Theor Appl Genet (1994) 88:548–556

H. Simianer

Derivation of single-locus relationship coefficients conditional on marker information

Received: 12 September 1993 / Accepted: 12 October 1993

Abstract The coefficient of relationship is defined as the correlation between the additive genetic values of two individuals. This coefficient can be defined specifically for a single quantitative trait locus (QTL) and may deviate considerably from the overall expectation if it is taken conditional on information from linked marker loci. Conditional halfsib correlations are derived under a simple genetic model with a biallelic QTL linked to a biallelic marker locus. The conditional relationship coefficients are shown to depend on the recombination rate between the marker and the QTL and the population frequency of the marker alleles, but not on parameters of the QTL, i.e. number and frequency of QTL alleles, degree of dominance etc., nor on the (usually unknown) QTL genotype of the sire. Extensions to less simplified cases (multiple alleles at the marker locus and the QTL, two marker loci flanking the QTL) are given. For arbitrary pedigrees, conditional relationship coefficients can also be derived from the conditional gametic covariance matrix suggested by Fernando and Grossman (1989). The connection of these two approaches is discussed. The conditional relationship coefficient can be used for marker-assisted genetic evaluation as well as for the detection of QTL and the estimation of their effects.

Key words Relationship coefficients · Genetic markers · Quantitative trait loci

Introduction

The concept of coefficients of inbreeding and relationship was introduced by Wright (1922) into quantitative genetics. The coefficient of relationship of two individuals x and y, R_{xy}, is defined as the correlation between the additive genetic values of the two individuals. There is also a probabilistic interpretation, since under panmictic conditions, R_{xy} equals twice the coefficient of coancestry, A_{xy} (Malécot 1948), which is the probability that a random gene in individual x is identical by descent to random gene in individual y (Wright 1969).

These coefficients are defined over the whole genome, i.e. on average for all loci or, equivalently, for a randomly sampled locus out of the population of all loci in the genome. It is also possible to define these coefficients specifically for a single locus. The relationship at this single locus may deviate substantially from the overall expectation if additional information can be used. This additional information can be the knowledge of genetic markers that are linked to the locus considered. If, say, two halfsibs have obtained the same marker allele from their sire, the probability that they have also obtained the same allele at a quantitative trait locus (QTL) linked to this marker locus is increased, and the expected coefficient of relationship for the QTL will be larger than 0.25.

Fernando and Grossman (1989) have suggested an algorithm to set up a gametic relationship matrix for a QTL conditional on information from one linked marker, which can be readily extended to account for several independent marker-QTL pairs (Fernando and Grossman 1989; Cantet and Smith 1991). Multiple dependent markers were considered by Goddard (1992). It will be shown later how to extract conditional coefficients of relationship from these gametic relationship matrices.

The approach presented here follows a different concept in that it derives the conditional correlation of the genetic values at the QTL considered rather than the variances and covariances of the two respective gametic values. The derivation will be given in detail for a simple case; the conditional halfsib correlation at a biallelic QTL with one linked biallelic marker. Extensions to less simplified situations and possible applications will be indicated afterwards.

Communicated by C. Smith

H. Simianer
Department of Animal Husbandry and Animal Breeding, Animal Genetics Group, University of Hohenheim (470/HG), D-70593 Stuttgart, Federal Republic of Germany

The genetic model

Consider a QTL and a marker locus, both located on the same autosomal chromosome. For simplicity and, as will be shown later, without loss of generality, only two alleles are assumed at both loci: Q and R at the QTL and A and B at the marker locus. Inheritance at the marker locus is assumed to be codominant, i.e. the genotype of an individual can be perfectly identified.

Let the population frequency of alleles Q and R be q and $(1-q)$ and the frequency of alleles A and B be m and $(1-m)$. Following the concept of genotypic values (Falconer 1989), homozygous QQ individuals have the genotypic value $\mu + a$, homozygous RR individuals have the value $\mu - a$ and heterozygous QR or RQ individuals have the value $\mu + d$. The two alleles at the marker locus are assumed to be neutral with respect to the quantitative trait considered. The distance between the QTL and the marker locus is expressed as recombination rate r, with $r = 0$ indicating complete linkage and $r = 0.5$ indicating free recombination.

In addition to this one locus affecting the quantitative trait, there may be an infinite number of loci that contribute to the phenotype by small additive effects. Their additive contribution (excluding the above described QTL) is summarized in the polygenic additive genetic variance σ_a^2. All these loci are assumed to be unlinked to both the marker locus and the QTL. In the further derivation, progeny groups of sires which are neither inbred nor related to each other and are mated to a random sample of dams are considered. With such a halfsib structure, all interactions between or within loci can be included in the residual error variance with the exception of the additive × additive interaction between two or more loci, which will be ignored. The population is assumed to be in Hardy-Weinberg equilibrium at all loci and no linkage disequilibrium between loci is assumed.

Marker information in a progeny test design

Let us consider a sire that is heterozygous AB at the marker locus. If the marker locus and the QTL are linked. i.e. $r < 0.5$, two offspring with marker genotype AA are on average more similar than one offspring with marker genotype AA and another one with marker genotype BB. This is so because the AA individuals both have obtained the same marked chromosome segment and, therefore, with probability $(1-r)^2 > 0.25$, the same QTL-allele from the father, while the BB offspring has inherited the other marked chromosome segment and, therefore, only with probability $(1-r)r < 0.25$ the same QTL-allele as the offspring with marker genotype AA.

Given a sire is heterozygous at the marker locus, he carries with probability $q(1-q)$ allele A and Q on one chromosome and B and R on the other. With the same conditional probability, the linkage phase may be A–R and B–Q, but let us restrict the considerations to the former case. With a given recombination rate r, gametes A–Q and B–R are produced with probability $P(A–Q) = P(B–R) = 0.5(1-r)$, and gametes A–R and B–Q are produced with probability $P(A–R) = P(B-Q) = 0.5r$. If this sire is mated to females that are randomly sampled from a population that is in Hardy–Weinberg equilibrium, the respective probabilities for the female gametes are $P(A–Q) = mq$, $P(A–R) = m(1-q)$, $P(B–Q) = (1-m)q$ and $P(B–R) = (1-m)(1-q)$. In Table 1, the 16 possible offspring genotypes and their expected frequencies are given.

If we now look at the offspring with marker genotype AA in the upper left 2 × 2 part of Table 1, their expected proportion in the whole offspring group is $0.5m$. Given this marker genotype, the conditional probabilities of QTL genotypes QQ, QR and RR are $(1-r)q$, $(1-r)(1-q)+rq$ and $r(1-q)$, respectively.

Let M_i^s be the marker genotype of the sire (indicated by the superscript s) with three different categories: $M_1^s = AA$, $M_2^s = AB$, and $M_3^s = BB$. Analogously, let M_i^o

Table 1 Genotypes and genotype frequencies arising under the assumed mating scheme for a sire with genotype A–Q/B–R

Paternal gamets	Frequency	Maternal gemetes			
		A–Q	A–R	B–Q	B–R
		mq	$m(1-q)$	$(1-m)q$	$(1-m)(1-q)$
A–Q	$0.5(1-r)$	A–Q A–Q $0.5(1-r)mq$	A–R A–Q $0.5(1-r)m(1-q)$	B–Q A–Q $0.5(1-r)(1-m)q$	B–R A–Q $0.5(1-r)(1-m)(1-q)$
A–R	$0.5r$	A–Q A–R $0.5rmq$	A–R A–R $0.5rm(1-q)$	B–Q A–R $0.5r(1-m)q$	B–R A–R $0.5r(1-m)(1-q)$
B–Q	$0.5r$	A–Q B–Q $0.5rmq$	A–R B–Q $0.5rm(1-q)$	B–Q B–Q $0.5r(1-m)q$	B–R B–Q $0.5r(1-m)(1-q)$
B–R	$0.5(1-r)$	A–Q B–R $0.5(1-r)mq$	A–R B–R $0.5(1-r)m(1-q)$	B–Q B–R $0.5(1-r)(1-m)q$	B–R B–R $0.5(1-r)(1-m)(1-q)$

be the marker genotype of the offspring (indicated by the superscript o) with three different categories: $M_1^o = \mathrm{AA}$, $M_2^o = \mathrm{AB}$, and $M_3^o = \mathrm{BB}$. Given the sire is heterozygous, the conditional probabilities of the daughter's marker genotype are

$$\mathrm{P}(M_1^o | M_2^s) = 0.5m$$

$$\mathrm{P}(M_2^o | M_2^s) = 0.5$$

$$\mathrm{P}(M_3^o | M_2^s) = 0.5(1-m).$$

Considering only sires that are heterozygous at the marker locus, there are four different sire marker-QTL genotype combinations MQ_j^s

$$MQ_1^s = \begin{matrix}\mathrm{A-Q}\\ \mathrm{B-Q}\end{matrix} \qquad MQ_2^s = \begin{matrix}\mathrm{A-Q}\\ \mathrm{B-R}\end{matrix}$$

$$MQ_3^s = \begin{matrix}\mathrm{A-R}\\ \mathrm{B-Q}\end{matrix} \qquad MQ_4^s = \begin{matrix}\mathrm{A-R}\\ \mathrm{B-R}\end{matrix}.$$

Finally, let Q_i^o be the QTL genotype of the offspring with three different categories: $Q_1^o = \mathrm{QQ}$, $Q_2^o = \mathrm{QR}$, and $Q_3^o = \mathrm{RR}$. Then, the conditional probabilities for the example considered here are

$$\mathrm{P}(Q_1^o | M_1^o, MQ_2^s) = (1-r)q$$

$$\mathrm{P}(Q_2^o | M_1^o, MQ_2^s) = (1-r)(1-q) + rq$$

$$\mathrm{P}(Q_3^o | M_1^o, MQ_2^s) = r(1-q).$$

The conditional probabilities for all possible constellations are given in Table 2. For sires that are homozygous at the QTL, probabilities only depend on the population frequency of the QTL alleles. Conditional probabilities in the offspring sires that are heterozygous at the QTL depend on allele frequencies at the marker locus and the QTL and also on the recombination rate.

The expected genotypic value of an individual with QTL genotype Q_m is

$$E(y|Q_1) = \mu + a$$

$$E(y|Q_2) = \mu + d$$

$$E(y|Q_3) = \mu - a.$$

The expected phenotypic value y_{ijkl} of progeny l with marker genotype k, from sire j with marker-QTL combination MQ_i^s, is

$$E(y_{ijkl} | M_k^o, MQ_i^s) = \sum_{m=1}^{3} [\mathrm{P}(Q_m^o | M_k^o, MQ_i^s) \times E(y_{ijkl} | Q_m^o)].$$

For the example given above, this yields

$$E(y_{2j1l} | M_1^o, MQ_2^s) = \mu + ((1-r)q - r(1-q))a + ((1-r) \times (1-q) + rq)d$$

$$= \mu + (q-r)a + (2qr - r - q + 1)d.$$

Let c_{ik} be the expected deviation of a daughter with marker genotype M_k^o of a sire with marker-QTL genotype MQ_i^s from the overall mean μ, i.e.

$$c_{ik} = E(y_{ijkl} - \mu | M_k^o, MQ_i^s).$$

These conditional expectations can be derived for all possible genetic constellations and are given in Table 3.

Table 2 Conditional probabilities of offspring QTL genotypes given offspring marker genotype and sire genotype ($\mathrm{P}(Q_m^o | M_k^o, MQ_i^s)$)

Sire genotype	Offspring QTL genotypes	Probability of offspring QTL genotype when offspring marker genotype is:		
		AA	AB	BB
A–Q B–Q	QQ	q	q	q
	QR	$1-q$	$1-q$	$1-q$
	RR	0	0	0
A–Q B–R	QQ	$(1-r)q$	$((1-r)(1-m)+rm)q$	rq
	QR	$(1-r)(1-q)+rq$	$((1-r)(1-m)+rm)(1-q)$ $+(r(1-m)+(1-r)m)q$	$(1-r)q+r(1-q)$
	RR	$r(1-q)$	$(r(1-m)+(1-r)m)(1-q)$	$(1-r)(1-q)$
A–R B–Q	QQ	rq	$(r(1-m)+(1-r)m)q$	$(1-r)q$
	QR	$(1-r)q+r(1-q)$	$(r(1-m)+(1-r)m)(1-q)$ $+((1-r)(1-m)+rm)q$	$(1-r)(1-q)+rq$
	RR	$(1-r)(1-q)$	$((1-r)(1-m)+rm)(1-q)$	$r(1-q)$
A–R B–R	QQ	0	0	0
	QR	q	q	q
	RR	$1-q$	$1-q$	$1-q$

Table 3 Expected genotypic deviation from the overall mean of offspring with defined marker genotype

Sire genotype	Frequency	Offspring marker genotype		
		AA $0.5m$	AB 0.5	BB $0.5(1-m)$
A–Q B–Q	q^2	qa $+(1-q)d$	qa $+(1-q)d$	qa $+(1-q)d$
A–Q B–R	$q(1-q)$	$(q-r)a$ $+(2qr-r-q+1)d$	$(q-m-r+2rm)a$ $+(1-r-m-q+2rm+2rq+2mq-4rmq)d$	$(r+q-1)a$ $+(r+q-2rq)d$
A–R B–Q	$q(1-q)$	$(r+q-1)a$ $+(r+q-2rq)d$	$(q+m+r-2rm-1)a$ $+(r+m+q-2rm-2rq-2mq+4rmq)d$	$(q-r)a$ $+(2qr-r-q+1)d$
A–R B–R	$(1-q)^2$	$(q-1)a$ $+qd$	$(q-1)a$ $+qd$	$(q-1)a$ $+qd$

The coefficient of relationship conditional on marker information

As was already stated in the introduction, the coefficient of relationship is defined as the correlation between additive genetic values of two individuals. Therefore, the conditional coefficient of relationship in the present case is equal to the conditional halfsib correlation.

Let g_x denote the unknown QTL effect of individual x. The covariance of these QTL effects for two daughters x and y with marker genotype m_x and m_y (each being either 1, 2 or 3) of a sire that is heterozygous at the marker locus then is

$$\operatorname{cov}(g_x, g_y | M^o_{m_x}, M^o_{m_y}) = \sum_{i=1}^{4} \mathrm{P}(MQ^s_i)(c_{im_x} - \bar{c})(c_{im_y} - \bar{c})$$

$$= \left[\sum_{i=1}^{4} \mathrm{P}(MQ^s_i) c_{im_x} c_{im_y}\right] - \bar{c}^2$$

where $\bar{c}$ is the expected overall genotypic mean deviation from μ; in the biallelic case

$$\bar{c} = (2q-1)a - 2(q-q^2)d.$$

The probabilities of all possible marker-QTL combinations for sires that are heterozygous AB at the marker locus, $P(MQ^s_i)$. $i = 1, \ldots, 4$, are given as sire genotype frequencies in Table 3. If we denote the total marker information as M, the coefficient of relationship conditional on M, $R_{xy}|M$, is the conditional covariance divided by the additive genetic variance at the QTL, σ^2_q, which is in the biallelic case (Falconer 1989)

$$\sigma^2_q = 2q(1-q)[a + (1-2q)d]^2$$

so that

$$R_{xy}|M = \frac{\operatorname{cov}(g_x, g_y | M^o_{m_x}, M^o_{m_y})}{\sigma^2_q}$$

$$= \frac{\left[\sum_{i=1}^{4} \mathrm{P}(MQ^s_i) c_{im_x} c_{im_y}\right] - \bar{c}^2}{2q(1-q)[a + (1-2q)d]^2}$$

According to this formula, the conditional relationship coefficients were derived for the general case with the aid of the program *Mathematica* (Wolfram 1991) and are given in Table 4.

One remarkable result is that the conditional relationship coefficients are independent of the parameters q, a and d of the QTL, which cancel out in the division of the conditional covariance by σ^2_q. $R_{xy}|M$ is also independent of the QTL genotype of the sire, which usually is unknown. Conditional halfsib correlations are functions of r and, if at least one individual with marker genotype AB is involved, of m (see also Dekkers and Dentine 1991; Hoeschele 1993). The reason for this may be illustrated by the following example: let us assume that the population frequency m of A is low. Then, an AB daughter of an AB sire has a higher probability of having obtained the B allele from the population and the A allele from the sire than vice versa. Hence, the AB individual is probably more similar to an AA halfsib than to one with marker genotype BB because the former has also obtained the A allele from the common sire. This is shown in Figs. 1 and 2, where conditional halfsib relationships for all marker genotypes depending on the recombination rate are plotted for a low ($m = 0.1$, Fig. 1) and medium ($m = 0.4$, Fig. 2) frequency of the A allele. With low frequencies of A (Fig. 1) halfsib correlations given

Table 4 Correlation of additive QTL effects between halfsibs conditional on known marker genotype combination

Marker combination of halfsibs	Conditional halfsib correlation
AA–AA or BB–BB	$\frac{1}{2} - (r - r^2)$
AA–BB	$(r - r^2)$
AA–AB	$\frac{1}{2} - (m/2 + (1-2m)(r - r^2))$
BB–AB	$m/2 + (1-2m)(r - r^2)$
AB–AB	$\frac{1}{4} + (\frac{1}{2} - m)^2 - (1-2m)^2(r - r^2)$

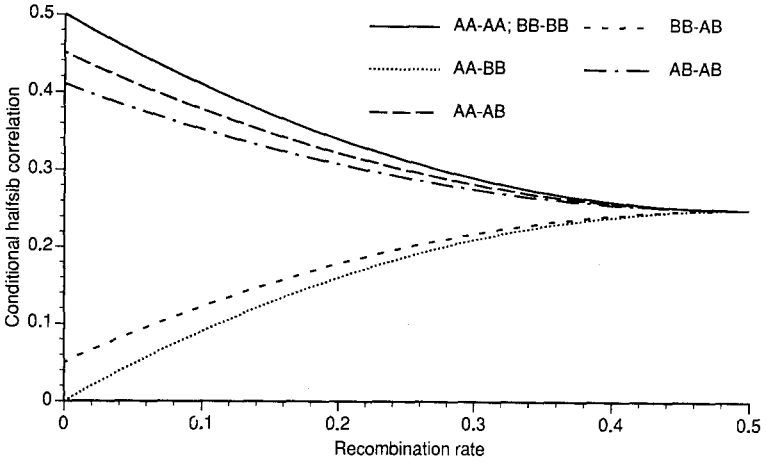


Fig. 1 Halfsib correlation conditional on marker genotype combination depending on the recombination rate r; frequency m of marker allele A is 0.1

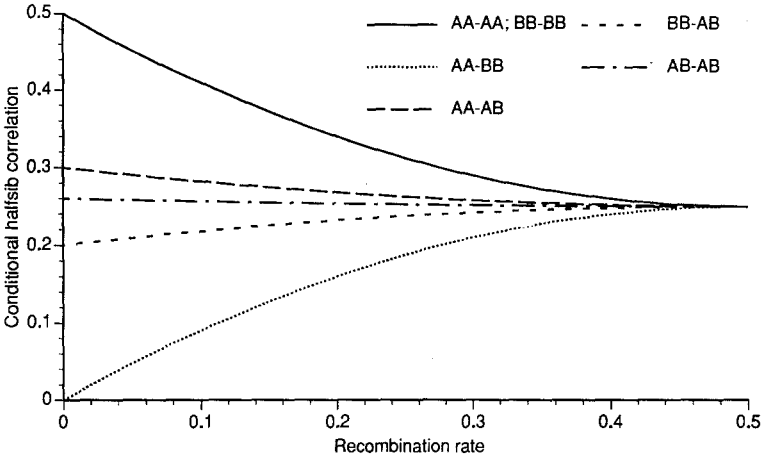


Fig. 2 Halfsib correlation conditional on marker genotype combination depending on the recombination rate r; frequency m of marker allele A is 0.4

the marker genotype combinations AA–AB and AB–AB are close to the one obtained with marker genotype combination AA–AA, while the combination BB–AB yields results similar to the combination AA–BB. With intermediate frequencies (Fig. 2), all correlations including AB animals are close to 0.25 (they are exactly 0.25 with $m = 0.5$).

In the extreme case of complete linkage ($r = 0$), the correlation of two identically homozygous halfsibs (AA–AA or BB–BB) is 0.5, while the correlation of differently homozygous halfsibs (AA–BB) is 0. This is so because in the former case the two progeny have obtained the same QTL-allele from the sire, while in the latter case they have obtained different alleles (which may be identical if the sire is homozygous at the QTL, but are not identical by descent if the sire is not inbred). That means, that identically homozygous halfsibs are as informative as fullsibs in the unconditional case. With decreasing linkage ($r \rightarrow 0.5$) all conditional relationship coefficients approach 0.25, the usual halfsib correlation, because with free recombination the marker provides no information about the QTL genotype.

Extension to more general situations

Multiple alleles at the marker locus

Offspring marker genotypes are only informative if the sire is heterozygous at the marker locus. If there are more than two marker alleles segregating in the population, a heterozygous sire still can only carry two of them. If, say, the sire has marker genotype AB but additional alleles (C,D,...) exist, three cases have to be distinguished:

- offspring with genotype AA or AC, AD... (denoted as 'A–' which means 'A and not B') definitely have obtained allele A from the sire and are to be treated like an AA offspring;
- offspring with genotype BB or BC, BD... (denoted as 'B–' which means 'B and not A') definitely have obtained allele B from the sire and are to be treated like a BB offspring;
- offspring with genotype AB remain ambiguous, i.e. it is not clear which allele stems from the sire. It is only

in this situation that the genotyping of the dam potentially adds information. If the dam is, say, AA, then it is clear that the B allele stems from the sire. If in the multiallelic case no information on the maternal marker genotype is available, the marker allele frequency m in Table 4 has to be replaced by the term $m_A/(m_A + m_B)$, where m_A and m_B are the population frequencies of the alleles A and B. If the dam is genotyped and also AB, use $m = 0.5$.

Multiple alleles at the QTL

The conditional relationship coefficient was shown to be independent of the parameters q, a and d of the QTL in the biallelic case. This property holds also for multiple alleles at the QTL, except for trivial cases where σ_q^2 becomes zero, i.e. the frequency of one allele being unity or all genotypic values being identical. This is an important result, since it means that in analyses based on the concept presented here there is no need to specify the nature of the potential QTL in terms of numbers of alleles, degree of dominance etc.

Several marker loci

The association of a QTL with one linked marker is less informative than its association with two flanking markers. In the former case a recombination between the marker and the QTL remains undetected, and by this a source of error. In the latter case, only double recombinations remain undetected. Since double recombinations are much less likely than single recombinations, flanking markers potentially are much more useful for practical applications.

If we assume one linked marker locus at either side of the QTL, only offspring groups of sires which are heterozygous at both marker loci are fully informative. If, say, a sire carries alleles A and B at the first marker locus and alleles C and D at the second one with known linkage phase A–C, two offspring with marker genotype AA–CC will have obtained the same paternal QTL allele unless a double recombination has taken place. If, however, an offspring has marker genotype AA–DD, it is certain that one recombination has taken place, but it remains unknown whether the first marker and the QTL or the QTL and the second marker are recombined.

For the flanking marker situation, conditional halfsib relationship coefficients can be derived along the exact same lines as demonstrated above for a single marker locus. Unfortunately, the derivation does not result in as simple terms as displayed in Table 4 for the single marker situation. With nine different offspring marker genotypes in the flanking marker situation, 45 different marker genotype combinations are possible, leading to 41 different conditional relationship coefficients (the reduction by 4 is due to the fact that some of the offspring combinations are equally informative). The respective conditional halfsib correlations depend on the recombination rates between the markers and the QTL and on the marker allele frequencies, but again not on parameters of the QTL. A FORTRAN subroutine to set up the conditional relationship coefficients for a given set of parameters was written and is available from the author upon request.

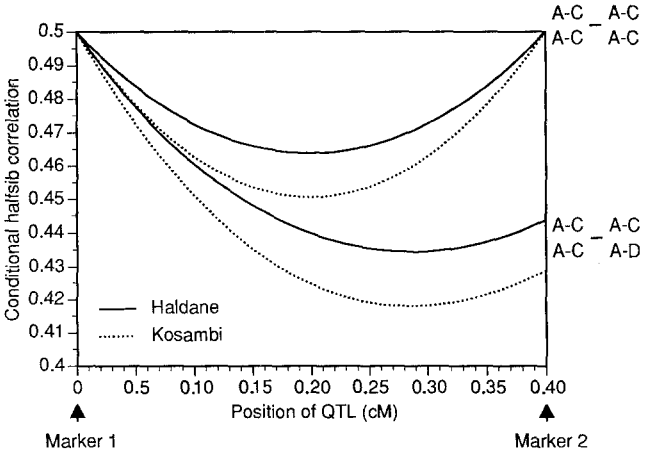
The effect of flanking markers will be demonstrated with the following example: two biallelic marker loci are assumed that are 40 centiMorgan (cM) apart. A sire is assumed to have marker genotype AB–CD with known linkage phase A–C. The position of a potential QTL is moved in steps of 1 cM from the position of marker 1 to the position of marker 2, and at each assumed QTL position the conditional halfsib correlation is calculated for

- two offspring with marker genotype AA–CC;
- two offspring with marker genotype AA–CC and AA–CD, respectively.

The results are given in Fig. 3. Since the conditional halfsib correlations are a function of recombination rates rather than genetic distances in cM, the map function has an impact on the results. In the example, the Haldane map function (Haldane 1919), assuming no interference, and the Kosambi map function (Kosambi 1944), assuming partial interference, were used. The latter leads to lower values of $R_{xy}|M$, since with the Kosambi map function the probability of recombination of loci with a given distance in cM is higher than with the Haldane map function.

For the first pair of equally homozygous offspring, the conditional halfsib correlation is 0.5 if the position of the QTL is at either of the two markers and is slightly less if the QTL is assumed to lie in the interval, the reduction accounting for the possibility of double recombination with maximum probability if the QTL lies

Fig. 3 Conditional halfsib relationship coefficients at a QTL flanked by two marker loci depending on the assumed position of the QTL, the marker genotype combination, and the map function (Haldane or Kosambi map function) used



exactly in the middle of the interval. For the second pair of offspring, the informativeness of the second marker locus depends on the population frequency of the alleles at this locus; the frequency of C was assumed to be 0.25. Although there is a certain amount of ambiguity, the conditional halfsib correlation remains consistently above 0.4 even close to the second marker locus due to the information on the first marker (the conditional halfsib correlation given only the information on the second marker would be $R_{xy}|M \leq 0.375$ with the maximum at $r = 0$).

Other types of pedigrees

The approach used in this study can be extended to other types of systematic pedigree structures, i.e. to fullsib populations or to hierarchically nested half- and fullsib structures as they arise in multiparous species like pigs, sheep, fish etc. An adaptation to experimental crosses of inbred lines seems possible but requires further research. An extension to arbitrary pedigrees (including all types of relationships and inbred individuals), however, certainly is not straightforward. Chevalet et al. (1984) have presented a general, but also very complex, approach to this problem. In order to avoid some kind of "combinatorial explosion" they suggested to restrict its use to sparse pedigrees over few generations. As an alternative, it is also possible to derive conditional relationship coefficients from the conditional gametic variance–covariance matrix suggested by Fernando and Grossman (1989).

Derivation of conditional relationship coefficients with the approach of Fernando and Grossman (1989)

Fernando and Grossman (1989) and subsequent studies (Cantet and Smith 1991; Goddard 1992; Hoeschele 1993) considered the covariance matrix G_v of gametic effects at the QTL. If N individuals are considered, this is a $2N \times 2N$ matrix containing the variances and covariances of the two gametic effects per individual. The genetic effect at a QTL is the sum of the gametic effects, and hence the variance of the genetic effect is the sum of the 2×2 submatrix on the diagonal of G_v pertaining to this individual, as was also shown by Hoeschele (1993). Similarly, the covariance of the genetic effects of two individuals is the sum of the 2×2 offdiagonal submatrix in G_v comprising the covariances of the respective gametes. Formally, this summation can be written $K'G_vK$ with $K' = I_N \otimes [11]$ where I_N is an identity matrix of dimension N and $\otimes$ symbolizes the direct of Kronecker product (Searle 1966).

G_v can be split into a dispersion matrix A_v times the additive variance of the QTL effects σ_v^2 pertaining to a single gamete. σ_q^2 as used in the present study is the variance of the sum of the two gametic effects at the QTL, hence $0.5\sigma_q^2 = \sigma_v^2$ if the base animals are not inbred. Making use of this, $K'G_vK$ can be written as

$$K'G_vK = K'A_vK\sigma_v^2 = 0.5\,K'A_vK\sigma_q^2 = Q\sigma_q^2.$$

$Q = 0.5\,K'A_vK$ is the single-locus numerator relationship matrix conditional on marker information.

For the numerical example given by Fernando and Grossman (1989), the matrix Q corresponding to the matrix G_v given in their Table II is

$$Q = \begin{bmatrix} 1.00 & 0.00 & 0.50 & 0.95 \\ & 1.00 & 0.50 & 0.05 \\ & & 1.00 & 0.75 \\ sym. & & & 1.45 \end{bmatrix}$$

In this pedigree, there is not only a conditional relationship, but also conditional inbreeding with a conditional inbreeding coefficient $F_4|M = 0.45$ for the last individual. Since the relationship coefficient is defined as a correlation, inbreeding has to be accounted for, and in the general case

$$R_{ij}|M = \frac{q_{ij}}{\sqrt{q_{ii} \times q_{jj}}} = \frac{q_{ij}}{\sqrt{(1 + F_i|M) \times (1 + F_j|M)}}.$$

This yields, e.g. for the conditional relationship of animal 1 and 4,

$$R_{14}|M = \frac{0.95}{\sqrt{1.00 \times 1.45}} = 0.789,$$

and correspondingly $R_{24}|M = 0.041$ and $R_{34}|M = 0.623$.

The approach suggested by Fernando and Grossman (1989) can hence be used to derive conditional relationship coefficients in complex pedigrees that cannot be handled with the approach suggested in the present paper. However, the basic algorithm suggested by Fernando and Grossman (1989) assumes that it is known unambiguously for each individual which marker allele was obtained from the sire and the dam. The above-mentioned case of an AB sire having an AB offspring with unknown dam cannot be handled in their approach or, at least, no clear rules are given how to do it. These practical limitations of the approach suggested by Fernando and Grossman (1989) have been eliminated by Hoeschele (1993).

Discussion and potential applications

The concept of single-locus relationship coefficients conditional on marker information can be used for the purposes of detection of QTLs and for marker-assisted selection. Consider the mixed linear model

$$y = X\beta + a + q + e$$

where y is the observation vector, $X\beta$ contains the set of fixed effects, a is the vector of random polygenic additive effects, q is a vector of random additive genetic effects of the QTL and e is the random error vector. The variance-covariance matrix of the random effects then is

$$Var\begin{bmatrix} a \\ q \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_a^2 & & \\ & Q\sigma_q^2 & \\ & & I\sigma_e^2 \end{bmatrix}$$

where A is the polygenic additive numerator relationship matrix and Q is the single-locus additive genetic numerator relationship matrix conditional on marker information. If observations are on halfsibs, simple rules can be given how to set up Q, since

$$q_{ij} = \begin{cases} 1.00 & \text{if } i = j \\ 0.00 & \text{if individuals } i \text{ and } j \text{ are not halfsibs} \\ 0.25 & \text{if individual } i \text{ and } j \text{ are halfsibs with a non-informative sire} \\ R_{ij}|M & \text{if individual } i \text{ and } j \text{ are halfsibs with an informative sire} \end{cases}$$

Based on this mixed linear model, best linear unbiased predictions of the polygenic and the single-locus breeding values can be computed from Henderson's mixed model equations (Henderson 1973).

It is also possible to estimate the variance components and to assess the association of a putative QTL with markers. The respective parameters can be estimated with a maximum likelihood or residual maximum likelihood approach (Patterson and Thompson 1971), and hypotheses (e.g. presence versus absence of a QTL in a marked chromosome segment) can be tested with the likelihood ratio test (Kendall and Stuart 1979). An application of this approach to detect potential QTLs for milk production traits in the region of the polymorphic casein genes will be reported elsewhere (Simianer 1993).

When using gametic covariance matrices, Cantet and Smith (1991) and Goddard (1992) suggested applying a reduced animal model to reduce the dimension of the mixed model equation system in the best linear unbiased prediction of random polygenic and gametic effects at one or several QTLs. Hoeschele (1993) suggested estimating gametic effects only for individuals with pertinent marker information, which yields a further reduction of the dimension of the system of equations to solve. A substantial reduction can also be achieved with the transformation of A_v to the conditional numerator relationship matrix Q and its use in a mixed linear model to predict QTL gene effects. Since only one effect per animal and QTL has to be estimated, the dimension of the system of linear equations to solve is reduced from $N_\beta + 3N_a$ to $N_\beta + 2N_a$, where N_β is the number of estimable fixed effects and N_a is the number of individuals considered. A further substantial reduction to dimension $N_\beta + N_a$ is possible if the equivalent model

$$y = X\beta + \alpha + e$$

is used, where $\alpha = a + q$ is the vector of combined polygenic and single-locus additive effects with $Var(\alpha) = A\sigma_a^2 + Q\sigma_q^2$.

Algorithms to set up Q^{-1} or $(A + cQ)^{-1}$, where c is a known constant, directly, however, are not available. In the application to halfsib pedigrees suggested in this study, this is not a problem since A and Q are blockdiagonal with block sizes equal to the number of progeny per sire. Such matrices can be numerically inverted in most cases.

The derivations presented here and in virtually all related studies are based on assumptions with respect to the genetic constellation, e.g. presence of Hardy-Weinberg and linkage phase equilibria, independence of polygenic and single-locus effects etc. These assumptions may be justified when non-selected base populations are considered. If, however, the data reflect a selection process or the base generation animals in the pedigree were selected, related to each other or inbred, these assumptions are certainly violated because, for example, selection introduces covariance between single-locus and polygenic additive effects. This latter situation is the more likely one when dealing with real life data. Hence, further research is needed to assess the impact of a violation of these assumptions or to derive approaches which apply to less restricted and more realistic situations.

Acknowledgements Two anonymous and one non-anonymous (Ina Hoeschele) reviewers are thanked for helpful comments and suggestions.

References

Cantet RJC, Smith C (1991) Reduced animal model for marker-assisted selection using best linear unbiased prediction. Genet Sel Evol 23:221–233

Chevalet C, Gillois M, Vu Tien Khang J (1984) Conditional probabilities of identity of genes at a locus linked to a marker. Genet Sel Evol 16:431–444

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

Falconer DS (1989) Introduction to quantitative genetics, 3rd edn. Longman Scientific & Technical, Longman House, Burnt Mill, Harlow, Essex, England

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

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

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

Henderson CR (1973) Sire evaluation and genetic trends. In: Anim Breed Genet Symp Dr JL Lush. ASAS and ADSA, Champaign, Ill. pp 10–41

Hoeschele I (1993) Elimination of quantitative trait loci equations in an animal model incorporating genetic marker data. J Dairy Sci 76:1693–1713

Kendall M, Stuart A (1979) The advanced theory of statistics, vol 2. Griffin, London
Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175
Malécot G (1948) Les mathématiques de l'hérédité. Masson et Cie, Paris
Patterson HD, Thompson R (1971) Recovery of interblock information when block sizes are unequal. Biometrika 58:545–554
Searle SR (1966) Matrix algebra for the biological sciences. Wiley & Sons, New York London Sydney
Simianer H (1993) Maximum likelihood estimation of the additive genetic variance at single loci using marker information. With an application to milk protein polymorphisms and milk production traits in dairy cattle. Habilitation thesis, University of Hohenheim
Wolfram S (1991) Mathematica, 2nd edn. The advanced book program. Redwood City, Calif.
Wright S (1922) Coefficients of inbreeding and relationship. Am Nat 56:330–338
Wright S (1969) Evolution and the genetics of populations, vol 2: the theory of gene frequencies. The University of Chicago Press, Chicago London