Mapping Quantitative Trait Loci Controlling Milk Production in Dairy Cattle by Exploiting Progeny Testing

Michel Georges,* Dahlia Nielsen,* Margaret Mackinnon,* Anuradha Mishra,* Ron Okimoto,*
Alan T. Pasquino,† Leslie S. Sargeant,* Anita Sorensen,* Michael R. Steele,*
Xuyun Zhao,* James E. Womack‡ and Ina Hoeschele†

*Genmark Inc., Salt Lake City, Utah 84108, †Department of Dairy Science, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0315, and ‡Department of Pathobiology, College of Veterinary Medicine, Texas A & M University, College Station, Texas 77843

ABSTRACT

We have exploited "progeny testing" to map quantitative trait loci (QTL) underlying the genetic variation of milk production in a selected dairy cattle population. A total of 1,518 sires, with progeny tests based on the milking performances of >150,000 daughters jointly, was genotyped for 159 autosomal microsatellites bracketing 1645 centimorgan or approximately two thirds of the bovine genome. Using a maximum likelihood multilocus linkage analysis accounting for variance heterogeneity of the phenotypes, we identified five chromosomes giving very strong evidence (LOD score ≥ 3) for the presence of a QTL controlling milk production: chromosomes 1, 6, 9, 10 and 20. These findings demonstrate that loci with considerable effects on milk production are still segregating in highly selected populations and pave the way toward marker-assisted selection in dairy cattle breeding.

IN dairy cattle, milk yield and composition are typical polygenic traits. Phenotypes are continuously distributed and reflect the joint action of large numbers of polygenes or quantitative trait loci (QTL) confounded with environmental effects. In the populations of interest, milk production traits have *narrow sense heritabilities* in the 25–50% range (PEARSON et al. 1990).

Despite early efforts to map QTL for milk production using small numbers of genetic markers (e.g., GELDERMANN et al. 1985; COWAN et al. 1990; HOESCHELE and MEINERT 1990; BOVENHUIS 1992; ANDERSSON-EKLUND and RENDEL 1993; SCHUTZ et al. 1993), the nature of the genes underlying the genetic variance of milk production remains essentially unknown. Since the discovery of microsatellite markers that can be typed using the polymerase chain reaction (WEBER and MAY 1989), a systematic genetic dissection of milk production and other production traits in livestock has become feasible. Characterization of these QTL may lead to more efficient breeding programs using marker-assisted selection (SOLLER and BECKMAN 1982) and may contribute to a better understanding of lactational physiology.

Most successful QTL mapping efforts described to date have exploited F2 or back-crosses obtained from parental populations divergent for the traits of interest (e.g., PATERSON et al. 1989; HILBERT et al. 1991). Although a similar approach might help to understand

the genetic differences between high and low producing breeds, our objective was to map QTL segregating *within* elite dairy cattle populations, as these are the molecular substrate of ongoing selection programs. However, because these populations have been intensely selected for milk production, it is generally assumed that polygenes with large effects are near or at fixation, whereas those still segregating are believed to have minor effects. As the individual contribution of such QTL to the overall phenotypic variance would be modest, their mapping is considerably complicated.

Recently, however, a number of strategies have been proposed to increase the power of QTL mapping. These strategies include selective genotyping (LANDER and BOTSTEIN 1989), progeny testing (LANDER and BOTSTEIN 1989), interval mapping (LANDER and BOTSTEIN 1989), the simultaneous search for multiple QTL (LANDER and BOTSTEIN 1989), the use of DNA pools (ARNHEIM et al. 1985) and the study of disease-tagged QTL (GEORGES et al. 1993). In this work, we illustrate the use of "progeny testing" in combination with interval mapping to map QTL controlling milk production in an elite Holstein dairy cattle population selected intensely for increased milk production for several generations.

MATERIALS AND METHODS

Exploiting progeny testing: the "granddaughter design": During the last 20 years, annual milk production per cow in the United States has increased from ~4,500 to 6,800 kg. This remarkable progress, which in recent years is mainly genetic in nature (PEARSON et al. 1990), is due to the extensive use

Corresponding author: Michel Georges, Department of Genetics, Faculty of Veterinary Medicine, B43, University of Liège, 20, Boulevard de Colonster, 4000 Liège (Sart Tilman), Belgium.
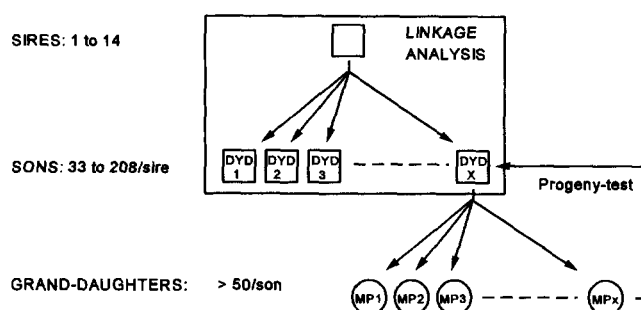E-mail: michel@stat.fmv.ulg.ac.be

FIGURE 1.—Schematic representation of the "granddaughter design" (GDD) (WELLER *et al.* 1991). Linkage analysis is performed within half-sib families each consisting of a founder sire and a large number of its half-sib sons. In this study, 14 such pedigrees were collected with a range of 33–208 sons. The trait values analyzed were the DYDx of the sons computed from the milking performances (MPx) of their respective daughters as part of the progeny-test procedure.

of artificial insemination (AI) and the resulting impact of superior sires on the genetic merit of the herd. Widespread use of a given sire, however, is only justified when its *breeding value* is estimated with sufficient reliability. This requires progeny testing: young sires, resulting from planned matings of sires and dams with highest breeding values, are tested based on the milking performances of 50–100 of their daughters. In the United States, the milking records of the daughters are collected as part of a nationwide record-keeping system known as the National Cooperative Dairy Herd Improvement Program, or NCDHIP, monitoring ~45% of the dairy herd or 4.5 million cows. During a monthly herd visit, Dairy Herd Improvement Association supervisors collect cow and herd data as well as milk samples that are forwarded to testing laboratories and Dairy Record Processing Centers. The ensuing standardized lactation records are used by the US Department of Agriculture to compute genetic evaluations (breeding value estimates) of bulls and cows using a statistical procedure referred to as Best Linear Unbiased Prediction with an "individual animal model" (VANRADEN and WIGGANS 1991). Similar data collection and genetic evaluation systems are in place in several other countries as well.

Because of the widespread implementation of this procedure, it is relatively straightforward to identify pedigrees characterized by the structure illustrated in Figure 1: large sets of progeny-tested paternal half-brothers. The experimental design, referred to as the "granddaughter design" (GDD) (WELLER *et al.* 1990), takes advantage of such pedigree material to map QTL underlying milk production traits. Marker genotyping and linkage analysis are performed in the sons, using averages of their respective daughter phenotypes as quantitative measurement. For traits with 25% heritability, this approach requires ~3.5 times less genotyping than the alternative "daughter" design, in which no advantage is taken from progeny testing (APPENDIX A).

For this study, we identified 14 such half-sib pedigrees, with between 33 and 208 sons per founder sire (mean 108) for a total of 1518 sons. None of the dams were available for analysis.

As many of the AI companies discard semen from sires culled after progeny testing, our samples are generally characterized by selection bias. Figure 2 compares for one of our half-sib families the distribution of protein yield for all sons progeny tested *vs.* the sample of sons available for analysis. We have previously studied the effect of such selection bias on the linkage analysis for quantitative traits (MACKINNON

and GEORGES 1992). It was shown that such bias substantially reduced the power to detect QTL because it decreased the apparent magnitude of the *average effect of the gene substitution* (FALCONER 1989) in the studied sample. Aware of this problem, many AI companies are now retaining semen samples of all progeny-tested sires for analysis.

Microsatellite genotyping: Two to four different microsatellite systems were amplified simultaneously in 10-μl reaction volumes, from 30 ng of each template DNA. Reagent concentrations were 75 mM KCl, 15 mM Tris-HCl (pH 8.4), 2.25 mM MgCl$_2$, 0.02% gelatin, 0.3 mM of each dNTP, 1 μM of each primer, 0.05 U AmpliTaq/μl and 0.1 μCi α-^{32}dCTP/μl. PCR reactions were set up with a Biomek 1000 robotic station (Beckman Instruments, Palo Alto, CA) and carried out in Techne MW2 devices (Techne, Cambridge, UK). Samples were denatured at 95° for 5 min and cycled 30 times under the following conditions: 93° for 1 min, 60° for 1 min and 72° for 1 min. After addition of 1 volume of formamide dye and denaturation at 95° for 5 min, 2 μl of each product was electrophoresed on a 7% acrylamide gel containing 32% formamide, 5.6 M urea, 135 mM Tris, 45 mM boric acid and 2.5 mM EDTA. The gels were autoradiographed for 2 hr to overnight.

The genotypes were interpreted by visual examination of the autoradiograms. For convenience, all systems were encoded as three allele systems: alleles *1* and *2* corresponding to the two alleles of the founder sire, whereas all other alleles encountered were pooled in allele group *3*. A first examiner called the genotypes and entered them twice in a database using custom-made data-management software. The two entries were automatically compared and discrepancies brought to the attention of the user. Interpretation and entry of these genotypes were then double-checked by a second examiner.

Map construction: All linkage analyses were performed with the ANIMAP programs (D. NIELSEN and M. GEORGES, unpublished data). These programs were designed to perform linkage studies in half-sib pedigrees. They can be used (1) to generate LOD score tables between pairs of markers with codominant alleles, (2) to perform multipoint linkage analysis with up to 16 markers (maximum likelihood recombination rates between adjacent markers are determined for all or a subset of marker orders) and (3) to generate LOD scores between a QTL whose position can be varied with respect to a set of up to 15 markers whose relative positions are held fixed (see QTL mapping).

For *x* informative markers in a given order, the likelihood of the corresponding pedigree was calculated as follows:

$$\sum_{i=1}^{2^x/2} P_i \times \prod_{j=1}^{n} \left[\sum_{k=1}^{2^x} \left[P(k|i) \times \prod_{m=1}^{x} AFM_m \right] \right],$$

where $\sum_{i=1}^{2^x/2}$ is summation over all possible sire linkage phases *i*, $\prod_{j=1}^{n}$ is product over all sons *j*, $\sum_{k=1}^{2^x}$ is summation over all paternal gametes *k* compatible with Mendelian laws, $\prod_{m=1}^{x}$ is product over all markers *m* within the synteny group, P_i is probability of phase *i* (the markers were assumed to be in linkage equilibrium and consequently all phases were considered equally likely), $P(k|i)$ is probability of gamete *k* given Mendelian laws, phase *i* and recombination rates between adjacent loci, θ_1 to θ_x, and AFM_m is allelic frequency of the obliged maternal allele of marker *m*, given the paternal gamete *k*.

Marker allele frequencies, required for the likelihood computation, were determined from the dam population separately for each pedigree as follows: $p1 = (1 - p3) \, n_{11}/(n_{11} + n_{22})$; $p2 = (1 - p3) \, n_{22}/(n_{11} + n_{22})$; $p3 = (n_{13} + n_{23}/n,$ with n_{xy} being the number of sons in the pedigree with genotype "*xy*," *n* the total number of sons in the pedigree and *pi* the frequency of marker allele *i*.
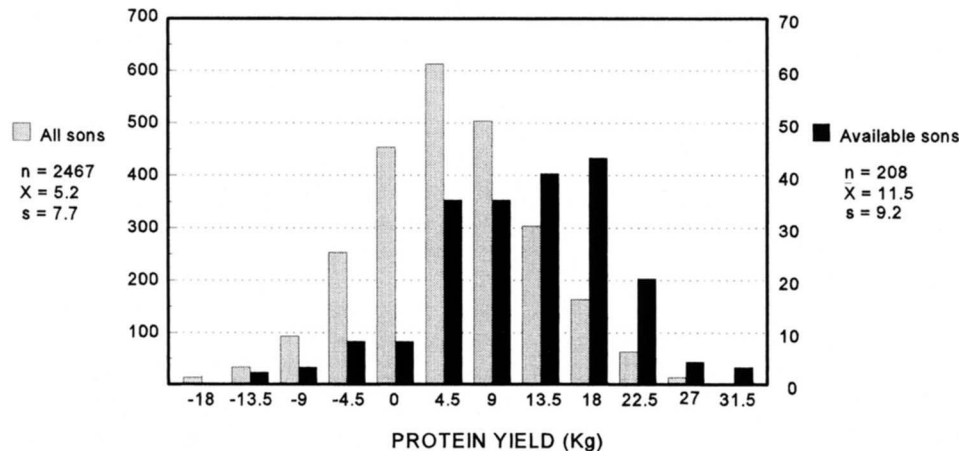
FIGURE 2.—Illustration of the selection bias characterizing the GDD in this study. The distribution of DYD for protein yield for all sons of founder sire 10 is compared with that for the sons available for analysis, in terms of number of sons (n), mean ($\times$) and standard deviation (s) in each group.

For pairwise linkage analyses, likelihoods of the pedigree data were computed for a range of fixed recombination rates. LOD scores were computed as $\log_{10}$ (likelihood of pedigree data for $\theta \neq 0.5$ / likelihood of pedigree data for $\theta = 0.5$). Pairwise linkage analysis was performed between all pairs of markers. For a given pair of markers, the LOD score tables were compared across pedigrees to check for heterogeneity. In cases of extreme heterogeneity, the genotypes were double checked and, in all cases, revealed artefacts of genotype collection. Marker pairs yielding LOD scores ≥ 3 were pooled into linkage groups.

The validity of these linkage groups was tested by analysis of the segregation patterns of the respective markers in a panel of somatic cell hybrids (DIETZ *et al.* 1992). Marker order within *bona fide* linkage groups was determined by multilocus linkage analysis. Maximum likelihood estimates of recombination rates between adjacent markers were computed for all possible orders using the GEMINI optimization routine (LALOUEL 1983). The most likely orders with corresponding estimated recombination rates are reported in Table 1.

QTL mapping: Five milk production traits were analyzed in this study: milk yield, fat yield, protein yield, fat percentage and protein percentage. Yield traits are the main components of the selection indices used in dairy cattle breeding programs. All five traits are characterized by quasinormal distributions in dairy populations. The different milk production traits are correlated to various degrees. Genetic correlations among yield traits are close to 0.8, whereas that among percentage traits is close to 0.5. Percentage traits show genetic correlations of approximately -0.3 with milk yield and of 0.2 with the corresponding yield trait (fat % and fat yield, and protein % and protein yield) (PEARSON *et al.* 1990).

The quantitative measurements used in the linkage analysis were sires' Daughter Yield Deviations (DYDs): unregressed weighted averages of their daughter's lactation performances (expressed as deviation from the population mean) (VANRADEN and WIGGANS 1991). Before averaging, the lactation yields are adjusted for systematic environmental effects and breeding values of the daughters' dams. The DYDs were obtained from the sire summary data base of January 1993 of the US Department of Agriculture.

The analyses were performed within (*vs.* across) half-sib pedigrees, *i.e.*, each of the 14 families was analyzed independently. Within a pedigree, the different linkage groups were

sequentially tested for the presence of a linked QTL affecting the traits studied. A separate analysis was performed for each trait. For each analysis, we postulated the presence of a single QTL within the studied linkage group, for which the founder sire was heterozygous "$+/-$". As is customary when generating location scores or performing interval mapping, the position of the postulated QTL was changed with respect to the markers composing the linkage group held in fixed positions according to Table 1. Only chromosomal segments bracketed by informative markers in the corresponding pedigree were scanned for the presence of QTL.

For each position of the hypothetical QTL and given x informative markers, the likelihood of the corresponding pedigree was calculated as follows:

$$\sum_{i=1}^{2^{x+1}/2} P_i \times \prod_{j=1}^{n} \left[\sum_{k=1}^{2^{x+1}} \left[P(k|i) \times \prod_{m=1}^{x} AFM_m \times P(DYD_j|k) \right] \right],$$

where $\sum_{i=1}^{2^{x+1}/2}$ is summation over all possible sire linkage phases i, $\prod_{j=1}^{n}$ is product over all sons j, $\sum_{k=1}^{2^{x+1}}$ is summation over all paternal gametes k compatible with Mendelian laws, $\sum_{m=1}^{x}$ is product over all markers m within the synteny group, P_i is probability of phase i (the markers were assumed to be in linkage equilibrium and consequently all phases were considered equally likely), $P(k|i)$ is probability of gamete k given Mendelian laws, phase i, and recombination of rates between adjacent loci, θ_1 to θ_x, AFM_m is allelic frequency of the obliged maternal allele of marker m, given the paternal gamete k and $P(DYD_j|k)$ is probability density of the DYD value of son j given the QTL allele ("$+$" or "$-$") of gamete k.

This probability density was obtained assuming a normal distribution of the son's DYDs:

$$P(DYD_j|k_{(+or-)}) = \frac{1}{\sqrt{2\pi}\,\sigma_j} e^{-(1/2)[(DYD_j-\mu_j)/\sigma_j]^2},$$

with mean $\mu_j = (0.25BV_{sire(j)} + 0.25BV_{dam(j)} \pm 0.25\alpha)$

$$= (0.5PTA_{sire(j)} + 0.5PTA_{dam(j)} \pm 0.25\alpha),$$

where PTA denotes predicted transmitting ability (VANRADEN and WIGGANS 1991) and is the best linear unbiased prediction of half of an animal's breeding value or of its transmitting ability (TA). PTAs were also obtained from the sire summary database of January 1993 of the US Department of Agricul-

M. Georges *et al.*

TABLE 1

Bovine autosomal microsatellite map

Chromosome	Linkage group
U01-16	[MGTG1-(2.0)-TGLA245-(31.1)-TGLA53-(38.9)-TGLA334]
U02-09	ETH225-(25.0)-AGLA300-(16.5)-TGLA261-(6.9)-TGLA427-(13.2)-TGLA73
U03-05	AGLA22-(16.3)-ETH152 [IGF1-(4.6)-AGLA254-(21.4)-TGLA124-(4.9)-AGLA293] AGLA248
U04-21	TGLA122-(5.6)-TGLA337-(21.5)-ETH131-(18.9)-[HEL5-(6.7)-AGLA233]
U05-10	TGLA111-(19.8)-TGLA131-(17.9)-[AGLA8-(1.3)-TGLA4-(0.0)-TGLA378]-(19.7)-[TGLA444-(0.8)-TGLA102]-(12.6)-TGLA433-(24.8)-TGLA272
U06-03	TGLA263-(23.4)-AGLA247-(13.5)-TGLA76-(30.5)-TGLA127
U07-25	TGLA414-(13.7)-{TGLA86}
U08-	ETH153
U09-18	TGLA227-(36.1)-ILSTOO2 TGLA357
U10-01	[AGLA17-(1.1)-TGLA49]-(42.5)-TGLA57-(3.4)-TGLA135-(12.1)-TGLA213-(14.2)-TGLA415-(24.7)-TGLA130-(21.0)-MAF46
U11-13	TGLA23-(11.1)-TGLA6 [AGLA232-(40.4)-AGLA285-(21.1)-TGLA381] TGLA342
U12-	AGLA13
U13-04	[TGLA116-(0.0)-MAF50]-(31.8)-TGLA420-(3.5)-TGLA215-(9.0)-{TGLA159}-(4.7)-TGLA60-(2.1)-AGLA227B-(8.0)-MGTG4B
U14-27	TGLA254
U15-06	C3H3-(20.0)-TGLA37
U16-11	TGLA436-(7.3)-TGLA77-(18.1)-TGLA340-(7.6)-TGLA58-(13.8)-TGLA327 HEL13
U17-02	[TGLA110-(7.8)-TGLA226]-(38.8)-ETH121-(11.1)-TGLA377-(4.0)-TGLA61-(10.0)-TGLA431-(7.0)-{TGLA44}
U18-08	TGLA341-(22.8)-{TGLA25}-(1.3)-HEL9-(2.0)-TGLA339-(5.8)-TGLA80-(14.4)-TGLA13-(9.3)-TGLA27-(2.9)-AGLA234-(6.5)-{TGLA10}
U19-15	AGLA259-(13.3)-MGTG13B-(22.9)-MAF65-(15.4)-HEL1-(24.7)-FSHB-(17.3)-TGLA75
U20-23	[AGLA212-(0.0)-TGLA142-(10.6)-MGTG7-(0.0)-{TGLA387}-(3.0)-CYP21-(1.9)-AGLA291]-(34.8)-TGLA147
U20-20	[AGLA29-(0.0)-TGLA214]-(5.3)-{TGLA153}-(13.0)-TGLA126-(11.6)-[{TGLA304}-(0.0)-TGLA443-(0.0)-TGLA172-(2.7)-AGLA267]
U21-19	HEL10-(7.0)-TGLA94-(12.7)-TGLA51
U22-07	ILSTOO1-(2.6)-TGLA176-(14.0)-[TGLA303-(2.5)-TGLA48]-(19.2)-TGLA164-(37.3)-[RASA-(2.7)-AGLA260]
U23-17	[TGLA26-(0.0)-AGLA299-(0.0)-TGLA188]-(23.7)-[TGLA231-(4.0)-ETH185]-(22.4)-TGLA170-(17.7)-TGLA322
U24-14	TGLA179 TGLA307
U25-	
U26-26	TGLA22-(23.2)-HEL11-(19.9)-TGLA134-(10.8)-TGLA429
U27-12	[TGLA9-(6.7)-TGLA36]-(30.5)-AGLA226-(7.3)-TGLA28-(14.4)-TGLA345-(37.5)-TGLA441
U28-24	[TGLA99-(1.7)-TGLA351]-(11.7)-AGLA269-(27.2)-TGLA435
U29-28	RBP3-(34.0)-[TGLA306-(0.0)-MGTG3]-(16.3)-[ETH1112-(5.4)-TGLA82]
??	AGLA206, AGLA209, AGLA217, MAF92, TGLA11, TGLA210, TGLA260, TGLA40, TGLA423, TGLA70

All markers have been previously described (FRIES *et al.* 1990; DIETZ 1992; GEORGES and MASSEY 1992; KAUKINEN and VARVIO 1992). Markers are sorted by autosomal synteny group: U1 to U29 (corresponding chromosomes are given following the hyphen when known). Adjacent markers in a linkage group are connected with hyphens and the estimated male recombination rate is given in parentheses. Marker sets for which the odds against alternative orders are ≤100:1 are bordered by square brackets. Markers that could not be mapped using the somatic cell hybrid panel are flanked by braces. Markers that could be mapped neither by linkage analysis nor in the somatic cell hybrid panel are reported in the last line. Note that the presence of a singleton marker does not necessarily mean that this marker is not linked to any of the other markers reported on that chromosome, but only that linkage could not been demonstrated with the available data.

ture. The term α represents the difference in average effects of the QTL alleles "+" and "−" or the average effect of the QTL gene substitution as defined by FALCONER (1989).

and variance $\sigma_j^2 = [0.25 / \text{Rel}_{\text{son}(j)}$

$$- 0.0625 \text{Rel}_{\text{sire}(j)} - 0.0625 \text{Rel}_{\text{dam}(j)}]\sigma_A^2,$$

where $\text{Rel}_{\text{son}(j)}$ (contributed by the progeny test only), $\text{Rel}_{\text{sire}(j)}$ and $\text{Rel}_{\text{dam}(j)}$ are the reliabilities of the breeding values estimated for son, sire and dam (FALCONER 1989). The reliability is the squared correlation between true and estimated

breeding value and reflects the amount of information available to estimate the breeding value of a given individual. These reliabilities were computed from $\text{Rel}_{\text{PTAsire}(j)}$, $\text{Rel}_{\text{PA}(j)}$ (reliability of average parent PTA) and $\text{Rel}_{\text{PTAson}(j)}$, obtained from the sire summary database of January 1993 of the US Department of Agriculture (see APPENDIX B).

Strictly speaking, σ_j^2, as defined here, is an approximation because it is further reduced by an amount $0.0625\alpha^2$, because the QTL segregation effect is included in the mean j. However, this is compensated for by a slight underestimation of σ_A^2, when estimated simultaneously with α.

The values of α and σ_A^2 maximizing the likelihood of the pedigree were determined using GEMINI (LALOUEL 1983). The resulting liklihood was divided by the likelihood of the pedigree maximized with respect to σ_A^2 but with the value of α fixed at 0, *i.e.*, assuming that there is no segregating QTL at the corresponding map position. Evidence for a QTL at the corresponding map position was expressed as a LOD score, *i.e.*, the $\log_{10}$ of the likelihood ratio.

The likelihood computed as described is a function of the absolute value of α and not of its sign. To compare the effects of a given QTL on the different production traits, the effects were given a sign as follows. Given the most likely linkage phase of the markers (paternal marker haplotypes M1 and M2) and given QTL heterozygosity of the sire $(+/-)$, the sire's most likely genotype can be either $(M1,+)/(M2,-)$ or $(M1,-)/(M2,+)$. When considering the effect of the QTL on two traits, obtaining the same most likely genotype $(M1,+)/(M2,-)$ or $(M1,-)/(M2,+)$ for both traits implies that the same QTL allele has a favorable effect on both traits; obtaining different most likely genotypes $(M1,+)/(M2,-)$ for one trait, $(M1,-)/(M2,+)$ for the other indicates that the allele with favorable effect on one trait has an unfavorable effect on the other trait. The substitution effect was considered positive for the trait yielding the highest LOD score (*e.g.*, A). For the other traits, the effect was considered positive if the same genotype was the most likely and negative if the alternative genotype was more likely. Stated otherwise, for the other traits the effect was considered positive if the higher mean was associated with the same marker haplotype that produced the higher mean for trait A and was negative otherwise.

RESULTS

Construction of a primary bovine DNA marker map: The 14 founder sires were genotyped for 181 previously described bovine microsatellite markers (FRIES *et al.* 1990; DIETZ 1992; GEORGES and MASSEY 1992; KAUKINEN *et al.* 1992). Markers known to reside on the sex chromosomes were not included in this study. Indeed, given the *paternal* half-sib pedigree structure, only male meioses are exploited, providing no linkage information for the sex chromosomes. At least one sire was found heterozygous for 159 of these markers. The mean heterozygosity for these 159 markers within the 14 founder sires was 56.4%. Informative families, *i.e.*, sibships for which the founder sire was heterozygous, were genotyped with the respective markers. The 104,523 resulting genotypes were used to construct the autosomal map shown in Table 1.

Of the 159 markers included in the analysis, 138 could be assigned to 27 linkage groups. These 27 linkage groups were assigned to 24 of the previously defined 29 autosomal synteny groups. Twenty-one of these synteny groups were characterized by one linkage group and 3 by two linkage groups: U3 and U11 and U20. The two linkage groups assigned to U20, representing 59 and 33 cM, respectively, correspond to chromosomes 23 and 20, respectively. These two chromosomes have been shown to have confounded segregation patterns in the somatic cell hybrids used (R. FRIES, personal communication). Of the remaining 21 markers, 2 were

discarded because they gave ambiguous patterns, whereas 19 remained as singletons, *i.e.*, showed no evidence for linkage with any of the other markers. Nine of these singletons could be assigned to a synteny group. With the exception of U25, there is at least one marker on each of the 29 previously defined autosomal synteny groups. It is likely that we have at least one marker on each of the 29 bovine autosomes, however, because the two chromosomes (20 and 23) with confounded segregation patterns in the panel of somatic cell hybrids used are represented in our map. The segregation pattern defining U25 could either be erroneous or correspond to a fragmented chromosome.

When converting recombination rates between adjacent markers to centimorgans using Kosambi's mapping function and summing over all linkage groups, we obtain a total of 1,645 autosomal centimorgans flanked by linked markers. Assuming that the male genome in cattle represents $\sim$2500 cM (as deduced from chiasmata counts OGUE and HARVEY 1978), this would correspond to a coverage of $\pm$66% of the genome. The mean bracket size equaled 14.8 cM.

Mapping QTL controlling milk production: We analyzed our data for the presence of detectable QTL affecting five milk production traits (milk yield, fat yield, protein yield, fat percentage, protein percentage) using a multilocus maximum likelihood approach (LATHROP *et al.* 1984; LANDER and BOTSTEIN 1986). The quantitative measurements used in the linkage analysis were DYDs (VANRADEN and WIGGANS 1991), obtained by progeny testing young dairy sires. Because the progeny test is based on a different number of daughters for each sire (from 50 to several thousands), an algorithm was developed that would account for variance heterogeneity of the phenotypes (see MATERIALS AND METHODS). Evidence for a QTL at the corresponding map position was expressed as a LOD score. Following LANDER and BOTSTEIN (1989) and knowing that we explored $\sim$16 Morgan with brackets of $\sim$15 cM, we chose a stringent LOD score threshold of three to reduce the chance of a false positive occurring anywhere in the genome to <5%.

Using this approach, we identified five chromosomes giving LOD scores $\geq$ 3 within bracketed segments: chromosomes 1 (U10), 6 (U15), 9 (U2), 10 (U5) and 20 (U20). The corresponding LOD score curves are shown in Figure 3, whereas the associated phenotypic effects are reported in Table 2.

Four of the 14 available pedigrees contributed to the detection of these QTL effects. An effect was found in each of the two families with 200 or more sons: pedigrees 1 (chromosome 1) and 10 (chromosome 20); one effect was identified in pedigree 9 with 115 sons (chromosome 6); and three linkages were found in pedigree 3 consisting of 103 sons (chromosomes 9, 10 and 20). Although the QTL effects on chromosomes 1, 6, 9 and 10 were only significant at the three LOD
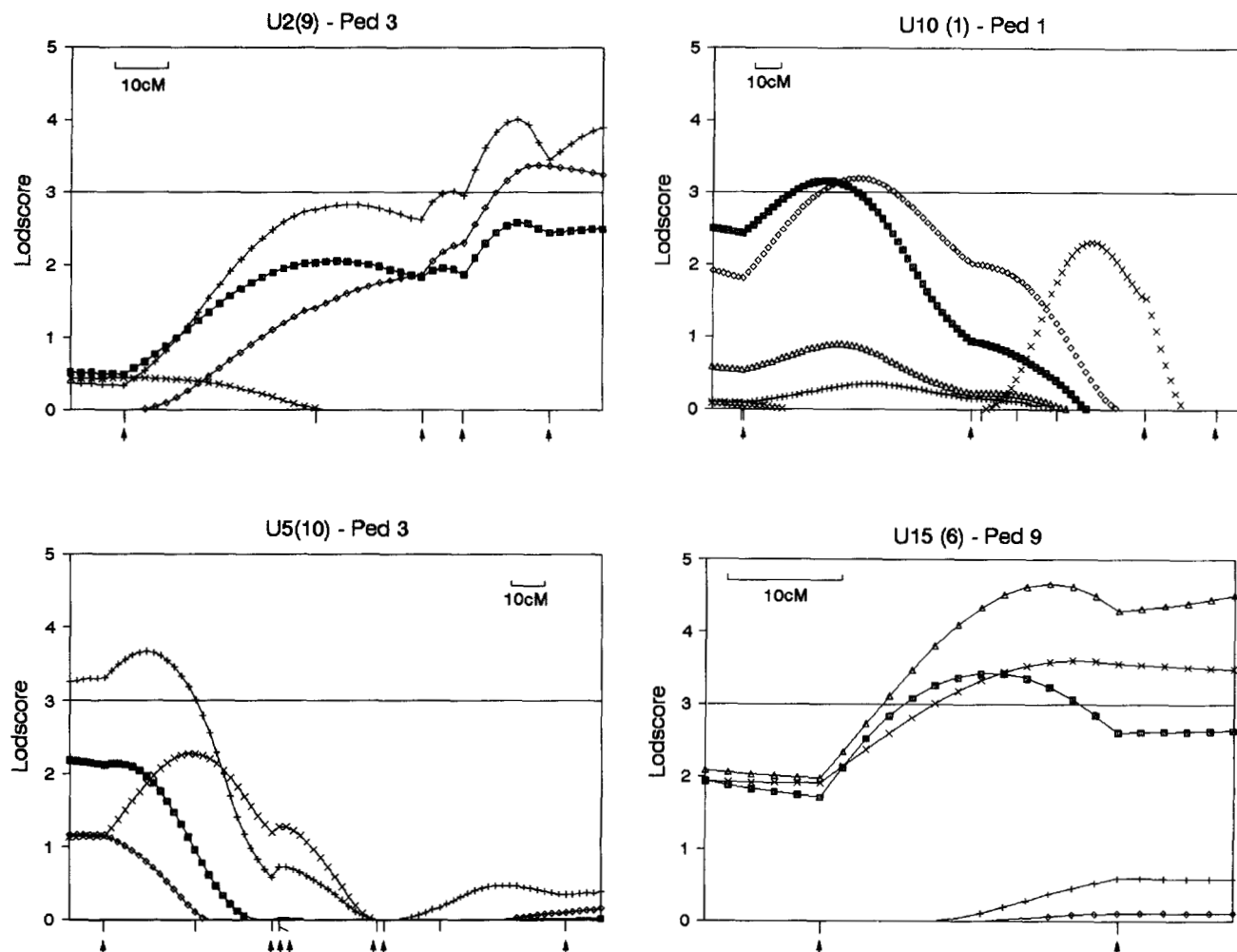
FIGURE 3.—LOD scores profiles for chromosome: pedigree combinations yielding LOD scores ≥ 3 for at least one trait. LOD score profiles are shown for the five studied DYDs: ■, milk yield; +, fat yield; ◇, protein yield; △, fat percentage; ×, protein percentage. The positions of the microsatellite markers in a linkage group are shown on the x-axis; the informative markers in the pedigree are indicated by arrows. The width of a 10-cM interval is indicated in each graph.

score threshold in a single pedigree, the effect on chromosome *20* was detected in pedigrees 3 and 10, pointing toward the segregation of the same QTL in both kindreds. For several of these chromosomes, however, additional pedigrees yielded LOD scores between 1 and 3 for the same or another of the milk production traits, which could reflect the segregation of the same or a related QTL in these families.

As expected, given the correlation observed between milk production traits, the identified QTL affect more than just one of the studied traits. In general, the LOD score curves pertaining to the different milk production traits always maximized in the same interval, indicating that a single QTL is likely responsible for all observed effects. One exception to this rule was observed on chromosome *1*, where the LOD scores maximized in the interval TGLA49–TGLA57 for all traits except protein percentage with a LOD score of 2.31 in the interval flanked by TGLA57 and TGLA130. Although this discrepancy might point toward two QTL segregating on

this chromosome, it is as likely to reflect inaccuracy in the estimation of QTL position.

Interestingly, each of the mapped QTL affects the different milk production traits in a distinct manner. The QTL on chromosome *9* increased the amount of milk produced without significantly altering its fat and protein composition; fat and protein yield were increased concomitantly. The two QTL on chromosomes *6* and *20*, on the other hand, appeared to increase milk yield but not fat or protein yield; fat and protein percentage were both reduced. The QTL on chromosomes *1* and *10* seemed to have differential effects on milk composition; whereas the higher milk yield was accompanied by a stronger increase in fat yield than in protein yield for the QTL on *10*, the opposite was observed for the QTL on *1*.

We found it prudent to limit ourselves to QTL candidates yielding LOD scores ≥ 3 within chromosomal segments bounded by informative markers. Figure 3, however, reports how the LOD score curves behaved
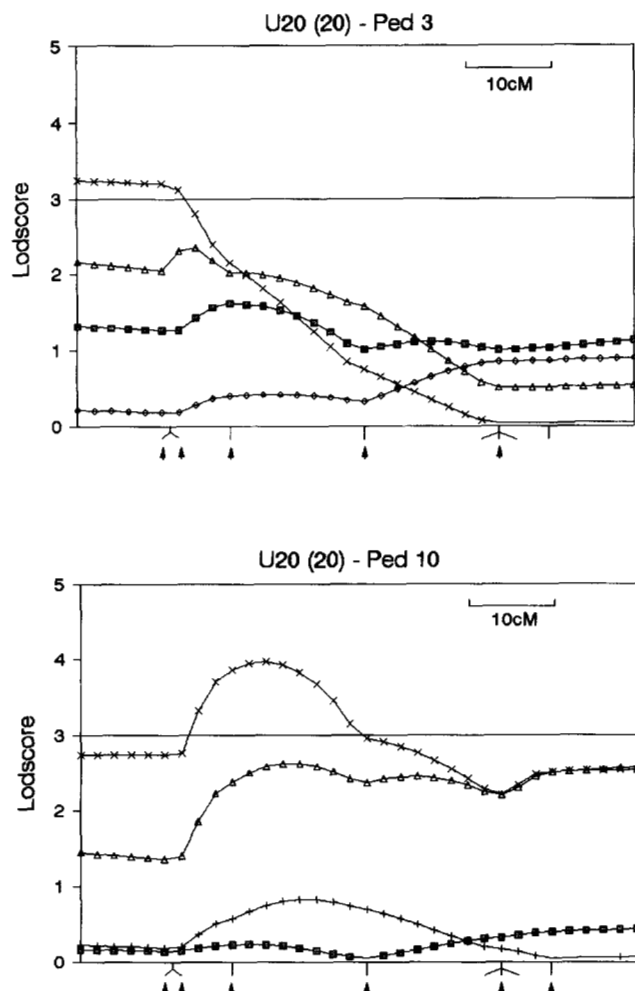
FIGURE 3.— *Continued*

of daughters for all sons) and assuming a trait with 30% heritability and a progeny test based on 100 daughters, 87% of σ^2_{DYD} is genetic in nature. One third of this genetic component, or 29%, of σ^2_{DYD} corresponds to the sire's Mendelian sampling variance. Consequently, the identified QTL would explain between 38 and 179% of the expected Mendelian segregation variation of the sire. Values above 100% obviously generate an interesting conundrum (see DISCUSSION).

Note that the *breeding value* (BV) of an individual is defined as *twice* the mean deviation of its progeny from the population mean (FALCONER 1989). Consequently, a DYD only contains *half* the BV of the bull, and therefore these QTL effects are estimates of half the *average effect of a gene substitution* $\alpha/2$ (FALCONER 1989). The same QTL would have effects twice as large if studying actual milking performances in the "daughter design." After multiplication by two, these effects consequently range from 0.62 to 1.34 additive genetic standard deviations, σ_A.

DISCUSSION

This study demonstrates that QTL with considerable effects on milk production traits can be found in dairy cattle populations intensely selected over several generations for increased milk production. The mapped QTL contribute to the genetic variance of milk production traits, which is exploited by artificial selection programs. Consequently, the identification of these QTL may open the way to more efficient marker-assisted selection schemes.

The availability of dinucleotide microsatellites as a convenient and abundant source of highly polymorphic genetic markers has made this project possible. Microsatellite genotyping based on the PCR is amenable to multiplex amplification and semiautomated procedures and allowed for a throughput of up to 1500 genotypes per person per week. Averaged over the entire duration of the project, the genotyping has been performed by four scientists in ~11 months with ~10,000 genotypes determined per month. The total costs of microsatellite genotyping for this project were estimated at $4–5 per genotype. Although the analysis of microsatellites requires a relatively tedious electrophoretic size separation, it is at the present time the preferred genetic marker system in mammals.

Based on the genotypes generated in this study, we constructed an autosomal microsatellite map composed of 147 markers placed on the map either by linkage analysis or by the use of somatic cell hybrids, or most often by combining both techniques. It is likely that there is at least one marker on each of the 29 bovine autosomes with a mean of five markers per chromosome. Twelve markers could not be positioned on the map. One hundred thirty-eight of the 147 positioned markers fell into 27 linkage groups, bracketing an esti-

beyond these boundaries. For the reported QTL, the LOD score curves always maximized within an interval bounded by two informative markers for the corresponding pedigree. The effect on protein percentage on chromosome *20* in pedigree 3 was the only exception to this rule. However, and because the LOD score curves on chromosome *20* peaked in the same interval (bounded by AGLA29/TGLA214 on one side and TGLA126 on the other) for all other traits in both pedigrees 3 and 10, this interval is the most likely location of the corresponding QTL.

For the significant linkages (LOD scores ≥ 3), the maximum likelihood estimates of the QTL effects ranged from 0.67 to 1.45 standard deviations σ_{DYD}, explaining from 11 to 52% of the total variance of DYD within a half-sib family, σ^2_{DYD}. As σ^2_{DYD} approximately equals $[0.1875\sigma^2_A + (0.5\sigma^2_A + \sigma^2_E)/n$ (where σ^2_A is the additive genetic variance of the corresponding trait in the population of interest, σ^2_E is environmental variance embracing all residual variation of nongenetic origin and n is the number of daughters included in the progeny test, assuming for simplicity an identical number

TABLE 2

Identified QTL effects

	DYD	Lods	$\alpha/2$	σ_{DYD}
U2(9)—pedigree 3	MY (kg)	2.58	+230	286
	FY (kg)	4.00	+10.1	10.9
	PY (kg)	3.38	+5.9	6.4
	F%	0.01	−0.22	0.98
	P%	0.44	−0.18	0.44
U5(10)—pedigree 3	MY (kg)	2.21	+336	286
	FY (kg)	3.67	+12.0	10.9
	PY (kg)	1.17	+5.2	6.4
	F%	0.00		0.98
	P%	2.27	−0.46	0.44
U10(1)—pedigree 1	MY (kg)	3.15	+266	299
	FY (kg)	0.34	+5	10.4
	PY (kg)	3.19	+8.2	9.1
	F%	0.90	−0.70	0.98
	P%	2.31	−0.70	0.51
U15(6)—pedigree 9	MY (kg)	3.42	−244	291
	FY (kg)	0.59	+3.2	10
	PY (kg)	0.10	+1.4	7.7
	F%	4.66	+1.58	1.13
	P%	3.60	+1.07	0.63
U20(20)—pedigree 10	MY (kg)	0.41	−78	294
	FY (kg)	0.83	+3.6	10.9
	PY (kg)	0.00		8.2
	F%	2.62	+0.77	1.29
	P%	3.99	+0.40	0.58
U20(20)—pedigree 3	MY (kg)	1.61	−171	286
	FY (kg)	0.00		10.9
	PY (kg)	0.84	+4.1	6.4
	F%	2.35	+0.77	0.98
	P%	3.20	+0.46	0.44

For each synteny group, pedigree combination yielding a LOD score ≥ 3 for at least one of the five traits studied, we report the highest LOD scores (LODs) observed for each trait, the maximum likelihood estimates of 0.5α at the corresponding positions (not necessarily the same positions for the different traits; see Figure 3), with α corresponding to the average effect of a QTL allele substitution, and the standard deviation for the respective DYDs in the corresponding half-sib family (σ_{DYD}).

mated 1645 cM (Kosambi). Assuming that the male genome in cattle represents ~2500 cM—as deduced from chiasmata counts—this would correspond to a coverage of ±66% of the genome. However, Monte-Carlo simulations (data not shown) indicate that the likelihood to cover 1645 cM (Kosambi) of the bovine genome with bracketed segments not larger than 35 recombination units, using 150 randomly selected markers, is <0.001. The expected coverage has a mean value of 1343 cM (Kosambi) (54% of the genome) with a standard deviation of 97 cM. A number of factors may underlie this apparent contradiction. Despite the multiple checkpoints imposed on the genotyping procedure, remaining typing errors may inflate the estimated map distances. Further, the level of interference assumed under Kosambi's mapping function may be insufficient for the bovine genome.

In conjunction with two recently published bovine maps (BARENDSE *et al.* 1994; BISHOP *et al.* 1994), this marker set should give very adequate genome coverage for further mapping studies in cattle. As both these maps included large numbers of markers used in the present study as well, cross-referencing between maps should be relatively straightforward.

To map QTL for milk production, we have used the "granddaughter design" (WELLER *et al.* 1991) that takes advantage of two features characterizing dairy cattle populations: (1) the occurrence of large paternal half-sib families, resulting from the widespread use of artificial insemination, allowing for within-family analysis and consequently simplifying issues of genetic heterogeneity and (2) the progeny-testing scheme used for the genetic evaluation of young dairy bulls, which reduces the environmental noise when searching for QTL.

We used a statistical method related to interval mapping (LATHROP *et al.* 1984; LANDER and BOTSTEIN 1986). However, our method differs from conventional interval mapping as described by LANDER and BOTSTEIN (1986) in that information from all markers composing the linkage group is used in computing the likelihood

at a putative QTL location instead of information from flanking markers only. A similar algorithm has been used previously by HILBERT *et al.* (1991). Although this approach is computationally more costly, it has a number of advantages over conventional interval mapping. Not only does it extract more information from the available data, it also eliminates the discontinuities in the LOD score curves that can be observed at the markers, caused by adjacent pairs of markers providing different amounts of information (KNOTT and HALEY 1992). Furthermore, our method differs from conventional interval mapping in the fact that we analyzed phenotypic averages in an unbalanced design instead of individual phenotypes, requiring us to account for variance heterogeneity of the phenotypes.

The objective of this study was to identify chromosomal segments, giving strong evidence for the presence of genes affecting milk production traits. A stringent LOD score threshold of three was chosen to reduce the type I error rate to <5% despite the large numbers of markers tested. Note that the type I error associated with a LOD score threshold of three is likely to be lower than 5% in this case. Indeed, the LOD score thresholds derived by LANDER and BOTSTEIN (1989) are based on the null hypothesis that no QTL exists. Although this approach is appropriate for a trait with unknown genetic base, it may not be the best choice for a trait with known significant heritability, for example, for milk production traits with well-estimated heritabilities ranging from 0.3 to 0.5. However, we deemed this very stringent LOD score threshold appropriate at this early stage of QTL research in livestock. It is indeed important to demonstrate convincingly that QTL can be mapped in segregating populations, which is a major milestone in the implementation of marker-assisted selection in livestock. Moreover, only regions achieving stringent significance levels as strong evidence in favor of the presence of QTL justify the considerable investments required for their further characterization. A better understanding of these loci, indispensable for their exploitation in breeding programs, will depend on the development of high density genetic and physical maps in the regions of interest and on the genotyping of additional samples to dissect the QTL into component alleles and accurately estimate their effects and population frequencies.

However, choosing a very high LOD score threshold has the disadvantage that it reduces the power of the design and results in overestimation of the QTL effects when using ML methods. This is illustrated by the following simulations. Paternal half-sib pedigrees with 50, 100 and 200 offspring were generated segregating for QTL with effects of 0.25, 0.5 and 1.0 σ_{DYD} (σ^2_{DYD} being the variance observed within half-sib families). Reliabilities were considered identical for all offspring. The QTL was placed within a 20-cM chromosomal segment

TABLE 3

QTL mapping simulations

	$n = 50$	$n = 100$	$n = 200$
$1 \times \sigma$			
(a)	40%	90%	100%
(b)	1.21	1.03	1.00
(c)	3.8	3.1	2.4
$0.75 \times \sigma$			
(a)	9%	44%	92%
(b)	1.18	0.92	0.77
(c)	4.0	3.8	3.1
$0.50 \times \sigma$			
(a)	1.5%	7.7%	37%
(b)	1.15	0.89	0.67
(c)	5.0	4.4	4.1
$0.25 \times \sigma$			
(a)	0%	0.1%	2.2%
(b)		0.84	0.62
(c)		10	4.7

Paternal half-sib pedigrees with 50, 100 and 200 offspring were generated segregating for QTL with effects of $1 \times \sigma$, $0.75 \times \sigma$, $0.5 \times \sigma$ and $0.25 \times \sigma$, where σ^2 is the within half-sib family variance; 1000 pedigrees were generated for each condition. The table reports (a) the % of runs yielding a LOD score ≥ 3, (b) the average maximum likelihood estimate of QTL effect in σ units for runs with lodscore ≥ 3 and (c) the average distance between maximum likelihood position of QTL and real position of QTL in cM for runs with LOD scores ≥ 3.

bounded by two markers, at 5 cM from one of these. The markers were characterized by ideal polymorphism information content so that the paternal allele could be determined unambiguously for all offspring. QTL yielding LOD scores ≥ 3 were considered significant. Table 3 summarizes the results of this simulation. With decreasing size of the actual effect, the power of detection was reduced and the magnitude of the effect was increasingly overestimated. The precision in the localization of the QTL diminished concomitantly. These results show that the effects of the identified QTL are likely to be inflated. Overestimation is further supported by the magnitude of the estimated segregation variances of some of the mapped QTL, which exceeded the entire Mendelian sampling variance in the sire. The overestimation of the QTL effects may have been somewhat reduced, however, by the selection bias in our pedigree material causing an underestimation of the QTL effects (MACKINNON and GEORGES 1992).

Nevertheless, the mere fact that we are detecting QTL from the available pedigree material indicates that genes with substantial effects on milk production traits are still segregating in these elite populations, even if we cannot confidently estimate their magnitude because of the limited power of the experimental design. This raises the question why QTL with such effects are still segregating in intensely selected populations. A number of hypotheses can be enumerated. First, DEKKERS

and DENTINE (1991) pointed out that even under strictly polygenic inheritance, large segregation effects would be commonplace. Indeed, some of the identified effects may reflect the combined action of two or more linked QTL. Occasionally, favorable haplotypes could be detected in the offspring of some individuals as large segregating effects, but the fixation of these haplotypes in the population would be hampered by recombination between the QTL. Second, despite the intense selection applied to dairy populations, the selection intensity on individual QTL may be limited. The two QTL mapping to chromosome 6 and 20, causing an increase in milk volume without significant increase in protein and fat yield, may actually be quite neutral with respect to the actual selection criteria that are emphasizing component yield. Moreover, artificial selection programs for livestock populations do not consider just one but usually several traits simultaneously, reducing selection pressure and progress in any particular trait. Third, possible associations of apparently favorable effects on milk production and unfavorable characteristics, whether monogenic disorders or polygenic health and fertility traits, must be considered as well. Such associations could be due to pleiotropic effects of the same genes or linkage between the QTL and genes with unwanted effects. A now classical example is the association found between milk production and progressive degenerative myeloencephalopathy or Weaver in Brown Swiss (HOESCHELE and MEINERT 1990; GEORGES *et al.* 1993). Finally, the identified effects could also be due to recent mutations, possibly appearing in a limited pool of target genes. Phenomena like anticipation, having recently received satisfactory explanation at the molecular level in the expansion of triplet repeats (KUHL and CASKEY 1993), allow one to speculate about the possible nonconventional origins of new mutations.

Although our approach has proven successful to map QTL, the figures reported in Table 3 indicate the need to develop superior methods of analysis. The overestimation of the QTL effects can be adjusted by treating them as random rather than fixed effects. Random treatment implies assuming a prior distribution of gene effects where minor are more likely than major effects, for example, a normal or exponential distribution. The estimated QTL effects are then regressed back to their mean prior to a degree reflecting the amount of available information. Random treatment of QTL effects has been proposed by FERNANDO and GROSSMAN (1989) and GODDARD (1992) in a best linear unbiased prediction type of analysis and more recently by HOESCHELE and VANRADEN (1993a,b) using a Bayesian approach. This Bayesian linkage analysis uses a prior probability of linkage of a QTL to a single marker or of location within a marker bracket, computed as a function of genome, chromosome and interval length

and heritability of a trait, in the test criterion rather than the null hypothesis that no QTL exists.

The underestimation of the QTL effects due to selection bias in our pedigree material (MACKINNON and GEORGES 1992) may be compensated for by including DYDs of nongenotyped sons in the analysis, as suggested by selection theory (IM *et al.* 1989).

Although in this work we have limited ourselves to a within-family analysis, *i.e.*, each half-sib family was treated separately, one can perform an across-pedigree analysis, considering all half-sib pedigrees jointly. A further extension would include more or all known relationships between individuals. This full-pedigree analysis would extract more information from the data, allow accurate modeling of the residual variances of the DYDs and permit the search of QTL on the *X* chromosome. However, some of the assumptions required in across- or full-pedigree analysis, for example, the segregation of a single biallelic QTL per chromosome, are questionable. Moreover, developing algorithms that would deal with the large number of inbreeding and marriage loops typical for cattle populations is far from trivial. Efficient algorithms for implementing a full-pedigree analysis may now be obtained from Markov chain Monte Carlo and data augmentation methods (GUO and THOMPSON 1992).

Increased efficiency may also be obtained from a multivariate analysis, *i.e.*, by analyzing different correlated traits jointly rather than separately as in our study. Finally, having identified a first set of QTL will allow us to include these effects in our models, which, by accounting for some of the "genetic noise," should facilitate the detection of subsequent QTL.

The complexity of lactational physiology and the relative crudeness of the measurements analyzed in the present study make it very difficult to speculate about the plausible mode of action of the identified QTL. However, some ideas may be obtained from their differential effect on the five studied traits. The two QTL mapping to chromosomes 6 and 20, respectively, apparently act by causing an increase in milk volume without concomitant increase of fat and protein yield and consequently with marked reduction in fat and protein percentage. Such an effect could result from an increase in osmotic pressure of the milk, causing an influx of water in the mammary gland to restore isosmolarity with blood plasma. As lactose is the major osmole, the observed effects could for instance result from an increase of lactose secretion in the milk. The effect observed on chromosome 9, consisting of an increase in milk volume without significant alteration of its fat and protein content, could be due to an increase in the number of mammary secretory cells or an increase in the mean milk yield per cell. The number of secretory cells reflects the equilibrium between mammary cell proliferation or mammogenesis and mammary involution. Both are controlled by a complex set of systemic

and local factors varying throughout an animal's life cycle. Similarly, lactogenesis or initiation of milk production and galactopoiesis or maintenance of milk production, both determining the mean milk yield per cell, are affected by systemic and local controls. What mechanisms might underlie an increase in milk production accompanied by a differential effect on fat and protein composition as observed on chromosomes *1* and *10* is unknown. It is interesting to note that an effect on protein yield was recently identified in goats that was confidently attributed to the α_{s1}-casein locus and accompanied by a significant effect on fat composition (MAHE *et al.* 1993). This finding points toward the existence of an as yet unexplained interaction between protein and fat composition.

Among the five chromosomes where we found strong evidence for the presence of a QTL, only chromosome *6* harbors known candidate genes for the observed effect: the casein locus. Although one cannot exclude that the effect reported on chromosome *6* is due to the casein locus, the most likely position of the QTL and the type of effect observed—different from all other effects of the casein locus reported in the literature—favor an independent QTL. It is noteworthy, however, that pedigree 12 showed preliminary evidence (LOD score 2.53) for the segregation of a QTL with a 4.4-kg substitution effect on protein yield compatible with a direct effect of the casein locus. We also found evidence for a QTL on chromosome *23* segregating in pedigree 10 in the vicinity of the locus harboring the family of prolactin-related genes (DIETZ *et al.* 1992). The effect was most pronounced on fat yield ($\alpha/2 = 5.9$ kg). Because the LOD score exceeded three only outside of a chromosomal segment bracketed by informative markers, this effect did not satisfy our significance criteria. Note that pedigree 10 is different from the family in which an effect of the corresponding region was reported previously (COWAN *et al.* 1990).

The identification of QTL segregating in elite dairy cattle populations is the first step toward the application of marker-assisted selection for milk production. The selection of young dairy sires, which presently relies on the expensive and time-consuming progeny-testing procedure, offers a unique opportunity for the usage of genetic markers in livestock production. Young dairy bulls result from planned matings of "bull sires" and "bull dams" with the highest BVs or PTAs (VANRADEN and WIGGANS 1991). The predicted BV of the offspring correspond to the average of the parental BVs. The actual BV of the offspring will, however, deviate from the predicted, because the estimates of the parental BV is not fully accurate and mainly because of Mendelian sampling effects, or the fact that different offspring receive a different sample of genes from their parents. Progeny testing has been implemented for that very reason.

Several years ago the use of genetic markers in the selection of young dairy bulls was proposed (SOLLER and BECKMAN 1982; SMITH and SIMPSON 1986). The benefit of markers has often been analyzed in terms of improved accuracy of selection. The gain to be made following this approach is generally accepted to be marginal (SMITH and SIMPSON 1986; MEUWISSEN and VAN ARENDONK 1992). Furthermore, to be effective, this approach requires a very detailed understanding of the identified QTL in terms of number of segregating alleles and their respective effects, which may be very difficult to achieve in the near future.

In the short term the major advantage of markers will likely result from predicting parts of the Mendelian sampling effects at a stage where current selection schemes do not provide any information to differentiate among full-sibs. As multiple ovulation and embryo transfer enables us to produce larger numbers of full-sibs in dairy cattle, markers will allow preselection among full-sib brothers before progeny testing, thus testing only those more likely to have BVs superior to the parental mean. To implement such a scheme, it must be determined for which of the identified QTL bull sire and bull dam are heterozygous. Indeed, these are the QTL contributing to differentiation among siblings due to Mendelian sampling in a given mating. In the short term, this analysis seems difficult to achieve for bull dams. However, segregation analysis (involving markers linked to identified QTL) using the progeny-test daughters may allow determination of heterozygosity in the bull sire. The feasibility of such a scheme has been examined by HOESCHELE and ROMANO (1994) and is presently under further study. KASHI *et al.* (1990) proposed to use information from paternal and maternal grandsire to select among QTL alleles from bull sire and bull dam, respectively. Although in the short term this might be the only feasible alternative to select among the QTL alleles of the bull dam, this approach has the disadvantage that selection potential is wasted on QTL for which the bull dam is homozygous and which consequently do not contribute to Mendelian sampling.

In the long term, a better understanding of QTL parameters may lead to more complex strategies combining phenotypic and QTL data into a single analysis.

Our demonstration that QTL can be mapped in highly selected segregating dairy populations should strongly encourage efforts to develop selection schemes incorporating marker information.

LITERATURE CITED

ANDERSSON-EKLUND, L., and J. RENDEL, 1993 Linkage between amylase-1 locus and a major gene for milk fat content in cattle. Anim. Genet. **24:** 101–103.

ARNHEIM, N., C. STRANGE and H. ERLICH, 1985 Use of pooled DNA samples to detect linkage disequilibrium of polymorphic restriction fragments and human disease: studies of the HLA class II loci. Proc. Natl. Acad. Sci. USA **88:** 507–511.

BARENDSE, W., S. M. ARMITAGE, L. M. KOSSAREK, A. SHALOM, B. KIRCKPATRICK *et al.*, 1994 A genetic linkage map of the bovine genome. Nature Genet. **6:** 227–235.

BISHOP, M. D., S. M. KAPPES, J. W. KEELE, R. T. STONE, S. L. F. SUNDEN *et al.*, 1994 A genetic linkage map for cattle. Genetics **136:** 619–639.

BOVENHUIS, H., 1992 The relevance of milk protein polymorphisms for dairy cattle breeding. Doctoral Thesis, Department of Animal Breeding, Wageningen Agricultural University, The Netherlands.

BOVENHUIS, H., and J. I. WELLER, 1994 Mapping and analysis of dairy cattle quantitative trait loci by maximum likelihood methodology using milk protein genes as genetic markers. Genetics **137:** 267–280.

COWAN, C. M., M. R. DENTINE, R. L. AX and L. A. SCHULER, 1990 Structural variation around prolactin gene linked to quantitative traits in an elite Holstein sire family. Theor. Appl. Genet. **79:** 577–582.

DEKKERS, J. C. M., and M. R. DENTINE, 1991 Quantitative genetic variance associated with chromosomal markers in segregating populations. Theor. Appl. Genet. **81:** 212–220.

DIETZ, A. B., 1992 Analysis of the prolactin related proteins and placental lactogen family and the development of a sequence tag site based bovine gene map. Dissertation, Texas A & M University.

DIETZ, A. B., M. GEORGES, D. W. THREADGILL, J. E. WOMACK and L. A. SCHULER, 1992 Somatic cell mapping, polymorphism, and linkage analysis of bovine prolactin-related proteins and placental lactogen. Genomics **14:** 137–143.

FALCONER, D. S. 1989 *Introduction to Quantitative Genetics*, Ed. 5. Longman Scientific and Technical, New York.

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

FRIES, R., A. EGGEN and G. STRANZINGER, 1990 The bovine genome contains polymorphic microsatellites. Genomics **8:** 403–406.

GELDERMANN, H., U. PIEPER and B. ROTH, 1985 Effects of marked chromosome sections on milk performance in cattle. Theor. Appl. Genet. **70:** 138–146.

GEORGES, M., and J. MASSEY, 1992 *Polymorphic DNA markers in Bovidae.* WO Pub. 92/13102, World Intellectual Property Org., Geneva.

GEORGES, M., A. B. DIETZ, A. MISHRA, D. NIELSEN, L. S. SARGEANT *et al.*, 1993 Microsatellite mapping of the gene causing weaver disease in cattle will allow the study of an associated quantitative trait locus. Proc. Natl. Acad. Sci. USA **90:** 1058–1062.

GODDARD, M. E., 1992 A mixed model for analysis of data on multiple genetic markers. Theor. Appl. Genet. **83:** 878–886.

GUO, S. W., and E. A. THOMPSON, 1992 A Monte Carlo method for combined segregation and linkage analysis. Am. J. Hum. Genet. **51:** 1111–1126.

HILBERT, P., K. LINDPAINTER, J. S. BECKMAN, T. SERIKAWA, F. SOUBRIER *et al.*, 1991 Chromosomal mapping of two genetic loci associated with blood-pressure regulation in hereditary hypertensive rats. Nature **353:** 521–529.

HOESCHELE, I., and T. R. MEINERT, 1990 Association of genetic defects with yield and type traits: the weaver locus effect on yield. J. Dairy Sci. **73:** 2503–2515.

HOESCHELE, I., and E. O. ROMANO, 1994 On the use of marker information from granddaughter designs. J. Anim. Breed. Genet. **110:** 429–449.

HOESCHELE, I., and P. M. VANRADEN, 1993a Bayesian analysis of linkage between genetic markers and quantitative trait loci. I. Prior knowledge. Theor. Appl. Genet. **85:** 946–952.

HOESCHELE, I., and P. M. VANRADEN, 1993b Bayesian analysis of linkage between genetic markers and quantitative trait loci. II. Combining prior knowledge with experimental evidence. Theor. Appl. Genet. **85:** 953–960.

IM, S., E. A. FERNANDO and D. GIANOLA, 1989 Likelihood inferences in animal breeding under selection: a missing data theory viewpoint. Genet. Sel. Evol. **21:** 399.

KASHI, Y., E. HALLERMAN and M. SOLLER, 1990 Marker-assisted selection of candidate bulls for progeny testing programmes. Anim. Prod. **51:** 63–74.

KAUKINEN, J., and S. L. VARVIO, 1992 Bovine microsatellites and SINE-elements. Anim. Genet. **23** (Suppl. 1): 61.

KNOTT, S. A., and C. S. HALEY, 1992 Maximum likelihood mapping of quantitative trait loci using full-sib families. Genetics **132:** 1211–1222.

KUHL, D. P. A., and C. T. CASKEY, 1993 Trinucleotide repeats and genome variation. Curr. Opin. Genet. Dev. **3:** 404–407.

LALOUEL, J. M., 1983 Optimization of functions. Contrib. Epidemiol. Biostat. **4:** 235–259.

LANDER, E.S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185–199.

LATHROP, G. M., J.-M. LALOUEL, C. JULIER and J. OTT, 1984 Strategies for multilocus linkage analysis in humans. Proc. Natl. Acad. Sci. USA **81:** 3443–3446.

LOGUE, D. N., and M. J. A. HARVEY, 1978 Meiosis and spermatogenesis in bulls heterozygous for a presumptive 1/29 Robertsonian translocation. J. Reprod. Fertil. **54:** 159–165.

MACKINNON, M. J., and M. GEORGES, 1992 The effects of selection on linkage analysis for quantitative traits. Genetics **132:** 1177–1185.

MAHE, M. F., E. MANFREDI, G. RICORDEAU, A. PIACERE and F. GROSCLAUDE, 1993 Effects of the s1 casein polymorphism on goat dairy performances: a within-sire analysis of alpine bucks. Genet. Sel. Evol. **26:** 151–158.

MEEUWISSEN, T. H. E., and J. A. M. VAN ARENDONK, 1992 Potential improvements in rate of genetic gain from marker assisted selection in dairy cattle breeding schemes. J. Dairy Sci. **75:** 1651–1659.

PATERSON, A. H., E. S. LANDER, J. D. HEWITT, S. PETERSON, S. E. LINCOLN *et al.*, 1989 Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature **335:** 721–726.

PEARSON, R.E., W. E. VINSON and T. R. MEINERT, 1990 The potential for increasing productivity through selection for increased milk and component yields, pp. 104–113 in Proceedings of the 4th World Congress on Genetics Applied to Livestock Production, Vol. XIV, edited by W. G. HILL, R. THOMPSON and J. A. WOOLIAMS.

SCHUTZ, M. M, A. E. FREEMAN, G. L. LINDBERG and D. C. BEITZ, 1993 Effects of maternal lineages grouped by mitochondrial genotypes on milk yield and composition. J. Dairy Sci. **76:** 621–629.

SMITH, C., and S. P. SIMPSON, 1986 The use of genetic polymorphism in livestock improvement. J. Anim. Breed. Genet. **103:** 205–217.

SOLLER, M., and J. S. BECKMAN, 1982 Restriction fragment length polymorphisms and genetic improvement, pp. 396–404 in Proceedings of the 2nd World Congress on Genetics Applied to Livestock Production, Vol. VI, Madrid.

SOLLER, M., T. BRODY and A. GENIZI, 1976 On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor. Appl. Genet **47:** 35–39.

VANRADEN, P. M., and G. R. WIGGANS, 1991 Derivation, Calculation, and use of National Animal Model Information. J. Dairy Sci. **74:** 2737–2746.

WEBER, J. L., and P. E. MAY, 1989 Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. **44:** 388–396.

WELLER, J. L., Y. KASHI and M. SOLLER, 1990 Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. J. Dairy Sci. **73:** 2525–2537.

Communicating editor: B. S. WEIR

APPENDIX A

The reduction in sample size resulting from progeny testing can be understood as follows. In the GDD, marker genotyping and linkage analysis are performed in half-sib pedigrees each consisting of a founder sire and a large number of his sons, which have been progeny tested for milk production and possibly other traits of economic importance. The quantitative traits considered in the analysis are the son's DYDs (VANRADEN and WIGGANS 1991). A DYD is an unregressed weighted average of the lactation performances (expressed as deviation from the population mean) of the daughters of a bull. Note that the BV of an individual is defined as twice the mean deviation of its progeny from the population mean (FALCONER 1989). Consequently, a DYD only contains *half* the BV of the bull. Before averaging, the lactation yields are adjusted for systematic environmental effects and PTAs of the daughter's dams. The variance of DYD within paternal half-brother families thus equals approximately $0.1875\sigma_A^2 + (0.625\sigma_A^2 + \sigma_E^2)/n]$, where σ_A^2 is the additive genetic variance of the corresponding trait in the population of interest, σ_E^2 is environmental variance embracing all residual variation of nongenetic origin, and n is the number of daughters included in the progeny test (assuming for simplicity an identical number of daughters for all sons). If the search for QTL is performed by comparing the DYD means of the sons sorted by paternal marker allele, the number of sons required to detect an *average effect of a gene (marker) substitution of* α (FALCONER 1989) equals $[4(t_0 + t_1)^2 \times \sigma_R^2/0.25\alpha^2]$ (SOLLER *et al.* 1976), where t_0 is the value of the t distribution associated with the desired type I error, t_1 is the value of the t distribution associated with twice the desired type II error, and σ_R^2 corresponds to the residual variance not accounted for by the marker, *i.e.*, $[0.1875\sigma_A^2 + (0.625\sigma_A^2 + \sigma_E^2)/n - 0.0625\alpha^2]$. Indeed, as DYD contains *half* the *breeding value*, a locus with average effect of a gene (marker) substitution of α generates an expected difference among sons DYDs of 0.5α.

In the alternative "daughter design" (DD), in which no advantage is taken of progeny testing, marker genotyping and linkage analysis are performed in half-sib pedigrees composed of a founder sire and a large number of its milking daughters. The quantitative traits used are the yield deviations (YD) of the daughters (VANRADEN and WIGGANS 1991). YD are weighted average lactation performances expressed as deviations from the population mean, adjusted for management group, permanent environmental effects and herd–sire interaction effects. The variance of YD within paternal half-sister families equals approximately $[0.75\sigma_A^2 + \sigma_E^2]$. Assuming again that the search for QTL is performed by comparing the YD means of the daughters sorted by paternal marker allele, the number of daughters required to detect an average effect of a gene (marker)

substitution of α would equal $[4(t_0 + t_1)^2 \times \sigma_R^2/\alpha^2]$, with $\sigma_R^2 = [0.75\sigma_A^2 + \sigma_E^2 - 0.25\alpha^2]$.

The ratio between the sample sizes required for DD *vs.* GDD then reduces to $(n - 0.25nh^2 - 0.25nh^2c^2)/(4 - 2.5h^2 - 3.25nh^2 - 0.25nh^2c^2)$, where h^2 corresponds to the heritability of the studied trait and c to the ratio α/σ_A. For a trait with $h^2 = 0.25$ and an average effect of a gene (marker) substitution α of $0.5\sigma_A$, a GDD with progeny test based on 100 daughters requires ~3.5 times less genotyping than the DD.

APPENDIX B

Variance $\sigma_j^2 = [0.25/Rel_{son(j)} - 0.0625Rel_{sire(j)} - 0.0625Rel_{dam(j)}] \sigma_A^2$: The population variance of true *half* breeding values or TAs equals $0.25\sigma_A^2$, with σ_A^2 corresponding to the additive genetic variance.

The variance of DYD_j equals:

$$Var(DYD_j) = 0.25\sigma_A^2 + (0.625\sigma_A^2 + \sigma_E^2)/n_j,$$

with σ_E^2 being the environmental variance and n_j the number of daughters used for the progeny test. The term $0.625\sigma_A^2$ accounts for the daughter yields being corrected for the PTA of the son's mates (the dams of the daughters), or $0.625\sigma_A^2 = 0.5 + 0.25(1 - Relm)$, with Relm representing the reliability of the PTA of the mate assumed to equal 0.5.

The reliability of DYD_j, $Rel_{son(j)}$, is the ratio between these two variances:

$$Rel_{son(j)} = Var(TA_j)/Var(DYD_j)$$
$$= 0.25\sigma_A^2/[0.25\sigma_A^2 + (0.625\sigma_A^2 + \sigma_E^2)/n_j].$$

Consequently, the variance of DYD_j:

$$Var(DYD_j) = 0.25\sigma_A^2/Rel_{son(j)}.$$

After adjustment for average parental transmitting ability, the variance of DYD_j would be reduced by $2 \times 0.0625\sigma_A^2$, if the parental transmitting abilities were known exactly ($Rel_{sire} = Rel_{dam} = 1$). With Rel_{sire} and $Rel_{dam} < 1$, the variance is reduced less to reflect uncertainty about the true parental transmitting abilities given their PTAs:

$$\sigma_j^2 = Var(DYD_j - 0.5PTA_{sire(j)} - 0.5PTA_{dam(j)})$$
$$= 0.25\sigma_A^2/Rel_{son(j)} - 0.0625\sigma_A^2$$
$$\times Rel_{sire(j)} - 0.0625\sigma_A^2 \times Rel_{dam(j)}.$$

Strictly speaking, σ_j^2 as defined here is an approximation because it is further reduced by an amount $0.0625\alpha^2$, because the QTL segregation effect is included in the mean μ_j. However, this is compensated for by a slight underestimation of σ_A^2, when estimated simultaneously with α in the ML analysis.

Calculation of $Rel_{sire(j)}$, $Rel_{dam(j)}$ and $Rel_{son(j)}$:

$$Rel_{sire(j)} = Rel_{PTAsire(j)}.$$

$\text{Rel}_{\text{sire}(j)}$, or the reliability of the sire's PTA, was obtained from the sire summary database of January 1993 of the US Department of Agriculture. As the founder sires have large numbers of daughters, sons and other relatives, their reliabilities approach one.

$$\text{Rel}_{\text{dam}(j)} = \text{Rel}_{\text{PTAdam}(j)}.$$

$\text{Rel}_{\text{dam}(j)}$, or the reliability of the dam's PTA, was calculated from the reliability of the average parental transmitting ability ($\text{Rel}_{\text{PA}(j)}$) and the reliability of the sire's PTA ($\text{Rel}_{\text{PTAsire}(j)}$), as follows:

$$\text{Rel}_{\text{dam}(j)} = 4 \times \text{Rel}_{\text{PA}(j)} - \text{Rel}_{\text{PTAsire}(j)}$$

(VANRADEN and WIGGANS 1991).

As for $\text{Rel}_{\text{PTAsire}(j)}$, $\text{Rel}_{\text{PA}(j)}$ was obtained from the sire summary database of January 1993 of the US Department of Agriculture.

$$\text{Rel}_{\text{son}(j)} = \text{Rel}_{\text{PTson}(j)}.$$

$\text{Rel}_{\text{son}(j)}$, or the reliability of the son's PTA *contributed by the progeny test* (DYD uses only information from progeny test daughters), was calculated from the reliability of the son's PTA, $\text{Rel}_{\text{PTAson}(j)}$ (including information of progeny test daughters *and* ancestors) as follows. $\text{Rel}_{\text{PTAson}(j)}$ and $\text{Rel}_{\text{PA}(j)}$ were converted to "daughter equivalents" (DE) (VANRADEN and WIGGANS 1991) using:

$$\text{Rel}_{\text{PTAson}(j)} = \text{DE}_{\text{PTAson}(j)} / [\text{DE}_{\text{PTAson}(j)} + 14]$$

and

$$\text{Rel}_{\text{PA}(j)} = \text{DE}_{\text{PA}(j)} / [\text{DE}_{\text{PA}(j)} + 14].$$

Then $\text{Rel}_{\text{son}(j)} = \text{Rel}_{\text{PTson}(j)}$ was computed using:

$$\text{DE}_{\text{PTson}(j)} = \text{DE}_{\text{PTAson}(j)} - \text{DE}_{\text{PA}(j)}$$

and

$$\text{Rel}_{\text{PTson}(j)} = \text{DE}_{\text{PTson}(j)} / [\text{DE}_{\text{PTson}(j)} + 14].$$

$\text{Rel}_{\text{PTAson}(j)}$ was obtained from the sire summary database of January 1993 of the US Department of Agriculture.