Copyright © 2003 by the Genetics Society of America

Molecular Dissection of a Quantitative Trait Locus: A Phenylalanine-to-Tyrosine Substitution in the Transmembrane Domain of the Bovine Growth Hormone Receptor Is Associated With a Major Effect on Milk Yield and Composition

Sarah Blott,* Jong-Joo Kim,* Sirja Moisio,† Anne Schmidt-Küntzel,* Anne Cornet,* Paulette Berzi,* Nadine Cambisano,* Christine Ford,‡ Bernard Grisart,* Dave Johnson,§ Latifa Karim,* Patricia Simon,* Russell Snell,‡ Richard Spelman,§ Jerry Wong,‡ Johanna Vilkki,† Michel Georges,*,[1] Frédéric Farnir* and Wouter Coppieters*

**Department of Genetics, Faculty of Veterinary Medicine, University of Liège, 4000 Liège, Belgium, †Animal Production Research, Agricultural Research Centre MTT, 31600 Jokioinen, Finland, §Livestock Improvement Corporation, Hamilton, New Zealand and ‡Vialactia Biosciences (NZ) Limited, Clinical Building, University of Auckland Medical School, Auckland, New Zealand*

Manuscript received July 1, 2002
Accepted for publication October 9, 2002

ABSTRACT

We herein report on our efforts to improve the mapping resolution of a QTL with major effect on milk yield and composition that was previously mapped to bovine chromosome 20. By using a denser chromosome 20 marker map and by exploiting linkage disequilibrium using two distinct approaches, we provide strong evidence that a chromosome segment including the gene coding for the growth hormone receptor accounts for at least part of the chromosome 20 QTL effect. By sequencing individuals with known QTL genotype, we identify an *F* to *Y* substitution in the transmembrane domain of the growth hormone receptor gene that is associated with a strong effect on milk yield and composition in the general population.

WITH the development of genome-wide marker maps for several species it is becoming possible to map quantitative trait loci (QTL) underlying the genetic variation for continuously distributed phenotypes of medical and agronomical importance (*e.g.*, Andersson 2001; Flint and Mott 2001; Mackay 2001; Mauricio 2001). The mapping resolution that is achieved with most experimental designs, however, is in the tens of centimorgans at best, therefore precluding efficient marker-assisted selection let alone positional cloning of the corresponding genes.

When working with model organisms or plants, strategies to improve the mapping resolution most often require breeding of a large number of progeny to increase the density of crossovers in the chromosome regions of interest (*e.g.*, Darvasi 1998). When working with humans or farm animals, this approach is not practical and alternative strategies need to be identified. One approach that has recently received considerable attention is linkage disequilibrium (LD) mapping, which aims at exploiting historical recombinants. However, as useful LD is expected to extend only over limited distances (*e.g.*, 60 kb in the human; Reich *et al.* 2001) this approach requires a commensurate increase in marker density.

In some livestock populations, including dairy cattle, LD has been shown to extend over very long chromosome segments when compared to human populations (Farnir *et al.* 2000). This has raised hope that LD would be readily exploitable in these species using the presently available medium density maps (*e.g.*, Kappes *et al.* 1997). These hopes have been supported by a series of initial, promising results dealing with both simple (*e.g.*, Charlier *et al.* 1996) and complex inherited traits (*e.g.*, Grisart *et al.* 2002). Potential downsides associated with this long-range LD are (i) a possible limited mapping resolution and (ii) the occurrence of association in the absence of linkage due to gametic association between nonsyntenic loci.

We herein use LD to improve our molecular understanding of a QTL influencing milk yield and composition that was previously mapped to bovine chromosome 20 (Georges *et al.* 1995; Arranz *et al.* 1998).

MATERIALS AND METHODS

Pedigree material: The pedigree material used in this study was composed of the following:

Data set I: a previously described Black-and-White Holstein-Friesian granddaughter design (GDD) sampled in the Netherlands and composed of 22 paternal half-sib families for a total of 987 bulls (Spelman *et al.* 1996; Coppieters *et al.* 1998a).

Data set II: 276 progeny-tested Holstein-Friesian sires sampled in the Netherlands.

Data set III: 1550 progeny-tested Holstein-Friesian sires sampled in New Zealand.

[1]*Corresponding author:* Department of Genetics, Faculty of Veterinary Medicine, University of Liège (B43), 20 Bd. de Colonster, 4000-Liège, Belgium. E-mail: michel.georges@ulg.ac.be.

Data set IV: 959 progeny-tested Jersey sires sampled in New Zealand.
Data set V: 485 Holstein-Friesian cows sampled in New Zealand.
Data set VI: 387 Jersey cows sampled in New Zealand.

Phenotypes: Phenotypes were, respectively, daughter yield deviations (DYD) for bulls, lactation values (LV, the unregressed first lactation yield deviations) for cows, average parental predicted transmitting abilities (PTA) for bulls and cows for milk, protein, and fat yield, as well as protein and fat percentage (Van Raden and Wiggans 1991). DYDs, lactation values, and PTA were obtained directly from CR-DELTA (Netherlands; data sets I and II) or Livestock Improvement Corporation (LIC, New Zealand; data sets III–VI), respectively.

Map construction: Microsatellite genotyping, map construction, and information content mapping were performed as previously described (Coppieters *et al.* 1998a). Sequence information for the primers used for PCR amplification of anonymous type II microsatellite markers can be obtained from *ArkDB* (http://www.thearkdb.org/species.html). The following primers were designed on the basis of Heap *et al.* (1995) to amplify a microsatellite in the promotor region of the growth hormone receptor gene: GHRJA.UP, 5′-TGCTCTAATCTTTT CTGGTACCAGG-3′, and GHRJA.DN, 5′-TCCTCCCCAAAT CAATTACATTTTCTC-3′.

Conventional QTL mapping: QTL mapping was performed by multimarker regression (Knott *et al.* 1996) using the previously described HSQM software (Coppieters *et al.* 1998b). Chromosome-wide significance thresholds were determined by permutation as previously described (Churchill and Doerge 1995; Coppieters *et al.* 1998b). Segregating sire families were identified on the basis of the results of within-family analyses as previously described (Coppieters *et al.* 1998a).

Haplotype-based test for association: *Assumptions:* Assume a QTL that is characterized by two additively acting alleles, Q and q, that segregate in the population of interest with respective allelic frequencies of f_Q and $(1 - f_Q)$. Assume that the Q allele appeared in the population by mutation or migration on a chromosome with *h*aplotype "H" for a series of flanking markers. All *o*ther haplotypes are pooled and referred to as "O." At the present generation the H haplotype may still be in LD with the Q allele by an amount D. The H to O haplotype substitution effect can then be shown to equal

$$\alpha = a\frac{D}{f_H(1 - f_H)},$$

where a corresponds to one-half the difference between the phenotypic values of QQ *vs.* qq individuals, and f_H corresponds to the population frequency of the H haplotype (Falconer and Mackay 1996).

Test for association: Knowing that in our GDD, phased-marker genotypes are available for all sons and their sires but *not* their dams as these are not marker genotyped, and defining T_i as $[DYD_i - PA_i]$, where DYD_i is the daughter yield deviation of son i and PA_i is the average predicted transmitting ability (Van Raden and Wiggans 1991) of the sire and dam of son i, one can express the expected value of T_i as a function of the marker genotype of the sire's chromosomes (SC) and the marker genotypes of the paternal (PC) and maternal gametes (MC) inherited by son i, as shown in Table 1.

Expected values of T_i can be seen to be linear functions of the unknown haplotype substitution effect, α. A least-squares estimator of α can therefore easily be obtained by linear regression, while the ratio

$$\frac{\text{SSR}}{\text{SSE}/(n - 2)},$$

which is distributed as an F-statistic with 1 and $n - 2$ d.f., can be used to measure the evidence in favor of a statistically significant haplotype substitution effect. In this n corresponds to the number of sons available in the GDD, SSE to the residual sum of squares, and SSR to the regression sum of squares.

By using T_i as phenotype, one essentially performs a transmission disequilibrium test (TDT; Spielman *et al.* 1993), which simultaneously tests for association and linkage. As the dams are not genotyped, however, our TDT reduces in part to a conventional association test.

Choice of markers and haplotypes: So far, we have not defined which of the m markers available on the chromosome have to be considered when defining a haplotype. As the exact location of the QTL and the size of the haplotype that will maximize α are both unknown, all possible windows comprising between one and m adjacent markers are tested separately. We thus examine m windows of one marker, $(m - 1)$ windows of two markers, $(m - 2)$ windows of three markers, . . . , and one window of m markers.

Having selected the markers comprising the haplotype, one still has to choose the H haplotype among all haplotypes encountered in the population. In the proposed approach, the haplotypes that were successively considered as H haplotypes correspond to the chromosomes of the s sires in the GDD that were known to be heterozygous Qq for the QTL on the basis of the results of a marker-assisted segregation analysis performed on their sons (see above). As one does not *a priori* know which of the sire's homologs carries the Q allele, the haplotypes corresponding to both chromosomes are examined, for a total of $2s$ homologs.

When estimating the substitution effect of the haplotypes of a given sire, his sons were eliminated from the data set, to avoid extracting information that would be redundant with the linkage analysis.

Significance thresholds: The F-ratio defined above does not account for the multiple tests that are performed, *i.e.*, the $(m^2 + m)/2$ marker windows tested for each of the $2s$ homologs. We accounted for multiple testing by applying a permutation test. The phenotypes and marker genotypes were shuffled 1000 times and the $2s(m^2 + m)/2$ tests were performed on each permutated data set. The highest F-ratios obtained with the real data were then compared with the highest F-ratios obtained across the 1000 permutations.

Simultaneous mining of linkage and linkage disequilibrium: *QTL fine-mapping exploiting both linkage and LD:* The utilized mapping method is implemented in the LD variance component mapping (LDVCM) programs and can be summarized as follows. Testing for the presence of a QTL at map position p of the studied chromosome was performed as follows:

1. For all markers on the studied chromosome, we determine the marker linkage phase of the sires and sons as described (Farnir *et al.* 2002). As a consequence, the marker data then consist of $2s$ sire chromosomes (SC), n paternally inherited chromosomes of the sons (PC), and n maternally inherited chromosomes of the sons (MC), where s and n correspond, respectively, to the number of sire families and the number of sons in the GDD. From the genotypes of the PC, we can easily compute the probability that son i inherited the "left" (λ_p) or "right" ($\rho_p = 1 - \lambda_p$) SC from its sire at map position p as described (Coppieters *et al.* 1998b).
2. We compute identity-by-descent (IBD) probabilities (ϕ_p) for all pairwise combinations of SC and MC using the method described by Meuwissen and Goddard (2001). This method approximates the probability that two chromosomes are IBD at a given map position conditional on the identity-by-state (IBS) status of flanking markers, on

TABLE 1

Expected values of T (= DYD − PA) as a function of the marker genotype of the sire and the marker genotypes of the paternal and maternal gametes inherited by the son

Paternal genotype (SC)	Paternal gamete (PC)	Maternal gamete (MC)	
		H	O
HH	H	$T = \frac{1}{2}\alpha\, f_H$	$T = -\frac{1}{2}\alpha(1 - f_H)$
HO	H	$T = \frac{1}{2}\alpha(1 + f_H)$	$T = \frac{1}{2}\alpha\, f_H$
	O	$T = -\frac{1}{2}\alpha(1 - f_H)$	$T = -\frac{1}{2}\alpha(2 - f_H)$
OO	O	$T = \frac{1}{2}\alpha\, f_H$	$T = -\frac{1}{2}\alpha(1 - f_H)$

SC, PC, MC, H, O, α, and f_H are as defined in MATERIALS AND METHODS.

the basis of coalescent theory (HUDSON 1985). Windows of 16 markers were considered to compute ϕ_p.

3. Using $(1 - \phi_p)$ as a distance measure, we apply the UPGMA hierarchical clustering algorithm (*e.g.*, MOUNT 2001) to generate a rooted dendrogram representing the genetic relationship—at position p—between all SC and MC haplotypes encountered in the population.
4. We use the logical framework provided by this dendrogram to group the SC and MC in functionally distinct clusters. A cluster is defined as a group of haplotypes that coalesce into a common node. A useful feature of UPGMA trees in this regard is that the distance $(1 - \phi_p)$ between all the haplotypes that coalesce into a given node is less than or equal to two times the distance between the node and any of these haplotypes. As a consequence, the tree is scanned downward from the root and branches are cut until nodes are reached such that all coalescing haplotypes (*i.e.*, all haplotypes within the cluster) have a distance measure $(1 - \phi_p) < T$ (KIM and GEORGES 2002).
5. We model the sons' phenotypes (DYDs) using the following linear model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z_h h} + \mathbf{Z_u u} + \mathbf{e}.$$

y is the vector of phenotype records of all sons. **b** is a vector of fixed effects, which in this study reduces to the overall mean. **X** is the incidence matrix relating fixed effects to individual sons, which in this study reduces to a vector of ones. **h** is the vector of random QTL effects corresponding to the defined haplotype clusters. $\mathbf{Z_h}$ is an incidence matrix relating haplotype clusters to individual sons. In $\mathbf{Z_h}$, a maximum of three elements per line can have nonzero value: "1" in the column corresponding to the cluster to which the MC haplotype belongs and λ_p and ρ_p in the columns corresponding, respectively, to the haplotype clusters of the right and left SC. If the SC and/or MC belong to the same cluster, the corresponding coefficients are added. **u** is the vector of random individual polygenic effects ("animal model", LYNCH and WALSH 1997). $\mathbf{Z_u}$ is a diagonal incidence matrix relating individual polygenic effects to individual sons. **e** is the vector of individual error terms.

Haplotype cluster effects with corresponding variance, σ^2_H, individual polygenic effects with corresponding variance, σ^2_A, and individual error terms with corresponding variance, σ^2_E, were estimated using average information restricted maximum likelihood (AIREML; JOHNSON and THOMPSON 1995), by maximizing the restricted log-likelihood function L:

$$L = -0.5\ln|\mathbf{V}| - 0.5\ln|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}| - 0.5(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}).$$

In this,

$$\mathbf{V} = \sigma^2_H\mathbf{Z_h}\mathbf{H}\mathbf{Z_h^T} + \sigma^2_A\mathbf{Z_u}\mathbf{A}\mathbf{Z_u^T} + \sigma^2_E\mathbf{I}.$$

Because we assume that the covariance between the QTL effects of the different haplotype clusters is zero, **H** reduces to an identity matrix. This differentiates our approach from that of MEUWISSEN and GODDARD (2000), in which **H** is the matrix of between-haplotype IBD probabilities. **A** is the additive genetic relationship matrix (LYNCH and WALSH 1997).

6. Steps 4 and 5 are repeated for all possible values of T (from 0 to 1), to identify a restricted maximum likelihood (REML) solution for map position p. By analogy with FARNIR *et al.* (2002), we denote the hypothesis corresponding to this REML solution as H_2.

QTL mapping exploiting linkage only: Note that the previous model can be extended with minor modifications to map QTL by exploiting linkage information only. This is simply achieved by ignoring all MCs and considering that all SCs belong to distinct haplotype clusters, irrespective of their marker genotype. REML solutions for the different parameters can be found as described in the previous section. Again by analogy with FARNIR *et al.* (2002), we refer to the corresponding hypothesis as H_1.

Hypothesis testing and significance thresholds: The log likelihoods of the data under the H_2 and H_1 hypotheses are compared with that under the null hypothesis, H_0, of no QTL at map position p. The latter is computed as described above but using the reduced model,

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{Z_u u} + \mathbf{e}.$$

Evidence in favor of a QTL at map position p can then be expressed as a lod score:

$$z_p = \frac{(L_{H_{1/2}} - L_{H_0})}{\ln(10)}.$$

As customary when performing interval mapping, we are sliding the hypothetical position of the QTL throughout the chromosome map and compute lod scores at each map position as described to generate chromosome-wide lod score profiles.

KIM and GEORGES (2002) have shown by simulation that when analyzing a chromosome of 100 cM with a marker density of one marker every 5 cM, $2 \times (L_{H_{1/2}} - L_{H_0})$ has (under the null hypothesis) an approximate chi-square distribution with 2 d.f. corrected (Bonferroni correction) for two and six independent tests when considering, respectively, H_1 and H_2. Chromosome-wide significance levels were computed from these distributions in this study.

Sequencing the coding portion of the growth hormone receptor from genomic DNA: To develop primers that would allow us to conveniently amplify and sequence the entire

TABLE 2

Primers used for amplification and sequencing of the GHR exons from bovine genomic DNA

GHRex3_F	TAG GAG TTC CTT TTA GAG GAT AGG TGC
GHRex3_R	GCC TTG TGG AGA AGT TGA CAA A
GHRex4_F	GCC CAG AGA AAC AGC ATT TCT A
GHRex4_R	TCA CTG CCA TAT TTC CAG CAT C
GHRex5_F	CTT GCT CAT AAA ATA CTC GTG TCC T
GHRex5_R	ATG CAA TGG CAA AGT CTT CCT AC
GHRex6_F	TGT ATG AAG TAA CTT AGT CGT CTT CG
GHRex6_R	GAG AGG GGT TGT TGA ACA CAA A
GHRex7_F	TCC TAC TTT CCA GAA ATT CAT TTT G
GHRex7_R	CTG AGG CTA ATG TAT ATT GAT CTG GAC
GHRex8_F	GTG GCT ATC AAG TGA AAT CAT TGA C
GHRex8_R	ACT GGG TTG ATG AAA CAC TTC ACT C
GHRex9_F	GCC TCA TCA TTC ACT GCT TA
GHRex9_R	GGT TTC AAC ATA AGG CTC TG
GHRex10_F	ACA TGG TTT GTT ATA TGA TTT TGT TAC
GHRex10_R	TTC ATA TTC CCC ACC CTC AAC T
GHRex10_1F	ACA TTC TGG AGG CTG ATT TC
GHRex10_2F	CAA AAG AAT AAG ACT GGG AA
GHRex10_1R	AGC TTG GCT CTA CGT GTG AT
GHRex10_2R	GAT AAC ACT GGG CTG CTG GT

All primer sequences are written 5′ → 3′. All exons were PCR amplified and sequenced with the same primers except for exon 10, which was amplified with GHRex10_F and GHRex10_R and then sequenced with these primers plus GHRex10_1F, GHRex10_1R, GHRex10_2F, and GHRex10_2R.

growth hormone receptor (GHR) coding sequence from bovine genomic DNA, we screened a bovine bacterial artificial chromosome (BAC) library (Warren *et al.* 2000) using standard procedures with an oligonucleotide probe complementary to exon 10 and isolated eight GHR-containing clones. DNA from one of these clones was used as template for sequencing the intron-exon boundaries using exonic primers designed on the basis of the bovine cDNA sequence (*e.g.*, Hauser *et al.* 1990) and predicted to flank exon-intron boundaries assuming conservation of intron position between human and cattle (*e.g.*, Godowski *et al.* 1989). On the basis of the obtained intronic information, we then designed primers (Table 2) to amplify and sequence most of the GHR coding sequence from genomic DNA using standard procedures. Sequence traces were analyzed with the POLYPHRED software (Nickerson *et al.* 1997).

Oligonucleotide ligation assay: An oligonucleotide ligation assay (OLA) test to genotype the GHR *F279Y*, *Nt864-33(T-G)*, *Nt933+21(A-G)*, *Nt1095(T-C)*, *N528T*, and *Nt1922(C-T)* single-nucleotide polymorphisms (SNPs) in multiplex was developed as previously described (Karim *et al.* 2000). The primers used for the PCR amplification step and the ligation reaction are reported in Table 3.

Estimating the effect on milk yield and composition associated with the *F279Y* polymorphism in the general dairy cattle population: The effect of the *F279Y* genotype on milk yield and composition was estimated using the model

$$y_i = \mu + g_i + a_i + e_i,$$

where y_i were DYDs when studying bulls or lactation values when studying cows, g_i is a fixed effect corresponding to the *F279Y* genotype (*FF*, *FY*, or *YY*), a_i is a random polygenic component accounting for all known pedigree relationships ["animal model" (Lynch and Walsh 1997) including ungenotyped individuals whose phenotypes were ignored], and e_i is a random residual. Maximum-likelihood solutions for g_i, a_i, and e_i were obtained using the MTDFREML program (Boldman 1993), setting $\sigma_a^2/(\sigma_a^2 + \sigma_e^2)$ for yield (percentage) traits at 70% (75%) and 35% (50%) for DYDs and LVs, respectively.

The statistical significance of the *F279Y* genotype effect was estimated from

$$\frac{(n - 3) \times (\mathrm{SSM_F} - \mathrm{SSM_R})}{3 \times \mathrm{SSE_F}},$$

where $\mathrm{SSM_F}$, $\mathrm{SSM_R}$, and $\mathrm{SSE_F}$ are the sum of squares due to the full model, reduced model, and error (full model), respectively, which is distributed as an *F*-statistic with 3 and $(n - 3)$ d.f.

The proportion of the trait variability due to the QTL genotype (r^2_{QTL}) was approximated as

$$r^2_{\mathrm{QTL}} = \frac{f_{FF}(\overline{FF} - \overline{T})^2 + f_{FY}(\overline{FY} - \overline{T})^2 + f_{YY}(\overline{YY} - \overline{T})^2}{\mathrm{SST}},$$

where f_{FF}, f_{FY}, and f_{YY} are the frequencies of the *FF*, *FY*, and *YY* genotypes; $\overline{FF}$, $\overline{FY}$, and $\overline{YY}$ are the genotype means estimated using the mixed model, SST is the total sum of squares, and $\overline{T}$ is the overall trait mean estimated as

$$\overline{T} = f_{FF}\overline{FF} + f_{FY}\overline{FY} + f_{YY}\overline{YY}.$$

RESULTS

Construction of a high-density microsatellite map of bovine chromosome 20: To refine the map position of the chromosome 20 QTL, we first increased the marker density on this chromosome. We genotyped data set I for 22 additional, publicly available microsatellites known to map to bovine chromosome 20 as well as for a microsatellite in the promotor region of the bovine growth hormone receptor gene (GHRJA). A male linkage map composed of 29 markers covering 85 cM [Kosambi (K)] was constructed with average marker interval of 3 cM (K; Figure 1). The information content of the corresponding map was computed as previously described (Coppieters *et al.* 1998a). It is >80% for most of the chromosome length. This is considerably larger than the average marker heterozygosity among the sires of data set I (45%), which illustrates the benefit of multipoint mapping methods. The map, shown in Figure 1, also reports the position of the prolactin receptor (PRLR) gene deduced from segregation data of prolactin receptor SNPs in the same pedigree material (S. Moisio, unpublished observations). Note that in the human, the GHR gene is located in band 5p13.3 at map position 37.4 Mb on the "golden path" human sequence (Ensembl Human Genome Server: http://www.ensembl.org), while the PRLR gene is located in band 5p13.1 at map position 50.9 Mb, *i.e.*, at ~15 Mb from the former. The genetic distance separating the bovine GHR and PRLR genes is therefore compatible with the human data.

Conventional QTL mapping using a dense marker map: We then used these novel microsatellite genotypes to repeat a QTL-mapping analysis in data set I. Figure 1 reports the location scores that were obtained by multimarker regression in the across-family analysis

TABLE 3

Primers (5′–3′) used for OLA multiplexing of GHR SNPs

	PCR	
SNP	UP	DN
F279Y	GTGGCTATCAAGTGAAATCATTGAC	ACTGGGTTGATGAAACACTTCACTC
Nt864-33(T-G)	GCCTCATCATTCACTGCTTA	GGTTTCAACATAAGGCTCTG
Nt933+21(A-G)	GCCTCATCATTCACTGCTTA	GGTTTCAACATAAGGCTCTG
Nt1095(T-C)	TCTTGGGTTGAATTTATTGAACTAG	ATCGCACATGTCACTGACATGGAA
N528T	ACATTACACCAGCAGGAAATGTGGT	TGGTAAGGCTTTCTGTGGTGATGT
Nt1922(C-T)	AGGTCGGGGACAGCAGAACAT	TCATATTCCCCACCCTCAACTCA

	OLA		
SNP	AS1	AS2	C
F279Y	Fam-GGGCTAGCATGTACATTATA	Hex-GGGCTAGCATGTACATTATT	P-TTTACTCATATTTTCTAAACAGC
Nt864-33(T-G)	Fam-GTCTTTTGAAATGAGATGAGG	Hex-GTCTTTTGAAATGAGAATGAGT	P-AACTTACATCAAAACAAAATTTTG
Nt933+21(A-G)	Fam-CAAGGTAATTAATAAAATATCTAAG	Hex-CAAGGTAATTAATAAAATATCTAAA	P-TTGTACATGACACTAATTAAATGT
Nt1095(T-C)	Fam-CAGACTTCTGAGCAATGAC	Hex-CAGACTTCTGAGCAATGAT	P-CATGAAAAATCACTCAATATC
N528T	Fam-TAACTTCATCGTGGACAC	Hex-TAACTTCATCGTGGACAA	P-CGCTTACTTCTGCGAG
Nt1922(C-T)	Fam-AGCTTTTCTTTGATTTCCC	Hex-AGCTTTTCTTTGATTTCCT	P-ATGAGCTACCCATTGAAT

UP and DN are the primers used for PCR amplification. AS1, AS2, and C are, respectively, the two allele-specific primers as well as the common primer used for the OLA reaction.

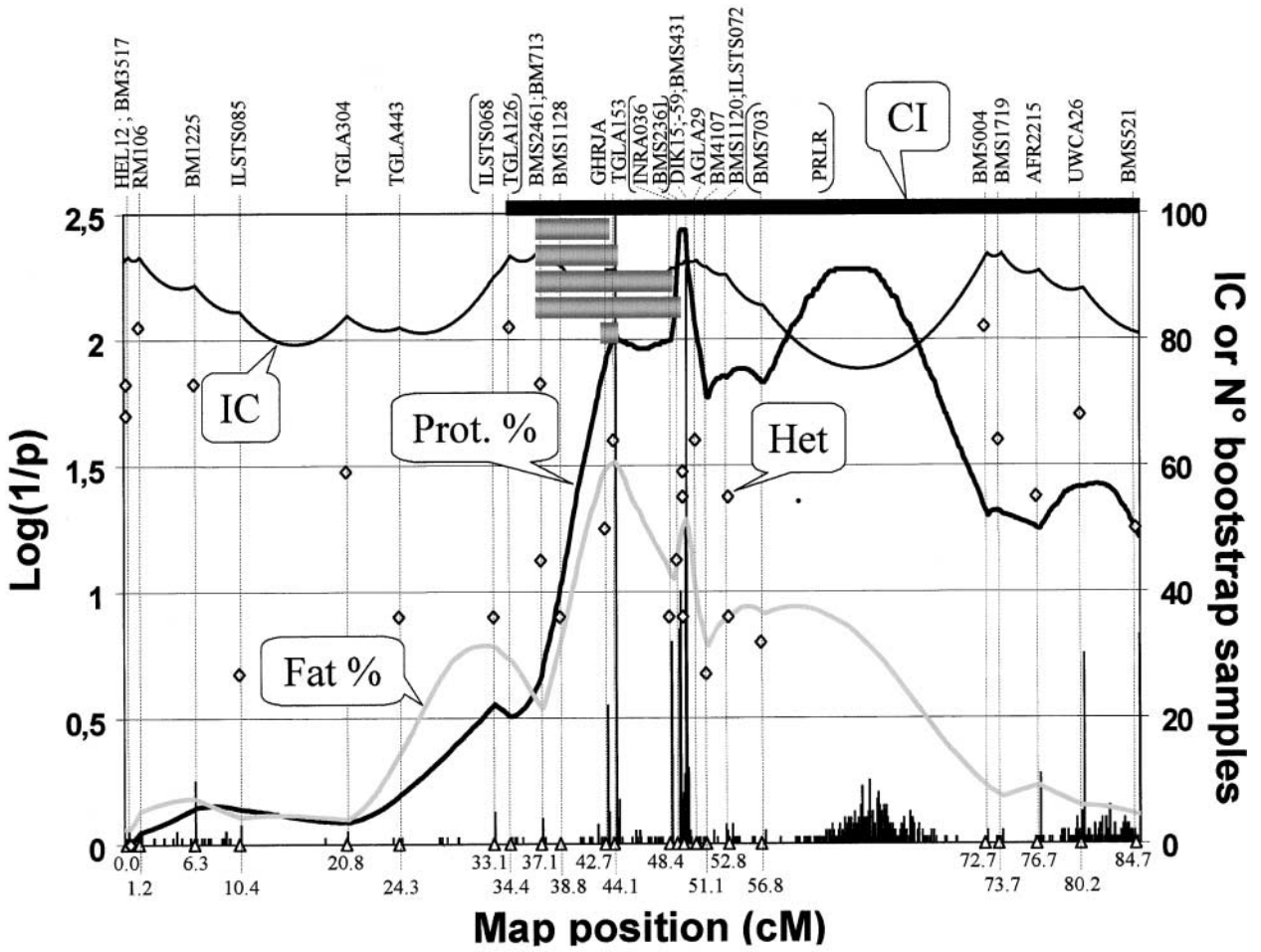

FIGURE 1.—Chromosome 20 microsatellite map, conventional QTL mapping, and haplotype-based test for association. Chromosome 20 microsatellite map: the name of the corresponding markers is given at the top and their respective position in centimorgans (Kosambi) at the bottom. GHRJA corresponds to a microsatellite marker in the promotor of the growth hormone receptor gene. The most likely position of the prolactin receptor (PRLR) inferred from the segregation of SNP markers (S. MOISIO, unpublished results) is given. Markers that could not be ordered with odds >1000 are in parentheses. The thin black curve corresponds to the information content (IC; expressed as a percentage—right *y*-axis) obtained in the GDD. The open diamonds correspond to the percentage of founder sires in the GDD that are heterozygous for the corresponding microsatellite marker. Conventional QTL mapping: the thick black and gray curves correspond to the location scores obtained, respectively, for milk protein percentage and milk fat percentage. Location scores are expressed as $\log(1/p)$ (left *y*-axis), where p corresponds to the chromosome-wide probability of obtaining the corresponding signal under the null hypothesis of no QTL determined by phenotype permutation. Most likely QTL positions obtained across 1000 bootstrap samples (left *y*-axis) are given as black vertical bars. The resulting 95% confidence interval is shown as a thick black horizontal bar at the top (C.I.). Haplotype-based test for association: marker windows showing significant effects in the haplotype-based association test are shown as gray cylinders. Their position with respect to the left *y*-axis corresponds approximately to their significance level determined as described in MATERIALS AND METHODS.

along the newly generated chromosome 20 marker map. As expected, these results confirm the presence of a QTL with strong effect on protein percentage at most likely position 49 cM. The QTL affects fat percentage to a lesser extent and has only very modest influence on the yield traits (data not shown).

Bootstrap analyses were performed for protein percentage according to VISSCHER *et al.* (1996) to estimate the 95% confidence interval (C.I.) for the position of the QTL. Figure 1 illustrates the distribution of the most likely position of the QTL across 1000 bootstrap samples as well as the deduced 95% C.I. It can be seen that the C.I. covers ~50 cM, which in essence corresponds to the distal half of chromosome 20 and therefore to a very poor location of the QTL.

We then performed within-family regression analyses on protein percentage as described (ARRANZ *et al.* 1998) to identify sire families that were segregating for this QTL. Two such families were identified in data set I: families 1 and 18 (data not shown).

Refining the map position of a QTL—use of a haplotype-based test for association: The previously described

within-family analyses indicate that sires 1 and 18 are heterozygous for QTL alleles with large substitution effects (Q) on chromosome 20. Previous work within the same population revealed extensive genome-wide linkage disequilibrium due to random drift (Farnir *et al.* 2000). We therefore hypothesized that the marker haplotypes flanking the Q alleles in the segregating sires might well be in linkage disequilibrium with the same Q alleles in the general population as well. To test this hypothesis, we measured the effect on protein percentage of the sire haplotypes in the general population using the haplotype-based test for association described in materials and methods.

By doing so, we identified five haplotype windows that were yielding significant F-ratios ($P < 0.01$ after correction for multiple testing), corresponding to substitution effects of ~0.03% milk protein. The corresponding haplotypes were all derived from a chromosome segment that was shared identical by state by sires 1 and 18. The sons of both sires were eliminated from the data set prior to performing the test for association. Figure 1 shows the position and statistical significance of the corresponding marker windows. It can be seen that their position centers around the TGLA153-GHRJ marker pair, corresponding to a minor peak for protein percentage, but the most likely QTL position when analyzing fat percentage. This result strongly suggests that a gene in the vicinity of these markers indeed contributes to the QTL effect observed on bovine chromosome 20.

Refining the map position of a QTL—combined linkage and LD analysis: To confirm the findings obtained with the haplotype-based test for association, we analyzed data set I using the LDVCM program for combined linkage and LD mapping. Figure 2 shows the location scores that were obtained with this approach for protein percentage. The profile obtained when considering only linkage information essentially parallels that obtained by multimarker regression (*cf.* Figure 1), although the lod scores are slightly less significant ($z_{max} = 1.8$; chromosome-wide P value = 0.016). When including linkage disequilibrium information, however, a very significant lod score of 8.5 corresponding to a chomosome-wise P value of 1.5E-8 was obtained at map position 43 cM, *i.e.*, very close to the chromosome region identified by the haplotype-based association test. Using the same approach, highly significant lod scores were obtained in the same chromosome region for fat percentage (position 43 cM, lod score 5.9, P value 7.5E-6), milk yield (position 43 cM, lod score 4.5, P value 0.00018), fat yield (position 46 cM, lod score 3.2, P value 0.0047), and protein yield (position 43 cM, lod score 5.2, P value 3.7E-5; data not shown). These results therefore supported the existence of a QTL influencing milk yield and composition in the vicinity of the GHR gene.

Scanning the bovine growth GHR gene for DNA sequence polymorphisms: Knowing the key role played by the growth hormone axis in the initiation and maintenance of lactation, the GHR obviously stood out as a strong positional candidate for the corresponding QTL effect. Assuming that the GHR gene would indeed account for at least part of the QTL effect, we predicted on the basis of the haplotype-based test for association that sires 1 and 18 would both be heterozygous for a mutation causing the GHR to be functionally different. We therefore decided to scan the coding portion of the GHR gene for DNA sequence polymorphisms in these animals. Intronic primers allowing for the convenient amplification and sequencing of exons 3–10 of the GHR were developed as described in materials and methods. Analysis of the sequence traces obtained from five Holstein-Friesian individuals including sires 1 and 18 revealed 10 SNPs in the GHR gene. Figure 3 reports the position and nature of the corresponding SNPs.

Four of these are SNPs located in introns [*Nt71-85(del1)*, *Nt71-12(T-C)*, *Nt864-33(T-G)*, and *Nt933+21(A-G)*], one is an SNP located in the 3′ UTR of the GHR gene [*Nt1922(C-T)*], and three are synonymous mutations in third codon positions [*Nt1095(C-T)*, *Nt1635(C-T)*, and *Nt1809(C-T)*]. None of these are *a priori* likely to affect the function of the GHR gene.

The two remaining SNPs, however, modify the amino acid sequence of the GHR receptor. A *T* to *A* substitution in exon VIII results in the nonconservative replacement of a neutral phenylalanine with an uncharged but polar tyrosine residue (*F279Y*). The corresponding phenylalanine residue is located within the transmembrane domain of the GHR and is conserved among all analyzed mammals (human, baboon, rabbit, mouse, rat, dog, pig, sheep, and opossum) except guinea pig, where it is nevertheless replaced by a neutral leucine residue. In chicken and pigeon, the corresponding residue is also a neutral isoleucine.

An *A* to *C* substitution in exon X results in the replacement of an asparagine with a threonine (*N528T*), both amino acids being polar uncharged residues. This residue is less conserved during evolution, being either an asparagine (human, rabbit, pig, and chicken) or a serine residue (ovine, mouse, and rat). It is, however, only one residue removed from one of the highly conserved, potentially phosphorylated intracellular tyrosine residues.

Sires 1 and 18, which were both heterozygous for the GHR-containing marker haplotype associated with a highly significant substitution effect on protein percentage in the association test, were heterozygous for SNPs *Nt71-85(del1)*, *Nt864-33(T-G)*, *Nt933+21(A-G)*, and most importantly *F279Y*. Given the effect of this SNP on the primary sequence of the GHR and therefore possibly on its function, *F279Y* stood out as a prime candidate for the mutation causing the observed QTL effect.

Inclusion of SNPs in the combined linkage and LD analysis dramatically increases the lod score at the GHR locus: We constructed an OLA as described (Karim *et*

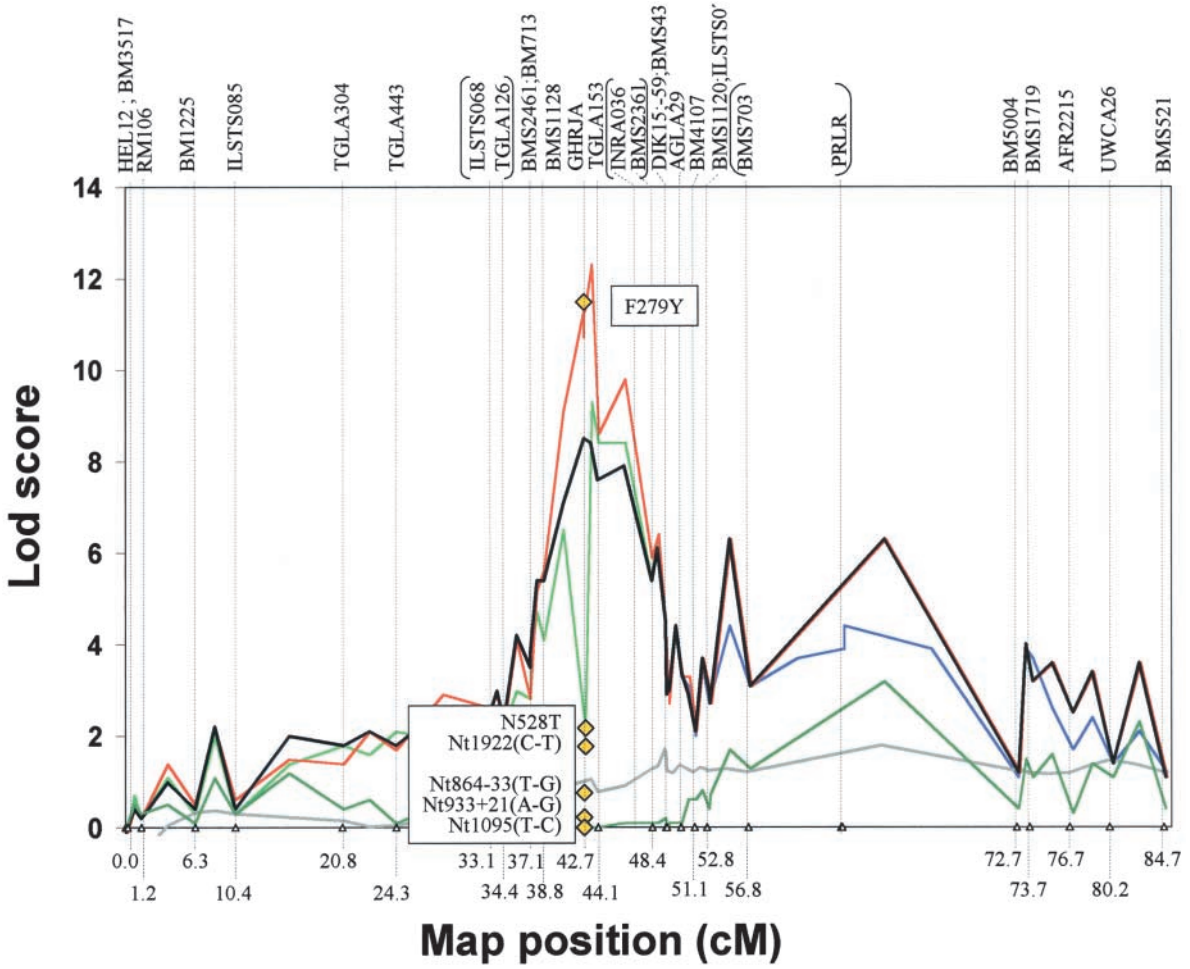

Figure 2.—Lod score profiles obtained for protein percentage along the chromosome 20 map using the LDVCM programs. The names of the markers comprising the map are given at the top and their respective positions in centimorgans (Kosambi) at the bottom. The gray curve is obtained by considering linkage information only, while all other curves are obtained by considering both linkage and LD. Black, basic chromosome 20 microsatellite marker map; red, chromosome 20 microsatellite marker map plus six GHR SNPs [*F279Y*, *Nt864-33(T-G)*, *Nt933+21(A-G)*, *Nt1095(T-C)*, *N528T*, and *Nt1922(C-T)*]; light green, chromosome 20 microsatellite marker map plus five GHR SNPs (*F279Y* dropped); Blue, chromosome 20 microsatellite marker map plus four PRLR SNPs; dark green, chromosome 20 microsatellite map plus *F279Y* treated as a fixed effect. The yellow diamonds correspond to the lod scores obtained by single-point analysis with the individual GHR SNPs. The names of the corresponding SNPs are given in the adjacent boxes.

al. 2000) for multiplex genotyping of the *F279Y*, *Nt864-33(T-G)*, *Nt933+21(A-G)*, *Nt1095(T-C)*, *N528T*, and *Nt1922 (C-T)* SNPs and applied it to data set I. The linkage phase was determined as described (Farnir *et al.* 2002). Figure 4 shows the frequency distribution of the GHR haplotypes as measured in the MCs (see above). It shows that at least 13 distinct haplotypes occur in the Dutch Holstein-Friesian population; however, 3 of these account for 85% of the chromosomes in this population.

We placed the GHR SNP haplotype by linkage analysis on the chromosome 20 marker map at position 42.7 cM, coinciding with the GHRJ microsatellite as expected.

A combined linkage and LD analysis was then performed using the LDVCM software, including the new GHR SNP genotypes. As shown in Figure 2 for protein percentage, inclusion of the GHR SNPs increased the maximum lod score by 3.8 units, yielding a maximum lod score of 12.3 at position 43 cM, *i.e.*, just distal of the GHR gene. Table 4 reports the corresponding variance component estimates.

Including the GHR SNPs in the LDVCM analysis had a comparable effect when analyzing fat percentage. The lod score increased from 5.9 to 7.8, maximizing exactly at the GHR gene (Table 4). The effect was more modest for milk yield and fat yield, increasing the lod scores by, respectively, 0.4 and 0.1 units but maximizing in both instances on the GHR gene (Table 4). For protein yield, however, inclusion of the GHR SNPs resulted in a marked decrease of the lod scores, dropping from 5.2 to 1.7 or less in the region of the GHR gene (Table 4).

For comparison, performing a combined linkage and LD analysis after inclusion of a haplotype composed of

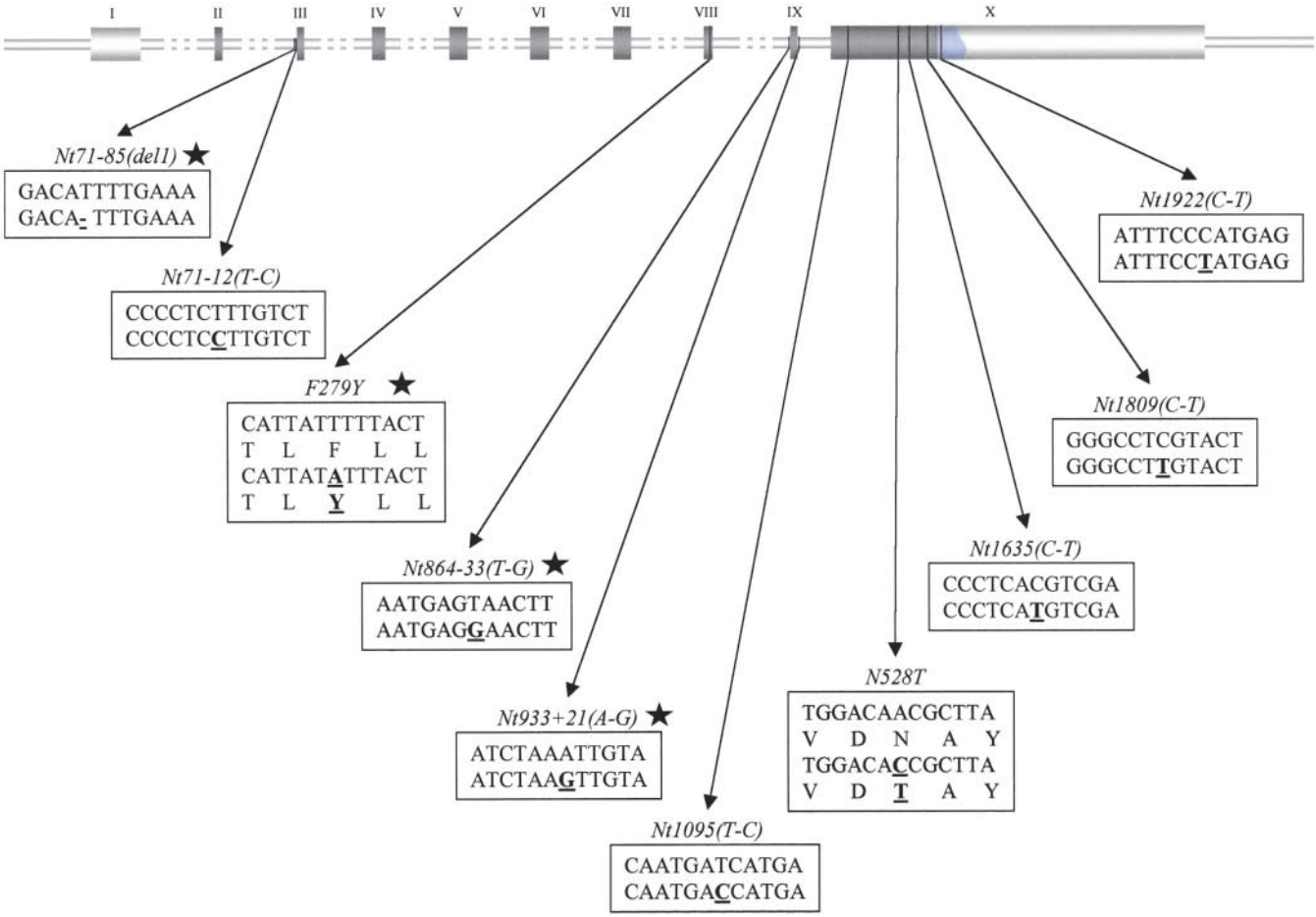

FIGURE 3.—Schematic representation of the bovine GHR gene. The 10 exons are shown as large cylinders and labeled by exon number. Coding sequences are shown in dark gray, sequenced 3′ and 5′ UTR sequences in gray, and unsequenced UTR regions in light gray. Introns are shown as interrupted thin cylinders. SNPs are marked as black lines connected with a box detailing the corresponding DNA sequences. The SNPs for which sires 1 and 18 were found to be hetereozygous are marked by stars.

four PRLR SNPs resulted in a local decrease in the lod score values for all traits (see Figure 2 for protein percentage).

These results strongly suggested that either the GHR itself or a gene in strong LD with GHR underlies the observed QTL effect.

Unique status of the *F279Y* polymorphism with regard to the chromosome 20 QTL effect: We then performed three tests to determine the relative contribution of the different SNPs to the increase in signal noted for protein percentage. First, we reran LDVCM analyses by sequentially dropping one of the six GHR SNPs comprising the GHR SNP haplotype. While dropping the *Nt864-33(T-G)*, *Nt933+21(A-G)*, *Nt1095(T-C)*, *N528T*, and *Nt1922 (C-T)* SNPs did not significantly alter the lod score profiles (data not shown), dropping the *F279Y* SNP virtually annihilated the entire gain obtained by considering the complete GHR SNP haplotype (Figure 2).

Second, we ran LDVCM analyses considering only the microstallite markers to compute pairwise IBD probabilities (ϕ_p) between SCs and MCs, but sequentially including the GHR and PRLR SNP genotypes as fixed effects in our mixed model. Inclusion of the *F279Y* SNP as a fixed effect completely erased the lod score peak in the GHR region (Figure 2). With none of the other SNPs, whether in the GHR or PRLR genes, did the lod scores drop below 5.7 (data not shown). Note that including the PRLR SNPs also did not markedly affect the minor lod scores peak around position 65 cM, *i.e.*, the approximate location of the PRL gene.

Finally, we used LDVCM to estimate the effects of the different GHR SNPs individually (*i.e.*, without considering flanking marker data): *F279Y* would yield a lod score of 11.5 while the other SNPs would respectively yield lod scores of only 0.75 [*Nt864-33(T-G)*], 0.22 [*Nt933+21 (A-G)*], 0 [*Nt1095(T-C)*], 2.18 (*N528T*), and 1.77 [*Nt1922 (C-T)*] (Figure 2).

Altogether, these results clearly point toward a unique status of the *F279Y* polymorphism with regard to the chromosome 20 QTL effect, indicating that this SNP is either directly responsible for the QTL effect or tightly associated with the actual causal mutation.

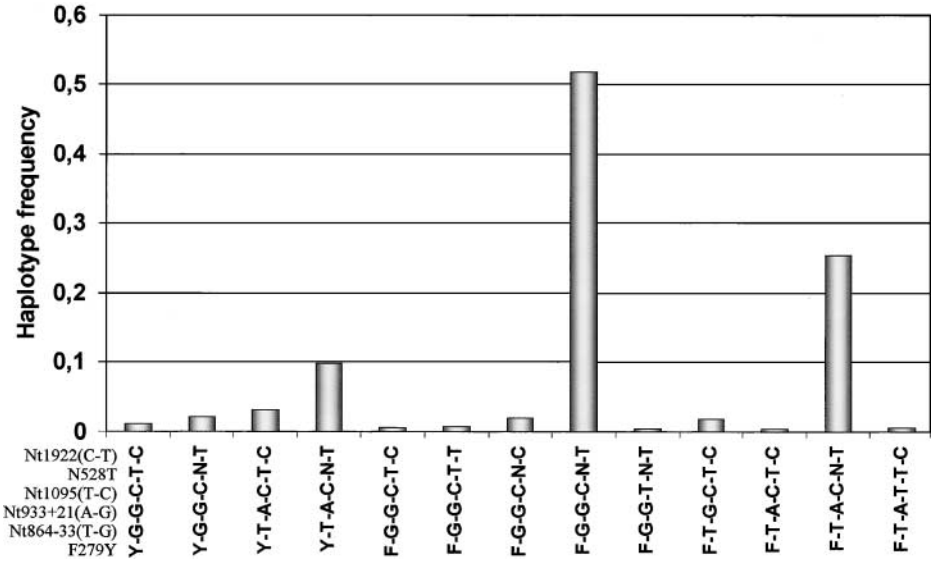

FIGURE 4.—Frequency distribution of the GHR SNP haplotypes in the Dutch Holstein-Friesian population.

Figure 5 shows the segregation of the *F* and *Y* alleles within the haplotype clusters maximizing the LDVCM lod scores when analyzing protein percentage including all six GHR SNPs. The REML solution is associated with a grouping in 22 haplotype clusters of which 17 are perfectly homogeneous with regard to the *F279Y* polymorphism.

Effect of the *F279Y* GHR polymorphism on milk yield and composition in the general dairy cattle population: To more accurately estimate the effect of the *F279* GHR polymorphism on milk yield and composition, we genotyped data sets II–VI—corresponding to an additional 2772 bulls and 872 cows—for this SNP. Effects of the *F279Y* genotype on DYDs and LVs for milk yield (kilograms), protein yield (kilograms), fat yield (kilograms), protein percentage, and fat percentage were estimated using a mixed model including a fixed genotype effect and a random animal model to account for the polygenic background.

It can be seen from Table 5 that the *F279Y* polymorphism behaved in a very similar fashion in all analyzed populations, whether Dutch or New Zealander, Holstein-Friesian or Jersey. As expected, the effect of the *F279Y* genotype was—in all five data sets—most pronounced on protein percentage, accounting for 5.3–9.4% of the trait variance. The effect of the *F279Y* mutation was also clearly detectable in all these populations on fat percentage and to a lesser extent on milk yield. It accounts for between 2.3 and 5.5% of the variance in fat percentage and between 0.7% and 2.9% of the variance in milk yield. For milk yield, inheriting one *Y* allele increases the DYD for milk yield by an estimated ±67 kg to ±112 kg and the LV for milk yield by ±87 kg to ±162 kg. Effects of the *F279Y* genotype on fat and protein yield were in essence nonsignificant although a tendency toward a decrease in fat yield of 1.5–2.5 kg for every dose of *Y* allele was notable.

The fact that the *F279Y* mutation showed very compa-

TABLE 4

Results of the LDVCM analysis after addition of the six GHR SNPs to the BTA20 microsatellite map

Trait	Map position	No. of clusters	r^2 QTL	Lod score	r^2 POLYG	r^2 RES
Milk yield (kg)	GHR	21	0.06	4.9	0.72	0.22
Fat yield (kg)	GHR	4	0.05	3.3	0.81	0.14
Protein yield (kg)	GHR	20	0.04	1.7	0.78	0.18
Fat %	GHR	5	0.09	7.8	0.88	0.02
Protein %	GHR-TGLA153	22	0.10	12.3	0.87	0.03

Map position, marker interval associated with the highest lod score for the considered trait; No. of clusters, the number of clusters in the haplotype dendrogram that yields the highest lod score; r^2 QTL, fraction of the trait variance due to the QTL, computed as $2\sigma^2_H/[2\sigma^2_H + \sigma^2_A + \sigma^2_E]$; r^2 POLYG, fraction of the trait variance due to the polygenic background, computed as $\sigma^2_A/[2\sigma^2_H + \sigma^2_A + \sigma^2_E]$; r^2 RES, fraction of the trait unexplained by the model, computed as $\sigma^2_E/[2\sigma^2_H + \sigma^2_A + \sigma^2_E]$.

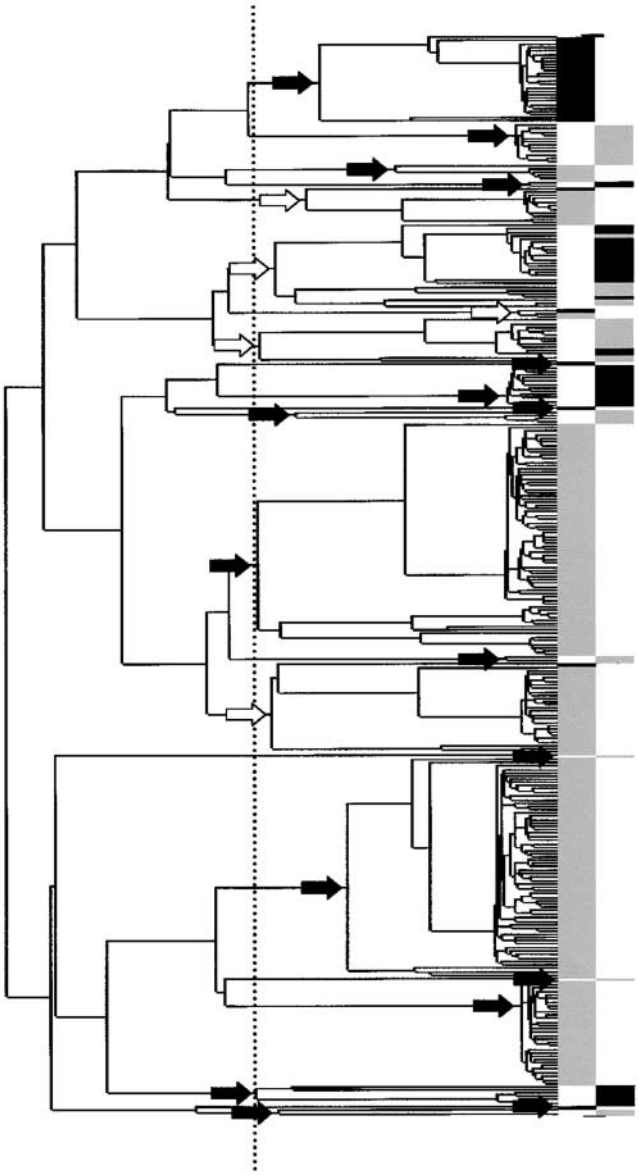

FIGURE 5.—UPGMA dendrogram representing the genetic relationship between the SC and MC haplotypes at position 43.4 cM (interval *GHR-TGLA53*), yielding the most likely REML solution when analyzing protein percentage. The dotted vertical line corresponds to the optimal threshold value, *T*. The arrows point toward the nodes in which all haplotypes belonging to a given haplotype cluster coalesce. The arrows are solid when the corresponding cluster contains either only the *F* or only the *Y* allele for the *F279Y* SNP. They are open when the cluster contains both alleles. The bars at the right of the tree indicate the *F279Y* genotype of the corresponding haplotype: black equals *Y* and gray equals *F*. The bar is broken to mark the separation between clusters.

rable effects in all five analyzed populations strongly supports their *bona fide* nature and the causality of either *F279Y* itself or a tightly linked mutation.

The effects on LVs were expected to be approximately twice as large as their equivalent on DYDs (as LVs and DYDs are by definition measured on a different scale; FALCONER and MACKAY 1996), but to explain a smaller proportion of the trait variability as the heritability of LVs is lower than that of DYDs. For reasons that are as yet not explained, this tendency was not systematically apparent from our data (Table 5). Note also that the absolute values of the effects on the percentage traits cannot be directly compared between data sets I and II (Netherlands) *vs.* data sets III–VI (New Zealand) as the percentage traits are computed from the yield traits using different formulas in both countries.

DISCUSSION

We herein illustrate how LD can be exploited to dissect and refine the map position of QTL in outbred dairy cattle populations. In this work, this led to the demonstration that a chromosomal region harboring the *GHR* gene contributes to the QTL effect on milk yield and composition that was previously identified on bovine chromosome 20 (*e.g.*, GEORGES *et al.* 1995; ARRANZ *et al.* 1998). No evidence was found, on the contrary, for an effect of the chromosomal region harboring the *PRLR* gene, an equally intriguing positional candidate in the initial stages of this analysis (ARRANZ *et al.* 1998).

Two distinct approaches aimed at capturing LD information were utilized in this study and led to similar conclusions: a haplotype-based association test and a variant component-mapping approach. The latter differs from the method that was recently described by MEUWISSEN and GODDARD (2000, 2001), by including a hierarchical haplotype-clustering step aimed at solving numerical issues and increasing the computation speed. This clustering step has the additional advantage that it allows for the identification of haplotypes that are likely to be functionally different, which could then become the focus for further molecular work. This proposition is presently being tested in the analysis of other QTL segregating in dairy cattle populations.

By sequencing a very small number of individuals that were carefully selected on the basis of their inferred QTL genotype, we subsequently identified a nonconservative substitution of a highly conserved *F* residue in the transmembrane domain of the *GHR*. We present a series of systematic tests that strongly suggest that the corresponding *F279Y* polymorphism may be the direct cause (quantitative trait nucleotide; MACKAY 2001) of the consistently associated effects on milk yield and composition. Preliminary evidence (R. SPELMAN, unpublished observations) suggests that the *F279Y* polymorphism affects live weight, which would also be compatible with a direct effect of the *GHR*.

However, these observations do not formally prove the causality of the *GHR* and *F279Y* mutation. The fact that the lod score is highest in the multipoint analysis and maximizes just distally from the *GHR* gene might indicate that the causal mutation is different, affecting either the *GHR* or another gene. Examination of the genes lo-

TABLE 5

Effect of the GHR *F279Y* mutation on milk yield and composition

Trait	*FY* − *FF* (±SE)	*YY* − *FF* (±SE)	r^2_{QTL}	P_{QTL}
A. Data sets I + II (Dutch Holstein-Friesian sires—DYDs) Genotype frequencies: *FF*, 0.67; *FY*, 0.31; *YY*, 0.02; *n* = 1263				
Milk yield (kg)	67 ± 16	128 ± 49	0.017	6.8E-05
Fat yield (kg)	−1.4 ± 0.6	−4.1 ± 1.8	0.007	0.015
Protein yield (kg)	−0.1 ± 0.4	0.2 ± 1.3	0.000	0.961
Fat %	−0.06 ± 0.01	−0.14 ± 0.03	0.035	2.2E-09
Protein %	−0.033 ± 0.01	−0.06 ± 0.01	0.057	7.3E-15
B. Data set III (New Zealand Holstein-Friesian sires—DYDs) Genotype frequencies: *FF*, 0.68; *FY*, 0.29; *YY*, 0.03; *n* = 1550				
Milk yield (kg)	89 ± 17	68 ± 45	0.010	8.4E-06
Fat yield (kg)	−1.5 ± 0.7	−5.2 ± 1.7	0.006	0.01
Protein yield (kg)	−1.0 ± 0.5	−4.4 ± 1.3	0.004	0.007
Fat %	−0.13 ± 0.02	−0.20 ± 0.02	0.032	1.2E-12
Protein %	−0.10 ± 0.01	−0.17 ± 0.04	0.094	8.0E-32
C. Data set IV (New Zealand Jersey sires—DYDs) Genotype frequencies: *FF*, 0.89; *FY*, 0.10; *YY*, 0.01; *n* = 959				
Milk yield (kg)	112 ± 33	441 ± 102	0.029	3.35E-05
Fat yield (kg)	0.1 ± 1.5	−2.3 ± 4.7	0.000	0.97
Protein yield (kg)	0.3 ± 1.1	6.3 ± 3.3	0.003	0.31
Fat %	−0.19 ± 0.05	−0.68 ± 0.02	0.030	5.62E-05
Protein %	−0.13 ± 0.02	−0.30 ± 0.08	0.059	1.1E-07
D. Data set V (New Zealand Holstein-Friesian cows—LVs) Genotype frequencies: *FF*, 0.73; *FY*, 0.24; *YY*, 0.03; *n* = 485				
Milk yield (kg)	87 ± 48	99 ± 129	0.007	0.17
Fat yield (kg)	−1.3 ± 2.2	−8.3 ± 5.9	0.004	0.34
Protein yield (kg)	−0.6 ± 1.5	−3.1 ± 4.1	0.001	0.73
Fat %	−0.14 ± 0.05	−0.31 ± 0.13	0.023	0.004
Protein %	−0.10 ± 0.02	−0.19 ± 0.06	0.053	2.3E-06
E. Data set VI (New Zealand Jersey cows—LVs) Genotype frequencies: *FF*, 0.81; *FY*, 0.17; *YY*, 0.02; *n* = 387				
Milk yield (kg)	162 ± 52	59 ± 135	0.025	0.009
Fat yield (kg)	−2.1 ± 2.7	−8.3 ± 6.9	0.005	0.40
Protein yield (kg)	1.2 ± 1.7	−1.4 ± 4.5	0.001	0.73
Fat %	−0.28 ± 0.06	−0.33 ± 0.17	0.055	6.7E-05
Protein %	−0.13 ± 0.03	−0.10 ± 0.07	0.058	1.6E-05

FY − *FF*, difference between the mean trait values of the *FY* and *FF* genotypes = effect of one *Y* dose; *YY* − *FF*, difference between the mean trait values of the *YY* and *FF* genotypes = effect of two *Y* doses; r^2_{QTL}, proportion of the trait variability explained by the GHR *F279Y* genotype, computed as described in MATERIALS AND METHODS; P_{QTL}, statistical significance of the GHR *F279Y* genotype effect.

cated in the vicinity of the *GHR* on the human genomic sequence (http://www.ensembl.org/homo_sapiens/), however, does not reveal any obvious alternative candidate. It is also noteworthy that administering growth hormone to lactating cows has a major effect on protein yield (BAUMAN *et al.* 1999), while this trait is in essence not affected by this QTL.

The effects of the *F279Y* genotype on the indexes that are used as the basis for selection in the Netherlands and New Zealand [net economic merit index (INET) and breeding worth (BW), respectively] are highly significant (R. SPELMAN, unpublished observations). As a matter of fact, a retrospective survey of the genotypes of the New Zealand sires clearly indicates that the frequency of the *F* allele has increased in recent years and that the *FF* genotype increases the likelihood for a sire to be selected for breeding (R. SPELMAN, unpublished observations). As a consequence, we anticipate that this marker has the potential to be very useful for marker-assisted selection and to more effectively increase the frequency of the favorable *F* allele.

Data sets V and VI (composed of cows) allowed for the analysis of potential dominance effects between the *F* and *Y* alleles. Modest evidence in favor of dominance

of the *Y* over the *F* allele was found for protein percentage ($P < 0.05$; data not shown). However, as the number of *YY* individuals was small, the power to detect significant dominance interactions was very limited. Preliminary analyses in these data sets also suggest that the *F279Y* mutation and the previously described *K232A* mutation in the bovine DGAT gene (Grisart *et al.* 2002) act in an additive manner.

We believe it unlikely that the *F279Y* or tightly associated polymorphism accounts for the entire chromosome 20 QTL effect. Indeed, examination of the location scores (*e.g.*, Figure 1), suggests that additional more distally located genes might contribute to the QTL effect on BTA20 as well. We have also identified two sires that would clearly be heterozygous for a QTL on BTA20 despite being homozygous for the *F279Y* polymorphism (data not shown). For one of these (sampled in the Netherlands), the QTL would primarily affect fat percentage and milk yield (rather than protein percentage) and the most likely QTL location would be at map position 48 cM. This could indicate either that the *F279Y* is not the causative mutation or that another closely linked QTL is segregating in this population as well. For the other sire (sampled in New Zealand), the QTL would primarily affect protein and fat percentage but the most likely QTL position would be at map position 85 cM. For the latter sire, the GHR gene falls outside of the 95% confidence interval for the QTL. Having identified a first polymorphism with major effect on milk yield and composition on BTA20 should facilitate the identification of these other putative QTL.

This work was funded by grants from Holland Genetics, Livestock Improvement Corporation, Via Lactia Biosciences, the Vlaamse Rundvee Vereniging, the Ministère des Classes Moyennes et de l'Agriculture (Belgium), and European Union grants B104-CT95-0073 and PL970471. Sarah Blott was supported by a Wellcome International research fellowship.

LITERATURE CITED

Andersson, L., 2001 Genetic dissection of phenotypic diversity in farm animals. Nat. Rev. Genet. **2:** 130–138.

Arranz, J.-J., W. Coppieters, P. Berzi, N. Cambisano, B. Grisart *et al.*, 1998 A QTL affecting milk yield and composition maps to bovine chromosome 20: a confirmation. Anim. Genet. **29:** 107–115.

Bauman, D. E., R. W. Everett, W. H. Weiland and R. J. Collier, 1999 Production responses to bovine somatotropin in Northeast dairy herds. J. Dairy Sci. **82:** 2564–2573.

Boldman, K. G., 1993 *A Manual for Use of MTDFREML: A Set of Programs to Obtain Estimates of Variances and Covariances.* U.S. Department of Agriculture, Agriculture Research Service.

Charlier, C., F. Farnir, P. Berzi, P. Vanmanshoven, B. Brouwers *et al.*, 1996 IBD mapping of recessive traits in livestock: application to map the bovine syndactyly locus to chromosome 15. Genome Res. **6:** 580–589.

Churchill, G. A., and R. W. Doerge, 1995 Empirical threshold values for quantitative trait mapping. Genetics **138:** 963–971.

Coppieters, W., J. Riquet, J.-J. Arranz, P. Berzi, N. Cambisano *et al.*, 1998a A QTL with major effect on milk yield and composition maps to bovine chromosome 14. Mamm. Genome **9:** 540–544.

Coppieters, W., A. Kvasz, J.-J. Arranz, B. Grisart, F. Farnir *et al.*, 1998b A rank-based nonparametric method to map QTL in outbred half-sib pedigrees: application to milk production in a granddaughter design. Genetics **149:** 1547–1555.

Darvasi, A., 1998 Experimental strategies for the genetic dissection of complex traits in animal models. Nat. Genet. **18:** 19–24.

Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics,* Ed. 4. Longman, New York.

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

Farnir, F., B. Grisart, W. Coppieters, J. Riquet, P. Berzi *et al.*, 2002 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. Genetics **161:** 275–287.

Flint, J., and R. Mott, 2001 Finding the molecular basis of quantitative traits: successes and pitfalls. Nat. Rev. Genet. **2:** 437–445.

Georges, M., D. Nielsen, M. Mackinnon, A. Mishra, R. Okimoto *et al.*, 1995 Mapping quantitative trait loci controlling milk production by exploiting progeny testing. Genetics **139:** 907–920.

Godowski, P. J., D. W. Leung, L. R. Meacham, J. P. Galgani, R. Hellmiss *et al.*, 1989 Characterization of the human growth hormone receptor gene and demonstration of a partial gene deletion in two patients with Laron-type dwarfism. Proc. Natl. Acad. Sci. USA **86:** 8083–8087.

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford *et al.*, 2002 Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT gene with major effect on milk yield and composition. Genome Res. **12:** 222–231.

Hauser, S. D., M. F. McGrath, R. J. Collier and R. J. Krivi, 1990 Cloning and in vivo expression of bovine growth hormone receptor mRNA. Mol. Cell. Endocrinol. **72:** 187–200.

Heap, D., M. C. Lucy, R. J. Collier, C. K. Boyd and W. C. Warren, 1995 Nucleotide sequence of the promoter and first exon of the somatotropin receptor gene in cattle. J. Anim. Sci. **73:** 1529.

Hudson, R. R., 1985 The sampling distribution of linkage disequilibrium under an infinite alleles model without selection. Genetics **109:** 611–631.

Johnson, D. L., and R. Thompson, 1995 Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. J. Dairy Sci. **78:** 449–456.

Kappes, S. M., J. W. Keele, R. T. Stone, R. A. McGraw, T. S. Sonstegard *et al.*, 1997 A second-generation linkage map of the bovine genome. Genome Res. **7:** 235–249.

Karim, L., W. Coppieters, L. Grobet, A. Valentini and M. Georges, 2000 Convenient genotyping of six myostatin mutations causing double-muscling in cattle using a multiplex oligonucleotide ligation assay. Anim. Genet. **31:** 396–399.

Kim, J. J., and M. Georges, 2002 Evaluation of a new fine-mapping method exploiting linkage disequilibrium: a case study analysing a QTL with major effect on milk composition on bovine chromosome 14. Asian-Aust. J. Anim. Sci. **15:** 1250–1256.

Knott, S., J. M. Elsen and C. Haley, 1996 Methods for multiple marker mapping of quantitative trait loci in half-sib populations. Theor. Appl. Genet. **93:** 71–80.

Lynch, M., and B. Walsh, 1997 *Genetics and Analysis of Quantitative Traits.* Sinauer Associates, Sunderland, MA.

Mackay, T. F. C., 2001 Quantitative trait loci in Drosophila. Nat. Rev. Genet. **2:** 11–20.

Mauricio, R., 2001 Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. Nat. Rev. Genet. **2:** 370–381.

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

Mount, D. W., 2001 *Bioinformatics: Sequence and Genome Analysis.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Nickerson, D. A., V. O. Tobe and S. A. Taylor, 1997 PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res. **25:** 2745–2751.

Reich, D. E., M. Cargill, S. Bolk, J. Ireland, P. C. Sabeti *et al.*,

2001 Linkage disequilibrium in the human genome. Nature **411:** 199–204.

Spelman, R. L., W. Coppieters, L. Karim, J. A. M. Van Arendonk and H. Bovenhuis, 1996 Quantitative trait loci analysis for five milk production traits on chromosome six in the dutch Holstein-Friesian population. Genetics **144:** 1799–1808.

Spielman, R. S., R. E. McGinnis and W. J. Ewens, 1993 Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am. J. Hum. Genet. **52:** 506–516.

Van Raden, P. M., and G. R. Wiggans, 1991 Derivation calculation and use of National Animal Model Information. J. Dairy Sci. **74:** 2737–2746.

Visscher, P. M., R. Thompson and C. S. Haley, 1996 Confidence intervals in QTL mapping by bootstrapping. Genetics **143:** 1013–1020.

Warren, W., T. P. Smith, C. E. Rexroad, III, S. C. Fahrenkrug, T. Allison *et al.*, 2000 Construction and characterization of a new bovine bacterial artificial chromosome library with 10 genome-equivalent coverage. Mamm. Genome **11:** 662–663.

Communicating editor: C. Haley