Haplotype analysis and linkage disequilibrium for *DGAT1*

EVA M. STRUCKEN[1,2], SIHAM RAHMATALLA[1], DIRK-JAN DE KONING[2] and GUDRUN A. BROCKMANN[1]

[1]Breeding Biology and Molecular Genetics, Department for Crop and Animal Sciences, Faculty of Agriculture and Horticulture, Humboldt-Universität zu Berlin, Berlin, Germany, [2]Genetics and Genomics, Roslin Institute and R(D)SVS, University of Edinburgh, Roslin, United Kingdom

Abstract

This study focused on haplotype effects and linkage disequilibrium (LD) for the K232A locus and the promoter VNTR in the *DGAT1* gene. Analyses were carried out in three German Holstein Frisian populations (including 492, 305, and 518 animals) for milk yield, milk fat and protein yield, and milk fat and protein content. We found that effects of the promoter VNTR were not significant and explain only a small amount of the variation of the QTL on BTA14. Haplotype effects were less significant than the K232A locus by itself, but the haplotype containing the A allele of the K232A locus and allele 3 with five repeats of the promoter VNTR showed negative effects on protein content when paternally inherited, whereas the haplotype with the A allele and VNTR allele 2 (with six repeats) increased the protein content. Significant differences between these two haplotypes occurred for protein yield as well, pointing to a linked effect that is picked up by the haplotypes rather than a direct effect of the VNTR. The linkage disequilibrium, estimated by *D',* showed values between 0.29 and 0.59 which is unexpectedly low for a distance of ~10 kb. Only a very low correlation between the two loci was observed due to the almost similar frequencies of haplotypes containing the A or K allele of the K232A locus.

Keywords: German Holstein Frisian, *DGAT1*, haplotype, linkage disequilibrium

Zusammenfassung

Haplotypenanalyse und Kopplungsungleichgewicht für *DGAT1* in Holstein Populationen

Die vorliegende Studie untersucht Haplotypeneffekte und das Kopplungsungleich-gewicht (LD) für den K232A Polymorphismus und den Promoter-VNTR im *DGAT1*-Gen. Die Analysen von in drei Deutschen Holstein Frisian Populationen (mit 492, 305 und 518 Milchkühen) für Milchleistung, Milchfett- und Milchproteinmenge sowie Milchfett- und Milchproteingehalt durchgeführt. Wir konnten zeigen, dass die Effekte in den einzelnen Herden signifikant für den K232A Polymorphismus jedoch nicht signifikant für den Promoter VNTR sind, welcher somit nur einen kleinen Anteil an der Variation des QTLs auf BTA14 darstellt. Haplotypeneffekte sind weniger signifikant als der K232A Locus für sich genommen. Der Haplotyp, der das A-Allel des K232A Locus sowie Allel 3 des Promoter-VNTRs mit fünf Wiederholungen enthält, hat negative Effekte auf den Proteingehalt, wenn er paternal vererbt wurde, wohingegen der Haplotyp mit Allel A und dem VNTR-

Allel 2 mit sechs Wiederholungen den Proteingehalt steigert. Signifikante Unterschiede zwischen diesen Haplotypen treten ebenfalls für Proteinmenge auf. Diese Haplotypen-effekte deuten darauf hin, dass der VNTR nicht allein den Effekt erklären kann. Das Kopplungsungleichgewicht wurde durch *D'* bestimmt und zeigt Werte zwischen 0.29 und 0.59, was für eine Distanz von ~10 kb unerwartet niedrig ist. Die geringe Kopplung kann durch die beinahe gleichen Haplotypenfrequenzen, die entweder das A oder das K Allel des K232A Locus enthalten, erklärt werden.

Schlüsselwörter: Deutsche Holstein Frisian, *DGAT1*, Haplotyp, Kopplungsungleichgewicht

Introduction

Several studies showed an association between gene variants of the acylCoA:diacylglycerol acyltransferase 1 gene (*DGAT1*) on BTA14 and milk composition and quality. A non-synonymous mutation (K232A) in exon 8 as well as a variable number of tandem repeats (VNTR) in the promoter region of *DGAT1* showed impacts on fat and protein content as well as on fatty acid composition (WINTER *et al.* 2002, GRISART *et al.* 2002, THALLER *et al.* 2003, KÜHN *et al.* 2004, CITEK *et al.* 2007, RAHMATALLA *et al.* 2008). Significant dominance effects for the K allele as well as for allele 1 (seven repeats) and 5 (three repeats) of the VNTR on milk fat content have been reported only by KÜHN *et al.* (2007). Additional association studies investigated the effects of haplotypes considering the two loci and showed different effects on the traits depending on which VNTR allele was present (SANDERS *et al.* 2006, KÜHN *et al.* 2007). Furthermore, a high *D'* value of 0.98 between the two loci was presented for Holstein breeds (GAUTIER *et al.* 2007) which is expected for a distance of about 10 kb. In the present study, we analyzed the two loci of the *DGAT1* gene in three German Holstein Frisian herds with respect to dominance effects and the association of maternal and paternal inherited haplotypes with milk, fat and protein yield, and fat and protein content. Furthermore, linkage disequilibrium between the two loci was calculated.

Material and methods

Genotype and haplotype data for the *DGAT1* K232A locus, the *DGAT1* promoter VNTR, as well as haplotype information were taken from RAHMATALLA *et al.* (2008) where all three herds were analysed as one. The three herds were formerly selected independently. Herd 2 was selected for high protein yield. Even if no allele differences are seen on the analyzed *DGAT1* polymorphisms, genome-wide differences are very likely, and therefore, this study treated the three herds separately. The alleles of the K232A mutation were denoted as K (Lysine) and A (Allanin). Alleles of the *DGAT1* promoter VNTR were denoted as 1 to 5, where allele 1 has seven repeats, allele 2 six, allele 3 five, allele 4 four, and allele 5 three repeats, respectively. Haplotypes were denoted as amino acid variant A or K and the number of the VNTR allele, e.g. A2. For the estimation of genotype and haplotype effects, only genotypes and haplotypes represented by at least 10 % of the animals were considered. Only complete records of milk production performance from the first to third lactation at the 305th day were considered. Thus, for the final analysis, herd 1 consisted of

492 animals descending from 106 sires, herd 2 of 305 animals from 79 sires, and herd 3 of 518 animals from 145 sires. Animal-specific performance data were provided by VIT (Vereinigte Informationssysteme Tierhaltung w.V., Verden, Germany). Because we did not want to include corrections outside the herds, data analysis was based on herd specific yield deviation (YD) for every lactation corrected for calving season based on the VIT data. Other factors had no significant influence. The mean YD over the first three lactations was calculated as the measure for all three lactations.

Genotype and allele substitution effects as well as haplotype effects for both loci were modelled according to RAHMATALLA *et al.* (2008; model 2, 3, 4, and 5). The herd effect was tested including the herd as a fixed effect in model 2 of RAHMATALLA *et al.* (2008) and applied to a data set including all herds.

To show whether the haplotype as a combination of the two loci explains more variation than the single K232A locus following model was used:

$$Y_{ijkl} = \mu + s_i + \text{K232A}_k + \text{HT}_l + e_{ijkl} \tag{1}$$

where Y_{ijkl} is the YD of daughter j within sire i, μ is the overall mean of the herd, s_i is the random effect of sire i, K232A_k is the fixed effect of the *DGAT1* K232A genotype k (AA, KA, KK), HT_l is the fixed effect of either the paternally or maternally inherited haplotype l and e_{ijkl} is the random residual effect.

All analyses were performed with the MIXED procedure of the SAS program version 9.2 (SAS 2008). Dominance values were estimated as the difference between the heterozygous and the mean of the homozygous genotypes for the *DGAT1* K232A locus only. For estimating linkage disequilibrium (LD), D' (normalized LD between the two loci weighted by the product of allele frequencies), r^2 (pooled correlation between the two loci based on allele frequencies), and χ^2 (chi-square statistic to test for independence between alleles at the two loci) were calculated for multiallelic loci according to ZHAO *et al.* (2007).

Results

The *DGAT1* promoter VNTR was not in HWE, except for herd 2. The deviation from HWE for the VNTR locus might be due to a lower animal number or the presence of family relationships. Genotype and allele frequencies over 10 % are shown in Table A1. Herd as a fixed effect in an analysis over all three herds did not show any significant influence on the *DGAT1* loci.

However, slightly different effects were found between the herds if they were analysed separately. Paternal as well as maternal haplotype K3 showed a positive effect on the content traits in all three herds. The effects of the haplotype K3 showed deviations from additive effects in such way that the sum of the effect of the K allele and the VNTR allele 3 on fat content was higher than the effect of haplotype K3. Similar deviations were observed for other traits.

Model 1 was used to show whether the single K232A locus or the haplotype as a combination of the K232A and VNTR polymorphism has a greater effect. Including the K232A locus revealed that the haplotypes were almost always insignificant and explained

less of the variance then the K232A locus on its own (Table 1). This shows that the influence of the VNTR can be assumed as very small.

Table 1
P-values for effects of the K232A locus and haplotypes (HT)
P-Werte für Effekte des K232A Locus und Haplotypen (HT)

Trait	Herd	K232A[1]	Maternal HT	K232A[2]	Paternal HT
YD_{mkg}	1	<0.0001	0.09	0.01	0.27
	2	0.07	0.37	0.0004	0.20
	3	0.05	0.94	0.01	0.08
YD_{fkg}	1	0.003	0.34	0.35	0.29
	2	0.24	0.28	0.65	0.14
	3	0.002	0.50	0.003	0.08
$YD_{f\%}$	1	<0.0001	0.08	<0.0001	0.26
	2	<0.0001	0.76	<0.0001	0.46
	3	<0.0001	0.59	<0.0001	0.99
YD_{pkg}	1	0.06	0.07	0.30	0.71
	2	0.83	0.42	0.03	0.29
	3	0.13	0.93	0.04	0.04
$YD_{p\%}$	1	<0.0001	0.59	<0.0001	0.54
	2	<0.0001	0.77	<0.0001	0.06
	3	0.01	0.70	0.0001	0.45

[1]*P*-value for the K232A locus in model 1, when the maternal HT was used, [2]*P*-value for the K232A locus in model 1, when the paternal HT was used

Table 2
Linkage Disequilibrium between the two multiallelic loci in *DGAT1* (K232A and VNTR) for three Holstein populations
Kopplungsungleichgewicht zwischen den beiden multiallelen Loci im DGAT1-Gen (K232A und VNTR) für drei Holstein Populationen

	Herd					
	Maternal HT			Paternal HT		
	1	2	3	1	2	3
n	507	327	546	507	327	546
D'	0.34	0.36	0.51	0.29	0.59	0.46
r^2	0.05	0.05	0.10	0.06	0.15	0.08
χ^2	116.75	87.03	170.30	88.50	180.81	178.50
P-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

The K allele showed significant dominance in all herds on all traits except for fat yield. Only herd 3 showed a significant dominance effect for the K allele on fat yield. Furthermore, we report a dominance effect of the A allele on protein yield for herd 3 (*P*=0.0004), which has to be proven by further analysis.

Relatively low *D'* values between 0.29 and 0.59 were estimated as a measurement of LD, and in addition very low r^2 values were observed. All estimates were highly significant with 4 degrees of freedom (Table 2).

Discussion

Frequencies of both *DGAT1* polymorphisms fulfilled the expectations for Holstein Frisians (SANDERS *et al.* 2006, GAUTIER *et al.* 2007, KÜHN *et al.* 2007). The effects of the K232A locus as reported for Holstein Frisian breeds could be confirmed for all three herds (KÜHN *et al.* 2007, RAHMATALLA *et al.* 2008). Effects reported for the VNTR locus on the content traits could not be confirmed. A major effect for allele 1 (with seven repeats) was reported by KÜHN *et al.* (2007). We did not find allele 1 with a sufficiently high frequency in our population. The missing VNTR allele 1, and thus, its absent allele effect might be restricted to this data set.

Taking the K232A locus into account, the VNTR locus alone as well as the haplotypes showed no significant effects, suggesting only a very small or no contribution of the promoter VNTR to the variation explained by the QTL on BTA14, which does not reach significance due to the smaller animal numbers in the separate herds. Nevertheless, excluding the K232A locus (model 3 by RAHMATALLA *et al.* [2008]), significant differences between the paternal haplotypes A3 and A2 were found in two herds. Effects of the haplotypes followed in most cases the reported effects for the K232A alleles (Figure A1, Table A2).

Significant positive dominance effects of the Lysine *DGAT1* variant (K) on fat content can be confirmed (KÜHN *et al.* 2007). A striking point is the significant dominance effect of the Alanin variant (A) for protein yield in herd 3 while for herd 1 and 2 dominance of the K allele is indicated (Figure A1, Table A3).

In our populations, the K232A alleles occurred with frequencies up to 18% with other VNTR alleles than allele 3 and did not show one exclusive haplotype, resulting in relatively low *D'* values compared to GAUTIER *et al.* (2007). Nevertheless, we showed that the K232A alleles were associated with allele 3 of the VNTR more frequently than with other VNTR alleles. The low r^2 values result from the almost equivalent frequency of A3 and K3 haplotypes. The genetic background of herd 2, which had been selected for high protein yield, seems to have no influence on the *DGAT1* effects. However, herd 3 shows deviation from the other two herds regarding dominance effects which can not be explained without more information about the population. This might reflect interaction of the *DGAT1* locus with other regions in the genome. Unidentified population substructures as a result of the population history including gene flow and founder effects could cause the deviation in haplotype frequency. The determination of the LD between the biallelic K232A polymorphism and the multiallelic VNTR depends on the representative sample of animals of the population. The three herds in the present examination represent an unbiased sample with *D'* estimates between 0.29 and 0.59 for the LD.

Acknowledgements

References

Citek J, Rehout V, Hradecka E, Vecerek L, Panicke L (2007) The breeding values of German Holstein sires and the *DGAT1* polymorphism. Arch Tierz 50, 136-46

Gautier M, Capitan A, Fritz S, Eggen A, Boichard D, Druet T (2007) Characterization of the *DGAT1* K232A and variable number of tandem repeat polymorphisms in French dairy cattle. J Dairy Sci 90, 2980-8

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 (2002) 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, 222-31

Kühn C, Thaller G, Winter A, Bininda-Emounds OR, Kaupe B, Erhardt G, Bennewitz J, Schwerin M, Fries R (2004) Evidence for multiple alleles at the *DGAT1* locus better explains a quantitative trait locus with major effects on milk fat content in cattle. Genetics 167,1873-81

Kuehn C, Edel C, Weikard R, Thaller G (2007) Dominance and parent-of-origin effects of coding and non-coding alleles at the acylCoA-diacylglycerol-acyltransferase (*DGAT1*) gene on milk production traits in German Holstein cows. BMC Genet 8, 62

Rahmatalla S, Mueller U, Strucken EM, Brockmann G (2008) Characterization of *DGAT1*-variants in German Holstein cattle under production conditions. Züchtungsk 80, 473-84 [in German]

Sanders K, Bennewitz J, Reinsch N, Thaller G, Prinzenberg EM, Kuhn C, Kalm E (2006) Characterization of the *DGAT1* mutations and the CSN1S1 promoter in the German Angeln dairy cattle population. J Dairy Sci 89, 3164-74

Thaller G, Kramer W, Winter A, Kaupe B, Erhardt G, Fries R (2003) Effects of *DGAT1* variants on milk production traits in German cattle breeds. J Anim Sci 81, 1911-8

Winter A, Kramer W, Werner FA, Kollers S, Kata S, Durstewitz G, Buitkamp J, Womack JE, Thaller G, Fries R (2002) 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 USA 99, 9300-5

Zhao H, Nettleton D, Dekkers JC (2007) Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between single nucleotide polymorphisms. Genet Res 89, 1-6

Appendix

Table A1
Genotype (GT), allele and haplotype (HT) frequencies with more than 10% for the two loci in *DGAT1* (K232A and VNTR) in three Holstein populations

Genotypen (GT), Allele- und Haplotypenfrequenzen (HT) der beiden Loci im DGAT1-Gen (K232A und VNTR) mit mehr als 10% in drei Holstein Frisian Populationen

Locus	GT	Herd 1	2	3	Allele	1	2	3
		Genotype frequencies				Allele frequencies		
K232A	AA	0.29	0.30	0.32	A	0.54	0.56	0.57
	KA	0.50	0.52	0.49				
	KK	0.21	0.18	0.18	K	0.46	0.44	0.43
VNTR					5	0.02	0.01	0.02
	34	0.28	0.23	0.19	4	0.22	0.17	0.15
	33	0.39	0.33	0.37	3	0.64	0.60	0.62
	23	0.19	0.26	0.24	2	0.11	0.18	0.17
					1	0.002	0.03	0.05
	HT*	Maternal HT frequency			HT*	Paternal HT frequency		
HT	A4	0.15	0.13	0.13	A4	0.13	0.09	0.07
	A3	0.27	0.25	0.25	A3	0.34	0.28	0.32
	A2	0.11	0.12	0.15	A2	0.05	0.18	0.12
	K4	0.11	0.10	0.06	K4	0.06	0.03	0.03
	K3	0.27	0.28	0.30	K3	0.39	0.41	0.39
	K2	0.06	0.06	0.04	K2	0.01	0.009	0.02

*The symbol of the haplotype consists of the allele at the K232A locus followed by the allele number of the VNTR allele.

Table A2
Significant trait differences between haplotype groups (HT) for maternal and paternal haplotypes for all herds (LSM ±SE, (P-value))

Signifikante Unterschiede zwischen Haplotypengruppen (HT) für maternale und paternal Haplotypen für alle drei Herden (LSM ±SE, (P-Wert))

Trait	HT	Herd 1	Herd 2	Herd 3	Herd 1	Herd 2	Herd 3
		Maternal HT differences			Paternal HT differences		
YD_{mkg}	A4-K3	–	694.40 ± 21.76*	454.94 ± 173.27*	365.27 ± 153.18*	–	–
	A3-K3	389.40 ± 126.27*	397.33 ± 80.84	350.30 ± 142.22	597.37 ± 113.73***	–	401.36 ± 128.52**
	A2-K3	692.86 ± 164.69***	738.90 ± 28.30**	390.54 ± 166.42	–	–	–
YD_{fkg}	A2-K3	–	–	−16.65 ± 5.64*	−16.49 ± 6.13*	–	−24.74 ± 6.24***
$YD_{f\%}$	A4-K3	−0.27 ± 0.06***	−0.28 ± 0.07**	−0.30 ± 0.06***	−0.41 ± 0.06***	–	–
	A3-K4	–	−0.30 ± 0.08**	–	–	–	–
	A3-K3	−0.24 ± 0.05***	−0.25 ± 0.06***	−0.30 ± 0.05***	−0.40 ± 0.05***	−0.22 ± 0.06***	−0.31 ± 0.04***
	A2-K4	−0.23 ± 0.08*	−0.34 ± 0.10**	–	–	–	–
	A2-K3	−0.33 ± 0.07***	−0.28 ± 0.08**	−0.40 ± 0.05***	–	−0.23 ± 0.08**	−0.30 ± 0.06***
YD_{pkg}	A3-K3	–	–	–	9.37 ± 3.70*	–	10.30 ± 3.92*
	A2-K3	17.34 ± 5.08**	–	–	–	–	–
	A3-A2	–	–	–	–	–	13.98 ± 5.53*
$YD_{p\%}$	A4-K3	−0.09 ± 0.02***	-0.09 ± 0.03*	−0.07 ± 0.02**	−0.10 ± 0.03**	–	–
	A3-K3	−0.07 ± 0.02**	-0.08 ± 0.03*	−0.07 ± 0.02**	−0.12 ± 0.02***	−0.11 ± 0.03***	–
	A2-K3	−0.08 ± 0.03*	-0.14 ± 0.03***	−0.08 ± 0.02***	–	–	–
	A3-A2	–	–	–	–	−0.09 ± 0.03**	–

*differences statistically significant at $P \leq 0.05$,   **differences statistically significant at $P \leq 0.01$,   ***differences statistically significant at $P \leq 0.001$

Table A3
Dominance values (±SE, (P-value)) for the *DGAT1* K232A locus in three Holstein Frisian herds

Dominanzwerte (±SE, (P-Wert)) für den DGAT1 K232A Locus in drei Holstein Frisian Herden

Trait	Herd 1	Herd 2	Herd 3
YD_{mkg}	−80.32 ± 64.05***	−187.60 ± 95.34***	−22.13 ± 69.43***
YD_{fkg}	3.35 ± 2.49	0.57 ± 3.53	4.48 ± 2.33*
$YD_{f\%}$	0.07 ± 0.02***	0.10 ± 0.03***	0.06 ± 0.02***
YD_{pkg}	−0.50 ± 2.00**	−3.03 ± 2.99**	**0.92 ± 2.15***
$YD_{p\%}$	0.02 ± 0.01***	0.04 ± 0.01**	0.02 ± 0.01***

*differences statistically significant at $P \leq 0.05$,   **differences statistically significant at $P \leq 0.01$,   ***differences statistically significant at $P \leq 0.001$
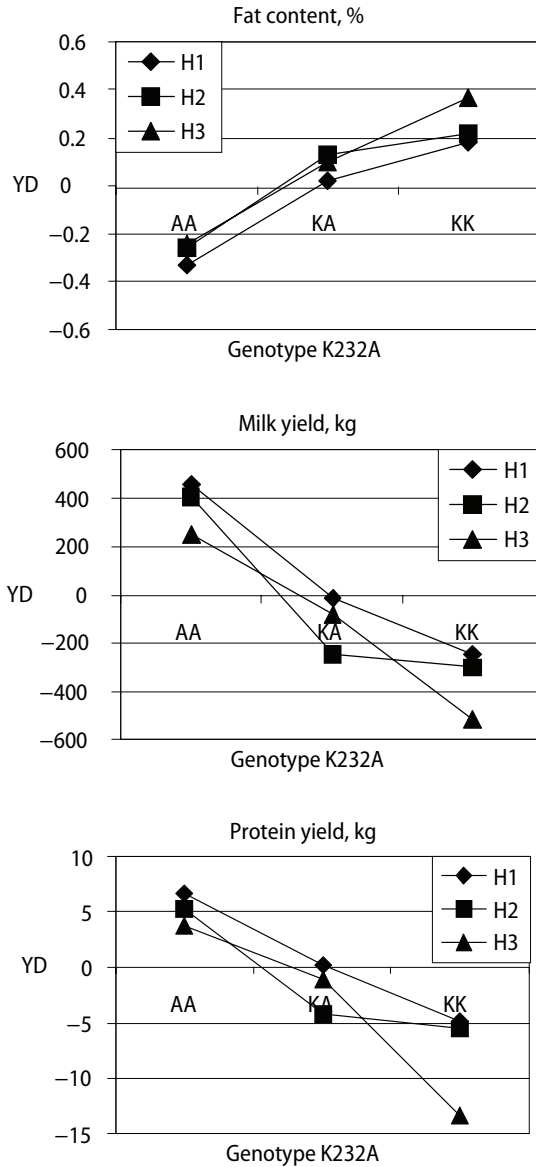
Figure A1
Least square means (LSM) of the yield deviation (YD) for the K232A genotypes in the three herds (H1=herd 1; H2= herd 2; H3=herd 3). Significant dominance effects were observed for the K allele except on fat yield. Herd 3 showed dominance effect on protein yield for the A allele.

Least square means (LSM) für die Mengenabweichung des K232A Genotyps in drei Herden (H1=Herde 1; H2=Herde 2; H3=Herde 3). Signifikante Dominanzeffekte konnten für das K-Allel beobachtet werden außer für Fettmenge. In Herde 3 konnte außerdem ein Dominanzeffekt auf Proteinmenge für das A-Allel gezeigt werden.
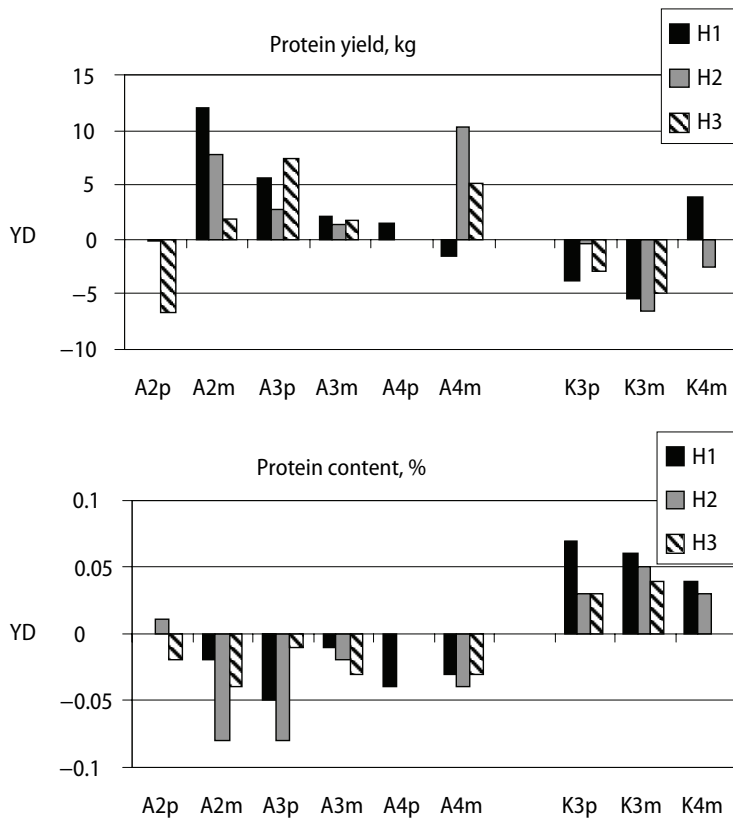
Figure A2
Differences between haplotype effects of *DGAT1* (K232A and VNTR locus) in three herds (H1=herd 1; H2= herd 2; H3=herd 3). Differences between A3 and A4 were significant for paternally (p) inherited haplotypes in herd 3 for protein yield and in herd 2 for protein content. All other significant differences occurred between haplotypes containing an A or a K allele and effects followed the observed allele effects of the K232A locus.

Unterschiede zwischen Haplotypeneffekten des DGAT1-Gens (K232A und VNTR) in drei Herden (H1=Herde 1; H2=Herde 2; H3=Herde 3). Differenzen zwischen A3 und A4 waren signifikant für den paternal (p) vererbten Haplotyp für Proteinmenge in Herde 3 und für Proteingehalt in Herde 2. Alle anderen signifikanten Unterschiede traten zwischen Haplotypen auf, die entweder ein A oder ein K Allel trugen. Die Effekte folgten den beobachteten Alleleffekten des K232A Lokus.

Corresponding author:

GUDRUN A. BROCKMANN
email: gudrun.brockmann@agrar.hu-berlin.de

Breeding Biology and Molecular Genetics, Department for Crop and Animal Sciences, Faculty of Agriculture and Horticulture, Humboldt-Universität zu Berlin, Invalidenstraße 42, 10115 Berlin, Germany