Genet. Sel. Evol. 35 (2003) 319–338

DOI: 10.1051/gse:2003011

Original article

Combined analysis of data from two granddaughter designs: A simple strategy for QTL confirmation and increasing experimental power in dairy cattle

Jörn Bennewitz[a*], Norbert Reinsch[a], Cécile Grohs[b], Hubert Levéziel[b], Alain Malafosse[c], Hauke Thomsen[a], Ningying Xu[a], Christian Looft[a], Christa Kühn[d], Gudrun A. Brockmann[d], Manfred Schwerin[d], Christina Weimann[e], Stefan Hiendleder[e], Georg Erhardt[e], Ivica Medjugorac[f], Ingolf Russ[f], Martin Förster[f], Bertram Brenig[g], Fritz Reinhardt[h], Reinhard Reents[h], Gottfried Averdunk[i], Jürgen Blümel[j], Didier Boichard[k], Ernst Kalm[a]

[a] Institut für Tierzucht und Tierhaltung, Christian-Albrechts-Universität, 24098 Kiel, Germany
[b] Laboratoire de génétique biochimique et de cytogénétique, Institut national de la recherche agronomique, 78352 Jouy-en-Josas Cedex, France
[c] Union nationale des coopératives d'élevage et d'insémination animale, 149 rue de Bercy, 75595 Paris Cedex 12, France
[d] Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere, 18196 Dummerstorf, Germany
[e] Institut für Tierzucht und Haustiergenetik der Justus-Liebig-Universität, 35390 Gießen, Germany
[f] Institut für Tierzucht der Ludwig-Maximilians-Universität, 80539 München, Germany
[g] Institut für Veterinärmedizin der Georg-August-Universität, 37073 Göttingen, Germany
[h] Vereinigte Informationssysteme Tierhaltung w.V., 27283 Verden, Germany
[i] Bayerische Landesanstalt für Tierzucht, 85586 Grub, Germany
[j] Institut für die Fortpflanzung landwirtschaftlicher Nutztiere, 16321 Schönow, Germany
[k] Station de génétique quantitative et appliquée, Institut national de la recherche agronomique, 78352 Jouy-en-Josas Cedex, France

(Received 14 June 2002; accepted 5 December 2002)

* Correspondence and reprints
E-mail: jbennewitz@tierzucht.uni-kiel.de

Abstract – A joint analysis of five paternal half-sib Holstein families that were part of two different granddaughter designs (ADR- or Inra-design) was carried out for five milk production traits and somatic cell score in order to conduct a QTL confirmation study and to increase the experimental power. Data were exchanged in a coded and standardised form. The combined data set (JOINT-design) consisted of on average 231 sires per grandsire. Genetic maps were calculated for 133 markers distributed over nine chromosomes. QTL analyses were performed separately for each design and each trait. The results revealed QTL for milk production on chromosome 14, for milk yield on chromosome 5, and for fat content on chromosome 19 in both the ADR- and the Inra-design (confirmed within this study). Some QTL could only be mapped in either the ADR- or in the Inra-design (not confirmed within this study). Additional QTL previously undetected in the single designs were mapped in the JOINT-design for fat yield (chromosome 19 and 26), protein yield (chromosome 26), protein content (chromosome 5), and somatic cell score (chromosome 2 and 19) with genomewide significance. This study demonstrated the potential benefits of a combined analysis of data from different granddaughter designs.

QTL mapping / granddaughter design / combined analysis / QTL confirmation / dairy cattle

1. INTRODUCTION

With the aid of genetic markers, it was possible in several studies to detect quantitative trait loci (QTL) involved in the variation of traits of economic interest. In dairy cattle, most QTL experiments used a granddaughter design [4, 7,23,25,33], where the number of sires genotyped in each family was typically below 150. The power to detect a QTL present in a granddaughter design is largely influenced by the number of families included in the experiment and by the size of the individual families [30]. Consequently, increasing family size in a granddaughter design is desirable but in many cases has its limitations in the availability of progeny tested sires and in the costs of determining genotypes.

Although the substitution effect estimates of the detected QTL tend to be overestimated [8], the most detected QTL are of sufficient magnitude to consider them in marker assisted selection (MAS), especially in preselection of young bulls entering progeny testing [13,18]. However, Lander and Kruglyak [16] postulated that a detected marker-QTL linkage must be replicated to be credible. Similarly Spelman and Bovenhuis [24] suggested that a QTL confirmation study prior to starting MAS should be conducted in order to prevent a selection for a non-existing QTL.

Therefore it could be useful to combine data from different experiments that use the same experimental design and the same or closely related breeds. The potential benefits of the extraction of additional information could be substantial: a higher experimental power to detect QTL, especially if they have a small phenotypic effect, a confirmation of QTL previously detected in only one experiment, and more precise conclusions about the QTL position. Walling *et al.* [28] mapped QTL in seven different F2 crosses with altogether

almost 3000 pigs. The QTL analysis was conducted for three different traits on chromosome 4 either separately for each individual cross, or jointly for the combined data set. Their results emphazised the potential benefit of a joint analysis of data from independent mapping experiments although the families used were nested within the population. Compared to other species, dairy cattle are particularly suitable for a joint analysis of data from QTL experiments, since many Holstein sires, but also Simmental and Brown Swiss sires, are used across countries and therefore may be included in several different granddaughter designs. In a combined analysis, not only the total size of the granddaughter design is expanded, but also the individual family size may be increased, and, therefore, the experimental power is significantly higher [30]. Furthermore, QTL can be confirmed when specific families are included in different designs, because the probability that the same families show a type one error in several QTL experiments is low when the level of significance is sufficiently stringent. Prior to a joint analysis, however, several problems have to be solved: the phenotypes of individuals in the different designs might not be comparable due to the different environments of the individuals and/or different calculation of phenotypic parameters like estimated breeding values (*e.g.* different models based on data from different testing schemes). Different markers might have been genotyped in different designs and/or the marker genotypes could have been recorded by different techniques.

In this study we conducted a joint analysis of five Holstein families that are included in two different granddaughter designs described by Thomsen *et al.* [25] and Boichard *et al.* [3]. Nine chromosomes and six traits were investigated. The first aim of this study was QTL confirmation, where the data were analysed separately and a confirmed QTL should show a significant effect in both data sets. A second aim was a joint analysis of the two data sets in order to increase family size and, hence, statistical power to detect further QTL.

2. MATERIALS AND METHODS

2.1. Pedigree

Five paternal half-sib Holstein families were selected that are included in two different granddaughter designs. The first design was part of the common QTL mapping project of German AI and breeding organisations, several German animal breeding institutes and animal computing centres initiated by the German cattle breeders federation (ADR) [25]. Final results of this project will be published elsewhere. The second design was part of the QTL mapping project of several French AI breeding organisations and Inra with results published by Boichard *et al.* [3]. The sires were progeny-tested either in Germany (subsequently denoted as the ADR-design) or in France (referred

Table I. Size of the five Holstein paternal half-sib families included in the experiment.

Family	ADR-design	Inra-design	JOINT-design
F1	42	102	144
F2	50	83	133
F3	127	95	222
F4	126	168	294
F5	126	236	362
Totals	471	684	1155

ADR-design: sons that were progeny-tested in Germany; Inra-design: sons that were progeny-tested in France; JOINT-design: ADR-design and Inra-design merged.

to as the Inra-design) but no sire was progeny-tested in both countries. The combined data of both designs is denoted as the JOINT-design. For the family sizes see Table I.

2.2. Marker data

For the present study nine chromosomes were selected from the whole genome scan data sets. According to the literature reports and to our own experience, these chromosomes were of special interest to us. On these chromosomes, the individuals of the ADR-design were originally genotyped for 75 microsatellite markers, three single strand conformation polymorphisms (SSCP) and the EAC blood group system. The individuals of the Inra-design were genotyped for 53 microsatellites and the EAC blood group system on the same chromosomes. Microsatellite and SSCP genotypes were determined in both designs by automated fragment analysis (ALF, Amersham-Pharmacia or ABI377, Perkin-Elmer). The routine blood typing laboratories (two in Germany and one in France) determined the genotypes for the EAC system according to standard procedures. Only 21 markers including EAC were genotyped in both experiments making the derivation of haplotypes of the five grandsires for all markers difficult. Therefore, to generate a number of co-informative meioses between markers genotyped in the ADR-design and markers genotyped in the Inra-design, roughly 30 sires of each family of the ADR-design were additionally genotyped for markers that previously had been genotyped in the Inra-design only. These sires were genotyped for all markers included in the experiment. According to our practical experience this number is sufficient to derive the haplotypes of a sire in a half-sib design. All alleles from microsatellites were coded as follows: short paternal allele '1'; long paternal allele '2'; and all deviating maternal alleles '3'. A unified coding of the EAC blood group marker was done with haplotype analysis of the grandsires. Allele coding was performed to avoid a common standardisation

of the genotypes and to ensure the anonymity of the individuals. All marker data were stored in the ADRDB database [20] and were checked for their agreement with the Mendelian laws of inheritance. Multipoint marker maps were computed using CRIMAP [9]. For chromosomes, markers, and the genetic maps see Table II. The marker order is in agreement with maps published previously [12].

Table II. Chromosomes (BTA) included in the experiment and genetic maps.

BTA	Marker, distance from the start of the chromosome in cM
02	*TGLA44 0.0*, BM3627 5.0, **TGLA431 13.9**, CSFM50 23.7, *TGLA377 36.8*, **CSSM42 44.9**, BMS1300 59.3, ILSTS98 74.5, *ILSTS82 78.6*, BMS778 90.4, *MM8 113.0*, TGLA110 113.1, BMS1987 124.9, Inra*135 127.3*, BM2113 135.2, Inra*231 136.7*, IDVGA2 157.8
05	**BM6026 0.0**, **RM103 27.3**, CSSM34 42.9, *RM500 60.0*, Lysmic 69.0, ETH10 76.3, *CSSM022 84.6*, RM29 98.3, BM1248 112.9, BM2830 133.8, **BM315 149.4**, *ETH2 167.6*, *ETH152 177.9*
06	**ILSTS93 0.0**, ILSTS90 13.0, URB16 38.6, **BM1329 41.1**, **DIK82 57.6**, IL97 79.5, FBN14 86.5, Inra*K 92.0*, CSN3 92.01, BP7 106.6, BMC4203 127.1, *BM2320 135.2*
14	KIEL_E8 0.0, ILSTS039 1.3, **CSSM66 9.0**, RM180 42.8, BM4630 57.6, **RM11 58.8**, RM192 77.4, *BMS1899 87.9*, BM4513 118.1, BL1036 147.5, Inra*092 148.7*, Inra*100 148.71*
18	**IDVGA31 0.0**, TGLA357 16.9, *ABS13 23.8*, *HAUT14 55.1*, BM7109 62.6, **ILSTS002 70.6**, *IDVGA55 94.7*, **EAC 109.7**, BM2078 117.6, **TGLA227 141.2**
19	*BM9202 0.0*, ILSTS73 2.0, **HEL10 17.9**, **URB44 43.6**, UWCA40 61.3, *BMS2389 61.6*, BM17132 67.1, URB32 72.4, DIK39 75.5, CSSM65 78.9, FBN501 82.0, **BMS1069 88.7**, *ETH3 92.0*, RM388 102.5, **BMC1013 110.7**
20	*HEL12 0.0*, BM3517 0.5, **BM1225 5.9**, *TGLA126 24.7*, BM713 30.1, *ILSTS072 43.5*, BM5004 66.9, *UWCA26 79.0*
23	*IOBT528 0.0*, Inra132 10.3, Inra064 10.4, BM47 26.2, **RM033 34.8**, UWCA1 44.8, BM1258 49.2, DRB3 60.1, *BOLADRBP 64.1*, BTN 64.2, MB026 64.3, MB025 65.0, CYP21 65.7, *BM7233 76.3*, BM1818 78.4, **BM1905 96.6**
26	*ABS12 0.0*, BMS651 5.0, *BMS907 13.9*, BM1314 24.6, Inra*081 27.6*, RM26 37.2, BM4505 41.1, BM6041 50.4, *IDVGA59 51.9*

Markers genotyped in both the ADR-design and Inra-design are written bold-face and underlined, markers genotyped only in the ADR design are written regularly. Markers genotyped in the Inra-design and additionally for a number of sons from the ADR-design are written in italics. The numbers indicate the chromosomal position of the markers.

2.3. Phenotypic data

Six traits were included: milk yield (MY), fat yield (FY), protein yield (PY), fat content (FC), protein content (PC), and somatic cell score (SCS). For the Inra-design, daughter yield deviations (DYD) were used for all six traits. DYD were by-products of the French genetic evaluation system based on a single trait repeatability animal model. For the ADR-design DYD were not available, therefore, estimated breeding values calculated with a BLUP animal model for all six traits were taken from the national sire evaluation and were de-regressed as described by Thomsen *et al.* [26]. To ensure the anonymity of the sires, the phenotypes were expressed in genetic standard deviation units as provided by each country and were slightly rounded (two decimal places). For further analysis the within family and design mean (ADR-design and Inra-design, respectively) was set to zero. It turned out that the phenotypes of the Inra-design were in general somewhat more variable than the phenotypes of the ADR-design (Tab. III).

2.4. Statistical analysis

The QTL analyses were performed for each trait separately across families with the programs BIGMAP and ADRQLT [20] in Germany and

Table III. Standard deviations, minimum and maximum of phenotypic parameters by design.

Trait	Design	Standard dev.	Minimum	Maximum
Milk yield	ADR	0.93	−2.58	3.00
	Inra	1.13	−3.97	3.93
Fat yield	ADR	0.84	−2.46	2.65
	Inra	1.30	−5.56	3.74
Protein yield	ADR	0.92	−2.83	2.82
	Inra	1.08	−3.88	3.36
Fat content	ADR	0.93	−2.47	2.88
	Inra	1.09	−2.87	3.29
Protein content	ADR	0.96	−2.71	3.00
	Inra	0.90	−3.08	2.73
Somatic cell score	ADR	0.92	−2.77	2.56
	Inra	1.07	−3.26	2.94

Phenotypes are expressed in genetic standard deviation units. The mean was set to zero. ADR-design: sons that were progeny-tested in Germany; Inra-design: sons that were progeny-tested in France; JOINT-design: ADR-design and Inra-design merged.

QTLMAP [6] in France. Both softwares are based on the multimarker regression approach [14]. For each cM on the chromosome, the phenotypes were regressed on the corresponding QTL transition probabilities and the position with the highest test statistic (F-ratio) was taken as the most likely position of the QTL on each chromosome. The following model was applied:

$$y_{ijk} = gs_i + b_{ik} * tp_{ijk} + e_{ijk}$$

where y_{ijk} is the trait value of the jth sire of the ith grandsire, gs_i is the fixed effect of the ith grandsire, b_{ik} is the regression coefficient for the ith grandsire at the kth chromosomal location, tp_{ijk} is the probability of the jth sire receiving the chromosomal segment for gamete one (gamete numbers were randomly assigned) from the ith grandsire at the kth chromosomal position, and e_{ijk} is the random residual. Test statistic critical values were calculated empirically for each trait separately using the permutation method [4]. Briefly, by shuffling the phenotypic data 10 000 times randomly while keeping the marker data constant the genotype-phenotype association was uncoupled and, hence, after applying the mapping procedure every QTL estimate indicated a type I error per definition. The chromosomewise critical values $\alpha = 1\%$, 5%, and 10% were calculated by taking the 99th, 95th, and 90th quantile from the corresponding distribution of the test statistic, respectively. Because only nine chromosomes were analysed, the genomewise significance levels were computed using the Bonferroni correction assuming 30 chromosomes:

$$p_g = 1 - (1 - p_c)^{30}$$

where p_g (p_c) is the genomewise (chromosomewise) error probability. Additionally, the QTL position estimates from each evaluated permuted data set were recorded for permutation bootstrapping. The distribution of the QTL position estimates along each chromosome was termed the *null distribution*. A single family was assumed as heterozygous at a significant QTL when it showed a significant haplotype contrast ($P \leq 0.05$, t-test) at the estimated position from the across family analysis.

For each QTL position that deemed to be significant ($p_c \leq 0.05$) a noncentral confidence interval was computed with permutation bootstrapping [2]. First 250 bootstrap samples were generated and analysed [27]. The distribution of the QTL position estimates along the chromosome was termed *linkage distribution* and was corrected for the marker impact in a second step. This was done by dividing the frequency of each chromosomal position of the linkage distribution by the corresponding frequency of the null distribution resulting in the marker *corrected distribution*. From this distribution, noncentral 95% confidence intervals were computed with an analogous highest posterior density method [2].

To account for the problem of multiple testing across traits, the expected false discovery rate (FDR), *i.e.* the expected proportion of a true null hypothesis within the class of rejected null hypotheses [31], was computed for all 54 hypotheses tested ($m = 54$, 6 traits $\times$ 9 chromosomes). The FDR was computed for the ordered set of chromosomewise error probabilities $p_c(1) \leq \cdots p_c(m)$ for each chromosomewise error probability $p_c(i)$ $(i = 1, \ldots, m)$ as:

$$q(i) = mp_c(i)/i.$$

The FDR can be controlled at some level q^* by determining the largest rank i for which: $q^* < mp_c(i)/i$. The results of the calculated FDR can be used as a guide to present a ranking of significant chromosomes, which could be suggested for a further analysis, *e.g.* fine mapping.

All statistical analyses were done once with the ADR-design and once with the Inra-design to carry out a QTL confirmation study. Furthermore, the same statistics were applied to the JOINT-design to detect QTL that were not declared as significant in the other two. This could be done because the additional genotypings of the sons from the ADR-design with markers used only in the Inra-design so far ensured a high accuracy of the sires haplotype derivation.

3. RESULTS

3.1. General results

The results will be presented only for the analysis carried out in Germany with the BIGMAP and ADRQLT programmes, which were essentially equal to those obtained in France with the QTLMAP programme. A QTL was considered as chromosomewise (genomewise) significant when the error probability p_c (p_g) was ≤ 0.05. In Table IV all significant QTL together with the error probabilities p_c and p_g, the estimated location, the 95% confidence interval, and the number of heterozygous families are presented. Note that p_g was only calculated when p_c was below 0.03. Significant QTL were found for all traits and all three designs, but no QTL for any trait could be mapped on the chromosomes 6 and 20. The highest number of QTL (genomewise significance) was found in the JOINT-design, followed by the Inra-Design, and the fewest in the ADR-design. The number of significant heterozygous families was in general 2 or 3 out of 5. The sign of the QTL effect of significant families was in the same direction in all three designs. When comparing confidence intervals of different designs for the same significant trait $\times$ chromosome combination, suprisingly they were in general the largest for the JOINT-design. For a few QTL with an estimated position at the start or at the end of the chromosome, the confidence interval did not include the estimated position. For an example see the QTL for SCS on chromosome 18 in

the ADR-design (Tab. IV). Confidence intervals computed with the classical bootstrap method [27] were always larger (not shown).

3.2. QTL-results for milk production traits

Chromosomewise significant QTL were found on chromosome **5** for MY in all three designs, for FY in the ADR-design, for FC for the JOINT-design, and additionally, a genomewise significant QTL for PC was found in the JOINT-design. Chromosome **14** harboured highly significant QTL for all three designs. The estimated positions were always at the start of the chromosome at position 0 cM for the QTL detected in the ADR-design, whereas for the Inra- and the JOINT-design it was in some cases a few cM closer to the middle part. However, the plot of the test statistic for all QTL detected on chromosome 14 and all designs were very similar, being on a high level in the first part and then dropping down rapidly between 30 and 50 cM (not shown). The sign of the QTL effects for the heterozygous families for PY and FY were in the opposite direction, *i.e.* the paternal haplotype that increased PY lowered FY. Genomewise significant QTL were found on chromosome **19** for FY in the Inra- and the JOINT-design at a very similar position. For FC, genomewise QTL were found on the same chromosome in the Inra- and the JOINT-design and a chromosomewise QTL in the ADR-design, with a similar shape of the test statistic (see Fig. 1). On chromosome **23** only chromosomewise significant QTL were found. The ADR-design showed an effect for MY and PY at similar positions (around 60 cM). Additionally the JOINT design showed an effect on this chromosome for PC with an estimated position at the start of the chromosome. In the Inra- and the JOINT-design, genomewise QTL were mapped on chromosome **26** for MY, FY, and PY, and the JOINT-design showed an additional chromosomewise significant effect for PC. The estimated QTL positions were 51 cM and around 23 cM for the Inra-design and the JOINT-design respectively. The plots of the test statistics for MY and chromosome 26 are presented in Figure 1.

3.3. QTL-results for somatic cell score

Chromosome **2** harboured a chromosomewise significant QTL for SCS in the ADR-design and a genomewise significant QTL in the JOINT-design. Note that the Inra-design showed an effect with $p_c = 0.1$ for SCS on the same chromosome. For the plots of the test statistics see Figure 1. The ADR-design showed a genomewise significant QTL for SCS on chromosome **18**, but no effect could be observed in the Inra- and the JOINT-design on this chromosome. In contrast, the Inra- and the JOINT-design showed a genomewise significant effect for SCS on chromosome **19** whereas no significant effects could be observed for SCS in the ADR-design on this chromosome.

Table IV. Results of QTL analyses with chromosomewise (p_c) and genomewise (p_g) error probabilities, estimated QTL location, 95% confidence interval (CI95) and the heterozygous families.

(continued on the next page)

BTA	Trait[a]	Design[b]	p_c	p_g	Location[c] (cM)	CI95 (cM)	Heterozygous Families
02	SCS	ADR	0.01	0.32	100	0–128	F3, F5
	SCS	JOINT	< 0.01	0.02	99	6–157	F3, F5
05	MY	ADR	0.05		109	7–150	F2, F4
	MY	Inra	0.05		157	7–168	F1, F2
	MY	JOINT	0.01	0.26	158	1–168	F1, F2
	FY	ADR	0.02	0.45	151	2–167	F2, F5
	FC	JOINT	0.05		154	28–149	F1
	PC	JOINT	< 0.01	0.03	42	13–149	F4
14	MY	ADR	< 0.01	< 0.01	0	1–78	F3, F4
	MY	Inra	< 0.01	< 0.01	7	0–7	F2, F3, F4
	MY	JOINT	< 0.01	< 0.01	0	1–119	F2, F3, F4
	FY	ADR	< 0.01	< 0.01	0	0–77	F3, F4
	FY	Inra	< 0.01	< 0.01	7	0–8	F3, F4
	FY	JOINT	< 0.01	< 0.01	8	8–97	F3, F4
	PY	ADR	0.03		0	2–78	F4
	PY	JOINT	< 0.01	0.14	0	0–87	F2, F4
	FC	ADR	< 0.01	< 0.01	0	0–2	F2, F3, F4, F5
	FC	Inra	< 0.01	< 0.01	0	0–7	F2, F3, F5
	FC	JOINT	< 0.01	< 0.01	9	9–9	F2, F3, F5
	PC	ADR	< 0.01	< 0.01	0	0–86	F3, F4
	PC	Inra	< 0.01	< 0.01	29	0–113	F2, F3, F4
	PC	JOINT	< 0.01	< 0.01	0	0–148	F2, F3, F4
18	SCS	ADR	< 0.01	0.05	141	11–137	F2, F3, F4
19	SCS	Inra	< 0.01	0.04	32	3–108	F3, F4
	SCS	JOINT	< 0.01	0.05	50	1–105	F3, F4
	FY	Inra	< 0.01	< 0.01	59	57–109	F4
	FY	JOINT	< 0.01	< 0.01	61	1–105	F2, F4
	FC	ADR	0.01	0.26	77	26–102	F2, F3
	FC	Inra	0.04		65	10–110	F2, F3
	FC	JOINT	< 0.01	< 0.01	76	1–108	F2, F4
23	MY	ADR	0.01	0.26	55	44–94	F4, F5
	PY	ADR	0.03		69	12–90	F4, F5
	PC	JOINT	0.01	0.26	4	1–95	F4, F5

Table IV. Continued.

BTA	Trait[a]	Design[b]	p_c	p_g	Location[c] (cM)	CI95 (cM)	Heteroygous Families
26	MY	Inra	< 0.01	0.06	51	6–51	F1, F5
	MY	JOINT	< 0.01	0.02	25	4–50	F1, F5
	FY	Inra	< 0.01	< 0.01	51	13–51	F1, F2, F5
	FY	JOINT	< 0.01	< 0.01	21	6–50	F1, F5
	PY	Inra	< 0.01	< 0.01	51	9–51	F1, F5
	PY	JOINT	< 0.01	0.05	23	4–50	F1, F5
	PC	JOINT	0.03		25	4–50	F2, F4

The results are listed for QTL that show a chromosomewise error probability $p_c \leq 0.05$. Genomewise error probabilities p_g were only computed when p_c was below 0.03. [a] MY: milk yield; FY: fat yield; PY: protein yield; FC: fat content; PC: protein content; SCS: somatic cell score. [b] ADR-design: sons that were progeny-tested in Germany; Inra-design: sons that were progeny-tested in France; JOINT-design: ADR-design and Inra-design merged. [c] Estimated QTL location (in cM map position).

4. DISCUSSION

4.1. Number of QTL

A critical question in QTL mapping studies is how many of the statistical significant QTL represent real QTL rather than a type I error. Assuming six uncorrelated traits and a genomewise error probability p_g of 0.05, the number of QTL that would be expected to be genomewise significant by chance under the null hypothesis (H_0) of no linked QTL is below 1 in each design. The number of expected QTL under H_0 would be even lower when taking the correlation of the traits into account [23, 33]. The number of actual QTL found on a genomewise significance level was above the expected number under H_0 in every design (Tab. IV). Controlling the FDR at a stringent level of 0.01, *i.e.* 1 out of 100 rejected null hypotheses are true, revealed 1 QTL of the ADR-design, 4 QTL of the Inra-design and 6 QTL of the JOINT-design (Tab. V). Note that with respect to Winter *et al.* [32] and Grisart *et al.* [10], it was assumed that the QTL of milk production traits on chromosome 14 was a single one. Additionally, because of the very similar shape of the test statistic for the QTL of the milk production traits on chromosome 26 (not shown) and the known genetic correlation of the five traits, these QTL were also considered as a single one. Three different QTL showed at least chromosomewise effects in all three designs (Tab. IV). Additionally, some families showed a significant effect in all three designs for QTL ($P < 0.05$) that were chromosomewise significant only in one or two designs in the initial analysis (not shown).

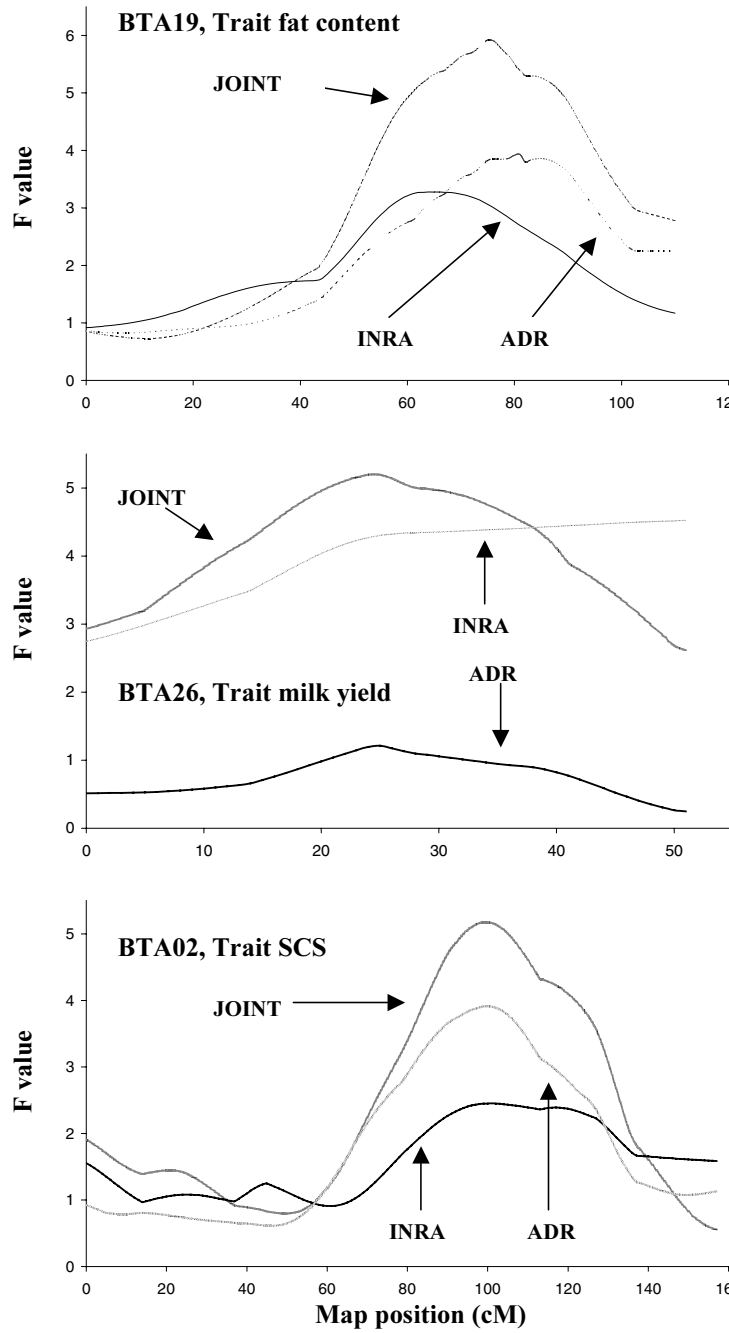

Figure 1. Test statistic profiles for all three designs (indicated by arrows) for fat content and chromosome 19 (top), milk yield and chromosome 26 (middle), and somatic cell score (SCS) and chromosome 2 (bottom).

Table V. Computation of the false discovery rate (q) across traits within each design.

	ADR-design[a]			Inra-design[a]			JOINT-design[a]		
i[b]	BTA	Trait[c]	q	BTA	Trait	q	BTA	Trait	q
1	14	MY	0.000	14	MY	0.000	14	MY	0.000
2	14	FC	0.000	26	FY	0.000	14	FY	0.000
3	14	PC	0.000	14	FC	0.000	19	FY	0.000
4	18	SCS	0.031	14	PC	0.000	14	FC	0.000
5	14	FY	0.031	14	FY	0.001	14	PC	0.000
6	19	FC	0.110	19	FY	0.001	19	FC	0.001
7	23	MY	0.131	26	PY	0.001	26	FY	0.002
8	02	SCS	0.158	19	SCS	0.010	26	PY	0.003
9	05	FY	0.148	26	MY	0.011	26	MY	0.003
10	14	PY	0.141	19	FC	0.242	02	SCS	0.006
11	23	PY	0.163	05	MY	0.241	19	SCS	0.006
12	05	MY	0.221	26	FC	0.437	14	PY	0.006
13	19	MY	0.335	05	FC	0.430	05	PC	0.039
14	05	PY	0.362	05	PC	0.409	05	MY	0.055
15	23	PC	0.364	23	PC	0.487	23	PC	0.057
16	06	FY	0.345	26	PC	0.475	26	PC	0.073
17	19	FY	0.372	14	PY	0.473	05	FC	0.129
18	19	PC	0.356	02	SCS	0.480	19	PC	0.154
19	02	PY	0.417	23	FY	0.513	23	MY	0.150
20	19	PY	0.419	19	PY	0.520	19	PY	0.168
21	06	PC	0.435	18	FY	0.509	23	FY	0.212
22	23	FY	0.428	20	PC	0.490	06	SCS	0.212
23	06	PY	0.498	06	SCS	0.584	18	PY	0.245

[a] ADR-design: sons that were progeny-tested in Germany; Inra-design: sons that were progeny-tested in France; JOINT-design: ADR-design and Inra-design merged.
[b] Rank. [c] MY: milk yield; FY: fat yield; PY: protein yield; FC: fat content; PC: protein content; SCS: somatic cell score.

As expected, the number of QTL that appeared to be genomewise significant increased with increasing size of the design. This was attributed to the increased experimental power of the Inra-design compared to the ADR-design, and of the JOINT-design compared with the two other designs. Additionally the more variable phenotypes of the Inra-design (Tab. II) resulted in a larger haplotype contrast compared to the ADR-design. The reasons for the higher variability could not be clarified completely. A possible explanation is the different unit

of measurement in the two designs (DYD *vs.* de-regressed estimated breeding values). Because of the higher power of the Inra-design, it dominated the JOINT-design. As a consequence all 6 chromosomewise significant QTL of the Inra-design, but only 4 of the ADR-design, were also detected in the JOINT-design. However, the number of detected QTL that were only chromosomewise significant was greater for the ADR-design than for the Inra-design. When the FDR was controlled at a statistically lower level, *e.g.* 0.2, it revealed 7 QTL for the ADR-design and only 4 QTL for the Inra-design (Tab. V). In our opinion, however, it seems to be appropriate for this point to put more emphasis on the stringent significance level of genomewise significance.

4.2. Comparison of QTL results across designs and with literature reports

For QTL confirmation, no fixed rules concerning the significance level have been established, mainly because the choice of an appropriate null hypothesis is unclear and the aim of the confirmation study can differ [24]. The aim of our confirmation study was to confirm that a significant QTL is a real effect. Therefore, we declared a QTL as confirmed within this study, if it showed a chromosomewise significant effect of $p_c \leq 0.05$ in both the ADR- and the Inra-design within similar chromosomal regions, although we are aware that this is a stringent criterion. Due to the applied stringent significance level for confirmation, the conclusion that a significant not confirmed QTL is a type I error should be treated with caution as long as single families show a significant effect in the same direction in both designs. The QTL results were also compared to some extent with literature reports, keeping in mind that an across study comparison of QTL results is difficult because different studies used different designs with different sets of markers and different levels of significance.

Chromosome 2: The chromosomewise significant QTL for **SCS** in the ADR-design was not confirmed by the Inra-design, this design showed only a p_c of 0.1. The large F5 family, however, was identified as QTL heterozygous ($P < 0.01$) in both designs, and the plots of the test statistics (Fig. 1) looked very similar with the highest value at 100 cM for both designs. The JOINT-design showed an increased test statistic (Fig. 1), which resulted in a genomewise significance. As far as we know, this QTL has not been reported previously. One of the most important findings of this study was this QTL, because SCS is very attractive for MAS due to the high economic importance of the correlated trait, mastitis, and due to its low heritability.

Chromosome 5: The chromosomewise significant QTL for **MY** on chromosome 5 was found to be significant in all three designs with the F2 family identified as QTL heterozygous; consequently this QTL was confirmed within the present study. The QTL for **PC** found in the JOINT-design on this

chromosome was not confirmed within the present study. The F4 family, however, showed a significant effect in all three designs ($P < 0.01$ in every design, calculated at the QTL position estimated in the JOINT-design).

In contrast to the estimated position for the **MY** QTL, the **PC** QTL was located closer to the start of the chromosome (marker interval *BM103 CSSM34*), indicating that different QTL are segregating for **MY** and **PC** on chromosome 5, which is also supported by visual comparison of the plots of the corresponding test statistics (not shown) and by the fact that different families showed an effect for these QTL.

The weak QTL for **FC** in the JOINT-design on this chromosome was previously reported near the same marker (*BM315*) [11].

Chromosome 14: Based on literature reports [11,33], the QTL for milk production traits on chromosome 14 was expected. Looft *et al.* [17] mapped the expressed sequence tag *KIEL_E8* to the proximal region of this chromosome and found a linkage disequilibrium to the milk production QTL. The estimated effects were in the opposite direction for fat yield and protein yield [17]. Identical results were obtained in the present study. *KIEL_E8* and *ILSTS039* was only genotyped in the ADR-design (Tab. II), probably generating small differences in the estimated positions of the QTL between the three designs. Recently Grisart *et al.* [10] and Winter *et al.* [32] reported the identification of a positional candidate gene (*DGAT1*) with a mutation most likely causing the chromosome 14 QTL effect.

Chromosome 18: The QTL on chromosome 18 for **SCS** found in the ADR-design has been reported in the literature. Ashwell *et al.* [1] found a significant effect on SCS in the neighbourhood of the *BM2078* marker. In our study this marker was located near the estimated SCS QTL position at the end of the chromosome (Tab. II). Schrooten *et al.* [22] mapped a QTL for SCS in the middle part of the chromosome (marker interval *BM7109 ILSTS002*).

Chromosome 19: The genomewise significant QTL for **FC** in the JOINT-design was confirmed in the present study (chromosomewise significant in both the ADR- and Inra-Design). The plots of the test statistics (Fig. 1) show that the shapes for the ADR- and the Inra-design were very similar. As expected, the test statistic of the JOINT-design showed the highest value. To our knowledge no **FC** QTL on chromosome 19 has been reported by any QTL study. The genomewise significant QTL for **FY** found in the Inra- and JOINT-design, could not be confirmed in the study. Probably the identified QTL for **FC** and **FY** represents a single one, because the plots of the test statistics had a similar shape (not shown).

Additionally we found a QTL for **SCS** on this chromosome in the Inra- and JOINT-design, which was not confirmed by the ADR-design. To our knowledge, this QTL has not been reported previously.

Chromosome 23: The QTL for **PC** was only significant in the JOINT-design, hence it was not confirmed within this study. Two families (F4 and F5), however, showed significant haplotype contrast in all three designs at the estimated QTL position in the JOINT-design. No families were heterozygous in the Inra-design for the QTL for **MY** and **PY** found in the ADR-design, hence it could not be confirmed in this study.

Chromosome 26: The results on this chromosome were very contradictory. While the QTL for **MY**, **FY** and **PY** was highly significant in the Inra- and in the JOINT-design, it was not found in the ADR-design, and hence, could not be confirmed within the study. The plots of the test statistics for **MY** (Fig. 1) show the substantial difference between the shapes of the Inra- and JOINT-design on the one hand and the ADR-design on the other hand. It was not possible to uncover the reason for these discrepancies from our data. The differences between the plots of the Inra- and the JOINT-design were attributed to the fact that the Inra-design was genotyped only for two markers in the distal chromosomal region (Inra*081* and *IDVGA59*), but the ADR-design was genotyped for three additional markers located between the two markers (see Tab. II) with the consequence that the average information content of the JOINT-design in this marker bracket was increased.

The test statistics for **FY**, **PY** and **FC** followed the test statistics presented in Figure 1 for all three designs very well (not shown). Therefore, it may be assumed that at this position there could be a single pleiotropic QTL which affects several milk production traits rather than different closely linked QTL, although we do not have definitive proof for this.

Chromosomes 6 and 20: Many studies report QTL for various milk production traits on chromosome 20 [7,33], and, so even more, on chromosome 6 [5,15,23,33]. Ron *et al.* [21] verified the hypothesis that two QTL are segregating on chromosome 6. With respect to these studies, it was surprising that no significant effect for any trait in any design could be detected on chromosomes 6 and 20, probably due to QTL homozygosity of the families included in the present study.

4.3. Discrepancies between the ADR- and Inra-design

The comparison of QTL results for the ADR- and Inra-design revealed three substantial discrepancies: For the SCS QTL on chromosomes 18 and 19 and for the milk production QTL on chromosome 26. Based on our data, the reasons for this lack of congruency could not be elucidated. An explanation might be that the QTL allele frequencies differ between the populations and the QTL shows a dominance effect. A further explanation might be that genotype by environment interaction plays an important role. This seems to be more likely for the SCS trait rather than for the milk production traits.

According to Interbull evaluations, the correlation between breeding values for this trait estimated in Germany and in France is above 0.9 [19], which can be interpreted as an argument against a genotype by environment interaction on a whole genome level. Nevertheless, a genotype by environment interaction might exist for a single QTL, especially when the interaction causes an allele substitution effect to show a different size but equal sign of effect in two environments. Additionally, a QTL by polygenic genotype interaction might exist, indicating that there might be different background genes for the traits with the consequence that some QTL appear only as significant in one of the two designs. Note that these hypotheses are very speculative.

4.4. Experimental strategy

This study demonstrated the advantages of a joint analysis of families that are included in different granddaughter designs. Because of the optimal allocation of limited resources it was not possible to include all families of both initial granddaughter designs, therefore a subset of families was chosen that were included in both experiments. However, Walling *et al.* [28] showed that there is even a benefit in a combined analysis of data from different designs when families are nested within experiments. Due to the coding of genotypes, phenotypes and animals, the information was only exchanged on a family level, and therefore, individual anonymity was ensured. The large family size of the JOINT-design resulted in a high experimental power. As a consequence, a number of QTL could be mapped, which met the stringent significance level of genomewise significance. Three QTL showed an at least chromosomewise significant effect in both the ADR- and in the Inra-design, and therefore, were confirmed in this study. However, it was not possible to draw more precise conclusions about the position of the QTL, because the confidence intervals were in general the largest for the JOINT-design. This was surprising, because it is known from simulation studies that bootstrap confidence intervals for the QTL positions become smaller with increasing population size [2,27]. An explanation is that the JOINT-design was genotyped heterogeneously, *i.e.* individuals from one family were genotyped for a different set of markers (see Tab. II). Following this, an increased number of markers genotyped in all individuals of the ADR- and the Inra-design would be necessary to reduce the interval width to a desired size of 10 to 20 cM. Hence, the low number of common markers is the weak point of this study, a higher number would have probably resulted in more precise estimates of the QTL position. However, for a derivation of haplotypes of the grandsires, the number of common markers was large enough due to the additional typings of sires from the ADR-design with markers that had been genotyped only for the sires of the Inra-design so far. The comparison of the haplotypes of the sires derived once from the Inra-design data and once from the ADR-design data did not reveal any conflicts.

The noninclusion of the estimated QTL position in the confidence intervals in some cases was attributed to the fact that these positions were also marker locations and, therefore, might be biased by the markers [29]. Permutation bootstrap produces confidence intervals that are corrected for the marker impact [2].

5. CONCLUSION

The potential benefit of a combined analysis of data from different granddaughter designs for QTL analysis was demonstrated. It was possible to detect and confirm a number of QTL simultaneously. Important prerequisites for this type of study are comparable phenotypes across studies and a sufficient number of common markers genotyped in all members of a family. This number was limited in the present study with the consequence, that it was not possible to estimate QTL positions more precisely. Since many Holstein families are bred worldwide and most QTL experiments used a granddaughter design, similar studies for traits that are comparable across experiments should be performed.

ACKNOWLEDGEMENTS

This research was supported by the German Cattle Breeders Federation (ADR) and the German Ministry of Education, Science, Research and Technology (BMBF). It has benefited from the helpful comments of two anonymous reviewers.

REFERENCES

[1] Ashwell M.S., Rexroad C.E., Miller R.H., Vanraden P.M., Da Y., Detection of loci affecting milk production and health traits in an elite US Holstein population using microsatellite markers, Anim. Genet. 28 (1997) 216–222.

[2] Bennewitz J., Reinsch N., Kalm E., Improved confindence intervals in Quantitative Trait Loci mapping by permutation bootstrapping, Genetics 160 (2002) 1673–1686.

[3] Boichard D., Grohs C., Bourgeois F., Cerqueira F., Faugeras R., Neau A., Rupp R., Amigues Y., Boscher M.Y., Levéziel H., Detection of genes influencing economic traits in three French dairy cattle breeds, Genet. Sel. Evol. 35 (2003) 77–101.

[4] Churchill G.A., Doerge R.W., Empirical threshold values for quantitative trait mapping, Genetics 138 (1994) 963–971.

[5] Coppieters W., Kvasz A., Arranz J.-J., Grisart B., Riquet J., Farnir F., Georges M., The great-grand-daughter design: a simple strategy to increase the power of a grand-daughter design for QTL mapping, Genet. Res. 74 (1999) 189–199.

[6] Elsen J.M., Mangin B., Goffinet B., Boichard D., Le Roy P., Alternative models for QTL detection in livestock, I. General information, Genet. Sel. Evol. 31 (1999) 213–224.

[7] Georges M., Nielsen D., Mackinnon M.J., 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) 907–920.

[8] Göring H.H.H., Terwilliger J.D., Blangero J., Large upward bias in estimation of locus-specific effects from genomwide scans, Amer. J. Hum. Genet. 69 (2001) 1357–1369.

[9] Green P., Falls K., Crooks S., Documentation of CRI-MAP, Version 2.4. Washington University School of Medicine, St. Louis, MO, USA, 1990.

[10] Grisart B., Coppieters W., Farnir F., Karim L., Ford C., Berzi P., Cambisano N., Mni M., Reid S., Simon P., Spelman R., Georges M., Snell R., Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition, Genome Res. 12 (2002) 222–231.

[11] Heyen D.W., Weller J.I., Ron M., Band M., Beever J.E., Feldmesser E., Da Y., Wiggans G.R., Vanraden P.M., Lewin H.A., A genome scan for QTL influencing milk production and health traits in dairy cattle, Physiol. Genomics 1 (1999) 165–175.

[12] Kappes S.M., Keele J.W., Stone R.T., Sonstegard T.S., Smith T.P.L., Mcgraw R.A., Lopezcorrales N.L., Beattie C.W., A second-generation linkage map of the bovine genome, Genome Res. 7 (1997) 235–249.

[13] Kashi Y., Hallerman E., Soller M., Marker assisted selection of candidate bulls for progeny testing programmes, Anim. Prod. 51 (1990) 63–74.

[14] Knott S.A., Elsen J.M., Haley C.S., Methods for multiple-marker mapping of quantitative trait loci in half-sib populations, Theor. Appl. Genet. 93 (1996) 71–80.

[15] 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, Anim. Genet. 30 (1999) 333–340.

[16] Lander E., Kruglyak L., Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results, Nat. Genet. 11 (1995) 241–247.

[17] Looft C., Reinsch R., Karall-Albrecht C., Paul S., Brink M., Thomsen H., Brockmann G., Kühn C., Schwerin M., Kalm E., A mammary gland EST showing linkage disequilibrium to a milk production QTL on bovine chromosome 14, Mamm. Genome 12 (2001) 646–650.

[18] Mackinnon M.J., Georges G., Marker-assisted preselection of young dairy bulls prior to progeny testing, Livest. Prod. Sci. 54 (1998) 229–250.

[19] Mark T., Fikse F., Banos G., Emanuelson U., Philipsson J., Summary of Mace pilot-runs for somatic cell count and clinical mastitis, Interbull Bull. 26 (2001) 43–52.

[20] Reinsch N., A multiple-species, multiple-project database for genotypes at codominant loci, J. Anim. Breed. Genet. 116 (1999) 425–435.

[21] Ron M., Kliger 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.

[22] Schrooten C., Bovenhuis H., Coppieters W., van Arendonk J.A.M., Whole genome scan to detect quantitative trait loci for conformation and functional traits in dairy cattle, J. Dairy Sci. 83 (2000) 795–806.

[23] Spelman R.J., Coppieters W., Karim L., van Arendonk J.A.M., Bovenhuis H., Quantitative trait locus analysis for five milk production traits on chromosome six in the Dutch Holstein Friesian population, Genetics 144 (1996) 1799–1808.

[24] Spelman R.J., Bovenhuis H., Moving from QTL experimental results to the utilisation of QTL in breeding programmes, Anim. Genet. 29 (1998) 77–84.

[25] Thomsen H., Reinsch N., Xu N., Looft C., Grupe S., Kühn C., Brockmann G.A., Schwerin M., Leyhe-Horn B., Hiendleder S., Erhardt G., Medjugorac I., Russ I., Förster M., Brenig B., Reinhardt F., Reents R., Blümel J., Averdunk G., Kalm E., A male bovine linkage map for the ADR granddaughter design, J. Anim. Breed. Genet. 117 (2000) 289–306.

[26] Thomsen H., Reinsch N., Xu N., Looft C., Grupe S., Kühn C., Brockmann G.A., Schwerin M., Leyhe-Horn B., Hiendleder S., Erhardt G., Medjugorac I., Russ I., Förster M., Brenig B., Reinhardt F., Reents R., Blümel J., Averdunk G., Kalm E., Comparison of estimated breeding values, daughter yield deviations and de-regressed proofs within a whole genome scan for QTL, J. Anim. Breed. Genet. 118 (2001) 357–370.

[27] Visscher P.M., Thompson R., Haley C.S., Confidence intervals in QTL mapping by bootstrapping, Genetics 143 (1996) 1013–1020.

[28] Walling G.A., Visscher P.M., Andersson L., Rothschild M.F., Wang L., Moser G., Groenen M.A.M., Bidanel J.-P., Cepica S., Archibald A.L., Geldermann H., de Koning D.J., Milan D., Haley C.S., Combined analyses of data from quantitative trait loci mapping studies: chromosome 4 effects on porcine growth and fatness, Genetics 155 (2000) 1369–1378.

[29] Walling G.A., Haley C.S., Perez-Enciso M., Thompson R., Visscher P., On the mapping of quantitative trait loci at marker and non-marker locations, Genet. Res. 79 (2002) 97–106.

[30] Weller J.I., Kashi Y., Soller M., Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle, J. Dairy Sci. 73 (1990) 2525–2537.

[31] Weller J.I., Song J.Z., Heyen D.W., Lewin H.A., Ron M., A new approach to the problem of multiple comparisons in the genetic dissection of complex traits, Genetics 150 (1998) 1699–1706.

[32] Winter A., Kramer W., Werner F.A., Kollers S., Kata S., Durstewitz G., Buitkamp J., Womack J.E., Thaller G., Fries R., Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content, Proc. Natl. Acad. Sci. 99 (2002) 9300–9305.

[33] Zhang Q., Boichard D., Hoeschele I., Ernst C., Eggen A., Murkve B., Pfister-Genskow M., Witte L.A., Grignola F.E., Uimari P., Thaller G., Bishop M.D., Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree, Genetics 149 (1998) 1959–1973.