Copyright © 2002 by the Genetics Society of America

Fine Mapping of a Quantitative Trait Locus for Twinning Rate Using Combined Linkage and Linkage Disequilibrium Mapping

Theo H. E. Meuwissen,*[,1] Astrid Karlsen,[†] Sigbjørn Lien,[†] Ingrid Olsaker[‡] and Mike E. Goddard[§]

**Institute for Animal Science and Health, Lelystad, 8200 AB The Netherlands, [†]Agricultural University of Norway, Ås, 1432, Norway, [‡]Department of Morphology, Genetics and Aquatic Biology, Norwegian School of Veterinary Science, 0033 Oslo, Norway, [§]Institute of Land and Food Resources, University of Melbourne, Parkville, 3052 Australia and [§]Victorian Institute of Animal Science, Attwood, Victoria, 3049 Australia*

Manuscript received January 9, 2001
Accepted for publication February 11, 2002

ABSTRACT

A novel and robust method for the fine-scale mapping of genes affecting complex traits, which combines linkage and linkage-disequilibrium information, is proposed. Linkage information refers to recombinations within the marker-genotyped generations and linkage disequilibrium to historical recombinations before genotyping started. The identity-by-descent (IBD) probabilities at the quantitative trait locus (QTL) between first generation haplotypes were obtained from the similarity of the marker alleles surrounding the QTL, whereas IBD probabilities at the QTL between later generation haplotypes were obtained by using the markers to trace the inheritance of the QTL. The variance explained by the QTL is estimated by residual maximum likelihood using the correlation structure defined by the IBD probabilities. Unlinked background genes were accounted for by fitting a polygenic variance component. The method was used to fine map a QTL for twinning rate in cattle, previously mapped on chromosome 5 by linkage analysis. The data consisted of large half-sib families, but the method could also handle more complex pedigrees. The likelihood of the putative QTL was very small along most of the chromosome, except for a sharp likelihood peak in the ninth marker bracket, which positioned the QTL within a region <1 cM in the middle part of bovine chromosome 5. The method was expected to be robust against multiple genes affecting the trait, multiple mutations at the QTL, and relatively low marker density.

LINKAGE mapping of genes affecting complex traits is hampered by the fact that genotypes cannot be inferred with confidence from phenotype because other genes and environmental effects influence the phenotype. Linkage disequilibrium (LD) has been suggested for high-resolution mapping of these genes, but serious reservations have been expressed about the power of this method mainly with respect to the required marker density (Terwilliger and Weiss 1998; Kruglyak 1999). Cattle populations have lower effective population sizes and very large families, which produce extensive genome-wide linkage disequilibrium (Farnir *et al.* 2000) and facilitate fine mapping with relatively low density marker maps. Here we take advantage of extensive data from dairy cattle and propose a novel, robust method of combining linkage and LD mapping.

A quantitative trait locus (QTL) for twinning was detected at chromosome 5 using a genome-wide linkage analysis (Lien *et al.* 2000). The aim here was to fine map this QTL by combining the information from linkage analysis and LD, accounting for unknown background genes and the pedigree of the cattle. The method is based on estimating variance components for the variances associated to the QTL and to the background genes (see Hoeschele *et al.* 1997, for a review).

METHODS

Data: The data consisted of 6 grandsire families of Norwegian cattle, their sons, which will be referred to as sires in the following, and the daughters of the sires whose twinning rate was recorded (in total 6 bull sires and 285 sires with an average of 845 daughters per sire). All bulls were genotyped for 15 markers on chromosome 5 according to Lien *et al.* (2000), Våge *et al.* (2000), and USMARC Genome Database (2000). Both paternal and maternal alleles were scored on the markers. The CHROMPIC option of CRI-MAP 2.4 (Green *et al.* 1990) was used to identify unlikely double crossovers, and the BUILD option of the program was used to construct the linkage map presented in Figure 1. In the data set, there were no recombinations found among the marker pairs AGLA254-ETH10, CSSM22-ILSTS66, and RM154-IGF1. The distances between these markers were set at a small distance of 1 cM to allow for some recombinations between these markers in the past. Except for 3 bulls, which contributed only to the genotyping information, all bulls had daughters whose numbers of (dead or

[1]*Corresponding author:* Institute for Animal Science and Health, Box 65, Lelystad, 8200 AB The Netherlands.
E-mail: t.h.e.meuwissen@id.dlo.nl

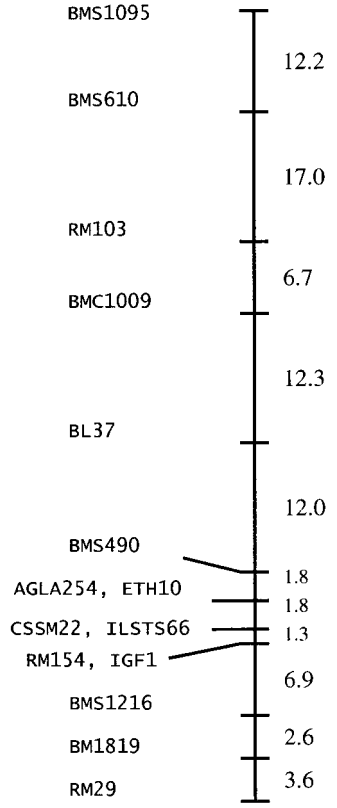

FIGURE 1.—The genetic map of the markers that were used to position the QTL.

TABLE 1

Calculation of the matrix of IBD probabilities, G_p

	Base haplotypes		
G_p =	GH	MHS	PHS
GH	[0]	[0]	[1a]
MHS	[0]	[0]	[1a]
PHS	[1a]	[1a]	[1b]

GH, grandsire haplotypes; MHS, maternal haplotypes of sires; PHS, paternal haplotypes of sires. G_p is symmetric and the diagonals = 1. The blocks denoted by [0] are calculated by the method of MEUWISSEN and GODDARD (2000). The recursive Equation 1 is used first to calculate blocks [1a] and then to calculate block [1b].

alive) offspring were recorded in the Norwegian Dairy Recording System. For a description of the data in the genotyped families, see LIEN *et al.* (2000). The pedigree of the genotyped animals was traced back for five generations (or as far as known). For a detailed description of the twinning data in the Norwegian cattle population, see KARLSEN *et al.* (2000).

Analysis: The analysis estimates variance components associated with the marked QTL, with background genes, and with a residual or error variance due to environmental effects on twinning rate. The polygenic variance is estimated using the covariance or relationship matrix among polygenic effects, which is known from pedigree. Similarly, the QTL variance is estimated using a matrix of covariances among QTL effects that is calculated from the marker haplotypes. If the marker haplotypes are similar there is an increased chance that the QTL alleles they carry are identical by descent (IBD). If the common ancestor occurs within the known pedigree, then this probability can be calculated from the markers by linkage analysis (LA). If the common ancestor is outside the known pedigree it is a source of LD. In this case the probability that the QTL alleles are IBD is calculated from the similarity between the marker haplotypes, *i.e.*, which marker alleles have both haplotypes in common, assuming LD is due to finite population size.

The first step of the analysis was to infer linkage phases of the marker alleles of the genotyped animals such that their paternally and maternally inherited marker haplotypes could be constructed. A Gibbs sampling program was written to estimate the haplotypes. The paternally and maternally inherited marker alleles were sampled simultaneously, given the genotype of the animal (if it was genotyped), the paternal and maternal marker alleles of the parents, mates, and progeny of the individual, and given the alleles of these animals and the individual at the adjacent marker loci. The marker alleles of all individuals were sampled in turn, and within an individual, all marker positions were sampled in turn. The Gibbs sampler was executed for 10,000 cycles. Early cycles, where the inheritance of the marker alleles did not yet follow Mendel's rules, were discarded from the Gibbs chain. The number of Gibbs cycles was relatively small because only linkage phases that were estimated with (almost) certainty were used in the analysis. All haplotypes of the 6 bull sires and 285 sires were used since the linkage phases of their marker alleles were estimated almost with certainty. In situations where there was uncertainty about which of the two marker alleles was linked to which of the two haplotypes, both haplotypes were considered to differ at this position from all other haplotypes (and from each other). This reduces its probability of being IBD at the QTL to other haplotypes; *i.e.*, this assumption yielded conservative predictions of IBD probabilities.

The second step is the prediction of IBD probabilities of pairs of haplotypes at putative QTL positions, which forms a matrix of IBD probabilities (Table 1). We distinguish base haplotypes, which are the haplotypes of the grandsires and maternal haplotypes of the sires, from the paternal haplotypes of the sires. The latter are (possibly recombined) copies of the grandsire's haplotypes. The IBD probabilities at the QTL of the base haplotypes given marker and pedigree information were predicted using the method of MEUWISSEN and GODDARD (2001), which accounts for the multilocus linkage disequilibria between the markers and the putative QTL position. Briefly, the IBD probability at the QTL between two

base haplotypes is based on the marker alleles that surround this locus; *i.e.*, many (non)identical marker alleles near the QTL imply a high (low) IBD probability at the QTL. The actual level of the IBD probabilities is affected by the effective population size, N_e. For the Norwegian cattle N_e was assumed to be 250 (E. Sehested, personal communication). The probability of coalescence between the current and an arbitrary base generation, $T = 100$ generations ago, is calculated given the marker alleles that both haplotypes (whose IBD probability at the QTL is calculated) have in common. Simulation studies show that the estimates of QTL position are relatively insensitive to choice of N_e and T (Meuwissen and Goddard 2000). If the two haplotypes occur in animals with a known common ancestor, then the calculation of IBD probability at the QTL is modified to account for this.

The IBD probabilities at the QTL of the base haplotypes with the paternal haplotypes of the sons, and among the paternal haplotypes, are obtained from the following equation, which states that the IBD probability, $P_{IBD}(X(p); Y)$, of the paternal QTL allele of son X, $X(p)$, with any other QTL allele, Y, equals

$$P_{IBD}(X(p); Y) = r\,P_{IBD}(S(p); Y) + (1 - r)\,P_{IBD}(S(m); Y) \quad (1)$$

(Fernando and Grossman 1989), where $S(p)$ and $S(m)$ denote the paternal and maternal alleles of the sire S, respectively, and r is the probability that the son inherited the paternal QTL allele of the sire. Hence, $X(p) = S(p)$ with probability r, and $X(p) = S(m)$, with probability $(1 - r)$. The probability r was predicted from the paternal or maternal inheritance of the nearest informative markers that flanked the putative QTL position. The above equation is used recurrently to fill in the missing IBD probabilities at the QTL of paternal haplotypes of sires using the known IBD probabilities among the base haplotypes. In conventional linkage analysis mapping by variance components (*e.g.*, Hoeschele *et al.* 1997), the base haplotypes are assumed unrelated and the IBD probabilities between the paternal haplotypes of the sires are the basis for LA mapping. Here a complete matrix of pairwise IBD probabilities between all haplotypes at the putative position of the QTL is obtained and is denoted by $\boldsymbol{G}_p$, where subscript p denotes the position at which the IBD probabilities are evaluated, *i.e.*, the position of the putative QTL (Table 1).

The third step is the calculation of the likelihood at putative QTL positions using variance component methods (Hoeschele *et al.* 1997). The average twinning frequencies of the daughters of the bulls were modeled by

$$y = \mu\mathbf{1} + \boldsymbol{Zh} + \boldsymbol{u} + \boldsymbol{e},$$

where μ is the overall mean, $\boldsymbol{h}$ is a vector of random haplotype effects, $\boldsymbol{u}$ is a vector of random polygenic effects (combined effect of background genes), $\boldsymbol{e}$ is a vector of random sampling errors, $\boldsymbol{Z}$ is an incidence matrix that indicates which haplotypes pertain to which records, and $\mathbf{1}$ is a vector of ones. The correlation matrices of $\boldsymbol{h}$ and $\boldsymbol{u}$ were $\boldsymbol{G}_p$ and $\boldsymbol{A}$, respectively, where $\boldsymbol{A}$ is the additive genetic relationship matrix based on the pedigree of the bulls (Falconer and Mackay 1996). The sampling errors, $\boldsymbol{e}$, were assumed independent with a variance proportional to $1/n_i$, where n_i is the number of daughters included in the average twinning frequency of bull i. Although the twinning frequencies follow a binomial distribution, their distribution is approximated by the Gaussian because the number of daughters per bull was large. The variances of the random effects, $\boldsymbol{h}$, $\boldsymbol{u}$, and $\boldsymbol{e}$, were estimated using the ASREML package (Gilmour *et al.* 2000), which also calculated the likelihood of the above model.

The IBD probabilities, $\boldsymbol{G}_p$, and the likelihood of the records were evaluated for a QTL at the midpoint of each marker bracket. Since there were 15 markers with 14 midpoints, likelihoods were obtained for 14 putative QTL positions.

RESULTS

The log-likelihood ratio of a model containing a QTL for twinning rate and polygenic background genes *vs.* a model containing only background genes is given in Figure 2. The log-likelihood ratio is plotted for the QTL located at the midpoint between each pair of markers. For all but one midpoint position the log-likelihood ratio of having a QTL *vs.* no QTL is close to zero, with the fifth midpoint yielding a slight negative log-likelihood ratio. The algorithm that searches maximum-likelihood parameter estimates for the QTL apparently found slightly suboptimal estimates, since a QTL variance of zero (no QTL) yielded a slightly higher likelihood at the fifth midpoint. This may be due to the QTL variance being at the boundary of the parameter space, where the maximum-likelihood search algorithm has difficulties due to the discontinuity of the likelihood function.

At the ninth marker bracket midpoint, between markers *CSSM22* and *ILSTS66*, the log-likelihood of having a QTL is increased by 3.3 log-likelihood units (yielding a nominal P value <0.01). The distinct peak in the ninth bracket and the flat likelihood in the surrounding brackets support the conclusion that there is a QTL present in this bracket and not in the surrounding brackets.

Figure 3 shows the frequency of the effects of the haplotypes when the QTL was in the ninth bracket. The distribution is clearly bimodal with a small peak on the right containing haplotypes that increased the twinning frequency. This narrow peak contained 59 haplotypes. All of these haplotypes carried the marker alleles coded 3 and 1 at the markers surrounding the ninth bracket,

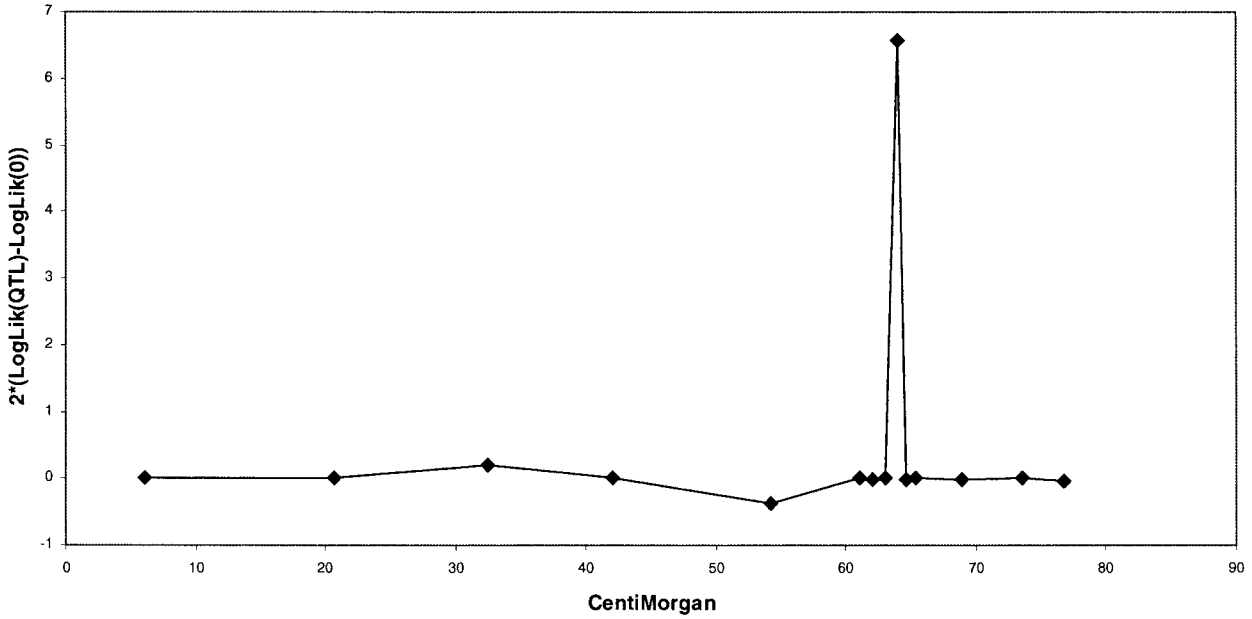

FIGURE 2.—Log-likelihood ratio of combined linkage disequilibrium and linkage analysis.

i.e., *CSSM22* and *ILSTS66*, respectively, while the alleles at other marker positions varied. Probably, the substantially larger estimates of effects for haplotypes that carried alleles [3 1] at *CSSM22* and *ILSTS66* resulted in the likelihood peak in this bracket.

Figure 4 shows the log-likelihood profiles when either the information from LD or that from linkage analysis was used alone. The likelihood profile with only LD information was obtained by calculating all IBD probabilities in $\boldsymbol{G}_p$ by the method of MEUWISSEN and GODDARD (2000); *i.e.*, known family relationships were ignored. However, this is not guaranteed to ignore all family relationships, because part of the family relationships can be recovered from the marker information (especially in the case of dense marker maps). The LD log-likelihood ratio shows again some small negative values due to slightly suboptimal maximum-likelihood parameter estimates. The likelihood profile with only linkage analysis information was obtained by setting all IBD probabilities between base haplotypes (grandsire haplotypes and maternal haplotypes of sires) to zero in $\boldsymbol{G}_p$; *i.e.*, this gives the GRIGNOLA *et al.* (1996) model for linkage analysis. The LA profile of Figure 4 gives a sharper peak than the linkage analysis of LIEN *et al.* (2000), probably due to the increased information (inclusion of extra markers; taking better account of family relationships). The LD and LA likelihood profiles of Figure 4 show more peaks than the combined LD-LA analysis (Figure 2), with a broad peak at the beginning of the chromosome and two sharp peaks at and near the ninth bracket. Because of the two peaks in the region of the ninth bracket, the chromosomal regions that contain the QTL are substantially larger when the likelihood profiles of Figure 4 are used compared to when that of Figure 2 is used. It also seems that the combined LD-LA analysis filters away spurious likelihood peaks, possibly because the data have to conform with both the LD and LA concepts at the putative QTL position. It is interesting to note that both the LD and LA profiles show a dip of 2–3 likelihood units around bracket 10, which may also have resulted in the strong drop in likelihood at bracket 10 in the combined LD/LA analysis (Figure 1).

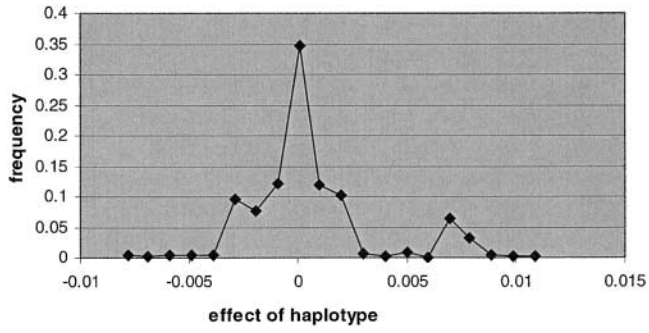

FIGURE 3.—Frequency distribution of haplotype effects at the midpoint of the ninth bracket.

Table 2 contains the estimated variances, when the QTL is at the maximum-likelihood position, due to the QTL, polygenic background genes, and environmental effects for twinning rate. The polygenic and, especially, QTL heritability are small. Also, the QTL explains only a small fraction of the total genetic variation. The average effect of the 59 haplotypes belonging to the small peak of Figure 2 was +0.74% twinning while that of the remaining 582 − 59 = 523 haplotypes was −0.03%; *i.e.*,

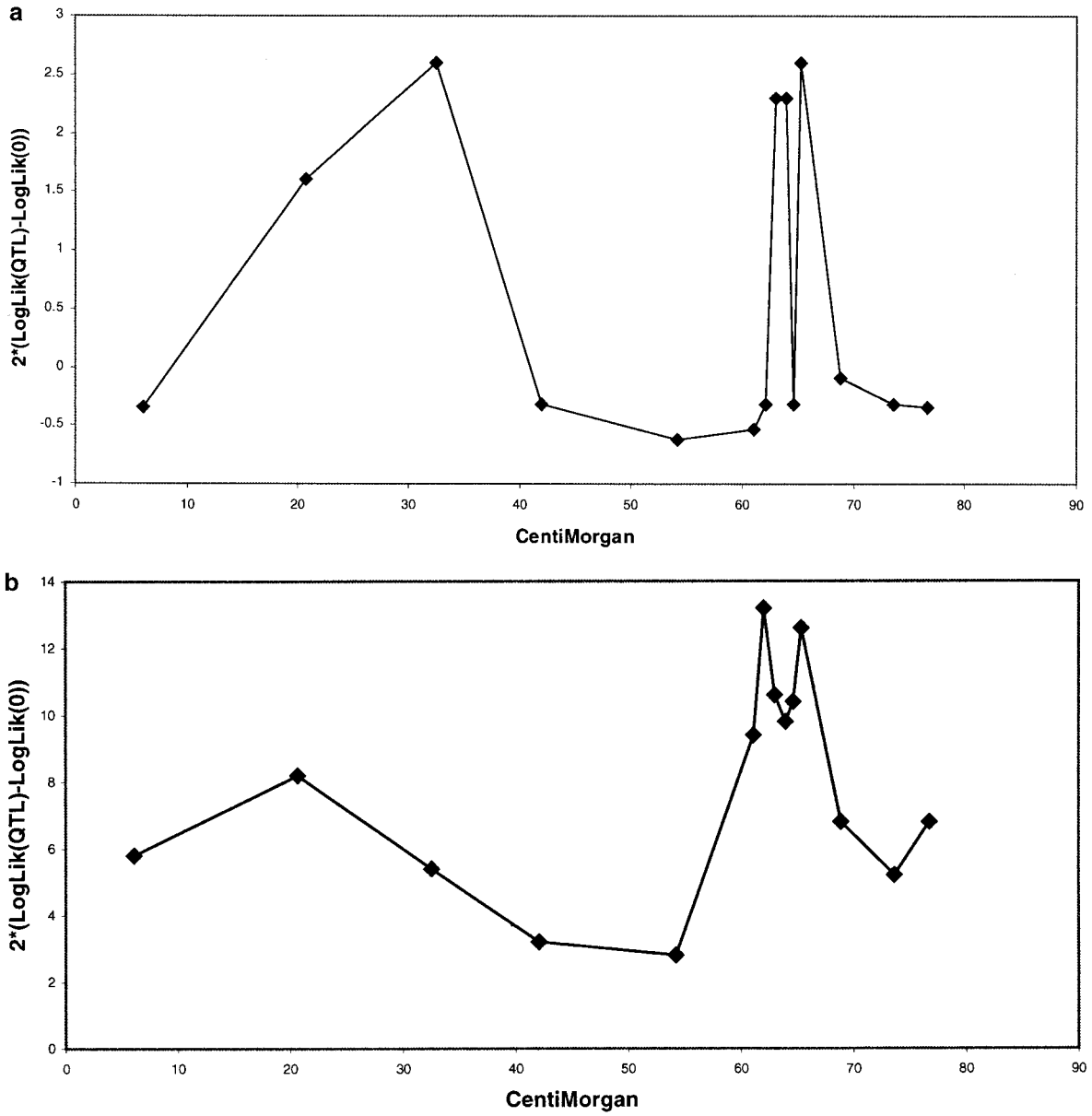

FIGURE 4.—Log-likelihood ratio of linkage disequilibrium analysis (a) or linkage analysis (b).

the difference between the two groups of haplotypes was 0.77 percentage points of twinning.

DISCUSSION

A QTL for twinning rate was mapped to marker interval *CSSM22-ILSTS66*. The close linkage of these two markers, indicated by no observed recombinants in our data (Figure 1), is supported by a 0.6-cM distance estimate in the USMARC GENOME DATABASE (2000). It seems safe to conclude that the QTL was mapped to a region <1 cM. Since the correlation between twinning rate and double ovulation rate is high (VANVLECK *et al.* 1991), the QTL probably affected dizygotic instead of monozygotic twinning. The genetic variance due to this QTL was rather small (Table 2). The difference between the group of high and low haplotype effects that resulted from the frequency distribution of Figure 3 was 0.77 percentage points of twinning, which is substantial since it equals 8% of the phenotypic standard deviation of twinning rate. About 0.1 of all haplotypes were in the group of haplotypes with high twinning frequency, which suggests that the small QTL variance was low due to a small frequency of the high twinning rate allele. The large half-sib family design seemed to have sufficient power to accurately map this rare QTL allele.

The power to map a gene explaining such a small proportion of the variance is due to the use of a granddaughter design and the large size of the families. Each son has, on average, 845 daughters so the mean twinning

TABLE 2

Variance and heritability due to the QTL and background genes in the ninth bracket

Component	Variance ($\times 10^{-4}$)	Heritability ($\times 10^{-3}$)	F_{QTL}
QTL	0.012	0.13	0.0325
Polygenic background genes	0.348	3.76	
Environmental	91.82		

F_{QTL}, fraction of the genetic variance due to the QTL.

rate of his daughters is little affected by the large error variance. Each grandsire has, on average, 47 sons so that it is possible to detect a QTL explaining only one-thirtieth of the genetic variance. The substantial difference between the groups of high and low haplotype effects of Figure 2 (*i.e.*, 0.77%) probably contributed to the power of detecting the QTL.

Some of the smallest marker brackets did not show recombination between the markers in the current data set, and their distances were assumed to be 1 cM. Although the actual distances may have been <1 cM, a distance of 1 cM already implies that recombination is a rare event, and the effect of marker distances <1 cM on IBD probabilities is small. For instance, two haplotypes that carry identical alleles for two flanking markers at a distance of 1 cM have an IBD probability of 0.62 at the midpoint of this marker bracket, while, if the distance between the markers is halved, the IBD probability is 0.66. Hence, a factor 2 difference in marker distance resulted in a difference in IBD probability of only 0.04.

The combined LA/LD mapping method that accounted for polygenic background genes is novel and resulted in a very clear signal for the position of the QTL in Figure 2. This clear signal may have resulted from (1) filtering of spurious likelihood peaks because a putative QTL is expected to give both LA and LD mapping information; (2) the correction for polygenic background genes; (3) the effective population size in the Norwegian cattle population, which results in IBD chromosome segments that are small enough for mapping at a centimorgan scale but not too small to be detected by markers at the spacing used; (4) the design and family sizes; and (5) the absence of other QTL for twinning on chromosome 5. This method of analysis can also be applied to other family structures that may contain more base haplotypes (but less information per haplotype). If the analysis indicates that there are more QTL on the chromosome, the model could be extended to include several QTL. The presented LA/LD mapping method is expected to be robust against multiple genes affecting the trait (because it accounts for background genes) and multiple mutations at the QTL (because it relies only on covariances between haplotypes that are IBD at the QTL). The method does not rely on detecting a single marker haplotype that is descended intact from the chromosome in which the most recent mutation occurred. Consequently it does not require markers that are extremely close to the QTL. The method requires only that marker haplotypes that are IBD at the QTL carry the same QTL allele. Multiple mutations or insufficiently dense markers will both result in more than one haplotype being associated with the same QTL allele but this does not destroy the mechanism on which the method is based.

On the basis of the results from this study, which maps the QTL to a relatively narrow region of <1 cM, it may be feasible to find causal gene(s) by positional cloning. An alternative, and probably more realistic, approach could be to apply the comparative positional candidate cloning strategy, which takes advantage of the recent advances in human genomics. The strategy is based on determination of evolutionary breakage points between humans and cattle, followed by extrapolation of positional information from the highly developed human map to lower density maps in cattle. However, the use of this strategy is complicated by a number of inversions and rearrangements showing up when aligning the bovine chromosome 5 with its counterparts on human chromosomes 12 and 22 (Ozawa *et al.* 2000). This is illustrated by the mapping of the *mast cell growth factor* (*MGF*) gene, which at one point was a very good positional candidate for the twinning QTL. Whereas human *MGF* is located only 1.1 cM centromeric of the *IGF1* gene on chromosome 12, the two genes are located >40 cM away on bovine chromosome 5 (Aasland *et al.* 2000). Work is currently in progress to determine at high resolution the evolutionary breakage points between species and test other comparative positional candidates from the human map in bovine.

The Research Council of Norway and GENO-Breeding and A.I. Association are acknowledged for providing financial support, The Norwegian Dairy Recording Service for providing phenotypical data, and GENO for providing relationship information for bulls.

LITERATURE CITED

Aasland, M., D. I. Våge, S. Lien and H. Klungland, 2000 Resolution of conflicting assignments for the bovine casein kinase II alpha (CSNK2A2) gene. Anim. Genet. **31:** 131–134.

Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics.* Addison-Wesley Longman, Harlow, England.

Farnir, F., W. Coppietersen, J.-J. Arranz, P. Berzi, N. Cambisano *et al.*, 2000 Extensive genome-wide linkage disequilibrium in cattle. Genome Res. **10:** 220–227.

Fernando, R. L., and M. Grossman, 1989 Marker-assisted selection using best linear unbiased prediction. Genet. Sel. Evol. **21:** 467–477.

Gilmour, A. R., B. R. Cullis, S. J. Welham and R. Thompson, 2000 *ASREML Reference Manual.* ftp.res.bbsrc.ac.uk/pub/aar.

Green, P., K. Falls and S. Crooks, 1990 *Documentation for CRI-MAP, Version 2.4.* Washington School of Medicine, St. Louis.

Grignola, F. E., I. Hoeschele and B. Tier, 1996 Mapping quantitative trait loci via residual maximum likelihood: I. Methodology. Genet. Sel. Evol. **28:** 479–490.

Hoeschele, I., P. Uimari, F. E. Grignola, Q. Zang and K. M. Gage, 1997 Advances in statistical methods to map quantitative trait loci in outbred populations. Genetics **147:** 1445–1457.

Karlsen, A., J. Ruane, G. Klemetsdal and B. Heringstad, 2000 Twinning rate in Norwegian cattle: frequency, (co)variance components, and genetic trends. J. Anim. Sci. **78:** 15–20.

Kruglyak, L., 1999 Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat. Genet. **22:** 139–144.

Lien, S., A. Karlsen, G. Klemetsdal, D. I. Våge, I. Olsaker *et al.*, 2000 A primary screen of the bovine genome for quantitative trait loci affecting twinning rate. Mamm. Genome **11:** 877–882.

Meuwissen, T. H. E., and M. E. Goddard, 2000 Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. Genetics **155:** 421–430.

Meuwissen, T. H. E., and M. E. Goddard, 2001 Prediction of identity by descent probabilities from marker-haplotypes. Genet. Sel. Evol. **33:** 605–634.

Ozawa, A., M. R. Band, J. H. Larson, J. Donovan, C. A. Green *et al.*, 2000 Comparative organization of cattle chromosome 5 revealed by comparative mapping by annotation and sequence similarity and radiation hybrid mapping. Proc. Natl. Acad. Sci. USA **97:** 4150–4155.

Terwilliger, J. D., and K. M. Weiss, 1998 Linkage disequilibrium mapping of complex disease: fantasy or reality? Curr. Opin. Biotechnol. **9:** 578–594.

USMARC Genome Database, 2000 U.S. Meat Animal Research Center, Clay Center, Nebraska. http://sol.marc.usda.gov/genome/cattle/cattle.html.

VanVleck, L. D., K. E. Gregory and S. E. Echternkamp, 1991 Ovulation rate and twinning rate in cattle: heritabilities and genetic correlation. J. Anim. Sci. **69:** 3213–3219.

Våge, D., I., Olsaker, H. Klungland, L. Gomez-Raya and S. Lien, 2000 A male genetic linkage map for QTL-mapping in Norwegian cattle. Acta Agric. Scand. **50:** 56–63.

Communicating editor: C. Haley