Research Institute for the Biology of Farm Animals, Dummerstorf, Germany

GERTRAUDE FREYER, CHRISTA KÜHN and ROSEMARIE WEIKARD

Comparison of different statistical-genetic approaches of QTL detection by Evaluating results from a real dairy cattle data set

Abstract

Recent reports on estimating QTL positions and effects on milk production traits show several chromosomal regions, for example on BTA6, being putative QTL regions, for milk yield and traits of ingredients. The statistical methods and genetic models on which these results were based on, show different advantages and limits. Thus, it is sometimes difficult to compare and evaluate such results. Confirmation studies are inevitable, before drawing conclusions towards finemapping analyses or practical use of such results. We compared three published approaches, realizing up to five different genetic models of QTL estimation analysing the traits fat yield (kg) and fat content (%) based on a real data set of five granddaughter families from the German Holstein dairy cattle population. The marker map contained 16 microsatellite markers, distributed across chromosome BTA6. The focus was mainly on the most likely map positions of the putative QTL and on its character. The significant results obtained from using different methods and adapting several genetic models were conclusive and comparable to QTL positions reported elsewhere.

Key Words: comparison of methods, QTL mapping, multivariate analysis

Zusammenfassung

Titel der Arbeit: **Vergleich verschiedener statistisch-genetischer Methoden der QTL-Schätzung anhand der Ergebnisse eines Beispiels aus der Milchrindpopulation**
Verschiedene Positionen für QTL mit Einfluss auf Milchmenge und Inhaltsstoffe wurden z.B. auf dem Chromosom BTA6 gefunden. Die statistischen Methoden und genetischen Modelle sind oft unterschiedlich, haben Vorteile und auch Grenzen. Bestätigungsstudien sind notwendig, bevor die nötigen Schlussfolgerungen gezogen werden können, die auf Analysen in Richtung Feinkartierung bzw. praktische Nutzung solcher Ergebnisse abzielen. Wir verglichen drei publizierte methodische Lösungen, die bis zu fünf unterschiedliche genetische Modelle zur QTL- Schätzung realisierten. Ein reales Datenmaterial, bestehend aus fünf Halbgeschwisterfamilien aus der Deutschen Holstein Friesian Population, typisiert für 16 Mikrosatellitenmarker auf dem Chromosom BTA6, wurde bezüglich Fettmenge (kg) und Fettgehalt (%) analysiert. Das Hauptaugenmerk lag auf der Schätzung der wahrscheinlichsten QTL- Positionen und dem Charakter der QTL. Die signifikanten Ergebnisse bezüglich der QTL- Positionen waren aus den unterschiedlichen Analysen schlüssig interpretierbar und vergleichbar mit QTL-Positionen, die anderswo publiziert wurden.

Schlüsselwörter: Methodenvergleich, QTL-Kartierung, multivariate Analyse

Introduction

Worldwide, many scientists on the field of statistical genetics are concerned with creating new and more appropriate solutions of QTL estimation. Therefore, when focusing on a QTL study, it is often a question to decide, which approach is most suited for the task one is confronted with. The current solutions partly show different facilities and it depends most likely on the data structure, which genetic model and statistical method is preferable in the special case. In animal breeding, we find the highest number of publications with respect to milk production traits. Usually, the underlying experimental design for QTL estimation in dairy cattle is a half sib

414

FREYER et al.: Different statistical-genetic approaches of QTL detection by Evaluating results from a real dairy cattle data set

structure, either a granddaughter or a daughter design (WELLER et al., 1990). The number of successful QTL studies is steadily growing since the beginning nineties. The type of analyses changed during this time. First, e.g. when focusing on BTA6, there was a concentration on the casein gene variants for estimating direct effects on milk traits (e.g. BOVENHUIS et al., 1992; BOICHARD et al., 1994; LIEN et al., 1995; ORTNER et al., 1995; PANICKE et al., 1996). Then, the casein genes were used as markers in a single marker segregation analysis (e.g. van ARENDONK et al., 1994; LIU, 1994). The next step was covering the whole chromosome with several markers, mostly in a distance of ~20 cM (e.g. GEORGES et al., 1995; BOVENHUIS and WELLER, 1994; SPELMAN et al., 1996) in order to obtain a more precise result for a putative QTL region. However, these early studies were based on single trait – single QTL models. ZHANG et al. (1998) and further studies based on modeling more than one QTL, what came closer to the real situation. The focus on multivariate estimation of QTL, taking the information of correlated traits into account, is new (e.g. ALMASY et al., 1997; SØRENSEN et al., 2003). A comprehensive overview on methods of QTL estimation was given by HOESCHELE (2001). In dairy cattle breeding, there are only very few published results based on simultaneous multivariate and multi-QTL-analyses (FREYER et al., 2003).

Obtaining results from analyses using different statistical methods, besides from different data sets, is valuable for reasons of confirming a QTL. This means that the own results can be verified and also compared with QTL findings obtained elsewhere. Mostly, current approaches for QTL estimation in animal breeding have been created with respect to the conditions of a special task. They do not allow for general use, e.g. they need a specified family design, limited on two generations genotyped, or excluding animals with incomplete marker information. QTL estimates might reflect these properties.

Thus, the goal of this paper was to compare and to evaluate three different statistical-genetic approaches of QTL estimation, partly based on the same genetic modelling in order to enable direct comparisons, and to show some directions in adapting them, monitoring the method-specific results. We demonstrate this here with results from analysing real data, because of combining the comparison of methods with their practical application. The focus was on finding and comparing QTL map positions. In particular, results obtained with recent reports (FREYER et al., 2002, 2003) were lighted up, for the two measures yield and percentage of basically the same substrate - milk fat, focusing on methical effects of estimation.

Material and Methods

Statistical methods

Our comparison of methods for multi-QTL estimation is based on three different advanced solutions, each created for the specific conditions of dairy cattle breeding. They all allow for the same family structure, which is basically a half sib design, were the (grand)sires are assumed to be unrelated. Further, they allow for a nearly unlimited number of markers on a chromosome, whose information is used for calculating the probability of the unknown QTL alleles. The way of using this information in particular, e.g. within IBD matrices, alters somewhat between the methods. In all cases, the test statistics may be calculated at 1 cM intervals within a marked region of a chromosome. They all allow for incomplete marker information as well. But the

effect of this surely depends on the number of markers and their information content. The effect of using DYD or EBV on QTL estimates of QTL positions did not show any difference between the methods.

We can provide a brief informative description of the statistical methods only. For a more extensive explanation see FREYER et al. (2003) and the relevant original papers cited, respectively.

Method 1: Least squares method (**LS,** e.g. HOESCHELE, 2001)

The model including n QTL based on the granddaughter design was

$$y_{ij} = \mu + s_i + \sum_{q=1}^{n} b_{ikq} P_{ijkl} + e_{ij} , \qquad [1]$$

where y_{ij} was the trait value of son j of (grand)sire i, μ was the overall mean, s_i was a fixed effect, b_{ikq} was a regression coefficient for QTL q nested within sire i at position k, P_{ijkl} means the probability of inheriting a QTL allele l of QTL q from sire i at position k, n was the number of QTL fitted at a time (1 or 2) and e_{ij} residual with variance approximately equal to variance/z_{ij}. Here, z_{ij} is a weighting factor according to the reliability of EBV of son j of sire i. Using the model above with q = 1, an F-statistic for H_0 (no QTL) versus H_A (one QTL present) was adapted. The test of the one-QTL model versus the two-QTL model was performed using F-statistics as well. Within a set of computer programs (DU and HOESCHELE, 1999) the LS analysis, optional incorporating epistatic effects between the two QTL, could be performed. Boots- trapping and permutation test of the original data to obtain confidence intervals CI (95%) of QTL positions and a chromosome-wide significance threshold was included.

Method 2: Residual maximum likelihood method (**REML,** GRIGNOLA et al., 1996; ZHANG et al., 1998)

The residual maximum likelihood method) based on variance component analysis (VC) was implemented in a computer program by ZHANG and HOESCHELE (1998). The model including q QTL was according to GRIGNOLA et al. (1997):

$$y_{ij} = \mu + u_{ij} + \sum_{q=1}^{n} (v_{qij}^{1} + v_{qij}^{2}) + e_{ij} \qquad [2]$$

with variance $var(e_{ij}) = \sigma_e^2/z_{ij}$, where in addition to the symbols described for [1] u_{ij} was the additive genetic effect of son j of sire i, v_{qij}^{k} means an effect of allele l (l=1,2) at QTL q of son j of sire i, and σ_e^2 is the residual variance. The likelihood ratio test (LR) was used to test for presence of a putative QTL and to discriminate between one- and two-QTL models. The distribution of the LR test statistics from REML analysis follows approximately a chi-square distribution (with 1 to 2 degree of freedom) used for testing, as most authors of such solutions presuppose.

Method 3: Multivariate VC method (**AIREML**, JENSEN et al., 1997 ; SØRENSEN et al., 2003)

This approach is based on (co)variance component analysis and a random QTL model, and was realized within the DMU package (JENSEN and MADSEN, 1994). The data was analysed by fitting the multivariate mixed model including the effect of n_q QTL,

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \sum_{q=1}^{n} \mathbf{W_q a_q} + \mathbf{e} \qquad [3]$$

where **y** was a vector of fat yield and fat content observed for each son, **b** was a vector of fixed overall mean effects, $\mathbf{a_q}$ was a vector of random effects for the QTL q

416

FREYER et al.: Different statistical-genetic approaches of QTL detection by Evaluating results from a real dairy cattle data set

assumed to be normally distributed with mean and variances ($\mathbf{0}$, $\mathbf{K}_0^i \otimes \mathbf{Q}_{i|M,p}$); correspondingly $\mathbf{u}$ was a vector of random breeding values ($\mathbf{0}$, $\mathbf{G} \otimes \mathbf{A}$) and $\mathbf{e}$ a vector of random residual effects ($\mathbf{0}$, $\frac{1}{n_j}\mathbf{E} \otimes \mathbf{I}$), adjusted for the number of sons daughters (n_j).

$\mathbf{K}_0^i$ was an unknown trait specific (co)variance matrix. $\mathbf{G}$ and $\mathbf{E}$ were unknown (co)variance matrices of polygenic and residual effects, $\mathbf{X}$, $\mathbf{W_i}$, $\mathbf{Z}$, are known design matrices associating the traits for each son with the fixed and random effects. $\mathbf{A}$ is the numerator relationship matrix and $\mathbf{I}$ an identity matrix. $\mathbf{Q}_{i|M,k}$ is the identity by descent (IBD) matrix of the QTL q, a function of marker data ($\mathbf{M}$) and the position (k_q) of the QTL q on the chromosome (e.g. GEORGE et al., 2000; SØRENSEN et al., 2003). The hypothesis tests H1, H2, H3 and H4 using the LR test again, assuming to be χ^2-distributed, based on five different models (see simplified in Table 1).

Table 1
Models and their parameters specified in the bivariate analysis in Method 3 (Modelle und deren spezifizierte Parameter der bivariaten Analyse in Method 3, SØRENSEN et al., 2003)

Model		QTL related parameters	Model image
L_{11} one QTL for both	$\mathbf{y} = \begin{pmatrix} \mathbf{q_1} \\ \mathbf{q_2} \end{pmatrix} + \begin{pmatrix} \mathbf{u_1} \\ \mathbf{u_2} \end{pmatrix} + \begin{pmatrix} \mathbf{e_1} \\ \mathbf{e_2} \end{pmatrix}$	$K_{11} \neq 0$, $K_{22} \neq 0$, $K_{12}=K_{21} \neq 0$ and k	pleiotropic
L_{10} QTL for FY	$\mathbf{y} = \begin{pmatrix} \mathbf{q_1} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{u_1} \\ \mathbf{u_2} \end{pmatrix} + \begin{pmatrix} \mathbf{e_1} \\ \mathbf{e_2} \end{pmatrix}$	$K_1 \neq 0$ and k_1	simple bivariate models
L_{01} QTL for FC	$\mathbf{y} = \begin{pmatrix} \mathbf{0} \\ \mathbf{q_2} \end{pmatrix} + \begin{pmatrix} \mathbf{u_1} \\ \mathbf{u_2} \end{pmatrix} + \begin{pmatrix} \mathbf{e_1} \\ \mathbf{e_2} \end{pmatrix}$	$K_2 \neq 0$ and k_2	
L_{11} trait specific QTL	$\mathbf{y} = \begin{pmatrix} \mathbf{q_1} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathbf{q_2} \end{pmatrix} + \begin{pmatrix} \mathbf{u_1} \\ \mathbf{u_2} \end{pmatrix} + \begin{pmatrix} \mathbf{e_1} \\ \mathbf{e_2} \end{pmatrix}$	$K_1 \neq 0$, $K_2 \neq 0$, k_1 and k_2	close linkage
L_{00} No QTL	$\mathbf{y} = \begin{pmatrix} \mathbf{u_1} \\ \mathbf{u_2} \end{pmatrix} + \begin{pmatrix} \mathbf{e_1} \\ \mathbf{e_2} \end{pmatrix}$	-	pure polygenic

K_{ij} - element in the genetic (co)variance matrix $\mathbf{K}$ (dimension=2) for traits i and j, K_i (dimension=1) - element in the genetic variance for trait i, k_1 and k_2 - QTL positions for trait i = 1 and j = 2, for hypothesis testing in a likelihood ratio (LR) test: H1: L_{00} and L_{11} with 3 degrees of freedom (df); H2: L_{00} and L_{10} for fat yield (L_{01} for fat content) with 2 df each; H3: L_{00} and L_{11} with 3 df; H4: L_{11} and L_{10} (L_{01}) with 2 df each, if H1 results in a significant finding.

The Akaike information criterion (AIC = - 2 ln MLH + 2 * number of parameters in the model) was used for identifying the more parsimonious model in comparison of the close linkage model and the pleiotropic model, because of the non-nested nature of both models.

Our comparison is not based on a simulated dataset, because these approaches have been verified and reviewed elsewhere. Here, an application to real conditions is of interest for practical demonstration. The focus was on comparing results from analysing fat yield and fat content, which are positively correlated traits, with r_p ranging from 0.23 to 0.55 (a cross section of the international literature) under a model- and method- analytical view. The computations were supported by VPI&SU, USA, and DIAS, DK, on the basis of a collaboration (FREYER et al., 2002, 2003).

Material

We show the analysis based on phenotypic data of 641 sons of five German Holstein sire families, in milk fat traits yield and content. The data was provided by the German Computing Centre for animal breeding purposes, VIT Verden. The standard deviation was 17 kg fat, varying from 11.1 (in family fa 4) to 18.4 (in family fa 2) and 0.28 % fat (varying from 0.22 in family fa 5 to 0.29 in family fa 1), for estimated breeding values

(EBV), respectively. Briefly, the marker map for BTA6 contained 16 micro satellite markers (KÜHN et al., 1999 extended) of the order and map position in cM Kosambi: 0, 16, 40, 43, 58, 59, 64, 66, 67, 71, 76, 79, 89, 95, 102, and 124 cM. The mean number of genotyped sons per sire was 60, with EBV from averagely 278 daughters per son (further details in FREYER et al., 2003).

Results

Single family analyses

For fat yield, we notice an agglomeration of family specific 'spots with highest test statistic values' in the test statistic profile between 62 and 70 cM, exceeding the threshold at 68 cM in family fa 2 (Figure 1). Further, a QTL for fat yield at 95 cM was suggested from fa 1. The same family yielded in a significant QTL estimate for fat content, at 46 cM (Figure 2), whereas the steep peak obtained from fa 3, at 60 cM, was not significant.

The single family QTL analysis of fat yield, based on a two-QTL, model was inferior in fa 2. The two suggested QTL positions confirmed the same QTL (Table 2). The analysis result from fa 3 showed QTL positions at 47 and 62 cM for fat yield, which were suggested to be significant here, and a superiority of the epistasis model.

Across family analyses

Analysing all families together by method 1, we found the QTL position at 46 cM for fat content, as for fa 1. Method 2 placed this QTL at 41 cM, respective 43 cM (from the two-QTL model, Table 3). A quite interesting result from method 2 is, that QTL positions at 41 and 68 cM were estimated for fat yield. The two-QTL model realised by method 3 yielded one significant QTL position at 46 cM and a further QTL suggestion at 95 cM for fat content. The first QTL at 46 cM for fat content was confirmed by the one-QTL model. The confidence intervals CI(95%) obtained by bootstrapping were comparably high, likely because of several family-specific QTL. But this was also reported by others (e.g. RON et al., 2001).
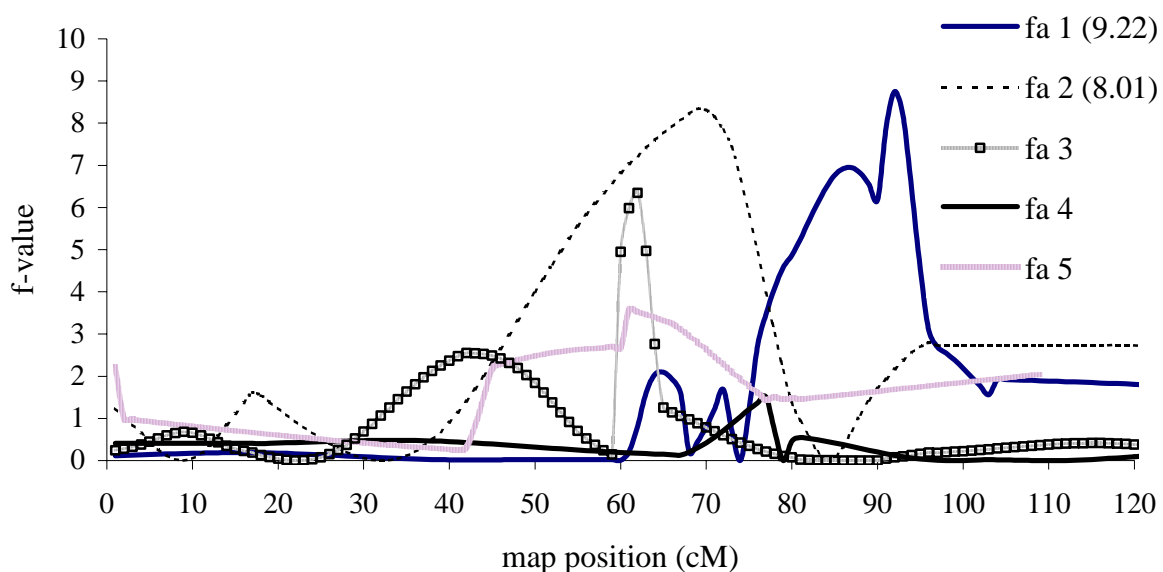

Fig. 1: Test statistic profile for fat yield from single family results (Method 1, threshold presented in brackets in relevant cases, exceeded for family fa 2) (Verläufe der Teststatistik für Fettmenge, Ergebnisse aus der Analyse einzelner Familien nach Methode 1, Signifikanzschwelle in Klammern, überschritten in Familie 2)
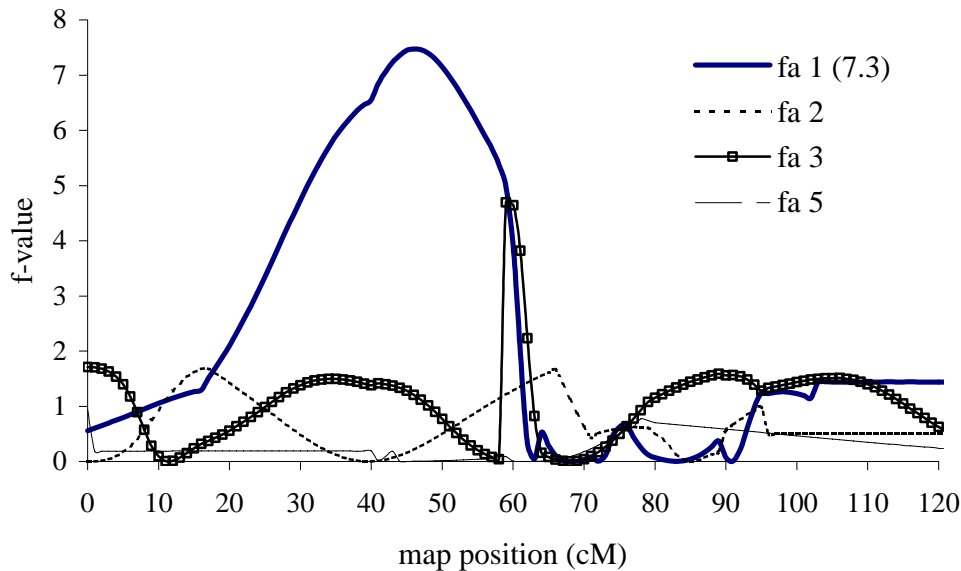
418

FREYER et al.: Different statistical-genetic approaches of QTL detection by Evaluating results from a real dairy cattle data set

Fig. 2: Test statistic profile for fat content from single family results (Method 1, threshold presented in brackets in cases of significance at the p < 0.05 level, reached for familiy fa 1) (Verläufe der Teststatistik für Fettgehalt, Ergebnisse aus der Analyse einzelner Familien nach Methode 1, Signifikanzschwelle in Klammern, überschritten in Familie 1)

Table 2
Single family results of the two QTL model analysis (Method 1) of fat yield FY and fat content FC (Ergebnisse von Einzel-Familien-Analysen mit dem Zwei-QTL-Modell in Methode 1 für Fettmenge FY und Fettgehalt FC)

Trait / family	QTL position (cM)		highest F-value		F-value Epistasis model	CI (95%)	
	QTL 1	QTL 2	QTL 1	QTL 2		QTL 1	QTL 2
FY fa 1	61[3]	91	2.56	4.47	6.51	2 ... 79	40 ...102
fa 2	64	66	4.32	4.71	5.73	0 ... 76	66 ... 95
fa 3	47[2]	62[2]	6.32	9.74	15.58[4]	3 ... 62	45 ...102
fa 4	76	80[3]	1.32	0.42	6.62	1 ... 89	66 ...102
fa 5	42[3]	60	0.96	3.71	4.44	0 ... 89	40 ...102
FC fa 1	47[3,1]	71[3]	11.93	3.92	0.82	17 ... 76	60 ...102
fa 2	64	66	2.12	2.24	4.68	0 ... 89	40 ...100
fa 3	61	102	5.94	3.35	13.63	0 ... 64	49 ...102
fa 4	76	102	5.42	2.73	10.07	1 ... 78	66 ...102
fa 5	16[3,2]	41[3,2]	5.50	5.31	3.66	0 ... 87	41 ... 97

Bold script indicates that both QTL are found to be significant according to the permutation test based on 10000 permutations.
[1] significant result for one position
[2] significant result for both positions
[3] significant epistatic effects
[4] meets the position ± 5cM obtained from the epistatic QTL model

Table 3
Estimated QTL positions (cM) from the across family analyses based on the one- and two- QTL model in yield (FY) and content (FC) (methods 1, 2 and 3) and bootstrap confidence interval CI(95%) from the one-QTL model (Geschätzte QTL-Position in cM, aus Analysen mit Modellen für ein und zwei QTL im Vergleich der Methoden 1, 2, und 3 für Fettmenge FY und Fettgehalt FC sowie bootstrap CI(95%) aus dem Ein-QTL-Modell)

Trait	one-QTL model				two-QTL model	
	Method 1	Method 2	Method 3	CI (95%)	Method 2	Method 3
FY	91	78	68[1]	58 …99	41and 68	66 and 79
FC	46[2]	41[2]	46	5…117	43[2] and 103	46[1] and 95

[1] indicates exceeding of 0.05 threshold (the ln "maximum likelihood" from the two QTL model was tested against the ln "maximum likelihood" from both univariate analyses each, not described in detail)
[2] highest F-value over the test statistic profile along the marked chromosome

Bivariate analyses

The results of bivariate analyses, on both fat traits, in principle confirm the previous significant findings, accompanied by additional information. The profile obtained from the pleiotropic model analysis shows its highest test statistic levels between 46 and 58 cM, suggesting a QTL affecting both traits simultaneously (Figure 3). This area contains the estimated QTL position for fat content. The QTL position for fat yield at 68 cM, significantly suggested by fa 2, was not included.
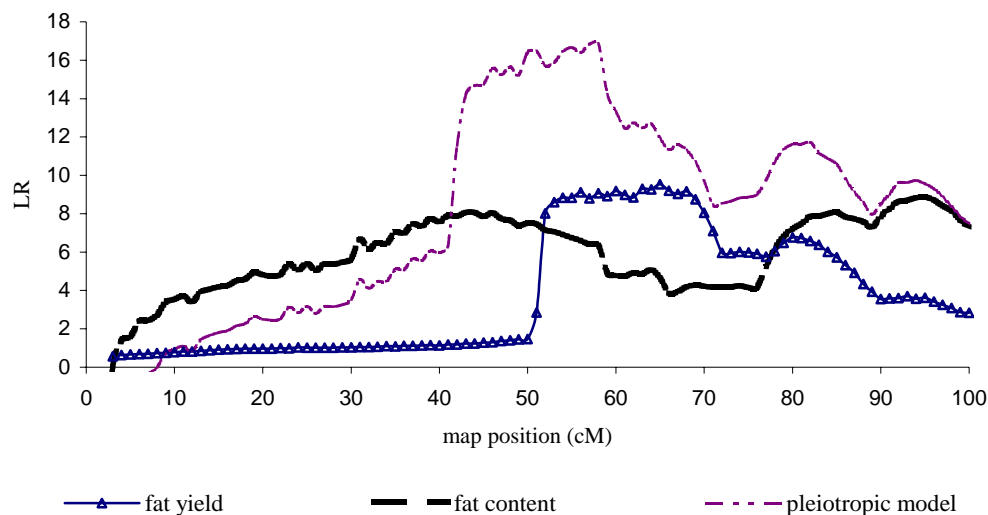

Fig. 3: Test statistics profiles obtained from the multitrait analysis, method 3 (significance for pleiotropic model, results for LR > 15) (Verläufe der Teststatistik aus der multivariaten Analyse, Methode 3, (Signifikanzschwelle für das pleiotrope Modell LR > 15))

The profile of the simple bivariate analysis of fat yield shows a flat course with high test statistic values between 52 and 70 cM, highest at 68 cM. This was covered partly by the profile obtained from the pleiotropic model analysis. The 'common part' of the pleiotropic profile for both traits was found between 52 and 58 cM. The QTL variance components were smallest when the data were analysed 'simply-bivariately' (Table 4). The errors of pleiotropic estimates were fairly high. In case of the close linkage model, the area with higher test statistic values was large for fat yield, roughly covering the region between 45 and 70 cM, and a second QTL for fat content at 95 cM seems to be likely (Figure 4). When comparing the AIC for both decision models (pleiotropic and close linkage model), we notice nearly equal AIC. The basic LR for these models (H1 and H3) were formally significant and similar.

We evaluated the simple- bivariate models to be the models of fit here (Table 5). This means, both 'strong' QTL for each of the single traits, from two different families, 'got through' in the bivariate analysis. There are some additional spots in between the 22 cM distance, likely reflecting 'inferior family specific QTL'. The test of H4 did not fully support a pleiotropic QTL. But the difference was small and the close linkage model received its highest test statistic values at 46 cM (for content of fat) and 50 cM (yield of fat) and also at 95 cM for fat content in combination with 68 cM for fat yield.

Table 4
Relative QTL- variances (var), non QTL variances from the simple bivariate models L_{01} and L_{10}, pleiotropic model L_{11} and close linkage model L_{1-1} at the most likely QTL position, Method 3 (Relative QTL-Varianzen (var), polygene + residuale Varianzen aus den 'einfach-bivariaten' Modellen L_{01}und L_{10}, dem pleiotropic Model L_{11} und dem Modell, enge Kopplung' L_{1-1} für die wahrscheinlichsten QTL-Positionen, Methode 3)

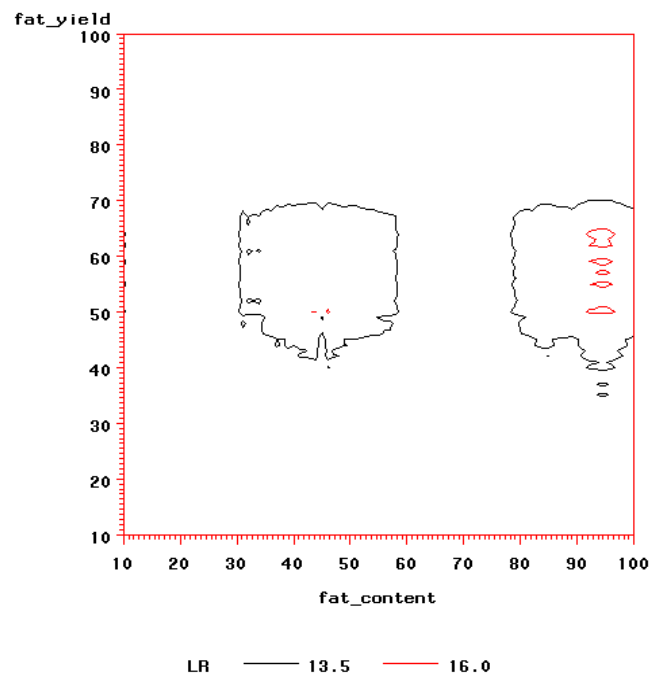| Model	QTL var fat content	non-QTL var fat content	QTL var fat yield	non-QTL var fat yield
$L_{01}$and L_{10}	0.147 ± 0.017	0.853 ± 0.068	0.274 ± 0.025	0.726 ± 0.055
L_{11}	0.458 ± 0.592	0.542 ± 0.088	0.355 ± 0.037	0.645 ± 0.072
L_{1-1}	0.213 ± 0.125	0.787 ± 0.044	0.402 ± 0.008	0.598 ± 0.033

Fig. 4: Two dimensional plot of the test statistic values obtained from the close linkage model analysis Method 3 (x-axis: map position scale for fat content in cM, y-axis: cM for fat yield), LR levels: marginal bounds indicates LR exceeding 13.5, and internal bounds indicates LR exceeding 16.0 (Zweidimensionale Darstellung der Teststatistikresultate aus dem Modell 'enge Kopplung' Methode 3 (x-Achse: Maßstab in cM für Fettgehalt, y-Achse: cM für Fettmenge), LR- Niveau: äußere 'Inseln' kennzeichnen LR > 13,5 und innere 'Inseln' kennzeichnen LR > 16,0)

Table 5
Akaike Information Criterion (AIC) resulting from the pleiotropic L_{11} and close linkage model L_{1-1} (Method 3) of fat yield (FY) and fat content (FC) at QTL positions (Akaike-Informations-Kriterium AIC aus dem pleiotropic Modell L_{11} und dem Modell enge Kopplung L_{1-1} (Methode 3) für Fettmenge FY und Fettgehalt FC an den gefundenen QTL Positionen)

Model	Number of parameters	Trait	Position (cM)	AIC
L_{10}	7	FY	66	**8684**
L_{01}	7	FC	46	**8692**
L_{11}	9	both	58	8710
L_{1-1}	9	both	58	8709
L_{1-1} (best result)	9	FC and FY	46 and 50	**8702**
L_{1-1} (second best result)	9	FC and FY	95 and 68	8715

Bold script indicates that the AIC obtained from the simple bivariate model analyses are in conformance with the best result of the close linkage model analysis.

Discussion

Confirmation studies and fine mapping of QTL are inevitably to avoid conclusions for practical purposes from results yielded by chance. Analyzing data repeatedly on the basis of varying advanced model assumptions and different statistical-genetic methods may be one step towards verifying QTL results. A QTL segregating in all families can not be presupposed. Here, the single families were analysed effectively by adapting the LS method. It enables distinguishing between putative family specific QTL. Furthermore, an overall analysis is better interpretable with knowledge about the single families. Different QTL suggestions between the families showed up in our example, taken from the family specific test statistic profiles (Figure 1 for fat yield, and Figure 2 for fat content) and from some significant results of the two-QTL model analyses (Table 2). For fat yield in fa 2 (significant at 68 cM from the one-QTL model 1), the two QTL model analysis resulted insignificantly in 64 and 66 cM basically for the same QTL. A small distance can not be verified at this level of QTL search. In case of estimating two significant QTL within a relatively small region of 15 cM, we surely may expect epistatic interactions of both QTL, as found in family fa 3 for fat yield (Table 2). But the test statistic profile (Figure 1) shows only one single peak at 62 cM and thus, it is hard to believe, that a second QTL should be located in this region. This particular result should be better interpreted in the way: If there would be two QTL, they show epistatic interactions.

On the basis of three different statistical approaches adapted to a real data set, we can show agreeing results. Mostly, the two-QTL model results were confirming the one position found by the one-QTL model analysis (Table 3). In single family analyses, we found QTL positions at 47 cM, for fat yield in family fa 3 and at 46 cM for fat content in family fa 1. The results of the overall analyses based on method 2 (or method 3) were similar (Table 3), suggesting the QTL at positions 43 cM (fat content) and 68 cM (fat yield). Additionally, we notice estimates for a second fat content QTL at 103 (or 95) cM.

The result of the multivariate analysis was most interesting. The increasing power by using extended information from correlated traits compared to univariate VC analysis was proved e.g. by KOROL et al. (1995) and SØRENSEN et al. (2003). Furthermore, this type of analysis enables discriminating between pleiotropic and closely linked trait specific QTL. The test statistic profile of the simple bivariate analysis of fat yield was flat between 52 and 70 cM (Figure 3) and reflecting the range of family specific QTL estimates. The overlapping area between 52 and 58 cM, suggesting one QTL for both traits, obtained by the pleiotropic model analysis in comparison to the simple bivariate analyses, could not be confirmed, neither by any single family result (Table 2), nor by any methodical result (Table 3). But astonishingly, this particular region is in conformance to other reports: RON et al. (2001) found a QTL for both traits at BM143, corresponding to 59 cM in our marker map. ZHANG et al. (1998) reported a QTL for fat content near TGLA37, at 66 cM. The strongest QTL estimates dominated the close linkage model result partly. The very similar AIC values from the pleiotropic and the close linkage model at 58 cM might be a suggestion of a further (pleiotropic) QTL, that would not be unexpected in the case of these both traits describing milk fat production. The results underline that further investigations, especially on finemapping QTL, are necessary for characterizing the QTL more precisely.

422

FREYER et al.: Different statistical-genetic approaches of QTL detection by Evaluating results from a real dairy cattle data set

References

ALMASY, L.; DYER, T. D.; BLANGERO, J.:
Bivariate quantitative trait linkage analysis: pleiotropy versus co-incident linkages. Genet Epidemiol. **14** (1997), 953-8

BOICHARD, D.; BISHOP, M. D.:
Detection of QTL influencing milk production and mastitis resistence with a granddaughter design in Holstein cattle. Book of abstracts 47th EAAP meeting (1997), G1.2

BOVENHUIS, H.; VAN ARENDONK, J.A.M.; KOVER, S.:
Associations between milk protein polymorphisms and milk production traits. J. Dairy Sci. **75** (1992), 2549-2559

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

DU, F.; HOESCHELE, I.:
Computer Program performing Least Squares Analysis for QTL Detection. VPI&SU DASC Blacksburg, USA (1999)

FREYER, G.; KÜHN, C; WEIKARD, R.; ZHANG, Q.; MAYER, M.; HOESCHELE, I.:
Multiple QTL on Chromosome six in Dairy Cattle affecting Yield and Content Traits. J. Anim Breed. Genet. **119** (2002), 69-82

FREYER, G.; SØRENSEN, P.; KÜHN, C; WEIKARD, R.; HOESCHELE, I:
Search for Pleiotropic QTL on Chromosome BTA6 Affecting Yield Traits of Milk Production. J. Dairy Sci. **86** (2003), 999-1008

GEORGE, A. W.; VISSCHER, P. M.; HALEY, C. S.:
Mapping quantitative Trait Loci in complex Pedigrees: A Two-Step Variance Component Approach. Genetics **156** (2000), 2081-2091

GEORGES,M.; NIELSEN, D.; MACKINNON, M.; MISHRA, A.; OKIMOTO, R.; PASQUINO, A. T.; SARGEANT, L. S.; SORENSEN, A.; STEELE, M. R.; ZHAO, X.; WOMACK, J. E.; HOESCHELE. I.:
Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics **139** (1995), 907-20

GRIGNOLA, F. E.; HOESCHELE, I.; TIER, B.:
Mapping quantitative trait loci via residual maximum likelihood: I Methodology. Genet. Sel. Evol. **28** (1996), 479-490

HOESCHELE, I.:
Mapping Quantitative Trait Loci in Outbred Pedigrees. In Handbook of Statistical Genetics. D.J. Balding Wiley & Sons, 2001

JENSEN, J.; MADSEN, P.:
DMU: A package for the analysis of multivariate mixed models. Proceedings of the 5th World Congress Genet Appl Livest Prod. (1994) 22, 45-46

JENSEN, J; MANTYSAARI, E.; MADSEN, P.; THOMPSON, R.:
Residual maximum likelihood estimation of (co)variance components in multivariate mixed linear models using average information. Journal of Indian Society of Agricultural Statistics **49** (1997), 215-236

KOROL, A. B.; RONIN, Y. I.; KIRZHNER, V. M.:
Interval mapping of quantitative trait loci employing correlated trait complexes. Genetics **140** (1995), 1137-47

KÜHN, C.; FREYER, G.; WEIKARD, R.; GOLDAMMER T.; SCHWERIN, M.:
Detection of QTL for milk production traits in cattle by application of a specifically developed marker map of BTA6. Animal Genetics **30** (1999), 1-8

LIEN, S.; GOMEZ-RAYA, L.; STEINE, T.; FIMLAND, E.; ROGNE, S.:
Associations between Casein Haplotypes and milk yield traits. J. Dairy Sci. **78** (1995), 2047-2057

LIU, Z.:
Markergestützte komplexe Segregationsanalyse der Milchleistungsmerkmale beim Rind. Univ. Hohenheim, Diss., 1994

ORTNER, M.; ESSL, A.; SÖLKNER, J.:
Zur züchterischen Bedeutung verschiedener Milchproteingenotypen beim Rind. Züchtungskunde **67** (1995), 353-367

PANICKE, L.; FREYER, G.; ERHARDT, G.; SCHLETTWEIN, K.:
Schätzung direkter Allel- und Genotypeneffekte der Milchproteinvarianten auf Milchleistungsmerkmale. Arch. Tierz., Dummerstorf **39** (1996), 3-16

RON, M.; KLINGER, D.; FELDMESSER, E.; SEROUSSI, E;. EZRA, E.; WELLER, J.I.:

     Multiple Quantitative Trait Locus Analysis of Bovine Chromosome 6 in the Israeli Holstein Population by a Daughter Design. Genetics **159** (2001), 727-735

SØRENSEN, P.; LUND, M. S.; GULDBRANDTSEN, B.; SORENSEN, D. A.; JENSEN, J.:
     Comparison of bivariate and univariate QTL mapping in livestock populations. Accepted by GSE (2003)

SPELMAN, R. J.; COPPIETERS, W.; KARIM, L.; VAN ARENDONK, J. A. M.; BOVENHUIS, H:
     Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population. Genetics **144** (1996), 1799-1808

VAN ARENDONK, J.A.M.; BOVENHUIS, H.; VAN DER BEEK, S.; GROEN, A.F.:
     Detection and exploitation of markers linked to quantitative traits in farm animals. Proc. 5th WCGALP **21** (1994), 193-200

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

ZHANG, Q.; HOESCHELE, I.:
     Computer program performing REML Analysis for QTL Detection. VPI & SU DASC Blacksburg, USA (1998)

ZHANG, Q.; BOICHARD, D.; HOESCHELE, I.; ERNST, C.; EGGEN, A.; MURKVE, B.; PFISTER-GENSKOW, M. et al.:
     Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. Genetics **149** (1998), 1959-1973

Corresponding author
Dr. GERTRAUDE FREYER
Research Institute for the Biology of Farm Animals
Wilhelm-Stahl-Allee 2
D-18196 Dummerstorf
Germany

E-Mail: freyer@fbn-dummerstorf.de