Multipoint genetic mapping of quantitative trait loci using a variable number of sibs per family

SHIZHONG XU* AND DAMIAN D. G. GESSLER
Department of Botany and Plant Sciences, University of California, Riverside, CA 92521, USA

(*Received 16 June 1997 and in revised form 30 October 1997*)

Summary

We present a multipoint algorithm to map quantitative trait loci (QTLs) using families from outbred populations with a variable number of sibs. The algorithm uses information from all markers on a chromosome simultaneously to extract information of QTL segregation. A previous multipoint method (Kruglyak & Lander (1995) *American Journal of Human Genetics* **57**, 439–454) extracts information using a hidden Markov model. However, this method is restricted to small families ($<$ 10 sibs). We present an approximate hidden Markov model approach that can handle large sibships while retaining similar efficiency to the previous method. Computer simulations support the notion that data sampled from a small number of large families provide more power than data obtained from a large number of small families, under the constraint that the total number of individuals for the two schemes is the same. This is further reflected in simulations with variable family sizes, where variance in family size improves the statistical power of QTL detection relative to a constant size control.

1. Introduction

The polymorphism of marker loci largely determines the efficiency and power of mapping quantitative trait loci (QTLs). In a line crossing experiment that involves two inbred parental strains, a (co-dominant) marker locus has two levels of polymorphism: fully informative and non-informative – partially informative markers do not exist. This yes-or-no characteristic warrants that the origins of QTL alleles can be traced solely with two flanking markers, an approach commonly referred to as interval mapping (Lander & Botstein, 1989; Haley & Knott, 1992). If a marker is not segregating in a given cross, an upstream marker can be used (Martínez & Curnow, 1992), but the procedure is still called interval mapping.

In less controlled populations, i.e. in the absence of inbred lines, markers may be partially informative. In such situations, two flanking markers may not extract the maximum amount of information about the segregation of a putative QTL and, because of this, markers outside the interval can provide additional information. In these situations, a more desirable procedure than interval mapping is to use all markers simultaneously, a procedure called multipoint mapping (Fulker *et al.*, 1995; Kruglyak & Lander, 1995; Olsen, 1995).

To map QTLs in outbred populations, variance component analysis has been used whereby the segregation variance of a putative QTL is estimated and tested (Haseman & Elston, 1972; Goldgar, 1990; Schork, 1993; Fulker & Cardon, 1994; Xu & Atchley 1995). The key to the variance component approach is to capitalize on the variance in the number of alleles identical-by-descent (IBD) shared by a pair of relatives at a particular locus. Kruglyak & Lander (1995) recently developed a multipoint method for QTL mapping using a hidden Markov model (HMM) approach. Their method uses all markers to predict the conditional *distribution* of the IBD value of a putative QTL. Fulker *et al.* (1995) and Olson (1995) also espouse a multipoint approach, though they use the estimated IBD values of all markers to infer the conditional *expectation* of the IBD at each putative QTL. With regard to testing the association of a chromosome position with a trait of interest, using the expectation of the IBD value is more flexible and

* Corresponding author. Tel: +1(909) 787-5898. Fax: +1(909) 787-4437. e-mail: xu@genetics.ucr.edu.

https://doi.org/10.1017/S0016672398003115 Published online by Cambridge University Press

convenient than using the distribution, with little difference in power and efficiency between the two (Fulker & Cherny, 1996; Gessler & Xu, 1996).

The multipoint mapping of Fulker *et al.* (1995) is an extension of the interval mapping of Fulker & Cardon (1994) where the squared phenotypic difference between pairs of sibs is regressed on the estimated IBD value. This sib-pair difference approach is less efficient than the maximum likelihood (ML) method that includes the multivariate relationships of sib data (Xu & Atchley, 1996; Fulker & Cherny, 1996). However, the idea of multipoint mapping of Fulker *et al.* (1995) is fundamentally important because the multipoint estimation of the IBD can be adapted to the ML analysis.

The multipoint method developed by Kruglyak & Lander (1995) considers all possible genotypic configurations of the offspring given all marker genotypes. Because of this, the computational load scales exponentially with family size, and for practical purposes this limits the algorithm to situations with fewer than 20 meioses (equivalent to < 10 sibs per family). Olson's (1995) multipoint method uses a similar idea to the hidden Markov model but it estimates the IBD states of a putative QTL using the probability distributions of IBDs at marker loci. The multipoint method of Fulker *et al.* (1995) takes a multiple regression approach to estimate the IBD of a putative QTL from the IBDs of the markers. The latter two methods can handle an arbitrary number of sibs and are computationally much faster than the multipoint method of Kruglyak & Lander (1995), though both are slightly less efficient.

In this paper, we will refer to Kruglyak & Lander's (1995) hidden Markov model as the *HMM method*, Fulker *et al.*'s (1995) method as the *regression method*, and the method we propose here as the *approximate HMM method*. In referring to Fulker *et al.*'s (1995) method as the regression method, we are referring only to the use of multiple regression in determining the IBDs; the determination of the QTL variance and its decomposition is done by maximum likelihood.

In many plants and laboratory or agricultural animals, family sizes can extend into the tens or hundreds. The motivation of this paper is to present an alternative multipoint method that can handle large families rapidly, while still retaining the comparative performance of previous multipoint models.

2. Theory

(i) *Linear model and likelihood function*

Consider a full-sib family with n siblings, where the phenotypic value of the jth individual is described by the following linear model:

$$y_j = \mu + g_j + a_j + e_j,$$

where μ is the overall mean, g_j is the additive effect of a putative QTL with mean 0 and variance σ_q^2, a_j is the polygenic effect (excluding g_j) with mean 0 and variance σ_a^2, and e_j is the residual effect distributed as $N(0, \sigma_e^2)$. A dominance effect is assumed absent.

In matrix notation, the phenotypic values of n siblings can be expressed as

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{g} + \mathbf{a} + \mathbf{e}$$

where $\mathbf{1}$ is a column vector of order n. The model has an expectation and variance-covariance matrix of

$$\mathrm{E}(\mathbf{y}) = \mathbf{1}\mu$$

and

$$\mathrm{Var}(\mathbf{y}) = \mathbf{V} = \Pi\sigma_q^2 + \mathbf{A}\sigma_a^2 + \mathbf{I}\sigma_e^2,$$

respectively, where $\Pi = \{\pi_{ij}\}_{n\times n}$ is an $n \times n$ matrix with the element of the ith row and the jth column being the shared IBD value for sibs i and j at the QTL, $\mathbf{A}$ is an $n \times n$ additive relationship matrix, and $\mathbf{I}$ is an identity matrix of order n.

Under the assumption that $\mathbf{y}$ is multivariate normal and Π is known, the likelihood function is

$$\mathrm{L}(\xi \mid \mathbf{y}\Pi) = |\mathbf{V}|^{-1/2} \exp\{-\tfrac{1}{2}(\mathbf{y} - \mathbf{1}\mu)^{\mathrm{T}}\, \mathrm{V}^{-1}(\mathbf{y} - \mathbf{1}\mu)\}, \qquad (1)$$

where $\xi = [\mu\, \sigma_q^2\, \sigma_a^2\, \sigma_e^2]^{\mathrm{T}}$ are the unknown parameters.

When N independent families are considered, the overall likelihood function is simply the product of these family-specific likelihoods. The test statistic is taken as the log likelihood ratio (Xu & Atchley, 1995).

(ii) *Elements of the additive relationship matrix **A** are the unconditional, expected IBDs for each sib-pair and, as such, the matrix is determined solely by the pedigree relationship*

In a full-sib family without inbreeding, the diagonal elements of $\mathbf{A}$ are unity and the off-diagonal elements are $\frac{1}{2}$. The diagonal elements of the matrix Π are also unity, but the off-diagonal elements vary depending on how many IBD alleles are shared by the two siblings. Since the genotype of a QTL cannot be seen, any given element π_{ij} is unobservable. Thus the rationale is to use markers in the same linkage group to infer the distribution of π_{ij}.

(iii) *Multipoint estimation of* π_{ij}

The IBD value can be partitioned into two components:

$$\pi_{ij} = \tfrac{1}{2}(\phi_{ij} + \gamma_{ij}), \qquad (2)$$

where ϕ_{ij} indicates that the sibs share a common paternal allele and γ_{ij} indicates they share a common maternal allele. The two components, ϕ_{ij} and γ_{ij}, may be referred to as the gametic IBD values. Each gametic IBD is a Bernoulli variable, i.e. $\phi_{ij} = 1$ if the

https://doi.org/10.1017/S0016672398003115 Published online by Cambridge University Press

sibs share their paternal allele and $\phi_{ij} = 0$ otherwise. When both parents are genotyped, ϕ_{ij} and γ_{ij} can be estimated from the markers separately. Let $\hat{\phi}_{ij} = \mathrm{E}(\phi_{ij} \,|\, \mathrm{I_M}) = \Pr(\phi_{ij} = 1 \,|\, \mathrm{I_M})$ and $\hat{\gamma}_{ij} = \mathrm{E}(\gamma_{ij} \,|\, \mathrm{I_M}) = \Pr(\gamma_{ij} = 1 \,|\, \mathrm{I_M})$ be the estimated ϕ_{ij} and γ_{ij} conditional on the marker genotypes $\mathrm{I_M}$; then the estimated IBD value is

$$\hat{\pi}_{ij} = \tfrac{1}{2}(\hat{\phi}_{ij} + \hat{\gamma}_{ij}) \tag{3}$$

(see also Wang *et al.*, 1995). We now discuss the multipoint estimation of the gametic IBD.

For convenience of presentation, we focus our discussion on one particular sib-pair, and thus replace the subscripts of ϕ_{ij} by q, meaning the gametic IBD at the putative QTL (denoted ϕ_q). Imagine that there are M ordered markers on the chromosome of interest and there is a gametic IBD for each marker locus, denoted by ϕ_k for the kth marker. The multipoint method essentially uses ϕ_k for $k = 1, \ldots, M$ to infer $\hat{\phi}_q = \mathrm{E}(\phi_q \,|\, \mathrm{I_M}) = \Pr(\phi_q = 1 \,|\, \mathrm{I_M})$.

Assume that the QTL is located between marker k and $k+1$ where $M-1 \geqslant k \geqslant 1$. In other words, there are k markers on the left and $M-k$ markers on the right of the QTL. What we want is to calculate $\Pr(\phi_q = 1 \,|\, \boldsymbol{\phi})$, the conditional probability that $\phi_q = 1$ given $\boldsymbol{\phi}$ of the markers. The sequence $\{\phi_1 \ldots \phi_k \phi_q \phi_{k+1} \ldots \phi_M\}$ forms a reversible Markov chain with a transition matrix between ϕ_k and ϕ_l of

$$\mathbf{T}_{kl} = \begin{bmatrix} (1-\theta_{kl})^2 + \theta_{kl}^2 & 2\theta_{kl}(1-\theta_{kl}) \\ 2(1-\theta_{kl})\,\theta_{kl} & (1-\theta_{kl})^2 + \theta_{kl}^2 \end{bmatrix}$$

(Guo, 1994), that is

$$\Pr(\phi_k = 1 \,|\, \phi_l = 1) = \Pr(\phi_k = 0 \,|\, \phi_l = 0) = (1-\theta_{kl})^2 + \theta_{kl}^2$$

and

$$\Pr(\phi_k = 0 \,|\, \phi_l = 1) = \Pr(\phi_k = 1 \,|\, \phi_l = 0) = 2(1-\theta_{kl})\,\theta_{kl},$$

where θ_{kl} is the recombination fraction between markers k and l. When this transition matrix is used, the linkage phase information of the parents is not required. Our purpose here is to estimate ϕ_q from $\boldsymbol{\phi} = [\phi_1 \phi_2 \ldots \phi_M]$ using a hidden Markov model approach (Lander & Green, 1987; Kruglyak & Lander, 1995).

The conditional probability is given by

$$\Pr(\phi_q \,|\, \phi_1 \ldots \phi_M) = \frac{\Pr(\phi_1 \ldots \phi_k \phi_q \phi_{k+1} \ldots \phi_M)}{\Pr(\phi_1 \ldots \phi_k \phi_{k+1} \ldots \phi_M)}, \tag{4}$$

where

$$\Pr(\phi_1 \ldots \phi_k \phi_q \phi_{k+1} \ldots \phi_M) = \Pr(\phi_q) \Pr(\phi_1 \ldots \phi_k \,|\, \phi_q) \Pr(\phi_{k+1} \ldots \phi_M \,|\, \phi_q)$$

and

$$\Pr(\phi_1 \ldots \phi_k \phi_{k+1} \ldots \phi_M) = \Sigma_{\phi_q} \Pr(\phi_q) \Pr(\phi_1 \ldots \phi_k \,|\, \phi_q) \Pr(\phi_{k+1} \ldots \phi_M \,|\, \phi_q).$$

$\Pr(\phi_q)$ is the prior probability with $\Pr(\phi_q = 1) = \Pr(\phi_q = 0) = \frac{1}{2}$. Substituting the above joint probabilities into (4), we have

$$\Pr(\phi_q \,|\, \phi_1 \ldots \phi_M) = \frac{\Pr(\phi_q) \Pr(\phi_1 \ldots \phi_k \,|\, \phi_q) \Pr(\phi_{k+1} \ldots \phi_M \,|\, \phi_q)}{\Sigma_{\phi_q} \Pr(\phi_q) \Pr(\phi_1 \ldots \phi_k \,|\, \phi_q) \Pr(\phi_{k+1} \ldots \phi_M \,|\, \phi_q)}.$$

Since $\{\phi_1 \ldots \phi_M\}$ is a Markov chain, we have

$$\Pr(\phi_1 \ldots \phi_k \,|\, \phi_q) = \Pr(\phi_1 \,|\, \phi_2) \Pr(\phi_2 \,|\, \phi_3) \ldots \Pr(\phi_{k-1} \,|\, \phi_k) \Pr(\phi_k \,|\, \phi_q)$$

and

$$\Pr(\phi_{k+1} \ldots \phi_M \,|\, \phi_q) = \Pr(\phi_M \,|\, \phi_{M-1}) \Pr(\phi_{M-1} \,|\, \phi_{M-2}) \ldots \Pr(\phi_{k+2} \,|\, \phi_{k+1}) \Pr(\phi_{k+1} \,|\, \phi_q).$$

Therefore,

$$\Pr(\phi_q \,|\, \phi_1 \ldots \phi_M) = \frac{\Pr(\phi_q) \Pr(\phi_k \,|\, \phi_q) \Pr(\phi_{k+1} \,|\, \phi_q)}{\Sigma_{\phi_q} \Pr(\phi_q) \Pr(\phi_k \,|\, \phi_q) \Pr(\phi_{k+1} \,|\, \phi_q)}.$$

It is clear that $\Pr(\phi_q \,|\, \phi_1 \ldots \phi_M) = \Pr(\phi_q \,|\, \phi_k \phi_{k+1})$, meaning that conditioning on flanking markers is equivalent to conditioning on all markers.

In reality, elements of $\boldsymbol{\phi}$ are not always observed. We must use the marker information to infer their distribution. Let $\Pr(\phi_1 \,|\, \mathrm{I_M}), \ldots, \Pr(\phi_M \,|\, \mathrm{I_M})$ be the prior distributions of markers $\boldsymbol{\phi}$ given $\mathrm{I_M}$ and write $\Pr(\mathrm{I_M} \,|\, \phi_q)$ by

$$\begin{aligned} \Pr(\mathrm{I_M} \,|\, \phi_q) &= [\Sigma_{\phi_1} \ldots \Sigma_{\phi_k} \Pr(\phi_1 \,|\, \mathrm{I_M}) \ldots \Pr(\phi_k \,|\, \mathrm{I_M}) \quad \Pr(\phi_1 \ldots \phi_k \,|\, \phi_q)] \\ &\quad \times [\Sigma_{\phi_{k+1}} \ldots \Sigma_{\phi_M} \Pr(\phi_{k+1} \,|\, \mathrm{I_M}) \ldots \Pr(\phi_M \,|\, \mathrm{I_M}) \\ &\quad \Pr(\phi_{k+1} \ldots \phi_M \,|\, \phi_q)]. \end{aligned}$$

Then, from Bayes' theorem, the hidden Markov model is written as

$$\Pr(\phi_q \,|\, \mathrm{I_M}) = \frac{\Pr(\phi_q) \Pr(\mathrm{I_M} \,|\, \phi_q)}{\Sigma_{\phi_q} \Pr(\phi_q) \Pr(\mathrm{I_M} \,|\, \phi_q)}.$$

Specifically,

$$\begin{aligned} \Pr(\phi_q = 1 \,|\, \mathrm{I_M}) &= \frac{0.5 \Pr(\mathrm{I_M} \,|\, \phi_q = 1)}{0.5 \Pr(\mathrm{I_M} \,|\, \phi_q = 1) + 0.5 \Pr(\mathrm{I_M} \,|\, \phi_q = 0)} \\ &= \frac{\Pr(\mathrm{I_M} \,|\, \phi_q = 1)}{\Pr(\mathrm{I_M} \,|\, \phi_q = 1) + \Pr(\mathrm{I_M} \,|\, \phi_q = 0)} = \hat{\phi}_q \end{aligned}$$

and

$$\Pr(\phi_q = 0 \,|\, \mathrm{I_M}) = 1 - \hat{\phi}_q.$$

In matrix notation, we have

$$\Pr(\mathrm{I_M} \,|\, \phi_q = 1) = \mathbf{1}^{\mathrm{T}} D_1 \mathbf{T}_{12} \mathbf{D}_2 \ldots \mathbf{T}_{kq} \mathbf{D}_{(1)} \mathbf{T}_{qk+1} \ldots \mathbf{D}_{M-1} \mathbf{T}_{M-1M} \mathbf{D}_M \mathbf{1}$$

and

$$\Pr(\mathrm{I_M} \,|\, \phi_q = 0) = \mathbf{1}^{\mathrm{T}} \mathbf{D}_1 \mathbf{T}_{12} \mathbf{D}_2 \ldots \mathbf{T}_{kq} \mathbf{D}_{(0)} \mathbf{T}_{qk+1} \ldots \mathbf{D}_{M-1} \mathbf{T}_{M-1M} \mathbf{D}_M \mathbf{1}$$

https://doi.org/10.1017/S0016672398003115 Published online by Cambridge University Press

where

$$\mathbf{D}_k = \begin{bmatrix} \Pr(\phi_k = 1 \mid \mathrm{I_M}) & 0 \\ 0 & \Pr(\phi_k = 0 \mid \mathrm{I_M}) \end{bmatrix},$$

$$\mathbf{D}_{(1)} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{D}_{(0)} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad \text{and } \mathbf{1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \qquad (5)$$

Note that when a marker is fully informative, $\Pr(\phi_q = 1 \mid \mathrm{I_M}) = 1$ and $\Pr(\phi_q = 0 \mid \mathrm{I_M}) = 0$ or $\Pr(\phi_q = 1 \mid \mathrm{I_M}) = 0$ and $\Pr(\phi_q = 0 \mid \mathrm{I_M}) = 1$. With the multipoint method one could get the same estimate of ϕ_q at a given position by using the pair of the nearest informative markers to calculate $\hat{\phi}_q$. One can consider the multipoint method implemented here as providing an automatic search for the nearest fully informative flanking markers and using them to estimate ϕ_q.

(iv) *Computing* $\Pr(\phi_k = 1 \mid \mathrm{I_M})$ *from marker genotypes*

$\Pr(\phi_k = 1 \mid \mathrm{I_M})$ is the probability that the two sibs share a common paternal allele given the genotypes of themselves and their parents at marker k. Define p_{j1} as the probability that the paternal (first) allele of the father has been inherited by sib j, given the observed genotypes at the marker under consideration. Of course, the probability that the jth sib has inherited the maternal (second) allele of the father is then $\mathrm{p}_{j2} = 1 - \mathrm{p}_{j1}$. These probabilities, listed in Table 1, determine the prior distribution of the gametic IBD at marker k:

$$\Pr(\phi_k = 1 \mid \mathrm{I_M}) = \mathrm{p}_{i1}\,\mathrm{p}_{j1} + \mathrm{p}_{i2}\,\mathrm{p}_{j2}. \qquad (6)$$

The multipoint estimation of the gametic IBD through the mother, $\hat{\gamma}_{ij} = \Pr(\gamma_{ij} \mid \mathrm{I_M})$, is similarly derived. Given $\hat{\phi}_{ij}$ and $\hat{\gamma}_{ij}$, the multipoint estimation of the IBD value takes the average of the two components, as given by (3)

3. Simulation methods

We employ the above theory to map a QTL in a series of simulated outbred populations. The simulation techniques are similar to those described in Gessler & Xu (1996), but we give a brief summary here to note those differences in map length, number of alleles, and so forth that are unique to these simulations.

To begin a simulation, we assign constant allele frequencies to pre-chosen marker and QTL positions to build a common grandparent population. These alleles constitute all marker loci and a single specific QTL that we are interested in mapping. In all cases, we work with one 50 cM chromosomal segment with six markers at positions 0, 10, 20, 30, 40 and 50 and a QTL at position 25 (between markers 3 and 4). Each marker segregates four, and the QTL segregates six, equally frequent alleles. From this grandparent population we generate four gametes and from these gametes we construct two parents. Each parent also receives an additional 12 unlinked loci that simulate the polygenic contribution to the trait. Each polygenic allele is drawn from a normal distribution with mean zero and variance $\sigma_a^2/24$. A random pick of one allele from each polygenic locus of each parent is contributed to each offspring as its polygenic effect. From each set of parents we generate two or more sibs, each sib being the product of an independent meiotic event. How we determine the actual number of sibs per family is described below. We repeat this for a total of N families, with parents for each family being picked anew from the invariant, infinite-size grandparent population. The number of families (N) is determined such that $N \times$ (the mean number of sibs/family) is approximately 500.

Using the mean and variance of the QTL effect in the grandparent population, we transform the QTL effect in each sib so that its genetic contribution is distributed with mean zero and variance $\sigma_q^2 = 12{\cdot}5$. The phenotypic value for each sib is the sum of this effect, its polygenic contribution, and an environmental effect. Each polygenic allele has a variance of $12{\cdot}5/24$, so that the total polygenic variance (contributed by 24 alleles) is $\sigma_a^2 = 12{\cdot}5$. The environmental effect is a normal variate with mean zero and variance $\sigma_e^2 = 25$. The QTL therefore has a heritability of $h_q^2 = \sigma_q^2/\sigma^2 = 0{\cdot}25$; the remaining 75% of the variance is divided between the polygenic $h_a^2 = \sigma_a^2/\sigma^2 = 0{\cdot}25$ and environmental terms $\sigma_e^2/\sigma^2 = 0{\cdot}50$.

For each sib-pair, an estimate of π_q is obtained as described in the previous section. These π_q values are then used in a maximum likelihood algorithm to infer the variance components σ_q^2, σ_a^2 and σ_e^2. This is repeated every 2 cM along the chromosomal segment, and a likelihood ratio, $\mathrm{LR} = -2(\mathrm{L}_0 - \mathrm{L}_1)$, is computed. L_0 and L_1 are the values of the log-likelihood functions under the null ($\sigma_q^2 = 0$) and alternative ($\sigma_q^2 > 0$) hypotheses respectively.

We call the above the 'standard' setup, and examine variants on it by modifying the number of sibs per family, variance in the number of sibs per family, the number of alleles at marker loci, the QTL heritability, and the method of IBD estimation. For each variant, all parameters are identical to the standard setup except for the change under inspection. To vary the number of sibs per family, we examine two, eight and 50 sibs per family for 250, 62/63 and 10 families respectively. The '62/63' reflects that for 8 sibs per family half the runs are with 62 families and half are with 63 families; we combine the results to estimate $62{\cdot}5 \times 8 = 500$ individuals. To add variance to the number of sibs per family, family size is distributed as $\mathrm{P}(\mathrm{X} = x) = \text{Poisson }(\mu_n - 2; x) + 2$ where μ_n is the expected number of sibs per family. For example, for $\mu_n = 4$, each run has exactly 125 families, so while any given run may deviate slightly in its total number of

https://doi.org/10.1017/S0016672398003115 Published online by Cambridge University Press

Table 1. *Probabilities of inheritance used to compute gametic IBDs of equation* (6) *in the text*

Mating type (parent × other parent)	Progeny type	p_{j1}
I: $M_iM_i \times$ —	—	$\frac{1}{2}$
II: $M_iM_j \times M_iM_i$	M_iM_i	1
	M_iM_j	0
III: $M_iM_j \times M_jM_j$	M_iM_j	1
	M_jM_j	0
IV: $M_iM_j \times M_kM_k$	M_iM_k	1
	M_jM_k	0
V: $M_iM_j \times M_iM_j$	M_iM_i	1
	M_iM_j	$\frac{1}{2}$[a]
	M_jM_j	0
VI: $M_iM_j \times M_iM_k$	M_iM_i	1
	M_iM_k	1
	M_iM_j	0
	M_jM_k	0
VII: $M_iM_j \times M_jM_k$	M_iM_j	1
	M_iM_k	1
	M_jM_j	0
	M_jM_k	0
VIII: $M_iM_j \times M_kM_l$	M_iM_k	1
	M_iM_l	1
	M_jM_k	0
	M_jM_l	0

p_{j1} is the probability that individual j has inherited the paternal (first) allele of the parent in consideration at a marker locus, given the genotypes of both parents and the progeny. Alleles within the parent are arranged as paternal followed by maternal. Dashes indicate any genotype.

[a] Under mating type V ($M_1M_2 \times M_1M_2$), if the sib-pair genotypic configuration is M_1M_2–M_1M_1 or M_1M_2–M_2M_2 then we know with certainty that the sibs share $\frac{1}{2}$ of their alleles. This, though, differs from the $\frac{1}{2}$ of mating type I, where $\frac{1}{2}$ is the expectation, and thus represents our ignorance of the actual number of alleles shared. This distinction requires a special treatment for mating type V. In (5) in the text, instead of using

$$\mathbf{D}_k = \begin{bmatrix} \Pr(\phi_k = 1 \mid I_M & 0 \\ 0 & \Pr(\phi_k = 0 \mid I_M) \end{bmatrix}$$

to compute $\Pr(\phi_q \mid I_M)$, we should break the chain into two chains. One is computed using $\mathbf{D}_k = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and the result is denoted by Pr $(\phi_q \mid I_M)_1$. The other is computed using $\mathbf{D}_k = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ and the result is denoted by $\Pr(\phi_q \mid I_M)_0$. The final result is then obtained by

$$\Pr(\phi_q \mid I_M) = \Pr(\phi_k = 1 \mid I_M) \Pr(\phi_q \mid I_M)_1 + \Pr(\phi_k = 0 \mid I_M) \quad \Pr(\phi_q \mid I_M)_0.$$

As pointed out by a reviewer, the probability of this event is $(\sum_{i=1}^{s} q_i^2)^2 - \sum_{i=1}^{s} q_i^4$ where q_i is the frequency of allele M_i for $i = 1, \ldots, s$. For alleles with equal frequencies ($q = q_1 = q_2 = \ldots = q_s$) this expression becomes $q^2 - q^3$. Therefore, for two alleles this probability is 0·125, for three alleles it is 0·074 and for four alleles it is 0·047. If more than one marker has mating type V, then the above is repeated recursively for all possible combinations.

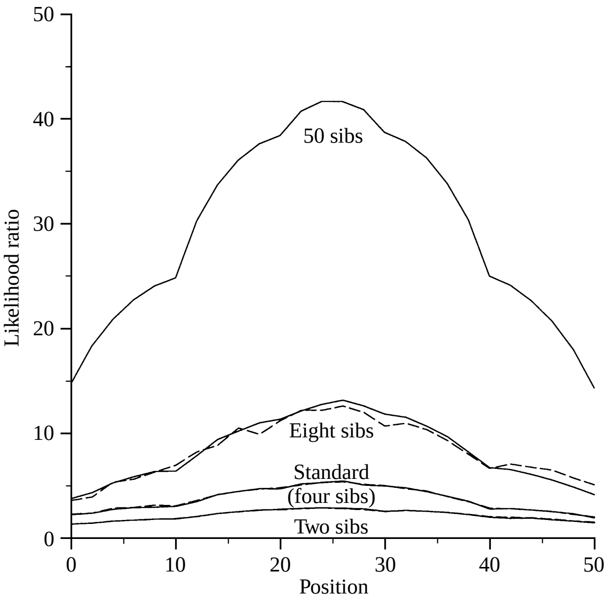

Fig. 1. Likelihood ratios across 50 putative QTL positions for families with two, four, eight and 50 sibs. Continuous line is the approximate HMM method, dashed line is the HMM method. Lines are obtained by averaging the likelihood ratio at each position over all runs and then connecting successive values. The true QTL position is at position 25.

individuals, across runs, the average total number of individuals is kept at approximately 500. To vary marker informativeness, we examine cases with two or eight alleles per maker. For heritability, we look at $h_q^2 = 0{\cdot}125$ and $h_q^2 = 0{\cdot}4$. In both cases we maintain a broad-sense heritability of 0·5. Lastly, we compare our method with the HMM and regression methods.

For the HMM method, we first generate the simulated data as described above, and then input this into MAPMAKER/SIBS (Kruglyak & Lander, 1995). We have MAPMAKER/SIBS calculate the π_{ij} values and then read these π_{ij} values back into our program for use by our maximum likelihood algorithm. This eliminates potential differences between the search engines, and focuses the comparisons on the core problem: generating accurate π_{ij} values.

For the regression method, we coded the algorithm directly. The algorithm is strictly only applicable to sib-pairs, so for families with four, eight and 50 sibs per family we treated each sib-pair within a family as an independent family. This *ad hoc* adjustment strictly invalidates the statistical justification of the method, so for comparison we imposed the same sib-pair treatment on our approximate HMM method even though it does not require it.

We report the result of 250 independent simulations for the standard and each variant, with the exception of only 25 runs for eight sibs under MAPMAKER/SIBS. This is because of the high computational expense for eight sibs under this method. For select cases, we estimate the strength of a false positive signal by

https://doi.org/10.1017/S0016672398003115 Published online by Cambridge University Press

running an additional 1000 simulations with no QTL segregating. For these simulations, we set $\sigma_a^2 = 25$ to maintain an overall heritability of 50%. From each simulation we choose the maximum observed likelihood ratio found across the chromosome segment, and then choose the 50th greatest value from the list of 1000 maximums as an estimate of the segment-wise critical value (with 95% confidence) when no QTL is segregating.

Table 2. *Observed 95th percentile likelihood ratios when no QTL is segregating*

Simulation	Test statistic
Standard	4·97
Two sibs per family	5·96
Eight sibs per family:	
Without sib-pairing	5·09
With sib-pairing	6·24
Regression method	6·10
Fifty sibs per family:	
Without sib-pairing	4·27
With sib-pairing	7·53
Two alleles at marker loci	5·65
Eight alleles at marker loci	5·04

All values are for the approximate HMM method unless otherwise noted. ‘Sib-pairing’ refers to treating each sib-pair within a family as an independent family when the likelihood ratio is calculated.

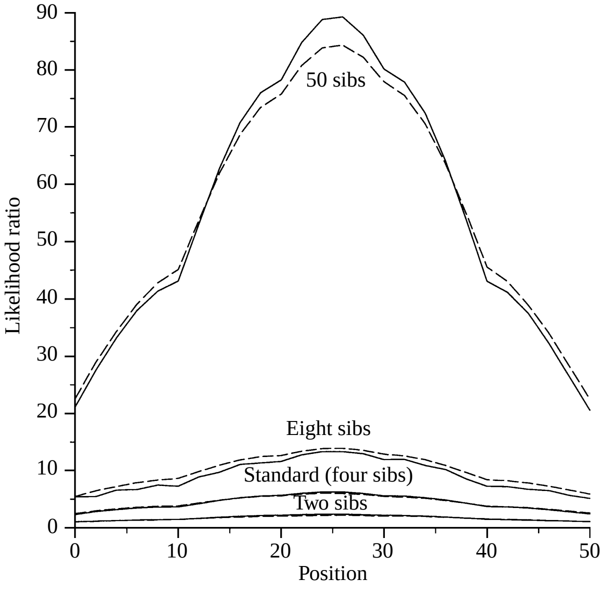

Fig. 2. Similar to Fig. 1, except that the dashed line is the regression method. For all lines, the maximum likelihood procedure decomposes families into multiple families of sib-pairs. The fact that the signal for the regression method for eight sibs exceeds the approximate HMM method across all positions means that correlations amongst positions artificially inflate the signal, even though this is not reflected by a higher critical test statistic in Table 2.

4. Results

Fig. 1 shows the large increase in signal when effort is partitioned into fewer families of larger size, rather than more families of smaller size. In the cases

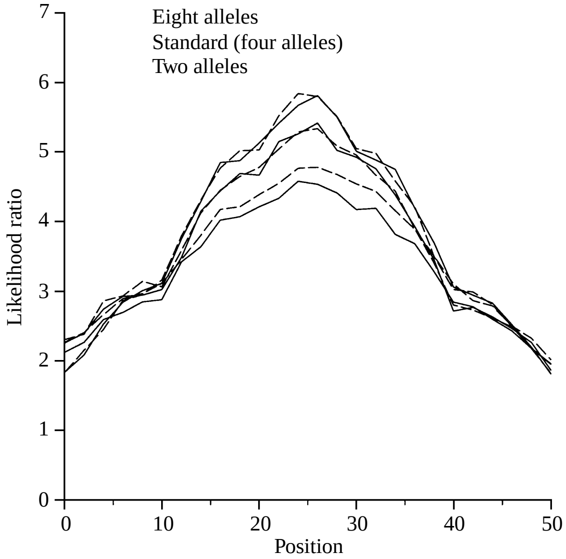

Fig. 3. Effect of varying the number of alleles at marker loci. Continuous line is the approximate HMM method, dashed line is the HMM method. Jaggedness is due to the combined effect of a relatively low signal and testing for putative position every 2 cM.

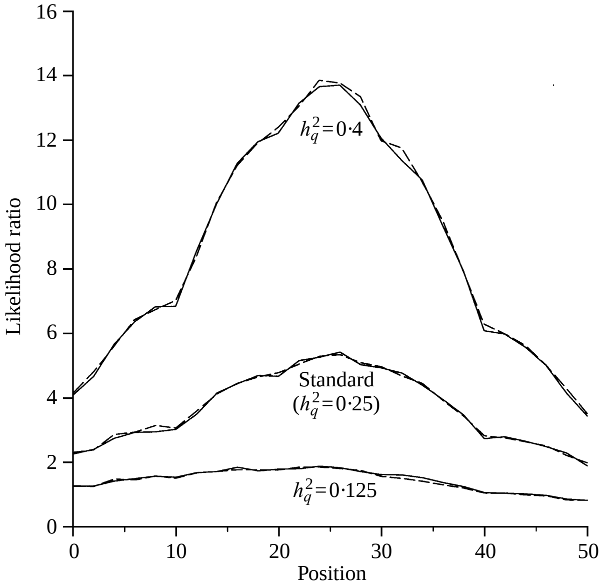

Fig. 4. Effect of varying the heritability of the trait due to the QTL. Continuous line is the approximate HMM method, dashed line is the HMM method.

https://doi.org/10.1017/S0016672398003115 Published online by Cambridge University Press

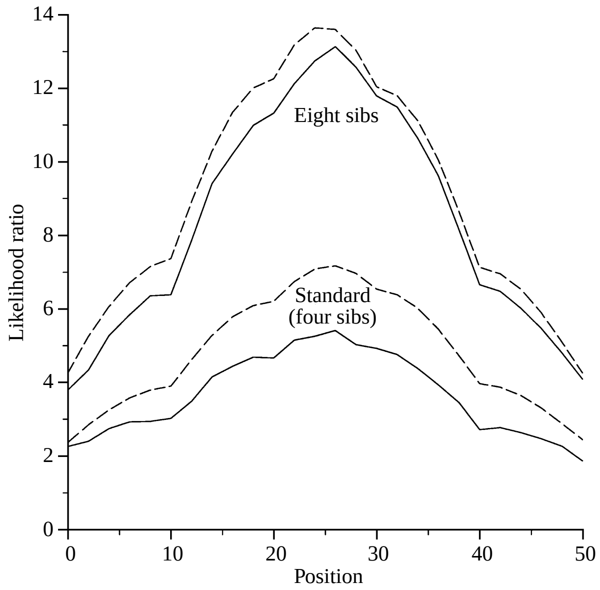

Fig. 5. Effect of adding variability to the number of sibs per family. Continuous line is with a constant number of sibs, dashed line is with a variable number of sibs as described in the text. Both lines use the approximate HMM method. Lines are averages of 300 runs.

examined, a signal that is barely significant with four sibs per family is unequivocally significant under a change in experimental design (critical test statistics are reported in Table 2). For the smaller family sizes there is virtually no difference between the approximate and the full HMM method.

Fig. 2 shows analogous results, but this time relative to the regression method. Two trends are evident from the figure. First, there is again virtually no difference between the regression and the approximate HMM methods. Second, there is a large gain in the likelihood ratio relative to Fig. 1. This is due to the breaking of each family into multiple families of sib-pairs. We address this in Section 5.

The similarity between the HMM methods is reiterated in Figs. 3 and 4, which show an expected increase in the signal as marker informativeness and heritability increase. Both signals for two alleles (Fig. 3) are below the critical test statistic (Table 2), so these lines and their differences give only suggestive evidence of a QTL.

Fig. 5 shows an increase in signal when one uses a variable number of sibs. The result is robust over

Table 3. *The approximate HMM method*

	Position	Total phenotypic variance	h_q^2	h_a^2
Standard	24·70	49·69	0·25	0·25
	(12·12)	(3·75)	(0·1149)	(0·1494)
No. of sibs per family				
2	24·52	50·30	0·31	0·19
	(13·94)	(3·11)	(0·1549)	(0·1756)
8	25·22	49·80	0·26	0·23
	(7·97)	(3·81)	(0·0800)	(0·1401)
50	25·34	49·17	0·23	0·24
	(4·68)	(6·33)	(0·0660)	(0·1979)
No. of alleles at marker loci				
2	24·39	49·56	0·28	0·21
	(12·12)	(3·53)	(0·1295)	(0·1698)
8	23·77	49·64	0·25	0·24
	(11·10)	(3·51)	(0·1089)	(0·1640)
QTL heritability				
0·125	21·43	49·65	016	0·33
	(14·61)	(3·75)	(0·1019)	(0·1553)
0·4	24·50	49·89	0·39	0·11
	(6·73)	(3·44)	(0·0977)	(0·1222)
Variable no. of sibs per family				
mean 4	24·85	49·81	0·27	0·22
	(11·03)	(3·38)	(0·1085)	(0·1521)
mean 8	25·12	49·87	0·26	0·23
	(8·53)	(3·85)	(0·0864)	(0·1466)

Estimates of the QTL position, total phenotypic variance, and the QTL and polygenic heritabilities. True values are 25, 50, 0·25 and 0·25 respectively unless otherwise noted. Estimates are obtained by selecting the position with the largest likelihood ratio from each run and then averaging over these values. Reported are the averages and (standard deviations) over 250 runs. Total number of individuals in all cases is 500. "Standard' refers to the standard setup as referred to in the text: four sibs per family, four alleles per marker locus, and $h_q^2 = 0{\cdot}25$.

https://doi.org/10.1017/S0016672398003115 Published online by Cambridge University Press

Table 4. *Similar to Table 3 but for the HMM method*

	Position	Total phenotypic variance	h_q^2	h_a^2
Standard	24·04	49·68	0·25	0·25
	(12·19)	(3·75)	(0·1136)	(0·1522)
No. of sibs per family				
2	24·48	50·31	0·31	0·19
	(14·01)	(3·12)	(0·1551)	(0·1740)
8	23·28	49·12	0·26	0·19
	(7·53)	(4·74)	(0·0745)	(0·0908)
No. of alleles at marker loci				
2	25·42	49·56	0·27	0·21
	(12·50)	(3·53)	(0·1180)	(0·1616)
8	23·62	49·63	0·25	0·24
	(10·61)	(3·51)	(0·1087)	(0·1615)
QTL heritability				
0·125	22·41	49·65	0·16	0·33
	(14·62)	(3·76)	(0·1038)	(0·1535)
0·4	24·65	49·88	0·39	0·12
	(6·67)	(3·44)	(0·1036)	(0·1260)

Table 5. *Similar to Tables 3 and 4, but when the likelihood ratio is calculated, each family is treated as multiple families of sib-pairs*

	Position	Total phenotypic variance	h_q^2	h_a^2
Approximate HMM method				
No. of sibs per family				
2	25·18	49·96	0·28	0·24
	(13·44)	(3·10)	(0·1644)	(0·1909)
4	25·44	49·67	0·27	0·22
	(11·73)	(3·42)	(0·1102)	(0·1418)
8	24·94	49·72	0·24	0·25
	(8·84)	(3·80)	(0·1063)	(0·1572)
50	25·49	49·03	0·24	0·22
	(5·16)	(6·05)	(0·0740)	(0·1872)
Regression method				
No. of sibs per family				
2	24·71	49·98	0·29	0·23
	(13·79)	(3·07)	(0·1717)	(0·1986)
4	24·47	49·64	0·28	0·21
	(12·11)	(3·39)	(0·1221)	(0·1474)
8	24·73	49·70	0·27	0·22
	(9·48)	(3·82)	(0·0955)	(0·1457)
50	25·38	49·02	0·25	0·21
	(6·28)	(6·03)	(0·0759)	(0·1864)

other family sizes not shown, with a relative gain increasing with family size.

Finally, we report specific statistics from all simulations in Tables 3–5. All methods are similar in their ability to locate the QTL and estimate the total phenotypic variance. All methods also conserve the sum $h_q^2 + h_a^2$ (showing that there is no confounding between the genetic and environmental sources of variation) and partition the total genetic variance into its two components. The success of the partition is qualitative, with some bias observed when the heritability is low or there are only two sibs per family.

5. Discussion

We present a method that aims at achieving the performance of the HMM method while extending its applicability to large families. We do this by using the

https://doi.org/10.1017/S0016672398003115 Published online by Cambridge University Press

conditional expectation of the IBD value instead of the conditional distribution, and this greatly simplifies the method. There is a link between this resultant method and the regression method of Fulker *et al.* (1995).

Essentially what Fulker *et al.* (1995) do is to solve the equation $\beta = \tilde{\mathbf{V}}^{-1}\mathbf{C}$, where $\tilde{\mathbf{V}}$ is the $M \times M$ variance–covariance matrix of marker IBD values and $\mathbf{C}$ is the $1 \times M$ covariance vector of IBD values of the putative QTL position and each of the M markers. The resultant elements β_m of β are the regression coefficients used to predict the IBD value of the putative QTL ($\hat{\pi}_q$) from the estimated IBD values of the markers:

$$\hat{\pi}_q = \beta_0 + \sum_{m=1}^{M} \beta_m \hat{\pi}_m,$$

where $\hat{\pi}_m$ is the estimated IBD value of the mth marker (Fulker *et al.*, 1995). The single vector β (derived from all families) is used repeatedly for each family's π_q. In this sense, β is the best-fit set of regression coefficients over all families: we will call it the *Global* β method. Of course, families differ in the information content of their sibs, so, as recognized by Fulker *et al.* (1995), this Global β method is only an approximation to a method where a separate β is computed for each individual family. We call this the *Individual* β method. This Individual β method cannot be worse than the Global β method, and indeed, by tailoring β to each family, it should be better whenever there is enough ambiguity such that different markers are partially informative in different families but not so much ambiguity as to render all methods powerless.

We have coded the Individual β method. In extensive simulations, including duplicates of the simulations we report here, the Individual β method is computationally identical to the fourth decimal place in virtually all respects with a variant of the approximate HMM method of this paper. Recall that (2) decomposes the IBD value π_{ij} into the paternal and maternal gametic IBD values ϕ_{ij} and γ_{ij}. Alternatively – although with the loss of some information – one could estimate the average gametic IBD value of the two parents in one routine: a routine that does not distinguish between the paternal and maternal lines. This single chain approach is computationally identical (as defined above) to the Individual β method.

There is, then, a progression of statistical improvements from the Global β method to the Individual β method to the approximate HMM to the HMM. Each method presents certain advantages as regards its computability versus its viable parameter space. Part of what we demonstrate in this paper is how closely these methods can converge, despite theoretical reasons for believing one to be superior to another. Our decision to use the approximate HMM method is based largely on the fact that it does not require the inversion of $\tilde{\mathbf{V}}$ and can be directly extended to large sibships (though note the special treatment required in Table 1). This increase in power is sufficient, under the parameter space we examine, to render the method virtually identical to the more exacting HMM method of Kruglyak & Lander (1995).

One may rationalize that if information content is low, it is imperative to use the most efficient method. We did not find this in our simulations. When information content was low (e.g. two alleles or $h_q^2 = 0{\cdot}125$) differences between methods were at best marginal. Strictly, although we do expect the regression method, the approximate HMM method and the HMM method to demonstrate a succession of increasing power, we found only two factors that produced notable increases in the test statistic: increasing the number of sibs per family and treating families as independent families of sib-pairs. Note that the latter is applied only when computing the likelihood function, not when estimating Π. Decomposing families into sib-pairs when estimating Π can cause a significant loss in power.

Increasing power by using more than two sibs per family has been previously demonstrated in both fixed and random models (Blackwelder & Elston, 1982; Götz & Ollivier, 1992; Amos *et al.*, 1996). The reason for the increase in power is primarily an increase in the total number of sib-pairs. Recall that the phenotypic covariance between sibs is

$$\mathrm{Cov(FS)} = \pi_q \sigma_q^2 + \tfrac{1}{2}\sigma_a^2.$$

The power to detect a QTL depends on deviations of π_q from $\frac{1}{2}$, thereby allowing a separation of the variance components. For two sibs per family there is only one sib-pair, and thus one π_q per family. In general, for N families each with n sibs, the total number of individuals is Nn and the total number of sib-pairs is $N[n(n-1)]/2$. As n increases while Nn is held constant, the number of sib-pairs increases faster than N decreases; consequently, so does the number of π_q values and the number of (non-independent) sources of information. For a total population size of 500, increasing the family size from two to 50 sibs per family increases the number of sib-pairs 49-fold: from 250 families × 1 sib-pair per family = 250 sib-pairs to 10 × 1225 = 12250.

A similar analysis can be extended to variable family sizes (see Götz & Ollivier, 1992). For variable family sizes, let n be a Poisson random variable with mean and variance λ. Then the expected total number of sib-pairs is

$$N\mathrm{E}\left[\frac{n(n-1)}{2}\right] = \tfrac{1}{2}N[\mathrm{E}(n^2) - \mathrm{E}(n)]$$

$$= \tfrac{1}{2}N[\mathrm{Var}(n) + \mathrm{E}^2(n) - \mathrm{E}(n)] = \tfrac{1}{2}N\lambda^2.$$

https://doi.org/10.1017/S0016672398003115 Published online by Cambridge University Press

In our simulations under the standard settings, $n = 2+x$, where x is Poisson variate of mean $\lambda = 2$. Thus

$$\tfrac{1}{2}N[\mathrm{Var}(n)+\mathrm{E}^2(n)-\mathrm{E}(n)] = \tfrac{1}{2}\times 125\times(2+16-4) = 875.$$

This exceeds the number of sib-pairs available under a constant family size of four $[\tfrac{1}{2}\times 125\times(0+16-4) = 750]$.

The conclusion that large families and variance in family sizes can increase the power in QTL mapping is somewhat mitigated by a limit beyond which the power may actually decrease (Muranty, 1996). This is the point where the number of families is reduced such that the resultant group of parents fails to capture an adequate sample of segregating marker and QTL alleles. At this point, the benefit due to increasing family size can be offset by a loss of information in the parental populations; i.e. sampling alone may yield a QTL strictly monomorphic in the studied population, and thus hidden from mapping. This, though, should be more of a problem in traditional line-crossing techniques than in the outbred techniques discussed here.

DNA heterozygosity studies (Nei, 1987) show that the majority of extant variation is captured with a relatively small number of individuals, though rare alleles are likely to be missed without considerable sampling effort. In allele-sharing QTL mapping algorithms such as the ones investigated here, allelic diversity *per se* is important only to the degree that one can use it to infer the meiotic history of a locus. This is reasonably attained with six to ten segregating alleles, though we have found weak dependencies up through as many as 50 alleles. Multipoint methods automatically skip non-informative and missing markers as they assess each marker's contribution. This, combined with the fact that once a fully informative marker is found all loci upstream contribute no new information, should make multipoint mapping relatively robust to all but severe reductions in the number of families.

In applying sib-pair algorithms to multiple sibs, previous studies treated each sib-pair in a large family as an independent family (Blackwelder & Elston, 1982; Götz & Ollivier, 1992; Amos *et al.*, 1996). To date there has been no explicit measure of the cost of this simplification *vis-à-vis* a multivariate analysis; to our surprise, we found that contrary to any net cost in signal strength, it actually increases the ability to detect the QTL by a considerable amount (compare the approximate HMM method in Figs. 1 and 2).

Statistically, such an *ad hoc* procedure violates the independence assumptions of the maximum likelihood model. Blackwelder & Elston (1982) showed that the cost in terms of false positives is minor for sib-trios, and this holds true for 50 sibs as well (Table 2). We did find, though, that our search engine was more likely to get caught in spurious maxima when there was sib-pairing. This computational cost is somewhat offset by the fact that decomposing families into sib-pairs replaces the $n\times n$ matrix of n sibs in the n-dimensional normal distribution, (1), by a series of bivariate normals. This circumvents the requirement of taking the inverse of the $n\times n$ variance–covariance matrix, and consequently reduces the computational burden to $O(n^2)$.

The sib-pair simplification means that each n-sib family likelihood is now the sum of $[n(n-1)]/2$ sib-pair likelihoods. Blackwelder & Elston (1982) discuss why an increase in family size increases the power relative to a comparable number of two-sib families, but why is this not also captured by the n-variate analysis? In an n-variate analysis, n sibs inflate the LR linearly with n. This is because the scalar returned by (1) increases with the dimension of $\mathbf{V}$. As seen above, though, sib-pairs increase the LR proportional to $O(n^2)$, and thus return a higher signal. Thus not only is the sib-pair simplification more computationally efficient, it is also powerful.

Wang *et al.* (1995) presented a method of QTL evaluation in general pedigrees using a single marker. Grignola *et al.* (1996*a*, *b*) extended the method of Wang *et al.* (1995) into flanking markers. Because some parents may be homozygous at the markers closest to the QTL, different flanking markers were used in different families. In other words, the flanking interval of the QTL varied so that the flanking markers were always fully informative. This method is similar to the multipoint algorithm proposed in this study. However, the former requires knowledge of the parental linkage phases of the markers whereas the method presented here does not. It would be useful to compare the efficiencies of the two methods in terms of statistical accuracy and computational convenience.

We end with a final caveat on the three procedures. There are parameter spaces that support the theoretical expectations of the three procedures: the HMM exceeding the approximate HMM, in turn exceeding the regression method (Fig. 6). These conditions tend to be under increased marker density and higher heritability, with average gains of each method of approximately 10–15%. Transitively, then, differences between the HMM method and the regression method can be considerable. However, for even moderate family sizes this comes at an exponentially high computational cost: for eight sibs, the HMM method runs two orders of magnitude slower than the approximate HMM method, and for 50 sibs, the HMM method is untenable.

Fulker *et al.*'s (1995) method is attractive for its ease of presentation and conceptual clarity. Kruglyak & Lander's (1995) HMM method is attractive for its thorough use of available information. The approximate HMM method is a medial design: it is

https://doi.org/10.1017/S0016672398003115 Published online by Cambridge University Press

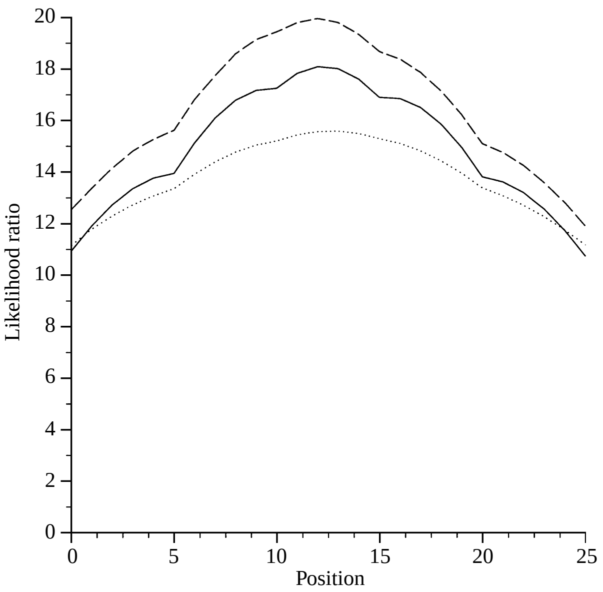

Fig. 6. Differences between the three methods. Top line is the HMM method, middle line is the approximate HMM method, bottom line is the regression method. Two sibs per family, 500 families, $h_q^2 = 0{\cdot}5$, six biallelic markers with alleles at equal frequencies and markers every 5 cM. Putative positions are tested every centimorgan. QTL is at position 12.

comparably as fast as Fulker *et al.*'s (1995) method, while it brings much of the power of Kruglyak & Lander's (1995) method to moderate and large sibships. The greatest improvements, though, come not in the choice of methods, but in the experimental design.

We wish to thank David Fulker and two reviewers for valuable comments on an earlier version of the manuscript. This research was supported by the National Institutes of Health Grant GM55321-01 and the USDA National Research Initiative Competitive Grants Program 95-37205-2313.

References

Amos, C. I., Zhu, D. K. & Boerwinkle, E. (1996). Assessing genetic linkage and association with robust components of variance approaches. *Annals of Human Genetics* **60**, 143–160.

Blackwelder, W. C. & Elston, R. C. (1982). Power and robustness of sib-pair linkage tests and extension to large sibships. *Communications in Statistics: Theory and Methods* **11**, 449–484.

Fulker, D. W. & Cardon, L. R. (1994). A sib-pair approach to interval mapping of quantitative trait loci. *American Journal of Human Genetics* **54**, 1092–1103.

Fulker, D. W. & Cherny, S. S. (1996). An improved multipoint sib-pair analysis of quantitative traits. *Behavior Genetics* **26**, 527–532.

Fulker, D. W., Cherny, S. S. & Cardon, L. R. (1995). Multipoint interval mapping of quantitative trait loci, using sib pairs. *American Journal of Human Genetics* **56**, 1224–1233.

Gessler, D. D. G. & Xu, S. (1996). Using the expectation or the distribution of the identity by descent for mapping quantitative trait loci under the random model. *American Journal of Human Genetics* **59**, 1382–1390.

Goldgar, D. E. (1990). Multipoint analysis of human quantitative genetic variation. *American Journal of Human Genetics* **47**, 957–967.

Götz, K. U. & Ollivier, L. (1992). Theoretical aspects of applying sib-pair linkage to livestock species. *Genetics*, *Selection*, *Evolution* **24**, 29–42.

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

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

Guo, S.-W. (1994). Computation of identity-by-descent proportions shared by two siblings. *American Journal of Human Genetics* **54**, 1104–1109.

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

Haseman, J. K. & Elston, R. C. (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behavior Genetics* **2**, 3–19.

Kruglyak, L. & Lander, E. S. (1995). Complete multipoint sib-pair analysis of qualitative and quantitative traits. *American Journal of Human Genetics* **57**, 439–454.

Lander, E. S. & Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199.

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

Martínez, O. & Curnow, R. N. (1992). Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theoretical and Applied Genetics* **85**, 480–488.

Muranty, H. (1996). Power of tests for quantitative trait loci detection using full-sib families in different schemes. *Heredity* **76**, 156–165.

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Olson, J. M. (1995). Multipoint linkage analysis using sib pairs: an interval mapping approach for dichotomous outcomes. *American Journal of Human Genetics* **56**, 788–798.

Schork, N. J. (1993). Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations. *American Journal of Human Genetics* **53**, 1306–1319.

Wang, T., Fernando, R. L. & van der Beek, S. (1995). Covariance between relatives for a marked quantitative trait locus. *Genetics*, *Selection*, *Evolution* **27**, 251–274.

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

https://doi.org/10.1017/S0016672398003115 Published online by Cambridge University Press