Theor Appl Genet (1993) 85:946–952

Bayesian analysis of linkage between genetic markers and quantitative trait loci. II. Combining prior knowledge with experimental evidence

I. Hoeschele[1] and P. M. VanRaden[2]

[1] Department of Dairy Science, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0315, USA
[2] Animal Improvement Programs Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD 20705-2350, USA

Received December 10, 1991; Accepted July 30, 1992
Communicated by D. Van Vleck

Summary. A Bayesian method was developed for identifying genetic markers linked to quantitative trait loci (QTL) by analyzing data from daughter or granddaughter designs and single markers or marker pairs. Traditional methods may yield unrealistic results because linkage tests depend on number of markers and QTL gene effects associated with selected markers are overestimated. The Bayesian or posterior probability of linkage combines information from a daughter or granddaughter design with the prior probability of linkage between a marker locus and a QTL. If the posterior probability exceeds a certain quantity, linkage is declared. Upon linkage acceptance, Bayesian estimates of marker-QTL recombination rate and QTL gene effects and frequencies are obtained. The Bayesian estimates of QTL gene effects account for different amounts of information by shrinking information from data toward the mean or mode of a prior exponential distribution of gene effects. Computation of the Bayesian analysis is feasible. Exact results are given for biallelic QTL, and extensions to multiallelic QTL are suggested.

Key words: Genetic markers – Quantitative trait loci – Linkage analysis – Bayesian method – *Prior information*

Introduction

With an increasing number of genetic markers for the dairy cattle genome (Georges et al. 1991; Steele and Georges 1991), marker-assisted selection (MAS) may be utilized in genetic improvement programs in the near future. MAS may be implemented in four stages: (1) development of a large number of DNA markers and establishment of a linkage map of the markers; (2) analysis of association between markers and quantitative traits to identify those markers linked to QTL of large effects; (3) routine genotyping of elite breeding stock or the entire population for the selected genetic markers; (4) direct use of this information in genetic evaluation.

The second stage can reduce the number of markers retained for routine MAS and provide initial estimates of additive genetic effects associated with these markers. Several designs and methods of anaysis have been used or suggested for this purpose.

The granddaughter design (Weller et al. 1990) seems most practical for linkage analysis in a dairy population because it requires the smallest number of animals to be genotyped. However, this design can be improved in several ways. Granddaughter averages by son can be modelled rather than individual records of granddaughters as done by Weller et al. (1990). Granddaughter averages are routinely computed in the genetic evaluation system for production as daughter yield deviations (DYDs) of bulls (VanRaden and Wiggans 1991) and are adjusted for systematic environmental effects and merits of mates of sons.

Methods of analysis of granddaughter or daughter designs that have been used in the past are linear fixed regression (Weller et al. 1990; Hoeschele and Meinert 1990), iteratively reweighted fixed regression (Dentine and Cowan 1990), random regression (Goddard 1992), and maximum likelihood (ML) (Weller 1990). ML simultaneously estimates quantitative trait loci (QTL) substitution effects, QTL allele or genotype frequencies, and recombination rates between marker and QTL.

Correspondence to: I. Hoeschele

Fixed regression analysis estimates marker allele substitution effects, which are a function of recombination rate with and substitution effects at a linked QTL. Random regression can predict the effects of QTL alleles provided that the recombination rate is known (Fernando and Grossman 1990; Goddard 1992), but it loses some information by assuming all animals have different QTL alleles except when inherited within a tight marker bracket.

In this paper, a Bayesian analysis of daughter or granddaughter designs is developed and illustrated with a simulated granddaughter design. In the Bayesian analysis, information from a daughter or granddaughter design is combined with the prior probability of linkage, which depends on the number of detectable QTL (Hoeschele and VanRaden 1993). Linkage is declared if the posterior probability exceeds a certain level. Subsequent to the demonstration of linkage, recombination rate, QTL gene or genotype frequencies, and QTL substitution effects are estimated. In contrast to ML, QTL effects are treated as random by combining the likelihood with a prior distribution of gene effects.

Methodology

Model for performance of offspring

Daughter or granddaughter designs provide the data necessary to detect linkage between markers and QTL within sire families. The sires are heterozygous at the markers, whereas their QTL genotypes are unknown. Phenotypic or genetic merit differences among offspring with different marker alleles indicate a linked, heterozygous QTL.

In general, the additive genetic value of an offspring is composed of the transmitting ability of the sire, the transmitting ability of the dam, sire Mendelian sampling, and dam Mendelian sampling. Sire Mendelian sampling may be partitioned into sampling at a particular QTL and sampling at all other loci. Working models for the phenotype of an offspring may combine sire Mendelian sampling at all other loci and dam Mendelian sampling with random environment in the residual,

$$y_{ijk} = s_i + bd_j + \alpha x^Q_{ijk} + e_{ijk} \quad (1)$$

or also include transmitting ability of the dam in the residual,

$$y_{ijk} = s_i + \alpha x^Q_{ijk} + \varepsilon_{ijk} \quad (2)$$

where y is the offspring's phenotype, s is transmitting ability of the sire, b is regression on dam's transmitting ability, d, (replaced by an estimate), α is the QTL allele substitution effect in the sire, e and ε are residuals, and x^Q_{ijk} is $+0.5$ or -0.5 depending on which QTL allele an offspring inherited from the sire. In subsequent discussion, a QTL substitution effect α will be referred to as an additive effect of a QTL. Effects of sires are considered as fixed for simplicity. Sires are highly selected, and differences among sires are less important than differences among offspring that inherited one or another marker allele.

Model (1) residuals are, at least approximately, uncorrelated and, therefore, permit working with a likelihood simpler than that of model (2), in which the residuals are correlated through additive genetic relationships among dams. The residual variances are generally heterogeneous because the y's represent average and adjusted records. Using identical models with the α-term omitted, we can calculate the residual variances (see Appendix). With the α-term in the model, residual variance is reduced by an amount of $0.25\alpha^2$ or $0.0625\alpha^2$, if the y's are daughter records or granddaughter averages by son, respectively. The calculation of the likelihood for analysis of linkage to a single QTL usually assumes that residuals are normally distributed (e.g., Paterson et al. 1991), although the true distribution is a mixture in the presence of other QTL and the QTL alleles inherited from the dams are not considered.

A QTL may be assumed to be either biallelic or highly polymorphic (Hoeschele and VanRanden 1993). At a biallelic QTL with allele frequency p, a fraction, $p^2 + (1-p)^2$, of sires are homozygous, and a fraction, $2p(1-p)$, of sires are heterozygous at the QTL with additive effect α. At a highly polymorphic QTL, each sire is assumed to carry two unique QTL alleles (Hoeschele and VanRaden 1993). QTL effects are then nested within sires in models (1) and (2), i.e., α is replaced by α_i. A fraction P of the sires is assumed to be heterozygous, i.e., to carry two alleles with distinguishable effects.

Linkage likelihood

Linkage likelihoods for segregating populations are derived from the theorem of total probability and are conditional upon the selection scheme determining the families used (e.g., MacLean et al. 1984). The conditional likelihood of the phenotypes y given the marker information M is employed here because only this part of the likelihood depends on the parameters of interest.

Alleles at a locus are considered to be random variables and denoted by X. Superscripts M and Q denote marker or QTL alleles, respectively. A marker allele inherited from the sire is termed sire-marker allele; a sire-QTL allele is defined analogously. The probability that a particular realization of X is x will be denoted by Prob(x).

Let there be S sires, identified by $i = 1,\ldots,S$. Marker and QTL genotypes of sire i are written in terms of the alleles as ${}^1x^M_i{:}{}^2x^M_i$ and ${}^1x^Q_i{:}{}^2x^Q_i$, respectively. Marker and QTL alleles with identical left superscripts are located on the same homologous chromosome in the sire. Let $\mathbf{y}$, $\mathbf{x}^M$, and $\mathbf{x}^Q$ be vectors of offspring phenotypes, sire-marker, and sire-QTL alleles, respectively, and $\mathbf{x}^M_*$ a subvector of $\mathbf{x}^M$ containing unknown sire-marker alleles of offspring. In the daughter design, $\mathbf{y}$ represents the vector of phenotypic records of daughters, and in the granddaughter design $\mathbf{y}$ is the vector of averages, by son, of the phenotypic records of granddaughters.

Known variables are the phenotypes ($\mathbf{y}$) and the sire-marker alleles of the offspring ($\mathbf{x}^M$). Unknown variables include the QTL genotypes of the S series (${}^1x^Q_i{:}{}^2x^Q_i$, $i = 1,\ldots,S$), the sire-marker alleles of some offspring that were not genotyped or whose sire-marker alleles cannot be determined with certainty ($\mathbf{x}^M_*$), and the sire-QTL alleles of all offspring ($\mathbf{x}^Q$). Summation is over all unknowns, weighted by appropriate probabilities.

The likelihood for linkage to one QTL in S sires is

$$L(\mathbf{y}|\mathbf{M}) = \sum_{{}^1x^Q_1{:}{}^2x^Q_1} \cdots \sum_{{}^1x^Q_S{:}{}^2x^Q_S} \Big\{ \mathrm{Prob}({}^1x^Q_1{:}{}^2x^Q_1,\ldots,{}^1x^Q_S{:}{}^2x^Q_S | {}^1x^M_1{:}{}^2x^M_1,\ldots,{}^1x^M_S{:}{}^2x^M_S) \cdot \sum_{\mathbf{x}^M_*} \mathrm{Prob}(\mathbf{x}^M_*|\mathbf{M}) \sum_{\mathbf{x}^Q} \mathrm{Prob}(\mathbf{x}^Q|\mathbf{x}^M, {}^1x^Q_1{:}{}^2x^Q_1,\ldots,{}^1x^Q_S{:}{}^2x^Q_S) \cdot f_\alpha(\mathbf{y}|\mathbf{x}^Q) \Big\} \quad (3)$$

Likelihood (3) is general and can accommodate relationships among sires; i.e., dependencies among the QTL genotypes of

the series. To compute the likelihood, (3) can often be simplified. For example, when there is linkage equilibrium across families, all sire-marker alleles are known, residuals of the y's are not correlated through dams, and sires are unrelated; the likelihood then is

$$L(\mathbf{y}|\mathbf{M}) = \prod_{i=1}^{s} \left\{ \sum_{{}_1x_i^Q:{}_2x_i^Q} \text{Prob}({}_1x_i^Q:{}_2x_i^Q) \cdot \prod_{j=1}^{n_i} \sum_{x_{ij}^Q} \text{Prob}(x_{ij}^Q|x_{jj}^M, {}_1x_i^Q:{}_2x_i^Q) f(y_{ij}|x_{ij}^Q) \right\} \quad (4)$$

where n_i is number of offspring of sire i, x_{ij}^Q is the sire-QTL allele of offspring j of sire i with $x_{ij}^Q = {}_1x_i^Q$ or $x_{ij}^Q = {}_2x_i^Q$, and x_{ij}^M is the sire-marker allele of the offspring. The offspring are either daughters (daughter design) or sons (granddaughter design) of sire i.

Parameterizations of the likelihood in terms of the unknown QTL parameters are given in the Appendix.

Posterior inferences about linkage and QTL parameters

The prior distribution of recombination rate given in Eq. (12) of Hoeschele and VanRaden (1993) consists of the prior probability of nonlinkage, Prob(r = 0.5), and a continuous prior density for r < 0.5. Combining the prior with the information from data yields the posterior distribution of the parameters of a QTL, given a single marker, which also consists of a discrete and a continuous part, or

$$\text{Posterior } (r, p, \alpha) = \begin{cases} \text{Prob}(r = 0.5|\mathbf{y}, \mathbf{M}) \\ (1 - \text{Prob}(r = 0.5|\mathbf{y}, \mathbf{M}) * f(r, p, \alpha|\mathbf{y}, \mathbf{M}) \end{cases} \quad (5)$$

where the posterior probability of no linkage is calculated as

$$\text{Prob}\,(r = 0.5|\mathbf{y}, \mathbf{M}) = \frac{\text{Prob}(r = 0.5) * \underset{\alpha,p}{E}\,[L(\mathbf{y}|\mathbf{M}; p, \alpha, r = 0.5)]}{\text{Prob}(r = 0.5) * \underset{\alpha,p}{E}\,[L(\mathbf{y}|\mathbf{M}; p, \alpha, r = 0.5)] + (1 - \text{Prob}(r = 0.5)) * \underset{r,\alpha,p}{E}\,[L(\mathbf{y}|\mathbf{M}; r, p, \alpha)]} \quad (6)$$

In (6), Prob (r = 0.5) represents the prior probability of nonlinkage as given in Eqs. (10) and (11) of Hoeschele and VanRaden (1993), and L(.) denotes the likelihood, e.g., (3) or (4), dependent on the QTL parameters r, p and α. Expectations in the denominator of (6) are taken with respect to the prior distribution of the parameters, e.g.,

$$\underset{r,\alpha,p}{E}\, L(\mathbf{y}|\mathbf{M}; r, p, \alpha) = \int_0^{0.5} \int_{\alpha_l}^{\alpha_u} \int_{p_l}^{p_u} L(\mathbf{y}|\mathbf{M}; r, p, \alpha) f(p, \alpha) f(r)\, dp d\alpha dr$$

In (5), (6), and from here on, allele frequency p may be replaced with the frequency of heterozygosity, P, as defined in Hoeschele and VanRaden (1993). For a polymorphic QTL, additive effect α should be replaced with a vector containing additive effects α_i nested within sires.

The posterior density of the QTL parameters r, p, and α in (5) is

$$f(r, p, \alpha|\mathbf{y}, \mathbf{M}) = L(\mathbf{y}|\mathbf{M}; r, p, \alpha) * f(\alpha, p) * f(r)/f(\mathbf{y}|\mathbf{M}) \quad (7)$$

where the marginal density $f(\mathbf{y}|\mathbf{M})$ equals the denominator of (6).

In testing the hypothesis of linkage versus no linkage, the Bayesian decision is simply the hypothesis with the larger posterior probability if the loss is 0–1 loss (Berger 1985). This decision rule implies acceptance of the linkage hypothesis if Prob (r = 0.5|y, M) < 0.5.

Upon linkage acceptance, point estimates of the QTL parameters can be obtained from (7). Point estimates may also be chosen based on a loss function with the (r, p, α)-mode of (7) computed under 0–1 loss and the r, p, and α marginal means under quadratic loss (Berger 1985, p 161). The Bayesian joint posterior modal estimates of r, p, and α are obtained by maximizing (7) apart from the constant $f(\mathbf{y}|\mathbf{M})$. Due to the inclusion of prior information in (7), the Bayesian estimates are different from the ML solutions. The information from data on α is shrunken toward mode α_1 of the prior distribution of α. If α were estimated by its marginal posterior mean, the information from data would be shrunken toward mean β of the prior distribution of α.

The likelihood depends on the additional parameters **s**, vector of sire effects, and b in models (1) and (2). To obtain the likelihood in (6) or (7), the corresponding likelihood conditional on **s** and b would have to be integrated with respect to these nuisance parameters. A computationally more feasible approximation consists of evaluating the likelihood at point estimates of **s** and b obtained; for example, by jointly maximizing the posterior density with respect to p, r, α, **s**, and b.

If a pair of linked marker loci M1 and M2 rather than a single marker is analyzed, the posterior probability of linkage equals the sum of the posterior probabilities that a QTL is located on the flank closest to M1 and linked to M1 (order Q:M1:M2), between M1 and M2 (order M1:Q:M2), or on the flank closest to M2 and linked (order M1:M2:Q). The posterior probability of nonlinkage can be represented as

$$\text{Prob}(r = 0.5|\mathbf{y}, \mathbf{M}) = \text{Prob}(r = 0.5) * L(r = 0.5) / \{ \text{Prob}(r = 0.5) * L(r = 0.5) + \text{Prob}(Q{:}M1{:}M2) * L(Q{:}M1{:}M2) + \text{Prob}(M1{:}Q{:}M2) \times * L(M1{:}Q{:}M2) + \text{Prob}(M1{:}M2{:}Q) * L(M1{:}M2{:}Q) \}. \quad (8)$$

In (8), r = 0.5 represents nonlinkage of a QTL to either marker, and the prior probabilities of nonlinkage and of the three linkage cases are given in Appendix B of Hoeschele and VanRaden (1993). The likelihoods in the denominator of (8) are evaluated at r = 0.5 and integrated with respect to α and p, integrated with respect to r_1, α and p, integrated with respect to r_1/r_M, α, and p, and integrated with respect to r_2, α, and p, respectively; r_1, r_M, and r_2 are recombination rates among the QTL and M1, among M1 and M2, and among the QTL and M2, respectively (see Appendix).

The posterior probability of a QTL on the flank closest to M1 and linked (Q:M1:M2), between the markers (M1:Q:M2), or on the flank closest to M2 and linked (M1:M2:Q) can be obtained by replacing the numerator of (8) with the corresponding denominator term. For example, the posterior probability of a QTL on the flank closest to M1 and linked equals Prob(Q:M1:M2)*L(Q:M1:M2)/denominator of (8). Neumann (1991) derived similar posterior probabilities for recombinant inbred strains and Mendelian traits with complete penetrance.

Results

A numerical study was conducted to investigate whether a granddaughter design of size and structure resembling an actual design in the US Holstein population would provide sufficient information to estimate all parameters of a marked QTL, and to assess the computational feasibility of the Bayesian analysis. An actual design in US Holsteins may include 13 bulls with an average of 120 proven sons and a high likelihood of detecting marker effects of at least 100 kg milk in individual sire families. With the average heterozygosity of variable number of tandem repeat (VNTR) sequences, restriction fragment length

polymorphism (RFLP) multisite haplotype, and polymerase chain reaction (PCR) markers in the order of 50–60% (Georges et al. 1991: Steele and Georges 1991), an average of 6 sires should be heterozygous at a marker.

Simulated granddaughter design

Phenotypic and marker data were generated for 6 sires with the characteristics listed in Table 1. The trait was milk yield with a phenotypic standard deviation $\sigma = 1134$ kg and heritability $h^2 = 0.25$. The parameters at a biallelic QTL were $p = 0.5$, $\alpha = 300$ kg, and $r = 0.05$, 0.20, or 0.50.

For each son, an average of milk yields of daughters, deviated from their mean, was simulated as

$$y_{ij} = s_i + \alpha z_{ij} + 0.5 d_{ij} + e_{ij}$$

where y_{ij} is the simulated average of daughter milk yield deviations for son j of sire i; sire effect s_i was generated as $s_i' + 0.25(\alpha_k + \alpha_l)$, $s_i' \sim N\{0, 0.0625\,[\sigma_a^2 - 2p(1-p)\alpha^2]\}$; α_k represents the additive effect of allele k (k, l = 1, 2) with $\alpha_1 = p\alpha$ and $\alpha_2 = -(1-p)\alpha$, $z_{ij} = 0.0$, -0.5, or 0.5 if the sire was homozygous, heterozygous and the son inherited q, or heterozygous and the son received Q, respectively; $d_{ij} \sim N(0, 0.25\sigma_a^2)$, $\mathrm{Var}(e_{ij}) = (0.25/\mathrm{Rel}_{ij} - 0.125)*h^2*\sigma^2 - 0.625*\alpha^2$, and Rel was Reliability (VanRaden and Wiggans 1991) or the squared correlation between true and estimated breeding value. Sire-marker alleles, assumed perfectly traceable, were randomly assigned to sons, and sire-QTL alleles were assigned subsequently depending on recombination rate r. Predicted transmitting ability of the dam, PTA_{ij}, was computed from her true transmitting ability d_{ij} using Rel = 0.70 as

$$PTA_{ij} = \mathrm{Rel}*d_{ij} + \sqrt{1 - \mathrm{Rel}}\,\sigma_a r; \quad r \sim N(0, 1).$$

The data were analyzed with model (1), except that PTA_{ij} replaced d_j.

Table 1. Characteristics of a simulated granddaughter design

	Sire					
	1	2	3	4	5	6
QTL genotype	qQ	qq	Qq	qq	qQ	QQ
Number of sons	134	140	123	84	74	115
△marker(kg)[a]	253	19	316	41	342	1

[a] Absolute difference in average y − 0.5 PTA(dam) among the sons grouped by the marker allele; y is an average of milk yields, deviated from their mean, of daughters of a son; PTA is predicted transmitting ability

Computational requirements

Maximum likelihood and joint posterior mode estimates were computed based on (4) and (7) apart from the constant $f(\mathbf{y}|\mathbf{M})$, respectively. The likelihood and posterior density were maximized with respect to 10 (r, P, α, $s_1, \ldots, s_6$, b) or 15 (r, P, $\alpha_1, \ldots, \alpha_6$, $s_1, \ldots, s_6$, b) parameters assuming a biallelic or multialleic QTL, respectively. Maximization was performed with the Quasi-Newton algorithm of IMSL (1987) subroutine DBCONF, which does not require analytical derivatives. Computation of the posterior probability of nonlinkage in (6) requires two- and three-dimensional integration, which was performed using IMSL subroutine DQAND. Computation of marginal posterior mean estimates of r, p, and α required evaluation of three additional three-dimensional integrals. Maximization and two-dimensional integration required 20–120 CPU s, and three-dimensional integration used 10–20 CPU minutes on an IBM 3090.

Numerical problems occur when likelihood or log likelihood functions are evaluated because both contain products of n_i densities for each sire i with n_i offspring. For many offspring per sire, values of the products are smaller than the smallest number storable or representable in floating-point mode, $10^{-\max}$, where max is close to 80 for many computers. Thus, products of densities within each sire were scaled upward by large constants. In (4), let

$$P_{ij}({}^1x_i^Q{:}{}^2x_i^Q) = \sum_{x_{ij}^Q} \mathrm{Prob}(x_{ij}^Q | x_{ij}^M, {}^1x_i^Q{:}{}^2x_i^Q) f(y_{ij}|x_{ij}^Q)$$

and let $({}^1x_i^Q{:}{}^2x_i^Q) = Q_mQ_m$, $Q_mQ_{m'}$, or $Q_{m'}Q_m$ be the three genotypic classes of the sires as described in the Appendix. Then, evaluating likelihood (4) requires accumulating the products

$$\prod_{j=1}^{n_i} P_{ij}(Q_mQ_m), \prod_{j=1}^{n_i} P_{ij}(Q_mQ_{m'}), \text{ and } \prod_{j=1}^{n_i} P_{ij}(Q_{m'}Q_m).$$

If during accumulation the smallest of these products became less than an arbitrarily small number (10^{-50}), all products were multiplied by its reciprocal (10^{50}), if the largest product was less than $10^{(\max-50)}$, with $10^{\max}$ being the largest storable number. Otherwise, the smallest product was set equal to zero. The number of multiplications per sire was counted and summed across sires at the end of accumulation. The log likelihood was computed as

$$\sum_{i=1}^{S} \log\left[(1-P)\prod_{j=1}^{n_i} P_{ij}(Q_mQ_m) + 0.5P\prod_{j=1}^{n_i} P_{ij}(Q_mQ_{m'}) + 0.5P\prod_{j=1}^{n_i} P_{ij}(Q_{m'}Q_m)\right]$$

where P is the frequency of heterozygosity (see Appendix). For maximization, parameter values

yielding the largest log likelihood or log posterior density (log likelihood plus log prior density) were determined.

For integration, the log likelihood or log posterior density had to be antilogged, requiring an additional number of multiplications to obtain storable values. The total number of multiplications was determined for each likelihood or posterior density evaluated at particular parameter values, and the minimum number of multiplications, min, was found. Likelihoods or posterior densities requiring more than min multiplications were set to zero. The resulting integral equals the desired integral multiplied by 10^{min*50}. This factor cancels out in evaluating the posterior probability of nonlinkage in (6) and also cancels out in computing marginal posterior means as point estimates of α, p (P), or r. For example, the marginal posterior mean of α may be represented as

$$E(\alpha|\mathbf{y},\mathbf{M}) = \left\{ \int_0^{\alpha_u} \int_{0.0}^{0.5} \int_{p_l}^{p_u} \alpha L(\mathbf{y}|\mathbf{M};p,r,\alpha) f(\alpha,p) f(r)\, dp\, dr\, d\alpha \right\} \Big/ f(\mathbf{y}|\mathbf{M})$$

where $f(\mathbf{y}|\mathbf{M})$ is the denominator of (6).

Note that for estimating α the lower limit α_l was set to zero to avoid a bias in the estimate. With $\alpha_l = 0$, $\text{Prob}(r = 0.5) = 0$.

Posterior probability of linkage

Posterior probabilities of linkage were computed for the granddaughter design described in Table 1, for a single marker, and for marker-QTL recombination rates of 0.05 and 0.5. Results are presented in Table 2.

There was very strong evidence in favor of linkage for the given granddaughter design, tight marker-QTL linkage ($r = 0.05$), and a QTL effect $\alpha = 300$ kg of the same magnitude as the effect estimated for the prolactin locus (Cowan et al. 1990). For the same QTL effect but not linked to the marker ($r = 0.50$), the probability of linkage was considerably decreased by the data, suggesting that the marker was not linked to a QTL.

Table 2. Posterior probabilities of linkage for the granddaughter design in Table 2 and assumed prior probabilities

Linkage	Prior probability (linkage)	Posterior probability (linkage)
yes ($r = 0.05$)	0.03	1.0
	0.50	1.0
no ($r = 0.50$)	0.03	0.01
	0.50	0.23

Estimation of QTL parameters

The QTL parameters α, r, and p or P were estimated by ML Bayesian estimators for the granddaughter design simulated with $r = 0.20$.

The likelihood function (4) was first parameterized in terms of α, r, and QTL gene frequency, p. The estimate of p obtained by ML and Bayesian methods from the granddaughter design in Table 1 was 0.75, which corresponds to the frequency of allele q estimated by gene counting, pretending that the QTL genotypes of the sires were known and that sire 6 was homozygous qq instead of QQ. Because the granddaughter design and the within-sire model of analysis could not distinguish between different homozygous genotypes in the sires, the likelihood was reparameterized in terms of frequency of heterozygosity, P.

Four additional granddaughter designs were simulated with the same parameter values as for the design in Table 1. Each design was generated by reassigning QTL genotypes to sires, sire-marker and sire-QTL alleles to sons, and regenerating DYD of sons. True values of r, p, and α were constant across designs. Averages of the ML and Bayes estimates of the QTL parameters from the five designs were close to the true values and are listed in Table 3.

Granddaughter designs used for the detection and estimation of QTL effects associated with different markers differ in the amount of information from data, i.e., in the number of sires and sons per sire (e.g.,

Table 3. Average estimates of the QTL parameters from five granddaughter designs

Parameter	true value	Method of estimation		
		ML	Bayes joint mode	Bayes marginal mean
r	0.20	0.21	0.21	0.19
P	0.57	0.56	0.56	0.56
α (kg)	300.00	311.00	307.0(305.0)[a]	295.0[b]

[a] $\beta = E(\alpha) = 200$ (120) kg, $\alpha_l = 100$ kg, $\alpha_u = 600$ kg
[b] $\beta = 200$ kg, $\alpha_l = 100$ kg, $\alpha_u = 600$ kg

Table 4. Estimates of QTL effect α for particular sires from the granddaughter design in Table 2

Sire	Number of sons	ML (kg)	Bayes joint mode[a]	
			$\beta = 200$ kg (kg)	$\beta = 120$ kg (kg)
1	134	256.0	246.0	239.0
3	123	331.0	320.0	313.0
5	30	348.0	303.0	274.0

[a] $\beta = E(\alpha)$

Dentine and Cowan 1990; Hoeschele and Meinert 1990). To mimic the analysis of several unbalanced granddaughter designs, the simulated data set was split by sires in Table 1. These sires differ in the number of sons, and to accentuate unbalancedness, only the first 30 sons of sire 5 were included. The QTL effect was then estimated separately for each sire in Table 1 by ML and the Bayesian method. Estimates are reported in Table 4 for those sires that were heterozygous at the QTL. Compared to ML, Bayesian estimates were more conservative and closer to the prior mode, which equalled $\alpha_1 = 100$ kg. Sire 5 had the least information, highest ML estimate, and also greatest shrinkage toward the prior information in the Bayesian analysis. In general, the Bayesian procedure should have largest advantage if granddaughter or daughter designs for different markers are small and variable in size.

Discussion

Standard least squares and ML analyses of daughter or granddaughter designs can only partially utilize prior information on marker-QTL linkage, e.g., in determining significance levels for linkage tests or by maximizing the likelihood only over reasonable values of the parameters. However, standard analyses still assume that a priori all values of QTL gene effects, minor and major, are equally likely, and may yield unrealistic estimates. The Bayesian analysis provides a formal framework for incorporating all of the available prior information.

The prior probability of linkage is determined from number and lengths of chromosomes, minimum detectable QTL effect, and total additive genetic variance (Hoeschele and VanRaden 1993). Linkage is declared based on the posterior probability of linkage. This criterion combines information from data with the prior expectation of linkage instead of with the number of other markers that have been or will be tested by the same or other persons.

Bayes estimates of QTL gene effects, α, account for unbalancedness of the data, i.e., different sizes of daughter or granddaughter designs used for different markers, by shrinking the data information toward the mode α_1 or mean $\beta = E(\alpha)$ of a prior exponential distribution of QTL gene effects.

Data to estimate the mean of the exponential distribution of gene effects are not yet available. If sufficient data become available, the mean might be estimated with an empirical Bayesian approach treating the mean as unknown with noninformative prior. One approximate strategy for computing an estimate of β with this approach consists of maximizing the posterior density with respect to the parameters (α, r, p) for a range of β values and storing the resulting density values for any marker analyzed. Once a number of unlinked markers have been analyzed, posterior densities should be multiplied across markers, separately for each value of β. The value of β yielding the largest posterior density across markers would be taken as its point estimate.

Computation of the posterior probability of linkage and of joint posterior modes or marginal posterior means as estimates of the parameters at a marked QTL from data on a typical granddaughter design is affordable.

Acknowledgements. Support of this research by a grant from the Holstein Association, Brattleboro VT, USA is gratefully acknowledged.

References

Berger JO (1985) Statistical decision theory and Bayesian analysis. Springer Berlin Heidelberg New York

Cowan CM, Dentine MR, Ax RL, Schuler LA (1990) Structural variation around prolactin gene linked to quantitative traits in an elite Holstein sire family. Theor Appl Genet 79: 577–582

Dentine MR, Cowan CM (1990) An analytical model for the estimation of chromosome substitution effects in the offspring of individuals heterozygous at a segregating marker locus. Theor Appl Genet 79:775–780

Fernando RL, Grossman M (1989) Marker-assisted selection using best linear unbiased prediction. Genet Sel Evol 21:467–477

Georges M, Gunarwardana A, Threadgill DW, Lathrop M, Olsaker I, Mishra A, Sargeant LL, Schoeberlein A, Steele MR, Terry C, Threadgill DS, Zhao X, Holm T, Fries R, Womack JE (1991) Characterization of a set of variable number of tandem repeat markers conserved in bovidae. Genomics 11:24–32

Goddard ME (1992) A mixed model for analyses of data on multiple genetic markers. Theor Appl Genet 83:878–886

Hoeschele I, Meinert TR (1990) Association of genetic defects with yield and type traits: The Weaver locus has a major effect on yield. J Dairy Sci 73:2503–2515

Hoeschele I, VanRaden PM (1993) Bayesian analysis of linkage between genetic markers and quantitative trait loci: I. Prior knowledge. Theor Appl Genet (in press)

IMSL user's manual (1987) Math/library, Fortran subroutines for mathematical applications, version 1.0

MacLean CJ, Morton NE, Yee S (1984) Combined analysis of genetic segregation and linkage under an oligogenic model. Comp Biomet Res 17:471–480

Neumann PE (1991) Three-locus linkage analysis using recombinant inbred strains and Bayes' theorem. Genetics 128: 631–638

Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE, Lander ES, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato: Comparison across species, generations, and environments. Genetics 127:181–197

Steele MR, Georges M (1991) Generation of bovine multisite haplotypes using random cosmid clones. Genomics 10: 889–904

Weller JI (1990) Experimental designs for mapping quantitative trait loci in segregating populations. In: Hill WG,

Thompson R, Woolliams JA (eds) Joyce Darling, Midlothian, Scotland, Proc 4th World Congr Genet Appl Livst Prod, vol XIII. Edinburgh, pp 113–116

Weller JI, Kashi Y, Soller M (1990) Power of daughter and granddaughter designs for genetic mapping of quantitative traits in dairy cattle using genetic markers. J Dairy Sci 73:2525–2537

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

Appendix

Components of the likelihood function

A.1 Prob$({}^1x_i^Q, {}^2x_i^Q | {}^1x_i^M, {}^2x_i^M)$

Under linkage equilibrium in the population, this conditional probability reduces to the frequency of the genotype ${}^1x_i^Q:{}^2x_i^Q$. Hardy-Weinberg genotype frequencies at a biallelic locus with alleles Q and q are p^2, $p(1-p)$, $p(1-p)$, and $(1-p)^2$ for genotypes qq, qQ, Qq, and QQ, respectively.

Because the model for phenotype [(1) or (2)] depends only on the difference between effects of the two alternative QTL alleles an offspring can inherit from the sire, α, and not on the sire's genotypic effect, genotypic classes of the sires may be reduced to three classes, a homozygous (Q_mQ_m) and two heterozygous classes ($Q_mQ_{m'}$ and $Q_{m'}Q_m$ with $m \neq m'$). Then, class frequencies are $1-P$, probability of homozygosity, and 0.5P for each of the two heterozygous classes, with P, frequency of heterozygosity. For a biallelic QTL, the homozygous class contains the genotypes QQ and qq, and the heterozygous classes correspond to the genotypes Qq and qQ. For a polymorphic QTL, the homozygous class contains all homozygous genotypes or genotypes with alleles of similar effect, and the heterozygous classes include all genotypes with distinguishable effects.

A.2 Prob$(x_{ij}^Q | x_{ij}^M, {}^1x_i^Q:{}^2x_i^Q)$

If the sire is homozygous, i.e., ${}^1x_i^Q = {}^2x_i^Q$, this probability is 1 for $x_{ij}^Q = {}^1x_i^Q$ and 0 for all other x_{ij}^Q. If the sire is heterozygous, the probability is a function of recombination rate r. For a single marker and if, for example, $x_{ij}^M = {}^1x_i^M$, the probability that $x_{ij}^Q = {}^1x_i^Q$ is $1-r$ and the probability that $x_{ij}^Q = {}^2x_i^Q$ is r. For a pair of markers, the probabilities are the same when assuming gene orders Q:M1:M2 or M1:M2:Q with r_1, recombination rate between QTL and M1, and r_2, recombination rate between QTL and M2, replacing r, respectively. For gene order M1:Q:M2, the two nonrecombinant marker haplotypes include different QTL alleles with probability 1 assuming no double recombinants, and each of the two recombinant marker haplotypes includes the alternative QTL alleles with probabilities $1-r_1/r$ and r_1/r.

A.3

The probability density function of an offspring phenotype, conditional on the sire-QTL allele and based on model (1), was obtained with the following distribution assumptions and parameterization:

$$y_{ij}|s_i, \alpha \sim \begin{cases} N(s_i, \sigma^2_{e_{ij}}) \text{ if sire's genotype is } Q_mQ_m \\ N(s_i + x\alpha, \sigma^2_{e_{ij}} - x^2\alpha^2) \text{ if sire's genotype is } Q_mQ_{m'}, \\ \quad m \neq m', \text{ and offspring inherited } Q_m \\ N(s_i - x\alpha, \sigma^2_{e_{ij}} - x^2\alpha^2) \text{ if sire's genotype is } Q_mQ_{m'}, \\ \quad m \neq m', \text{ and offspring inherited } Q_{m'} \end{cases}$$

where $x = 0.5$ or $x = 0.25$ for daughter or granddaughter designs, respectively, and $\sigma^2_{e_{ij}}$ was assumed known. For the granddaughter design, $\sigma^2_{e_{ij}}$ was calculated as $[0.25/\text{Rel(son)} - 0.125 + 0.0625 (1 - \text{Rel(dam)})] \sigma^2_a$ with Rel, reliability of the estimate of total additive genetic value of son ij (VanRaden and Wiggans 1991), or squared correlation between true and estimated breeding value of son ij. For a polymorphic QTL, α was replaced by α_i.