Am. J. Hum. Genet. 61:748–760, 1997

Markov Chain Monte Carlo Segregation and Linkage Analysis for Oligogenic Models

Simon C. Heath*

Department of Statistics, University of Washington, Seattle

Summary

A new method for segregation and linkage analysis, with pedigree data, is described. Reversible jump Markov chain Monte Carlo methods are used to implement a sampling scheme in which the Markov chain can jump between parameter subspaces corresponding to models with different numbers of quantitative-trait loci (QTL's). Joint estimation of QTL number, position, and effects is possible, avoiding the problems that can arise from misspecification of the number of QTL's in a linkage analysis. The method is illustrated by use of a data set simulated for the 9th Genetic Analysis Workshop; this data set had several oligogenic traits, generated by use of a 1,497-member pedigree. The mixing characteristics of the method appear to be good, and the method correctly recovers the simulated model from the test data set. The approach appears to have great potential both for robust linkage analysis and for the answering of more general questions regarding the genetic control of complex traits.

Introduction

Linkage analysis can be a computationally demanding problem, particularly if multipoint likelihoods, jointly involving many loci, are required. Likelihoods can be calculated, for large numbers of loci and very small pedigrees (typically <16 meioses of interest), by use of the Lander-Green algorithm (Lander and Green 1987; Kruglyak et al. 1995) or, for a few loci and larger pedigrees, by use of peeling-based algorithms (Elston and Stewart 1971; Cannings et al. 1978). For multipoint linkage analysis with many loci and larger pedigrees than the Lander-Green algorithms can handle, sampling-based approaches are required.

Markov chain Monte Carlo (MCMC) (Metropolis et al. 1953; Hastings 1970) methods have been used to calculate Monte Carlo estimates of likelihoods, for linkage mapping of problems for which exact likelihood calculation is infeasible, owing to pedigree and/or model complexity (Guo and Thompson 1992; Thompson 1994*a*, 1994*b*). Bayesian approaches to genetic analyses also have been implemented, by use of MCMC techniques (Stephens and Smith 1993; Hoeschele 1994; Heath 1995; Satagopan et al. 1996); these analyses would be impractical for all but the simplest genetic systems, without use of MCMC or some other simulation-based method, for the approximation of the required high-dimensional integrals.

A recent development in MCMC methodology is the use of samplers that allow moves to be made between different models, simplifying estimation of the relative posterior probabilities of competing models (Green 1994; Carlin and Chib 1995; Phillips and Smith 1996). The method described here uses a reversible jump MCMC sampler (Green 1994, 1995), to enable moves to be made between models with different numbers of quantitative-trait loci (QTL's). Incorrect prior specification of the number of QTL's can lead to biased estimates of QTL position and effects. Dizier et al. (1993) used simulated data to show that, when there were two QTL's present, a single QTL segregation analysis detected a single gene with parameters that did not correspond to either of the simulated genes; a similar effect was noticed by Haley and Knott (1992). The allowance of the variation of the number of QTL's allows estimation of the relative probability of models with different numbers of QTL's, as well as an assessment of how the numbers of QTL's affect the model predictions. Robust estimates of QTL parameters can be made by averaging of the predictions of models with different numbers of QTL's.

A method for QTL segregation and linkage analysis that uses reversible jump MCMC methods, to allow the number of QTL's and the linkage status of the QTL's currently in the model to vary, is described. This allows estimation of the number of segregating QTL's and of the number of QTL's that are linked to a particular chromosome or chromosome region. The map positions of the linked QTL's, the effects and frequencies of all the QTL's, and other model parameters, such as the residual variance, also can be estimated by use of this approach. The method can handle very large numbers

Received January 9, 1997; accepted for publication July 1, 1997.

Address for correspondence and reprints: Simon C. Heath, Laboratory of Statistical Genetics, Rockefeller University, Box 192, 1230 York Avenue, New York, NY 10021. E-mail: heath@linkage.rockefeller.edu

* Present affiliation: Laboratory of Statistical Genetics, The Rockefeller University, New York.

0002-9297/97/6103-0032$02.00

of highly polymorphic marker loci and, so, can take advantage of the large amounts of marker data now becoming available for linkage studies. The capabilities of the method could lead to a new approach to linkage analysis. Instead of the searching of small regions of chromosomes, for evidence of linkage for an individual QTL, a joint analysis of QTL number and position can be performed, when a large number of markers spaced throughout the genome is considered.

Single-locus peeling (Elston and Stewart 1971; Cannings et al. 1978) forms an integral part of the algorithm, both in the sampling of genotypes and in the improvement of the efficiency of the MCMC sampler. Although this restricts the types of pedigrees that can be handled, the class of problems that can be addressed by this method is broader than that of any of the current exact methods. Potential methods to relax this restriction will be discussed. The method is illustrated by use of a simulated data set that was produced for the 9th Genetic Analysis Workshop (GAW9) (MacCluer et al. 1995). This data set, which consists of 23 extended families, has a quantitative trait that is controlled by three QTL's and also is affected by a number of covariates and by a residual polygenic effect. The ability of the method to recover the simulated genetic model is investigated.

Material and Methods

Test Data Set

The test data set was generated for GAW9; the simulation model is described in detail in MacCluer et al. (1995). The pedigree consisted of 1,497 individuals from 23 extended families. Four quantitative traits (Q1, Q2, Q3, and Q4) were simulated; for this analysis, only the first trait, Q1, was analyzed. Q1 was affected directly by two major genes (MG_1 and MG_2) and indirectly by a third gene (MG_3), through Q3. MG_2 and MG_3 were diallelic; MG_1 was triallelic. Age and Q3 had linear effects on Q1; Q3 likewise was affected by a continuous environmental covariate (EF). Both Q1 and Q3 had small polygenic contributions. The contributions of MG_1, MG_2, MG_3, and the polygenes to Q1, as a percentage of the phenotypic variances, were 8%, 16%, 11%, and 3%, respectively. All pedigree members were typed for 180 highly polymorphic marker loci, with the marker loci having between two and nine alleles. These marker loci were located on six chromosomes, with each chromosome having 30 markers spaced at 2-cM intervals. MG_1, MG_2, and MG_3 were located on chromosomes 5, 1, and 2, respectively.

Model

A quantitative trait is modeled as being genetically controlled by k diallelic QTL's. The trait also can be affected by environmental covariates. The model currently does not allow for interactions among the QTL's or between the QTL's and the environmental covariates, although extension of the model, to allow such interactions, would be straightforward.

For QTL i, genotypes A_1A_1, A_1A_2, and A_2A_2 have effects a_i, d_i, and $-a_i$, respectively. The additive (a_i) and dominance (d_i) effects for QTL i are collected together in the vector $\boldsymbol{\alpha}_i \equiv \begin{pmatrix} a_i \\ d_i \end{pmatrix}$. The model for a quantitative trait $\mathbf{y}$ is therefore:

$$\boldsymbol{y} = \mu + \mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^{k} \mathbf{Q}_i\boldsymbol{\alpha}_i + \boldsymbol{e}\,, \qquad (1)$$

where μ is the overall mean, $\boldsymbol{\beta}$ is an $(m \times 1)$ vector of covariate effects (kept separate from μ, for convenience), $\boldsymbol{\alpha}_i$ is a (2×1) vector of effects for the ith QTL, $\boldsymbol{e}$ is an $(n \times 1)$ vector of normally distributed residual effects, k is the number of QTL's in the model, and $\mathbf{X}$ $(n \times m)$ and $\mathbf{Q}_i$ $(n \times 2)$ are incidence matrices for the covariate and the QTL effects, respectively. $\mathbf{Q}_i$ is derived directly from the genotypes for QTL i. For example, if the first five observations for $\boldsymbol{y}$ were of individuals whose genotypes at QTL i were (A_1A_1, A_1A_2, A_1A_1, A_2A_2, A_1A_2), then the first five rows of $\mathbf{Q}_i$ would be:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

This model easily can be extended to allow for residual polygenic effects and for additional random effects, such as shared environmental effects.

Reversible jump MCMC methods (Green 1995) are used to produce samples from the joint posterior distribution of all unknown parameters (including k). Samples of individual parameters can be regarded as being drawn from the marginal posterior distributions, and these estimated marginal distributions are used to draw inferences about parameters of interest (Tierney 1994).

The data $\mathbf{Y}$ consist of observations regarding the quantitative trait, the covariates, and the marker data. It is assumed that marker data, when present, are correct, although this restriction could be lifted to allow for the possibility of typing errors. Marker positions are assumed to be known, with markers being grouped into a number of chromosomes. Each QTL in the model has an equal prior probability of being on any chromosome or of being unlinked. Within a chromosome, each QTL has an equal probability of being anywhere on that chromosome.

The joint distribution of all variables is given by

$$p(k, \mathbf{G}, \mathbf{M}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \sigma_e^2, \mu, \mathbf{Y}) , \qquad (2)$$

where **G** and **M** are the complete genotypes (including phase) of all QTL's and markers, **δ** denotes which QTL's are currently linked, **λ** is the vector of the QTL map positions (including an indicator of which chromosome the QTL's are on) for the linked QTL's, **η** is the vector of allele frequencies for the QTL's and the markers, and σ_e^2 is the variance of the residual environmental effects *e*. Other parameters are as in equation (1). Note that the incidence matrix for the QTL effects (**Q**) can be obtained directly from **G.** Map positions were converted into recombination fractions by use of Haldane's mapping function (Haldane 1919); alternative functions could be used, but Haldane's function is the simplest function that works with multiple loci. Map distances were assumed to be the same for both sexes, although the analysis simply could be extended to allow for sex-specific maps (the same map order would be imposed for both sexes, but distances between loci would be allowed to vary).

All parameters were assigned independent priors that were mostly uniform. The exception was for **α,** which was assigned independent normal priors:

$$a_i \sim \mathrm{N}(0, \tau^2), \qquad d_i \sim \mathrm{N}(0, \tau^2) . \qquad (3)$$

Assignment of a proper prior for **α** is necessary because, during the sampling process, one or more QTL genotypes may not appear in the population. For the analyses presented here, τ^2 was set to a constant value roughly corresponding to the phenotypic variation present in the data. Marker frequencies were assigned Dirichlet (1, 1, . . .) priors: this specifies a uniform prior probability for all combinations of allele frequencies at a locus. The prior for *k* was uniform on 0, 1, 2, . . . , $k_{\max}$; $k_{\max}$ was set to 10 for all analyses reported here. Only when *k* was started at 10 did *k* ever approach $k_{\max}$. The prior for the QTL position took into account *L,* the total length of the genome, with the prior probability of any individual QTL being located in a chromosome region of length *t* being *t*/*L*.

MCMC Sampling

MCMC samplers are used to generate samples of $\boldsymbol{x} = (x_1, \ldots, x_l)$ from a joint distribution $P(\boldsymbol{x})$. A Markov chain having an equilibrium distribution of $P(\boldsymbol{x})$ is constructed, and the samples of $\boldsymbol{x}$ from the Markov chain are used to make inferences about the posterior distribution of $\boldsymbol{x}$ (Tierney 1994). In general, MCMC samplers start with an initial realization of $\boldsymbol{x}$. A move from $\boldsymbol{x}$ to a new state, $\boldsymbol{x}'$, is proposed, and an acceptance ratio, *A,* is calculated. With probability min (1, *A*), the move to $\boldsymbol{x}'$ is accepted, and, with probability 1 − min (1, *A*), the Markov chain stays at $\boldsymbol{x}$. Throughout this paper, $q(\ \ )$ is used to indicate proposal probabilities; therefore, $q(\boldsymbol{x}'; \boldsymbol{x})$ is the probability when a move to $\boldsymbol{x}'$ is proposed, when currently in state $\boldsymbol{x}$. Also, the notation $\boldsymbol{x}_{-i}$ will be used to indicate all elements of $\boldsymbol{x}$ *apart* from x_i. Introduction to the various MCMC samplers used in this paper are given in Appendix A.

Sampling Scheme

The complete sampling scheme used for the method described here has the following update steps:

1. Update complete (i.e., including phase) marker genotypes **M** for each locus in turn;
2. For each QTL *i:*
   (a) Update QTL effects $\boldsymbol{\alpha}_i$;
   (b) Update QTL position λ_i and linkage status δ_i;
   (c) Update QTL genotypes $\mathbf{G}_i$;
3. Update QTL and marker frequencies **η;**
4. Update covariate effects **β** and overall mean μ;
5. Update residual variance σ_e^2;
6. Birth or death of a QTL; and
7. Split one QTL into two; combine two QTL's into one.

A sampling iteration is a complete pass through this scheme.

The parameters (μ, **β,** σ_e^2, **η**) are updated by use of Gibbs steps, that is, by the sampling of each parameter, in turn, from its full conditional distribution. This has been described in several papers (e.g., Wang et al. 1993; Heath 1994) and, so, will not be discussed further here.

The genotypes for all loci (markers and QTL's) also are updated by use of Gibbs steps, with the genotypes at a given locus being updated *simultaneously* for all individuals (although only one locus at a time); this sampling scheme was suggested by Kong (1991). This differs from the MCMC schemes typically used in genetics, for which the genotypes at a given locus are updated on an individual-by-individual basis (e.g., Guo and Thompson 1992; Heath 1994). The genotype sampling method, in this article called "reverse peeling," uses a modification of the peeling algorithm, to calculate the required genotype sampling distributions (Ott 1989). The pedigree, therefore, is required to be peelable, although only for each locus separately. This scheme is computationally more complex than an individual-by-individual updating scheme but has the benefits of greatly improved mixing, of not requiring an initial genotype configuration, and of avoiding irreducibility problems when dealing with multiallelic loci (Sheehan and Thomas 1993; Lin et al. 1993, 1994).

Updates of map position and linkage status for QTL *i* (λ_i, δ_i) are made unconditionally on the current genotypes for QTL *i* (G_i). This is done by use of peeling, to integrate out $\mathbf{G}_i$ from the MCMC acceptance ratio. The use of this approach allows large moves (i.e., be-

tween marker intervals or between different chromosomes) to be made with reasonable frequency. Since these moves use partial conditioning on a subset of the model parameters (Besag et al. 1995), they must be followed by a Gibbs update of the QTL genotypes, as discussed in Appendix A. The updating of QTL linkage status changes model dimension (a linked QTL has a parameter describing its location, whereas an unlinked QTL does not) and, so, uses a reversible jump step (Green 1995).

Reversible jump MCMC steps also are used to change the number of QTL's in the model. Two pairs of reversible jump steps are utilized: birth/death steps and split/combine steps (Richardson and Green 1997). With a birth step, a new QTL is proposed independently of existing QTL's in the model. A death step is the reverse process, whereby an existing QTL is selected at random and is removed from the model. With a split step, an existing QTL is selected, and its effect distributed between two QTL's. For a combine step, therefore, two QTL's are selected, and their effects combined to form a single QTL.

Moves that increase the number of QTL's require the generation of parameters for the so-called new QTL's. The efficiency of the proposed move depends, to a large extent, on how "good" the proposed parameters for the new QTL's are. The sampling of the parameters independently can result in parameters that jointly are highly unlikely. An alternative approach is used here; the variance contributed by the new QTL's is sampled, and this variance is transformed to yield the QTL effect. The additive (σ_a^2) and dominance (σ_d^2) variance contributed by a QTL can be estimated (when Hardy-Weinberg equilibrium is assumed) by

$$\begin{aligned} \sigma_a^2 &= 2\eta(1-\eta)[a + d(1-2\eta)]^2 , \\ \sigma_d^2 &= [2\eta(1-\eta)d]^2 , \end{aligned} \tag{4}$$

where a and d are the additive and dominance effects of the QTL, as defined in the Model section, and η is the frequency of allele A_1 (Falconer 1989). Exponential distributions are used as proposal probabilities for the variances, whereas the allele frequency, as stated before, is sampled from its prior. The mean for the exponential distribution is set to some fixed fraction c of the current value of σ_e^2, the residual variance. This is because the new QTL's should account for some of the residual variance, so it seems sensible to propose a QTL that produces some fraction of σ_e^2. For all results presented here, c was set to $^1/_3$; the variation of c around this value appeared to have little effect, although it could be worthwhile to investigate further the effect of c on sampling efficiency.

The sampled values of σ_a^2 and σ_d^2, are transformed by use of the inverses of equation (4), to yield values for a and d. This is not a one-to-one mapping; for each combination of σ_a^2, σ_d^2, and η, there are four possible combinations of a and d, one of which is selected at random.

The calculation of the acceptance probabilities for the reversible jump steps (changing linkage status, the birth/death step, and the split/combine step) are given in Appendix B.

Segregation and Linkage Analysis

Two models were used for the analysis: for model 1, age and EF were fitted as covariates, and, for model 2, age and Q3 were fitted as covariates. EF was not fitted in model 2 because it only has an effect on Q1 through Q3. Model 1 should allow all three major genes to be detected, whereas for model 2 only MG_1 and MG_2 should be detectable, because MG_3 affects Q1 only through Q3. Although it was known that MG_1 had three alleles, only diallelic QTL's were fitted.

Initially, a segregation analysis, fitting none of the markers, was performed, to get estimates of the numbers and effects of any segregating genes. Inspection of the output from trial runs indicated that the sampler appeared to reach convergence after, at most, 200 iterations. The first 200 iterations, therefore, were discarded from all runs, after which all samples were used for estimation. For all analyses presented here, 20,000 additional sampling iterations were performed after the initial 200.

The segregation analysis fitting model 1 was performed with starting values of 0 and 10 for k; this was to check for an effect of the starting value, on the estimated posterior distribution of k. The segregation analysis with model 2, and with all subsequent analyses, used a starting value of 0 for k.

The segregation analysis was followed by a genome scan; each chromosome was fitted individually (i.e., all markers on a given chromosome were fitted) for both models, and the probability of linkage to that chromosome was estimated. Further analyses simultaneously fitting multiple chromosomes then were performed, to determine whether this had an effect on the results.

Results

Segregation Analyses

The estimated posterior distributions of k, from the segregation analyses, are given in table 1. It can be seen that changing the starting value for k had little effect on $\hat{p}(k)$. For the analysis starting with $k = 10$, after 40 iterations k had dropped to ~2 and subsequently behaved the same as for the analysis started from $k = 0$.

The estimated posterior distribution for k, from model 2, appears to have shifted to the left by 1, when

Table 1

Estimate of the Posterior Distribution of the QTL Number, from Segregation Analyses

		Estimated Distribution, for k =					
Model	k_0[a]	0	1	2	3	4	5
1	0	.00	.00	.71	.25	.04	<.01
1	10	.00	.00	.75	.21	.04	<.01
2	0	.00	.75	.22	.02	<.01	.00

[a] Starting value for k.

compared with the estimate from model 1, having a mode at $k = 2$ and $k = 1$ for models 1 and 2, respectively. This difference is expected because, when a correction is made for Q3 in model 2, the effect of MG_3 should be removed. The estimated frequencies and additive effects of the QTL's from the two analyses are shown in figure 1. From the figure, it appears that a QTL with frequency of ~.2 and an additive effect of ~13 was present in the analyses using both models. An additional QTL, with frequency of .4 and an effect of ~5, was present only in the model 1 analysis. The simulated frequency and additive effect (on Q1) of MG_2 were .2 and 15, respectively, and for MG_3 were .49 and 7.6, respectively. This would indicate that the method is picking up MG_2 and MG_3, by use of model 1, and only MG_2, by use of model 2. MG_1 apparently was not being detected by use of either model.

Single-Chromosome Linkage Analyses

The posterior probabilities for linkage, from the genome scans, are shown in table 2. The values shown are the probabilities that at least one QTL was linked to the chromosome being tested, when results from models with different numbers of QTL's were averaged. A more detailed picture is given by figures 2 and 3, which show the estimated log probability of linkage, as a function of position along each chromosome, for models 1 and 2, respectively. Arrows indicate the simulated positions of the QTL's on chromosomes 1, 2, and 5. Under model 1, chromosomes 1 and 2 showed strong support for linkage. The most likely position for a QTL on chromosome 1 is shown, in figure 2, to be close to the simulated location of MG_2. There are two likely positions for a QTL on chromosome 2, one of which closely corresponds to the simulated position of MG_3. Under model 2, there was again strong support for linkage around the location of MG_2, but the support for linkage to chromosome 2 had disappeared. The correction for Q3,

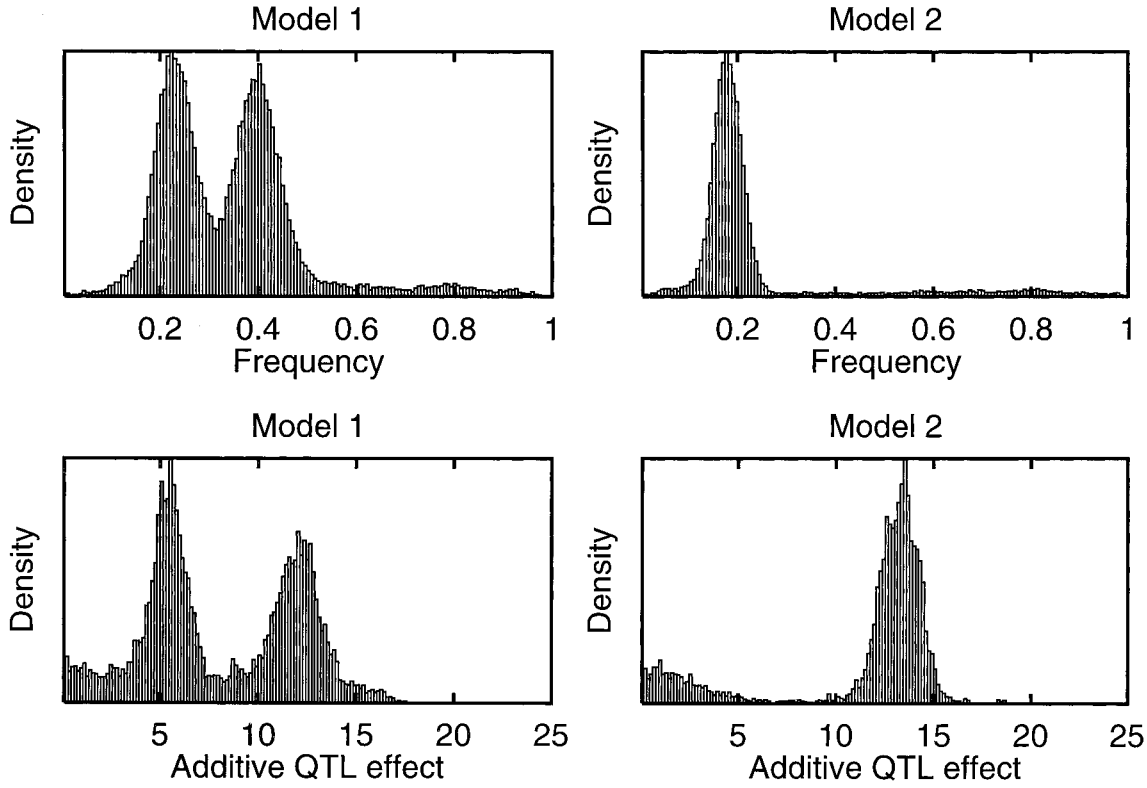

Figure 1 Estimates of the posterior densities of QTL frequency and additive effect, for the segregation analyses under model 1 and model 2.

Table 2

Estimates of the Posterior Distribution of the QTL Number and of the Probabilities of Linkage, by the Fitting of Each Chromosome Separately

Model and Chromosome	Estimated Distribution, for k =						
	0	1	2	3	4	>4	p
1:							
1	.00	.00	.38	.47	.12	.03	.995
2	.00	.00	.28	.55	.15	.02	.970
3	.00	.00	.70	.24	.05	.01	.042
4	.00	.00	.72	.23	.05	<.01	.061
5	.00	.00	.72	.24	.04	<.01	.067
6	.00	.00	.72	.23	.04	.01	.046
2:							
1	.00	.10	.60	.25	.04	.01	.999
2	.00	.79	.19	.02	<.01	.00	.026
3	.00	.78	.20	.02	<.01	.00	.016
4	.00	.80	.18	.02	<.01	.00	.026
5	.00	.70	.26	.04	<.01	.00	.065
6	.00	.74	.23	.03	<.01	.00	.038

therefore, prevented the detection of MG_3 on chromosome 2, as it should. For both models, there was no strong support for linkage to chromosome 5, although figure 3 does show a peak in the log probability of linkage at the simulated location of MG_1.

Table 2 also shows the estimated posterior distribution of k. As with the segregation analyses, $\hat{p}(k)$ from model 2 shifted to the left by 1, with respect to the model 1 estimate, and, when a linked QTL was detected, there was a shift of $\hat{p}(k)$ to the right.

Multiple-Chromosome Linkage Analyses

To investigate the effect of fitting multiple chromosomes, the model 2 analyses were repeated by the fitting of pairs of chromosomes, with chromosome 1 being paired in turn with each of the other chromosomes. The estimated probability of linkage along each chromosome is shown in figure 4; the chromosome 1 plot is from the analysis fitting chromosomes 1 and 5 jointly, but there was little difference (in the results for chromosome 1) when the other chromosomes were analyzed. The results show strong support for linkage to chromosome 1 (p = 1.000) and to chromosome 5 (p = .937) but not

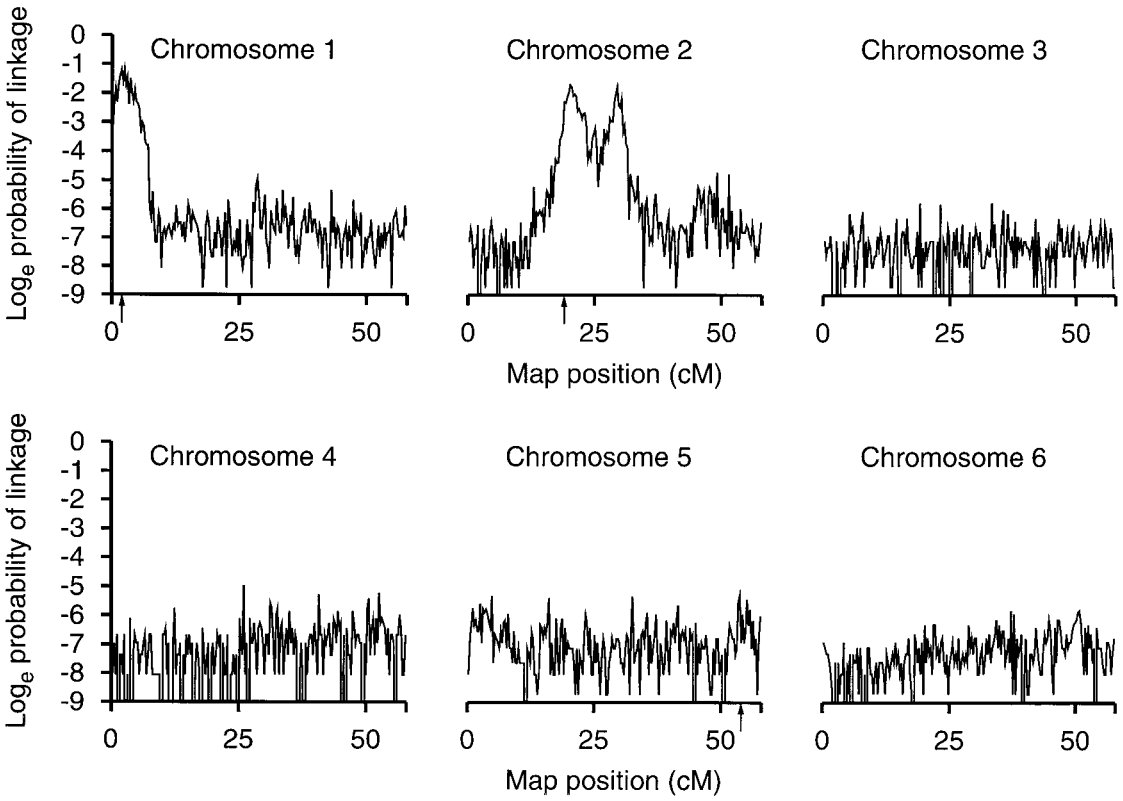

Figure 2 Estimates of the log posterior probability, when model 1 is fitted for at least one QTL being linked to a given chromosomal region, for all six chromosomes. The positions of the simulated QTL's are indicated by arrows.

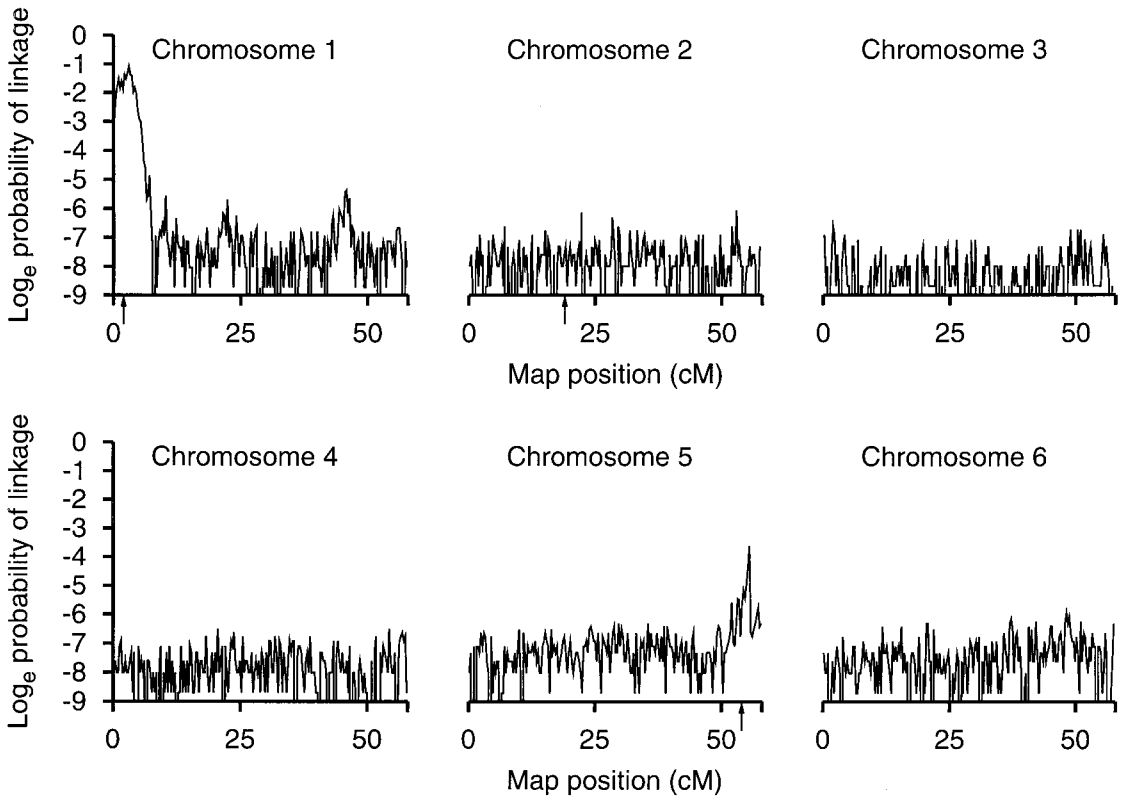

Figure 3 Estimates of the log posterior probability, when Model 2 is fitted for at least one QTL being linked to a given chromosomal region, for all six chromosomes. The positions of the simulated QTL's are indicated by arrows.

to any of the other chromosomes. The most likely positions for the linked chromosomes coincided with the simulated locations of MG_1 and MG_2. As with the single-chromosome analyses, MG_3 was (correctly) not being detected.

In the final analysis performed, chromosomes 1, 2, and 5 were fitted jointly under model 1. This combination was fitted because these were the three chromosomes known to contain the simulated QTL's. This analysis involved the fitting of 90 markers simultaneously and was computationally expensive, and, for this reason, other combinations of three chromosomes were not tried. Support for linkage to chromosome 1 ($p = .989$) and to chromosome 2 ($p = .972$) is still shown, but there was still little support for linkage to chromosome 5 ($p = .227$), although more so than in the single-chromosome analysis. Plots of the probability of linkage along the chromosomes (fig. 5, *top*) show a similar story as that shown for the single-chromosome analyses, although there is now a slight peak at the location of MG_1 on chromosome 5. The plot for chromosome 2 still shows two peaks, as it did in the single-chromosome analysis.

Figure 5 (*bottom*) also shows a trace from the same analysis, showing both the cumulative probabilities for the number of QTL's in the model, at each sample iteration, and the estimated posterior density of the QTL number. The cumulative probability plot appears to have leveled off about halfway through the experiment; this indicates that the model was mixing well between models with different numbers of QTL's. Also shown in figure 5 is the estimated density of the QTL number from the analysis. The mode is at the correct value of three, although models with two or four QTL's cannot be ruled out.

Discussion

The advent of MCMC has made many changes in the types of linkage analyses that are possible. The development of reversible jump MCMC could have just as big an impact, changing the way in which linkage analysis is performed. Instead of looking for a single QTL by use of a small number of markers, a very large number of marker loci can be considered together in a joint analysis of QTL number and position. The methodology could be extended to allow more-flexible models, with reversible jump MCMC being used to add or to remove epistatic interactions between QTL's or to change the number of alleles for a QTL. Another feature of this

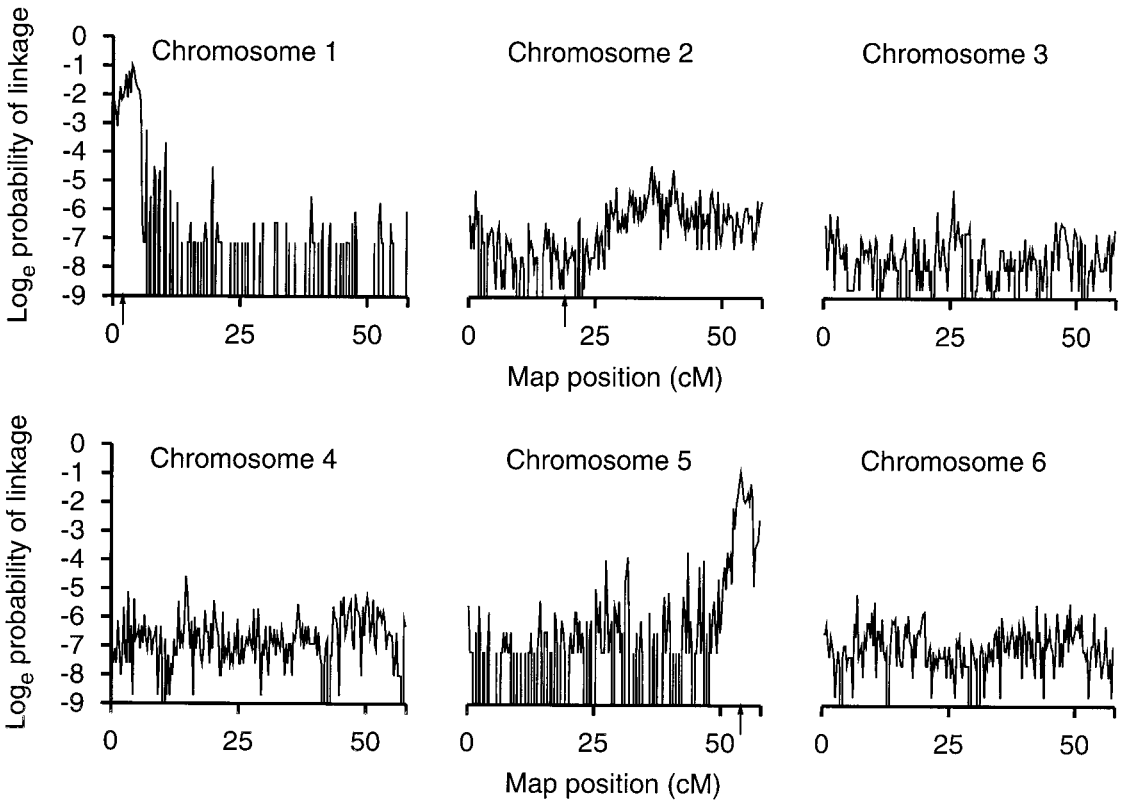

Figure 4 Estimates of the log posterior probability, when Model 2 is fitted for at least one QTL being linked to a given chromosomal region. Pairs of chromosomes are fitted simultaneously, with chromosome 1 being paired, in turn, with each of the remaining five chromosomes. The positions of the simulated QTL's are indicated by arrows.

type of analysis is the natural modeling of genetic heterogeneity; the genetic model for all families is not forced to be the same, so the disease could be caused by different loci in different families.

When applied to the GAW9 data set, the method detected and estimated the effects and positions of the two larger loci, MG_2 and MG_3, without difficulty. The other locus, MG_1, only could be detected with confidence when correction for Q3 was made and chromosomes 1 and 5 were fitted simultaneously. This could be because of the relatively small effect of MG_1 or because MG_1 had three alleles. When MG_1 and MG_2 were detected, the most likely positions indicated by the analysis were centered on the simulated locations of the loci. There appeared to be two, almost equally likely, locations for a QTL on chromosome 2, one of which corresponded to the simulated location of MG_3. During the analyses of chromosome 2, for the majority of sampling iterations, there was only one QTL linked to the chromosome. The bimodal plots for chromosome 2, therefore, indicated two possible locations for a single QTL, rather than the presence of two QTL's on the chromosome. Note that in a conventional interval mapping approach it would not be possible to distinguish between these two possibilities.

Further work on this and other problems is necessary, to investigate the power and the limitations of this approach. It should be noted that none of the penetrance parameters or the gene frequencies were assumed to be known, for this analysis; instead, the MCMC sampler was used to integrate over all of these so-called nuisance parameters. The results were not dependent, therefore, on a point estimate of the genetic model but accounted for the uncertainty about the model.

The estimated probabilities of linkage, because they take into account the length of the entire genome, can be interpreted as genomewide probabilities. Results have been given for the probabilities of linkage to entire chromosomes, but probabilities for linkage to smaller regions also could be found easily. For example, the probability of linkage of one or more QTL's to the region between the first and fourth markers on chromosome 1 (a distance of 6 cM) is $\sim$.94. The prior probability for the QTL location that was used here assumes that each QTL in the model has an individual uniform probability of being located anywhere in the entire genome. If information about the distribution of coding regions along the genome were available, then this could be factored easily into the analysis.

By comparing the segregation, the single-chromosome

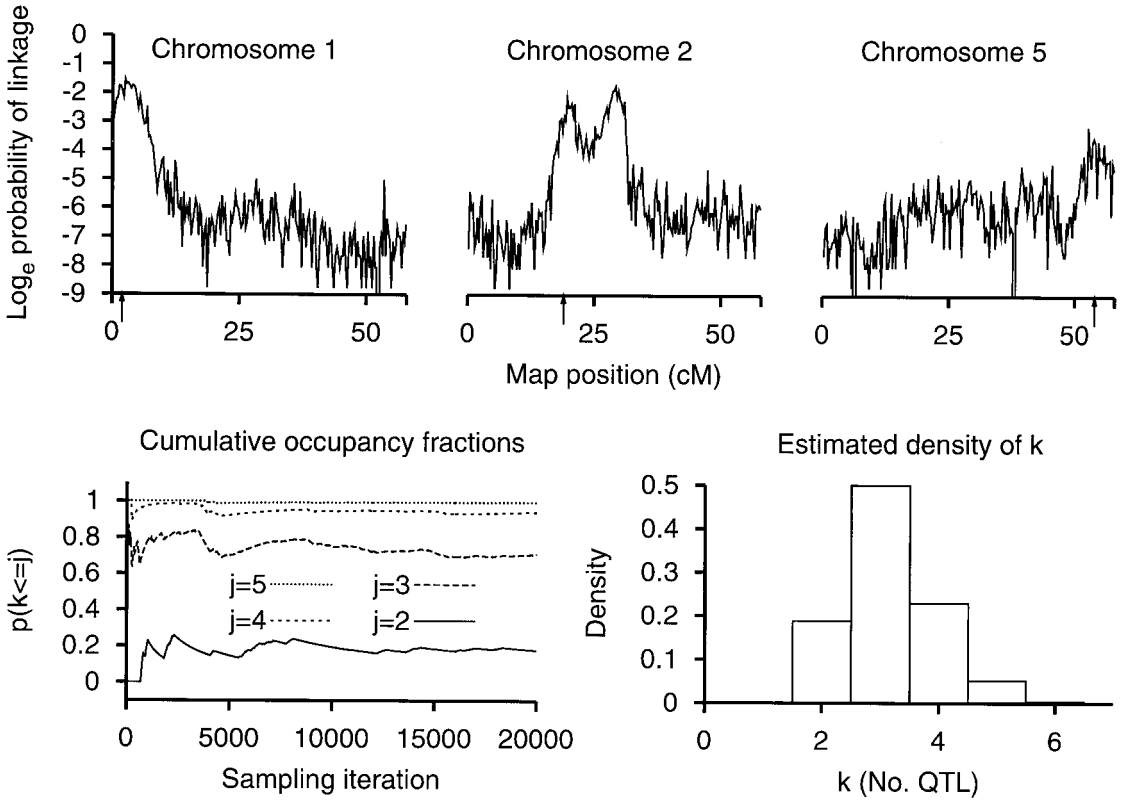

Figure 5 *Top,* Estimates of the log posterior probability, when Model 1 is fitted for at least one QTL being linked to a given chromosomal region. The three displayed chromosomes (1, 2, and 5) were fitted simultaneously. The positions of the simulated QTL's are indicated by arrows. *Bottom,* Cumulative probabilities for k, against sample iteration and the estimated posterior density of the QTL number, from the same analysis as described for the *top panel.*

analysis, and the multiple-chromosome analysis, we can see a general trend toward increasing k when there is more marker data. This is to be expected: without marker data there can be little information that distinguishes between several small QTL's or a few larger QTL's. As more marker data becomes available, the resolving power of the analysis should increase, allowing the detection of QTL's of smaller effect. The single-chromosome analyses (table 2) also show that, when there was support for a linked QTL, this also tended to shift the distribution of k to the right. This indicates that, after the fitting of a single QTL to a chromosome, there still was support for additional segregating QTL's.

It is interesting to compare the performance of the method described here, which was used on the GAW9 data set, with that of the methods presented at the GAW9 meeting. More details on these methods are given in the summary paper for GAW9 (Blangero 1995). The current method performs well, when compared with the other methods used to analyze trait Q1. The method provided strong support for linkage to the correct chromosome regions, for all three QTL's affecting Q1, with no false positives. This performance is better than that achieved by any of the methods presented at GAW9. This is not too surprising, since only a few of the methods involved any form of multipoint linkage analysis.

The most computationally demanding analysis described in this paper fitted three chromosomes (90 markers) simultaneously and required ~10 Mb of memory and 2 d of computing time on a Digital Alphastation 400. The single-chromosome analyses took approximately one-third of the time and memory used for the multiple-chromosome analysis. Although these memory and time requirements, therefore, are greater than those for the majority of analyses presented at GAW9, the extra information that can be obtained by the performance of multipoint analyses on these large data sets would seem to justify the use of the method. It should be noted that a comparable exact linkage analysis (the simultaneous fitting of all markers to even a single chromosome) would be infeasible with pedigrees of the size found in the GAW9 data set.

Results obtained from variation of the starting value for k and from the plot of the cumulative occupancy fractions (fig. 5) indicate that the sampler appears to mix well without excessive numbers of iterations being required. This is important because the computational costs of each iteration for this analysis are quite high;

so, there is a requirement to keep the iteration number as low as possible. The high costs per iteration stem both from the inherent complexity of the model, which simultaneously models up to almost 100 discrete loci, and from the use of peeling to improve mixing. Some of the uses of peeling—for example, the generation of genotypes for new QTL's—are unavoidable. Other uses (e.g., the updating of QTL location) could be avoided, but earlier work suggests that, without the use of peeling, models with tightly linked loci would fail to mix at all. The use of peeling, therefore, can be justified as an important part of the implementation of a reversible jump sampler, as well as being useful in getting the more conventional part of the sampling algorithm to mix effectively.

The method described here is computationally more costly than a simple MCMC scheme, but the advantages in functionality and mixing are great. The reliance on peeling means that pedigrees must be single-locus peelable; even with this restriction, the method can address a much wider range of problems than can be handled by existing exact methods such as multilocus peeling or Lander-Green–based algorithms. Note also that the capabilities of the approach used here (e.g., the ability to estimate the number of QTL's) exceed those of existing methods. There are, however, reasons for the need for a method that could handle arbitrarily complex pedigrees. Such pedigrees can arise in isolated human populations or in animal populations and potentially could allow inferences to be made, about the genetic control of complex traits, that could not be answered by use of simpler pedigrees.

A potential approach to the extension of the method to general pedigrees would be to use an approximate reverse peeling method for the genotype sampling, with the error from the approximation corrected by use of a Metropolis-Hastings acceptance/rejection step. To be able to use an approximate sampling procedure, in an MCMC sampling scheme, the sampling distribution must be known. The approximate peeling method of Thomas (1986) appears amenable to extension, resulting in a sampling method that would satisfy this criterion. Further work in this area clearly is warranted.

In conclusion, the application of reversible jump MCMC to linkage analysis allows the fitting of highly flexible models, for which the details of the model can be altered by the sampling procedure. The methodology allows both the robust estimation of QTL effects and the answering of questions about the distribution of QTL numbers affecting a trait, which previously would have been extremely difficult to do. The method described here appears to work well when tested against a complex simulated data set. Many improvements to the method no doubt could be made, in terms of mixing, computational efficiency, and functionality; there also remains much work to be done in the testing of the limitations of the method and in the extension of it to general pedigrees. As it stands, however, the method already has functionality beyond what is currently available and gives an indication of what is possible by use of this approach.

Acknowledgments

The author is grateful to Elizabeth Thompson for her helpful discussion and comments, to Charlie Geyer for his discussions on MCMC, and to the referees for their useful comments on an earlier version of this paper. This work was supported by NIH grant GM46255. The simulated data used in this analysis were generated for GAW9 with the support of NIH grant GM31575.

Appendix A

MCMC Sampling

Metropolis-Hastings Sampler

The Metropolis-Hastings sampler (Hastings 1970) has an acceptance ratio for a move from state $\boldsymbol{x}$ to $\boldsymbol{x}'$ of:

$$A = \frac{p(\boldsymbol{x}')q(\boldsymbol{x};\, \boldsymbol{x}')}{p(\boldsymbol{x})q(\boldsymbol{x}';\, \boldsymbol{x})}. \tag{A1}$$

This is just the product of the probability ratio $p(\boldsymbol{x}')/p(\boldsymbol{x})$ and the ratio of the probability of the proposition of the *reverse move* from $\boldsymbol{x}'$ to $\boldsymbol{x}$ against the probability of the proposition of the *forward move* from $\boldsymbol{x}$ to $\boldsymbol{x}'$. Note that $p(\boldsymbol{x})$ only needs to be known up to a multiplicative constant. Proposed moves to $\boldsymbol{x}$ can change either single elements or groups of elements.

Gibbs Sampler

A special case of the Metropolis-Hastings sampler is the Gibbs sampler (Geman and Geman 1984). With this sampler, changes typically (but not necessarily) are made to one element of $\boldsymbol{x}$ at a time. When x_i is updated, the new value for x_i is sampled from the conditional distribution $p(x_i|\boldsymbol{x}_{-i})$. In this case, the acceptance probability is always 1, as is shown below for a proposed change from $\{x_i, \boldsymbol{x}_{-i}\}$ to $\{x_i', \boldsymbol{x}_{-i}\}$:

$$\begin{aligned} A &= \frac{p(x_i', \boldsymbol{x}_{-i})p(x_i|\boldsymbol{x}_{-i})}{p(x_i, \boldsymbol{x}_{-i})p(x_i'|\boldsymbol{x}_{-i})} \\ &= \left[\frac{p(x_i', \boldsymbol{x}_{-i})}{p(x_i, \boldsymbol{x}_{-i})}\right]\left[\frac{p(x_i, \boldsymbol{x}_{-i})}{p(\boldsymbol{x}_{-i})}\right]\left[\frac{p(\boldsymbol{x}_{-i})}{p(x_i', \boldsymbol{x}_{-i})}\right] = 1. \end{aligned} \tag{A2}$$

A sampling scheme does not have to consist of all Metropolis-Hastings steps or of all Gibbs steps but, instead,

can be a mixture, with some elements of $\boldsymbol{x}$ being updated with Gibbs steps and others with Metropolis-Hastings steps.

Partial Conditioning

The example above shows x_i being updated by use of its *full conditional* distribution, that is, its distribution conditional on all the other elements of $\boldsymbol{x}$. This is not necessary, and, with certain restrictions, updates can be made by sampling from *reduced* conditionals, conditioning on only a subset of $\boldsymbol{x}_{-i}$ (Besag et al. 1995). The same applies to general Metropolis-Hastings update steps; updates to x_i can be made with respect to a subset of $\boldsymbol{x}_{-i}$, by integration of the unused variables out of the acceptance ratio. This can improve the efficiency of the sampler and is used for some of the update steps described in this paper. An important restriction is that if an update step is made unconditional on x_j, then only $\boldsymbol{x}_{-j}$ can be guaranteed to have the desired joint distribution; in effect, the current value of x_j is invalid. To ensure the correct equilibrium distribution for *all* elements of $\boldsymbol{x}$, after such a step, x_j must be updated by use of a Gibbs update (Besag et al. 1995). This works because a Gibbs update for x_j does not depend on the current, invalid value for x_j.

Reversible Jump MCMC

The sampling schemes outlined above require the length of $\boldsymbol{x}$ to be fixed. Reversible jump MCMC (Green 1995; Richardson and Green 1997) is an extension to the Metropolis-Hastings sampler, permitting moves to be made that change the dimension of $\boldsymbol{x}$. The sampler then can move between models of different dimension, allowing the sampler to select between, or to average over, alternative models. The acceptance probabilities for reversible jump steps are calculated in a way analogous to those for Metropolis-Hastings update steps, the difference being that the proposals must now take account of the change in dimension. For example, consider a move from $\boldsymbol{x}$ to $\boldsymbol{x}'$, where $\boldsymbol{x}$ has dimension l_0 and where $\boldsymbol{x}'$ has dimension l_1, with $l_1 > l_0$. To make up the difference in length between $\boldsymbol{x}$ and $\boldsymbol{x}'$, a random vector $\boldsymbol{u}$, of length $l_1 - l_0$, is sampled and then is transformed to yield the extra elements of $\boldsymbol{x}'$. When the reverse step is made, the extra elements are simply discarded. The acceptance ratio for this step is given by

$$\frac{p(\boldsymbol{x}')q(l_0;\, l_1)}{p(\boldsymbol{x})q(l_1;\, l_0)q(\boldsymbol{u})}\left|\frac{\partial \boldsymbol{x}'}{\partial(\boldsymbol{x},\, \boldsymbol{u})}\right|, \tag{A3}$$

where $q(l_1;\, l_0)$ is the probability of the proposition of the move as described before, $q(u)$ is the proposal probability for the $\boldsymbol{u}$, and the last term is the Jacobian of the transformation from $(\boldsymbol{x},\, \boldsymbol{u})$ to $\boldsymbol{x}'$.

Appendix B

Acceptance Probabilities

Changing QTL Linkage Status and Position

If a QTL is currently linked and no change in δ_i is proposed, the update simply changes the QTL position (λ_i). This is therefore a standard Metropolis-Hastings step, and the acceptance probability for the change from λ_i to λ_i' will be $\min(1, A)$, where

$$A = \frac{p(\boldsymbol{Y}|k, \boldsymbol{G}_{-i}, \boldsymbol{M}, \boldsymbol{\beta}, \lambda_i', \boldsymbol{\lambda}_{-i}, \boldsymbol{\delta}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \sigma_e^2, \mu)\, p(\lambda_i')q(\lambda_i;\, \lambda_i')}{p(\boldsymbol{Y}|k, \boldsymbol{G}_{-i}, \boldsymbol{M}, \boldsymbol{\beta}, \lambda_i, \boldsymbol{\lambda}_{-i}, \boldsymbol{\delta}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \sigma_e^2, \mu)p(\lambda_i)q(\lambda_i';\, \lambda_i)}. \tag{B1}$$

Equation (B1) is the product of the likelihood ratio of QTL i being at position λ_i' versus position λ_i (when the genotype of QTL i is integrated out by use of peeling), the prior ratio for the two positions, and the proposal ratio.

The acceptance probability for a move that changes δ_i uses a reversible jump step. Despite this, it is very similar to equation (B1). When the QTL's are moved from unlinked to linked states, a map position for the QTL's must be proposed, and, when the reverse move is made, the map position is simply discarded. Therefore, a move from an unlinked to a linked state would have an acceptance probability of $\min(1, A)$, where

$$A = \frac{p(\boldsymbol{Y}|k, \boldsymbol{G}_{-i}, \boldsymbol{M}, \boldsymbol{\beta}, \lambda_i', \boldsymbol{\lambda}_{-i}, \delta_i', \boldsymbol{\delta}_{-i}\, \boldsymbol{\eta}, \boldsymbol{\alpha}, \sigma_e^2, \mu)}{p(\boldsymbol{Y}|k, \boldsymbol{G}_{-i}, \boldsymbol{M}, \boldsymbol{\beta}, \boldsymbol{\lambda}_{-i}, \delta_i, \boldsymbol{\delta}_{-i}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \sigma_e^2, \mu)} \times \frac{p(\delta_i')}{p(\delta_i)}\, \frac{p(\lambda_i')q(\delta_i;\, \delta_i')}{q(\lambda_i', \delta_i';\, \delta_i)}. \tag{B2}$$

Birth/Death Steps

A birth step requires generation of the QTL effects, frequency, linkage status, map position if linked, and genotypes for all pedigree members, for the new QTL's. None of the existing QTL's are affected. With a death step, the parameters of the selected QTL's are simply discarded. As with the location updates, peeling is used so that the genotypes for the selected QTL's do not enter into the acceptance probability for the move. If a birth step is successful, genotypes for the new QTL's are sampled by use of reverse peeling. The effects for the new QTL's are generated both by the sampling of the variances contributed by the QTL's and by transformation to yield the effects. The estimated effect of the new QTL's on μ is used to propose a new value, μ', for the mean. The acceptance probability for a birth [death] step is therefore $\min(1, A)$ [$\min(1, 1/A)$], where:

$$A = \frac{p(\mathbf{Y}|k+1, \mathbf{G}_{-i}, \mathbf{M}, \boldsymbol{\beta}, \lambda_i', \boldsymbol{\lambda}_{-i}, \delta_i', \boldsymbol{\delta}_{-i}, \eta_i', \boldsymbol{\eta}_{-i}, \boldsymbol{\alpha}_i', \boldsymbol{\alpha}_{-i}, \sigma_e^2, \mu')}{p(\mathbf{Y}|k, \mathbf{G}, \mathbf{M}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \sigma_e^2, \mu)} \times \left\{\frac{p(\boldsymbol{\alpha}_i')}{[1/(c\sigma_e^2)]^2 e^{-(\sigma_a^2+\sigma_d^2)/(c\sigma_e^2)}(1/4)}\right\} \times \left[\frac{\sqrt{2}}{16\sqrt{\sigma_a^2\sigma_d^2(\eta_i')^3(1-\eta_i')^3}}\right] \times \left[\frac{p(k+1)}{(k+1)p(k)}\right]\left[\frac{q(\text{death}; k+1)}{q(\text{birth}; k)}\right]. \quad \text{(B2)}$$

Note that i refers to the new QTL's in a birth step or to the QTL's to be removed in a death step. The first line is the ratio of the probability of the model with the new QTL's against the current model; note that the genotypes for the new QTL's have been integrated out of the numerator. The second line is the ratio of the priors for the new QTL effect to the proposal probabilities for σ_a^2 and σ_d^2, and the third line is the Jacobian of the transformation from $(\sigma_a^2, \sigma_d^2, \mu)$ to (a_i', d_i', μ'). In the proposal probability expression, $c\sigma_e^2$ is the mean of the exponential distribution, where c is a constant between 0 and 1. The new parameters $(\lambda_i', \delta_i', \eta_i')$ were sampled from their priors, so the expressions for these cancel out. The factor $(1/4)$ in the denominator arises because the transformation from variances to QTL effects is not a one-to-one mapping, and one of the four possible combinations of QTL effects is picked at random. The last line is the product of the prior ratio for k and the proposal ratio for a death step versus a birth step. The $(k+1)$ factor is in the denominator of the last line because, in order to reverse a birth step, the same QTL must be selected in a death step.

Split/Combine Steps

The split/combine steps are more complicated than the birth/death steps, because they involve the changing of two QTL's. For a split step, one of the current QTL's is picked at random. The additive and dominance variances produced by that QTL are estimated by use of equation (4). Two variables (u_a, u_d) then are sampled from independent U(0:1) distributions and are used to partition the variances between the original QTL (i) and a new QTL (j). An allele frequency, linkage status, and map position then are proposed for the new QTL, and the effects for both QTL's are derived from the partitioned variances, as described for the birth step. New genotypes for QTL i are sampled conditional on the new effects, and then the proposed addition of QTL j is handled in a fashion similar to a normal birth step but conditional on the new genotypes for QTL i. If the change is accepted, then genotypes for QTL j will be generated by use of the reverse-peeling algorithm; otherwise, the original genotypes and effects for QTL i will be restored.

The combine step is the reverse of this process; two QTL's are selected, with the order of selection being noted. The second QTL (j) is discarded, and the first QTL (i) is given an effect so that its estimated variance contribution is the same as that of both the original QTL's.

The acceptance ratio for a split move (and, analogously, for a combine move) is then $\min(1, A)$, where

$$A = \frac{p(\mathbf{Y}|k, \mathbf{G}_{-i}, \mathbf{M}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\eta}, \boldsymbol{\alpha}_i', \boldsymbol{\alpha}_{-i}, \sigma_e^2)p(\boldsymbol{\alpha}_i')}{p(\mathbf{Y}|k, \mathbf{G}_{-i}, \mathbf{M}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\eta}, \boldsymbol{\alpha}_i, \boldsymbol{\alpha}_{-i}, \sigma_e^2)p(\boldsymbol{\alpha}_i)} \times \left[\frac{p(\mathbf{Y}|k+1, \mathbf{G}_i', \mathbf{G}_{-i}, \mathbf{M}, \boldsymbol{\beta}, \lambda_j', \boldsymbol{\lambda}_{-j}, \delta_j', \boldsymbol{\delta}_{-j}, \eta_j', \boldsymbol{\eta}_{-j}, \boldsymbol{\alpha}_i', \boldsymbol{\alpha}_j', \boldsymbol{\alpha}_{-i,j}, \sigma_e^2)}{p(\mathbf{Y}|k, \mathbf{G}_i', \mathbf{G}_{-i}, \mathbf{M}, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\eta}, \boldsymbol{\alpha}_i', \boldsymbol{\alpha}_{-i}, \sigma_e^2)}\right] \times \left[\frac{p(k+1)}{(k+1)p(k)}\frac{q(\text{combine}; k+1)}{q(\text{split}; k)}\right] \times \left[\frac{p(\boldsymbol{\alpha}_j')p(\lambda_j')p(\delta_j')p(\eta_j')}{q(u_a)q(u_d)q(\lambda_j')q(\delta_j')q(\eta_j')(1/4)}\right] \times \left|\frac{d(a+d-2d\eta_i)}{4} \times \frac{\sqrt{(1-\eta_i)^3\eta_i^3}}{\sqrt{(1-\eta_j')^3(\eta_j')^3 u_a u_d(1-u_a)(1-u_d)}}\right| \quad \text{(B3)}$$

and where (a, d) refer to the original effect of QTL i. The first line in equation (B3) is the probability ratio for the new effect versus the original effect, for QTL i (when the genotype for QTL i is integrated out). The second line, then, is the probability ratio of having QTL j in the model versus not having it in the model (conditional on the new sampled genotypes for QTL i). The third and fourth lines are the ratio of priors and proposal probabilities for the move, and the last two lines are the Jacobian of the transformation from (a, d, u_a, u_d) to (a_i', d_i', a_j', d_j').

References

Besag J, Green P, Higdon D, Mengersen K (1995) Bayesian computation and stochastic systems. Stat Sci 10:3–66

Blangero J (1995) Genetic analysis of a common oligogenic trait with quantitative correlates: summary of GAW9 results. Genet Epidemiol 12:689–706

Cannings C, Thompson EA, Skolnick MH (1978) Probability functions on complex pedigrees. Adv Appl Prob 10:26–61

Carlin B, Chib S (1995) Bayesian model choice via Markov chain Monte Carlo. J R Stat Soc B 57:473–484

Dizier M-H, Bonaïti-Pellié C, Clerget-Darpoux F (1993) Conclusions of segregation analysis for family data generated under two-locus models. Am J Hum Genet 53:1338–1346

Elston RC, Stewart J (1971) A general model for the genetic analysis of pedigree data. Hum Hered 21:523–542

Falconer DS (1989) Introduction to quantitative genetics, 3rd ed. Wiley, New York

Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Trans Patt Anal Mach Intell 6:721–741

Green PJ (1994) Contribution to the discussion of paper by Grenander and Miller. J R Stat Soc B 56:589–590

——— (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82:711–732

Guo SW, Thompson EA (1992) A Monte Carlo method for combined segregation and linkage analysis. Am J Hum Genet 51:1111–1126

Haldane JBS (1919) The combination of linkage values, and the calculation of distance between the loci of linked factors. J Genet 8:299–309

Haley CS, Knott S (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109

Heath SC (1994) Estimation of linked QTL effects with an animal model using Gibbs sampling. In: Smith C, Gavora JS, Benkel B, Chesnais J, Fairfull W, Gibson JP, Kennedy BW, Burnside EB (eds) Proceedings of the Fifth World Congress on Genetics Applied to Livestock Production. Vol 18. University of Guelph, Guelph, Ontario, pp 398–401

——— (1995) Inferences on the genetic control of quantitative traits from selection experiments. PhD thesis, Edinburgh University, Edinburgh

Hoeschele I (1994) Bayesian QTL mapping via the Gibbs Sampler. In: Smith C, Gavora JS, Benkel B, Chesnais J, Fairfull W, Gibson JP, Kennedy BW, Burnside EB (eds) Proceedings of the Fifth World Congress on Genetics Applied to Livestock Production. Vol 21. University of Guelph, Guelph, Ontario, pp 241–244

Kong A (1991) Analysis of pedigree data using methods combining peeling and Gibbs sampling. In: Keramidas EM, Kaufman SM (eds) Computer Science and Statistics Proceedings of the 23rd Symposium on the Interface. Interface Foundation, Fairfax Station, VA, pp 379–385

Kruglyak L, Daly MJ, Lander ES (1995) Rapid multipoint linkage analysis of recessive traits in nuclear families, including homozygosity mapping. Am J Hum Genet 56:519–527

Lander ES, Green P (1987) Construction of multilocus genetic maps in humans. Proc Natl Acad Sci USA 84:2363–2367

Lin S, Thompson EA, Wijsman E (1993) Achieving irreducibility of the Markov chain Monte Carlo method applied to pedigree data. IMA J Math Appl Med Biol 10:1–17

——— (1994) Finding noncommunicating sets for Markov chain Monte Carlo estimations on pedigrees. Am J Hum Genet 54:695–704

MacCluer JW, Blangero J, Dyer TD, Kammerer CM (1995) Simulation of a common oligogenic disease with quantitative risk factors. GAW9 problem 2: the answers. Genet Epidemiol 12:707–712

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. J Chem Physiol 21:1087–1091

Ott J (1989) Computer-simulation methods in human linkage analysis. Proc Natl Acad Sci USA 86:4175–4178

Phillips D, Smith A (1996) Bayesian model comparison via jump diffusions. In: Gilks W, Richardson S, Spiegelhalter D (eds) Markov chain Monte Carlo in practice. Chapman & Hall, London, pp 215–219

Richardson S, Green PJ On Bayesian analysis of mixtures with an unknown number of components. J R Stat Soc B 59: 731–792

Satagopan JM, Yandell BS, Newton MA, Osborn TC (1996) A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. Genetics 144:805–816

Sheehan N, Thomas A (1993) On the irreducibility of a Markov chain defined on a space of genotype configurations by a sampling scheme. Biometrics 49:163–175

Stephens DA, Smith AFM (1993) Bayesian inference in multipoint gene mapping. Ann Hum Genet 57:65–82

Thomas A (1986) Approximate computation of probability functions for pedigree analysis. IMA J Math Appl Med Biol 3:157–166

Thompson EA (1994*a*) Monte Carlo likelihood in genetic mapping. Stat Sci 9:355–366

——— (1994*b*) Monte Carlo likelihood in the genetic mapping of complex traits. Phil Trans R Soc Lond B 344:345–351

Tierney L (1994) Markov chains for exploring posterior distributions. Ann Stat 22:1701–1762

Wang CS, Rutledge JJ, Gianola D (1993) Marginal inferences about variance components in a mixed linear model using Gibbs sampling. Genet Sel Evol 25:41–62