let X_{cm} denote the map location of marker cm on the continuum along

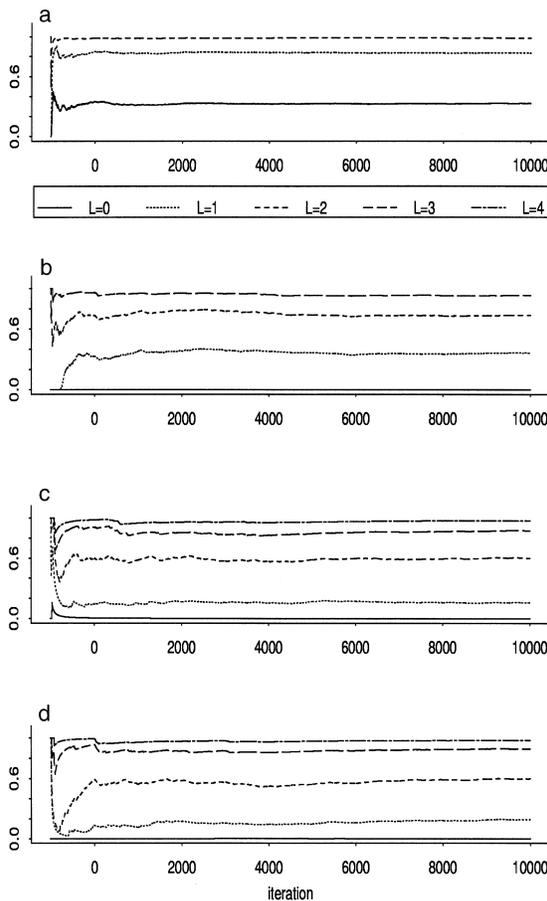


Figure 2 Convergence plots for the distribution of L for simulations with sibship size $I = 50 \times J = 8$ for (a) true $L = 0$, (b) true $L = 1$ and heritability $b^2 = 0.75$, (c) true $L = 2$ and $b^2 = 0.75$, and (d) true $L = 3$ and $b^2 = 0.75$. The lines represent the cumulative probabilities for the indicated values of the fitted L from the start of the iterations (iteration $-1,000$).

Table 1

BFs for the Number of Trait Loci when True L Varies from 0 to 5

MODEL	$I \times J$	HERITABILITY (h^2)	BFs FOR FITTED L					
			0	1	2	3	4	5+
0	50 × 8	0	6.60	3.35	.67	.08	.00	.00
1a	50 × 8	.10	5.92	3.44	.74	.10	.01	.00
1b	50 × 8	.25	.12	3.10	1.79	.50	.09	.00
1c	50 × 8	.50	.01	2.11	1.89	.93	.29	.06
1d	50 × 8	.75	.00	2.43	1.66	.87	.36	.09
1e	80 × 5	.75	.79	3.17	1.66	.45	.09	.00
1f	200 × 2	.75	6.20	3.36	.73	.11	.01	.00
1g	14 × 8	.75	.01	2.49	1.75	.76	.31	.09
1h	40 × 5	.75	.00	.53	1.73	1.45	.90	.42
1i	400 × 2	.75	.00	1.68	1.87	1.03	.45	.17
2a	50 × 8	.75	.00	.03	.95	1.53	1.68	1.29
2b	50 × 8	.75	.00	.00	.71	1.96	1.61	1.02
2c	50 × 8	.75	.00	.01	1.65	1.91	.92	.36
2d	50 × 8	.75	.00	.01	.29	1.24	2.09	1.91
2e	50 × 8	.75	.00	.00	.14	.83	2.18	2.65
3a	50 × 8	.75	.00	.00	.06	.65	2.27	2.93
3b	50 × 8	.75	.00	.01	.03	1.36	2.21	1.75
3c	50 × 8	.75	.00	.01	.88	1.80	1.72	.85
4	50 × 8	.75	.00	.00	.39	1.41	2.42	1.38
5	50 × 8	.75	.00	.00	.15	1.19	2.29	1.88
I1	50 × 8	.75	.00	.23	1.64	1.75	.98	.33
I2	50 × 8	.75	.00	.05	1.03	1.77	1.51	.87

the chromosome, and let x_ℓ denote the corresponding map location of trait locus ℓ . Thus, if $C_\ell = c$ and $0 < P_\ell = m < M_c$, then $X_{cm} < x_\ell < X_{cn}$, where $n = m + 1$. (This distinction between positions and locations allows us to replace the difficult task of sampling from the highly multimodal distribution of locations by two successive steps—a multinomial distribution of positions, followed by a unimodal continuous distribution of locations within a position.) Finally, let \tilde{X}_c denote the total length (in map units) of chromosome c , $\tilde{X} = \sum \tilde{X}_c$ be the total length of the genome, and let $\tilde{P} = \sum M_c + C$ be the total number of possible trait locus positions.

Models

The directed acyclic graph for the proposed model is shown in figure 1. The number of trait loci is assumed to have some prior distribution with hyperparameter λ . In our implementation, we have assumed a Poisson distribution, but the Bayes factors (BFs) for L are relatively insensitive to the choice of prior distribution or its hyperparameter—see, for example, the work of Thomas et al. (1997). Given L , there are two sets of parameters $\Theta = (\mathbf{x}, \mathbf{q})$ and $\Omega = (\beta, \delta)$, describing, respectively, the distributions of the genotypes and the displacements, the latter expressed in terms of genetic displacements β_ℓ and dominances δ_ℓ . Each component of these parameter vectors is independently distributed, with certain priors, which are described below. The trait genotypes \mathbf{G}_ℓ at

each locus are independently determined by the location x_ℓ relative to flanking markers, the allele frequency q_ℓ , and the observed marker information H_{cm}, H_{cn} . Finally, the trait phenotype is determined by the genotypes \mathbf{G} , parameters Θ , and Γ , and covariates \mathbf{Z} , where Γ denotes a vector of global parameters not related to the trait genotypes, such as overall means μ , variances σ^2 , and covariate coefficients γ .

We assume the trait loci are, a priori, independently uniformly distributed across the entire genome, $x_\ell \sim U(0, \tilde{X})$. For computational ease, we require that there be at most one trait locus between any pair of marker loci (or to the left of the first and right of the last markers on any chromosome), but this is not a fundamental requirement. The allele frequencies q_ℓ are assumed to have exchangeable Beta distributions with hyperpriors (ϕ_1, ϕ_2) . We used the Haldane map function to relate map distances to recombination fractions, but any map function could be used.

Within a nuclear family, we assume that the parental genotypes are in Hardy-Weinberg and linkage equilibrium across all trait and marker loci, with no assortive mating. Thus, letting $G_{i\ell}^p = (G_{i\ell 1}^p, G_{i\ell 2}^p)$ denote the parental alleles at trait locus ℓ for parent p and grandparental source s with $p, s = m, f$, and letting \mathbf{G}^p denote the vector of all parental genotypes, then

$$\Pr(\mathbf{G}^p | \mathbf{q}) = \prod_{\ell} \prod_i \prod_p \prod_s \Pr(G_{i\ell s}^p | \mathbf{q}_\ell), \quad (1)$$

where the latter probabilities are simply q_ℓ or $(1 - q_\ell)$. Then, conditional on the parents' genotypes and their

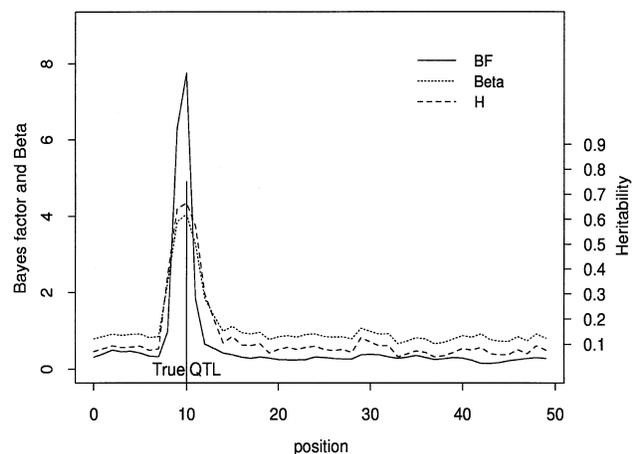


Figure 3 BFs, heritability ($H = h^2$), and displacement (Beta = β) estimates across 50 marker positions, when there is only a single trait locus with heritability 0.75 and pedigree size 50×8 . The true simulated trait locus, its displacement β , and heritability ($\times 10$) are shown as a vertical line.

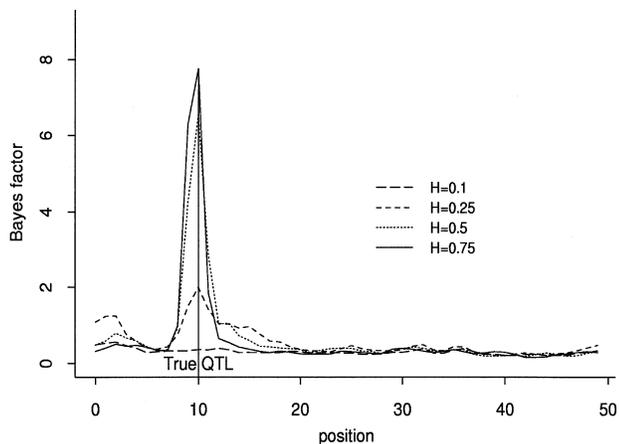


Figure 4 BFs across 50 marker positions for different degrees of heritability—0.1, 0.25, 0.5, and 0.75 ($\times 10$), when there is only a single trait locus (fixed pedigree size, 50×8).

marker sources, the offsprings’ genotypes are independent across trait loci, sibs, and parental sources, with probabilities for sib j at locus ℓ from parent p being proportional to

$$\Pr(S_{ij\ell p} | T_{ijcnp}, T_{ijcp}; x_\ell, X_{cm}, X_{cn}) = \theta_1^{r_1} (1 - \theta_1)^{1-r_1} \theta_2^{r_2} (1 - \theta_2)^{1-r_2}, \quad (2)$$

where $\theta_1 = \theta(x_\ell - X_{cm})$ and $\theta_2 = \theta(X_{cn} - x_\ell)$ indicate the recombination fractions corresponding to the indicated map distances, $r_1 = 1$ if there is a recombination with marker cm (i.e., $S_{ij\ell p} \neq T_{ijcnp}$) and 0 otherwise, and r_2 is defined similarly for marker cn . For some meioses, it may not be possible to uniquely determine the grandparental sources of marker alleles. Conditional on the marker haplotypes, the transition probabilities are obtained by summing over all possible sources,

$$\Pr(S_{ij\ell p} | H_{ijcm}, H_{ijcn}; x_\ell, X_{cm}, X_{cn}) = \sum \Pr(S_{ij\ell p} | T_{ijcnp}, T_{ijcp}; x_\ell, X_{cm}, X_{cn}) \times \Pr(T_{ijcm}, T_{ijcn} | H_{ijcm}, H_{ijcn}; X_{cm} - X_{cn}), \quad (3)$$

where the sum is taken over the four possible parental haplotypings at the two adjacent marker loci.

For illustration, we consider the phenotype to be a vector of continuous multivariate normally distributed traits Y , determined by a linear penetrance model of the form

$$Y \sim \text{MVN}(\alpha + \gamma'Z + \sum_{\ell=1}^L \beta_\ell f(G_\ell), \Sigma), \quad (4)$$

where

$$f(G) = \begin{cases} -\Delta_\ell & \text{if } G = 0 \\ \delta_\ell - \Delta_\ell & \text{if } G = 1 \\ 1 - \Delta_\ell & \text{if } G = 2 \end{cases}$$

and

$$\Delta_\ell = 2q_\ell(1 - q_\ell)\delta_\ell + q_\ell^2. \quad (5)$$

The term Δ_ℓ is chosen so that $f(G)$ will have expectation 0 over the population distribution of genotypes. Σ denotes a matrix of residual variances and covariances and γ a matrix of regression coefficients for covariates Z . We assume flat priors on α , γ , and δ and independent negative exponential priors on $\beta_\ell > 0$ with expectation ω . However, we emphasize that the phenotype could take any form with the appropriate specification of the penetrance model; see, for example, Daw et al. (1999a) for an application of such methods to censored age-at-onset phenotypes.

Thus, the joint probability of the phenotypes and the model parameters, conditional on the observed marker haplotypes and covariates, is

$$\begin{aligned} & \Pr(L, \Theta, \Omega, \Gamma, G, Y | \lambda, \phi, \omega, H, Z) \\ &= \Pr(L | \lambda) \Pr(\Theta | \phi, L) \Pr(\Omega | \omega, L) \\ & \quad \times \Pr(\Gamma) \Pr(G | H; \Theta, L) \Pr(Y | G, Z; \Omega, \Gamma) \\ &= \Pr(L | \lambda) \Pr(\Gamma) \Pr(Y | G, Z; \Omega, \Gamma) \\ & \quad \times \prod_{\ell=1}^L \Pr(q_\ell | \phi) \Pr(x_\ell | P_\ell) \Pr(P_\ell) \Pr(\beta_\ell | \omega) \\ & \quad \times \Pr(\delta_\ell) \Pr(G_\ell | H_{cm}, H_{cn}; q_\ell, x_\ell, X_{cm}, X_{cn}). \end{aligned} \quad (6)$$

Reversible-Jump MCMC

The model is fitted using “reversible-jump Markov chain Monte Carlo” methods (Green 1995), a variant of the Metropolis algorithm for problems where the number of parameters is itself one of the unknown quantities. The process involves iterative application of the following sequence of operations:

1. Parental marker haplotypes are first reassigned by randomly selecting a grandparental source at each locus in turn, conditional on the current haplotype assignment of the previous locus, the offsprings’ marker phenotypes at the pair of loci, and the known map distance between them. We have developed a Metropolis-Hastings approach with this proposal to ensure that haplotypes are generated from the correct sampling distribution, conditional on *all* the loci. However, we found that the

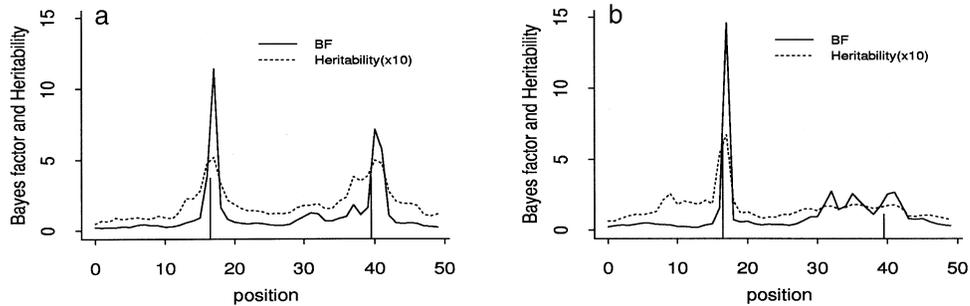


Figure 5 BFs and estimates of heritability ($\times 10$) across 50 marker positions when there are two trait loci with distance 25 markers apart. *a*, Equal-heritability model ($b^2 = 0.37, 0.37$). *b*, Unequal-heritability model ($b^2 = 0.64, 0.11$).

acceptance rates were close to 100% in most cases and that the results were virtually identical with and without the acceptance step; most of the simulations described below were conducted without this additional complexity. Results from more extensive investigations on this haplotype update for sparse marker spacings will be found in the Discussion section.

2. For each currently existing locus $\ell = 1, \dots, L$:

2.1. The (ordered) genotypes for each nuclear family are reassigned by sampling from their full conditional distributions, given the phenotypes and flanking markers of the entire sibship, the genotypes at the other trait loci, and the current values of the model parameters (including the location of the locus). This step involves peeling the genotype information onto the parents, after random gene dropping (Ploughman and Boehnke 1989).

2.2. A change in position P_ℓ either to the right or the left by one marker is proposed randomly with equal probabilities, unless the current position is at the beginning or at the end of a chromosome (when the proposal is always away from the end). The Metropolis-Hastings acceptance probability is then a function of the ratio of the peeled likelihoods at

the new and old positions and the ratio of the lengths of the two segments.

2.3. A new map location x_ℓ is sampled from the interval $[X_{cm}, X_{cn}]$, conditional on the current position. A Metropolis-Hastings algorithm is used based on a Beta proposal distribution.

2.4. The penetrance parameters (β_ℓ, δ_ℓ) are updated by sampling from their full conditional distributions, given phenotypes and genotypes.

3. The global parameters α , γ , and Σ are updated by sampling from their respective full conditional distributions.

4. Finally, the number of trait loci L is updated by randomly deciding whether to propose (*a*) adding a new locus or (*b*) deleting an existing one with equal probability (unless $L = 0$). If an addition is proposed, then a position is selected at random from those not already occupied (favoring locations for which the trait residuals show some evidence of linkage), together with parameters x , q , β , and δ drawn at random from their respective priors, and genotypes are assigned as in step 2.1. If a deletion is proposed, then an existing locus is selected with equal probability. The Metropolis-Hastings ratio of true to proposed probabilities of the entire state of

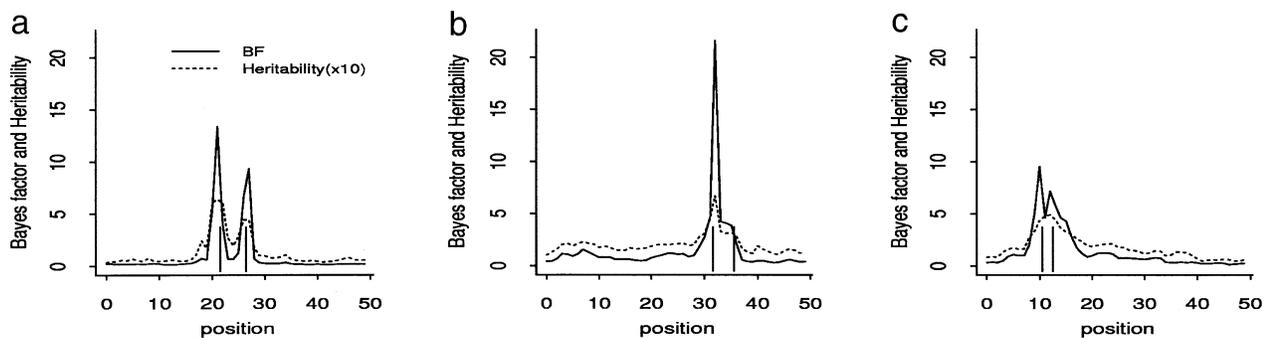


Figure 6 BFs and estimates of heritability ($\times 10$) across 50 marker positions when there are two trait loci with heritability 0.5 and 0.25, varying their spacing (fixed pedigree size, 50×8). *a*, Distance between two loci = 5. *b*, Distance between two loci = 4. *c*, Distance between two loci = 2.

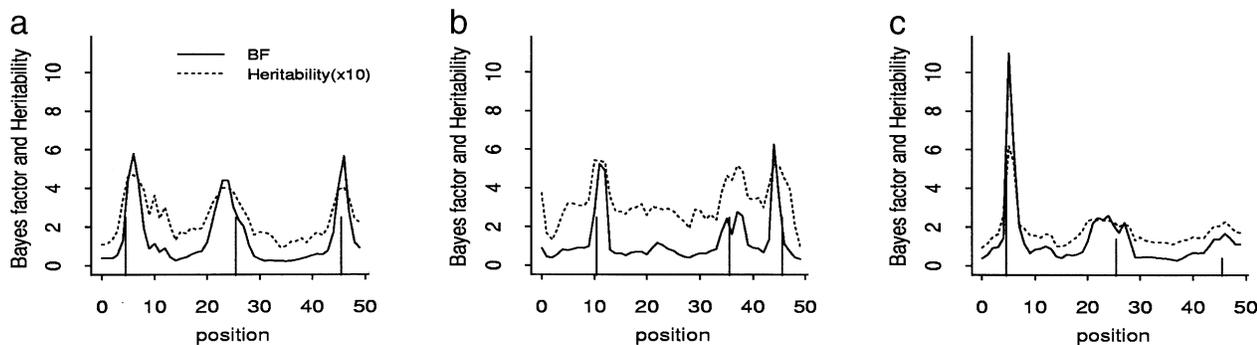


Figure 7 BFs and estimates of heritability ($\times 10$) when there are three trait loci with total heritability 0.75 (fixed pedigree size, 50×8). *a*, Moderate segregation parameters and same heritability. *b*, Rare allele frequencies ($q = .02, .05, .2$). *c*, Disproportionate heritability ($b^2 = 0.57, 0.14, \text{ and } 0.04$).

the system under the new and old parameters is then computed to decide whether to accept the proposed change.

The process is initialized by a random selection of L from its Poisson prior, followed by random assignment of displacement parameters and genotypes in the same way as for new loci (step 4). A more detailed description of our updating schemes can be found in the appendix. Many parts of our updating algorithms are similar to those in Heath (1997), but there are several important differences, especially in the methods of assigning marker haplotypes and of adding, deleting, and moving trait loci.

In the first step, we generate an independent assignment of marker haplotypes at each iteration, conditional only on the marker genotypes, whereas Heath updates the marker haplotypes by sampling from their full conditional distributions. Details of our approach to haplotyping are provided in the appendix, together with an argument for why our approach might be expected to generate approximately the same marginal distribution, but with less dependency between iterations.

Steps 2.2 and 4 of our algorithm also differ from Heath's. In step 4, we propose to add loci by scanning

the entire genome for evidence of linkage of the trait residuals (adjusting for the loci in the current model) using a two-marker variant of the Haseman-Elston algorithm described in the appendix, whereas the reverse move—deletion of an existing locus—is proposed with equal probability among all current loci. In contrast with Heath's proposal of simple random sampling from all currently unoccupied positions, our approach is more likely to propose a locus which will be accepted in the Metropolis step. Likewise, in step 2.2 we propose a simple move of an existing locus by one marker position to the right or left (supplemented by another move of the exact location within its current marker interval), whereas Heath proposes to split an existing locus into two—one at the same locus and one at an entirely new locus. Again, details are provided in the appendix.

Simulation Study

We performed a simulation study to assess the efficacy of our MCMC method for identifying—under various combinations of sibship size, genetic parameters, and heritability—the number of trait loci, their locations,

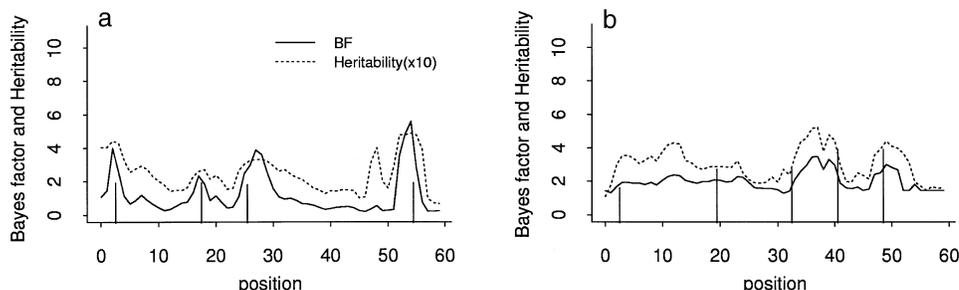


Figure 8 BFs and estimates of heritability ($\times 10$) across 60 marker positions when there are four and five trait loci, respectively (total heritability 0.75). *a*, Four-locus model. *b*, Five-locus model.

Table 2

Simulated Nuclear-Family Data with a Single Trait Locus at Various Combinations of Pedigree Size, QTL Location (Left Marker), and Segregation Parameters

Model	$I \times J$	Location (Fitted)	(BF)	q (Fitted \pm SD)	β (Fitted \pm SD)	δ (Fitted \pm SD)	h^2 (Fitted)
1a	50 \times 8	10 (10)	(.36)	.5 (.33 \pm .23)	.94 (.54 \pm .36)	.5 (.48 \pm .28)	.10 (.03)
1b	50 \times 8	10 (10)	(1.99)	.5 (.39 \pm .16)	1.63 (1.08 \pm .37)	.5 (.52 \pm .23)	.25 (.12)
1c	50 \times 8	10 (10)	(6.55)	.5 (.40 \pm .11)	2.83 (2.12 \pm .49)	.5 (.56 \pm .14)	.50 (.36)
1d	50 \times 8	10 (10)	(7.75)	.5 (.41 \pm .09)	4.90 (4.05 \pm .84)	.5 (.50 \pm .11)	.75 (.66)
1d.r1-r10	50 \times 8	10 (10)	(7.32–7.75)	.5 (.40–.42)	4.90 (4.02–4.14)	.5 (.49–.50)	.75 (.66–.68)
1e	80 \times 5	10 (10)	(8.54)	.5 (.39 \pm .10)	4.90 (4.26 \pm .92)	.5 (.51 \pm .10)	.75 (.69)
1f	200 \times 2	10 (11)	(5.91)	.5 (.41 \pm .12)	4.90 (3.97 \pm 1.13)	.5 (.55 \pm .15)	.75 (.66)
1g	14 \times 8	10 (9)	(7.75)	.5 (.44 \pm .12)	4.90 (4.05 \pm .94)	.5 (.50 \pm .12)	.75 (.67)
1h	40 \times 5	10 (10)	(7.10)	.5 (.37 \pm .12)	4.90 (4.06 \pm 1.09)	.5 (.49 \pm .12)	.75 (.66)
1i	400 \times 2	10 (11)	(10.56)	.5 (.44 \pm .10)	4.90 (4.45 \pm .98)	.5 (.50 \pm .11)	.75 (.71)

NOTE.—In the parentheses the identified location with the highest BF, its BF, MCMC-fitted estimates of the segregation parameters, and their corresponding sample SDs are shown. In 1d.r1-r10, we showed the minimum and maximum estimated values (in the parentheses) of BF and segregation parameters from 10 independent runs on 1d data set with different starting points.

and segregation parameters. For this study, we simulated nuclear-family data, fixing the total number of offspring at 400 in most cases—50 families with sibship size 8, 80 families with sibship size 5, or 200 families with sibship size 2. We used a fixed map of 50 markers, randomly spaced an average of 2 cM apart on a single chromosome, with 4–8 alleles and randomly chosen allele frequencies at these markers. We then set L to 0, 1, ..., 5 and chose their locations x_ℓ at random (subject to the constraint that at most one trait locus was located between any two markers). For single-QTL models, we first set the trait allele frequency $q_\ell = 0.5$ and dominance $\delta_\ell = 0.5$ and then chose β_ℓ to produce heritabilities h_ℓ^2 of 10%, 25%, 50%, and 75%, respectively, where heritability is defined as

$$h_\ell^2 = \frac{\beta_\ell^2 \sum_G f^2(G) Q_G(q_\ell)}{\sigma^2 + \sum_\ell \beta_\ell^2 \sum_G f^2(G) Q_G(q_\ell)}$$

and $Q_G(q)$ denotes the Hardy-Weinberg probabilities for genotype G corresponding to allele frequency q . For models with two or more QTLs, we generated similar nuclear-family data with various combinations of parameter values q_ℓ , δ_ℓ , and β_ℓ , chosen to attain specified values h_ℓ^2 adding up to 75%. Marker alleles were assigned at each locus to each set of parents, assuming Hardy-Weinberg and linkage equilibrium and random mating. Parental marker sources were assigned to their offspring locus-by-locus, conditional on the previous marker locus. Trait alleles were then assigned independently to parents and transmitted to their offspring, given the markers at the flanking loci. Finally, normally distributed phenotypes were assigned to parents and their offspring, conditional on their trait genotypes.

Each data set was analyzed by our MCMC approach described above, retaining 10,000 samples after dis-

carding the first 1,000 iterations to allow for convergence. Figure 2 displays convergence diagnostics for four runs with $L = 0, 1, 2,$ and 3 , respectively, providing the cumulative occupancy fractions for the fitted posterior distribution of L . It is evident that these percentages have stabilized before the first 1,000 iterations ($-1,000-0$ in fig. 2). In general, convergence was faster than illustrated here for models with weaker heritabilities, as exemplified in the panel with simulated $L = 0$. Each run required 64–87 min on a multiprocessor SUN Ultra-4 Sparc Unix system. We now describe our results in each model with a different number of true trait loci in turn.

Null Model

We first tested whether any spurious trait loci were detected when there is no true locus. In the plots which follow for other models, we show the BFs for location (defined as the posterior frequency with which a trait locus at each position P appeared in the model divided by its prior probability $\lambda(X_{cm} - X_{c,m-1})/\tilde{X}$) and the mean heritability parameter h^2 over all assignments to that position. For the null model, the BFs for location never exceeded 0.5 and the displacement estimates were uniformly <1.0 over all positions (data not shown). The other parameter estimates were also uniformly estimated, with similar values and high sample SDs. Table 1 provides a summary of the BFs for L for this model and those that follow. When the true $L = 0$, the highest BFs were for $L = 0$, although $L = 1$ also had $BF > 1$.

Single-Locus Models

In models with $L = 1$, we were primarily interested in assessing the detectability of loci with various degrees of heritability and different sample-size distributions. Initially, we fixed the latter at 50 sibships of size 8 and varied

Table 3

Simulated Nuclear-Family Data with Two Trait Loci: Five Models of Different Combinations of QTL Location (Left-Marker) and Segregation Parameters

Model	Location (Fitted)	(BF)	q (Fitted \pm SD)	β (Fitted \pm SD)	δ (Fitted \pm SD)	h^2 (Fitted)
2a	16 (17)	(11.41)	.5 (.40 \pm .09)	3.50 (3.02 \pm .52)	.5 (.45 \pm .11)	.37 (.35)
	39 (40)	(7.14)	.5 (.41 \pm .09)	2.82 (2.24 \pm .44)	1.0 (.90 \pm .13)	.37 (.32)
2b	16 (17)	(14.59)	.5 (.40 \pm .07)	3.70 (3.16 \pm .47)	1.0 (.93 \pm .08)	.64 (.63)
	39 (32)	(2.74)	.5 (.33 \pm .16)	1.85 (1.46 \pm .55)	.5 (.36 \pm .24)	.11 (.06)
2c	21 (21)	(13.37)	.3 (.27 \pm .06)	2.83 (2.85 \pm .40)	1.0 (.92 \pm .09)	.50 (.49)
	26 (27)	(9.34)	.5 (.43 \pm .11)	2.83 (2.46 \pm .61)	.5 (.60 \pm .15)	.25 (.22)
2d	31 (31)	(4.65)	.3 (.32 \pm .15)	2.83 (2.26 \pm .95)	1.0 (.74 \pm .23)	.50 (.21)
	35 (32)	(22.60)	.5 (.26 \pm .08)	2.83 (3.16 \pm .53)	.5 (.89 \pm .11)	.25 (.53)
2e	10 (10)	(9.53)	.3 (.31 \pm .13)	2.83 (2.32 \pm .67)	1.0 (.68 \pm .26)	.50 (.28)
	12 (12)	(7.16)	.5 (.31 \pm .12)	2.83 (2.46 \pm .74)	.5 (.73 \pm .24)	.25 (.35)

NOTE.—Total heritability is fixed at 0.75, and pedigree size is fixed at 50×8 . In the parentheses, the identified locations with the highest *BFs*, their *BFs*, MCMC-fitted estimates of the segregation parameters, and their corresponding sample SDs are shown.

the heritability between 0.1 and 0.75. (For all these comparisons, we fixed $q = \delta = 0.5$; we explore the effects of these parameters in later models.) Figure 3 illustrates how we will summarize the results for the comparisons to follow. Plotted as a function of location x are $BF(x)$, the estimated value of β , and the heritability $h^2(x)$ defined above, together with the simulated location and value of h^2 . The main plot of $BF(x)$ is based on the ratio of the posterior density for each marker interval, divided by its prior (which is proportional to the length of that interval).

Note that all three plots demonstrate a clear peak in this simulation at the location of the simulated locus, but the trace for $BF(x)$ shows the narrowest peak and the highest change in relative magnitude between baseline and peak. In particular, the $BF(x)$ drops to <1 within two marker loci of the simulated location (approximately ± 5 cM). In what follows, therefore, we will focus on the plots of $BF(x)$. Note that our plots show only the *BFs* for each marker position, not the exact location within each interval—although the latter could easily be obtained by smoothing the densities for the actual x_i assignments.

Figure 4 compares the $BF(x)$ traces for heritabilities of 10%, 25%, 50%, and 75% (models 1a–d in tables 1 and 2). Not surprisingly, the stronger the heritability, the higher and narrower the peak becomes. For $h^2 = 10\%$ there is no trace of a peak, but for all larger values the peak is centered at the simulated locus. The *BFs* for L (table 1) similarly show somewhat greater support for $L = 0$ than $L = 1$ in the case of $h^2 = 10\%$. For larger heritabilities, however, the highest $BF(L)$ is always for $L = 1$, and $BF(L = 0)$ drops precipitously; nevertheless, there also appears to be some weak support for $L = 2$. Table 2 provides the simulated and fitted segregation parameters for the subset of MCMC samples corresponding to the highest *BF* in the neighborhood of the simulated locus. The estimates of β —and, hence, also the estimates of h^2 —appear to be biased slightly down-

ward, probably because of the inclusion of additional spurious loci which absorb some of the effect of the linked locus in some of the MCMC samples. To check the convergence of our MCMC algorithm from different starting points, it was run 10 times independently with different starting points for a simulated data set (1d.r1–r10), from which the true QTL location was always correctly identified and the mean estimates were remarkably closely estimated (second block in table 2).

We also compared the effects of varying sibship sizes and numbers of families in two ways: first, holding the total sample size (IJ) fixed at 400 (models 1e–f in tables 1 and 2); second, holding the number of sib pairs, $IJ(J - 1)/2$, fixed at ~ 400 (models 1g–i in tables 1 and 2), fixing h^2 at 75%. In all cases, the plots of $BF(x)$ showed clear peaks at the simulated locus. (Additional comparisons were carried out at 10%, 25%, and 50% heritabilities, which showed essentially the same results [data not shown].) The magnitude of the *BFs* and the estimates and SDs of the fitted β s varied between the sample sizes in a somewhat unpredictable manner, however (table 2). Holding the total sample size fixed, the estimates of β clearly became more precise as the sibship size increased, but the trend in maximum *BFs* was not so clear; for the 200×2 case, in addition to the peak close to the true simulated locus ($BF = 5.91$), two false peaks in $BF(x)$ appeared well away from the simulated locus, but neither exceeded 1.0 (data not shown). Holding the total number of sib pairs fixed, the SDs of β were similar, but the highest, $BF = 10.56$, was attained for $I = 200, J = 2$. There were no clear differences in the heritabilities.

Two-Locus Models

In two-locus models, we explored the resolution of the method as a function of the heritabilities and the

Table 4

Simulated Nuclear-Family Data for Three- and Four-Locus Models with Heritability 0.75 and $I \times J = 50 \times 8$

Model and Location (Fitted)	(BF)	q (Fitted \pm SD)	β (Fitted \pm SD)	δ (Fitted \pm SD)	b^2 (Fitted)
3a:					
4 (6)	(5.78)	.3 (.34 \pm .12)	2.2 (2.29 \pm .60)	.9 (.77 \pm .17)	.25 (.27)
25 (23)	(4.41)	.45 (.42 \pm .13)	2.2 (2.10 \pm .65)	1.0 (.72 \pm .22)	.25 (.21)
45 (46)	(5.65)	.55 (.49 \pm .13)	2.2 (2.17 \pm .50)	.0 (.20 \pm .18)	.25 (.21)
3b:					
10 (11)	(5.25)	.02 (.16 \pm .16)	8.9 (4.07 \pm 1.71)	.5 (.51 \pm .34)	.25 (.27)
35 (37)	(2.78)	.05 (.22 \pm .16)	5.7 (3.45 \pm 1.52)	.5 (.51 \pm .19)	.25 (.24)
45 (44)	(6.25)	.2 (.21 \pm .13)	4.1 (3.54 \pm 1.12)	.5 (.52 \pm .16)	.25 (.25)
3c:					
4 (5)	(10.99)	.7 (.63 \pm .09)	3.02 (2.74 \pm .43)	.0 (.07 \pm .11)	.57 (.50)
25 (24)	(2.61)	.7 (.40 \pm .17)	1.51 (1.62 \pm .65)	.0 (.45 \pm .23)	.14 (.09)
45 (46)	(1.67)	.7 (.38 \pm .18)	.76 (1.57 \pm .73)	.0 (.52 \pm .24)	.04 (.09)
4:					
2 (2)	(4.01)	.2 (.33 \pm .15)	4.4 (2.76 \pm 1.39)	.0 (.16 \pm .18)	.19 (.20)
17 (17)	(2.33)	.2 (.32 \pm .18)	3.0 (1.77 \pm .80)	.5 (.52 \pm .24)	.19 (.12)
25 (27)	(3.87)	.2 (.31 \pm .15)	1.8 (1.96 \pm .91)	1.0 (.61 \pm .21)	.18 (.14)
54 (54)	(5.58)	.2 (.29 \pm .10)	4.4 (3.59 \pm .84)	.0 (.10 \pm .10)	.19 (.24)

NOTE.—In the parentheses, the locations with the highest BFs, MCMC-fitted estimates, and their corresponding sample SDs are shown.

distance between the two loci, from here on fixing the pedigree size at 50×8 . We first show $BF(x)$ for two widely separated loci, in one case with the same heritability ($b_1^2 = b_2^2 = 37\%$; model 2a in tables 1 and 3 and fig. 5a) in and the other with very different heritabilities ($b_1^2 = 64\%, b_2^2 = 11\%$; model 2b in tables 1 and 3 and fig. 5b). In the equal-heritability case, both QTLs were well localized and their parameter estimates (table 3) were close to the true values. Not surprisingly, in the unequal-heritability case, the weaker locus showed a more broadly dispersed peak, in terms of BFs, and a $\hat{\beta}$ that was only modestly elevated above that at other locations but was still detectable, despite the coexistence of a much stronger locus.

The number of loci was not well estimated. Table 1 shows that in equal-heritability case, for example, the peak $BF(L)$ was attained for $L = 4$, whereas the BF for $L = 2$ was only 0.95; however, $L = 0$ and $L = 1$ could both be convincingly rejected by their BFs. Results for other $L = 2$ models were similar.

To investigate the effect of spacing, we considered two loci with more-comparable effects, one dominant and one additive, with heritabilities of 50% and 25%, respectively. In these cases, we varied the distance between the two trait loci from two to five marker positions (models 2c–e in tables 1 and 3). When the two loci were five marker positions apart, both were well localized (fig. 6a), and all the parameters were well estimated (table 3). When we tried to identify two trait loci placed at a gradually smaller distances (four and two markers apart in figs. 6b and 6c, respectively), we could find only a single mode of the BFs for location in between the two loci, lumping their effects

as if they were from a single-trait locus, so that their estimated total heritability was close to the simulated value. Thus, it appears that two trait loci having similar effects can only be distinguished if they are ≥ 10 cM apart in data sets of comparable size and heritability.

Models with Three or More Loci

Figures 7 and 8 illustrate the results for a variety of models having from three to five loci, all reasonably well separated, for a range of segregation parameters adding up to a total heritability of 75%. The corresponding parameter values and their estimates are given in table 4. For the three three-locus cases, we varied the dominance parameter in model 3a, simulated rare alleles with large displacements in model 3b, and varied the heritabilities for three recessive loci in model 3c (tables 1 and 4). In all models, the three loci were identified at or close to their simulated locations, with the possible exception of the minor locus with only 4% heritability in the third model, for which the fitted BF was only 1.67. The segregation parameters were also well estimated in the first and third models; however, in the rare-allele model, the allele frequencies were generally overestimated and the displacements underestimated, but they produced about the right heritability estimates. The four-locus model also produced good estimates of both locations and segregation parameters (tables 1 and 4 and fig. 8a). However, in the five-locus model (tables 1 for BFs), although five peaks attained a $BF(x) > 2$ (fig. 8b), the locations and the parameter values were not estimated well, so we do not report their actual data here. There could be several rea-

Table 5
Simulated Nuclear-Family Data with Two Trait Loci with Multiplicative Interaction Effects: High- (I1) and Low-Interaction (I2) Effect Models

Model and Location (Fitted)	(BF)	q (Fitted \pm SD)	β (Fitted \pm SD)	δ (Fitted \pm SD)	b^2 (Fitted)
I1:					
16 (17)	(7.79)	.5 (.38 \pm .11)	2.19 (2.37 \pm .46)	.5 (.39 \pm .12)	.15 (.31)
Interaction Effects	2.2445
39 (37)	(2.45)	.5 (.38 \pm .15)	1.79 (1.54 \pm .68)	1 (.63 \pm .27)	.15 (.17)
I2:					
16 (17)	(10.19)	.5 (.41 \pm .09)	3.15 (2.80 \pm .48)	.5 (.44 \pm .11)	.30 (.35)
Interaction Effects6415
39 (40)	(5.30)	.5 (.38 \pm .11)	2.55 (1.96 \pm .48)	1 (.84 \pm .17)	.30 (.27)

NOTE.—Total heritability is fixed at 0.75 and pedigree size is at 50×8 . In the parentheses the identified locations with the highest *BFs*, their *BFs*, MCMC-fitted estimates of the segregation parameters, and their corresponding sample SDs are shown.

sons for this, most likely their weaker heritabilities and the closer distances between them.

Two-Locus Models with Interaction Effects

We studied the robustness of our approach for the case that our assumption of locus-additivity was misspecified. We simulated data including a multiplicative interaction between two trait loci in the model of the form

$$Y_i \sim N[\beta_0 + \beta_1 f(G_{i1}) + \beta_2 f(G_{i2}) + \beta_3 f(G_{i1})f(G_{i2}), \sigma^2],$$

where f is defined as in equation (5). Fixing the total heritability at 75%, we performed two experiments setting the interaction parameter (β_3) to provide 45% and 15% heritability, respectively (models I1–2 in tables 1 and 5; fig. 9). The additive effects of the two trait loci (after interaction effects were subtracted) were then divided evenly between the two. Thus, the first model of strong interaction (I1) had two QTLs, each with heritability 15%, and the second model of weak interaction (I2) had two, each with heritability 30%. As expected, in the weak interaction case the true locations, parameter values, and heritability of the two QTLs were precisely predicted (fig. 9b1–b2 and the second block in table 5). In the strong-interaction case, their locations and individual displacement parameters (β) were still reasonably well captured but were much more widely dispersed.

Comparisons with LOKI

We also compared the performance of our algorithm with that of Heath’s (1997) program LOKI. For this purpose, we ran that program on a subset of the simulated data sets described below, using 40,000 iterations after a burn-in period of 10,000 iterations. For both methods, we computed the autocorrelation function (ACF) of the k th lag, r_k , and from this we estimated the “variance in-

flation factor” (VIF) using the time series method suggested by Geyer (1992):

$$VIF = 1 + \sum_{k=1}^K r_k [1 + \cos(\pi k/K)].$$

This can be used to compute the effective sample size (ESS) as R/VIF , which can be interpreted as the equivalent number of samples if they were independent, where R is the total length of each MCMC run. Figure 10 shows plots of the ACF of allele frequency q and the parameter L from two runs for the one-locus model 1b and a three-locus model similar to 3c but with $\delta_\ell = 1$ (dominant trait alleles) and $q_\ell = 0.3$. For both cases, the autocorrelations die away much faster for our algorithm than for LOKI. This translates into a much smaller sample size required to attain the same degree of Monte Carlo error. Table 6 shows VIF, ESS, and computer time required for these runs by LOKI and our algorithms at the same UNIX station as for our simulation studies. It is evident that, although the simpler LOKI algorithm runs about twice as fast as ours per iteration (348 vs. 171 iterations/minute for the one-locus model), the VIF for L is 3.5 times higher, so that the 40,000 LOKI iterations are roughly equivalent to the 10,000 of ours (ESS for L 775 and 680, respectively). Thus, our algorithm appeared to be at least twice as efficient as LOKI in terms of MCMC sampling (total length of run 64.4 vs. 143.4 minutes for the one-locus model).

In LOKI, there is no constraint that β_ℓ be positive, so the resulting distributions are often multimodal, with the different modes essentially corresponding to the same model with the designation of the high-risk allele reversed for one or more loci, with q and δ redefined correspondingly (S. Heath, personal communication). To avoid this problem, our approach constrains each of the β s to be positive. In addition, LOKI estimates the displacements of the aA and AA genotypes from aa sepa-

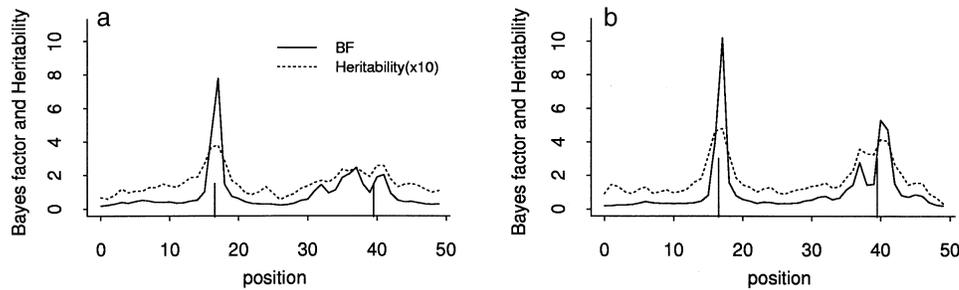


Figure 9 BFs and estimates of heritability ($\times 10$) across 50 marker positions when there are two trait loci with multiplicative interaction effects (total heritability 0.75; 25 marker distance between two loci). *a*, Strong interaction effects (interaction heritability 0.45). *b*, Weak interaction effects (interaction heritability 0.15).

rately, whereas our estimates are related via the parameters β and δ . For these reasons, the two sets of parameter estimates are not strictly comparable. For a comparison of parameter estimation, we therefore have reported, in table 6, the estimates and SDs of L , q , effect of heterozygote (E_{12} , deviation from E_{11}), effect of second homozygote (E_{22} , deviation from E_{11}), and derived quantities h_c^2 , whose distributions are not subject to the problem and tend to be unimodal. For the single-QTL case, LOKI identified the trait locus with a higher heritability (0.17; closer to the true $h^2 = 0.25$) than our approach, but for the three-QTL cases, LOKI failed to capture the effects of the major trait locus, whereas our algorithm was able to identify it close to the true simulated effects. Furthermore, the VIFs for q (or q_1 , the allele frequency of the largest trait locus) are substantially smaller from our program than from LOKI.

Notice also that the posterior distributions for L are somewhat broader. To determine which has the more appropriate coverage would require many replicate simulations, which would be beyond our computing resources. Nevertheless, it is evident from the comparisons of the iteration results for x shown in figure 11 that our algorithm appears to be mixing better.

Discussion

The reversible-jump MCMC methods for fitting QTL models with an unknown number of trait loci that have been developed quite independently for line crosses (Sattogopan et al. 1996; Stevens and Fisch 1998; Sillanpää and Arjas 1998, 1999) and for human nuclear-family and extended-pedigree data (Heath 1997; Thomas et al. 1997) are quite similar. The methods for experimental crosses do not involve many of the complexities of sampling the unobserved trait genotypes that arise in the human context, but the methods of updating the number of trait loci and their parameters are essentially the same. In the human context, the sampling of trait genotypes and the ar-

rangement of markers and trait genotypes into haplotypes is not so straightforward and requires an additional step.

Our approach differs from that of Heath (1997) in two essential ways. First, Heath does not distinguish between trait and marker loci in sampling haplotypes, and the sampling of parental sources at each locus is based on their respective full conditional distributions. In our approach, marker haplotypes are sampled first, without reference to the trait phenotypes; despite this potential loss of information, the implementation of this approach is somewhat simpler and may improve convergence, since the samples of haplotypes are independent from one cycle to the next (or almost independent, if the Metropolis-Hastings step is used). Second, our proposals for adding, deleting, and moving trait loci differ. Heath uses two different proposals for adding loci: either pick a new location completely at random over the entire genome or split an existing locus, retaining the position of the one and sampling a second position again at random over the entire genome. Our approach uses only the first type of proposal to add a locus but bases the choice of location parameter on a variant of the Haseman-Elston (1972) method to preferentially sample regions where the residuals from the current model suggest that another trait locus is likely to exist. This produces acceptance rates that are typically 18%–20% if the true L is >0 (if the true $L = 0$, the acceptance rate for “add moves” is 32%, and, for “delete moves” where $L = 1$, the acceptance rate is 73%). In contrast, the acceptance rates of LOKI’s add/delete move of a QTL by its reversible-jump algorithm was 1.3%–1.4%. Instead of split/join moves to change the position of a trait locus, we allow moves from one pair of flanking markers to an immediately adjoining pair, by means of a Metropolis-Hastings step. The acceptance rates for these moves is typically ~62%–75%, and location updates (within positions) generally have acceptance rates of 91%–94%. These two types of moves accomplish quite different purposes; an add

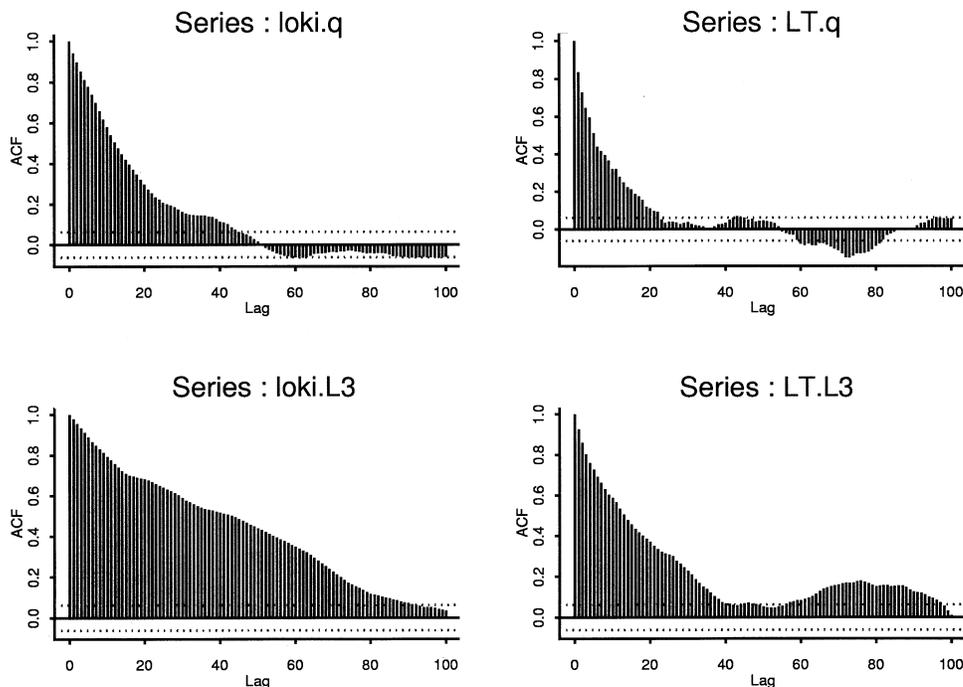


Figure 10 Auto-correlation functions (ACF) of LOKI and Lee-Thomas MCMC samples of q for one-locus model 1b and of the number of trait loci (L) for the three-locus model in table 6.

move allows new loci to be created far from any existing locus (but more likely in regions where there is some evidence of linkage in the residuals), whereas the position update allows an existing locus to jump over the boundaries defined by its flanking markers. Given that a locus already has been placed in a particular location, we feel that local moves (to adjacent positions or within an interval) are more likely to be accepted than are moves to more-distant locations, as are done in LOKI, as well as being computationally more efficient. We have not attempted a formal comparison of the separate effects of these methodological differences on their performance, but when these differences are taken together, it appears that they substantially improve the overall mixing of the algorithm.

A reviewer queried whether it would be necessary to add a constraint on the total variance predicted by the model. We do not feel this is necessary, since this will tend to be taken care of automatically in the updates of the parameters. However, it does suggest that the proposal to add loci might be improved by addition of such a constraint or by reallocation of either the displacement parameters for all loci or the residual variance to maintain the same total variance. We have not explored whether this would improve the performance, but the relatively high acceptance rates of our current proposal suggests that this is not really needed. A final difference is that Heath’s program LOKI is applicable to extended

pedigrees, whereas ours has been implemented so far only for nuclear families. However, ours could be extended to pedigrees without any fundamental changes in logic by modifying three parts—marker-haplotype updating, genotype assignment, and the proposal for adding loci.

Similar to the nuclear-family case, the haplotype update could be done locus-by-locus, calculating the probabilities of all possible configurations of the founders’ haplotypes for a pair of loci, conditional on the observed genotypes for the entire pedigree at these two loci, choosing a new haplotype configuration with these probabilities and then using the Hastings ratio to decide whether to accept the proposed change at that pair using the entire pedigree of haplotypes. One could also update one founder’s haplotype at a time by proposing a random change to the entire haplotype (a swap of a single marker or a whole segment) and using the Hastings ratio again to accept or reject that proposal. Additional moves to non-founders’ segregation indicators are also needed, to allow for recombination. We have implemented an MCMC haplotyping algorithm that incorporates such approaches and appears to work for general pedigrees, including those with inbreeding loops. However, further research is needed to compare the mixing performance of these or other approaches in complex pedigrees, such as that used in LOKI, as there are well-known difficulties with MCMC methods in multiallelic systems in complex pedigrees (Lin 1993).

Table 6

Comparison of Performance of LOKI and Lee-Thomas (LT) Methods in a One-QTL Model (1b) and a Three-QTL Model Similar to 3c: Computer Time, VIF, ESS, and Parameter Estimates and Their SDs

Summary Statistic and Parameter	LOKI	LT
Single-QTL model (1b):		
No. of iterations (burn-in)	50,000 (10,000)	11,000 (1,000)
Computer time (min)	143.4	64.4
VIF (ESS):		
L	51.6 (775)	14.7 (680)
q_1^*	27.6 (1,449)	14.4 (694)
E_{12}^*	28.5 (1,404)	8.0 (1,250)
E_{22}^*	37.7 (1,061)	8.5 (1,176)
Estimate (SD):		
$L = 1$	1.21 (.48)	1.67 (.86)
$q_1^* = 0.50$.40 (.17)	.39 (.16)
$E_{12}^* = 0.82$.58 (.40)	.55 (.27)
$E_{22}^* = 1.63$	1.29 (.39)	1.08 (.37)
$h_1^2 = 0.25$.17 (.05)	.15 (.06)
Three-QTL model:		
No. of iterations (burn-in)	50,000 (10,000)	11,000 (1,000)
Computer time (min)	186.1	86.3
VIF (ESS):		
L	60.2 (664)	31.8 (314)
q_1^*	38.0 (1,053)	9.3 (1,075)
E_{12}^*	27.0 (1,481)	13.3 (752)
E_{22}^*	28.1 (1,423)	11.0 (909)
Estimate (SD):		
$L = 3$	1.18 (.43)	3.27 (.91)
$q_1^* = 0.30$.40 (.21)	.36 (.10)
$E_{12}^* = 3.02$.47 (.49)	2.75 (.47)
$E_{22}^* = 3.02$	1.13 (.45)	3.04 (.51)
$h_1^2 = 0.57$.14 (.06)	.64 (.11)

^a LOKI's output E_{12} , effect of heterozygote, and E_{22} , effect of "22" homozygote, are essentially equivalent to LT's $\beta \times \delta$ and β , respectively. For the comparison between the two methods, we reparameterized LOKI's output as $q_1^* = q_1$ if $E_{22} > 0$, $1 - q_1$ otherwise, $E_{12}^* = E_{12} - \min\{0, E_{22}\}$, and $E_{22}^* = |E_{22}|$, since LOKI's output has multimodality caused by aliasing of the two alleles at each trait locus.

We also investigated the performance of our haplotype algorithm for larger marker spacings—5, 10, and 20 cM, on average. For nuclear-family data, our approximation appeared to be very good. For example, in the 5-cM case, the BF and estimates of segregation parameters were closely estimated: BF (6.98, 6.92), $q = 0.5$ (0.41, 0.41), $\beta = 2.83$ (2.23, 2.22), and $\delta = 0.5$ (0.52, 0.52), correcting and not correcting the Hastings ratio, respectively; similar results were observed in 10- and 20-cM cases. Thus, the conditional haplotype update using two flanking markers seems to be fairly close to the full conditional distribution of the trait loci and marker data in these nuclear-family data. As expected, when larger spacings were used, the BF s became slightly smaller—6.98, 6.08, and 5.47 for 5-, 10-, and 20-cM cases, respectively—reflecting the loss of information from using a sparser marker map.

The genotype update can be extended using the same peeling and gene-dropping algorithm of Ploughman and Boehnke (1989) as in our program and LOKI, but this would require that pedigrees be peelable. For complex pedigrees, MCMC methods could be used but might require many more iterations.

Our proposal for creating new loci, based on a Haseman-Elston regression of the squared differences in trait residuals (adjusted for loci already in the model) on identical-by-descent (IBD) sharing probabilities, could be extended to larger pedigrees in a number of ways. The simplest would be to include all possible sib pairs within each pedigree, treating them as independent, but one might also consider including more-distant relative pairs. Even though these contributions would not be independent, it is not necessary for this purpose that the resulting proposal be the correct likelihood but only that it be able to suggest areas of relatively higher likelihood for finding genes. The real likelihoods, based on a peeling calculation under the new and old models, would then be used to accept or reject the proposed new locus.

Over a wide range of true model specifications, we found that our method performed well at identifying the location of trait loci and at estimating their penetrances and allele frequencies, provided that their contribution to heritability was sufficiently large and that they were not too close together. When the model was misspecified by ignoring epistasis, the locations and segregation parameters of the trait loci were still well estimated, especially when the interaction effects were small. We have not extended our algorithm to estimation of epistatic models, but in principle this could be done by adding terms to the penetrance model; this might be accomplished by further reversible-jump operations to add and delete interaction terms, including higher-order interactions in a hierarchical manner.

The approach was less successful at identifying the true number of trait loci; in most cases we considered, the true value of L had the highest BF , but the variation in BF s was often not very large. Not surprisingly, when the heritability of one or more true loci was small (either low allele frequency or low penetrance), the method tended to underestimate L , and, for the more informative data sets (50×8), the BF s for L were more strongly peaked than for the less informative data sets (200×2) for the same heritability. Another situation where the method tended to break down was when two loci were too close to each other, where the posterior distribution of locations tended to coalesce into a single broad peak. Similar experience was reported by Richardson and Green (1997) and Lee et al. (1998) in applying reversible-jump MCMC to the problem of estimating mixture distributions with an unknown number of components. They found that in several real data sets, their mixture of normal densities provided no clear inferences on the number of compo-

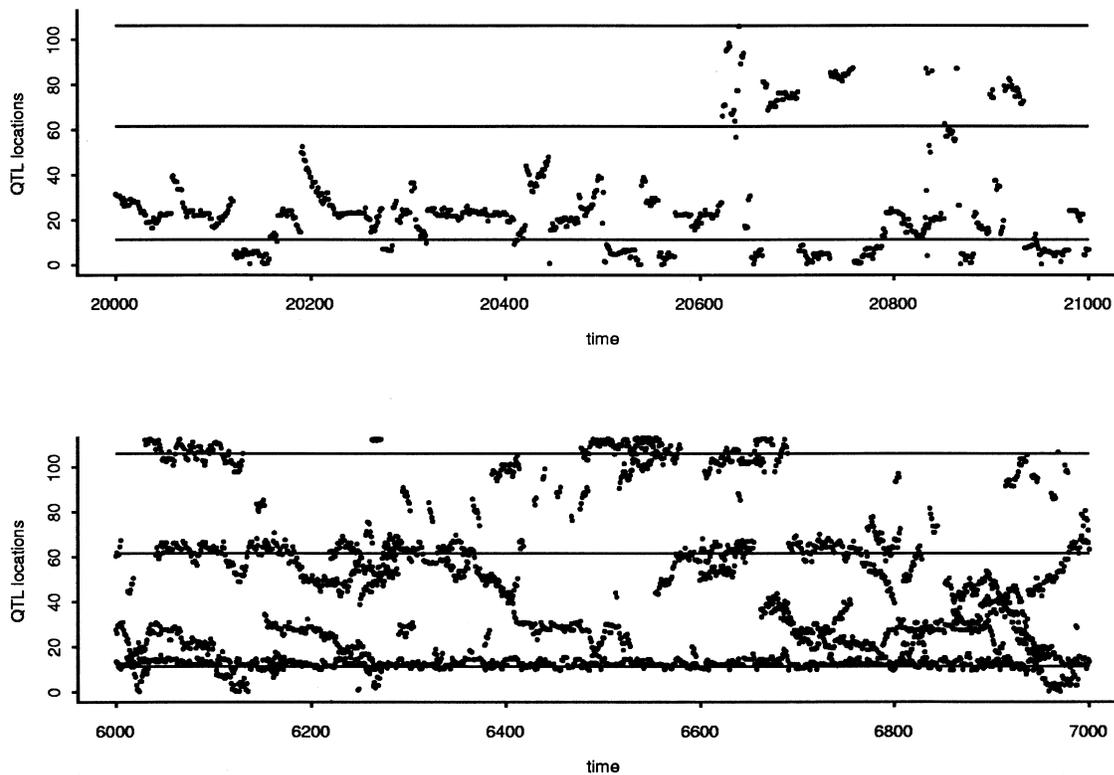


Figure 11 Time series plots of QTL locations of LOKI and Lee-Thomas methods for the simulated data of the three-locus model in table 6. Each of 1,000 iterations (randomly chosen) was shown for visibility.

nents, although the fitted *marginal* densities (averaging over the various numbers of components) provided an excellent fit to the observed density. The former authors speculate that the reason for this behavior is that the prior model does not penalize overfitting of many components and that a “decision theory” approach with a cost function or a combination with other criteria of overfit might be more appropriate.

The *BF* can be interpreted as a summary of the evidence provided by the data in favor of a postulated statistical model against a null model—presence of a QTL against no QTL at a particular locus in our case. Kass and Raftery (1995) have suggested guidelines for interpretation of this statistic: a *BF* of 1–3 can be considered as “very mild evidence,” 3–20 as “positive,” 20–150 as “strong,” and >150 as “very strong.” From our simulation results, most (25/26) of the simulated trait loci with heritability $\geq 25\%$ were identified by *BFs* >3, and some (7/11) of the simulated trait loci with heritability of 15%–25% were also identified by *BFs* >3. Therefore, our MCMC approach identified the trait loci with reasonably high heritability.

The computational demands of the method precluded a full-scale simulation study with many replicates of each choice of parameters. Thus, we are unable to address the type I and II error rates (although we hasten to point out

that the repeated sampling concepts of test size and power are not really meaningful in the Bayesian context in which our methods are set). However, the absence of any *BFs* >1.0 in the $L = 0$ case for any of our current and two other simulation conditions (153 possible positions in total; data not shown) suggest that that type I error rate is acceptably low. Similarly, the clear signals at or near most of the true simulated loci (38 true loci out of 41 with heritability >10% produced a *BF* of ≥ 1.0 within two marker positions of the true location) also suggest that power is good, provided that the heritability is sufficiently large. More-ambitious simulation studies with replication would be helpful to better understand the limits of detection of the method and its robustness to model misspecification. It is also worth noting the excellent performance of LOKI in its applications to the GAW 10 and 11 data sets, in which quite complex models were simulated and the fitted models were necessarily somewhat misspecified.

Acknowledgements

This work was supported in part by grants from the National Cancer Institute of the United States Public Health Service (CA 52862 and GM 58897) and National Science Foundation grant

BIR 95-04393. We thank Dr. Sylvia Richardson for help in the development of the algorithm, Dr. Simon Heath for useful discussions on LOKI's output, and the reviewers for helpful comments and many editorial suggestions on the manuscript.

Appendix A

MCMC Updating Procedures

Haplotype Assignment

The marker phenotypes are randomly reassigned to haplotypes at the beginning of each cycle. This is done sequentially on each chromosome, beginning with an arbitrary assignment of the marker alleles at the first locus to grandparental sources and then conditioning the assignment of subsequent loci on the grandparental sources of the previous locus. Let us consider two marker loci, each with four different allele types: a, b, c, d for the first locus and A, B, C, D for the second one. We then distinguish three potentially informative configurations of parental marker phenotypes at the second locus: (1) both parents heterozygous and sharing, at most, one allele; (2) one parent heterozygous; (3) parents sharing two alleles with subtypes (a) if the offspring is homozygous and (b) if the offspring is heterozygous. In each of these situations, the haplotype probabilities depend on the recombination fraction, as products over all the offspring of $\theta/2$, $(1 - \theta)/2$, and $\Theta = [\theta^2 + (1 - \theta)^2]/2$. A pair of parental haplotypes is then sampled with these probabilities. Grandparental sources T_{ijcmp} are assigned to each of the offspring where they can be inferred directly by matching the alleles; for example, in configuration (1), if the parents were assigned haplotype aA|bB × cC|dC, then an offspring with genotype ac,AC could only be assigned haplotype aA|cC with sources fF|fM (where “f” and “m” denote grandpaternal and grandmaternal sources, respectively, for the first locus, and likewise for the second locus). Similar situations arise in configuration (3a) and for the heterozygous parent in configuration (2). In the ambiguous situations, the two possible haplotypes are assigned at random with the appropriate probabilities. For example, in configuration (2), the source for the homozygous parent would be assigned with probability θ or $1 - \theta$, depending upon the assignment at the first locus. In configuration (3b), if the parents were assigned haplotypes aA|bB × cB|dA, then an offspring with genotype ac,AB could be assigned either haplotype aA|cB (fF|fF) or haplotype aB|cA (fM|fM) with probabilities $(1 - \theta)^2/2\Theta$ and $\theta^2/2\Theta$ respectively; on the other hand, if the parents were assigned aA|bB × cA|dB, the two offspring haplotypes would be assigned with equal probability.

As noted by a reviewer, sampling locus by locus in this fashion does not exactly generate the correct haplotype distribution, because we do not use the full phenotype data. Specifically, the true distribution, $P(\mathbf{T}|\mathbf{H})$, can be decomposed as

$$[T_1|\mathbf{H}][T_2|T_1, \dots][T_L|T_1, \dots, T_{L-1}, \mathbf{H}] ,$$

but what we actually sample from is

$$Q(\mathbf{T}|\mathbf{H}) = [T_1|H_1][T_2|T_1, H_1, H_2] \dots P[T_\ell|T_{\ell-1}, H_{\ell-1}, H_\ell] \dots P[T_L|T_{L-1}, H_{L-1}] .$$

The appropriate fix is to either accept or reject a new haplotyping based on the Hastings ratio, $R = \min[1, P(\mathbf{T}'|\mathbf{H})Q(\mathbf{T}|\mathbf{H})/P(\mathbf{T}|\mathbf{H})Q(\mathbf{T}'|\mathbf{H})]$. However, in our simulations, it appears that this ratio is generally so close to 1, in most cases, that this additional step is not needed.

Note that the trait phenotype is not used in making these marker-haplotype assignments. The resulting Markov chain thus entails sampling from $[\mathbf{T}|\mathbf{H}][\mathbf{G}|\mathbf{Y}, \mathbf{T}]$, which is approximately proportional to $[\mathbf{G}, \mathbf{T}|\mathbf{Y}, \mathbf{H}]$ if $[\mathbf{T}|\mathbf{H}] \approx [\mathbf{T}|\mathbf{H}, \mathbf{Y}]$ (since $[\mathbf{G}|\mathbf{Y}, \mathbf{T}] = [\mathbf{G}|\mathbf{Y}, \mathbf{T}, \mathbf{H}]$ in our construction). Heath (1997) instead samples from the full conditional distributions $[\mathbf{T}|\mathbf{H}, \mathbf{G}, \mathbf{Y}]$ and $[\mathbf{G}|\mathbf{Y}, \mathbf{T}, \mathbf{H}]$. Both samplers thus generate the same marginal distributions $[\mathbf{G}, \mathbf{T}|\mathbf{Y}, \mathbf{H}]$ but may have different time-series performance. Even though the sampling of marker haplotypes by our method may be less efficient, the samples are independent from one cycle to the next, which we speculate should reduce the auto-correlation in the series of $[\mathbf{G}|\mathbf{Y}, \mathbf{T}, \mathbf{H}]$, which should accelerate convergence and require fewer samples to tabulate marginal distributions.

Genotype Assignment

For nuclear families, it is straightforward to compute the joint probability of all possible genotype vectors at a single locus, conditional on the markers and on the genotypes at all other loci, and make a random draw from that distribution. First, we compute the peeled probabilities for each of the 16 possible genotypes $G_{i\ell}^p = (G_{i\ell s}^p)_{p,s=m,f}$ by summing over the 4 possible genotypes that could have been passed to each of the offspring, and select a parental genotype with the corresponding probability:

$$\Pr(G_{i\ell}^p | \mathbf{Y}_i, \mathbf{T}_i; \boldsymbol{\theta}, \boldsymbol{\Omega}) = \Pr(G_{i\ell}^p | q_\ell) \prod_j \sum_{(g_1, g_2)} \text{Tr}(g_1) \text{Tr}(g_2) R(g_1, g_2), \quad (\text{A1})$$

$$\text{where } \text{Tr}(g_p) = \Pr(G_{ij\ell p} = g_p | T_{ijcm p}, T_{ijcn p}; x_\ell, X_{cm}, X_{cn})$$

$$\text{and } R(g_1, g_2) = \Pr(y_{ij\ell} | G_{ij\ell} = (g_1, g_2); \beta_\ell, \delta_\ell, \boldsymbol{\Sigma}),$$

the three probabilities being given by equations (1), (2), and (4) respectively, with y indicating the phenotype residuals after the effects of all of the other trait loci are subtracted,

$$y_{ij\ell} = \mathbf{Y}_{ij} - \boldsymbol{\alpha} - \boldsymbol{\gamma}' Z_{ij} - \sum_{k \neq \ell} \beta_k f(G_{ijk}).$$

Then, conditional on the parental genotypes, we select a genotype for each offspring with probability given by the terms inside the summation in equation (A1).

Positions

The position probabilities cannot be computed using the current assignment of genotypes, since recombinants are only meaningful with respect to their currently flanking markers. Instead, the position probabilities must be computed by peeling the trait genotypes for each position considered. Since this would be too computationally intensive, we use instead the Metropolis-Hastings algorithm, based on a proposal to move the present position one marker to the left or right. If both positions are currently unoccupied, the choice of which direction to propose is made with equal probability. The acceptance probabilities are then computed using the ratios of $\Pr(\mathbf{Y} | \mathbf{T}_{cm}, \mathbf{T}_{cn})$ (summing over all possible trait genotypes) at the old and new positions, divided by the corresponding ratio of proposal probabilities. If the new position is accepted, new trait genotypes are assigned by sampling from the peeled probabilities, as described under Genotype Assignment above. We have also explored proposal probabilities based on the ratio of the numbers of recombinants with the right and left markers under the current genotype assignment but have not found any improvement in performance.

Locations

The conditional distribution of x_ℓ , given a particular position (C_ℓ, P_ℓ) , is proportional to

$$\theta(x - X_{cm})^{R_{cm\ell}} [1 - \theta(x - X_{cm})]^{N_{cm\ell}} \theta(x - X_{cn})^{R_{cn\ell}} [1 - \theta(x - X_{cn})]^{N_{cn\ell}}.$$

If the current position is at the end of a chromosome, this reduces to an easily sampled Beta distribution in θ . Otherwise, we use a Metropolis-Hastings step, proposing as the new $x_\ell = X_{cm} + (X_{cn} - X_{cm}) \text{Beta}(R_{cm\ell}, R_{cn\ell})$.

Penetrance Parameters

The sufficient statistics for estimating β_ℓ and δ_ℓ are the sample means $\bar{y}_{g\ell}$ of the residuals $y_{ij\ell}$ for the three possible genotypes g and the numbers $n_{g\ell}$ of subjects assigned to each genotype. The log-likelihood is then proportional to

$$n_{0\ell} (\bar{y}_{0\ell} + \beta \Delta_\ell)^2 + n_{1\ell} [\bar{y}_{1\ell} - \beta(\delta_\ell - \Delta_\ell)]^2 + n_{2\ell} [\bar{y}_{2\ell} - \beta(1 - \Delta_\ell)]^2, \quad (\text{A2})$$

where Δ_ℓ is treated as a function of δ_ℓ (eq. [5]). Conditional on δ , equation (A2) is easily expressed as a normal log-likelihood in β , and vice versa. We have found it convenient to use the Metropolis-Hastings algorithm, proposing new values of β and δ from their conditional distributions, truncated at 0 and 1 for δ , and then computing the Hastings

ratio using equation (A2) to decide whether to accept the new parameter values. For β , we allow for the prior in the proposal step by drawing from a normal distribution $N[\hat{\beta}, V(\hat{\beta})]$, restricted to $\beta > 0$, where

$$\tilde{\beta} = \hat{\beta} - \sqrt{V(\hat{\beta})/\omega} .$$

Allele Frequencies

The full conditional distribution for q_ℓ is a Beta($N_\ell, 4I - N_\ell$) distribution, where N_ℓ is the number of G alleles among the parents at locus ℓ , multiplied by the likelihood function for \mathbf{Y} as a function of $\Delta(q)$. However, since the dependence on $\Delta(q)$ is relatively weak, the simplest procedure is to use the Metropolis-Hastings method, with a proposal based on the Beta distribution part. q_ℓ having been updated, a new value of Δ_ℓ is computed, following equation (5).

Number of Trait Loci

Following the reversible-jump MCMC approach of Green (1995), we propose to increase L by 1, with probability b_L , or to decrease it, with probability $d_L = 1 - b_L$, where $b_L = 1/2$ unless $L = 0$ ($b_L = 1$) or $L = L_{\max}$ ($b_L = 0$). To increase L , we create a new trait locus with Θ_{L+1} and Ω_{L+1} drawn from their respective priors, as described above, and then assign $G_{i,L+1}$ by sampling genotypes (as described earlier) with probabilities

$$\Pr(G_{i,L+1} | \mathbf{Y}_i, \mathbf{G}_i, \mathbf{H}_i; \Theta_{L+1}, \Omega) = \frac{\Pr(G_{i,L+1} | \mathbf{H}_i; \Theta_{L+1}) \Pr(\mathbf{Y}_i | \mathbf{G}'_i, \mathbf{Z}_i; \Omega', \Gamma)}{\sum_{G_{i,L+1}} \Pr(G_{i,L+1} | \mathbf{H}_i; \Theta_{L+1}) \Pr(\mathbf{Y}_i | \mathbf{G}'_i, \mathbf{Z}_i; \Omega', \Gamma)} , \tag{A3}$$

where $\mathbf{G}' = (\mathbf{G}, \mathbf{G}_{L+1})$ and similarly for Ω' . To decrease L , we simply propose to eliminate an existing trait locus, selected with equal probability $1/(L + 1)$.

With this proposal, the calculation of the Metropolis-Hastings ratio becomes particularly simple. The proposal probabilities are

$$\begin{aligned} Q' &= \Pr(L \rightarrow L + 1) = b_L \Pr(x_{L+1} | \mathbf{P}) \Pr(q_{L+1}) \Pr(\beta_{L+1}) \Pr(\mathbf{G}_{L+1} | \mathbf{Y}, \mathbf{G}, \mathbf{H}; \Theta_{L+1}, \Omega) , \\ Q &= \Pr(L + 1 \rightarrow L) = d_{L+1}/(L + 1) . \end{aligned} \tag{A4}$$

The ratio of true model probabilities (eq. [6]) is

$$\frac{P'}{P} = \frac{\Pr(L + 1 | \lambda)}{\Pr(L | \lambda)} \Pr(x_{L+1} | \mathbf{P}) \Pr(q_{L+1}) \Pr(\beta_{L+1}) \frac{\Pr(\mathbf{G}_{L+1} | \mathbf{H}; \Theta_{L+1}) \Pr(\mathbf{Y} | \mathbf{G}', \mathbf{Z}; \Omega', \Gamma)}{\Pr(\mathbf{Y} | \mathbf{G}; \mathbf{Z}; \Omega, \Gamma)} .$$

The Hastings ratio thus reduces to

$$R = \frac{P'Q}{PQ'} = \frac{d_{L+1}\lambda}{b_L(L + 1)} LR , \tag{A5}$$

$$\text{where } LR = \frac{\sum_{G_{L+1}} \Pr(\mathbf{G}_{L+1} | \mathbf{H}; \Theta_{L+1}) \Pr(\mathbf{Y} | \mathbf{G}', \mathbf{Z}; \Omega', \Gamma)}{\Pr(\mathbf{Y} | \mathbf{G}, \mathbf{Z}; \Omega, \Gamma)}$$

is the likelihood ratio comparing the probability $\Pr(\mathbf{Y} | \mathbf{G}; \Omega')$ under the model with $L + 1$ loci, peeling over \mathbf{G}_{L+1} to the corresponding probability under the model with only L loci. An add move is then accepted with probability $\min(1, R)$ and a delete move with probability $\min(1, R^{-1})$.

The difficulty with this proposal is that the probability of creating a new locus in a linked region is very small. We have therefore developed an alternative proposal based on standard sib-pair methods (Haseman and Elston 1972). In this approach, we consider all presently unoccupied positions and compute the likelihood for the regression of the squared differences between sib pairs in their trait residuals (after removing the effects of all the loci presently in the model) on the number of alleles they share IBD at the two flanking markers. This likelihood is then multiplied by a

negative exponential prior for the two regression coefficients (so as to penalize the positions where one or both regressions have a positive sign). Letting Q_{cm} denote the resulting proposal probability, the acceptance probability, equation (A5), is modified by multiplying it by P_{cm}/Q_{cm} , where $P_{cm} = (dX_{cm})/(X - \Sigma_{\epsilon} dX_{\epsilon})$ and dX is the length of the indicated interval.

It is also possible to refine the deletion proposal—for example, by using the predictive value of each existing locus $\Pr(Y|G \setminus G_{\ell})$ or the sib-pair likelihoods for each locus. However, we have found that after these changes are allowed for in the Hastings ratio, the resulting acceptance rates were not improved enough to justify the additional computation required. Provided that the number of loci in the model is relatively small, each locus will be considered for deletion often enough, even under simple random sampling. This contrasts with the situation for additions, in which the probability of creating a new locus in a useful position is very small and, if created in an unlinked region, is unlikely to remain in existence long enough to move to a linked region.

Global Penetrance Parameters Γ

Update of the overall means α and regression coefficients γ for the fixed covariates are straightforward, simply entailing sampling from their normal full conditional distributions, which are functions of the sums, sums of squares, and sums of cross products of the trait residuals, after all other effects are subtracted (means, covariate effects, and all genetic loci). The residual covariance matrix update also follows standard Gibbs sampling principles. For a univariate trait, this is an inverse gamma distribution (i.e., dividing the sum of squared residuals by a random χ^2 with the corresponding degrees of freedom). For a multivariate trait, an inverse Wishart distribution is used—that is, $\Sigma = (s\mathbf{X}^{-1}s)^{-1}$, where $s's = S = \sum_{ij} y_{ij}y_{ij}$ is the Cholesky decomposition of the covariance matrix of residuals, \mathbf{X} is a random Wishart matrix $X = \sum_{ij} \mathbf{x}'\mathbf{x}$, and \mathbf{x} is a vector of independent identically distributed unit normal deviates.

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

D.C.T.'s Web site, <http://hydra.usc.edu/thomas/mcmc> (for the C++ program described here)

References

- Daw EW, Heath SC, Wijnsman EM (1999a) Multipoint oligogenic analysis of age-at-onset data with applications to Alzheimer disease pedigrees. *Am J Hum Genet* 64:839–851
- Daw EW, Kumm J, Snow GL, Thompson EA, Wijnsman EM (1999b) MCMC methods for genome screening. *Genet Epidemiol Suppl* 17:S133–S138
- George AW, Mengersen KL, Davis GP (2000) A Bayesian approach to ordering gene markers. *Biometrics* 55:419–429
- Geyer CJ (1992) Practical Markov chain Monte Carlo. *Stat Sci* 7:473–482
- Green PJ (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82:711–732
- Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 2:3–19
- Heath S (1997) Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Hum Genet* 61:748–760
- Heath S, Snow G, Thompson E (1997) MCMC segregation and linkage analysis. *Genet Epidemiol* 14:1011–1016
- Hinrichs A, Lin JH, Reich T, Bierut L, Suarez B (1999) Markov chain Monte Carlo linkage analysis of a complex qualitative phenotype. *Genet Epidemiol Suppl* 17:S615–S620
- Jansen RA (1996) General Monte Carlo method for mapping multiple quantitative trait loci. *Genetics* 142:305–311
- Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* 136:1447–1455
- Kass RE, Raftery AE (1995) Bayes factor. *J Am Stat Assoc* 90:773–795
- Kruglyak L, Lander ES (1995) A nonparametric approach for mapping quantitative trait loci. *Genetics* 139:1421–1428
- Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185–199
- Lee JK, Dancik V, Waterman MS (1998) Estimation for restriction sites observed by optical mapping using reversible-jump Markov chain Monte Carlo. *J Comput Biol* 5:505–515
- Lin S, Thompson E, Wijnsman E (1993) Achieving irreducibility of the Markov chain Monte Carlo method applied to pedigree data. *IMA J Math Appl Med Biol* 10:1–17
- Ploughman LM, Boehnke M (1989) Estimating the power of a proposed linkage study for a complete genetic trait. *Am J Hum Genet* 44:543–551
- Richardson S, Green P (1997) On Bayesian analysis of mixtures with an unknown number of components (with discussion). *J R Stat Soc Ser B* 59:731–792
- Satagopan J, Yandell BS, Newton MA, Osborn TC (1996) A Bayesian approach to detect quantitative trait loci using Markov Chain Monte Carlo. *Genetics* 144:805–816
- Sillanpää MJ, Arjas E (1998) Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics* 148:1373–1388
- (1999) Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. *Genetics* 151:1605–1619

- Stephens DA, Fisch RD (1998) Bayesian analysis of quantitative trait locus data using reversible jump Markov chain Monte Carlo. *Biometrics* 54:1334–1347
- Stephens D, Smith A (1993) Bayesian inference in multipoint gene mapping. *Ann Hum Genet* 57:65–82
- Thomas DC, Richardson S, Gauderman J, Pitkaniemi J (1997) A Bayesian approach to multipoint mapping in nuclear families. *Genet Epidemiol* 14:903–908
- Uimari P, Pitkaniemi J, Onkamo P (1999) A Bayesian MCMC approach to map disease genes in simulated GAW11 data. *Genet Epidemiol* 17:S743–S748
- Uimari P, Thaller G, Hoeschele I (1996) The use of multiple markers in a Bayesian method for mapping quantitative trait loci. *Genetics* 143:1831–1842
- Yuan B, Neuman R, Duan SH, Weber JL, Kwok PY, Saccone NL, Wu JS, Liu K-Y, Schonfeld G (2000) Linkage of a gene for familial hypobetalipoproteinemia to chromosome 3p21.1–22. *Am J Hum Genet* 66:1699–1704