Estimation of Quantitative Trait Loci Parameters for Milk Production Traits in German Holstein Dairy Cattle Population

J. Szyda,^{1,2} **Z. Liu**,¹ **F. Reinhardt**,¹ **and R. Reents**¹ ¹VIT, Heideweg 1, 27-283 Verden, Germany ²Department of Animal Genetics, Wrocław Agricultural University, 51-631 Wrocław, Poland

ABSTRACT

The main objective of this study was to estimate the proportion of total genetic variance attributed to a quantitative trait locus (QTL) on Bos taurus autosome 6 (BTA6) for milk production traits in the German Holstein dairy cattle population. The analyzed chromosomal region on BTA6 spanned approximately 70 cM, and contained 6 microsatellite markers. Milk production data were obtained from routine genetic evaluation for 4500 genotyped German Holstein bulls. Technical aspects related to the estimation of model parameters for a large data set from routine genotype recording were outlined. A fixed QTL model and a random QTL model were introduced to incorporate marker information into parameter estimation and genetic evaluation. Estimated QTL variances, expressed as the ratio of QTL to polygenic variances, were 0.04, 0.03, and 0.07 for milk yield; 0.06, 0.08, and 0.14 for fat yield; and 0.04, 0.04, and 0.11 for protein yield, in the first 3 parities, respectively. The estimated QTL positions, expressed as distances from the leftmost marker DIK82, were 18, 31, and 17 cM for milk yield; 25, 17, and 9 cM for fat yield; and 16, 30, and 17 cM for protein yield in the 3 respective parities. Because the data for the parameter estimation well represented the current population of active German Holstein bulls, the QTL parameter estimates have been used in routine marker-assisted genetic evaluation for German Holsteins.

(**Key words:** quantitative trait locus, variance component, milk production trait, chromosome 6)

Abbreviation key: AIA = average information algorithm, **BTA6** = *Bos taurus* autosome 6, **DYD** = daughter yield deviations, **EDC** = effective daughter contribution, **IBD** = identical by descent, **MAS** = marker-assisted selection, **MCMC** = Monte Carlo Markov Chain, **RRTDM** = random regression test day model, **VIT** = genetic evaluation center.

356

INTRODUCTION

The mixed inheritance model with random oligogenic (i.e., originating from QTL) and polygenic effects has been regarded as a promising statistical description of genetic variation of quantitative traits (George et al., 2000; Meuwissen and Goddard, 2001; Liu et al., 2002; de Koning et al., 2003; Freyer et al., 2003). The possibility of incorporating information on (co)variances between individuals on average genome level (i.e., polygenic) as well as at a specific position within genome (i.e., QTL) directly into the statistical model is especially suited for data with complex multigenerational pedigrees such as in dairy cattle, whereas statistical models with fixed QTL effects are less realistic approximations of the underlying modes of inheritance for outbred populations. However, proper separation between variance components corresponding to those 2 random effects is among the most important challenges in the statistical analysis of data under the mixed inheritance model. The separation of 2 random effects is done through the differentiation of genetic similarity between individuals into 2 parts: 1) at a site-specific level, related to the oligogenic component of genetic variance due to similarity at a given position of the genome, and expressed by the proportion of identity by descent (**IBD**) between individuals, and 2) at a genome-average level, related to the polygenic component of genetic variance and expressed by the standard numerator relationship matrix (Henderson, 1963). Because the vast majority of QTL have not yet been physically mapped on the bovine genome, indirect information coming from molecular markers linked to the QTL must be used to trace their inheritance patterns and corresponding similarities between individuals.

The mixed inheritance models with random QTL effects can be used in preselection of young bulls for progeny testing programs, selection of young animals without or with little progeny information, or selection of heifers as candidates for dams of bulls, following the marker-assisted selection (**MAS**) scheme. For routine application of MAS in dairy cattle populations, knowledge of parameters of polygenic and QTL components

Received April 6, 2004.

Accepted September 15, 2004.

Corresponding author: Joanna Szyda; e-mail: Joanna.Szyda@ vit.de.

of the genetic variance is the prerequisite. Such models have been applied to data from the US Holstein population by Zhang et al. (1998) and from a selected subset of the German Holstein population by Freyer et al., in univariate (2002) and multivariate (2003) frameworks. Therefore, the main objective of the current study was to estimate genetic parameters of the mixed inheritance model for the German Holstein dairy population using all available marker and phenotypic information. Moreover, technical aspects related the estimation of model parameters for a large data set from routine genotype recording were discussed.

MATERIALS AND METHODS

The MAS System for German Holsteins

For German Holstein cattle, MAS was initiated in 1995. In the first phase of the project, a genome scan of 339 markers was conducted using a granddaughter design, and applied to 1602 bulls from 19 grand-sire families. The marker and phenotypic data were used to localize the most significant QTL to be followed in upcoming generations. Regions on 3 chromosomes were chosen to represent QTL for milk production traits. Consequently, the second phase of MAS covered further genotyping of all young bulls and selected candidates for bull dams for 13 markers mapped to the 3 chromosomes. Meanwhile, genotypes were updated for older animals to achieve higher completeness of marker information across generations. Up to February 2003, genotype information was available for 10,152 individuals. The implementation of MAS in Germany involved (1)breeding organizations providing tissues for genotyping, (2) universities conducting research, (3) laboratories providing genotypes, and (4) a genetic evaluation center developing and implementing actual MAS evaluations. Cooperation was coordinated by the German Cattle Breeders Federation and by the genetic evaluation center (VIT). A national system for management of genotypic data, QTL mapping, and MAS was set up at VIT (for details see Szyda et al., 2002). Its main components comprised the (1) management of marker data, (2) data analysis, and (3) interchange of data and results between genotyping laboratories, breeding organizations, and VIT.

The analyzed material was a subset of bulls from the active population of the German Holstein and reflects the contents of the marker database in February 2003. The data consisted of 4500 genotyped animals born between 1985 and 2000. With nongenotyped parents of the genotyped animals included, the full pedigree contained 7841 animals.

Genotypic information. Based on the analysis of the genome scan data, a region representing QTL for

milk, fat, and protein yields was identified on chromosome 6 (**BTA6**). The 6 markers mapped to this region covered approximately 70 cM with an average intermarker distance of 14 cM (Thomsen et al., 2000). Table 1 gives distances between the markers and their basic characteristics based on all 10,152 animals.

Phenotypic information. The phenotypic records (Table 2) were daughter yield deviations (**DYD**) obtained from the official release of the May 2003 genetic evaluation, and were based on random-regression and a test-day model (**RRTDM**; Liu et al., 2001). The DYD of the first 3 parities for milk, protein, and fat yields were selected. The reliability associated with DYD was expressed by effective daughter contributions (**EDC**), approximated following Liu et al. (2004). The DYD in the genotyped population were slightly higher than the means of the whole population, indicating that, in general, genetically superior individuals were chosen for genotyping in the early stage of the MAS program.

Statistical Models

A fixed QTL model (model 1). The first model applied to the data included a fixed QTL effect specific to each grandsire, expressed as a difference between a heterozygous (say, Qq) and a homozygous (say, qq) genotype:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_{s}\mathbf{q}_{s} + \mathbf{Z}\mathbf{a} + \mathbf{e}$$
 [1]

where **y** is a vector of DYD for bulls, expressed on a 305-d lactation basis, β is a vector of fixed effects for year of birth, \mathbf{q}_s is a vector of fixed QTL effects for all grand-sires, **a** is a vector of random polygenic effects assuming $\mathbf{a} \sim N(0, \mathbf{G}_a \sigma_a^2)$ with \mathbf{G}_a representing polygenic relationships among individuals and σ_a^2 being a component of the total additive genetic variance attributed to polygenes, **e** is a vector of random errors assuming $\mathbf{e} \sim N(0, \mathbf{D}\sigma_e^2)$ with σ_e^2 denoting the error variance and matrix **D** containing a function of EDC (which is specified later in the text) on the diagonal, and **X**, \mathbf{Z}_s , and **Z** are corresponding design matrices. The (co)variance structure corresponding to model [1] is given by $\begin{bmatrix} \mathbf{a} \end{bmatrix} \begin{bmatrix} \mathbf{G}_a \sigma_a^2 & 0 \end{bmatrix}$

$$\operatorname{var}\begin{bmatrix}\mathbf{a}\\\mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{G}_a\sigma_a^2 & \mathbf{0}\\\mathbf{0} & \mathbf{D}\sigma_e^2\end{bmatrix}.$$

A random QTL model (model 2). The model included a random QTL effect specific to each animal:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_q \mathbf{q} + \mathbf{Z} \mathbf{a} + \mathbf{e}$$
 [2]

where **q** is a vector of random QTL effects; assuming $\mathbf{q} \sim N(0, \mathbf{G}_q \sigma_q^2)$ with \mathbf{G}_q representing relationship among individuals at a QTL position expressed by proportions

(2000).							
Marker	Map location No. of (cM) alleles		Genotyped individuals (no.) PIC ¹		Heterozygous individuals (%)		
DIK82	00.0	11	8184	0.766	40.2		
IL97	22.1	4	4882	0.374	24.9		
FBN14	30.6	6	5112	0.426	22.0		
CSN3	38.6	4	5218	0.367	20.6		
BP7	49.3	9	7138	0.645	35.1		
BMC4203	69.8	10	6078	0.602	33.5		

Table 1. Marker information in the genotyped population. Map locations estimated by Thomsen et al. (2000).

¹PIC = Polymorphic information content.

of alleles being IBD and σ_q^2 is the component of the total additive genetic variance due to the QTL; **X**, **Z**_q, **Z** are corresponding design matrices, and the other effects are as specified above. Covariance structure between the random effects of model [2] is defined as:

$$\operatorname{var}\begin{bmatrix} \mathbf{q} \\ \mathbf{a} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_q \sigma_q^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_a \sigma_a^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D} \sigma_e^2 \end{bmatrix}.$$

Derivation of DYD. By definition, DYD of bulls are averages of daughters' performance adjusted for all fixed and nongenetic random effects of the daughters and genetic effects of their dams. The DYD are deregressed and are a more independent measure of phenotypic performance of bull's daughters than are EBV (VanRaden and Wiggans, 1991). The DYD were derived following Liu et al. (2004), by absorbing the genetic effects of daughters by using their records adjusted for all other effects and for the EBV of the dams. For the RRTDM, the DYD were modeled with the same mathematical function as additive genetic effects and expressed in regression coefficients. The vector **y** contained DYD on a 305-d lactation basis, which were computed by summing up individual DYD of all DIM. **Derivation of weights on DYD.** Reliability of a bull due to daughter performance, denoted as R_b^2 , was obtained from the approximation procedure of Liu et al. (2004). The calculation of R_b^2 under an RRTDM accounted for numbers of daughters, numbers of lactations per daughter, numbers of tests per lactation, and the reliability of the EBV of the mates of the bull. The least squares part, denoted ϕ_b , of the left-hand side in the mixed model equation system corresponding to the bull was derived as:

$$\phi_b = \frac{R_b^2}{1 - R_b^2} \sigma_a^{-2} = \frac{kR_b^2}{1 - R_b^2} \sigma_e^{-2} = n_a \sigma_e^{-2}$$
[3]

where n_a is the weight on DYD of the bull and represents the diagonal element in **D** for the bull, and $k = \sigma_e^2/\sigma_a^2$ is the ratio of residual to genetic variance.

Estimation of IBD Proportion

Estimation of the IBD matrix of the QTL was based on the reversible jump Monte Carlo Markov Chain (**MCMC**; Green, 1995) algorithm along the whole marked chromosome region at a step-size of 1 cM. Because no formal monitoring of the MCMC algorithm

	••			
Parity	$Statistic^1$	Milk yield	Fat yield	Protein yield
1	$\overline{x} s_x s_x min/max}$	$\begin{array}{c} 659.65 \pm 18.68 \\ 568.90 \\ (-1426.41) \ / \ 2484.82 \end{array}$	$\begin{array}{c} 20.36 \pm 0.69 \\ 21.07 \\ (-45.57) \ / \ 95.96 \end{array}$	$\begin{array}{c} 21.54 \pm 0.56 \\ 17.18 \\ (-45.32) \ / \ 85.71 \end{array}$
2	\overline{x} s_x min/max	$\begin{array}{c} 671.23 \pm 23.19 \\ 625.62 \\ (-1858.44) \ / \ 3044.72 \end{array}$	$\begin{array}{c} 19.61 \pm 0.88 \\ 23.63 \\ (-66.48) \ / \ 114.10 \end{array}$	$\begin{array}{c} 21.70 \pm 0.73 \\ 19.74 \\ (-57.10) \ / \ 89.50 \end{array}$
3	\overline{x}_{s_x}	$\begin{array}{c} 631.87 \pm 27.65 \\ 638.97 \\ (-1955.58) \ / \ 3079.23 \end{array}$	$\begin{array}{c} 19.31 \pm 1.07 \\ 24.74 \\ (-68.78) \ / \ 118.72 \end{array}$	$\begin{array}{c} 19.12 \pm 0.85 \\ 19.71 \\ (-64.08) / 96.77 \end{array}$

Table 2. Phenotypic information in the analyzed sample.

¹Statistics: \overline{x} = average yield with 95% confidence interval, s_x = standard deviation, ^{min}/_{max} denotes border values.

convergence was performed, a long burn-in phase of 1000 rounds was used to allow sampling from the proper marginal distributions. Also, a long spacing of 100 rounds was chosen between scored IBD matrix realizations to account for the fact that mixing of the parameter values might be poor for a large multigenerational pedigree with many closely linked markers. The mode of 91 realizations was considered as the final estimate of the IBD matrix.

Estimation of Model Parameters

The restricted maximum likelihood (REML; Patterson and Thompson, 1971) approach was applied for estimating parameters of models [1] and [2]. Following Gilmour et al. (1995), average information algorithm (AIA) was used to maximize REML likelihood. The likelihood functions of models [1] and [2] were respectively defined as:

$$\begin{aligned} \ln L &= -\frac{n-r}{2} \ln(2\pi) - \frac{1}{2} (\ln|\mathbf{R}| + \ln|\mathbf{G}_a \sigma_a^2| - \ln|\mathbf{C}| \\ &+ \mathbf{y'} \mathbf{R}^{-1} \mathbf{y} - \mathbf{y'} \mathbf{R}^{-1} \mathbf{X} \beta - \mathbf{y'} \mathbf{R}^{-1} \mathbf{Z}_s \mathbf{q}_s - \mathbf{y} \mathbf{R}^{-1} \mathbf{Z} \mathbf{a}) \end{aligned}$$

and

$$\ln L = -\frac{n-r}{2}\ln(2\pi) - \frac{1}{2}(\ln|\mathbf{R}| + \ln|\mathbf{G}_q \sigma_q^2| + \ln|\mathbf{G}_a \sigma_a^2| - \ln|\mathbf{C}| + \mathbf{y'}\mathbf{R}^{-1}\mathbf{y} - \mathbf{y'}\mathbf{R}^{-1}\mathbf{X}\boldsymbol{\beta}$$

$$- \mathbf{y'}\mathbf{R}^{-1}\mathbf{Z}_q\mathbf{q} - \mathbf{y}\mathbf{R}^{-1}\mathbf{Z}\mathbf{a})$$

$$(4)$$

where n is the number of phenotypic records, r is the rank of the design matrix for fixed effects, $\mathbf{R} = \mathbf{D}\sigma_e^2$ is residual (co)variance matrix, C is coefficient matrix of the mixed model equations of model [2]. The estimated effects comprised $[\beta \quad \mathbf{q} \quad \mathbf{a}]$, and estimated variances are σ_a^2 , σ_e^2 for model [1] and, additionally, σ_a^2 for model [2]. For all the traits considered, the most probable position of the QTL was estimated using model [1] based on a likelihood profile constructed every 1 cM along the marked region of BTA6. Parameters of model [2] were estimated for the most probable QTL location estimated by model [1], except for second lactation fat yield, for which the parameters were estimated along the whole marked chromosome region every 1 cM. Confidence intervals for QTL position, $\hat{\sigma}_a^2$ and $\hat{\sigma}_a^2$ were obtained based on a normal approximation of the asymptotic distribution of maximum likelihood estimates:

$$(\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} < \theta < \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}})$$

where $\hat{\theta}$ is the estimate of QTL position, σ_a^2 , or σ_q^2 , α is the probability of type I error, $z_{\alpha/2}$ is the critical value corresponding to $\frac{\alpha}{2}$ type I error rate based on the standard normal distribution, and $\sigma_{\hat{\theta}}$ is the standard of $\hat{\theta}$ approximated deviation bv $\sqrt{\frac{\varDelta^2}{\ln L(\hat{\theta}-\varDelta) + \ln L(\hat{\theta}+\varDelta) - 2 \ln L(\hat{\theta})}} \text{ with } \varDelta \text{ set to } 1$

cM for QTL position, 0.3 for σ_a^2 , and 0.04 for σ_a^2 (Meyer and Hill, 1992).

Hypotheses Testing

The likelihood ratio test statistic was used as a testing criterion:

$$\lambda = -2 \ln \frac{L(M_0)}{L(M_1)}$$
^[5]

where $L(M_1)$ and $L(M_0)$ are the maximum values of likelihood functions underlying the unrestricted model given above and a more parsimonious model without QTL effects. Corresponding null and alternative hypotheses were:

H₁: $\mathbf{q}_s \neq 0$ and H₀: $\mathbf{q}_s = 0$ for model [1], with the

asymptotic null distribution of λ being χ^2_{1df} . H₁: $\sigma_q^2 > 0$ and H₀: $\sigma_q^2 = 0$ for model [2]. Because σ_q^2 was constrained to positive values, the asymptotic null distribution of λ followed a 50:50 mixture of 0 and χ^2_{1df} [for theoretical derivation, see Self and Liang (1987); for empirical results based on livestock data structure, see George et al. (2000)].

Implementation

The routine implementation of the estimation procedure involved 5 steps:

- 1. Checking for data consistency by database routine programs.
- 2. Data preparation, check, and preliminary analysis by a UNIX shell script, SAS (SAS Institute, 1999), and Fortran programs.
- 3. Estimation of IBD matrices by the package LOKI (Heath, 1997).
- 4. Inverting the IBD matrices using sparse matrix routines implemented in the FSPAK package (Misztal and Perez-Enciso, 1998).
- 5. Estimation of model parameters by the ASREML package (Gilmour et al., 1995) and prediction of estimated polygenic and QTL effects.

Table 3. Estimates of QTL positions with 95% confidence intervals based on fixed QTL model and estimates of variance components underlying a random QTL model.¹

Trait	Parity	Positions from DIK82 (cM)	Significance	$\hat{\sigma}_a^2/\hat{\sigma}_e^2$	$\hat{\sigma}_q^2 / \hat{\sigma}_e^2$	$\hat{\sigma}_q^2 / \hat{\sigma}_a^2$	$\hat{\sigma}_q^2$ in kg yield
Milk yield	1	18 ± 10	0.01129	4.58	0.20	0.04	10,111
	2	31 ± 04	0.03682	14.44	0.44	0.03	9959
	3	17 ± 10	0.00041	24.28	1.80	0.07	25,815
Fat yield	1	25 ± 09	0.00801	3.92	0.23	0.06	18.7
	2	17 ± 09	0.00019	9.96	0.82	0.08	38.4
	3	9 ± 12	0.00007	9.47	1.35	0.14	72.2
Protein yield	1	16 ± 11	0.02890	3.79	0.15	0.04	7.8
	2	30 ± 08	0.01030	12.61	0.53	0.04	12.9
	3	17 ± 10	0.00006	18.17	2.05	0.11	35.6

¹Significance relates to a random QTL model and is expressed as the nominal type I error rate, $\hat{\sigma}_e^2$ = residual variance component, $\hat{\sigma}_a^2$ = additive polygenic variance component, $\hat{\sigma}_a^2$ = QTL variance component.

RESULTS

QTL Position

Due to its relatively low computational requirements, model [1] was used to estimate location of QTL along BTA6. The chromosome-wise significance levels, expressed as nominal significance penalized for the number of marker intervals following Bonferroni correction, were very high (P < 0.00001) for all 9 parity × trait combinations. The estimated positions of particular QTL and their 95% confidence intervals obtained following Meyer and Hill (1992) are in Table 3. All the positions were expressed as distances from the leftmost marker DIK82. In summary, the estimated QTL positions for all 3 yield traits in first and third parities and fat yield in second parity were located within the leftmost interval flanked by DIK82 and IL97. For milk and protein yields in second parity, the estimated QTL positions were located within the next interval flanked by IL97 and FBN14. Corresponding likelihood profiles, shown in Figures 1, 2, and 3 for milk, fat, and protein yields, respectively, indicated high levels of significance for the first 2 marker intervals and high degrees of similarity between traits and parities, with the exception of milk and protein yields in second parity.

Variance Components

Model [2] was applied to the data to partition total variance into QTL, polygenic, and residual components. Because estimation of these parameters required an iterative solving of 15,706 equations at each QTL position, it was computationally intensive and time consuming. Thus, variance components were estimated only for those chromosome positions that exhibited the highest probability for the QTL, based on results from model [1]. The resulting estimates are in Table 3. Among 3 parities, the ratio of QTL to polygenic variance was greatest for the third parity: 0.07 for milk yield, 0.14 for fat yield, and 0.11 for protein yield. Corresponding QTL variance estimates for first and second lactations were similar and varied between $0.03\hat{\sigma}_a^2$ for milk yield in second parity and $0.08\hat{\sigma}_a^2$ for fat yield in second parity. Despite the relatively low estimates of QTL variance, the QTL term was significant for all parities and all yield traits.

To examine how inaccurate QTL position estimates affected estimates of QTL variance components, parameters of model [2] were estimated for second-parity fat yield along the entire 70 cM chromosome region. Results presented in Figure 4 indicated that with increasing distance from the more probable QTL position, the QTL variance diminished and a significant portion of QTL variance was absorbed into the polygenic component, whereas the residual component remained relatively stable along the chromosome. This effect was summarized by a linear regression of $\hat{\sigma}_e^2$ and $\hat{\sigma}_a^2$ on distance from the most probable QTL position (in cM), which yield regression coefficients of 0.04 ± 0.007 and 0.07 ± 0.007, respectively.

Likelihood of the Random QTL Model

As previously mentioned, a likelihood profile of the random QTL model [2] was constructed for second-lactation fat yield. It can be seen in Figure 5 that the likelihood profile curve did not have a smooth shape along the chromosome, indicating problems with finding global maxima of the multidimensional likelihood surface. Nevertheless, the test statistic was highest throughout the first and a part of the second marker interval (i.e., BTA6 region showing the highest probability of QTL location based on model [1]), followed by a clear drop thereafter.

Because the likelihood profile in Figure 5 indicated some difficulties in maximizing the likelihood function



Figure 1. Bos taurus autosome 6 test statistic (λ) profiles for milk yield of first (\Box), second (×), and third (\bullet) parities based on a fixed QTL model.

underlying model [2], it was important to know how much information, i.e., curvature, was available for estimating the variance components. For that purpose, a 2-D surface was constructed over the grid of σ_a^2 and σ_q^2 around the $\hat{\sigma}_a^2$ and $\hat{\sigma}_q^2$ values for second-lactation fat yield, while the other parameters of model [2] were estimated using the AIA as implemented in the AS-REML package. As shown in Figure 6, in the proximity of the $\hat{\sigma}_a^2$ and $\hat{\sigma}_q^2$ values, the likelihood surface remained rather flat, and equal likelihood values were obtained for 34 different combinations of σ_a^2 and σ_q^2 . The 95% confidence intervals approximated for the 2 variance estimates were (5.8 and 14.2) for σ_a^2 and (0.27 and 1.37) for σ_q^2 .

DISCUSSION

QTL for milk production traits on BTA6. Evidence for QTL affecting milk production traits in the neighborhood of the CN gene cluster has been reported by many authors and found across various cattle breeds and populations including Finnish Ayrshire (Velmala et al., 1999; de Koning et al., 2001; Viitala et al., 2003), Canadian Holstein (Nadesalingam et al., 2001), Dutch Holstein (Spelman et al., 1996), German Holstein

Holstein (Ron et al., 2001), US Holstein (Georges et al. 1995; Zhang et al., 1998; Rodriguez-Zas et al., 2002), and Norwegian Dairy Cattle (Olsen et al., 2002). Also in our study, the location of QTL in the region marked by DIK82-IL97-FBN14 corresponded with QTL positions estimated by other authors. Unlike many other QTL, the chromosomal location of these QTL remained surprisingly consistent among various populations despite different materials and methods used in the analyses. This result suggested that the QTL on BTA6 is one of the most important loci to be considered in MAS. Recently, Khatkar et al. (2004) performed a meta-analvsis of available QTL mapping results, and concluded that 2 QTL for milk yield and a single QTL for fat or protein yield were segregating in the CN region on BTA6. Variance component estimates. The main goal of

(Kuhn et al., 1999; Freyer et al., 2002, 2003), Israeli

Variance component estimates. The main goal of our study was to estimate genetic parameters underlying the random QTL model used for MAS, i.e., additive polygenic and QTL variances for the German Holstein population. Quantitative trait loci variance was already estimated for by Zhang et al. (1998) for US Holsteins and for the German Holsteins by Freyer et al. (2002, 2003). However, an important advantage of the current study was the large size of the data sets: (1) more than



Figure 2. Bos taurus autosome 6 test statistic (λ) profiles for fat yield of first (\Box), second (×), and third (\bullet) parities based on a fixed QTL model.

2 generations of genotyped individuals were available, (2) the young genotyped bulls represented almost all of the young bulls tested in German Holstein population (although not all of those bulls have DYD available yet). Those features mean that the data well represent the current German Holstein population. The variance components for milk and fat yields estimated by Zhang et al. (1998) based on the sample of 1794 sons of 14 sires remained in close agreement with our variance estimates. In general, QTL variances explained by the markers on BTA6 represented only a small proportion (<10%) of the additive genetic variance of milk production traits in both populations. In contrast, the study by Freyer et al. (2002, 2003) based on data from 562 sons of 5 sires gave much higher estimates of QTL variance, especially for fat and protein yields, varying between 23 and 50% of the total genetic variance depending on the model and trait studied. We noticed that there was a significant change in QTL variance estimates in the 2 investigations by Freyer et al. using the same data set. The QTL variance estimate for milk yield increased from 8 to 16% of variance of DYD, but significantly decreased from 50 to 20% for fat yield, and from 50 to 28% for protein yield between the 2 studies, even though the same univariate QTL model was used. The differences in QTL variance estimates between our

study and both studies of Freyer et al. were due to the following reasons: 1) The trait used by Freyer et al. was a yield trait on a combined lactation basis, whereas we analyzed the yield traits in 3 lactations separately. Because the 3 lactations were highly positively correlated, higher QTL component can be expected for combined lactation than for single lactations; 2) The dependent variable in our study was DYD, whereas the dependent variable in Frever et al. (2002) was EBV. They noted that DYD had higher proportion of nonQTL variance and lower QTL component than did EBV. In our opinion, DYD is an unregressed measure of daughter performance and should be preferred to EBV, despite the fact that the use of DYD led to a lower QTL variance component; 3) The DYD and EBV used by Freyer et al. (2003) corresponded to a different genetic evaluation model (Reents et al., 1995) and a different DYD calculation method; 4) The numbers of genotyped animals and animals in pedigree were much higher in our study than in the studies by Frever et al. (2003) and the data in their studies might not have been representative of the overall population. Sampling variances of QTL parameter estimates were significantly smaller in our studies than in Freyer et al. (2003); and 5) We had multiple generations of genotyped animals, in comparison with only 2 generations in both studies of Freyer



Figure 3. *Bos taurus* autosome 6 test statistic (λ) profiles for protein yield of first (\Box), second (×), and third (\bullet) parities based on a fixed QTL model.

et al., (ancestors of the 5 grandsires were ignored in their studies). However, adding more genotyped ancestors can significantly improve the tracing of origin of marker alleles and IBD calculation for QTL.

Considering the estimates of variance components along the marked region of BTA6 in Figure 4, we concluded that model [2] was well able to differentiate between the QTL and polygenic variances, because with increasing distance from the most probable QTL position, the proportion of QTL component in the total genetic variance diminished. In contrast, the residual component was not confounded with genetic components, as its estimates remained stable along the chromosome.

Compared with yield deviations of cow, DYD of bulls had a much smaller proportion of residual variance, due to higher reliabilities associated with the DYD. Error variance of the original polygenic RRTDM could not be accurately estimated based on the DYD in the parameter estimation for QTL variance. The difference between the estimated polygenic variance and original RRTDM genetic variance amounted to 5 to 10%, which can be explained by different samples of selected animals in both parameter estimations. The RRTMD was applied to cows with test-day records (Liu et al., 2001) vs. genotyped bulls with DYD for this study. In addition different pedigree structures were considered: sire and dam of animals for RRTDM (Liu et al., 2001) vs. male ancestors of bulls only in this study. Another aspect of using DYD as trait representation was weighting of residual variances in the parameter estimation. In our case, initial attempts to use weights based on EDC resulted in convergence problems by the AIA. Because DYD of all bulls had comparable and high reliabilities in current study, the fact that DYD were assigned to equal weights should not affect much the ratios of QTL to polygenic variances.

Likelihood surface. Comparison of likelihood profile shapes based on model [1] and model [2] (Figure 5) showed good agreement, because both models indicated the same 2 intervals as the most probable QTL location. In terms of QTL position, Zhang et al. (1998) obtained similar estimates using both least squares and variance component models as well. In addition, test statistics profiles for the 2 methods shown by Freyer et al. (2002) remained in good agreement. Figure 5 revealed an unfavorable feature of the likelihood profile resulting from model [2], as the likelihood values were not smooth along the analyzed region, indicating problems in convergence by the AIA in the neighborhood of maximum of the likelihood function. As shown in Figure 6, for the axes defined by variance components, the likelihood



Figure 4. Partitioning of the total variance for fat yield of second parity between residual (\blacksquare), additive polygenic (\blacksquare), and QTL (\blacksquare), based on a random QTL model. H₀ represents partitioning under the null hypothesis of no QTL. The QTL position estimated under a fixed QTL model is marked with an arrow.

surface was rather flat, not only at the very maximum but also in its proximity, which most likely caused the observed optimization problems.

Estimation of IBD proportions. As already pointed out by Grignola et al. (1996), the estimation of IBD relationships for large, multigenerational pedigrees was computationally very demanding. The first method for calculating IBD proportions for such pedigrees was proposed by Fernando and Grossman (1989) for a single marker scenario. However, using one marker at a time is not well suited for analyzing data from actual dairy cattle populations with a complex pedigree structure, because there are often uninformative or missing marker genotypes. Since the above-mentioned seminal study, a number of methods for calculating IBD coefficients have been developed, which can be mainly classified into deterministic and MCMC-based approaches, based on the estimation method, or into marker interval and multiple-marker based approaches, according to the use of marker information. The main advantage of the deterministic approach lies in its speed of computation; however, no method exists at present that is able

to use all available marker haplotype and pedigree information. Pong-Wong et al. (2001) and Liu et al. (2002) proposed methods that can account for a marker bracket. The multiple marker approach was developed by Almasy and Blangero (1998), but it was only applicable to special types of relationships between individuals. Meuwissen and Goddard (2001) presented a multimarker approach capable of tracing historical relationships not contained in the recorded part of a pedigree, but available relationship information was not explicitly used in the method. Recently, Lund et al. (2003) combined the information on historical and observed relationships following Meuwissen and Goddard (2001) and Wang et al. (1995), respectively for the 2 components. An overview of deterministic IBD estimation methods was given by George et al. (2000). The MCMC-based approaches were more flexible in terms of using marker and complex pedigree information, but they were time consuming, which limited their application to the analysis of large data sets. Additionally, issues related to monitoring convergence and irreducibility of the algorithms has not been well defined yet.



Figure 5. *Bos taurus* autosome 6 test statistic (λ) profiles for fat yield of second parity based on a fixed QTL model (\Box) and a random QTL model (\bullet).

The approach of Heath (1997) was among the first applications of MCMC to the analysis of large complex pedigrees, followed for example, by Xu and Gessler (1998) and Perez-Enciso et al. (2000). Grignola et al. (1996) proposed an intermediate approach, in which MCMC was used for the estimation of IBD among the genotyped part of the parental pedigree and then a deterministic approach to obtain IBD coefficient was used for nonparents.

The approach of Heath (1997) as implemented in the LOKI package, was used for the estimation of IBD proportions in the current study, which were then used in the parameter estimation under model [2] via the ASREML package. The simulation results by George et al. (2000) demonstrated that by using the aforementioned implementation, both variance components and QTL positions can be accurately estimated even in case of missing marker data. An application of this approach to livestock data was presented by de Koning et al. (2003). The results of simulation studies carried by Sørensen et al. (2002) showed that the algorithm provides accurate estimates of IBD proportions.

Using genomic information. Although the original plan in Germany was to acquire not only male but also female genotypes in the development of routine genotyping and MAS, data available for the current analysis consisted of genotypes of bulls only. Because of the large number of genotyped bulls, genotypes of some dams and maternal grandsires could have been reconstructed. Recently, Bolard and Boichard (2002) showed how the information on maternal grand-sire genotypes and, consequently, on QTL transmissions can be incorporated in the QTL mapping, albeit in the fixed QTL framework. The marker information on female animals should be, whenever possible, incorporated in both parameter estimation and routine genetic evaluation.

CONCLUSIONS

Based on a random QTL model, QTL variance for milk production traits was estimated using a large marker and phenotypic data set for the German Holstein population. The estimated QTL variances accounted for 4 to 14% of polygenic variances for milk, fat, and protein yields. These QTL parameter estimates have been used in routine marker-assisted genetic evaluations of production traits in Germany, based on DYD of bulls and yield deviations of cows. Our experience showed that the MCMC-based algorithm for estimating IBD proportions was able to use the most of available pedigree and marker information. On the other hand, the MCMC-based methods seemed to be rather time



Figure 6. Likelihood surface over σ_a^2 and σ_a^2 around the $\hat{\sigma}_a^2$ and $\hat{\sigma}_a^2$ for fat yield of second parity based on a random QTL model.

consuming for routine MAS in large data sets currently. Therefore, efficient and accurate deterministic approaches would be more appealing for routine MAS application. The random QTL model with both random QTL and polygenic effects accounted for relationships among animals on the genome level as well as at a specific genome position, which makes it applicable for routine genetic evaluation of dairy cattle. The DYD were used as dependent variables and associated EDC as a function for weighting factors in the marker-assisted genetic evaluation system. Our current MAS genetic evaluation system analyzes data from one lactation at a time only, although routine genetic evaluation of production traits has been conducted with a multiplelactation RRTDM (Liu et al., 2004). A logical further development of the random QTL model would be to extend it from the current single-lactation model to a multiple-lactation model to account for the correlations between lactations.

REFERENCES

- Almasy, L., and J. Blangero. 1998. Multipoint quantitative-trait linkage analysis in general pedigrees. Am. J. Hum. Genet. 62:1198–1211.
- Bolard, M., and D. Boichard. 2002. Use of maternal information for QTL detection in a (grand)daughter design. Genet. Sel. Evol. 34:335–352.
- de Koning, D. J., R. Pong-Wong, L. Varona, G. J. Evans, E. Giuffra, A. Sanchez, G. Plastow, J. L. Noguera, L. Andersson, and C. S. Haley. 2003. Full pedigree quantitative trait locus analysis in commercial pigs using variance components. J. Anim. Sci. 81:2155-2163.
- de Koning, D. J., N. F. Schulman, K. Elo, S. Moisio, R. Kinos, J. Vilkki, and A. Maki-Tanila. 2001. Mapping of multiple quantitative trait loci by simple regression in half-sib designs. J. Anim. Sci. 79:616–622.
- Fernando, R. L., and M. Grossman. 1989. Marker-assisted selection using best linear unbiased prediction. Genet. Sel. Evol. 21:467–477.
- Freyer, G., C. Kühn, R. Weikard, Q. Zhang, M. Mayer, and I. Hoeschele. 2002. Multiple QTL on chromosome six in dairy cattle affecting yield and content traits. J. Anim. Breed. Genet. 119:69-82.
- Freyer, G., P. Sorensen, C. Kuhn, R. Weikard, and I. Hoeschele. 2003. Search for pleiotropic QTL on chromosome BTA6 affecting yield traits of milk production. J. Dairy Sci. 86:999–1008.

- George, A. W., P. M. Visscher, and C. S. Haley. 2000. Mapping quantitative trait loci in complex pedigrees: A two-step variance component approach. Genetics 156:2081–2092.
- Georges, M., D. Nielsen, M. Mackinnon, A. Mishra, R. Okimoto, A. T. Pasquino, L. S. Sargeant, A. Sorensen, M. R. Steele, X. Zhao, J. E. Womack, and I. Hoeschele. 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139:907–920.
- Gilmour, A. R., R. Thompson, and B. R. Cullis. 1995. AI, an efficient algorithm for REML estimation in linear mixed models. J. Dairy Sci. 51:1440–1450.
- Green, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82:711– 732.
- Grignola, F. E., I. Hoeschele, and B. Tier. 1996. Mapping quantitative trait loci via residual maximum likelihood. I. Methodology. Genet. Sel. Evol. 28:479–490.
- Heath, S. C. 1997. Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. Am. J. Hum. Genet. 61:748–760.
- Henderson, C. R. 1963. Page 141 in Statistical Genetics and Plant Breeding. W. D. Hanson and H. F. Robinson, ed. Natl. Acad. Sci., National Research Council, Washington, DC.
- Khatkar, M. S., P. C. Thomson, I. Tammen, and H. W. Raadsma. 2004. Quantitative trait loci mapping in dairy cattle: Review and meta-analysis. Genet. Sel. Evol. 36:163–190.
- Kuhn, C., G. Freyer, R. Weikard, T. Goldammer, and M. Schwerin. 1999. Detection of QTL for milk production traits in cattle by application of a specifically developed marker map of BTA6. Anim. Genet. 30:333–340.
- Liu, Y., G. B. Jansen, and C. Y. Lin. 2002. The covariance between relatives conditional on genetic markers. Genet. Sel. Evol. 34:657-678.
- Liu, Z., F. Reinhardt, A. Bünger, and R. Reents. 2004. Derivation and calculation of approximate reliabilities and daughter yield deviations of a random regression test-day model for genetic evaluation of dairy cattle. J. Dairy Sci. 87:1896–1907.
- Liu, Z., F. Reinhardt, and R. Reents. 2001. Application of a random regression model to genetic evaluations test-day yields and somatic cell scores in dairy cattle. Interbull Bull. 27:159–166.
- Lund, M. S., P. Sørensen, B. Guldbrandtsen, and D. A. Sorensen. 2003. Multitrait fine mapping of quantitative trait loci using combined linkage disequilibria and linkage analysis. Genetics 163:405-410.
- Meuwissen, T. H. E., and M. E. Goddard. 2001. Prediction of identity by descent probabilities from marker-haplotypes. Genet. Sel. Evol. 33:605–634.
- Meyer, K., and W. G. Hill. 1992. Approximation of sampling variances and confidence intervals for maximum likelihood estimates of variance components. J. Anim. Breed. Genet. 109:264–280.
- Misztal, I., and M. Perez-Enciso. 1998. FSPAK90 A Fortran 90 interface to sparse-matrix package FSPAK with dynamic memory allocation and sparse matrix structure. Proc. 6th World Congr. Genet. Appl. Livest. Prod., Armidale, Australia XXVII:467-468.
- Nadesalingam, J., Y. Plante, and J. P. Gibson. 2001. Detection of QTL for milk production on chromosomes 1 and 6 of Holstein cattle. Mamm. Genome 12:27-31.
- Olsen, H. G., L. Gomez-Raya, D. I. Våge, I. Olsaker, H. Klungland, M. Svendsen, T. Ådnøy, A. Sabry, G. Klemetsdal, N. Schulman, W. Krämer, G. Thaller, K. Rønningen, and S. Lien. 2002. A genome scan for quantitative trait loci affecting milk production in Norwegian Dairy Cattle. J. Dairy Sci. 85:3124–3130.
- Patterson, H. D., and R. Thompson. 1971. Recovery of interblock information when block sizes are unequal. Biometrika 58:545– 554.

- Perez-Enciso, M., L. Varona, and M. F. Rothschild. 2000. Computation of identity by descent probabilities conditional on DNA markers via a Monte Carlo Markov Chain method. Genet. Sel. Evol. 32:467–482.
- Pong-Wong, R., A. W. George, J. A. Woolliams, and C. S. Haley. 2001. A simple and rapid method for calculating identity-by-descent matrices using multiple markers. Genet. Sel. Evol. 33:453–471.
- Reents, R., J. Jamrozik, L. R. Schaeffer, and J. C. M. Dekkers. 1995. Estimation of genetic parameters for test day records of somatic cell score. J. Dairy Sci. 78:2847–2857.
- Rodriguez-Zas, S. L., B. R. Southey, D. W. Heyen, and H. A. Lewin. 2002. Interval and composite interval mapping of somatic cell score, yield, and components of milk in dairy cattle. J. Dairy Sci. 85:2681–2691.
- Ron, M., D. Klinger, E. Feldmesser, E. Seroussi, E. Ezra, and J. I. Weller. 2001. Multiple quantitative trait locus analysis of bovine chromosome 6 in the Israeli Holstein population by a daughter design. Genetics 159:727–735.
- SAS Institute. 1999. SAS User's Guide. Statistics, Version 8.2 ed. SAS Inst., Inc., Cary, NC.
- Self, S. C., and K. Y. Liang. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. J. Am. Stat. Assoc. 82:605–610.
- Sørensen, A. C., R. Pong-Wong, J. J. Windig, and J. A. Woolliams. 2002. Precision of methods for calculating identity-by-descent matrices using multiple markers. Genet. Sel. Evol. 34:557–579.
- Spelman, R. J., W. Coppieters, L. Karim, J. A. 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.
- Szyda, J., Z. Liu, R. Maschka, F. Reinhardt, and R. Reents. 2002. Computer system for routine QTL detection and genetic evaluation under a mixed inheritance model in dairy cattle. Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France, XXXIII:249-250.
- Thomsen, H., N. Reinsch, N. Xu, C. Looft, S. Grupe, C. Kühn, G. A. Brockmann, M. Schwerin, B. Leyhe-Horn, S. Hiendleder, G. Erhardt, I. Medjugorac, I. Russ, M. Förster, B. Brenig, F. Reinhardt, R. Reents, J. Blümel, G. Averdunk, and E. Kalm. 2000. A male bovine linkage map for the ADR granddaughter design. J. Anim. Breed. Genet. 117:289–360.
- VanRaden, P. M., and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. 74:2737-2746.
- Velmala, R. J., H. J. Vilkki, K. T. Elo, D. J. de Koning, and A. V. Maki-Tanila. 1999. A search for quantitative trait loci for milk production traits on chromosome 6 in Finnish Ayrshire cattle. Anim. Genet. 30:136–143.
- Viitala, S. M., N. F. Schulman, D. J. de Koning, K. Elo, R. Kinos, A. Virta, J. Virta, A. Maki-Tanila, and J. H. Vilkki. 2003. Quantitative trait loci affecting milk production traits in Finnish Ayrshire dairy cattle. J. Dairy Sci. 86:1828–1836.
- Wang, T., R. H. Fernando, S. van der Beek, and J. A. M. van Arendonk. 1995. Covariance between relatives for a marked quantitative trait locus. Genet. Sel. Evol. 27:251–274.
- Xu, S., and D. D. G. Gessler. 1998. Multipoint genetic mapping of quantitative trait loci using a variable number of sibs per family. Genet. Res. 71:73–83.
- Zhang, Q., D. Boichard, I. Hoeschele, C. Ernst, A. Eggen, B. Murkve, M. Pfister-Genskow, L. A. Witte, F. E. Grignola, P. Uimari, G. Thaller, and M. D. Bishop. 1998. Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. Genetics 149:1959–1973.