Theor Appl Genet (1993) 85:953–960

THEORETICAL AND APPLIED GENETICS

Bayesian analysis of linkage between genetic markers and quantitative trait loci. I. Prior knowledge

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. Prior information on gene effects at individual quantitative trait loci (QTL) and on recombination rates between marker loci and QTL is derived. The prior distribution of QTL gene effects is assumed to be exponential with major effects less likely than minor ones. The prior probability of linkage between a marker and another single locus is a function of the number and length of chromosomes, and of the map function relating recombination rate to genetic distance among loci. The prior probability of linkage between a marker locus and a quantitative trait depends additionally on the number of detectable QTL, which may be determined from total additive genetic variance and minimum detectable QTL effect. The use of this prior information should improve linkage tests and estimates of QTL effects.

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

Introduction

The number of genetic markers for the dairy cattle genome has increased rapidly in recent years. One laboratory has developed approximately 200 random markers covering an estimated 60–70% of the genome (Georges et al. 1991; Steele and Georges 1991). Random markers include multisite restriction fragment length polymorphism (RFLP) haplotypes, variable number of tandem repeat (VNTR) sequences, and polymerase chain reaction (PCR)-based polymorphisms. Targeted markers include polymorphisms developed in the vicinity of known loci such as the prolactin locus. Linkage between genetic markers and quantitative trait loci (QTL) is likely to be investigated in great detail in the near future.

The granddaughter design (Weller et al. 1990) seems the most practical for linkage analysis in a dairy population because it requires the smallest number of animals to be genotyped. 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).

The fixed effects methods, fixed regression and ML, will overestimate the additive effects associated with selected markers, as noted by Smith and Simpson (1986). Paterson et al. (1991) applied these methods to an RFLP map of the tomato genome to detect and estimate QTL effects on traits measured in F_2 and F_3 tomato crosses. In the F_2, seven, four, and five marked QTL explained 72%, 44%, and 34% of the phenotypic variance of three traits with narrow-sense heritabilities of 45%, 25% and 15%, respectively. The authors explained the high variances attributed to the QTL by upward biases due to nonadditive genetic and genotype by environment interaction variances, but no quantitative evidence of nonadditive genetic variance was given.

In animal breeding, estimates of marker effects from linkage analyses are employed in advertising genetic markers by biotech companies for use by breeders and artificial insemination firms as an aid in selection. The Chro-Mo-Probe with an estimated

Correspondance to: I. Hoeschele

effect of 283 kg milk (Cowan et al. 1990) is an example. Advertised markers will be a selected subset of those initially tested, and estimates of marker effects will be based on different number of families and family sizes. Effects of marked QTL may be estimated more realistically by including prior information, similar to treating sire or animal effects as random instead of fixed in genetic evaluation.

Too many marker-QTL linkages may be declared if many single markers or marker pairs are tested for the same traits. Under the null hypothesis that no detectable QTL are linked to any of the markers, at least one false linkage will be declared with high probability. This has not been considered to be a problem in linkage analysis for a trait known to be monogenic (Edwards et al. 1987), because linkage is declared only for the most significant marker. This argument does not apply to polygenic traits (Clerget-Darpoux et al. 1990), because many linkages can be declared.

The approach of Lander and Botstein (1989) achieves a small type-I error rate (e.g., 0.05) for the entire set of linkage tests and requires very small type-I error rates for individual tests, depending on the number and density of markers. The prolactin locus was found to have a significant effect of 283 kg milk in one Holstein sire family (Cowan et al. 1990) at a type-I error rate of 0.05. This effect would not have been declared significant at a type-I error probability of, e.g., $0.0005 = 0.05/100$, derived from Lander and Botstein's approach and a large number of genetic markers (e.g., 100) now available for dairy cattle. Dekkers and Dentine (1991) showed that even for a strictly polygenic trait detectable chromosome substitution effects exist. Hence, given prior knowledge of sufficient heritability or additive genetic variance, the null hypothesis of no detectable QTL segregating appears inappropriate.

An alternative approach uses the concept of prior probability of linkage, which is a function of the number of detectable QTL instead of number of potential markers. The more detectable QTL that are segregating in a population, the larger the probability of linkage of any given marker or marker pair to at least one QTL. This approach allows the utilization of prior knowledge on the genetic base of quantitative or complex traits, with the number of detectable QTL depending on total additive genetic variance, which is well estimated for important traits of livestock.

In this paper, prior distributions of QTL substitution effects, QTL gene or genotype frequencies, and recombination rate between a marker locus and a QTL are presented. Subsequently, the prior probability of linkage between a single or a pair of marker loci and a QTL is derived. In a companion paper (Hoeschele and VanRaden 1993), this prior knowledge will be combined with data from daughter or granddaughter designs in a Bayesian analysis to compute the posterior probability of linkage between a marker locus and a QTL and to estimate QTL parameters.

Prior knowledge

Additive effects of QTL

Several hypothesis have been formed about the inheritance of quantitative traits. Most of the limited empirical evidence supports the hypothesis that quantitative traits are under the control of few major genes and numerous genes with smaller effects (Gregory 1965; Spickett and Thoday 1966; Thompson 1975; Edwards et al. 1987; Paterson et al. 1988, 1991; Shrimpton and Robertson 1988). Shrimpton and Robertson (1988) provided a frequency histogram for effects of genes at loci affecting bristle number in *Drosophila melanogaster*. They conjectured a hypothetical frequency distribution of bristle gene effects, which may be modelled using the exponential distribution.

In populations undergoing selection, alleles of large effects are likely to be fixed. However, relatively large effects have been found in association with the prolactin (Cowan et al. 1990) and weaver (Hoeschele and Meinert 1990) loci in dairy cattle. These effects may be due to single QTL or QTL clusters. Clusters of QTL or superloci may be treated as single QTL. On the assumption that the exponential distribution adequately describes the distribution of QTL effects, the additive or substitution effect α of a biallelic QTL has the prior density function

$$f(\alpha) = 1/\beta * \exp(-\alpha/\beta)$$

where β is the only parameter of the exponential distribution,

$$\beta = E(\alpha).$$

Imposition of an upper limit on α, α_u, avoids integration over a range of large α values with near zero density. As a lower limit, α_l, the smallest α likely to be detected in a given design may be chosen. Imposition of a lower limit leaves all QTL with smaller effects in the polygenic part of an individual's additive genetic value. The truncated exponential distribution has a positive density over the range α_l to α_u equal to

$$f(\alpha) = c/\beta * \exp(-\alpha/\beta) \quad (1)$$

where $c = 1/[\exp(-\alpha_l/\beta) - \exp(-\alpha_u/\beta)]$ is an integration constant.

The assumption of a biallelic QTL may serve as a good approximation if multiple alleles can be grouped into favorable and unfavorable with favorable or un-

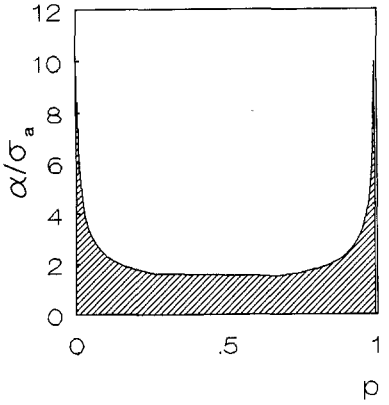

Fig. 1. Joint permissible parameter space for QTL gene frequency p and gene effect α expressed in total additive genetic standard deviation units

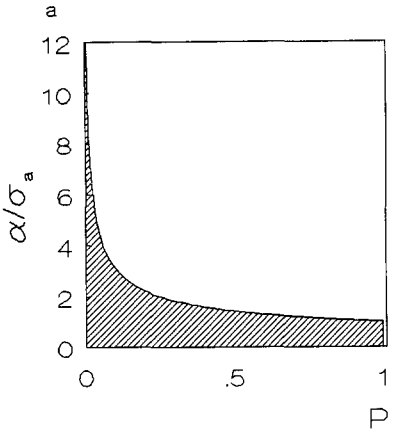

Fig. 2. Joint permissible parameter space for QTL frequency of heterozygosity P and gene effect α expressed in total additive genetic standard deviation units

favorable alleles having similar effects. Alternatively, the QTL may be assumed to be highly polymorphic, with each sire carrying different QTL alleles. Numbers of QTL alleles intermediate between two and many would be difficult to model, because the exact number of alleles is unknown and alleles of similar effects would be difficult to distinguish. If each sire is assumed to carry different QTL alleles, additive effects can be treated as nested within sires and modelled as independent a priori, or

$$f(\boldsymbol{\alpha}) = \prod_{i=1}^{S} f(\alpha_i) \tag{2}$$

where $\boldsymbol{\alpha}$ is the vector of α_i effects, S is number of sires, and the α_i are independently and identically exponentially distributed with density as in (1).

QTL allele or genotype frequencies

Absence of prior information about allele frequency p at a biallelic QTL may be expressed by letting p be distributed as beta Be(1, 1), which is the uniform distribution U(0, 1). This prior implies that all values of p are equally likely a priori. The additive genetic variance at a biallelic QTL, equal to $2p(1-p)\alpha^2$ (Falconer 1985), may exceed the total additive genetic variance σ_a^2 for certain values of p if α is large. Therefore, the joint density of α and p should be positive only for those combinations of α and p values yielding a QTL variance less than or equal to σ_a^2. The joint parameter space containing all α and p values with positive density is depicted in Fig. 1. The prior density for the truncated, joint distribution of α and p may be written as

$$f(\alpha, p) = \begin{cases} k * f(\alpha) & \text{if } 2p(1-p)\alpha^2 \le \sigma_a^2 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

with k, the reciprocal of the integral of $f(\alpha,p)$ over the restricted space of α and p values, or

$$k = 1 \Big/ \left\{ \int_{\alpha_l}^{\alpha_u} f(\alpha) \left[\int_0^{p_\alpha} f(p)dp + \int_{(1-p_\alpha)}^{1} f(p)dp \right] d\alpha \right\}$$

where p_α is the smaller root of the equation $\sigma_a^2 - 2p(1-p)\alpha^2 = 0$ if $0.5\alpha^2 > \sigma_a^2$, and $p_\alpha = 0.5$ otherwise; the prior density f(p) is Be(1,1) or, equivalently, U(0, 1).

For a multiallelic QTL, the Mendelian sampling variance in the population of sires is

$$\sigma^2_{m(QTL)} = (0.25)\sum_k \sum_l p_k p_l \alpha^2_{kl}$$

where p_k and p_l are allelic frequencies at the QTL in the sires, α_{kl} is the effect of substituting QTL allele k for l with $\alpha_{kk} = 0$, and the summation extends over all alleles at the QTL. This expression represents the total single parent Mendelian segregation variance at a QTL in a population in linkage equilibrium (Dekkers and Dentine 1991). Assuming that there are many alleles and $p_k = p$ for all k,

$$\sigma^2_{m(QTL)} = P(0.25)\,\overline{\alpha^2} \tag{4}$$

where

$$P = \sum_{k \neq l} \sum p_k p_l = p^2 * h$$

with h, number of heterozygotes, and α^2, average of all α^2_{kl} with $k \neq l$. The parameter space containing all combinations of $\bar{\alpha}$ and P values yielding a value of (4) equal to or less than the total sire Mendelian sampling variance or $0.25\sigma_a^2$ is shown in Fig. 2. The prior density for the truncated, joint distribution of α and P may then be written as

$$f(\alpha, P) = \begin{cases} k * f(\alpha) & \text{if } (4) \le 0.25\sigma_a^2 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

with k, the reciprocal of the integral of $f(\alpha, P)$ over the restricted space of α and P values.

Recombination rate

The prior distribution of recombination rate r between two loci may be modelled as a discrete and a continuous part, similar to the prior in the Bayesian analysis of a general point null hypothesis (Berger 1985). This prior assigns a certain probability mass to $r = 0.5$, the probability of nonlinkage Prob($r = 0.5$), while giving $r < 0.5$ the continuous density $[1 - \text{Prob}(r = 0.5)] * f(r)$.

This type of prior was first employed by Smith (1953, 1959) in his Bayesian method of estimating recombination frequency from data consisting of recombinant counts. Renwick (1971) and Elston and Lange (1975) used a similar prior for the distance between loci. A rederivation of the prior for recombination rate among two arbitrary loci is presented in this section preceding an extension to quantitative traits.

Let M be a marker locus and Q another locus with known genotypes but unknown location. A priori, M and Q can be located anywhere in the genome. Let C_M and C_Q denote chromosomes carrying M and Q, respectively. Given that M and Q are on the same chromosome, or $C_M = C_Q$, let δ be the genetic distance between M and Q, and δ_{min} the minimum distance at which the loci segregate independently, i.e., $r = 0.5$ for $\delta \geq \delta_{min}$. Then, the prior distribution of r may be represented as

$$\text{Prior}(r) = \begin{cases} \text{Prob}(r = 0.5) = 1 - \text{Prob}(C_M = C_Q \text{ and } \delta < \delta_{min}) \\ [1 - \text{Prob}(r = 0.5)] * f(r) \end{cases} \quad (6)$$

Evaluation of this prior requires determining $\text{Prob}(C_M = C_Q)$, $\text{Prob}(\delta < \delta_{min})$, and $f(r)$, which are given in Appendix A. The minimum distance for independent segregation, δ_{min}, can be obtained from an assumed map function $\delta = g(r)$ (Appendix D).

Assuming that all chromosomes have length L = 1 Morgan and combining (A.1) and (A.6) yields an approximate prior probability of nonlinkage, or

$$\text{Prob}(r = 0.5) = 1 - (1/n_c) * (2\delta_{min} - \delta_{min}^2) \quad (7)$$

where n_c is number of chromosomes.

Allowing chromosomes to differ in length, using (A.2), and letting $\text{Prob}(\delta < \delta_{min})$ depend on the length of individual chromosomes yields the prior probability of nonlinkage, or

$$\text{Prob}(r = 0.5) = 1 - \sum_{i=1}^{n_c} (L_i/L_T)^2 \, \text{Prob}(\delta < \delta_{min} | L_i) \quad (8)$$

where

$$\text{Prob}(\delta < \delta_{min} | L_i) = \begin{cases} 1 \text{ if } L_i \leq \delta_{min} \\ (2/L_i)\delta_{min} - (1/L_i)^2 \delta_{min}^2 \text{ otherwise} \end{cases}$$

and L_i and L_T are the genetic lengths of chromosome i and the genome, respectively. The same result was obtained by Elston and Lange (1975) and generalized to more than one marker. Let $\delta_{min} = 1$ Morgan. Then, $\text{Prob}(r = 0.5) = 0.955$ from (7) for the human genome with 22 autosomal chromosomes, and $\text{Prob}(r = 0.5) = 0.956$ from (8). Appendix D gives $\text{Prob}(r = 0.5)$ and $f_L(r)$ for the map function assuming complete interference (Haldane 1919), $\delta = r$, and for the function of Pascoe and Morton (1987), which fitted recombination data on both *Drosophila* and humans well.

The prior probability of linkage between a locus Q and a pair of linked marker loci M1 and M2 located on chromosome C_M is equal to $\text{Prob}(C_M = C_Q)$ times the probability of Q linked to M1 or M2, given that all three loci are on the same chromosome of length L. The latter probability can be partitioned into three components, Prob(Q:M1:M2), Prob(M1:Q:M2), Prob(M1:M2:Q), which are the probabilities of Q located on the flank closest to M1 and linked to M1, of Q located between M1 and M2, and of Q located on the flank closest to M2 and linked to M2, respectively. These prior probabilities and prior density functions of recombination rate for the different linkage situations are given in Appendix B.

Table 1. Number of detectable QTL (n_Q) for fractions F of additive genetic variance σ_a^2 explained by QTL, different parameter values of the prior distribution of QTL gene effects, α; number of sires, S; and minimum gene frequency, p_l. $\sigma_a = 560$ kg milk yield

F	$\alpha_l, \alpha_u, \beta$[a]	S	p_l[b]	n_Q
1.0	100, 600, 200	2	0.25	8
		10	0.05	10
	100, 600, 120	2	0.25	12
		10	0.05	15
	50, 600, 200	2	0.25	11
		10	0.05	13
	50, 600, 100	2	0.25	22
		10	0.05	28
0.5	100, 600, 200	2	0.25	4
		10	0.05	5
	100, 600, 120	2	0.25	6
		10	0.05	8
	50, 600, 200	2	0.25	4
		10	0.05	6
	50, 600, 100	2	0.25	11
		10	0.05	14
0.33	100, 600, 200	2	0.25	3
		10	0.05	4
	100, 600, 120	2	0.25	4
		10	0.05	5
	50, 600, 200	2	0.25	4
		10	0.05	5
	50, 600, 100	2	0.25	8
		10	0.05	10

[a] $\alpha_l \leq \alpha \leq \alpha_u$; $E(\alpha) = \beta$, values in kg milk yield
[b] $p_l \leq p \leq (1 - p_l)$, $p_l = 1/(2S)$

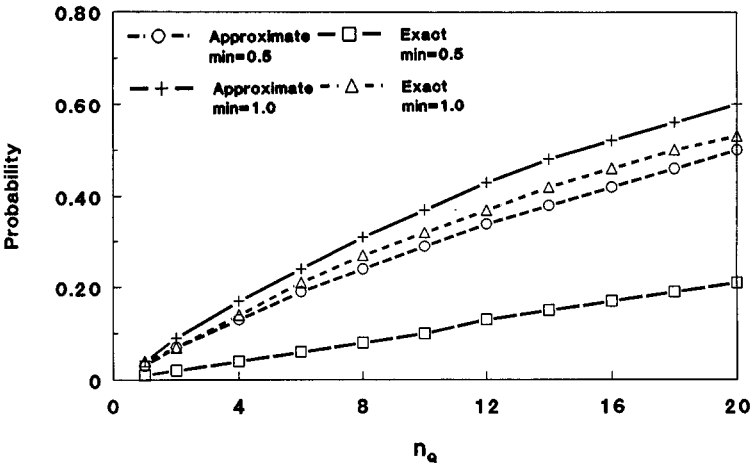

Fig. 3. Prior probability of linkage as a function of number of detectable QTL (n_Q) map function determining minimum distance between syntenic loci for independent segregation, δ_{min}, and approximate (constant chromosome length) versus exact computation for the human genome

Joint prior knowledge about QTL additive effect, allele frequency, and recombination

A prior for recombination frequency r was given in (6) for a single marker or marker pair and a single trait locus. For a quantitative trait, there may be several QTL with detectable effects, which would increase the prior probability of linkage. If the number of detectable QTL, n_Q, were known a priori, the prior probability of linkage could be obtained by straightforward extension of the results in the previous section.

A range for number of detectable QTL can be obtained for biallelic loci by setting the fraction F of the total additive genetic variance explained by detectable QTL equal to n_Q times the expected variance due to one detectable QTL, or

$$F\sigma_a^2 = n_Q E(V_Q) \tag{9}$$

where

$$E(V_Q) = k\int_{\alpha_l}^{\alpha_u} f(\alpha)\left\{\int_0^{p_\alpha}[2p(1-p)\alpha^2]dp + \int_{1-p_\alpha}^{1}[2p(1-p)\alpha^2]dp\right\}d\alpha$$

From these two equations, n_Q can be computed for different values of F and of the parameters of the prior density of α and p. Because allelic frequency at a QTL segregating in a small number of sires cannot closely approach zero or one, lower (p_l) and upper (p_u) limits may replace the integration limits for p of 0 and 1 in $E(V_Q)$. With number of sires denoted by S, $p_l = 1/(2S)$ and $p_u = 1 - p_l$. Then, performing the inner integration in $E(V_Q)$ yields

$$E(V_Q) = k\int_{\alpha_l}^{\alpha_u} \alpha^2[2(p_\alpha^2 - p_l^2) - 4/3(p_\alpha^3 - p_l^3)]f(\alpha)\,d\alpha$$

In evaluating $E(V_Q)$, the condition $p_\alpha > p_l$ must be satisfied, else $E(V_Q)$ is set to zero. The numbers of detectable QTL are given in Table 1 for a typical quantitative trait, milk yield, and then suggest a likely range of $n_Q = 5$ to $n_Q = 30$.

The prior probability of nonlinkage, for a given value of n_Q and from results (C.2) and (C.3) in Appendix C, is

$$\text{Prob}(r = 0.5) = 1 - \left[\sum_{i=1}^{n_c}(L_i/L_T)\sum_{j=1}^{n_Q}\binom{n_Q}{j}(L_i/L_T)^j \cdot(1 - L_i/L_T)^{(n_Q - j)} * (C.3)\right] \tag{10}$$

By assuming a constant length of chromosomes, L, letting $\delta_{min} > L/2$, and using (C.1), (12) may be approximated by

$$\text{Prob}(r = 0.5) = 1 - \sum_{j=1}^{n_Q}\binom{n_Q}{j}(1/n_c)^j(1 - 1/n_c)^{(n_Q - j)} \cdot[1 - (2/(j+1))(L - \delta_{min})^j] \tag{11}$$

The prior probability of linkage can now be evaluated for a range of values for number of detectable QTL, n_Q, using (10) or (11). Figure 3 shows the prior probability of linkage as a function of n_Q, exact [(10)] versus approximate [(11)] calculation, and map function determining δ_{min}. The prior probabilities were computed for the human genome with 22 autosomal chromosomes of genetic lengths given in Ott (1985). For n_Q ranging from 1 to 20, prior probabilities of linkage increased from 1% to 21% and from 4% to 53% for the map function with complete interference (Haldane 1919) and that of Pascoe and Morton (1987), respectively. Substantial differences in the prior probability of linkage were also observed among map functions and exact versus approximate calculations.

The joint prior distribution of QTL parameters may now be represented as

$$\text{Prior}(\alpha, p, r) = \begin{cases}\text{Prob}(r = 0.5) = (10) \\ [1 - \text{Prob}(r = 0.5)] * f(\alpha, p) * f(r)\end{cases} \tag{12}$$

Prior (12) is an extension of (6) for the case of linkage to a QTL with parameters α, p, and r.

Conclusions

Prior information on recombination rates between loci and on the distribution of QTL gene effects is available. The prior probability that a randomly chosen marker or marker pair is linked to any QTL is a function of lengths of chromosomes, map function relating distance to recombination rate among loci, and number of QTL. The number of detectable QTL may be determined from total additive genetic variance and minimum detectable QTL effect. Prior probability of linkage of a marker locus to at least one QTL ranged from 1% to 53% for the human genome using two alternative map functions with 1–20 detectable QTL present.

An exponential prior distribution of QTL gene effects was assumed, with minor effects more likely than major effects. The distribution depends only on the average gene effect, which is unknown. Data to estimate this mean are not yet available. However, if the distribution of single gene effects on quantitative traits of livestock is truly exponential, then a good guess of the mean should produce a prior closer to reality than the assumption of minor and major effects being equally likely. All effects being equally likely is assumed in fixed regression or ML linkage analysis. A Bayesian linkage analysis presented in a subsequent paper allows the available prior information to be combined with experimental evidence from data.

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

Appendix A. Components of the prior for recombination rate between one trait and one marker locus

Consider first $\text{Prob}(C_M = C_Q)$ and assume locus Q can be situated anywhere on the genome with equal probability. Then, if all chromosomes have equal length, the prior probability that M and Q are on the same chromosome is

$$\text{Prob}(C_M = C_Q) = 1/n_c \tag{A.1}$$

where n_c is number of chromosomes. For $n_c = 30$, $\text{Prob}(C_M = C_Q) = 0.33$. If chromosomes differ in length with L_T, the length of the genome, and L_i, the length of chromosome i ($i = 1,\ldots,n_c$), then the prior probability of M and Q located on the same chromosome is

$$\text{Prob}(C_M = C_Q) = \sum_{i=1}^{n_c} (L_i/L_T)^2 \tag{A.2}$$

For the human genome, estimates of the physical and genetic lengths of the 22 autosomal chromosomes are available (Ott 1985, p 11). Applying (A.1) and (A.2) yields values of 0.046, 0.049, and 0.052 for equal length, physical lengths, and genetic lengths, respectively. If C_M and its length were known,

$$\text{Prob}(C_M = C_Q) = L_{C_M}/L_T \tag{A.3}$$

The $\text{Prob}(\delta < d_{min})$ and the density f(r) are obtained by integrating the density $f(\delta)$ and by a change of variable from δ to r in $f(\delta)$, respectively. Let l_M and l_Q be random variables representing the location of M and Q given $C_M = C_Q$. Assuming M and Q can be located anywhere on the chromosome of length L Morgans with equal probability implies $l_M \sim U(0, L)$ and $l_Q \sim U(0, L)$. The change of variables from (l_M, l_Q) to (l_M, δ) and integration with respect to l_M yields the prior density of δ with parameter L, or

$$f_L(\delta) = \frac{4}{L^2}\max\left(0, \frac{L}{2} - \delta\right) + \frac{2}{L^2}\min(\delta, L - \delta) \tag{A.4}$$

It follows that

$$\begin{aligned}\text{Prob}(\delta \geq \delta_{min}) = \int_{\delta_{min}}^{L} f_L(\delta)\,d\delta &= \frac{4}{L^2}\int_{\delta_{min}}^{L/2}\left(\frac{L}{2} - \delta\right)d\delta \\ &+ \frac{2}{L^2}\int_{\delta_{min}}^{L/2}\delta\,d\delta + \frac{2}{L^2}\int_{\max(\delta_{min}, L/2)}^{L}(L - \delta)\,d\delta\end{aligned} \tag{A.5}$$

After some algebra,

$$\text{Prob}(\delta < \delta_{min}) = \frac{2}{L}\delta_{min} - \frac{1}{L^2}\delta_{min}^2. \tag{A.6}$$

Let $\delta = g(r)$ represent any explicit map function. A change of variable from δ to r in (A.4) yields the prior density of r for $r < 0.5$,

$$f_L(r) = f_L(g(r))\,|d\delta/dr|/\text{Prob}(\delta < d_{min}) \tag{A.7}$$

If chromosomes differ in length and M has not been assigned to a specific chromosome, the prior density of δ in (A.4) must be replaced by the density of a mixture distribution, or

$$f(\delta) = \sum_{i=1}^{n_c} \text{Prob}(C_M = C_Q = i) f_{L_i}(\delta) \tag{A.8}$$

where the density in the right-hand-side of (A.8) is that of (A.4) with $L = L_i$. A simple approximation to f(r) is to evaluate $f_L(r)$ in (A.7) at L equal to the average length of chromosomes. An alternative approximation is given in Elston (1975).

Appendix B. Components of the prior for recombination rate between one trait and two linked marker loci

Continue to assume that locus Q can be situated anywhere on the genome with equal probability. Let Δ be the map distance between M1 and M2. Then,

$$\text{Prob}(M1{:}Q{:}M2) = \Delta/L \tag{B.1}$$

and the prior density of r_1/r_M, the ratio of recombination rate among M1 and Q to recombination rate among M1 and M2, may be taken as the density of U(0, 1), given that Q is between M1 and M2.

The probability that Q is not between M1 and M2 is $(L - \Delta)/L$ from (B.1). The probability that locus Q is on the flank closer to M1, given Q is not between M1 and M2, is 0.5. Given that Q is on the flank closer to M1, the probability that the distance between Q and M1 is less than δ_{min} can be computed by assuming the locus order $0 \leq l_Q \leq l_{M1} < l_{M2} \leq L$ as

$$\begin{aligned}&\text{Prob}((l_{M1} - l_Q) < \delta_{min} \mid 0 \leq l_Q \leq l_{M1} < l_{M2} \leq L) = \delta_{min}/(L - \Delta) \\ &+ \int_{\delta_{min}}^{L-\Delta} 1/(L - \Delta) \int_{l_{M1} - \delta_{min}}^{l_{M1}} (1/l_{M1})\,dl_Q\,dl_{M1} \\ &= \delta_{min}/(L - \Delta)\,[1 + \log\{(L - \Delta)/\delta_{min}\}]\end{aligned}$$

Multiplication of the probabilities of Q not between M1 and M2, on the flank closer to M1 given not between to M1 and M2, and linked to M1 yields

$$\text{Prob}(Q{:}M1{:}M2) = [\delta_{min}/(2L)][1 + \log\{(L-\Delta)/\delta_{min}\}] \quad (B.2)$$

Probability Prob(M1:M2:Q) is identical to Prob(Q:M1:M2) in (B.2), and the prior probability of no linkage to either M1 or M2 is

$$\text{Prob}(r = 0.5) = (L - \Delta - \delta_{min}[1 + \log\{(L-\Delta)/\delta_{min}\}])/L \quad (B.3)$$

The prior density of r_1 or r_2, recombination rate between Q and M1 or M2, respectively, given linkage to either M1 or M2, is obtained by using (B.2) and $f(\delta) = 1/(L-\Delta)$ in Eq. A.7.

Appendix C. Components of the prior probability of nonlinkage between n_Q trait loci and one marker or marker pair

QTL are assumed to be randomly and independently distributed across the genome. Then, for a given value of n_Q, the probability that at least one QTL is located on the same chromosome as the marker or marker pair is, for equal length of chromosomes,

$$\text{Prob}(C_M = C_Q) = \sum_{j=1}^{n_Q} \binom{n_Q}{j} (1/n_c)^j (1 - 1/n_c)^{(n_Q - j)} \quad (C.1)$$

and, for unequal length of chromosomes,

$$\text{Prob}(C_M = C_Q) = \sum_{i=1}^{n_c} (L_i/L_T) \sum_{j=1}^{n_Q} \binom{n_Q}{j} (L_i/L_T)^j (1 - L_i/L_T)^{(n_Q - j)} \quad (C.2)$$

The probability $\text{Prob}(\delta < \delta_{min})$ now represents the probability that the distance between the marker and the closest QTL, δ_s, is less than δ_{min}. Elston and Lange (1975) derived this probability for one trait locus and several marker loci. Following their procedure, let the marker and some QTL be situated on chromosome C_i, let the position of the marker on the chromosome be l_M, and let $l'_M = l_M/L_i$ and $\delta'_{min} = \delta_{min}/L_i$. Then, conditional on l_m,

$$\text{Prob}(\delta_s > \delta_{min} | \text{j QTL on } C_i) = \begin{cases} (1 - l'_M - \delta'_{min})^j & \text{if } \delta_{min} > L_i/2 \text{ or } \delta_{min} < L_i/2 \text{ and } l_M < \delta_{min} \\ (1 - 2\delta'_{min})^j & \text{if } \delta_{min} < L_i/2 \text{ and } l_M > \delta_{min} \end{cases}$$

for $j = 1,\ldots,n_Q$. Integration with respect to l'_M yields, after some algebra,

$$\text{Prob}(\delta_s < \delta_{min} | \text{j QTL on } C_i) = \begin{cases} 1 - (2/(j+1)) * (1 - \delta'_{min})^{j+1} \text{ if } \delta_{min} > L_i/2 \\ 1 - (2/(j+1)) * (1 - \delta'_{min})^{j+1} - ((j-1)/(j+1)) * (1 - 2\delta'_{min})^{j+1} \text{ if } \delta_{min} < L_i/2 \end{cases} \quad (C.3)$$

Appendix D. Map functions applied to the prior for recombination rate

(D.1) Map function: $\delta = g(r) = r$ if $\delta < 0.5$; $r = 0.5$ if $\delta > 0.5$ (Haldane 1919).

When $\delta_{min} = 0.5$, $n_c = 30$, and $L = 1$ for all chromosomes, the use of Eqs. A.3, A.4, A.6, and A.7 yields

$$\begin{aligned} \text{Prob}(r = 0.5) &= 1 - \text{Prob}(\delta < \delta_{min}) * \text{Prob}(C_M = C_Q) \\ &= 1 - [1/L - 1/(4L^2)] * \text{Prob}(C_M = C_Q) \\ &= 0.975; \\ f_L(r) &= (0.75)^{-1} * (2 - 2r). \end{aligned}$$

(D.2) Map function: $\delta = g(r) = \sqrt{3/6} * \arctan[(1 + 4r)/\sqrt{3}] - 1/12 * \ln[(1-2r)^2] - 1/12 * \ln(1 + 2r + 4r^2) - 0.15115$ (Pascoe and Morton 1989).

When $\delta_{min} \approx 1.$, $n_c = 30$ and $L = 1$ for all chromosomes, the use of (A.3), (A.4), (A.6), and (A.7) yields

$$\begin{aligned} \text{Prob}(r = 0.5) &= 1 - [2/L - 1/L^2] * \text{Prob}(C_M = C_Q) \\ &= 0.967; \\ f_L(r) &= [4\max\{0, 1 - g(r)\} + 2\min\{g(r), 1 - g(r)\}]/(1 - 8r^3). \end{aligned}$$

References

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

Clerget-Darpoux F, Babron MC, Bonaiti-Pellie C (1990) Assessing the effect of multiple linkage tests in complex diseases. Genet Epidem 7:245–253

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

Dekkers JCM, Dentine MR (1991) Quantitative genetic variance associated with chromosomal markers in segregating populations. Theor Appl Genet 81:212–220

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

Edwards MD, Stuber CW, Wendel JF (1987) Molecular-marker-facilitated investigation of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116:113–125

Elston RC (1975) An approximation for the prior probability of autosomal linkage. HGM 2:290–292

Elston RC, Lange K (1975) The prior probability of autosomal linkage. Ann Hum Genet 38:341–350

Falconer DS (1985) Introduction to quantitative genetics, 2nd edn. Longman Scientific and Technical, New York

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

Gregory WC (1965) Mutation frequency, magnitude of change, and the probability of improvement in adaptation. Radiat Bot 5 [Suppl]:429–441

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

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. II. Combining prior knowledge with experimental evidence. Theor Appl Genet (in press)

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Ott J (1985) Analysis of human genetic linkage. The Johns Hopkins University Press, Baltimore, Md.

Pascoe L, Morton NE (1987) The use of map functions in multipoint mapping. Am J Hum Genet 40:174–183

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphism. Nature 335:712–726

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

Renwick JH (1971) The mapping of human chromosomes. Annu Rev Genet 5:81–120

Shrimpton AE, Robertson A (1988) The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. II. Distribution of third chromosome bristle effects within chromosome sections. Genetics 118:445–459

Smith CAB (1953) The detection of linkage in human genetics. J R Stat Soc Ser B 15:153–184

Smith CAB (1959) Some comments on the statistical methods used in linkage investigations. Am J Hum Genet 11:289–304

Smith C, Simpson SP (1986) The use of genetic polymorphisms in livestock improvement. J Anim Breed Genet 103: 205–217

Spickett SG, Thoday JM (1966) Regular responses to selection. 3. Interaction between located polygenes. Genet Res 7: 96–121

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

Thompson JN (1975) Quantitative variation and gene number. Nature 258:665–668

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