Simultaneous Mining of Linkage and Linkage Disequilibrium to Fine Map Quantitative Trait Loci in Outbred Half-Sib Pedigrees: Revisiting the Location of a Quantitative Trait Locus With Major Effect on Milk Production on Bovine Chromosome 14

Frédéric Farnir,*1 Bernard Grisart,*1 Wouter Coppiepers,* Juliette Riquet,* Paulette Berzi,* Nadine Cambisano,* Latifa Karim,* Myriam Mni,* Sirja Moisio,† Patricia Simon,* Danny Wagenaar,* Johanna Vilkki† and Michel Georges*2

*Department of Genetics, Faculty of Veterinary Medicine, University of Liége (B43), 4000-Liége, Belgium and
†Animal Production Research, Agricultural Research Centre MTT, 31600 Jokioinen, Finland

Manuscript received September 10, 2001
Accepted for publication February 22, 2002

ABSTRACT

A maximum-likelihood QTL mapping method that simultaneously exploits linkage and linkage disequilibrium and that is applicable in outbred half-sib pedigrees is described. The method is applied to fine map a QTL with major effect on milk fat content in a 3-cM marker interval on proximal BTA14. This proximal location is confirmed by applying a haplotype-based association method referred to as recombinant ancestral haplotype analysis. The origin of the discrepancy between the QTL position derived in this work and that of a previous analysis is examined and shown to be due to the existence of distinct marker haplotypes associated with QTL alleles having large substitution effects.

With the development of comprehensive genetic marker maps in many species, it has become possible to map quantitative trait loci (QTL) underlying the genetic variance of medically and economically important complex phenotypes (e.g., lander and schork 1994; Paterson 1995; Andersson 2001; Flint and Mott 2001; Mackay 2001; Mauricio 2001). Generally speaking, however, the resolution that is achieved when mapping QTL (typically of the order of 20–30 cM) is insufficient if one intends to positionally clone the corresponding genes. Also, this level of mapping accuracy limits the benefits of QTL mapping information for marker-assisted selection (e.g., Spelman and van Arenndonk 1997). There is therefore a need to develop methods that improve the resolution of QTL mapping.

Most strategies proposed to achieve this goal imply breeding of large numbers of additional offspring to increase the density of recombinants in the chromosome regions of interest (e.g., Darvasi 1998). In man as well as most livestock species this proposition is not practical. An alternative to the generation of recombinants de novo is to exploit historical recombinants, i.e., use linkage disequilibrium (LD; e.g., Terwilliger and Weiss 1998). As in most human populations useful levels of LD might not extend for >10–100 kb (Kruglyak 1999; Reich et al. 2001); this approach implies very dense maps counting maybe as many as 500,000 polymorphic markers.

We recently made the unexpected observation that, contrary to the situation in human, LD extends over several tens of centimorgans in cattle (Dunner et al. 1997; Farnir et al. 2000). This suggests that in cattle, and maybe in other livestock species, LD mapping might be effective using the available medium density microsatellite maps (e.g., Kappes et al. 1998). This prediction has been supported by the mapping of the bovine syndactyly gene (Charlier et al. 1996) as well as the fine mapping of a QTL influencing milk composition (Riquet et al. 1999) using strategies based on identity-by-descent.

In the Riquet et al. (1999) study, we applied a twotiered approach for the fine mapping of QTL in outbred populations. In the first stage, heterozygous “Qq” individuals were identified by performing marker-assisted segregation analysis in their offspring. In the second stage, shared haplotypes flanking the “Q” allele predicted to be identical-by-descent among Qq individuals were identified, thereby defining the QTL position.

While applied successfully to a QTL previously mapped to bovine chromosome 14 (Coppiepers et al. 1998a; Heyen et al. 1999; Riquet et al. 1999; Looft et al. 2001), extending this approach to other QTL has been hampered mainly because of the difficulty in identifying sufficient numbers of individuals with an unambiguous Qq genotype. In this work, we describe the development of a mapping method that simultaneously exploits linkage and linkage disequilibrium and that allows inclusion of individuals with an ambiguous QTL genotype.
The method has been applied to the QTL influencing milk yield and composition located at the centromeric end of chromosome 14. The results suggest a more proximal map position than the one that was previously identified (Riquet et al. 1999). A preliminary analysis of recombinant ancestral haplotypes (RAH) supports this proximal location of the QTL. We provide evidence that our initial erroneous localization is due to the heterogeneity of haplotypes associated with large QTL substitution effects.

**MATERIALS AND METHODS**

**Pedigree material:** The pedigree material used for QTL mapping comprised the following:

- **Data set I:** A Black-and-White Holstein-Friesian granddaughter design sampled in the Netherlands and composed of 22 paternal half-sib families for a total of 987 bulls.
- **Data set II:** A Black-and-White Holstein-Friesian granddaughter design sampled in New Zealand and composed of seven paternal half-sib families for a total of 227 bulls. Data sets I and II have been described previously (Spelman et al. 1996; Coppieters et al. 1998a).
- **Data set III:** A Red-and-White Holstein-Friesian granddaughter design sampled in the Netherlands and Belgium comprising 23 paternal half-sib families for a total of 401 bulls (B. Grisart, unpublished data).
- **Data set IV:** A novel granddaughter design composed of 39 Dutch Black-and-White Holstein-Friesian paternal half-sib families for a total of 430 bulls.
- **Data set V:** A daughter design composed of 51 New Zealand Holstein-Friesian paternal half-sib families for a total of 529 cows.

**Phenotypes:** Phenotypes were respectively daughter yield deviations (DYD) for bulls, lactation values (unregressed first lactation) for cows, and predicted transmitting abilities (PTA) for bulls and cows for lactation yield deviations) for cows, as well as average parental predicted transmitting ability (PTA) as the parental average. This likelihood is a function of two unknown parameters: the allele substitution effect, \( \alpha \), and the residual variation, \( \sigma^2 \). The values of \( \alpha \) and \( \sigma^2 \) maximizing \( L_{\text{ped}} \) can be found by means of iterative procedures such as the “quasi-Newton” methods (e.g., GEMINI; Lalouel 1983) or expectation-maximization (EM) methods.

Note that in this study, we have used the average parental predicted transmitting ability (see above) as the parental average (PA). An estimate of the dam’s breeding value that would not depend on her son’s record would actually be preferred.

**Likelihood of multiple half-sib pedigrees assuming linkage between marker map and QTL:** If \( n \) half-sib pedigrees are sampled in a given outbred population, one cannot assume that all founder sires will be segregating for the QTL at map position (\( p \)). Assuming a biallelic QTL with alleles \( Q \) and \( q \) in Hardy-Weinberg equilibrium and having respective frequencies \( f_Q \) and \( f_q \) (allele \( Q \) and \( q \)), a proportion \( 2 f_Q f_q \) of sires will be heterozygous \( Qq \), the others being either \( QQ \) or \( qQ \). Because one does not know a priori which founder sires are heterozygous, the overall likelihood of the data can be computed as

\[
L_{\text{ped}} = \prod_{i=1}^{n} \left[ P(\text{Son}_i, A(p)|M) \phi(\text{Ph}_i, PA_i + \frac{\alpha}{2}, \sigma^2) + P(\text{Son}_i, B(p)|M) \phi(\text{Ph}_i, PA_i - \frac{\alpha}{2}, \sigma^2) \right]
\]

\[\text{GEORGES et al. 1995}, \text{where } \prod_{i=1}^{n} \text{ is the product over all sons, and } P(\text{Son}_i, A(p)|M) \text{ and } P(\text{Son}_i, B(p)|M) \text{ are the probabilities that son } j \text{ has inherited homologs } A \text{ and } B \text{ from its sire at map position } (p) \text{ given its composite genotype } M_j \text{ for all markers in the map. These probabilities can be computed as previously described (Coppieters et al. 1998b). } \phi(\text{Ph}_i, PA_i \pm \frac{\alpha}{2}, \sigma^2) \text{ is the probability density function for a normally distributed random variable with mean } PA_i \pm \frac{\alpha}{2} \text{ and variance } \sigma^2, \text{ computed as }\]

\[
\phi(\text{Ph}_i, PA_i \pm \frac{\alpha}{2}, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{(\text{Ph}_i - PA_i \pm \alpha/2)^2}{2\sigma^2} \right].
\]

This likelihood is a function of two unknown parameters: the allele substitution effect, \( \alpha \), and the residual variation, \( \sigma^2 \). The values of \( \alpha \) and \( \sigma^2 \) maximizing \( L_{\text{ped}} \) can be found by means of iterative procedures such as the “quasi-Newton” methods (e.g., GEMINI; Lalouel 1983) or expectation-maximization (EM) methods.

Note that in this study, we have used the average parental predicted transmitting ability (see above) as the parental average (PA). An estimate of the dam’s breeding value that would not depend on her son’s record would actually be preferred.

**Likelihood of multiple half-sib pedigrees assuming linkage between marker map and QTL:** If \( n \) half-sib pedigrees are sampled in a given outbred population, one cannot assume that all founder sires will be segregating for the QTL at map position (\( p \)). Assuming a biallelic QTL with alleles \( Q \) and \( q \) in Hardy-Weinberg equilibrium and having respective frequencies \( f_Q \) and \( f_q \) (allele \( Q \) and \( q \)), a proportion \( 2 f_Q f_q \) of sires will be heterozygous \( Qq \), the others being either \( QQ \) or \( qQ \). Because one does not know a priori which founder sires are heterozygous, the overall likelihood of the data can be computed as

\[
L_{\text{ped}} = \prod_{i=1}^{n} \left[ f_Q f_i P(\text{Ped}|\text{Sire} = Q_i Q_o) + f_q f_i P(\text{Ped}|\text{Sire} = q_i q_o) + f_Q f_q P(\text{Ped}|\text{Sire} = q_i Q_o) \right].
\]

The \( Q_i Q_o \) vs. \( q_i q_o \) notation refer to QTL genotypes in which the \( Q \) allele is located on the \( A \) vs. \( B \) homolog of the sire, respectively. The probabilities of the \( i \)th pedigree conditional on the QTL genotype of the \( i \)th sire can be computed as

\[
P(\text{Ped}|\text{Sire} = Q_i Q_o) = P(\text{Ped}|\text{Sire} = q_i q_o) = \prod_{j=1}^{o} \phi(\text{Ph}_j, PA_j, \sigma^2)
\]

\[
P(\text{Ped}|\text{Sire} = Q_i q_o) = \prod_{j=1}^{o} \left[ P(\text{Son}_j, A(p)|M) \phi(\text{Ph}_j, PA_j + \frac{\alpha}{2}, \sigma^2) + P(\text{Son}_j, B(p)|M) \phi(\text{Ph}_j, PA_j - \frac{\alpha}{2}, \sigma^2) \right].
\]

\[
P(\text{Ped}|\text{Sire} = q_i Q_o) = \prod_{j=1}^{o} \left[ P(\text{Son}_j, A(p)|M) \phi(\text{Ph}_j, PA_j - \frac{\alpha}{2}, \sigma^2) + P(\text{Son}_j, B(p)|M) \phi(\text{Ph}_j, PA_j + \frac{\alpha}{2}, \sigma^2) \right].
\]

where \( \prod_{j=1}^{o} \) is the product over all \( O \) offspring of sire \( i \), \( \phi(\text{Ph}_j, PA_j, \sigma^2) \) is the probability density function for a normally
distributed random variable with mean $\mu$ and variance $\sigma^2$, and $\mathcal{M}_i$ denotes known marker information on son $i$.

The likelihood expression in Equation 2 is a function of $\alpha$, $\sigma^2$, and an additional unknown parameter, $f_i$. As before, the values of $\alpha$, $\sigma^2$, and $f_i$ jointly maximizing $L_{\text{Ped}}$ can be found using optimization routines such as GEMINI (Lalouel 1983).

Likelihood of multiple half-sib pedigrees assuming linkage disequilibrium between marker map and QTL—extracting information from the sires only: If we single out one of the microsatellite markers in the linkage group ($m$), we can assume that QTL allele $Q$ appeared in the population of interest—by mutation or migration—on a chromosome carrying allele $1$ at marker locus $m$, and that this occurred $g$ generations ago. Pooling all other alleles at marker locus $m$ in a single class, 0, the haplotype frequencies at the present generations have expected values

\[
\begin{align*}
  f(Q - 1) &= \gamma_1 \\
  f(Q - 0) &= \gamma_2 = f_0 - \gamma_1 \\
  f(q - 1) &= \gamma_3 = f_1 - \gamma_2 \\
  f(q - 0) &= \gamma_4 = f_2 - \gamma_3,
\end{align*}
\]

where

\[
\gamma_1 = f_0 \left[ f_0 + \rho (1 - \theta) f_0 \right].
\]

$\rho$ is a heterogeneity parameter corresponding to the proportion of Q-bearing chromosomes that are identical-by-descent with the founder chromosome bearing marker allele $1$ (Terwilliger 1995), $\theta$ is the distance in recombination units between marker locus $m$ and map position $p$, and $f_0$ and $f_2$ ($=1 - f_0$) are the respective frequencies of marker alleles $0$ and $2$. The latter can be estimated directly from the marker genotype data.

These haplotype frequencies can be used to calculate QTL genotype probabilities conditional on marker genotype $m$, as shown in Table 1.

From these, the likelihood of the data can now be computed as

\[
L_{\text{Ped}} = \prod_{i=1}^{N} \left[ P(\text{Sire} = Q, Q_1, m) P(\text{Ped} | \text{Sire} = Q, Q_0) + P(\text{Sire} = Q, Q_0, m) P(\text{Ped} | \text{Sire} = Q, Q_0) + P(\text{Sire} = q, q_1, m) P(\text{Ped} | \text{Sire} = q, q_0) + P(\text{Sire} = q, q_0, m) P(\text{Ped} | \text{Sire} = q, q_0) \right].
\]

In the absence of linkage disequilibrium, i.e., if $(1 - \theta)\rho = 0$, $P(\text{Sire} = Q, Q_1, m)$ reduces to $f_0$, $P(\text{Sire} = q, q_0, m)$ to $f_0$, and $P(\text{Sire} = Q, Q_0, m)$ and $P(\text{Sire} = q, q_1, m)$ to $f_0$, thereby reducing (7) to (2).

In reality, we do not know which marker allele was initially associated with the QTL allele. We therefore have to compute (7) for all $A$ alleles of marker $m$ and sum over all likelihoods,

\[
L_{\text{Ped}} = \sum_{k=1}^{A} \prod_{i=1}^{N} \left[ f_k \phi(P_h, PA, \sigma^2) + (1 - f_k) \phi(P_h, PA + \alpha, \sigma^2) \right] P(\text{Son}_{ij}, M_j, M_i)
\]

where $f_k$ is the population frequency of allele $k$ and $P(\text{Sire} = Q, Q_1, m)$ corresponds to the probability that sire $\alpha$ has QTL genotype $Q, Q_1, Q_0$ given its genotype for marker $m$ and the assumption that marker allele $k$ was initially associated with QTL allele $Q$.

So far, we have considered only the association between the QTL and one of the markers in the linkage group: $m$. Following Terwilliger (1995), we multiplied the likelihoods computed separately for the different markers in the map according to (8) over all $M$ markers to yield the following overall likelihood expression:

\[
L_{\text{Ped}} = \prod_{m=1}^{M} \prod_{k=1}^{A} f_k \phi(P_h, PA, \sigma^2) + (1 - f_k) \phi(P_h, PA + \alpha, \sigma^2) \]

The likelihood expression in Equation 9 involves two new unknown parameters, $g$ and $\rho$, which can be estimated jointly with $\alpha$, $\sigma^2$, and $f_i$ to maximize $L_{\text{Ped}}$. Note that we follow Terwilliger (1995) in estimating only a single value for $\rho$. Strictly speaking, however, different marker alleles may be associated with distinct subsets of $Q$ alleles corresponding to distinct proportions of $Q$-bearing chromosomes.

Likelihood of multiple half-sib pedigrees assuming linkage disequilibrium between marker map and QTL—extracting information from sire and ungenotyped dams: In half-sib designs, the majority of dams will have only one offspring. The most commonly applied QTL mapping methods are therefore not able to extract information from the dam’s meioses. This would require inclusion of additional pedigree relationships linking the dams to other pedigree members. This is, for instance, done in the great-granddaughter design (Coppieters et al. 1999) or methods that target full-pedigree analysis (e.g., Bink and van Arendonk 1999). When exploiting LD, however, even meioses from parents having a single offspring are potentially informative. Even if the dams have not been genotyped, as would often be the case in, for instance, the granddaughter design, the marker genotypes of the maternal gametes transmitted to the offspring can easily be inferred. This information can be utilized in Equation 9 by modifying Equations 3–5 as follows:

\[
P(\text{Ped} | \text{Sire} = q, q_1, m) = \prod_{j=1}^{\theta} \left[ f_0 \phi(P_h, PA, \sigma^2) + (1 - f_0) \phi(P_h, PA + \alpha, \sigma^2) \right] P(\text{Son}_{ij}, M_j, M_i)
\]
given its marker genotype $m_{278}$. F. Farnir

if $Q$ is the product over all

This can be accomplished using a conventional interval mapping approach, i.e., by sliding the position of the hypothetical QTL through the marker map and computing the likelihood of the data at each position according to Equation 9. The position on the map associated with the highest likelihood can then be considered as the most likely position of the QTL.

Table 2

<table>
<thead>
<tr>
<th>$P(\text{Son}, Q_k)$</th>
<th>$P(\text{Son}, -Q_k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{ij}/\phi_j$</td>
<td>$\gamma_{ij}/\phi_j$</td>
</tr>
</tbody>
</table>

In addition, one can compute the likelihood of the data under the null hypothesis of no QTL at the corresponding map position. We distinguish two distinct null hypotheses: $H_0$, in which $\alpha$ is set to zero, and $H_N$ in which the QTL is positioned at 50% recombination rate from the examined location, $p$.

The significance of these alternative hypotheses can be evaluated by generating different likelihood ratio statistics: $H_0/H_1$ tests the combined linkage + linkage disequilibrium signal, $H_1/H_2$ tests the linkage signal, and $H_2/H_1$ tests the linkage disequilibrium signal.

The approach chosen to combine the LD information from the different markers into a quasi-multipoint expression (Terwilliger 1995) is an approximation. As a consequence, the theoretical distributions of these likelihood-ratio statistics are not precisely known. One can demonstrate that in the absence of LD, $\log_{10}(L_{N-Po}(H_1)/L_{N-Po}(H_0))$, computed according to Equation 9, corresponds to $M \times$ lod score. We therefore report the corresponding likelihood ratios as $L(1/M)\log_{10}(L_{N-Po}(H_1)/L_{N-Po}(H_0))$ in this work.

The LINKLD programs: The described linkage and linkage disequilibrium analyses are implemented in a computer program (LINKLD), which is available from the authors upon request.

Recombinant ancestral haplotype analysis: Using the maximum-likelihood approach outlined above, one can estimate the most likely genotype of the ancestral marker haplotype (AH) on which the mutation generating the $Q$ allele occurred. Indeed, using Equation 7, one can compute the likelihood of the pedigree data assuming that a given allele ($k$) at the considered marker ($m$) was initially associated with the $Q$ QTL allele. The allele yielding the highest likelihood can then be
considered to be the most likely ancestral allele for that marker. The combination of the most likely ancestral alleles over all markers then yields the most likely AH. All other haplotypes are referred to as OH. Alternatively, the AH can be identified on the basis of the haplotype sharing among putative QTL sires as outlined in Riquet et al. (1999).

On the basis of the knowledge of this AH, we then define 2(n−1) classes of RAH, including (n−1) telomeric (T)RAH haplotypes that are identical-by-state with the AH from the telomeric end up to one of the n marker positions and (n−1) centromeric (C)RAH haplotypes that are identical-by-state with the AH from the centromeric end up to one of the n marker positions. Figure 3 illustrates this concept. The genotype of individuals with phase-known marker genotype can then be expressed as a function of their haplotype class. Individuals could, for instance, be of OH/TRAH(n−1) genotype, of AH/RAH(3) genotype, etc. The effect on phenotype of each haplotype class can then be estimated using the linear model

\[ \text{Ph}_i = H_i^p + 0.5(H_i^m - 0.5(SH_i^p + SH_i^m)) + e, \]

where \( \text{Ph}_i \) is the phenotypic value of individual \( i \), \( PA_i \) is the average phenotypic value of the sire and dam of individual \( i \), \( H_i^p \) and \( H_i^m \) are, respectively, the effect of paternal and maternal marker haplotypes of individual \( i \) as defined above, \( SH_i^p \) and \( SH_i^m \) are, respectively, the paternal and maternal marker haplotypes of the sire of individual \( i \), and \( e \) is a residual error term.

Indeed, the phenotype of each individual can be partitioned into the contribution of the QTL of interest, a residual "polygenic" (PG) component, and an error term as

\[ \text{Ph}_i = H_i^p + H_i^m + \text{PG}_i + e. \]

The expectation of the polygenic component of individual \( i \) is estimated as the average of the polygenic components of sire (\( \text{SPh}_i \)) and dam (\( \text{DPG}_i \)),

\[ \text{Ph}_i = H_i^p + H_i^m + 0.5(\text{SPh}_i + \text{DPG}_i) + e. \]

The polygenic components of sire and dam can be expressed as functions of their respective phenotypic values (\( \text{SPh}_i \) and \( \text{DPG}_i \)) and marker haplotypes (\( SH_i \) and \( DH_i \)):

\[ \text{Ph}_i = H_i^p + H_i^m \]
\[ + 0.5(\text{SPh}_i - SH_i + \text{DPG}_i - DH_i) + e, \]
\[ \text{Ph}_i = H_i^p + H_i^m + 0.5(\text{SPh}_i + \text{DPG}_i) - 0.5(SH_i + DH_i) \]
\[ - 0.5(DH_i - \text{DPG}_i) + e. \]

In the case of the granddaughter design the dams are not genotyped and the only known marker haplotype of the dam is \( H_i^m \). Therefore,

\[ \text{Ph}_i = H_i^p + H_i^m + PA_i - 0.5(SH_i + \text{SH}_i) - 0.5H_i^m + e, \]

which is identical to Equation 15. Note that the error terms, \( e \), in Equations 15–19 are not equivalent.

Following Thaller and Hoeschele (2000), but using historical rather than current recombinations, we then combine the different haplotype classes such as to derive specific contrasts for each marker interval. The expectation is that the interval-specific contrast will be maximal in the true interval.

Corresponding ANOVA analyses were performed with the GLM procedure of the SAS package.

**RESULTS**

**Joint linkage and linkage disequilibrium mapping point toward proximal location of the chromosome 14 QTL:** We initially applied the LINKLD programs to data sets I and III jointly. Although data sets I and III correspond to Black-and-White vs. Red-and-White animals, respectively, the corresponding phenotypes (DYD) are obtained from a joint breeding value estimation (NRS, The Netherlands), allowing us to treat data sets I and III as a single, homogeneous population. Figure 1 summarizes the results that were obtained with LINKLD when analyzing the effect of chromosome BTA14 on fat percentage. The choice of fat percentage is justified by the fact that this trait was shown previously to be the most profoundly affected by this QTL. (Coppieters et al. 1998a). Figure 1, A and B, illustrates the results obtained when considering only the information from the sires. It can be seen that the results are quite similar whether considering no QTL (\( \alpha = 0 \)) or an unlinked QTL (\( \theta = 0.5 \)) as null hypothesis. When considering linkage information only, very significant maximum lod scores of, respectively, 30 (\( H_{aq} \)) and 28 (\( H_{aq} : \theta = 0.5 \)) are obtained at map position 5 (i.e., in the second marker bracket). The additional information obtained from including linkage disequilibrium is also shown in Figure 1, A and B, yielding under both null hypotheses a maximum of 1.9 additional lod score at map position 6, i.e., the third marker. The corresponding maximum-likelihood estimators of \( \alpha, \rho, \sigma, g \) and \( f_0 \) are reported in Table 3.

It can be seen that the \( Q \) to \( q \) allele substitution effect is estimated at 1.23 residual standard deviations, which clearly confirms the major effect of the corresponding QTL. The population frequency of the \( Q \) alleles is estimated at 0.5 when ignoring LD and 0.39 when considering LD, therefore implying that \( \sim 50\% \) of the studied sire families might be segregating for this QTL. This high percentage is not really unexpected given the fact that at least 11 out of the 45 paternal half-sib families comprising data sets I and III were shown clearly to segregate on the basis of marker-assisted segregation analysis performed within families as described (Riquet et al. 1999; B. Grisart, unpublished observations) and despite the intrinsically low power to detect such segregating families. The average number of generations to coalescence was estimated at 3.8. Note that we have previously estimated that \( bi \) individuals sharing an identical-by-descent mutation inherited independently from a common founder \( g \) generations back will share an identical-by-descent flanking segment of size \( 2/gb \) Morgans on average (Dunne et al. 1997). The fat-increasing chromosome of the 11 Dutch segregating sires that were previously mentioned shared a common haplotype over a chromosome length of 13 cM (see Figure 2). This length has to be compared with the expected 4.7 cM. This discrepancy could be due to the close relatedness
Figure 1.—Lod scores obtained when analyzing data sets I and III jointly. Only sire information is considered in A and B, while sire and dam information is considered in C and D. The null hypothesis is that of no QTL \( H_0: \alpha = 0 \) for A and C and that of an unlinked QTL \( H_2: \theta = 0.5 \) in B and D. The black curves (solid diamond) correspond to \( (1/M) \times \log(H_2/H_0) \), therefore extracting information from both linkage and LD, the light gray curves (shaded square) to \( (1/M) \times \log(H_1/H_0) \), therefore extracting information from linkage only, and the dark gray curves (shaded triangle; scaled to 1/10) to \( \log(H_2/H_1) \) correspond to LD only. The cross-hatched area corresponds to the interval selected by Riquet et al. (1999) and subsequently refined by B. Grisart (unpublished data).
between some of the sires. The value of 92% for the heterogeneity parameter indicates the relative allelic homogeneity of the hypothetical Q population in the sires.

Figure 1, C and D, summarizes the results obtained from the same data sets when exploiting the information from the maternal chromosomes. When assuming no QTL (\(\alpha = 0\)) as a null hypothesis (Figure 1C) and when considering only linkage information (i.e., ignoring linkage disequilibrium), lod scores reach a maximum value of 79 at the first marker locus. This is a surprising result as this type of analysis exploits essentially the same mapping information as in the analysis in Figure 1A. It is notable that this lod score inflation is not limited to the proximal end of chromosome 14 but applies to its entire length. We attribute this to the fact that this analysis attempts to estimate the QTL genotype of the maternal gamete (see Equations 10–12 in MATERIALS AND METHODS). If, after correction for the inheritance of the paternal homologues, the sons’ phenotypic distribution clearly departs from normality, the null hypothesis of no QTL (\(\alpha = 0\)) becomes very unlikely, thereby increasing the lod score values across the entire chromosome. This explanation is supported by the examination of the lod score curves obtained when using the alternate null hypothesis of an unlinked QTL (\(\theta = 0.5\); Figure 1D). Although the lod score profile is parallel to the previous ones across the entire chromosome, the actual values have dropped by >40 lod score units when ignoring LD. Lod scores even become negative in the middle of the chromosome, indicating the higher likelihood of an unlinked QTL for that position. A maximal lod score of 36 is obtained at the position of the first marker under this scenario. This value is still higher than the one obtained when ignoring dam information (\(z = 30\); Figure 1, A and B). Assuming comparable distributions of lod score values under \(H_0\), this might indicate that even in the absence of LD, estimating the QTL genotype of the maternal gamete might increase the QTL mapping power.

The additional signal obtained from the consideration of LD also has a very similar profile across the chromosome under both null hypothesis options, maximizing in both cases just proximal from the first marker locus. At this position 7.4 additional lod score units are gained under the null hypothesis of no QTL vs. 3.3 units under the null hypothesis of an unlinked QTL. The reason for this discrepancy remains unclear. Table 3 reports the ML estimates of \(\alpha, \rho, \sigma, g, f_q\) and \(\tilde{f}_q\). Considering the dams’ gametic contributions has reduced the residual variance, thereby increasing the Q to q QTL alleles substitution effect to 1.8 standard deviations. The estimate of the Q allelic frequency has dropped from 0.40–0.49 to 0.22, indicating that the Q alleles might be less common in the dam than in the sire population. The number of generations to coalescence increases slightly (from 3.8 to 4.4), while the heterogeneity parameter, \(\rho\), drops slightly (from 0.92 to 0.81), which might be evidence for the existence of distinct Q alleles in the dam population.

For each marker locus, we identified the allele that yielded the highest likelihood (computed according to Equation 7) when considered associated with the hypothetical Q allele. When doing this at the respective positions that were associated with the highest signal from linkage disequilibrium, the same “ancestral haplotype” was obtained whether including information from the dams (map position 1; Figure 1, C and D) or not (map position 5; Figure 1, A and B) and is represented in Figure 2. As expected, it corresponds to the haplotype that was previously identified as being shared in total or in part by the fat-increasing chromosomes of the known \(Q_f\) founder sires. (RIQUET et al. 1999; B. GRISART, unpublished observations).

Whether considering only information from the sires’ gametes (Figure 1, A and B) or including information from the dams’ gametes (Figure 1, C and D), both the linkage and linkage disequilibrium signals obtained with the LINKLD program point toward a terminal location of the QTL on the centromeric end of chromosome 14. The BULGE36-BULGE17 interval, previously identified on the basis of haplotype sharing among \(Q_f\) sires (RIQUET et al. 1999; B. GRISART, unpublished observations), clearly lies outside and distal to the lod-2 support interval for the QTL.

**TABLE 3**

<table>
<thead>
<tr>
<th>Sires</th>
<th>(\alpha) (%)</th>
<th>(\sigma) (%)</th>
<th>(f_q)</th>
<th>(g)</th>
<th>(\rho)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>0.10</td>
<td>0.08</td>
<td>0.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L + LD</td>
<td>0.10</td>
<td>0.08</td>
<td>0.39</td>
<td>3.81</td>
<td>0.92</td>
</tr>
<tr>
<td>Sires + dam</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>0.11</td>
<td>0.06</td>
<td>0.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L + LD</td>
<td>0.11</td>
<td>0.06</td>
<td>0.22</td>
<td>4.45</td>
<td>0.81</td>
</tr>
</tbody>
</table>

The reason for this discrepancy remains unclear. Table 3 reports the ML estimates of \(\alpha, \rho, \sigma, g, f_q\) and \(\tilde{f}_q\). Considering the dams’ gametic contributions has reduced the residual variance, thereby increasing the Q to q QTL alleles substitution effect to 1.8 standard deviations. The estimate of the Q allelic frequency has dropped from 0.40–0.49 to 0.22, indicating that the Q alleles might be less common in the dam than in the sire population. The number of generations to coalescence increases slightly (from 3.8 to 4.4), while the heterogeneity parameter, \(\rho\), drops slightly (from 0.92 to 0.81), which might be evidence for the existence of distinct Q alleles in the dam population.

For each marker locus, we identified the allele that yielded the highest likelihood (computed according to Equation 7) when considered associated with the hypothetical Q allele. When doing this at the respective positions that were associated with the highest signal from linkage disequilibrium, the same “ancestral haplotype” was obtained whether including information from the dams (map position 1; Figure 1, C and D) or not (map position 5; Figure 1, A and B) and is represented in Figure 2. As expected, it corresponds to the haplotype that was previously identified as being shared in total or in part by the fat-increasing chromosomes of the known \(Q_f\) founder sires. (RIQUET et al. 1999; B. GRISART, unpublished observations).

Whether considering only information from the sires’ gametes (Figure 1, A and B) or including information from the dams’ gametes (Figure 1, C and D), both the linkage and linkage disequilibrium signals obtained with the LINKLD program point toward a terminal location of the QTL on the centromeric end of chromosome 14. The BULGE36-BULGE17 interval, previously identified on the basis of haplotype sharing among \(Q_f\) sires (RIQUET et al. 1999; B. GRISART, unpublished observations), clearly lies outside and distal to the lod-2 support interval for the QTL.

**Recombinant ancestral haplotype analysis supports the proximal location of the QTL:** To resolve this discrepancy, we performed a RAH analysis for the BULGE9-CSSM66 marker interval that spans the two conflicting QTL positions following the procedure described in MATERIALS AND METHODS. To increase the number of RAH, we genotyped an additional data set (data set IV) comprising 401 bulls corresponding to 39 paternal half-sib pedigrees for the markers in the region.

Figure 3 summarizes the obtained results. It can be seen that for the marker interval ILSTS39-CSSM66, the allelic effects of the centromeric RAH (CRAH-3 to CRAH-6) are superior to that of the corresponding telomeric RAH (TRAH-3 to TRAH-6), while for the most centromeric marker position (BULGE09) this tendency
Simultaneous Mining of Linkage and Linkage Disequilibrium

Figure 2.—Schematic representation of the ancestral haplotype identified with the LINKLD program in data sets I and III (AH-Dutch) in black and in data set V (AH-NZ) in gray. Also represented are the fat-increasing haplotypes of 11 sires shown to be heterozygous Qq by marker-assisted segregation analysis (Riquet et al. 1999; B. Grisart, unpublished data). Chromosome segments identical-by-state with the AH-Dutch haplotype are highlighted in black and those identical-by-state with the AH-NZ haplotype are highlighted in gray. DBWx are Dutch Black-and-White sires (data set I), DRWx are Dutch Red-and-White sires (data set III), and NZBWx correspond to the New Zealand Black-and-White sires (data set II).

These results therefore corroborate the LINKLD analysis and provide strong evidence against the location of the QTL in the ILSTS39-CSSM66 interval as initially inferred (Riquet et al. 1999; B. Grisart, unpublished data).

Understanding the origin of the discrepancy between the former (Riquet et al. 1999) and new estimates of the QTL map position: Assuming that the QTL is indeed positioned proximal to the ILSTS39 marker, this raises the issue of what could have caused its initial erroneous localization in the BULGE36-BULGE17 interval on the basis of the haplotype sharing among putative Qq sires (Riquet et al. 1999; B. Grisart, unpublished data). It is noteworthy in this respect that the sire (NZ1) that defined the proximal boundary of the BULGE36-BULGE17 interval was sampled in the New Zealand dairy cattle population (Figure 2). We therefore analyzed data set II, corresponding to the New Zealand portion of our original granddaughter design (Spelman et al. 1996; Coppieters et al. 1998a), with the LINKLD program. While the linkage component (H1) of this analysis generated convincing evidence for the segregation of a QTL with major effect on fat percentage on the centromeric end of chromosome 2 (zmax = 8.65 at the BULGE11 position) in this population as well, no clear supplementary signal could be obtained when counting for the linkage disequilibrium component (H2) in the analysis whether considering maternal chromosomes or not (data not shown). The reasons for this remain unclear.

To nevertheless obtain linkage disequilibrium information in the New Zealand population, we collected a daughter design comprising 51 paternal half-sib families for a total of 529 cows (data set V). This data set was
Figure 3.—Recombinant ancestral haplotype (RAH) analysis. The compositions of the ancestral haplotype (AH) and centromeric (CRAH-n) and telomeric (TRAH-n) recombinant ancestral haplotypes are shown as bars, respectively, at the top and bottom, aligned with respect to the positions of the proximal markers for chromosome 14. The marker interval identified in the Riquet et al. (1999) study corresponds to the stippled area. The light gray (shaded triangle) and black (solid diamond) lines in the graph connect the least-squares estimates of the allelic effects of the respective TRAH and CRAH. Least-squares estimates are flanked by error bars defining 95% confidence intervals (±2 standard errors of the estimate). The dotted line marks the estimate of the allelic effect of the OH (all-white) haplotype class (see materials and methods). The dark gray (shaded box) line connects the least-squares estimates of the interval-specific contrasts flanked by error bars defined as above. The grouping of the RAH to generate the respective interval-specific contrast is marked by the shaded ovals underneath the haplotype bars. Haplotypes marked by a dark oval belong to group I and those marked by a light oval to group II.

Genotyped for proximal chromosome 14 markers and the resulting genotypes were analyzed with the LINKLD programs using lactation values as phenotype. Figure 4 summarizes the results that were obtained, using an unlinked QTL as null hypothesis ($H_0$) and when considering the maternal haplotypes in the analysis. The signal from linkage maximizes at the position of the second marker (BULGE11; $z_{max} = 3.37$). A maximum of 0.5 lod score units were gained from considering linkage disequilibrium at the position of the third marker (ILSTS39). Despite the relatively modest lod score values, these results were again pointing toward a location
Simultaneous Mining of Linkage and Linkage Disequilibrium

Figure 4.—Lod scores obtained when analyzing data set V, using both sire and dam information and an unlinked QTL as null hypothesis \( (H_0) \). The black curve (solid diamond) corresponds to \( (1/M) \times \log(H_2/H_0) \), therefore extracting information from both linkage and LD, the light gray curve (shaded square) to \( (1/M) \times \log(H_1/H_0) \), therefore extracting information from linkage only, and the dark gray curve (shaded triangle; scaled to 1/10) to \( (1/M) \times \log(H_2/H_1) \) corresponds to LD only. The cross-hatched area corresponds to the interval selected in Riquet et al. (1999) and subsequently refined by B. Grisart (unpublished data).

of the QTL proximal from the BULGE34-BULGE17 interval.

The corresponding maximum-likelihood (ML) estimators of \( \alpha, \sigma, f_q, g, \) and \( p \) were, respectively, 0.15%, 0.13%, 0.32, 6, and 0.36. The allele substitution effect has increased from 0.11% to 0.15%. This is not unexpected given the fact that in this case we analyzed lactation values rather than daughter yield deviations (which correspond to one-half breeding values). The increased residual variance is as expected as well given the lower reliability of lactation values when compared to daughter yield deviations. The frequency of the \( Q \) allele is estimated at 0.32 in the New Zealand population, which is therefore in the same range as in the Dutch population. The value of 0.36 for \( p \) suggests the occurrence of distinct subgroups of \( Q \) QTL alleles in the New Zealand population.

We then identified the most likely ancestral haplotype from the analysis of data set V following the procedure described above. As shown in Figure 2, in the BULGE09-BULGE36 interval the resulting AH proved to be identical with the haplotype from sire NZ1 that was associated with an increase in fat percentage. This therefore suggested that in the New Zealand population, a fat-increasing allele might be preferentially associated with a marker haplotype distinct from the one identified in the Dutch population.

To verify this, we computed the average substitution effect of this newly identified haplotype (BULGE9-BULGE30), using daughters whose sire was heterozygous for this haplotype. Seventeen such heterozygous sires were identified in data set V for a total of 183 daughters. Average lactation values corrected for paternal breeding value were compared between daughters that received the haplotype \( vs. \) those that did not. This yielded a highly significant substitution effect of +0.12% \( (p < 1E^-7) \) as expected. This observation, associated with the fact that the fat-increasing haplotype of the NZ1 sire was fortuitously identical-by-state with the Dutch AH in the BULGE34-BULGE4 interval, likely caused the erroneous QTL localization in the previous analyses (Riquet et al. 1999; B. Grisart, unpublished data).

The average substitution effect of the Dutch AH (BULGE09-BULGE30) in the New Zealand population
was estimated using the same approach; i.e., average lactation values corrected for paternal breeding value were compared between daughters of heterozygous sires that received the Dutch haplotypes vs. those that did not. Seventeen sires heterozygous for the Dutch haplotype were identified in data set V for a total of 172 daughters. This analysis yielded an average substitution effect of +0.073% (p < 1E - 3), therefore confirming the increasing effect on fat percentage of the Dutch AH in the NZ population as well.

**DISCUSSION**

We herein describe a method aimed at simultaneously exploiting information from linkage and linkage disequilibrium to improve the power and resolution of QTL mapping in outbred half-sib pedigrees. The proposed approach is an extension of a multipoint association method developed by Terwilliger (1995) for the fine mapping of monogenic disease traits. It has been extended to (i) allow for the combined analysis of linkage and linkage disequilibrium in half-sib pedigrees and (ii) be applicable to quantitative traits. The principle underlying the proposed method consists of establishing a parallel between a disease-causing allele, \( D \), in the case of a monogenic disease, and a QTL allele with large substitution effect, \( Q \), in the case of a quantitative trait.

We demonstrate in this work that additional mapping information can indeed be obtained from linkage disequilibrium using the proposed approach. A drawback of the method as presently described is the unknown distribution of the generated lod scores under the null hypotheses of no QTL or unlinked QTL. As a consequence, one cannot confidently assign a significance level to the obtained LD signal. We are presently attempting to address this issue “empirically” using permutation tests (e.g., Churchill and Doerge 1995). In these, the transmission probabilities of alternate paternal homologs to individual sons \( P(\text{Son}_n, A(p)|M) \) and \( P(\text{Son}_n, B(p)|M) \) are maintained, thereby conserving the mapping information from linkage. However, marker haplotypes of the sires and maternal haplotypes of the sons are subjected to permutation. This effectively disconnects QTL and marker genotype while maintaining the actual LD between marker loci. This empirical approach is tedious, however, given the requirement for long computation times and the difficulties encountered to find the likelihood maxima using the available numerical recipes.

The proposed approach allows for the extraction of information from the maternal gametes. This information is ignored by most of the analysis methods that are the most commonly applied to granddaughter designs (e.g., Mackinnon and Weller 1995; Knott et al. 1996; Coppiters et al. 1998b). As presently formulated, the information from maternal chromosomes is obtained using an association study. This contrasts with the linkage disequilibrium information extracted from the sire chromosomes, which in essence corresponds to the more robust transmission disequilibrium test (Spielman et al. 1993). This approach is therefore susceptible to yielding associations independent of chromosomal localization in the case of population stratification. We therefore recommend considering the obtained result with caution. The algorithm could easily be reformulated to extract the maternal information using a transmission disequilibrium approach as well. Such an approach would, however, require genotyping of the dams, which is usually not practical.

We demonstrate in this work that the choice of null hypothesis, e.g., \( H_0: \alpha = 0 \) or \( H_\alpha: \theta = 0.5 \), can have a major effect on the resulting lod scores. More specifically, we show an instance in which the use of \( H_\alpha: \alpha = 0 \) tends to inflate the lod scores by up to 40 units. Note that this null hypothesis is the one that is generally used by QTL mappers. Alternatively, we have encountered situations (data not shown) where the real presence of an unlinked QTL with major effect on the examined trait substantially decreases the obtained lod scores when looking for QTL on a given chromosome and using \( H_\alpha: \theta = 0.5 \). These observations therefore stress the need to generate significance thresholds adapted to specific circumstances (for instance by permutation), particularly when using ML approaches.

Estimation of the heterogeneity parameter, \( \rho \), supports a relatively high homogeneity of marker haplotypes flanking the \( Q \) allele in the Dutch population (\( \rho = 0.81-0.92 \)). We provide strong evidence, however, for the association of \( Q \) alleles increasing fat percentage with distinct marker haplotypes in the New Zealand population. It has been argued that one of the advantages of Terwilliger’s method is that it would be robust against haplotype heterogeneity within the population of \( D \) bearing chromosomes (Terwilliger and Weiss 1998). A distinct difference between disease-causing alleles (\( D \)) and QTL alleles with large substitution effects (\( Q \)), however, is that one can envisage situations where distinct haplotypes are associated with \( Q \) alleles, some having positive effects on the trait of interest and others having negative effects. Preliminary evidence suggests that this might occur in the New Zealand population. It remains to be established how the proposed method would behave in such situations.

While this work demonstrates that it is possible to extract linkage disequilibrium information using the available medium density maps, it remains uncertain whether this will yield a significant gain in power and—more importantly—mapping resolution. While we believe that this work provides strong evidence that the QTL on bovine chromosome 14 is actually located proximal to the previously identified BULGE34-BULGE17 interval, this evidence is obtained as much from linkage as from linkage disequilibrium information. There seems to be as much variation in the most likely position of the
QTL across experiments, whether one considers information from linkage or linkage disequilibrium. We nevertheless believe that, at present, exploiting linkage disequilibrium remains one of the most promising strategies to increase mapping resolution and power in, for instance, dairy cattle. This work, as well as that from others (e.g., Meuwissen and Goddard 2000), is a first step toward that goal.

We are grateful to CR-DELTA and LIC for providing us with breeding values. We thank Richard Spelman and Johan van Arendonk for fruitful discussions. This work was funded by grants from Holland Genetics, Livestock Improvement Corporation (LIC), the Vlaamse Rundvee Vereniging, the Ministère des Classes Moyennes et de l’Agriculture, Belgium, and E.U. grants B104-CT95-0073 and PL970471.

LITERATURE CITED


