Detection of Quantitative Trait Loci Influencing Dairy Traits Using a Model for Longitudinal Data

S. L. Rodriguez-Zas, B. R. Southey, D. W. Heyen, and H. A. Lewin

Department of Animal Sciences
University of Illinois, Urbana 61801

ABSTRACT

A longitudinal-linkage analysis approach was developed and applied to an outbred population. Nonlinear mixed-effects models were used to describe the lactation patterns and were extended to include marker information following single-marker and interval mapping models. Quantitative trait loci (QTL) affecting the shape and scale of lactation curves for production and health traits in dairy cattle were mapped in three U.S. Holstein families (Dairy Bull DNA Repository families one, four, and five) using the granddaughter design. Information on 81 informative markers on six *Bos taurus* autosomes (BTA) was combined with milk yield, fat, and protein percentage and somatic cell score (SCS) test-day records. Six percent of the single-marker tests surpassed the experiment-wise significance threshold. Marker BL41 on BTA3 was associated with decrease in milk yield during mid-lactation in family one. The scale and shape of the protein percentage lactation curve in family four varied with BMC4203 (BTA6) allele that the son received from the grandsire. Some map locations were associated with variation in the lactation pattern of multiple traits. In family four, the marker HUII177 (BTA3) was associated with changes in the milk yield and protein percentage curves suggesting a QTL with pleiotropic effects or multiple QTL in the region. The interval mapping model uncovered a QTL on BTA7 associated with variation in milk-yield pattern in family four and a QTL on BTA21 affecting SCS in family five. The developed approach can be extended to random regressions, covariance functions, spline, genetic and variance component models. The results from the longitudinal-QTL approach will help to understand the genetic factors acting at different stages of lactation and will assist in positional candidate gene research. Identified positions can be incorporated into marker-assisted selection decisions to alter the persistency and

peak production or the fluctuation of SCS during a lactation.

(Key words: lactation curve, longitudinal data, nonlinear mixed-effects model, quantitative trait loci)

Abbreviation key: BTA = *Bos taurus* autosome, QTL = Quantitative trait loci.

INTRODUCTION

The detection of quantitative trait loci (QTL) that influence production and health traits in dairy cattle is possible by combining phenotypic and genetic records from resource populations, such as the Dairy Bull DNA Repository (Heyen et al., 1999). The granddaughter design utilizes this concept and reflects the structure of the dairy cattle population with genetic records (e.g., markers) on grandsires and their multiple sons and phenotypic records on the granddaughters (Geldermann, 1975; Weller et al., 1990). Previous QTL detection studies analyzed single-measurement phenotypic indicators based on the accumulation of monthly records, such as predicted transmitting ability or daughter yield deviation (e.g., Ashwell et al., 1998; Zhang et al., 1998; Heyen et al., 1999).

Test-day dairy records are examples of longitudinal or repeated measurements because they are obtained on a cow basis at multiple time points (Rodriguez-Zas et al., 2000b). Dairy cattle yield traits typically show a rapid increase to a maximum peak early in lactation (approximately 2 mo after calving) and then slowly decline until the end of the lactation (Gengler, 1996).

Concentration traits such as percentage of milk fat and protein show a rapid decline from the start of lactation until a minimum value early in lactation with a slow increase until the end of lactation. A goal of the dairy industry is to improve the efficiency of milk production by minimizing costs and maximizing returns (Freeze and Richards, 1992). This can be attained by improving milk production persistency since cows with flat curves are less prone to experience physiological stress, less susceptible to metabolic and reproductive disorders, and the smaller production-to-input ratio required permits inclusion of less expensive roughage in their diets (Gengler, 1996).

Received December 19, 2001.

Accepted March 8, 2002.

Corresponding author: S. Rodriguez-Zas; e-mail: rodrzzs@uiuc.edu.

The mapping of QTL that are manifested at different stages of the lactation or that are differentially expressed during lactation is hampered by the use of total lactation measurements or functions thereof (Rodriguez-Zas et al., 2000a). Study of monthly test-day records (instead of cumulative lactation records) could uncover genetic factors solely expressed at specific stages of the lactation, provide more accurate estimates of marker and QTL associations, or increase the statistical power to detect small effects. The reported genetic variability of the parameters on nonlinear models describing the lactation curve (Varona et al. 1998; Rodriguez-Zas et al., 2000a) indicates the possibility of multiple QTL affecting the complete lactation or particular stages. The utilization of models for repeated measurements in QTL mapping studies can increase the chances to detect stage-specific QTL when compared to cumulative lactation record models since the effects of these QTL are likely to be diluted across the lactation. The objectives of this study are to extend the single-marker and interval mapping models to describe repeated measurements using nonlinear functions and to apply this novel approach to the detection of associations between markers and QTL and production and health lactation patterns in U.S. Holstein sire families.

MATERIALS AND METHODS

Data

Test-day milk yield, fat and protein percentage, and SCS records from first-lactation granddaughters of the Dairy Bull DNA Repository (2001) grandsire families one, four, and five (Heyen et al., 1999) through January 2000 were provided by the Animal Improvement Programs Laboratory (ARS, USDA, 2001). Records were deleted if the cow was sick on test-day or if the records were estimated or obtained passed 365 DIM. Although nonlinear models can describe trends based on incomplete lactation records, lactations with less than five measurements were removed from the analysis to favor the convergence of the system of equations. The minimum and maximum number of test-day records and granddaughters are summarized in Table 1 with highest number of records associated with milk yield and lowest with SCS. The number of protein and fat percentage records closely followed the milk yield ones. The average test-day milk yield (kg), fat and protein percentage, and SCS (base two logarithmic transformation) were: 26.44, 3.72, 3.21, 2.42; 27.11, 3.73, 3.31, 2.66; 25.17, 3.77, 3.23, and 2.72 for families one, four and five, respectively. A principal components analysis of the four traits, by DIM, within family indicated that the first three principal components explained jointly between 86.9 and 93.2 percentage of the total variation.

Six *Bos taurus* (BTA) autosomes were selected based on significant associations reported in the literature (e.g., Ashwell et al., 1998; Zhang et al., 1998; Heyen et al., 1999). Heyen et al. (1999) offered a detailed description of the microsatellite marker information pertaining to the three families considered in this study. The total numbers of markers with information by chromosome across all families were: 11, 10, 6, 9, 6, and 4 corresponding to BTA3, 6, 7, 14, 21 and 22, respectively. The number of informative markers per autosome and family averaged 4.55 and ranged from two to seven. The total map length was 650 cM (Kosambi mapping function) with the average between-marker interval of 17.6 cM. The average heterozygosity was 59% with an average of 6.6 alleles per marker across all families. Table 1 presents the number of markers per family and autosome and the range of the number of informative sons per marker and family for all traits considered.

Model

Linkage analyses were performed for each trait within grandparental family using the general nonlinear mixed-effects model:

$$y_{ijk} = \mu_{ij} + f_{ijk} + e_{ijk} \quad [1]$$

where y_{ijk} was the test-day record measured on the k^{th} lactation day t_{ijk} , on daughter j of son i ; μ_{ij} represents the linear combination of the systematic fixed effects of herd-year-season and age at calving (covariate expressed in month) corresponding to daughter j ; f_{ijk} was a nonlinear function of cow-specific parameters or lactation curve descriptors and lactation day; and e_{ijk} was a random residual term. Preliminary analyses compared the fit provided by three nonlinear functions: Wood's (1967), inverse quadratic (Rodriguez-Zas et al., 2000a), and Morant and Gnanasakthy's (1989) using likelihood-based criteria (Akaike, 1974). The function that was best supported by the data and was used in following studies was that based on Morant and Gnanasakthy's (1989) function:

$$f_{ijk} = \alpha_{ij} \exp[\beta_{ij} t_{ijk}' + \delta_{ij} (t_{ijk}')^2 + \phi_{ij}/t_{ijk}] \quad [2]$$

The term t_{ijk}' denoted a positive monotone function of t_{ijk} that improves the convergence. The parameter α_{ij} was a scale parameter that characterizes the overall level of the trait (e.g., milk yield) during the lactation curve and had a major effect on the total cumulative value or area under the lactation curve. The parameter β_{ij} was the main shape parameter and describes the rate of change of the trait at mid-lactation or percentage per day. The parameter δ_{ij} was a secondary shape pa-

Table 1. Number of markers and range of number of sons per marker and range of number of granddaughters and test-day records per family.

Autosome	One ¹		Four		Five	
	Markers ²	Sons ³	Markers	Sons	Markers	Sons
3	7	35–173	6	41–95	5	40–86
6	6	31–159	5	39–109	7	44–110
7	4	78–143	4	35–101	4	36–88
14	2	33–134	7	45–110	7	46–106
21	4	65–155	4	28–83	4	43–115
22	2	54–134	2	26–48	2	53–98

¹Ranges of number of granddaughters per trait were 3235 to 28,817, 1967 to 22,651 and 1267 to 574 for families one, four and five, respectively. Ranges of number of test-day records per trait were 34,898 to 305,528, 20,695 to 239,521 and 13,202 to 60,788 for families one, four and five, respectively.

²Number of markers per autosome and family.

³Number of sons per marker and family. Lower bounds correspond to SCS and upper bounds correspond to milk yield. Fat and protein content numbers are equal or very similar to milk yield.

parameter and represents the change in the β_{ij} rate. The parameter ϕ_{ij} described the rate of change of the trait at the start of the lactation (e.g., rate of increase in milk production prior to the peak yield).

Figure 1 portrays the information on the lactation pattern summarized by the different parameters applied to milk yield. The trend denoted “standard” represents a typical lactation curve. The percentages of increase in the four parameters considered were selected based on the variability observed in the present data set, to reflect biologically meaningful lactation curves. The standard errors of the parameter estimates, expressed as a percentage of the estimates, were 15% for parameter α , 40% for parameters β and δ and 61% for parameter ϕ . The “ $\alpha+$ trend” depicts the increase in the lactation curve scale or level associated with a 4% increase in the parameter α , with the rest of the parameter values equal to those in the standard curve. The

Figure 1. Standard milk production lactation curve (—) and deviations due to positive treatments on the parameters $\alpha+$ ($\diamond$), $\beta+$, $\delta+$ ($\blacktriangle$) and $\phi+$ ($\blacksquare$).

scale effect of α is evident in the similarity of shape between the $\alpha+$ trend and standard trends, with the “ $\alpha+$ trend” being higher than the standard one. An increase in the parameter β value, with the rest of the parameter values equal to those in the standard curve, is associated with an increase in the persistency of the lactation curve scale. The curve associated with an increase in parameter β (not presented due to space limitations) exhibits a behavior similar to the standard curve up to the peak lactation and parallel but higher persistency thereafter. An increase in parameter δ , with the rest of the parameter values equal to those in the standard curve, results in augmented persistency. The lactation curve resulting from an increase in parameter δ (not presented due to space limitations) exhibits a nonparallel behavior (lesser decay) with respect to the standard trend in the middle- and late-lactation. The “ $\phi+$ trend” depicts the acceleration in the increments of yield prior to the lactation peak associated with a 40% increase in the parameter ϕ , with the rest of the parameter values equal to those in the standard curve. Although parameter ϕ mainly influences the shape of the curve early in the lactation, it has a small effect on the scale as suggested by the “ $\phi+$ trend” that exhibits a slight increase in the peak production. The correlation between parameters α , β , δ , and ϕ is low to moderate (0.23 to 0.52) except for the higher correlation between the shape parameters β and δ (0.87 to 0.93). For this reason, Figure 1 also depicts the changes in the milk production lactation curve due to a simultaneous increase of 40% in the β and δ parameters. The increase in persistency (deceleration of the yield reduction rate) can be observed in the “ $\beta+$, $\delta+$ trend”.

A linear model was used to describe each parameter using single marker and interval mapping approaches. The single-marker model was:

$$\lambda_{ijmn} = \mu_n + M_{\min} + u_{ijmn} \quad [3]$$

where λ_{ijmn} was any one of the n parameters of the nonlinear function ($n = 1, 2, 3$, and 4 or α , β , δ , and ϕ , respectively) describing the lactation curve corresponding to daughter j of son i that inherited marker allele m from the grandsire, μ_n was an overall mean for parameter n , $M_{\min}$ was the deviation associated with the fixed effect of marker allele m inherited by the i th son, and u_{ijmn} was the random effect of daughter j that includes unlinked polygenic and permanent environmental effects.

The interval mapping model was:

$$\lambda_{ijn} = \mu_n + b P_{in} + u_{ijn} \quad [4]$$

where λ_{ijn} was the parameter n corresponding to daughter j of son i , μ_n was an overall mean for parameter n , b was the regression coefficient associated with the fixed-effect probability that son i received the designated QTL allele (P_{in}) computed as described by Spelman et al. (1996), and u_{ijn} was the random effect of daughter j that included unlinked polygenic and permanent environmental effects.

The random terms u_{ijmn} and e_{ijk} were assumed to be Normal, independent, and identically distributed with mean zero and an $n \times n$ variance-covariance matrix for the polygenic effects, and a common residual variance. The nonlinear mixed-effects model was fitted using the method of Wolfinger and Lin (1997) using a restricted maximum likelihood (REML) approach implemented in the NLINMIX macro (Little et al., 1996).

An experiment-wise critical value $P < 0.05$ was chosen to protect against false positive and false negative results and because the nonrandom sample of Dairy Bull DNA Repository families and sons could result in a reduction of power to detect QTL (Georges et al., 1995). The Bonferroni and Hochberg methods were used to control the experiment-wise error in the strong sense (Westfall et al., 1999). Although principal component analysis within family indicated that three independent principal components explained, on average 90% of the variation of all four traits, multiple comparison adjustments assumed four independent traits. Two sets of confidence intervals for the QTL location on an interval mapping nonlinear mixed-effects model were computed using a LOD 1 drop, and P values lesser or equal to 0.001 surrounding the location estimate that maximized the likelihood. Resampling based thresholds were not computed due to the large computational demands associated with the large data set and model considered.

RESULTS AND DISCUSSION

General Results

A total of 324 tests, 81 map positions across three families evaluated at each of the four traits, were performed. Twenty-one tests (approximately 6% of all tests) were significant for at least one parameter descriptor of the lactation curve, at the experiment-wise level $P < 0.05$, and 32 were significant at nominal comparison $P < 0.001$. Only Bonferroni critical values are reported because the Bonferroni and Hochberg thresholds were very similar. Enumeration of the most experiment-wise significant positions together with family and trait information for all four descriptors of the lactation curves is shown in Table 2. No markers were significantly associated with all four parameters, 16 (76%) were associated with one parameter; 4 (19%) were associated with two parameters, and 1 (5%) was associated with three parameters describing the trend of the trait. Among the positions with a significant association with one parameter, nine, three, two, and two were associated with parameters α , β , δ , and ϕ , respectively. Considering the markers associated with two parameters, three were related to variations in parameters β and δ and one to α and β . The correlation between β and δ parameter estimates was high (0.85 to 0.92), and the common significant marker effect suggests a QTL with pleiotropic effect on both scale descriptors of the lactation curve and hence with effects during a prolonged stage of lactation. Only one map position was associated with significant variation of three parameters: β , δ , and ϕ . Among all map positions associated with significant variations in the lactation patterns, 10 (48%), 5 (24%), 4 (20%), and 2 (10%) were related to protein percentage, fat percentage, milk yield, and SCS, respectively. Ten of the markers with significant associations were found in family four, seven in family five, and four in family one.

An investigation of the incidence of false positive results in the proposed longitudinal linkage approach was conducted through the study of positions found nonsignificant (comparison-wise $P < 0.05$) in Heyen et al. (1999). In family five, information from all markers on BTA19 for fat percentage and SCS and on BTA13 for protein percentage trends was analyzed. No significant associations were detected between the map positions used as negative controls and the descriptors of the lactation curves. The P values ranged between 0.26 and 0.92 suggesting that the proposed longitudinal-linkage model does not artificially overestimate the association between genome positions and variation in the lactation curve patterns.

Considering all tests significant at unadjusted $P < 0.001$, 21 (68%) were associated with one parameter,

Table 2. Experiment-wise probability values with significant ($P < 0.05$) marker associations with at least one descriptor of the lactation curve by trait and family.

Autosome	Family	Trait	Marker name	Position (cM) ¹	Parameter			
					α	β	δ	ϕ
3	1	Protein	ILST096	16	9.0×10^{-3}	1.0×10^{-2}	1.6×10^{-1}	1.6×10^{-4}
3	1	Milk	BL41	32	6.9×10^{-2}	1.3×10^{-5}	5.6×10^{-4}	3.4×10^{-1}
3	1	Fat	BL41	32	1.5×10^{-1}	1.2×10^{-4}	7.4×10^{-4}	2.4×10^{-1}
3	1	Protein ²	BL41	32	4.9×10^{-2}	2.2×10^{-2}	3.2×10^{-1}	3.3×10^{-4}
3	4	Milk ²	HUJI177	100	2.3×10^{-4}	8.3×10^{-2}	9.4×10^{-2}	9.1×10^{-1}
3	4	Protein	HUJI177	100	1.3×10^{-1}	1.8×10^{-4}	1.4×10^{-3}	2.6×10^{-1}
3	4	Milk	BR4502	113	3.0×10^{-5}	4.3×10^{-1}	4.5×10^{-1}	4.9×10^{-1}
6	4	Protein	BM143	64	4.1×10^{-1}	5.1×10^{-7}	1.5×10^{-9}	4.7×10^{-1}
6	5	Protein	BM143	64	3.2×10^{-5}	7.1×10^{-2}	1.7×10^{-1}	1.0×10^{-1}
6	4	Protein	BMC4203	120	1.9×10^{-6}	5.0×10^{-6}	9.4×10^{-4}	1.3×10^{-1}
7	1	Protein	BMS2258	81	6.7×10^{-6}	5.1×10^{-1}	9.9×10^{-1}	3.7×10^{-1}
14	5	Fat ²	ILSTS039	0	2.3×10^{-4}	1.0×10^{-1}	1.7×10^{-1}	4.1×10^{-4}
14	4	Milk	CSSM066	13	1.6×10^{-1}	7.6×10^{-6}	2.1×10^{-6}	8.5×10^{-1}
14	4	Milk	CSSM066	13	1.6×10^{-1}	7.6×10^{-6}	2.1×10^{-6}	8.5×10^{-1}
14	4	Protein	BM4513	96	1.6×10^{-4}	1.4×10^{-1}	2.5×10^{-1}	1.0×10^{-1}
14	4	SCS	BM6425	113	5.9×10^{-5}	4.0×10^{-2}	9.2×10^{-2}	1.9×10^{-2}
14	5	Protein	BM6425	96	2.3×10^{-5}	6.8×10^{-2}	2.7×10^{-1}	4.8×10^{-1}
21	5	Fat	BM8115	0	3.1×10^{-3}	4.9×10^{-2}	9.9×10^{-2}	2.0×10^{-4}
21	4	Milk	ILSTS103	56	4.7×10^{-2}	4.2×10^{-3}	1.9×10^{-4}	8.5×10^{-1}
21	5	Fat	ILSTS103	56	8.8×10^{-4}	6.6×10^{-2}	2.7×10^{-1}	1.6×10^{-4}
21	5	Fat	TGLA122	78	2.5×10^{-5}	2.0×10^{-1}	9.3×10^{-1}	8.0×10^{-4}
21	4	Protein	ILSTS054	85	4.7×10^{-1}	8.6×10^{-6}	3.7×10^{-8}	9.4×10^{-1}
22	5	Fat	INRA194	41	9.8×10^{-2}	1.4×10^{-6}	2.7×10^{-8}	1.5×10^{-9}
22	5	Protein	CSSM041	65	9.25×10^{-6}	9.5×10^{-1}	6.1×10^{-1}	1.6×10^{-2}

¹Estimated position relative to first marker genotyped (Heyen et al., 1999). Positions of the first marker are: 25.8 cM, 0 cM, 22.1 cM, 0 cM, 0 cM, 0 cM for autosomes three, six, seven, 14, 21 and 22, respectively (U S Meat Animal Research Center, 2001).

²Experiment-wise significance of marker was $P < 0.07$.

eight (26%) were associated with two parameters, two (6%) were associated with three parameters, and none was associated with all four parameters describing the trend of the trait. Among the positions with a significant association with one parameter, ten, three, three, and five were associated with α , β , δ , and ϕ , respectively. No marker associated to two parameters was significantly related to parameter δ . Only one map position was associated with significant variation of three parameters: β , δ and ϕ . Among all positions associated to significant variations in the lactation patterns, 13 (41%), 8 (26%), 7 (23%), and 3 (10%) were related to protein percentage, fat percentage, milk yield, and SCS, respectively. Among the markers with significant associations, 15 were found in family four, 11 in family five, and 5 in family one. Of all markers associated with significant variation of lactation patterns, eight were on BTA3, five were located on BTA6, one was located on BTA7, nine were on BTA14, six were located on BTA21, and two were on BTA22.

The interval mapping approach confirmed many of the single-marker findings and uncovered additional genome positions with QTL. A summary of the locations significant at comparison-wise $P < 0.001$ is presented in Table 3.

The analysis of BTA3 showed significant associations between marker alleles and production traits for different parameters within different families (Table 2). Marker BL41 was associated with variation in milk yield and fat percentage patterns and was almost significant for protein percentage in family one (Table 2). A QTL near markers ILST096 and BL41 may be responsible for the significant association between these positions and the parameter ϕ that describes the change of the trait early in the lactation, in family one. The latter map position was also associated with variations of the parameter β that describes the persistency of milk yield and fat percentages in family one. The common significant marker effect suggests a QTL with pleiotropic effect on these traits or many QTL in the region. The interval mapping approach detected a QTL influencing the shape parameter β of the protein lactation curve at an intermediate position (62 to 72 cM) on same autosome in family one (Figure 2). A second QTL of lesser P value influencing parameter ϕ was also detected between 20 and 34 cM on BTA3 in family one. Heyen et al. (1999) reported a significant association between DYD for protein percentage with markers ILST096 and BL41 in the same family.

Table 3. Probability values, estimates of locations, LOD-1 drop off and lower and upper $P < 0.001$ limits of QTL positions significant at $P < 0.001$ by family and trait.

Autosome	Family	Trait	λ^1	P value	Position (cM) ²	Limits			
						LOD-1 drop off		$P < 0.001$	
						Lower	Upper	Lower	Upper
3	4	Milk	α	5.3×10^{-4}	9	0	40	0	36
3	4	Milk	α	4.1×10^{-4}	125	88	130	91	130
3	1	Protein	β	8.3×10^{-4}	70	20	89	62	72
6	4	Milk	α	6.4×10^{-4}	2	0	24	0	21
6	4	Milk	β	3.8×10^{-4}	129	106	129	108	129
7	4	Milk	α	2.3×10^{-5}	24	19	27	15	44
7	5	Milk	α	7.0×10^{-4}	41	31	53	37	45
7	5	SCS	α	6.4×10^{-4}	50	41	54	48	52
14	5	Fat	ϕ	2.1×10^{-5}	0	0	8	0	9
14	5	Fat	ϕ	7.5×10^{-5}	30	29	32	24	39
14	5	SCS	α	6.8×10^{-4}	48	24	75	38	53
21	5	SCS	α	6.4×10^{-4}	32	19	55	27	38
21	4	Protein	β	7.1×10^{-4}	43	32	58	37	45
22	1	Protein	β	6.9×10^{-4}	9	0	24	0	19

¹Parameter descriptor of the lactation curve.

²Estimated position relative to first marker genotyped (Heyen et al., 1999). Positions of the first marker are: 25.8 cM, 0 cM, 22.1 cM, 0 cM, 0 cM, 0 cM for autosomes three, six, seven, 14, 21 and 22, respectively (U S Meat Animal Research Center, 2001).

In family four, marker HUJI177 on BTA3 was associated with almost significant variation of the milk yield and protein percentage lactation curves. This marker had a significant association with the scale parameter α that describes the milk yield curve and a shape parameter β for protein percentage. These findings suggest either a QTL with pleiotropic effects across traits or multiple QTL in the vicinity of the marker. The pres-

ence of a QTL on the region is supported by the significant association between the next marker (BR4502) and the milk lactation curve scale parameter α on the same family. Interval mapping confirmed the presence of a QTL positioned between 91 and 130 cM on BTA3 in family four influencing the scale parameter α for milk production (Figure 3). In addition, the interval mapping approach identified a second QTL located on the other side of the autosome (0 to 36 cM), influencing the same trait descriptor in the same family. The latter findings

Figure 2. Interval mapping probability value for protein percentage lactation curve parameters β and ϕ (dotted and continuous curves, respectively) across autosome three in family one. The LOD1 drop off and $P < 0.001$ thresholds are denoted with straight lines (continuous and dotted, respectively). The $\blacktriangle$ symbol indicates the marker positions.

Figure 3. Interval mapping probability value for milk yield lactation curve parameter α (continuous curve) across autosome three in family four. The LOD1 drop off and $P < 0.001$ thresholds are denoted with straight lines (continuous and dotted, respectively). The $\blacktriangle$ symbol indicates marker positions.

Figure 4. Interval mapping probability value for milk yield lactation curve parameters α , β , and δ (continuous, dotted unmarked and $\bullet$ marked curves, respectively) across autosome six in family four. The LOD1 drop off and $P < 0.001$ thresholds are denoted with straight lines (continuous and dotted, respectively). The $\blacktriangle$ symbol indicates the marker positions.

are consistent with results reported by Heyen et al. (1999).

The analysis of BTA6 showed multiple significant associations between marker alleles and trait lactation patterns in different families (Table 2). The centromeric marker BM143 was associated with significant variation of the protein pattern in families four (β and δ) and five (α) although the affected parameters vary between families. The different lactation pattern descriptors affected on both families could be indicative of more than one QTL for protein located on the region. On BTA6 a QTL affecting parameter β was detected, with interval mapping, between 108 and 129 cM in family four. The same or a second QTL in the proximity influenced the second shape parameter δ as indicated in Figure 4. A second QTL on BTA6 influencing α on milk production (family four) was detected between 0 and 21 cM (Figure 4).

These results for BTA6 are consistent with previous studies based on cumulative single measurements (Georges et al. 1995; Spelman et al. 1996; Lipkin et al., 1998; Ashwell and Van Tassell, 1999; Kuhn et al., 1999; Van Tassell et al., 2000). The ability of the longitudinal mapping models to detect associations between markers and lactation patterns at specific stages of the lactation may explain some inconsistency between the present study and Heyen et al. (1999). These authors did not detect any experiment-wise significant association between markers on BTA6 and DYD of multiple production and health dairy traits, although there were multiple associations below the suggestive threshold in the QTL critical region (Heyen et al., 1999; <http://cagst.animal.uiuc.edu/genemap/WEB/Table1.html>).

Only one marker on BTA7 (BMS2258), located at 81 cM, had a significant association with the scale, cumulative parameter α describing the protein curve in family one (Table 2). This result was consistent with Heyen et al. (1999), who found an association between ILSTS006 positioned at 115 cM and DYD for protein yield in family one. Also, Lipkin et al. (1998) reported a significant association between marker ILSTS006, located 10 cM from the *RASA* gene (protein activator, 103 cM) and the predicted breeding value for protein production. A QTL with influences on the scale (parameter α) of the lactation curve was located between 15 and 44 cM on BTA7 in family four using interval mapping. Likewise, a QTL located between 37 and 45 cM affecting parameters α and β was identified in family five. Although the location estimate that maximizes the likelihood differs between families four and five (24 and 41 cM, respectively), it is not possible to conclude whether one or two QTL had been identified due to the difference in marker information availability and the single-QTL model used. Meanwhile marker RM006 (4 cM) flanks the centromeric interval in both families, markers BMS2258 (81 cM) and UWCA20 (43 cM) right-flank the interval in families four and five, respectively. A QTL affecting the scale of the SCS lactation pattern was detected on BTA7 in family five, within the 48 to 52 cM nonoverlapping interval.

Markers on BTA14 were associated with variation of the lactation curves for all traits in families four and five (Table 2). A QTL with effects on the fat percentage pattern could be responsible for the significant association between marker ILSTS039 and the scale parameter α in family five. This result is consistent with the analysis of DYD records by Heyen et al. (1999). The two significant genome positions influencing the pattern of fat percentage detected with an interval mapping model on BTA14 in family five are summarized in Table 3. The first QTL was located between zero and eight cM and the second between 24 and 39 cM, both influencing parameter ϕ and in coupling phase (Figure 5). The presence of one QTL in one interval influencing the estimates of the QTL location or effect in the other interval is unlikely since the marker information was available at 0, 13, 21, and 31 cM (Figure 6).

Marker CSSM066 had a significant association with the shape parameters β and δ that describe changes in milk yield during mid- and late-lactation in family four. Heyen et al. (1999) reported an association between this marker and fat yield DYD. Variation in the protein pattern main-scale parameter α was associated with marker BM4513 positioned at 96 cM and marker BM6425 positioned at 113 cM in families four and five, respectively. The consistency of these results strongly suggests that a QTL for protein percentage is located

Figure 5. Interval mapping probability value for fat percentage lactation curve parameter ϕ (dotted curves) across autosome 14 in family five. The LOD1 drop off and experiment-wise P thresholds are denoted with straight lines (continuous and dotted, respectively). The $\blacktriangle$ symbol indicates the marker positions.

in the region. The latter marker exhibited the only experiment-wise significant association with variation in the SCS pattern. The association between the marker and the scale parameter was detected in family four. Interval mapping indicated that variations of the parameters α and ϕ describing the SCS lactation pattern were associated to a QTL between 38 and 53 cM on BTA14 in family five. Previous research suggested similar results (Coppeters et al., 1998; Ron et al., 1998; Ashwell and Van Tassell, 1999; Riquet et al., 1999).

Figure 6. Interval mapping probability value for SCS lactation curve parameters α and ϕ (continuous and dotted curves, respectively) across autosome 21 in family five. The LOD1 drop off and $P < 0.0001$ thresholds are denoted with straight lines (continuous and dotted, respectively). The $\blacktriangle$ symbol indicates the marker positions.

Markers on BTA21 were associated with significant variation of the parameters describing all production trait curves in families four and five. Markers BM8115 and ETH131 had a significant association with variation on the fat percentage pattern in family five as indicated in Table 2. The first marker had an impact on parameter ϕ for fat percentage, and the second marker influenced parameter δ . The same parameters were associated with marker ILSTS103 that had a significant association with δ for milk yield in family four and with ϕ for fat percentage in family five. These findings are supported by significant results in consecutive marker positions. There was a significant association between marker TGLA122 and the fat percentage scale parameter α in family five and between marker ILSTS054 and the protein shape parameters β and δ in family four. Ashwell and VanTassell (1999) reported that marker BM103 (31 cM) on BTA21 was associated with PTA for productive life, a trait closely related to production. A significant effect of a QTL located between 37 and 45 cM on BTA21 on the parameter β for protein percentage was identified by the interval mapping approach in family four (Table 3).

On BTA21, QTL with significant effects on the parameters α and ϕ for SCS were detected between 27 and 38 cM in family five using the interval mapping model (Figure 6). The inability of the single-marker model to detect a significant association between the nearby marker ETH131 (33 cM) and SCS may be caused by the reduced number of informative sons in this family, marker, and nearby markers.

The significant results observed on BTA22 are limited to family five (Table 2). The marker INRA194 was associated with significant variation of the shape parameters β , δ , and ϕ describing the fat percentage lactation curve. The following marker (CSSM041) had a significant association with variation of the cumulative scale parameter α describing the protein percentage lactation curve. On BTA22, a QTL with significant effect on the parameter β for protein percentage was detected between 0 and 19 cM in family one with interval mapping.

Lactation Patterns

A major advantage of some nonlinear models over other models for repeated measurements is the ease of interpretation of the estimates as they relate to marker or QTL effects. Estimates of the marker allele deviations, adjusted by the remainder of the factors in the model and obtained using the single-marker approach, were used to estimate the lactation curve of cows receiving alternative alleles from the grandsire. A significant association between marker ILST103 (BTA21) and the

Figure 7. Estimated milk yield associated with marker ILST103 (autosome 21) alternative alleles (+ and ◆, respectively) in family four.

rate of change in milk yield during mid- and late-lactation is evident in the differential persistency observed in both curves (Figure 7). The association between marker BM4513 (BTA14) and the scale parameter α is evident in Figure 8 with the estimated protein pattern of daughters from a son that received one of the marker BM4513 (BTA14) alleles lower than the one corresponding to the other allele. These findings suggest the potential impact of genetic loci at specific lactation phases and provide the basis for targeted modification of particular aspects of the lactation curve. In terms of the direction of allelic deviations, some marker alleles were identified as having opposite associations on the shape or scale of the lactation curve across families. An example of this situation, typically observed in outbred populations, is the alleles of marker BM8115 on BTA21 that had opposite influence on the protein percentage scale and shape in families four and five. The estimated ef-

Figure 8. Estimated protein percentage associated with marker BM4513 (autosome 14) alternative alleles (+ and ◆, respectively) in family four.

fects (and standard error) on parameters α , β , δ , and ϕ were -1.04 (0.44) percentage, 0.032 (0.024) percentage change per mid-lactation day, -0.0026 (0.04) rate of percentage change per mid-late lactation day, 0.583 (0.253) percentage change per early-lactation day, 0.37 (0.21) percentage, -0.018 (0.008) percentage change per mid-lactation day, 0.005 (0.003) rate of percentage change per mid-late lactation day, and 0.017 (0.07) percentage change per early-lactation day in families four and five, respectively. The lack of data for the subsequent marker and different alleles at the following markers prevents the evaluation of haplotype effects at this map position.

Further Considerations

Alternative models to analyze the repeated measurements are available. Wu et al. (1999) utilized a multiple trait model to map QTL in rice, and each time point was treated as a separate trait. The multiple trait approach accounts for the correlation between records taken at different times and can identify genetic factors acting at the measured points, but it cannot directly identify genes exerting effect during a particular stage. Linear mixed-effects animal models such as random regression models (Jamrozik and Schaeffer, 1997), covariance functions (Meyer, 1998), and splines (White et al., 1999) have been used to estimate genetic parameters and provide genetic predictions without specific modeling of QTL or markers. Kirkpatrick (1997) succinctly proposed to combine covariance functions and QTL detection models, but the interpretation of the results (e.g., eigenvalues and orthogonal polynomials) may be cumbersome. In this study, nonlinear functions were chosen as an intuitive and comprehensive way to portray the lactation curves. All the parameters of the nonlinear function considered, as opposed to some parameters in some random regression and covariance functions, provide a direct description of a geometric property of the curve or have some biological interpretation (e.g., persistency) that can explain the underlying biology. In addition, the proposed longitudinal-QTL model can be extended to other traits (i.e., fat and protein yield) and records from multiple lactations. Similarly, the proposed model can include a family effect and can be used in an across-family study. The differences in linkage phase and marker information across-family and the family-specific estimated QTL positions suggest that an across-family study would not provide information substantially different from that already uncovered in the within-family studies.

The detection of QTL influencing the lactation curve patterns can be undertaken using alternative Bayesian approaches (Gianola and Kachman, 1983; Rodriguez-

Zas et al., 2000b) that can model genetic relationships between animals and incorporate prior information about the unknown parameters. This approach was not undertaken in this study because our aim was to demonstrate that the proposed methodology could be implemented with existing software. In addition, the large number of observations and animals might have, in the end, overwhelmed the prior distributions. Previous results from a study of SCS lactation patterns on a reduced number of cows showed very similar estimates of fixed effects between the approach implemented in this study (Rodriguez-Zas et al., 2000a) and a more complex model (Rodriguez-Zas et al., 2000b). Further, Zhang et al. (1998) reported no substantial differences between the results from least squares and variance component QTL detection approaches. Therefore, similar results are expected if a more complex model were fitted. Our results indicate that multiple models (i.e., single-marker and interval mapping) need to be examined to confirm the presence of QTL and also to identify time-dependent QTL.

CONCLUSIONS

This study revealed that QTL affecting the scale and shape of the lactation curve for production and health traits are segregating in the dairy cattle population. The results suggest the presence of QTL that influence the scale and shape parameters that describe the lactation pattern. Most results were consistent with QTL detected using single-lactation measurements. The proposed approach provided additional insight on gene function based on the lactation stage when the effect was detected. Significant marker and QTL associations were found for all parameters, traits, and autosomes studied, but varied among families and traits. Most map positions were associated with significant variation in one descriptor, suggesting that the correlated QTL exerts its effect at a particular stage of the lactation. The detection of QTL with stage-specific effects (e.g., a QTL located at 30 cM on BTA14 influencing the decay on fat percentage early in the lactation, a QTL located at 70 cM on BTA3 influencing the persistency in protein percentage) associated with parameters may explain why some of the significant positions have not been previously reported or discrepancies among previous studies that used cumulative measurements. The association of positions with more than one parameter suggests that some QTL exert their effect across an extended lactation period. These QTL are more likely to be detected than time-specific QTL with traditional single-measurement linkage models. The proposed model complements the findings from single-measurement models. The nonlinear model fitted allowed for

meaningful inferences while being sufficiently flexible to account for the main sources of variation without high computational demands. The repeated-measurement mapping approach developed in this study has great potential in that it can be applied to other well-studied longitudinal data such as growth in pigs, poultry, and beef cattle or immunological response indicators in health-related studies. The statistical approach described may lead to greater biological understanding of marker associations and a more intelligent approach for selection of candidate genes. The identified positions can be incorporated into marker-assisted selection decisions to alter the persistency and peak production or the fluctuation of SCS during a lactation.

ACKNOWLEDGMENTS

We thank Jill Philpot (Animal Improvement Programs Laboratory, USDA) for providing the test-day production and SCS records and collaborators of regional project NC209 and the contributing AI organizations for initiating, maintaining, and contributing to the Dairy Bull DNA Repository. Continuous support from CSREES, project number ILLU-35-0350 is greatly appreciated. This work was partly funded by grants from the National Center for Supercomputing Application numbers MCB990004N and MCB990029N.

REFERENCES

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Auto. Cont.* AC-19:716–723.
- Ashwell, M. S., Y. Da, C. P. Van Tassell, P. M. VanRaden, R. H. Miller, and C. E. Jr. Rexroad. 1998. Detection of loci affecting milk production and composition, health, and type traits in a United States Holstein population. *J. Dairy Sci.* 81:3309–3314.
- Ashwell M. S., and C. P. Van Tassell. 1999. Detection of putative loci affecting milk, health, and type traits in a US Holstein population using 70 microsatellite markers in a genome scan. *J. Dairy Sci.* 82:2497–2502.
- Coppieters, W., J. Riquet, J. J. Arranz, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marcq, P. Simon, P. Vanmanshoven, D. Wagenaar, and M. Georges. 1998. A QTL with major effect on milk yield and composition maps to bovine chromosome 14. *Mamm. Genome* 9:540–544.
- Dairy Bull DNA Repository. 2001. Available: <http://cagst.animal.uiuc.edu/genemap/WEB/intro.html>. Accessed Sep. 20, 2001.
- Freeze B. S., and T. J. Richards. 1992. Lactation curve estimation for use in economic optimization models in the dairy industry. *J. Dairy Sci.* 75:2984–2989.
- Geldermann, H. 1975. Investigations on inheritance of quantitative characters in animals by gene markers. *Theor. Appl. Genet.* 46:319–330.
- Gengler, N. 1996. Persistency of lactation yields: a review. Pages 87–96 in *Proc. Int. Workshop Genet. Impr. Functional Traits in Cattle*, Gembloux, Belgium. *Int. Bull. Eval. Serv. Bull.* No. 12. Dep. Anim. Breed. Genet., SLU, Uppsala, Sweden.
- 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 by exploiting progeny testing. *Genetics* 139: 907–920.

- Gianola, D., and S. D. Kachman. 1983. Prediction of breeding value in situations with nonlinear structure, categorical responses, growth functions and lactation curves. Pages 172–176 in Proc. Thirty-fourth Ann. Meet. Europ. Assoc. Anim. Prod., Madrid, Spain.
- Heyen, D. W., J. I. Weller, M. Ron, M. Band, J. E. Beever, E. Feldmesser, Y. Da, G. R. Wiggans, P. M. VanRaden, and H. A. Lewin. 1999. A genome scan for QTL influencing milk production and health traits in dairy cattle. *Physiol. Genomics* 11:165–175.
- Jamrozik, J., and L. R. Schaeffer. 1997. Estimates of genetic parameters for a test day model with random regression for yield traits of first lactation Holsteins. *J. Dairy Sci.* 80:762–770.
- Kirkpatrick, M. 1997. Genetic improvement of livestock growth using infinite-dimensional analysis. *Anim. Biotech.* 8:55–61.
- 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.
- Lipkin, E., M. O. Mosig, A. Darvasi, E. Ezra, A. Shalom, A. Friedmann, and M. Soller. 1998. Quantitative trait locus mapping in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers: analysis of milk protein percentage. *Genetics* 149:1557–1567.
- Littell, R. C., G. A. Milliken, W. W. Stroup, and R. R. Wolfinger. 1996. SAS® System for Mixed Models. SAS Institute Inc., Cary, NC.
- Meyer, K. 1998. Estimating covariance functions for longitudinal data using a random regression model. *Genet. Sel. Evol.* 30:221–240.
- Morant, S. V., and A. Gnanasakthy. 1989. A new approach to the mathematical formulation of lactation curves. *Anim. Prod.* 49:151–162.
- Riquet, J., W. Coppieters, N. Cambisano, J.-J. Arranz, P. Berzi, S. K. Davis, B. Grisart, F. Farnir, L. Karim, M. Mni, P. Simon, J. F. Taylor, P. Vanmanshoven, D. Wagenaar, J. E. Womack, and M. Georges. 1999. Fine-mapping of quantitative trait loci by identity by descent in outbred populations: application to milk production in dairy cattle. *Genetics* 96:9252–9257.
- Rodriguez-Zas, S. L., D. Gianola, and G. E. Shook. 2000a. Evaluation of models for somatic cell score lactation patterns in Holsteins. *Livest. Prod. Sci.* 67:19–30.
- Rodriguez-Zas, S. L., D. Gianola, and G. E. Shook. 2000b. An approximate Bayesian analysis of somatic cell score curves in Holsteins. *Acta Agric. Scand. Sec. A* 50:291–299.
- Ron, M., D. W. Heyen, J. I. Weller, M. Band, E. Feldmesser, H. Pasternak, Y. Da, G. R. Wiggans, P. M. Vanraden, E. Ezra, E., and H. A. Lewin. 1998. Proc. 6th World Congr. Genet. Appl. Livest. Prod., Armidale, Australia 26:422–425.
- Spelman, R. J., W. Coppieters, L. Karim, J. A. M. van Arendonk, and H. Bovenhuis. 1996. Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population. *Genetics* 144:1799–1808.
- USDA Meat Animal Research Center. 2001. Available: <http://sol.marc.usda.gov/>. Accessed Sep. 20, 2001.
- Van Tassell, C. P., M. S. Ashwell, and T. S. Sonstegard. 2000. Detection of putative loci affecting milk, health, and conformation traits in a US Holstein population using 105 microsatellite markers. *J. Dairy Sci.* 83:1865–1872.
- Varona, L., C. Moreno, L. A. Cortez, and J. Altarriba. 1998. Bayesian analysis of Wood's lactation curve for Spanish dairy cows. *J. Dairy Sci.* 81:1469–1478.
- Weller, J. I., 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.
- Westfall, P. H., R. D. Tobias, D. Rom, R. D. Wolfinger, and Y. Hochberg. 1999. Multiple comparisons and multiple tests using the SAS® system. SAS Institute Inc., Cary, NC.
- White, I. M. S., R. Thompson, and S. Brotherstone. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. *J. Dairy Sci.* 82:632–638.
- Wolfinger, R., and X. Lin. 1997. Two Taylor-series approximation methods for nonlinear mixed models. *Comp. Stat. Data An.* 25:465–490.
- Wood, P. D. P. 1967. Algebraic model of the lactation curve in cattle. *Nature* 216:164–165.
- Wu, W., W. Li, D.-Z. Tang, H.-R. Lu, and A. J. Worland. 1999. Time-related mapping of quantitative trait loci underlying tiller number in rice. *Genetics* 151:297–303.
- 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.