Institut für Tierhaltung und Tierzüchtung der Universität Hohenheim
Fachgebiet Genetik und Züchtung landwirtschaftlicher Nutztiere
Prof. Dr. Jörn Bennewitz

Investigations on major gene by polygene and gene by environment interaction in German Holstein dairy cattle

DISSERTATION

zur Erlangung des Doktorgrades
der Fakultät Agrarwissenschaften
der Universität Hohenheim

vorgelegt von

MELANIE ELISABETH STREIT
M. Sc. (Agr)
aus Mondorf
Saarland

Hohenheim, Juni 2013

Die Dissertation wurde mit dankenswerter Unterstützung der
Deutschen Forschungsgemeinschaft (DFG) angefertigt.

Institut für Tierhaltung und Tierzüchtung der Universität Hohenheim
Fachgebiet Genetik und Züchtung landwirtschaftlicher Nutztiere
Prof. Dr. Jörn Bennewitz

Investigations on major gene by polygene and gene by environment interaction in German Holstein dairy cattle

DISSERTATION

zur Erlangung des Doktorgrades
der Fakultät Agrarwissenschaften
der Universität Hohenheim

vorgelegt von

MELANIE ELISABETH STREIT
M. Sc. (Agr)
aus Mondorf
Saarland

Hohenheim, Juni 2013

Dekan:	Prof. Dr. Martina Brockmeier
Prüfungsvorsitzender:	Prof. Dr. Dr. h.c. Rainer Mosenthin
1. Prüfer, Berichterstatter:	Prof. Dr. Jörn Bennewitz
2. Prüfer, Mitberichterstatter:	Prof. Dr. Georg Thaller
3. Prüfer:	Prof. Dr. Hans-Peter Piepho
Tag der mündlichen Prüfung:	31. Januar 2014

Die Dissertation wurde mit dankenswerter Unterstützung der
Deutschen Forschungsgemeinschaft (DFG) angefertigt.

TABLE OF CONTENTS

GENERAL INTRODUCTION

In dairy cattle breeding, daughter records are used. Because of the widespread use of artificial insemination, the accuracy of sire breeding values is high. Often, daughters of a sire are milked in a wide range of environments, which raises the question of the importance of genotype-by-environment interaction (GxE). This means, that genotypes respond different to changes of the environment. For GxE, some studies (Kolmodin et al., 2002; König et al., 2005; Strandberg et al., 2009) already exist. To study GxE, reaction norms are used if the distribution of the environment is continuous. The intercept of a reaction norm shows the level of general production (GP) and the slope shows the environmental sensitivity (ES) (de Jong, 1995; Lynch and Walsh, 1998; James, 2009). As an environmental descriptor, mean performance of all animals has frequently been used (James, 2009). In dairy cattle, average herd production level is a common environmental descriptor (Kolmodin et al., 2002; Calus et al., 2002; Lillehammer et al., 2009).
A problem in dairy cattle is that the cows are very sensitive. Especially, when they are fed at a high level, they are very susceptible and reduce production or show diseases. The aim of this thesis was to identify SNPs which influence environmental sensitivity. Special attention was given to DGAT1 K232A, which is a major gene influencing milk yield and composition.

Putative interaction effects between DGAT1 K232A mutation and a polygenic term (other gene) were investigated in **chapter one**. This was done for milk yield, fat yield, protein yield, fat percentage and protein percentage in the German Holstein dairy cattle population. For this, we used mixed models and the test for interaction relied on the comparison of polygenic variance components depending on sire's genotype at DGAT1 K232A. Significant interaction effect were found for milk fat and protein percentage.

Reaction norm random regression sire models were used in the **second chapter** to study GxE in the German Holstein dairy cattle population. Around 2300 sires with at least 50 daughters per sire and at least seven first-lactation test day observations per daughter were analyzed. As observations, corrected test day records for milk yield, fat yield, protein yield and somatic cell score (SCS) were used. As environmental descriptors, we used herd test day solutions (htds) for milk traits, milk energy yield or SCS. Second-order orthogonal polynomial regressions were applied to the sire effects. The results show significant slope variances of reaction norms, which caused a non-constant additive genetic variance across the environmental

ranges considered and point to the presence of GxE effects. When the environment is improved, we found an increased additive genetic variance that is higher (lower) for milk traits (SCS). Because non-genetic variance increases in an improved environment, this was also influenced by pure scaling effects. The heritability was less influenced by the environment. Very little re-ranking of the sires could be found for the environmental range considered.

The aim of **chapter three** was to conduct a large scale genome-wide association analysis to identify SNPs that affect GP and ES of milk traits in the German Holstein population. Around 13 million daughter records were used in a linear reaction norm model to calculate sire estimates for GP and ES. The daughters were from 2297 sires, which were genotyped with a 54k chip. As an environmental descriptor, we used average milk energy yield performance of herds when the daughter observations were recorded. In a genome-wide association analysis, sire estimates of 1797 sires were used. With help of an independent validation set (500 sires of the same population), we confirmed significant SNPs. To separate GxE scaling and other GxE effects, a log-transformation of the observations was performed. Results from the reaction norm model showed GxE effects and some significant SNPs were validated for GP and ES. Many SNPs which affected GP also affected ES. We showed that ES of milk traits is a typical quantitative trait, which is genetically controlled by many genes with small effects and few genes with a larger effect. The log-transformation of the observations reduced the number of validated SNPs for ES, suggesting genes that not only caused scaling GxE effects.

Average herd milk production level is frequently used as an environmental descriptor, which is mainly influenced by level of feeding or feeding regime. Another important environmental factor is the level of udder health and hygiene, for which the average herd level of somatic cell count might be a descriptor. In **chapter four**, we conducted a genome-wide association analysis to identify SNPs that affect intercept and slope of milk protein yield reaction norms, when the average htds for SCS was used as environmental descriptor. Again around 12 million daughter records were used to calculate sire estimates for intercept and slope with a reaction norm model. The sires were genotyped for 54k SNPs as well. Discovery and validation of SNPs were similar as in chapter three. DGAT1 K232A is a known major gene which influences protein yield; it was included in the statistical model as a covariable. Some SNPs could be detected and validated. Most of them influence intercept and slope. In comparison to chapter three, many of the detected SNPs which affect slope were the same,

although the two environmental descriptors (average herd milk production level and SCS) had a low correlation.

Calus, M. P. L., A. F. Groen, G. de Jong. 2002. Genotype x environment interaction for protein yield in dutch dairy cattle as quantified by different models. J. Dairy Sci. 85 : 3115 – 3123.

de Jong, G. 1995. Phenotypic plasticity as a product of selection in a variable environment. Am. Nat. 145: 493 – 512.

James, J. W. 2009. Genotype by environment interaction in farm animals, pp. 151 – 167 in Adaptation and Fitness in Animal Populations. Evolutionary and Breeding Perspectives on Genetic Resource Management edited by J. van der Werf, H.-U. Graser, R. Frankham, C. Gondro, Springer Science+Business Media B. V.

König, S., G. Dietl, I. Raeder, H. H. Swalve. 2005. Genetic relationships for dairy performance between large-scale and small-scale farm conditions. J. Dairy Sci. 88: 4087 – 4096.

Kolmodin, R., E. Strandberg, P. Madsen, J. Jensen, H. Jorjani. 2002 . Genotype by environment interaction in nordic dairy cattle studied using reaction norms. Acta Agric. Scand. Sect. A Anim Sci. 52: 11 – 24.

Lillehammer, M., J. Ødegård, T. H. E. Meuwissen. 2009. Reducing the bias of estimates of genotype by environment interactions in random regression sire models. Gen. Sel. Evol. 41:30. doi:10.1186/1297-9686-41-30.

Lynch, M., B. Walsh. 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associates Inc Publishers, Sunderland, MA, USA.

Strandberg, E., S. Brotherstone, E. Wall, M. P. Coffey. 2009. Genotype by environment interaction for first-lactation female fertility traits in UK dairy cattle. J. Dairy Sci. 92: 3437 – 3446.

CHAPTER ONE

Short communication: Evidence for a major gene by polygene interaction for milk production traits in German Holstein dairy cattle

MELANIE STREIT,* NADINE NEUGEBAUER,* THEO H. E. MEUWISSEN,† and JÖRN BENNEWITZ*

*Institute of Animal Husbandry and Breeding,
University of Hohenheim, Garbenstraße 17,
70599 Stuttgart, Germany
†Department of Animal and Aquacultural Sciences,
Norwegian University of Life Sciences,
Box 1432, Ås, Norway

published in:
Journal of Dairy Science (2011) 94: 1597 – 1600
doi: 10.3168/jds.2010-3834

SHORT COMMUNICATION: Major gene by polygene interaction in dairy cattle

The present study investigated putative interaction effects between the major gene *DGAT1 K232A* and the polygenic component (i.e. all other genes except *DGAT1 K232A*) for milk production traits in the German Holstein dairy cattle population. A mixed model approach was used. Significant interaction effects were found for milk fat percentage and protein percentage.

Short communication: Evidence for a major gene by polygene interaction for milk production traits in German Holstein dairy cattle

M. Streit,* N. Neugebauer,* T. H. E. Meuwissen,† and J. Bennewitz*

*Institute of Animal Husbandry and Breeding, University of Hohenheim, Garbenstraße 17, 70599 Stuttgart, Germany

†Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Box 1432, Ås, Norway

Corresponding author:
Jörn Bennewitz
Institute of Animal Husbandry and Breeding
University of Hohenheim,
Garbenstraße 17
D-70599 Stuttgart, Germany
Phone: +49 711 459 23570
Fax: +49 711 459 23001
e-mail: j.bennewitz@uni-hohenheim.de

Abstract

The present study investigated putative interaction effects between the *DGAT1 K232A* mutation and the polygenic term (i.e. all genes except *DGAT1*) for five milk production traits in the German Holstein dairy cattle population. Mixed models were used, and the test for interaction relied on the comparison of polygenic variance components depending on the sire's genotypes at *DGAT1 K232A*. Substitution effects were highly significant for all traits. Significant interaction effects were found for milk fat and protein percentage.

Key words: *DGAT1*, epistatsis, dairy cattle

The genetic variation of quantitative traits is due to polymorphic loci with additive and non-additive genetic effects (Falconer and Mackay 1996). Non-additive genetic effects comprise usually interactions between alleles at the same loci (dominance) or at different loci (epistasis). In outbred populations, most of the genetic variance is additive, which is mainly due to the U-shaped distribution of gene frequency (Hill et al. 2008). On the gene or QTL level, epistatic effects were frequently found in experimental crosses (e.g. Carlborg and Haley 2004). They occasionally explained a substantial proportion of the genetic variance, which is also due to the intermediate gene frequencies in these crosses. In outbred livestock populations (e.g. dairy cattle), however, almost no attempts were made to map epistatic effects between QTL, which is most likely due to restricted experimental power. However, recently Hinrichs et al. (2010) reported interaction effects between a major gene (i.e. *DGAT1 K232A*) and QTL on BTA5 and BTA14 for fat and protein percentage in a dairy cattle population. The *DGAT1* gene is known to affect milk production traits in dairy cattle (Grisart et al. 2002, Winter et al. 2002). *DGAT1* encodes an enzyme that catalyzes the reaction of diacylglycerol and fatty acyl-CoA to form triglycerides. Both studies found a non-conservative substitution of lysine by alanine (*K232A*) in *DGAT1* caused by an adenine/adenine to guanine/cytosine dinucleotide substitution. *DGAT1 K232A* effects in the German Holstein dairy cattle population were estimated by Thaller et al. (2003) and Bennewitz et al. (2004). Both authors found a strong allele substitution effect for milk production traits. The lysine variant increased fat yield and percentage as well as protein percentage, the alanine variant increased milk and protein yield.

The aim of the present study was to investigate interaction effects between the *DGAT1 K232A* mutation and the polygenic component (i.e. all other genes) for milk production traits in German Holsteins.

The pedigree contained 1153 progeny-tested and *DGAT1 K232A* genotyped German Holsteins sires. Data from the first lactation for the following five traits were considered: milk yield, fat yield, protein yield, fat percentage and protein percentage. For the yield traits, daughter yield deviations (DYD) were used, which were multiplied by two. For the percentage traits no DYD were available; therefore estimated breeding values (EBV) were used. The EBV were not de-regressed, because they showed a high reliability. The DYD and EBV were taken from the routine national sire evaluation from 2009. For a summary statistic see Table 1.

Table 1. Summary statistics of the dependent variables (n = 1153).

Trait	unit	mean	sd	min	max
Fat (kg)	DYD	4.71	20.08	-56.19	79.05
Fat (%)	EBV	-0.01	0.27	-0.84	0.99
Protein (kg)	DYD	4.01	16.76	-47.77	54.68
Protein (%)	EBV	-0.01	0.11	-0.39	0.43
Milk (kg)	DYD	156.72	574.78	-1571.88	2192.88

Two statistical models were applied. The first one was

$$y = Xb + Zu + e, \quad (1)$$

with y as the vector of phenotypes, b is a 2 x 1 vector containing a fixed mean effect and a fixed regression coefficient on the number of lysine alleles (0, 1 or 2) of the animals at *DGAT1 K232A*. The regression coefficient in b represents the average allele substitution effect. X is the corresponding incidence matrix, u is the vector of polygenic effects with Z as the corresponding design matrix, and e is a vector of residuals. The expectation of y is $E(y) = Xb$ and the variance is $\text{var}(y) = ZA\sigma_u^2 Z' + I\sigma_e^2$, with A being the numerator relationship matrix among the animals, σ_u^2 is the polygenic variance, I is an identity matrix and σ_e^2 the residual variance.

In the second model, the polygenic term (i.e. the genetic term corrected for the effect of *DGAT1 K232A* mutation, u in model (1)) was split into a polygametic term of gametes associated with the lysine (u_L) and with the alanine (u_A) variant at *DGAT1 K232A*. For sires being homozygous for lysine (alanine) at *DGAT1 K232A*, the polygametic lysine (alanine) variant affected the phenotype two times, for heterozygous sires each polygametic term affected the phenotype one time. Subsequently, the model was

$$y = Xb + Z_1 u_L + Z_2 u_A + e, \quad (2)$$

with Z_1 and Z_2 are design matrices linking the phenotypes to the corresponding polygametic effects. Z_1 (Z_2) contains a 0, 1 or 2, if the sire carries 0, 1 or 2 copies of the lysine (alanine) alleles at *DGAT1 K232A*. The remaining terms are as defined above. The expectation of *y* is again $E(y) = Xb$, and the variance is $\text{var}(y) = \begin{bmatrix} Z_1 & Z_2 \end{bmatrix} \begin{bmatrix} A\sigma^2_{u_L} & A\sigma_{u_L,u_A} \\ A\sigma_{u_L,u_A} & A\sigma^2_{u_A} \end{bmatrix} \begin{bmatrix} Z_1' \\ Z_2' \end{bmatrix} + I\sigma^2_e$, where $\sigma^2_{u_L}$ is the lysine polygametic variance, $\sigma^2_{u_A}$ is the alanine polygametic variance and $\sigma_{u_L u_A}$ is the covariance between both. Parameters of both models were estimated using ASReml (Gilmour et al., 2006). The REML log-likelihood of model (1) and (2) was calculated and was denoted as $\log l_{m1}$ and $\log l_{m2}$, respectively.

For the test of putative *DGAT1 K232A* by polygene interaction effects, the null hypothesis was $H_0 : \sigma^2_{u_L} = \sigma^2_{u_A} = \sigma_{u_L,u_A}$. The alternative hypothesis was $H_1 : \sigma^2_{u_L} \neq \sigma^2_{u_A}$ or $\sigma^2_{u_L} = \sigma^2_{u_A} \neq \sigma_{u_L,u_A}$. Under the assumption of the null hypothesis, all three variance components in model (2) are the same and $\sigma^2_{u_L} + \sigma^2_{u_A} + 2 * \sigma_{u_L u_A}$ equals σ^2_u from model (1). In this case the expectations and variances of the observations are the same for each individual, and, hence, both models are equivalent. Therefore, the $\log l_{m1}$ is the log-likelihood under the null hypothesis. This leads to the restricted likelihood ratio test with $RLRT = 2(\log l_{m2} - \log l_{m1})$. With the assumption of between-subject independence, the asymptotic distribution of the *RLRT* under the null hypothesis follows a mixture of two χ^2-distributions with one and two degrees of freedom (Self and Liang 1987). Because the animals are related and hence the assumption of between-subject independence is not fulfilled, we used a conservative test with a χ^2-distribution with two degrees of freedom. The resulting comparisonwise error probability was denoted as p_c. Five tests were conducted, for each trait one. The correction for the resulting multiple testing was done using the Bonferroni correction, resulting in experimentwise error probabilities, $p_e = 1 - (1 - p_c)^5$.

Table 2. Average *DGAT1 K232A* lysine allele substitution effect (α) and polygenic variance component ($\hat{\sigma}_u^2$), results from model with one polygenic effect (i.e. model (1)).

Trait	$\hat{\alpha}$		$\hat{\sigma}_u^2$	
Fat (kg)	7.668	(0.857)	305.207	(39.844)
Fat (%)	0.281	(0.009)	416.567[1]	(42.205)
Protein (kg)	-6.508	(0.714)	209.009	(22.843)
Protein (%)	0.059	(0.005)	114.790[1]	(9.250)
Milk (kg)	-302	(23.741)	256367	(29672)

[1] Multiplied by 1000

Standard errors are given in parenthesis

The results of model (1) are presented in Table 2. The average substitution effect of the lysine variant is positive for fat yield and the two percentage traits and negative for protein and milk yield. These estimates are in agreement with Thaller et al. (2003) and Bennewitz et al. (2004) and also with those estimated in the French Holstein breed (Gautier et al. 2007). Allele substitution effects were all highly significant ($p < 0.001$). Results of model (2) and of the test for interaction effects are shown in Table 3. Experimentwise significant interaction effects were found for fat percentage ($p_e < 0.05$). Additional comparisonwise interaction effects were found for protein percentage ($p_c < 0.05$). These results agree with the interaction effects detected by Hinrichs et al. (2010). An additional weak, but not significant, interaction was found for fat yield. For these traits, the polygametic alanine variance is between 37% and 41% higher compared to the lysine variant (Table 3). Hence, the *DGAT1 K232A* lysine variant increased the mean of the phenotypes (Table 2) but decreased the polygametic variance. The correlation between u_L and u_A was below one for fat yield and especially fat percentage. This is not the case for protein percentage ($\hat{r} \approx 0.99$), which implies that the improved fit of model (2) is due to heterogeneous variances, i.e. $u_A = c * u_L$ (where $\hat{c} = 0.84$, not shown elsewhere). One possible explanation for the detected interaction effects might be the relationship between the enzyme activity and amount of end-product (i.e. the amount of milk, fat yield, fat percentage, protein yield and protein percentage). *DGAT1* catalyses the final step of the triglyceride synthesis. Grisart et al. (2004) found a higher enzyme activity level in producing triglycerides for the *DGAT1 K232A* lysine variant. Kacser and Burns (1973) developed a metabolic control theory, which modelled the phenotype as an end-product of enzyme activity. The enzyme activity causes a flux through metabolic pathways with a hyperbolic relationship. At a low flux level, an infinitesimal change of the enzyme activity results in a comparable larger change of the flux, and hence of the enzyme end-product (the

milk traits in our study). The opposite holds, if the flux level is already on a high level. Kacser and Burns (1981) used this theory to derive some conclusions for the dominance phenomenon, but it might also be valid in this case to explain the detected interaction effects. The limiting factor for a higher triglyceride synthesis might be the limited availability of the two substrates diacylglycerol and fatty acyl-CoA when forming triglycerides. If animals carry the favourable lysine alleles, the effect of additional favourable alleles at other genes might be reduced, because they have to compete for the two substrates.

Intuitively one might argue that it is not possible with these data to test for interaction effects. Genotypes and phenotypes are not recorded within the same generation and gene combinations break down during meiosis. However, the *DGAT1 K232A* frequencies among daughters of *DGAT1 K232A* homozygous sires and *DGAT1 K232A* heterozygous sires differ by one quarter, and among daughters of alternative *DGAT1 K232A* homozygous sires by one half. These daughters largely determine the DYD and EBV of the sires used in the study. With respect to this, it would have been better to estimate the DYD and EBV using a sire model and ignoring the relationships between the sires as done by Seidenspinner et al. (2009). However, this was not possible in this study. In theory, interaction between genes can be classified in additive x additive, additive x dominance etc. interactions (Falconer and Mackay 1996). The identified interaction terms are most likely due to an interaction between the additive *DGAT1 K232A* effect and the additive polygenic effect, because the phenotypes used (DYD and EBV) should contain a part of the additive x additive interaction (Falconer and Mackay 1996, p. 154). It may be noted that it is not possible with this data set to detect interaction effects within *DGAT1 K232A* (dominance), because the daughters only inherit the additive effect of the sires, but not their dominance effect.

Theoretically, the significant interactions could be used in breeding schemes. For each sire, three polygenic breeding values can be calculated, depending on the *DGAT1 K232A* genotypes of the sample, where the sire will be used. In a homozygous lysine (alanine) sample, the polygenic EBV would be $2*\hat{u}_L$ ($2*\hat{u}_A$) and in heterozygous sample it would be $\hat{u}_L + \hat{u}_A$. However, since the *DGAT1 K232A* genotypes of the samples are often unknown and usually heterogeneous, this is not a practical approach. Additionally, rank correlations between these three polygenic EBVs were always above 0.95 (data not shown), suggesting very little re-ranking of the sires.

Table 3. Polygametic variance and covariance components ($\hat{\sigma}^2_{u_L}, \hat{\sigma}^2_{u_A} \hat{\sigma}_{u_L u_A}$), correlation between the two polygametic terms ($\hat{r}$) (results from the model with two polygametic effects, i.e. model (2)), and restricted likelihood ratio test statistics (*RLRT*) with comparisonwise and experimentwise error probability (p_c and p_e, respectively).

Trait	$\hat{\sigma}^2_{u_L}$		$\hat{\sigma}^2_{u_A}$		$\hat{\sigma}_{u_L u_A}$		$\hat{r}$	*RLRT*	p_c	p_e
Fat (kg)	70.14	(10.60)	95.81	(14.07)	74.23	(11.37)	0.906	5.24	0.073	0.315
Fat (%)	94.42[1]	(12.01)	133.00[1]	(16.80)	87.06[1]	(14.51)	0.777	10.94	0.004	0.019
Protein (kg)	51.24	(2.14)	57.09	(2.38)	54.05	(2.25)	0.999	n.s.	n.s.	n.s.
Protein (%)	24.94[1]	(2.52)	35.28[1]	(3.76)	29.29[1]	(2.86)	0.988	6.98	0.031	0.146
Milk (kg)	64460	(2687)	63770	(2658)	60300	(2513)	0.941	n.s.	n.s.	n.s.

[1] Multiplied by 1000

Standard errors are given in parenthesis

Acknowledgements

The authors thank the VIT Verden w.V. for providing the data. The study was conducted within the FUGATO-plus project GeneDialog. MS was partly supported by a grant from the Deutsche Forschungsgemeinschaft, DFG.

References

Bennewitz, J., N. Reinsch, S. Paul, C. Looft, B. Kaupe et al. 2004. The DGAT1 K232A mutation is not solely responsible for the milk production quantitative trait locus on the bovine chromosome 14. J. Dairy Sci. 87:431 – 442.

Carlborg, Ö. and C. S. Haley. 2004. Epistasis: too often neglected in complex trait studies? Genetics 5:618 – 625.

Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to quantitative genetics. 4th ed. Longman Scientific and Technical, New York, NY.

Gautier, M., A. Capitan, S. Fritz, A. Eggen, D. Boichard and T. Druet. 2007. Characterization of the DGAT1 K232A and variable number of tandem repeat polymorphisms in French dairy cattle. J. Dairy Sci. 90:2980-2988.

Gilmore, A. R., B. J. Gogel, B. R. Cullis and R. Thompson. 2006. ASReml User Guide 2.0. VSN International Ltd., Hemel Hempstead, HP1 1ES, UK.

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford et al. 2002. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 K232A gene with major effect on milk yield and composition. Genome Res. 12:222 – 231.

Grisart, B., F. Farnir, L. Karim, N. Cambisano, J.-J. Kim, A. Kvasz, M. Mni, P. Simon, J.-M. Frère, W. Coppieters, and M. Georges. 2004. Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and confirmation. Proc. Natl. Acad. Sci. USA 101:2398–2403.

Hill, W. G., M. E. Goddard, and P. M. Visscher. 2008. Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 4(2): e1000008.doi:10.1371/journal.pgen.1000008.

Hinrichs, D., J. Bennewitz, J. Tetens, N. Reinsch, and G. Thaller. 2010. A two step approach to detect interaction between QTL and a known major gene in dairy cattle. Communication No. 379, 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany.

Kacser, H., and J. A. Burns. 1973. The control of flux. Symp. Soc. Exp. Biol. 27:65-104.

Kacser, H., and J. A. Burns. 1981. The molecular basis of dominance. Genetics 97:639-666.

Seidenspinner, T., J. Bennewitz, F. Reinhardt, and G. Thaller. 2009. The need for sharp phenotypes in QTL mapping for calving traits in dairy cattle. J. Anim. Breed. Genet. 126:455-462.

Self, S. G., and K. Liang. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. Journal of the American Statistical Association 82:605-610.

Thaller, G., W. Krämer, A. Winter, B. Kaupe, G. Erhardt and R. Fries. 2003. Effects of DGAT1 K232A variants on milk production traits in German cattle breeds. J. Anim. Sci. 81:1911 – 1918.

Winter, A., W. Krämer, F. A. O. Werner, S. Kollers, S. Kata et al. 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 percentage. Proc. Natl. Acad. Sci. USA 99:9300 – 9305.

CHAPTER TWO

Reaction norms and genotype by environment interaction in the German Holstein dairy cattle

MELANIE STREIT[1], FRIEDRICH REINHARDT[2], GEORG THALLER[3], and JÖRN BENNEWITZ[1]

[1]Institute of Animal Husbandry and Breeding,
University of Hohenheim,
Stuttgart, Germany
[2]Vereinigte Informationssysteme Tierhaltung w.V. (vit),
Verden, Germany
[3]Institute of Animal Breeding and Husbandry,
Christian-Albrechts-University,
Kiel, Germany

Published in:
Journal of Animal Breeding and Genetics (2012) 129: 380 – 389
doi:10.1111/j.1439-0388.2012.00999.x
ISSN 0931-2668

Reaction norms and genotype by environment interaction in the German Holstein dairy cattle

M. Streit[1], F. Reinhardt[2], G. Thaller[3], J. Bennewitz[1]

[1]Institute of Animal Husbandry and Breeding, University of Hohenheim, Stuttgart, Germany
[2]Vereinigte Informationssysteme Tierhaltung w.V. (vit), Verden, Germany
[3]Institute of Animal Breeding and Husbandry, Christian-Albrechts-University, Kiel, Germany

Corresponding author:
Jörn Bennewitz
Garbenstraße 17
70599 Hohenheim
Germany
phone: +49 711-459 23570
fax: +49 711-459 23101
E-mail: j.bennewitz@uni-hohenheim.de

Summary

Reaction norm random regression sire models were used to study genotype by environment interactions (GxE) in the German Holstein dairy cattle population. Around 2300 sires with a minimum of 50 daughters per sire and seven first lactation test day observations per daughter were analysed. Corrected test day records for milk yield, protein yield, fat yield and somatic cell score were used. Herd test day solutions for milk traits, milk energy yield or somatic cell score were used as environmental descriptors. Second order orthogonal polynomial regressions were applied to the sire effects. The results revealed significant slope variances of the reaction norms, which caused a non-constant additive genetic variance across the environmental ranges considered. This pointed to the presence of minor GxE effects. The additive genetic variance increased when the environment improved, i.e. higher (lower) herd test day solutions for milk traits (somatic cell score). This was also influenced by pure scaling effects, because the non-genetic variance increased in an improved environment and the heritability was less influenced by the environment. The GxE effects caused very little re-ranking of the sires for the environmental range considered in this study.

Key words: environmental sensitivity, random regression, sire model

Introduction

Genotype by environment interaction (GxE) refers to differences in response of genotypes to changes in the environment (Lynch and Walsh, 1998). If GxE effects are not taken into account, estimated breeding values (EBVs) may be biased and selection response reduced. Additionally, knowledge about existing GxE effects can be used in breeding schemes, e.g. to breed robust animals (Knap, 2005). Consequently, GxE effects have received considerable attention in dairy cattle breeding (see König et al. (2005), Kolmodin et al. (2002), Kolmodin et al. (2004) and references therein). Dairy farm environments in Germany vary in several ways, e.g. north vs. south, west vs. east, small vs. large farms or flat vs. mountainous areas. König et al. (2005) investigated putative GxE effects in this population. They defined geographical region and herd size as distinct environments and applied a multi-trait approach. Genetic correlations between eastern and western Germany were between 0.9 and 0.95 for protein yield, indicating only minor GxE re-ranking effects. Larger effects were reported when herd size was considered as an environmental descriptor.

The use of a multi-trait approach is a logical choice if the environment can be regarded as a distinct variable. The use of reaction norms, however, might be more appropriate if the environment changes gradually and can be measured on a continuous scale. This is because fewer parameters have to be estimated and there is no need to cluster individuals in different environmental classes. A reaction norm describes the performance of a genotype as a function of a gradually changing environment (Lynch and Walsh, 1998). The first derivative of the reaction norm function, the slope, is the environmental sensitivity. A non-zero genetic variation of the slope indicates the presence of GxE effects. In dairy cattle breeding, analysis of GxE effects with reaction norm models was applied by several authors (e.g. Fikse et al. (2003), Hayes et al. (2003), Calus et al. (2002), Kolmodin et al. (2002), Strandberg et al. (2009), or Lillehammer et al. (2009)). In those studies, the average herd production level served as a continuous environmental descriptor and the reaction norms were estimated using random regression models. Schaeffer (2004) advised to apply not only linear but also higher order orthogonal polynomial regression models in GxE studies. Calus and Veerkamp (2003) and Lillehammer et al. (2009) pointed out the importance of modelling heterogeneous residual variances, because otherwise matter of scales (Lynch and Walsh 1998) might be interpreted as GxE effects.

The aim of the present study was to analyse GxE effects for milk production traits and somatic cell score in German Holstein dairy cattle using non-linear reaction norms implemented in random regression sire models.

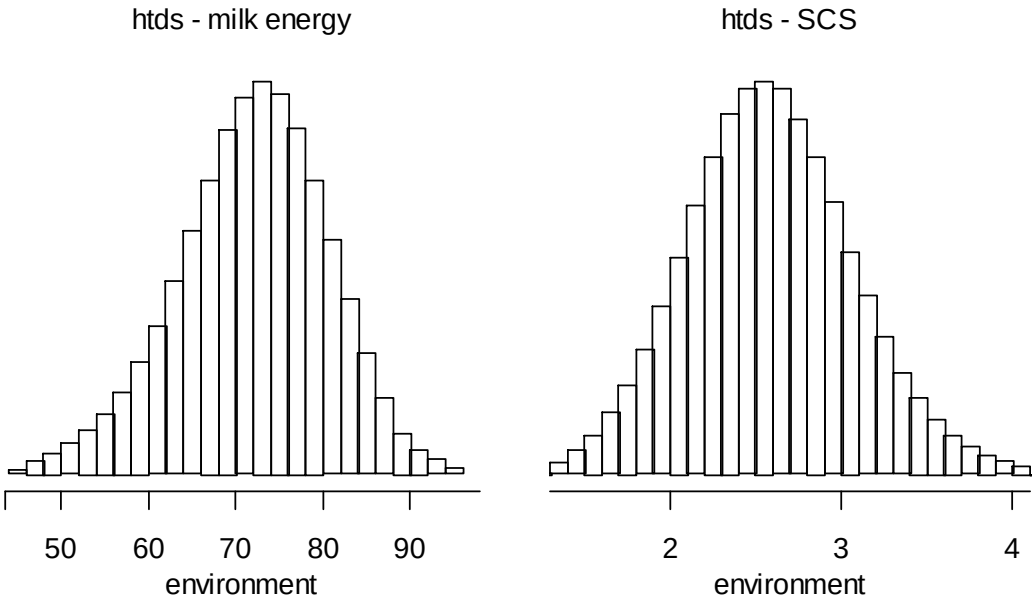

Figure 1: Histogram of the environmental descriptor milk energy yield and SCS.

Materials and Methods

Data and data editing

First lactation test day records of cows were taken from the routine animal recording scheme for milk yield, fat yield, protein yield and somatic cell score (SCS). The observations were corrected for the fixed effects herd test day, days in milk, age at calving, calving season and the random permanent environment effect. These correction factors were obtained from the routine animal evaluation, which is an animal test day model. Only daughters with at least seven observations per year and with a variation of the corresponding environmental descriptor above a certain threshold level were considered. Observations in extreme and rare environments were discarded (around two percent of the observations). These restrictions ensured that there were enough observations and variation of the environmental descriptor to apply reaction norms within individual cows and that the results were not affected by (unreliable) extreme and rare environments. The total number of daughters was around 1.3 million and the number of observations was around 12 million. The number of daughters per bull ranged from 50 to 74842 and the number of bulls was around 2300, see Table 1. The sires were born between 1983 and 2003.

Table 1: Summary statistics of the herd test day solutions (htds) used as environmental descriptors and number of sires and daughters.

environmental descriptor	unit	n	μ	sd	min	max	n_{sire}	$n_{daughter}$
htds – protein	kg	11927970	0.755	0.101	0.272	1.675	2279	1284531
htds – fat	kg	11927970	0.938	0.111	0.295	1.583	2279	1284531
htds – milk	kg	11927970	22.188	2.742	8.305	35.859	2279	1284531
htds – energy in milk	MJ	11927970	71.631	8.364	45.000	95.999	2279	1284531
htds – SCS	-	12056462	2.583	0.475	1.398	4.052	2291	1300833

The environmental descriptors were herd test day solutions (*htds*) for milk, protein and fat yield (all in kg), and SCS. These were obtained from routine animal evaluation (see Table 1 for summary statistics). An initial comparison of the three milk trait environmental descriptors revealed a high correlation between them (Table 2). Furthermore, it widely known that different levels of feeding or housing conditions result in different milk production levels. It is desirable to have only one environmental descriptor instead of three highly correlated descriptors that reflect these kinds of environmental forces. Therefore, we calculated the *htds* for milk energy yield (in MJ) from the *htds* of the three milk yields by a linear combination (following Nostitz and Mielke, 1995) as

$$energy\ yield = 0.802 * milk\ yield + 38.4 * fat\ yield + 23.6 * protein\ yield.$$

The equation assumes constant lactose content of 4.8 %. It is known that lactose content is not constant, but lactose content was not available.

The *htds* were scaled between -1 and 1 following Calus et al. (2002). The following random regression sire model was applied to all combinations of traits and environmental descriptors:

$$cy_{ijk} = \mu + b_1 * htds_k + b_2 * htds_k^2 + \sum_{m=0}^{u} s_{jm} * P_{km} + \sum_{m=0}^{v} d_{ijm} * P_{km} + e_{ijk},$$

where *cy* are the corrected yields of daughter *i* of sire *j* at herd test day *k*, μ is the overall mean, *htds* (*htds*^{2}) is the herd test day solution at herd test day *k* (herd test day solution at herd test day *k* squared) with the fixed regression coefficients b_1 (b_2), P_{km} are the covariables based on Legendre polynomials and related to the standardised *htds* at test day *k* of order *m*, *s* is the random sire effect of sire *j* of order *m*, *d* the random daughter effect of daughter *i* of sire *j*, *u* and *v* are the highest order of the polynomial regressions, and *e* is the random residual. The variance of the sire regression effects is (for $u = 1$, index *j* is suppressed for simplicity)

$$Var\begin{bmatrix} s_0 \\ s_1 \end{bmatrix} = A \otimes S = A \otimes \begin{bmatrix} \sigma^2_{s_0} & \sigma_{s_0 s_1} \\ \sigma_{s_0 s_1} & \sigma^2_{s_1} \end{bmatrix},$$

where *A* is the numerator relationship matrix and *S* is the covariance matrix of the random regression coefficients for sire effects. The variance structure of the regression coefficients for daughter effects is (for $v = 1$, indices *i* and *j* are suppressed)

$$Var\begin{bmatrix} d_0 \\ d_1 \end{bmatrix} = I \otimes D = I \otimes \begin{bmatrix} \sigma^2_{d_0} & \sigma_{d_0 d_1} \\ \sigma_{d_0 d_1} & \sigma^2_{d_1} \end{bmatrix},$$

where *I* is the identity matrix and *D* is the covariance matrix of the random regression coefficients for daughter effects. For higher orders ($u > 1$, $v > 1$) the terms are expanded straightforwardly. In order to model heterogeneous residual variance across the environments the observations were ordered according to the environmental descriptor and grouped into ten classes of equal size based on the environmental values. Residual variances were estimated for each class and residual covariance was assumed to be zero. Hence, $Var(e) = X'KX$, with *X* being a known design matrix that assigns each observation to an environmental class *i*, and $K = Diag\{\sigma^2_{e_i}\}$ (Lillehammer et al. 2009).

Table 2: Correlations between herd test day solutions (htds) environmental descriptors.

	fat yield	milk yield	SCS	milk energy yield
protein yield	0.846	0.956	-0.176	0.955
fat yield		0.839	-0.162	0.961
milk yield			-0.197	0.949
SCS				-0.182

Random regression models were solved using ASReml 3.0 (Gilmour et al., 2009). A critical question was the choice of the appropriate order of the random regression models, i.e. of *u* and *v*. For the uncorrelated daughter effects this was very clearly $v = 1$ for all trait and environment combinations except for both SCS as trait and as environment (see results section). For SCS as trait and environment $v = 0$ was used. The choice of the appropriate order of the random sire effect was made by applying a restricted likelihood ratio test, because a model of a lower order is nested within a model of a higher order for a given *v*. Additionally this was done by visual inspection of the eigenvalues of the estimated covariance matrix, as suggested by Kirkpatrick et al. (1990).

Once variance components were calculated, several genetic parameters were estimated for a defined environmental value *E* to evaluate possible GxE effects. Heritability as a function of *E* was estimated as:

$$h^2 \mid E = \frac{\hat{\sigma}^2_{a|E}}{\hat{\sigma}^2_{s|E} + \hat{\sigma}^2_{d|E} + \hat{\sigma}^2_{e|E}},$$

where $\sigma^2_{s|E}$ ($\sigma^2_{d|E}$) is the variance of the sire (daughter) as a function of the environment, $\sigma^2_{a|E} = 4 * \sigma^2_{s|E}$ is the additive genetic variance, and $\sigma^2_{e|E}$ is the residual variance of the environmental class that included the environmental value *E*. The $\sigma^2_{s|E}$ was calculated as $\sigma^2_{s|E} = L \mid E * \hat{S} * L \mid E'$ and the $\sigma^2_{d|E}$ as $\sigma^2_{d|E} = L \mid E * \hat{D} * L \mid E'$. The vector $L \mid E$ contains the Legendre polynomials for *E*. The genetic correlation between different environmental values was calculated from the entries of the covariance matrix G of the additive effects at the environmental values. The matrix G was calculated as

$$\hat{G} = 4 * \Phi \hat{S} \Phi',$$

with Φ being a matrix with polynomial coefficients for a defined value *E* on each row. These calculations are based on Kirkpatrick et al. (1990) and are nicely described in Schaeffer (2010). *EBV|E* of a sire was calculated as

$$EBV \mid E = L \mid E' * \hat{s},$$

with $\hat{s}$ being a vector of order u containing the estimated random regression coefficients of the sire. Putative GxE effects were evaluated by investigating the change in additive genetic variance along *E*, the genetic correlation and the *EBV|E* rank correlation between several environmental values.

Table 3: Model numbers with corresponding trait names and environmental descriptors.

model number	trait	environmental descriptor
1	protein	htds – milk energy yield
2	fat	htds – milk energy yield
3	milk	htds – milk energy yield
4	protein	htds – SCS
5	fat	htds – SCS
6	milk	htds – SCS
7	SCS	htds – SCS

Results

The restricted likelihood ratio test showed that a second order sire effect (u = 2) was significantly superior to a first order one ($p < 0.001$ in all cases). Higher order sire effects (u = 3) were significant for some traits and environments, but the variance component of the higher order level was very small and convergence was very slow (not shown). Additionally, the last eigenvalues of the covariance matrix *S* for u = 3 were very small compared to those of all other models (not shown), indicating that an additional order explains only a small amount of the observed phenotypic variation. Therefore, a second order sire effect (u = 2) was used in all models. For the uncorrelated daughter effect, v = 1 in nearly all cases, except for the model that used SCS as trait and as environment, as stated above. Higher order models did not converge and a lower order (v = 0) resulted in a substantially lower restricted log-likelihood value (not shown).

The results of the models using the milk traits and the *htds* of the corresponding milk traits were almost the same as when using the *htds* of the milk energy yield as an environmental descriptor. This was expected due to the high correlation between the environmental descriptors (Table 2). Therefore, only the results of the seven models with milk energy yield or SCS as environment (see Table 3) are reported. The distribution of these two environmental descriptors is presented in Figure 1. Both are approximately normally distributed, but with some skewness.

Table 4: Sire variance components of the random regression analyses. Standard errors are shown in parenthesis.

model number*	$\hat{\sigma}^2_{s_0}$ **	$\hat{\sigma}^2_{s_1}$ **	$\hat{\sigma}^2_{s_2}$ **	$\hat{\sigma}_{s_0 s_1}$ ***	$\hat{\sigma}_{s_0 s_2}$ ***	$\hat{\sigma}_{s_1 s_2}$ ***
1	1822.55 (56.2)	52.571 (3.550)	1.572 (0.337)	180.041 (10.698)	7.549 (4.241)	6.644 (0.999)
2	3132 (96.966)	63.487 (4.731)	1.223 (0.379)	326.907 (17.472)	2.355 (7.598)	3.843 (1.294)
3	2.18 (0.067)	0.056 (0.004)	0.02 (<0.001)	0.245 (0.012)	0.005 (0.005)	0.006 (0.001)
4	1968.490 (60.439)	5.620 (0.914)	0.78 (0.162)	<-0.001 (<0.001)	0.211 (3.512)	-0.637 (0.308)
5	3293.2 (101.548)	19.694 (2.293)	0.77 (0.297)	-134.035 (13.179)	-13.058 (6.983)	-1.567 (0.693)
6	2.359 (0.072)	0.007 (0.001)	<0.001 (<0.001)	-0.068 (0.007)	-0.001 (0.003)	-0.001 (<0.001)
7	0.099 (0.003)	0.001 (<0.001)	<0.001 (<0.001)	0.003 (<0.001)	-0.001 (<0.001)	<-0.001 (<0.001)

* The trait and environmental descriptor of the models are given in Table 3.

** $\hat{\sigma}^2_{s_0}$ is the sire variance of the intercept, $\hat{\sigma}^2_{s_1}$ is the sire variance of the slope, and $\hat{\sigma}^2_{s_2}$ is the second order sire variance.

*** $\hat{\sigma}_{s_0 s_1}$, $\hat{\sigma}_{s_0 s_2}$ and $\hat{\sigma}_{s_1 s_2}$ are the covariances between intercept, slope and the second order sire variance

The estimated variance components for the sire effects are shown in Table 4 and for the daughter effect in Table 5. The significance of the higher order sire effects point to the presence of GxE effects for all combinations of traits and environments. The linear sire effect was much more important than the second order effect in explaining the variation of the observations. The additive genetic variance as a function of the environmental value, $\sigma^2_{a|E} = 4 * \sigma^2_{s|E}$, is shown in Figure 2. For the milk traits and the milk energy as environment this variance was increasing substantially with an increase in the environment (e.g. from 3 to 6.6 kg^2 for milk yield and from 280 to 530 g^2 for fat yield). The opposite was observed for fat and milk yield and SCS as the environment. The additive genetic variance of protein yield was almost constant over the range of SCS. For SCS the additive genetic variance increased slightly with an increase in SCS. A similar pattern could be observed for the daughter variance $\sigma^2_{d|E}$ (Figure 3) when using milk energy as environment. For SCS there seems to be the lowest daughter variance for intermediate SCS levels (Figure 3). Note that for both, SCS as trait and environment, $v = 0$, hence, the variance was not affected by the environment.

Table 5: Daughter variance components of the random regression analyses. Standard errors are shown in parenthesis.

model number*	$\hat{\sigma}^2_{d_0}$ **	$\hat{\sigma}^2_{d_1}$ **	$\hat{\sigma}_{d_0 d_1}$ **
1	3212.67 (6.464)	544.457 (7.973)	645.489 (4.329)
2	5703.13 (11.892)	864.755 (15.065)	991.44 (8.041)
3	4.41 (0.008)	0.913 (0.008)	0.871 (0.005)
4	3814.31 (6.741)	163.842 (7.035)	<0.001 (<0.001)
5	6502.85 (12.170)	505.530 (14.031)	-204.036 (8.616)
6	5.271 (<0.001)	0.335 (<0.001)	<0.001 (<0.001)
7	0.01 (<0.001)	–	–

* The trait and environmental descriptor of the models are given in Table 3.

** $\hat{\sigma}^2_{d_0}$ is the variance of the intercept of the daughter, $\hat{\sigma}^2_{d_1}$ is the variance of the slope of the daughter and $\hat{\sigma}_{d_0 d_1}$ is the covariance between intercept and slope

The residual variance components, $\sigma^2_{e|E}$, are shown in Table 6. The residual variance was clearly heterogeneous and increased with environmental values for those models that used milk energy as an environment. Due to the abrupt change of the residual variance across the classes the trace of the heritability along the environmental values, $h^2 \mid E$, was not a smooth plot, but was peaked at several points. Because these peaks were an artefact of the somewhat arbitrary choice of the ten environmental values, they are not reported. The trace would become smoother if the number of environmental classes was increased. But in general the heritability is much less affected by the environment as the additive genetic variance. The heterogeneous residual variance was less obvious if SCS was the environment.

The genetic correlations of the traits at selected percentile of environments are shown in Table 7. They are generally high (>0.9) even for the two extreme environments (0^{th} – 100^{th} quantile). *EBV|E* rank correlations were all above 0.95 (not shown), indicating almost no re-ranking.

Table 6: Residual variance components within the ordered environmental classes.

environmental	model number*						
class**	1	2	3	4	5	6	7
1	5069.2	10161.7	3.948	6104.9	11928.2	4.497	0.585
2	5565.2	10921.2	4.182	6242.9	12094.7	4.581	0.643
3	5906.9	11455.1	4.355	6318.4	12256.1	4.626	0.694
4	6154.6	11782.2	4.497	6381.4	12355.6	4.661	0.728
5	6328.2	12079.4	4.564	6348	12345	4.657	0.768
6	6520.6	12429.1	4.674	6377.1	12393.4	4.683	0.807
7	6676.5	12767.8	4.763	6332.1	12376.3	4.659	0.849
8	6865.7	13164.9	4.902	6333.3	12365.3	4.669	0.901
9	7048.5	13705.2	5.036	6258.8	12344.3	4.636	0.979
10	7388.4	14940.1	5.322	6005.3	12102.8	4.492	1.139

* The trait and environmental descriptor of the models are given in Table 3.

** Class 1 (10) contains the lowest (highest) environmental values.

Table 7: Genetic correlation between selected percentile of the environmental descriptor.

model number*	r_g $0^{th} - 100^{th}$ percentile	r_g $5^{th} - 95^{th}$ percentile	r_g $25^{th} - 75^{th}$ percentile
1	0.895	0.904	0.938
2	0.945	0.950	0.968
3	0.925	0.932	0.956
4	0.983	0.985	0.990
5	0.974	0.977	0.986
6	0.987	0.989	0.993
7	0.960	0.964	0.978

* The trait and environmental descriptor of the models are given in Table 3.

Discussion

The present study investigates GxE effects in German Holsteins using non-linear reaction norm models. A significant substantial slope variance resulted in a varying additive genetic variance as a function of the changing environment and in a genetic correlation slightly lower than one between a trait evaluated at different environmental values. This points to the presence of minor GxE effects. In animal breeding GxE scaling and re-ranking effects are usually distinguished. The reaction norm models do not differentiate between scaling and re-ranking effects and therefore putative re-ranking effects were evaluated in this study by rank-correlations between EBVs as a function of the environment. The GxE effects did not result in re-ranking of sires for the environmental range considered in this study.

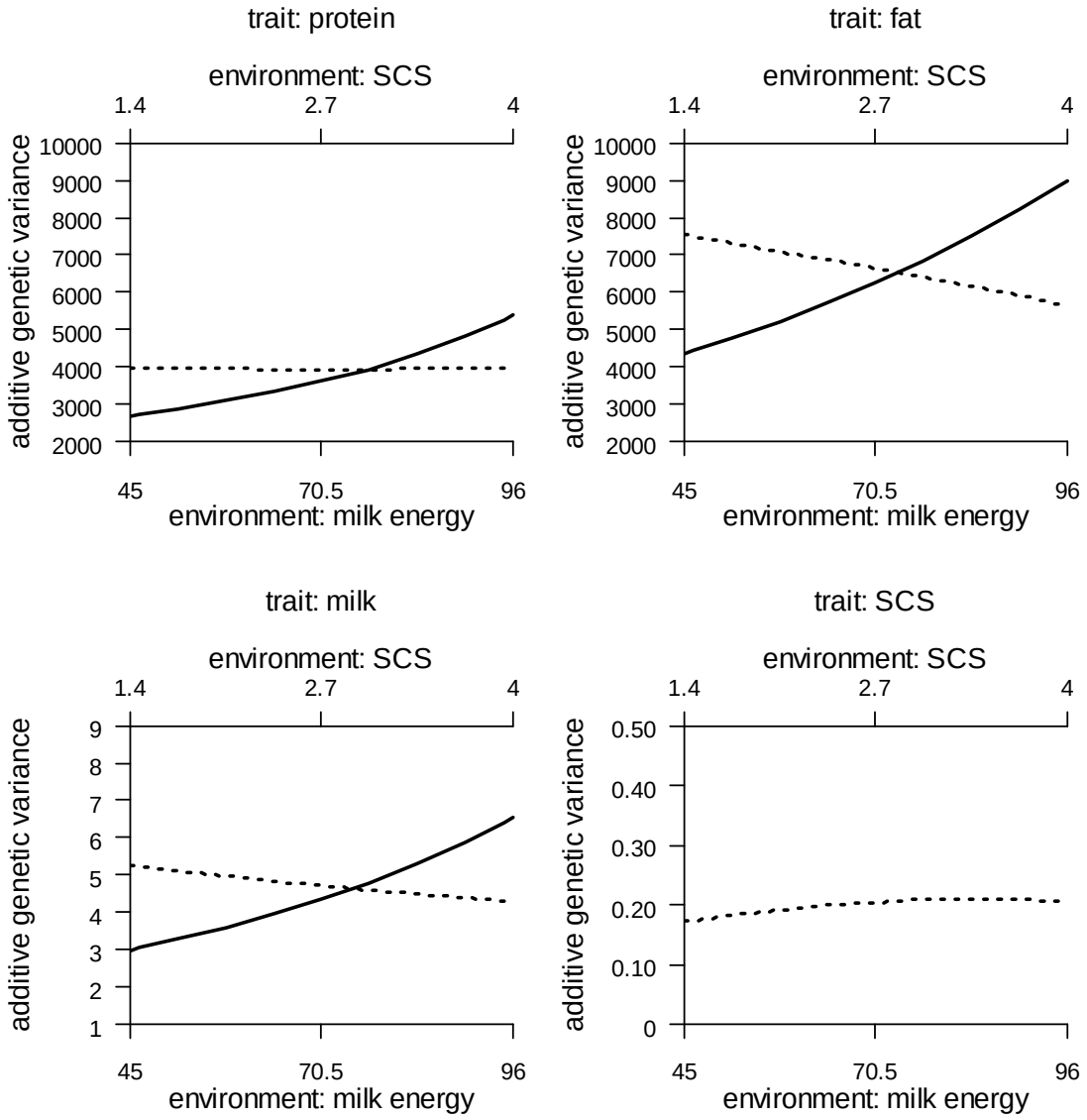

Figure 2: Additive genetic variance as a function of the environment (solid – milk energy environment, dotted – SCS environment)

The reaction norm models applied can be seen as an alternative to the multi-trait approach applied by König et al. (2005) to the same population. They also found few re-ranking effects for the environments they considered. In their study the environment was regarded as a distinct variable and the observations had to be grouped in the environmental classes; individual differences within the classes were not considered. In contrast, the reaction norm model uses the entire environmental variance. A critical question is the appropriate choice of the environmental descriptor. Several authors (e.g. Hayes et al. (2003), Calus et al. (2002), Kolmodin et al. (2002), Strandberg et al. (2009), or Lillehammer et al. (2009)) showed that

the production level of the herd can be a useful environmental descriptor. This is also a very convenient choice, because reliable herd production levels can be obtained from the routine sire evaluation. Following this, we used the average herd production level for milk traits or SCS, which were available as herd test day solutions from the routine breeding value evaluation. It is postulated that different levels of feeding result in different milk production levels. The milk energy production level might be the most appropriate descriptor, because it includes milk, fat and protein yields and should thus be more sensitive to the feeding level than one of the milk traits alone.

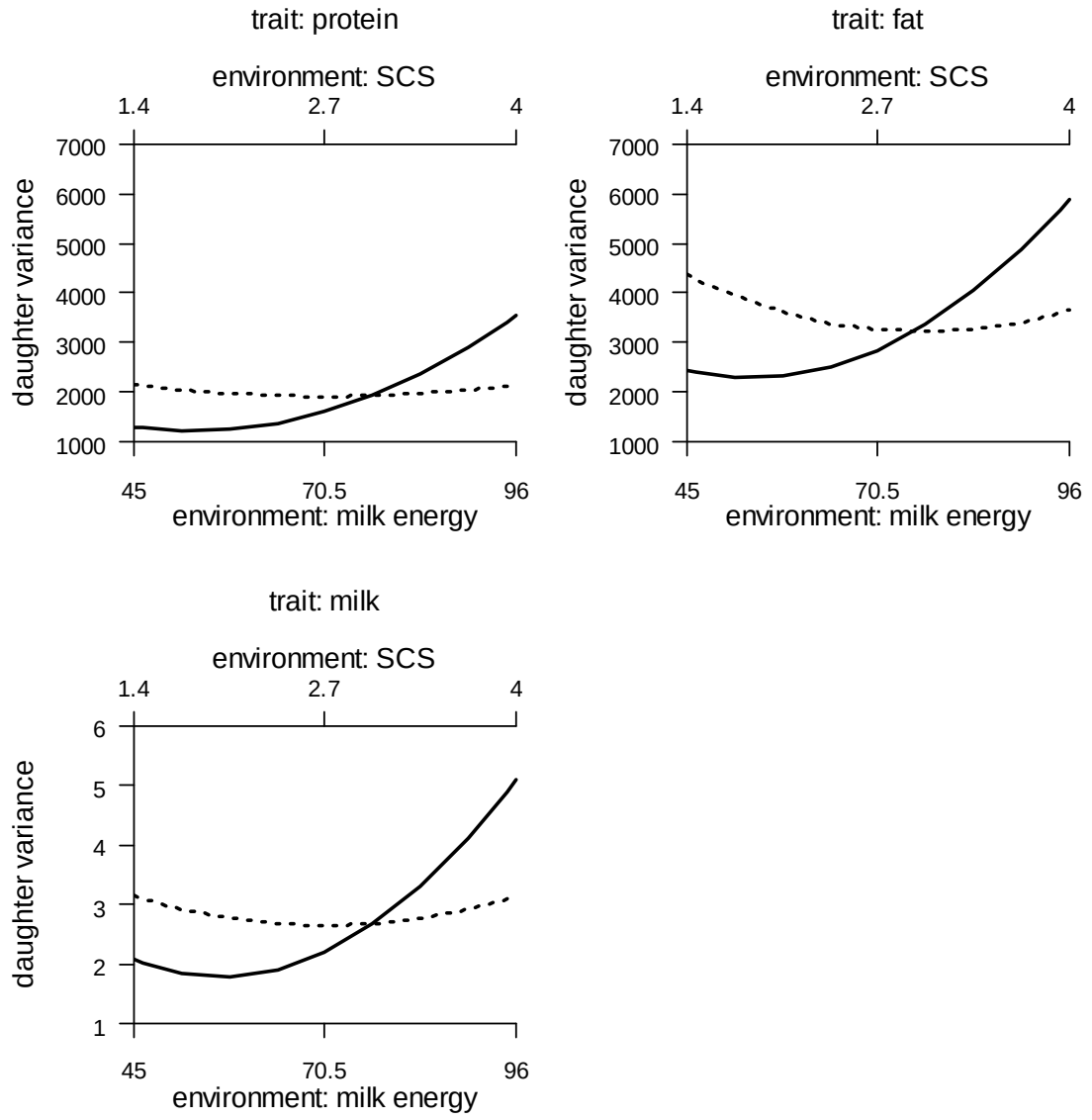

Figure 3: Daughter variance as a function of the environment (solid – milk energy environment, dotted – SCS environment). Note that for the trait SCS the daughter variance was constant over the range of the environment.

For all milk traits and milk environmental descriptors the correlation between sire intercept and first order sire effect is positive ($\hat{\sigma}_{s_0 s_1} > 0$, see Table 4), which is in agreement with Kolmodin et al. (2002), Calus et al. (2002), and Hayes et al. (2003). In Figure 4 the reaction norms of a sample of 15 sires (five with a steep positive, five with a steep negative and five with a flat slope) are presented. The reaction norms are almost linear, which underlines the dominance of the first order sire effect in explaining the environmental sensitivity. For milk traits as observation and milk energy yield as environment it can be seen that in a 'bad' environment the differences are smaller compared to a 'good' environment. Hence, it seems more beneficial for dairy farmers with a 'good' environment to invest in high quality semen than for farmers in a 'bad' environment.

When using herd test day solutions for SCS as an environmental descriptor, it is assumed that the level of hygiene on the farms, especially the infection pressure on the udder, is reflected. Thus, the decreased additive genetic variance for milk and fat yield with an increase of SCS environmental level (Figure 2) was expected. The negative correlation between sire intercept and first order sire effect ($\hat{\sigma}_{s_0 s_1} < 0$, see Table 4) indicates that bulls with a high (low) intercept generally have a negative (positive) slope. This can also be visualized by looking at the reaction norms of a sample of bulls (Figure 5). The reaction norms are generally somewhat flatter compared to the results obtained when milk traits are used as the environmental descriptor (Figure 4). Hence, GxE effects are more obvious when using milk energy yield as environmental descriptors. This is also underlined by the plots of the additive genetic variance, where variance is less sensitive for a changing SCS environment (Figure 2), especially for protein yield. The low daughter variance at intermediate SCS-levels (Figure 3) might be due to a non-linear relationship between SCS and udder health. In 'bad' environments, SCS may measure the ability of the cow to avoid mastitis, which is heritable. In 'good' environments, somatic cells may be considered as background noise, which is also heritable. In intermediate environments both are confounded, resulting in a lower heritability. This hypothesis is, however, highly speculative and is not supported by the trace of the additive genetic variance.

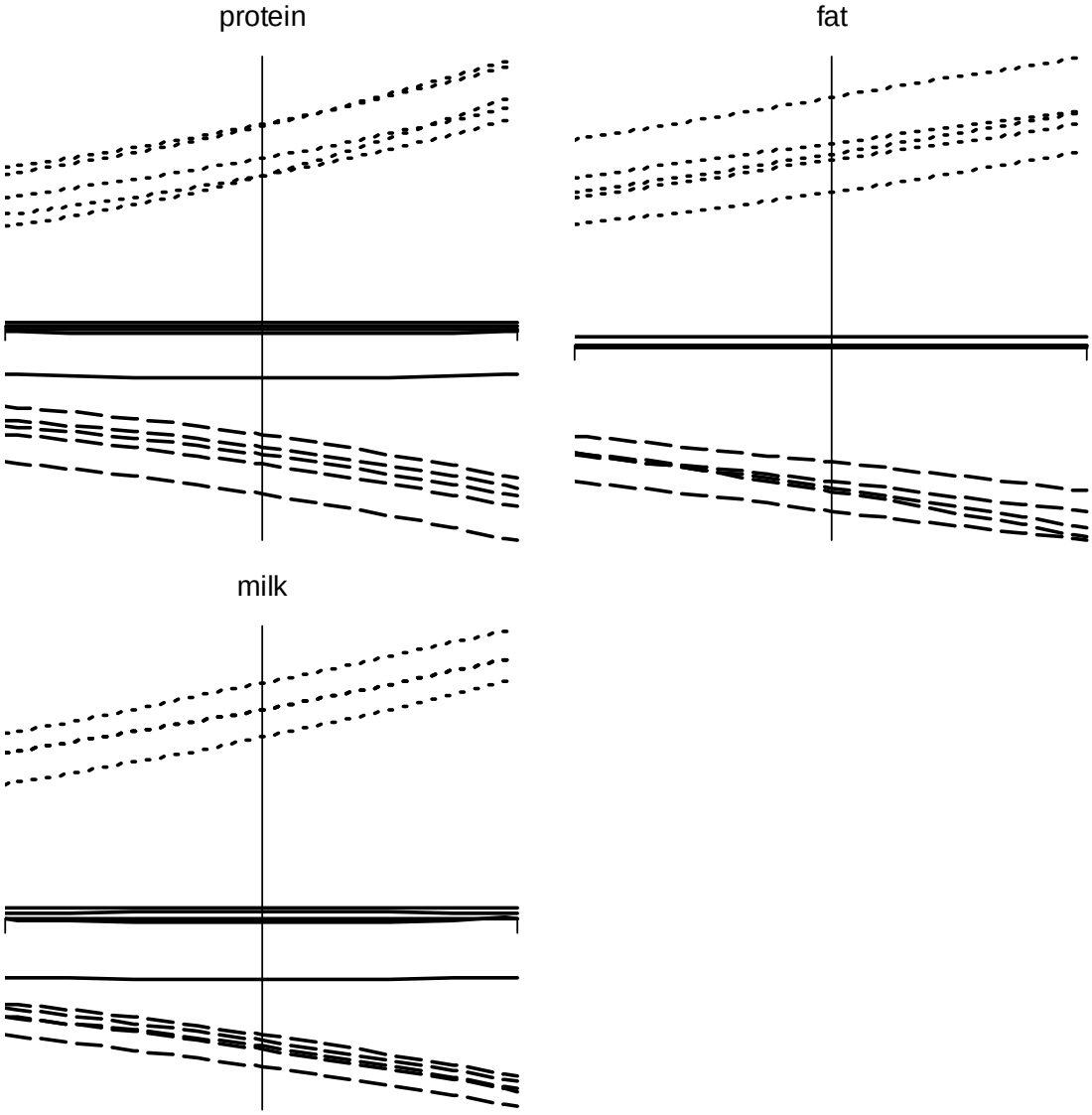

Figure 4: Reaction norms of a sample of 15 sires with a steep positive, a steep negative, and a flat slope, respectively, with milk energy yield as environmental descriptor.

Lillehammer et al. (2009) showed that the application of a sire reaction norm model might result in heterogeneous error variance, because usually three quarters of the additive genetic variance becomes part of the error variance in a sire model. Hence, if the sire variance is heterogeneous across the environmental range, the error variance is heterogeneous, too. In order to avoid this, we fitted uncorrelated daughter effects, which captured a remaining part of the additive genetic variance, and a part of the within-cow variance. Indeed, the daughter variance component was generally quite substantial (Table 5). However, this did not remove all the heterogeneity of the residual variance and a class-wise residual variance estimation was necessary (Table 6). This modelling of the residual variance is in close agreement with the so-

called IC-model of Lillehammer et al. (2009). For a sensible estimation of the heritability across the environment it would have been better to build more classes than only ten (as done in this study), because the estimated residual variance component would better match the environment under consideration.

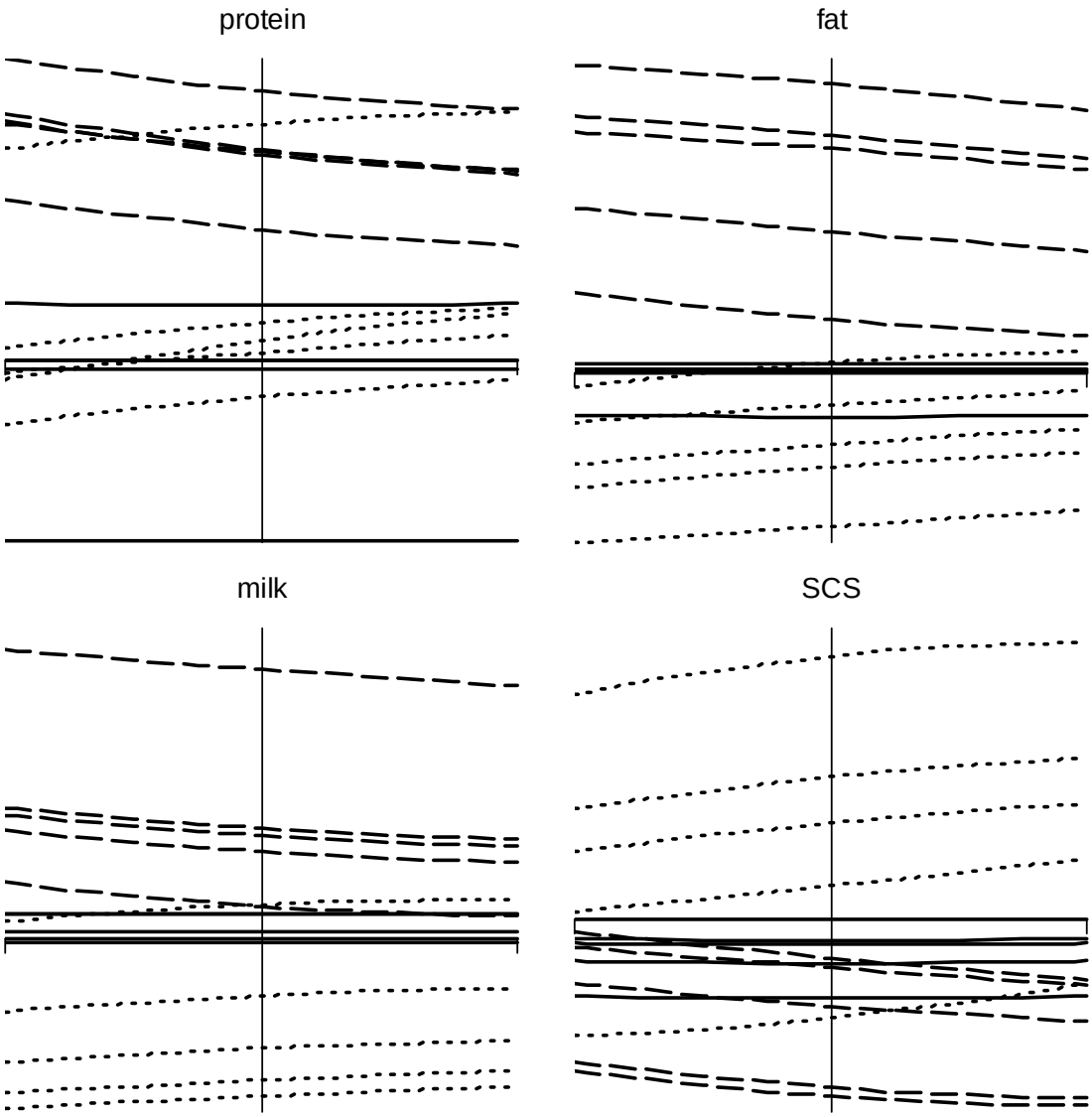

Figure 5: Reaction norms of a sample of 15 sires with a steep positive, a steep negative, and a flat slope, respectively, with SCS as environmental descriptor.

The increase of the residual variance with an increase in the environmental value and the less sensitive heritability with respect to the environment indicates that a part of the increase in the additive genetic variance is due to scaling effects, also caused by GxE. It is, however, difficult to quantify how much of the change is due to GxE effects. This problem of interpreting

reaction norm results was also described by Calus and Veerkamp (2003). The observations used were corrected for some systematic effects and for the effect of the herd test day. The herd test day solutions were adjusted for heterogeneous variances. Hence, it can be assumed that the GxE results are somewhat conservative, because some GxE effects were removed during the adjustment of the herd variances.

Conclusions

The results of the reaction norm models point towards the presence of minor GxE effects for milk traits and SCS in the German Holstein population. These did not result in re-ranking effects of sires for the environmental range investigated. Modelling heterogeneous residual variance played an important role in obtaining unbiased genetic parameters.

Acknowledgments

The manuscript has benefited enormously from the critical and helpful comments of three anonymous reviewers. The authors thank Chris Baes for language corrections. MS was supported by a grant from the Deutsche Forschungsgemeinschaft, DFG.

References

Calus, M. P. L., Groen, A. F., Jong, G. de (2002) Genotype x Environment Interaction for Protein Yield in Dutch Dairy Cattle as Quantified by Different Models. J. Dairy Sci., 85, 3115 – 3123.

Calus, M. P. L., Veerkamp, R. F. (2003) Estimation of Environmental Sensitivity of Genetic Merit for Milk Production Traits Using a Random Regression Model. J. Dairy Sci., 86, 3756 – 3764.

Fikse, W. F., Rekaya, R., Weigel, K. A. (2003) Assessment of environmental descriptors for studying genotype by environment interaction. Livest. Prod. Sci., 82, 223 – 231.

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Thompson, R. (2009) ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Hayes, B. J., Carrick, M., Bowman, P., Goddard, M. E. (2003) Genotype x Environment Interaction for Milk Production of Daughters of Australian Dairy Sires from Test-Day Records. J. Dairy Sci., 86, 3736 – 3744.

Kirkpatrick, M., Lofsvold, D., Bulmer, M. (1990) Analysis of the Inheritance, Selection and Evolution of Growth Trajectories. Genetics, 124, 979 – 993.

Knap, P. W. (2005) Breeding robust pigs. Aust. J. Exp. Agr., 45, 763 – 773.

König, S., Dietl, G., Raeder, I., Swalve, H. H. (2005) Genetic Relationships for Dairy Performance Between Large-Scale and Small-Scale Farm Conditions. J. Dairy Sci., 88, 4087 – 4096.

Kolmodin, R., Strandberg, E., Madsen, P., Jensen J., Jorjani, H. (2002) Genotype by Environment Interaction in Nordic Dairy Cattle Studied Using Reaction Norms. Acta Agric. Scand., Sect. A, Animal Sci., 52, 11 – 24.

Kolmodin, R., Strandberg, E., Danell, B., Jorjani, H. (2004) Reaction Norms for Protein Yield and Days Open in Swedish Red and White Dairy Cattle in Relation to Various Environmental Variables. Acta Agric. Scand., Sect. A, Animal Sci., 54, 139 – 151.

Knap, P. W. (2005) Breeding robust pigs. Aust. J. Exp. Agr., 45, 763 – 773.

Lillehammer, M., Hayes, B. J., Meuwissen, T. H. E., Goddard, M. E. (2009) Gene by environment interactions for productions traits in Australian dairy cattle. J. Dairy Sci., 92, 4008 – 4017.

Lynch, M., Walsh, B. (1998) Genetics and Analysis of Quantitative Traits. Sinauer Associates, Inc. Publishers. Sunderland, Massachusetts, 01375 USA.

Nostitz, B., Mielke, H. (1995) Vergleich verschiedener Methoden der Bestimmung des Milchenergiegehaltes beim Schwarzbunten Milchrind. J. Anim. Physiol. a. Anim. Nutr., 73, 9 – 18.

Schaeffer, L. R. (2004) Application of random regression models in animal breeding. Livest. Prod. Sci., 86, 35 – 45,

Schaeffer, L. R. (2010) Linear Models and Animal Breeding. Summer School Course. Germany.

Strandberg, E., Brotherstone, S., Wall, E., Coffey, M. P. (2009) Genotype by environment interaction for first-lactation female fertility in UK dairy cattle. J. Dairy Sci., 92, 3437 – 3446.

CHAPTER THREE

Using genome-wide association analysis to characterize environmental sensitivity of milk traits in dairy cattle

MELANIE STREIT*, ROBIN WELLMANN*, FRIEDRICH REINHARDT[§], GEORG THALLER[†], HANS-PETER PIEPHO[‡], and JÖRN BENNEWITZ*

* Institute of Animal Husbandry and Breeding,
University of Hohenheim,
70593 Stuttgart, Germany
[§] Vereinigte Informationssysteme Tierhaltung w.V. (VIT),
27283 Verden, Germany
[†] Institute of Animal Breeding and Husbandry,
Christian-Albrechts-University,
24098 Kiel, Germany
[‡] Institute of Crop Science,
University of Hohenheim,
70593 Stuttgart, Germany

Published in:
G3
doi:10.1534/g3.113.006536

Using genome-wide association analysis to characterize environmental sensitivity of milk traits in dairy cattle

M. Streit*, R. Wellmann*, F. Reinhardt[§], G. Thaller[†], H.-P. Piepho[‡], and J. Bennewitz*

* Institute of Animal Husbandry and Breeding,
University of Hohenheim,
70593 Stuttgart, Germany
[§] Vereinigte Informationssysteme Tierhaltung w.V. (VIT),
27283 Verden, Germany
[†] Institute of Animal Breeding and Husbandry,
Christian-Albrechts-University,
24098 Kiel, Germany
[‡] Institute of Crop Science,
University of Hohenheim,
70593 Stuttgart, Germany

key words: environmental sensitivity, genotype-by-environment interaction, association analysis, dairy cattle

Corresponding author:
Jörn Bennewitz
Garbenstraße 17
70599 Hohenheim
Germany
phone: +49 711-459 23570
fax: +49 711-459 23101
E-mail: j.bennewitz@uni-hohenheim.de

ABSTRACT

Genotype-by-environment interaction (GxE) has been widely reported in dairy cattle. One way to analyse GxE is to apply reaction norm models. The first derivative of a reaction norm is the environmental sensitivity (ES). In the present study we conducted a large scale genome-wide association analysis to identify SNPs that affect general production (GP) and ES of milk traits in the German Holstein population. Sire estimates for GP and for ES were calculated from around 13 million daughter records, using linear reaction norm models. The daughters were offspring from 2,297 sires. Sires were genotyped for 54k SNPs. The environment was defined as the average milk energy yield performance of the herds at the time where the daughter observations were recorded. The sire estimates were used as observations in a genome-wide association analysis, using 1,797 sires. Significant SNPs were confirmed in an independent validation set (500 sires of the same population). In order to separate GxE scaling and other GxE effects, the observations were log-transformed in some analyses. Results from the reaction norm model revealed GxE effects. Numerous significant SNPs were validated for both GP and ES. Many SNPs affecting GP also affect ES. We showed that ES of milk traits is a typical quantitative trait, genetically controlled by many genes with small effects and few genes with larger effect. A log-transformation of the observation resulted in a reduced number of validated SNPs for ES, pointing to genes that not only caused scaling GxE effects. The results will have implications for breeding for robustness in dairy cattle.

INTRODUCTION

Breeding cattle for milking traits relies on the use of daughter records for estimation of breeding values of their sires. Because sires are used widely through artificial insemination, their breeding values are estimable with a high accuracy, which resulted in a substantial genetic gain for milking traits over the last decades (Dekkers and Hospital 2002). It is expected, that this gain will be even further accelerated with the introduction of genomic selection methods (Meuwissen et al. 2001, Goddard and Hayes 2009). Often, frequently used sires have daughters that are milked in a wide range of environments. This raises the question about the importance of genotype-by-environment interaction (GxE). GxE refers to a variable response of genotypes to changes in the environment. Many studies have been conducted to quantify putative GxE effects in dairy cattle (e.g. Kolmodin *et al.* 2002, König *et al.* 2005, Strandberg *et al.* 2009 and references therein). The use of reaction norms is a powerful approach to study GxE effects if the environment can be described as a continuous variable. The slope of a reaction norm, i.e. the first derivative, is the environmental sensitivity (ES) and

genetic variation of ES can be interpreted as the existence of GxE (de Jong 1995, Lynch and Walsh 1998, James 2009). A frequently used environmental descriptor is the mean performance of all individuals in the environment (James 2009). It is assumed that various, unknown or unobservable environmental forces affect the mean performance. Mean performance is therefore a descriptor that captures these effects and weights them in a 'natural' way, i.e. by their effects on the performance. In dairy cattle, reaction norm models which include the average herd production level as a continuous environmental descriptor are widely used to study GxE (Kolmodin *et al.* 2002, Calus *et al.* 2002, Fikse *et al.* 2003, Hayes *et al.* 2003, Strandberg *et al.* 2009, Lillehammer *et al.* 2009a, Streit *et al.* 2012). Reaction norms are frequently fitted using random regression sire models. The daughter's observations are regressed on the corresponding herd solution. The regression is nested within sires, yielding a random sire estimate for the slope and for the intercept. The correlation between intercept and slope depends on where the intersection point of the reaction norm model is placed. It is recommended to place it in the average environment (van Tienderen and Koelewijn 1994, Kolmodin and Bijma 2004). In this case the intercept estimate can be interpreted as an estimate for average or general production (GP) and the slope as an estimate for ES for individual sires. A positive correlation between intercept and slope under this conditions was frequently reported (e.g. Kolmodin et al. 2002, Lillehammer et al. 2009b).

It might be worthwhile to consider ES in livestock breeding schemes (de Jong and Bijma 2002, Knap 2005, Veerkamp et al. 2009). Breeding for high yielding and sensitive individuals might be beneficial in high-producing and non-fluctuating environments, because sensitive individuals are able to benefit form these environmental conditions. In poor, fluctuating or unforeseeable environments, robust individuals are desired, if the robustness does not come at the expense of a decline in fitness and increase in health problems. One way to breed simultaneously for robustness and GP is to find genes that affect GP and ES of one trait in opposite directions, and to apply marker-assisted selection using these genes (Lillehammer *et al.* 2009b). Lillehammer *et al.* (2009b) applied association analysis using approximately 10,000 SNPs in the Australian dairy cattle population to find significant SNPs affecting GP and ES. Several SNPs were significant and around one third affected GP and ES in opposite directions; these SNPs are of special interest with regards to breeding for robustness.

The genetic architecture of dairy cattle milk traits has frequently been analysed (e.g. Cole et al. 2009, Hayes et al. 2010, Wellmann and Bennewitz 2011). However, it is unknown how

many genes affect ES, what the sizes and distributions of the effects are, and where they are located on the genome. In a recent study we applied higher order reaction norm random regression sire models to investigate GxE effects in German Holsteins (Streit *et al.* 2012). Herd test day solutions for production were used as environmental descriptors. We found highly significant GxE for milk traits, which resulted in substantial scaling and few re-ranking effects. For a deeper understanding of the nature of GxE effects, a partitioning of GxE effects into that due to scaling and due to changes in the rank of individuals across environments is desirable (e.g. Muir *et al.* 1992, Dutilleul and Potvin 1995, James 2009). An obvious method to reduce or eliminate scaling effects is to apply a data transformation (James 2009). This would allow partitioning of removable by data transformation and non-removable interaction.

The aim of the present study was to conduct a validated genome-wide association analysis to identify SNPs that affect GP and ES, and based on the results, to infer some knowledge of the genetic architecture of GP and ES. We were especially interested in the number of validated SNPs and the size and the sign of the effects on GP and ES. We applied a three-step procedure. In the first step, sire estimates for GP and for ES were calculated using first-order random regression sire models. These estimates were used in a second step as observations in an association analysis. In the third step, significant SNP associations were confirmed in an independent validation set of the same population. In order to remove GxE causing scaling effects, the observations were log-transformed in some analyses.

MATERIALS AND METHODS

Data and data editing

In total 2,356 progeny tested German Holstein sires were genotyped with the Illumina BovineSNP50 BeadChip, which contains a total of 54001 SNPs (Illumina, San Diego, CA; Matukumalli *et al.* 2009). The sires were born between 1983 and 2003 and reflect a representative sample of the population (Qanbari *et al.* 2010). Individuals with more than 10% missing marker genotypes were removed, resulting in 2,297 sires. An SNP was excluded if it had a minor allele frequency less than 3%, a call rate less than 90%, a significant deviation from the Hardy-Weinberg-equilibrium ($p<0.001$), or if the position on the genome was unknown. SNPs on the sex chromosome were also excluded. This data filtering was done using PLINK (Purcell *et al.* 2007). A total of 41,349 SNPs remained in the data set. Sporadic missing genotypes were imputed using fastPHASE (Scheet and Stephens 2006). The linkage disequilibrium (LD) structure in this population was investigated by Qanbari *et al.* (2010).

Around 13 million first lactation test day records for protein yield, fat yield and milk yield from daughters of the sires were used. The number of daughters per bull ranged from 50 to 74,842 and totalled around 1.3 million. Test day records were corrected for the fixed effects herd test day, days in milk, age at calving, calving season and the random permanent environment effect. These correction factors were obtained from the routine animal genetic evaluation, which is an animal test day model. After this adjustment, the trait population mean was added to the observations in order to obtain predicted trait values.

The environment was described by the mean herd test day performance for milk energy yield. It was calculated as a linear combination of milk yield, fat yield and protein yield, i.e. $energy\ yield = 0.802 * milk\ yield + 38.4 * fat\ yield + 23.6 * protein\ yield$, where the yields are measured in kg (Nostitz and Mielke 1995). We preferred this single parameter to describe the environment, because it combined the highly correlated herd test day performances for the three milk yield traits; see Streit *et al.* (2012) for further details. It is assumed that this parameter captures important unobservable and unknown environmental effects. The environmental descriptor was rescaled to have a mean equal to 0 and a standard deviation of 1. Hence, superior (inferior) environments show positive (negative) values for the environmental descriptor, and the 'average' environment shows a value close to zero. The distribution of the environmental descriptor is shown in the Supplemental. It is approximately normally distributed. Mean herd test day performances of milk yield, protein yield, and fat yield were obtained from the routine animal genetic evaluation, see VIT (2013) for a detailed description.

Statistical analysis

In a previous study we applied a second-order sire model, which gave an improved fit compared to a first-order model. However, a first-order sire effect explained most of the variation of ES (Streit *et al.* 2012). Therefore, we decided to apply a fist-order sire model in the present study. The following random regression model was applied in the first step for all three milk yield traits:

$$cy_{ijk} = \mu + b * htdsme_k + \sum_{m=0}^{1} s_{jm} * htdsme_k^m + \sum_{m=0}^{1} d_{ijm} * htdsme_k^m + e_{ijk}, \quad (1)$$

where cy_{ijk} is the corrected yield of daughter i of sire j at herd test day k, μ is the overall mean, $htdsme_k$ is the herd test day solution for milk energy yield at herd test day k with the fixed regression coefficient b, s_{jm} is the random sire effect of sire j of order m, d_{ijm} the random

daughter effect of daughter *i* of sire *j* of order *m*, and *e* is the random residual. The covariance structure of the sire regression effects is $Var\begin{bmatrix} s_0 \\ s_1 \end{bmatrix} = I \otimes \begin{bmatrix} \sigma^2_{s_0} & \sigma_{s_0 s_1} \\ \sigma_{s_0 s_1} & \sigma^2_{s_1} \end{bmatrix}$, and that of the daughter effects is $Var\begin{bmatrix} d_0 \\ d_1 \end{bmatrix} = I \otimes \begin{bmatrix} \sigma^2_{d_0} & \sigma_{d_0 d_1} \\ \sigma_{d_0 d_1} & \sigma^2_{d_1} \end{bmatrix}$, with *I* being the identity matrix. The estimated sire effects were used as observations in an association analysis (see below). In contrast to classical sire models, the relationship among sires was ignored. This could be done, because there was much progeny information available for each sire, and hence, the sire estimates were largely influenced by the progeny records and only very little by the pedigree. Indeed, preliminary results showed that the correlation between sire estimates with and without considering the pedigree in model (1) was >0.98 (not shown).

In order to model heterogeneous residual variance across the environments, the observations were ordered according to the environmental descriptor and grouped into ten classes of equal size based on the environmental values. Residual variances were estimated for each class, assuming the residual covariance to be zero. The uncorrelated daughter effects reduce the heterogeneity of residual variance if GxE effects are present (Lillehammer *et al.* 2009a). The models were fitted using ASReml 3.0 (Gilmour *et al.* 2009). Because the mean of the environmental descriptor was zero, the intercept solutions of the sire regression coefficients were used as sire estimates for GP, i.e. production level in the average environment. Furthermore, the slope solutions were used as estimates for ES.

Table 1 Sire variance components of the random regression analyses. Standard errors are shown in parentheses. $\sigma^2_{S_0}$ ($\sigma^2_{S_1}$) denotes the intercept (slope) sire variance, with correlation $\rho_{S_0 S_1}$.

Trait	Unit	$\sigma^2_{S_0}$	$\sigma^2_{S_1}$	$\rho_{S_0 S_1}$
Protein yield	g	2379.37 (87.48)	17.02 (0.98)	0.79
Fat yield	g	7883.41 (257.12)	46.76 (2.43)	0.93
Milk yield	kg	1.30 (0.04)	0.02 (< 0.01)	0.72
ln(protein yield)*	-	9.50 (< 0.01)	0.11 (< 0.01)	0.61
ln(fat yield)*	-	12.70 (< 0.01)	0.13 (< 0.01)	0.73
ln(milk yield)*	-	10.55 (< 0.01)	0.14 (< 0.01)	0.68

***** values are multiplied by 10,000

The whole data set was randomly split into a discovery data set (n = 1,797 bulls) and a validation data set (n = 500 bulls). In the second step of the statistical analysis, we performed genome-wide association analyses using the discovery data set. To do so, we applied the following mixed linear model:

$$\hat{s}_{jm} = \mu_m + sire_{jm} + b_{km} * x_{jk} + e_{jm}, \quad (2)$$

where $\hat{s}_{jm}$ is the estimated random sire effect for GP (m = 0) and ES (m = 1). These estimates were taken from the results of model (1). The model was applied for the two traits (m = 0 for GP and m = 1 for ES) separately. The effect of the SNP k was modelled as a regression on the number of copies of the allele with the higher frequency (x = 0, 1, or 2), with b_{km} being the regression coefficient. In order to control for the population structure, we fitted a random sire effect with the covariance structure $A\sigma^2_{sm}$, where A is the numerator relationship matrix calculated from high-quality pedigree information and σ^2_{sm} a variance attributable to the sires. This model was applied for each SNP k in turn, resulting in 41,349 association tests per trait. We declared each SNP with a pointwise error probability below p<0.001 as significant. In order to judge how many false positives were among the significant associations we applied the false discovery rate (FDR) technique. We calculated for each association test an FDR q-value using the software QVALUE (Storey and Tibshirani 2003). The FDR q-value of the significant SNP with the lowest test statistic (p≈0.001) provided an estimate of the proportion of false positives among the significant associations.

Table 2 Number of discovered and validated SNPs for intercept and slope for the traits on the observed scale. The FDR q-values (FDR) of the significant SNP with the largest error probability (p≈0.001) in the discovery dataset are shown.

Trait	Discovery dataset ($p \leq 0.001$)	FDR	Validation dataset ($p \leq 0.01$)
Intercept protein yield	450	0.07	69
Slope protein yield	351	0.09	44
Intercept fat yield	465	0.07	118
Slope fat yield	385	0.08	99
Intercept milk yield	415	0.08	104
Slope milk yield	416	0.08	98

In the third step, we confirmed significant SNP associations within the same population in the validation set. The same statistical model was applied, but only to significant SNPs. We declared an SNP as confirmed if the p-value in the validation set was p<0.01 and the signs of the effects were the same in both sets. This relaxed significance criterion was used in the

validation set, because less multiple testing was performed, and a more stringent significance level would reduce the power to validate SNPs. A similar protocol was used by Pryce et al. (2010). For the interpretation of the effects, the estimates of the validation set were used, because it can be assumed that these suffer less from the Beavis-effect and are less upwardly biased (Beavis 1994).

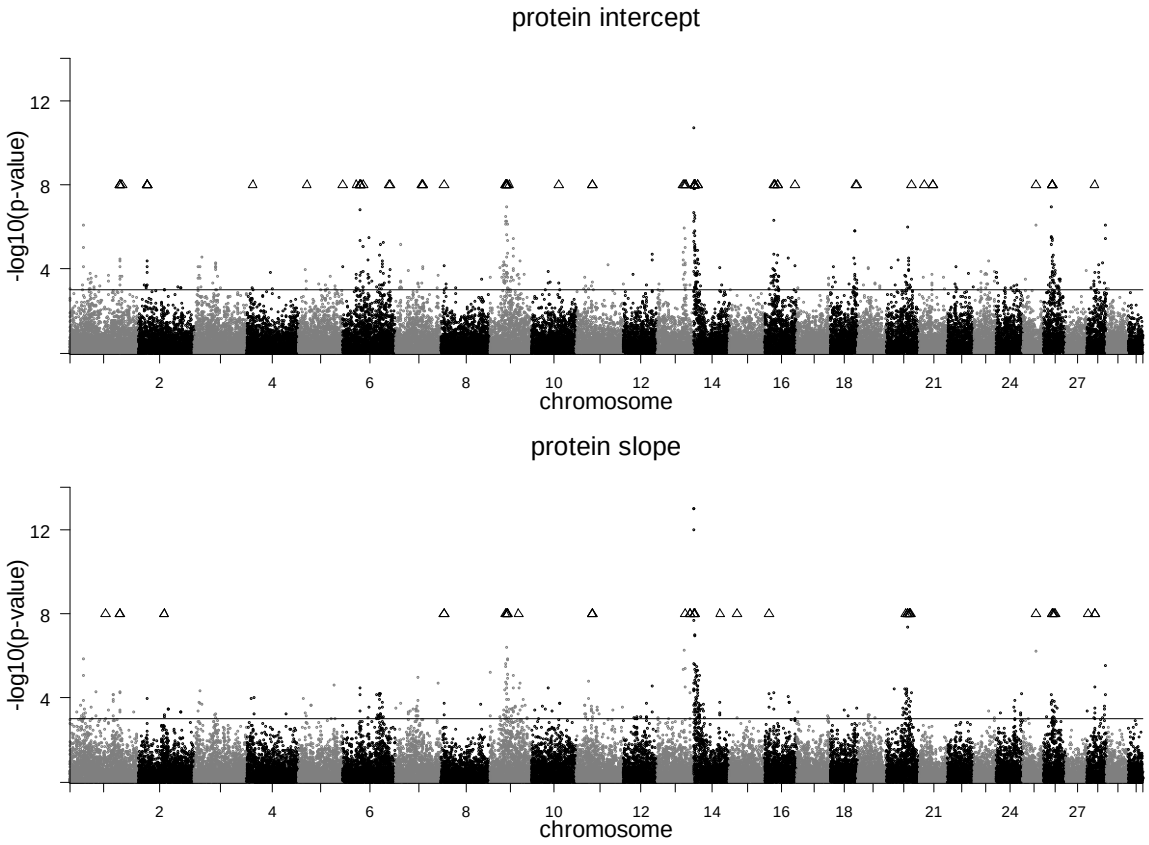

Figure 1 Test statistic profile of SNP effects for protein yield intercept (top) and protein yield slope (bottom) in the discovery data set. The nominal significance level (p<0.001) is indicated by a solid line. Positions of validated SNPs are indicated by a triangle.

In order to identify SNPs that not only cause scaling effects within the environmental range considered in our study, we applied the models also to log-transformed observations (Hayes *et al.* 2003, Lillehammer *et al.* 2009b). Preliminary results revealed convergence problems of model (1) with log-transformed observations (not shown), which was caused by the random regression of the daughter on the environment. Therefore, to ensure convergence, the random daughter effect was fitted without regression on the environment. The residual variance was homogeneous, so only one residual variance component was estimated. The sire solutions obtained from model (1) were used subsequently in model (2).

RESULTS

The main results of the variance component estimation are shown in Table 1. There is slope variance for all traits on both the observed and the log-transformed scales, pointing to the presence of GxE effects. These GxE effects were analysed in details and also tested for significance in an earlier study (Streit et al. 2012). On the observed scale, the correlation between intercept and slope was high and positive. The log-transformation reduced this correlation. As expected, the daughter variance was substantial and the residual variance was heterogeneous across the environmental classes for traits on the observed scale (not shown).

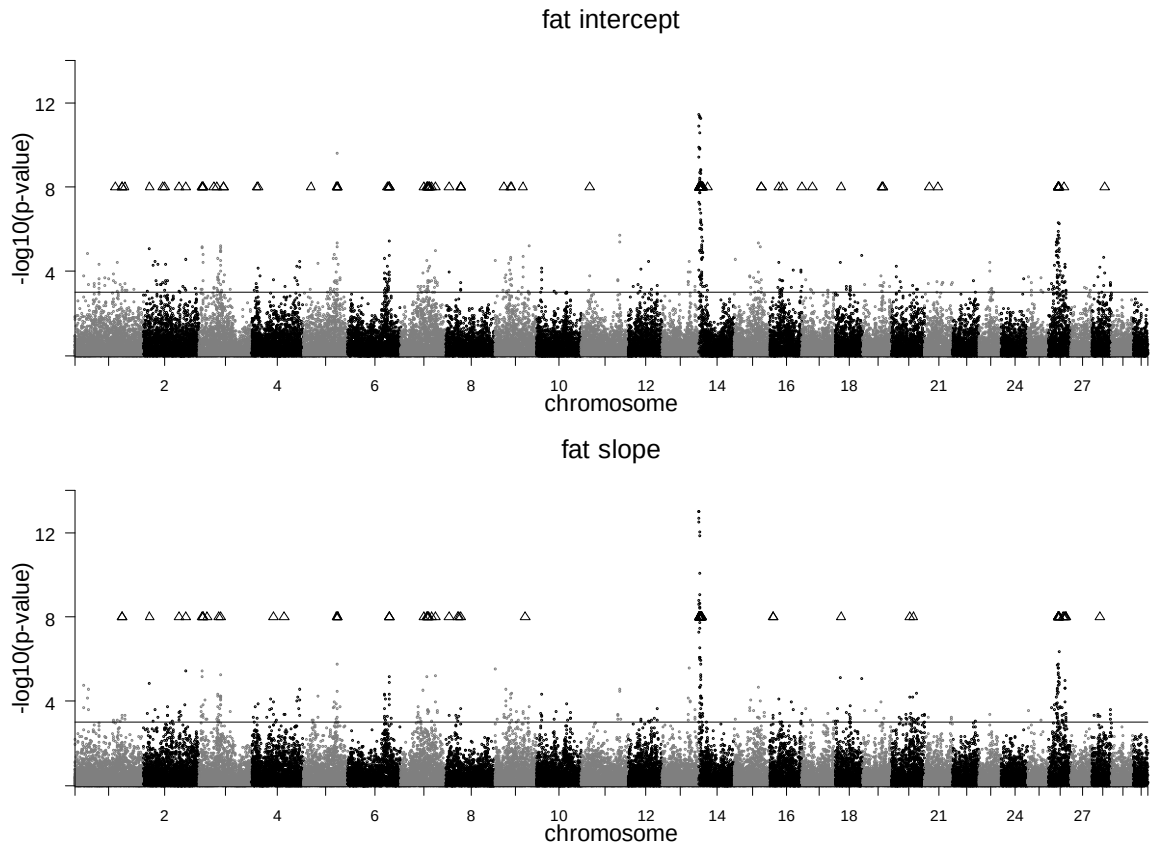

Figure 2 Test statistic profile of SNP effects for fat yield intercept (top) and fat yield slope (bottom) in the discovery data set. The nominal significance level (p<0.001) is indicated by a solid line. Positions of validated SNPs are indicated by a triangle.

The results of the association analysis for the traits on the observed scale are shown in Table 2. For all traits, 350 to 450 SNPs showed a nominal significant association in the discovery data set; the FDR-analysis revealed that around 7-9 % of these are false positives. For fat and protein yield, more trait-associated SNPs could be found for intercept than for slope. The number of validated SNPs was between 44 (protein slope) and 118 (fat intercept). The results

for the log-transformed data sets are shown in Table 3. For the intercepts, almost the same number of significant SNPs was found as on the observed scale, but fewer could be confirmed. For the slopes, the number of significant SNPs was reduced. The FDR *q*-values of the significant associations were similar or slightly higher.

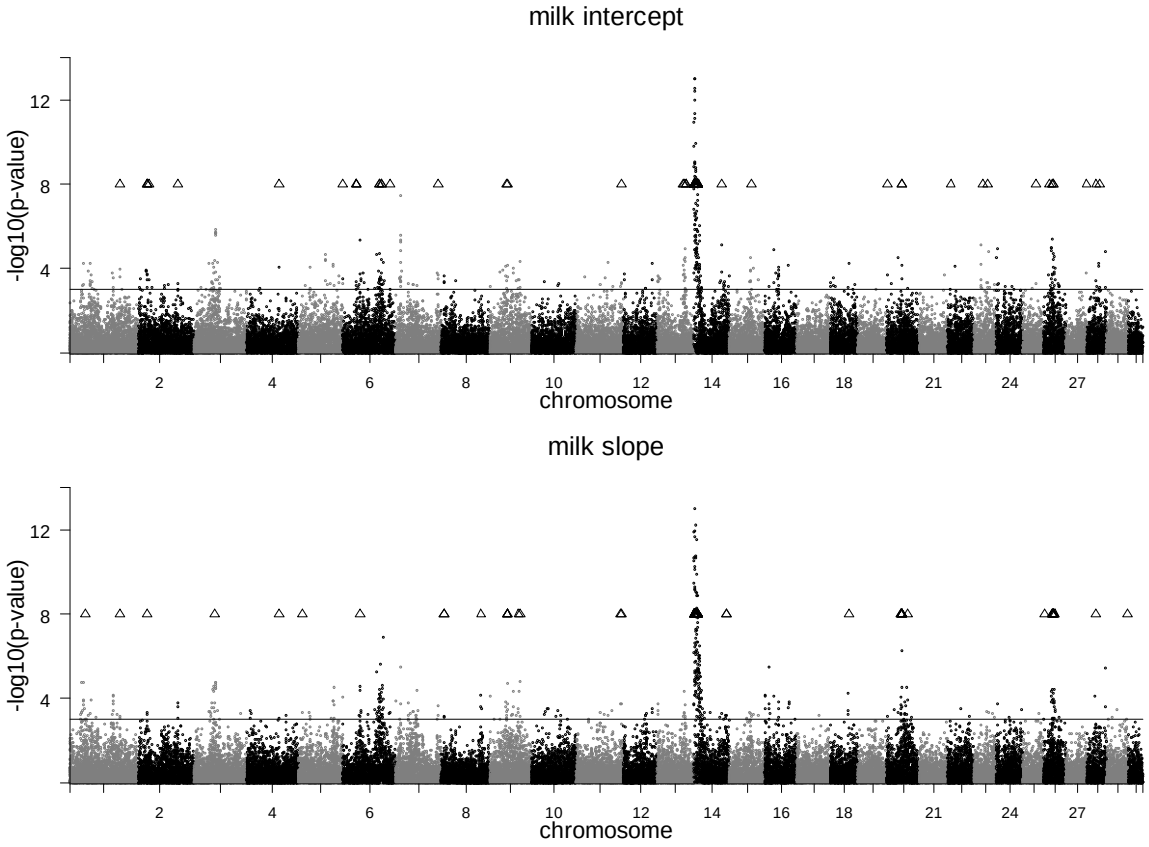

Figure 3 Test statistic profile of SNP effects for milk yield intercept (top) and milk yield slope (bottom) in the discovery data set. The nominal significance level (p<0.001) is indicated by a solid line. Positions of validated SNPs are indicated by a triangle.

The plots of the test statistic along the chromosomes are shown in Figure 1 to 3 for the traits on the observed scale. Chromosomal positions of validated SNPs are indicated by a triangle symbol. The pattern of the test statistic was similar for the intercept and slope within the traits, although for intercept the signals were generally more pronounced, leading to the higher number of significant associations. Significant SNPs were found on many chromosomes, and the clearest signals were observed on BTA14. Promising SNP clusters affecting intercept and slope of all traits were also identified on BTA26. Chromosome 9 is interesting with regard to protein, as it contains a validated SNP cluster for both intercept and slope. For slope, validated SNPs with a remarkably high test statistic were found on BTA20 and BTA25. For fat intercept, a highly significant SNP was found on BTA5, which also

affected slope to a lesser extent. For milk slope validated SNPs were mapped on BTA6 and BTA20. The test statistic plots for intercept on the observed and on the log-scale are almost identical for all three traits (not shown). For slope, however, the plots differ between the scales (see Figure 4). Again, SNPs on BTA14 showed the strongest signals for all three log-transformed traits for slope.

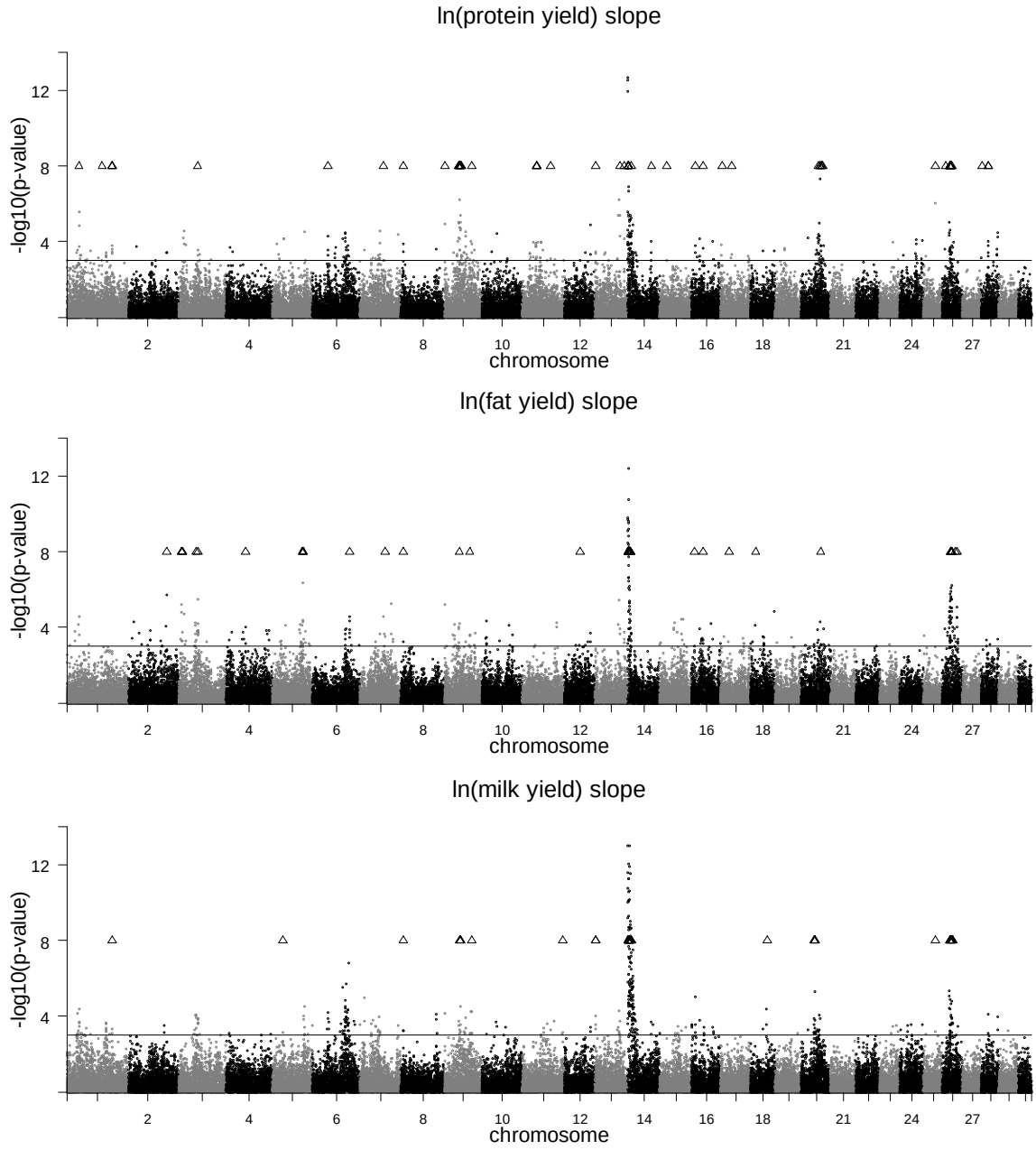

Figure 4 Test statistic profile of SNP effects for ln(protein yield) slope (top), ln(fat yield) slope (middle), and ln(milk yield) slope (bottom) in the discovery data set. The nominal significance level (p<0.001) is indicated by a solid line. Positions of validated SNPs are indicated by a triangle.

In Figure 5 the estimates of the validation set are shown for SNPs that were either significant for intercept, or for slope, or for both. The slope effect of the allele that increases the intercept is shown. It can be seen that every validated SNP affects both traits in the same direction, and the correlation between the solutions is highly positive. This was less pronounced if the data were log-transformed (Figure 6). For ln(protein yield), many validated SNPs for intercept showed a small but mostly non-significant negative effect for slope. In general, the largest SNP effects (in units of the standard deviation, σ) were observed for milk yield, with 11 (4) SNPs showing an intercept (slope) effect larger than 0.3σ. For the log-transformed data sets, the intercept effects are generally larger. This was not observed for slope effects. The estimates of each validated SNP for the traits on the observed scale are presented in Supplemental Table 1; estimates for the log-transformed observations are presented in Supplemental Table 2.

Table 3 Number of discovered and validated SNPs for intercept and slope for the traits on the log-scale. The FDR *q*-values (FDR) of the significant SNP with the largest error probability ($p \approx 0.001$) in the discovery dataset are shown.

Trait	Discovery dataset ($p \leq 0.001$)	FDR	Validation dataset ($p \leq 0.01$)
Intercept ln(protein yield)	463	0.07	56
Slope ln(protein yield)	313	0.11	64
Intercept ln(fat yield)	469	0.07	118
Slope ln(fat yield)	320	0.11	80
Intercept ln(milk yield)	419	0.08	87
Slope ln(milk yield)	386	0.09	68

DISCUSSION

In this study we attempted to identify and confirm SNPs for intercept (reflecting GP) and slope (reflecting ES) of milk traits in the German Holstein dairy cattle population. Numerous SNPs were identified and confirmed for both GP and ES. Many SNPs affecting GP also affect ES. We showed that ES of milk traits has a similar genetic architecture as GP and is a typical quantitative trait, genetically controlled by many genes with small effects and few genes with larger effect (Figure 5, Supplemental Table 1). Given the FDR *q*-values of the SNPs in the discovery set (Tables 2 and 3) it seems that some SNPs with true associations were not confirmed. This might be due to the reduced power of the validation set with 500 sires. A more stringent validation would be to also test if the SNP is significant in another population (Hayes *et al.* 2009, Pryce *et al.* 2010). Such a validation study would also increase mapping precision, because mapping resolution is increased when using an across-breed approach and

only those SNPs being in LD with the mutation in both breeds would be validated. No independent population was available, however, to do an across-population validation in this study.

In our study, the mapping precision is limited due to the LD structure observed in this population (Qanbari *et al.* 2010) in combination with the applied single marker association analysis. Alternatively, a combined linkage and LD mapping approach could have been applied, which predicts IBD-probabilities at putative QTL regions using multi-marker and pedigree information and uses these probabilities for QTL fine-mapping (Meuwissen et al. 2002). This method is, however, computationally demanding and needs higher marker densities. Another multi-marker approach that could have been applied is a Bayes-method originally developed for genomic selection (Meuwissen et al. 2001, Goddard and Hayes 2009). These Bayes-methods make use of the LD of the markers and the mutation and additionally of the LD between the markers. It is not completely clear how to test for significance when using these methods. Olsen et al. (2011) applied the three approaches mentioned above to map genes for fertility and milk production in dairy cattle. They applied single marker association analysis for a first screen, fine-mapped the regions using combined linkage and LD mapping and confirmed the putative positions by using BayesA from Meuwissen et al. (2001).

Some interesting SNP clusters affecting GP are located closely to well known candidate genes which segregate in the German Holstein population. This is most obvious on BTA14, were the clear signals for all milk traits for GP and ES probably reflect the effect of *DGAT1* (Grisart *et al.* 2002, Winter *et al.* 2002). This gene is known to segregate and affect all milk traits in this population (Bennewitz *et al.* 2004a). Several SNPs affecting GP of all three investigated milk traits were found on BTA6. From previous linkage analyses it is known that BTA6 harbours QTL affecting milk traits in this population (Kühn *et al.* 1999, Bennewitz *et al.* 2004b). Putative candidate genes underlying mapped QTL are discussed in Weikard *et al.* (2005). The *PPARGC1A* gene was postulated as the most plausible gene underlying a QTL for fat yield. Additionally, the casein gene complex is located on this chromosome, with an effect on protein yield and protein percentage traits in this population (Prinzenberg *et al.* 2003). On BTA5 we found a single SNP with a remarkably high test statistic for fat GP, which was also validated for fat ES. Wang *et al.* (2012) reported the gene *EPS8* to be most

likely causative for this association. The significant SNPs on BTA20 is very likely to be associated with the *GHR* gene (Blott *et al.* 2003, Wang *et al.* 2012).

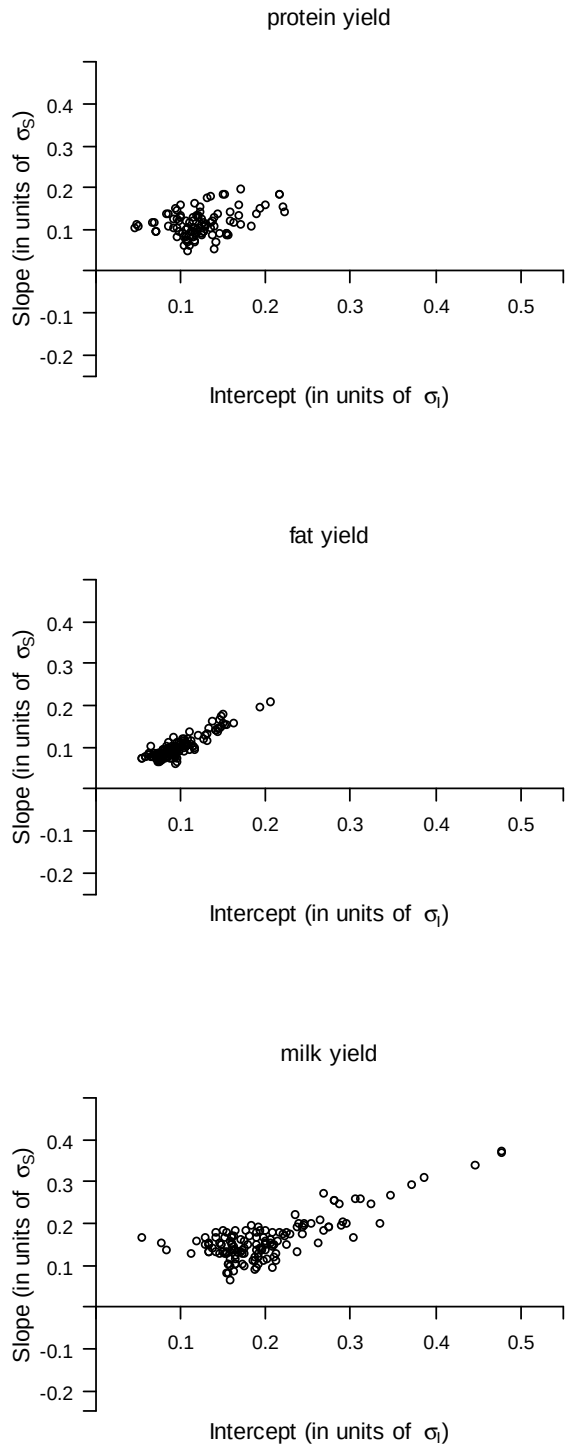

Figure 5 Estimated SNP effects for the traits on the observed scale. The term σ_S (σ_i) denotes the sire intercept (slope) standard deviation. Each SNP was validated within the population either for intercept, slope or both. Estimates were taken from the validation set.

Some validated SNPs for ES are in chromosomal regions similar to those found in other dairy cattle populations harbouring genes with GxE effects. In the Norwegian Red, milk production QTL for ES on BTA2, BTA6, BTA7, and BTA16 were reported by Lillehammer *et al.* (2007, 2008). A detailed analysis of BTA6 with a high marker density revealed two QTL for milk yield with an effect on ES, but no QTL with an ES effect for fat and protein yield. In our population, we were able to validate ES SNPs on BTA6 for milk and fat yield, but not for protein yield. In the Australian Holstein population Lillehammer *et al.* (2009b) found several SNPs with ES effects. Roughly one third of their significant associations affected GP and ES in opposite directions, which is in contrast to our findings. They stated, however, that this proportion is probably smaller than one third, because it is generally more difficult to find SNPs that affect GP and ES in the same direction rather than in opposite directions. Our study is considerably more powerful than that of Lillehammer *et al.* (2009b), hence we were likely able to detect more SNPs with effects in the same direction.

We previously reported significant GxE resulting in substantial scaling effects (Streit *et al.* 2012). In order to remove these scaling effects, a log-transformation was applied. The results from the association analysis applied to the log-transformed data revealed SNP that were not removable by this kind of transformation. These validated SNPs are of special interest, because they point to chromosomal regions harbouring genes with an effect on ES which are not or not solely due to scaling effects. Some regions with clear signals for ES on the observed scale could not be found on the log-scale. This was especially observed for ln(protein yield) and SNPs on BTA14 close to the *DGAT1* gene, where positive effects on ES were turned into small negative effects, although mostly not significant (Figure 6, Supplemental Table 2). Hence, these effects were completely removable by the log-transformation. It may be noted that the log-transformation is frequently applied, but maybe another transformation function (e.g. from the Box-Cox-family of transformation) would be able to eliminate scaling effects more effectively. This was not investigated further in this study. The reduced correlation between intercept and slope when applying the log-transformation (Table 1) was also observed by Lillehammer *et al.* (2009b). This decreased correlation has the following reason. For large yields the intercept of a regression is large as well. Since the logarithm is a concave function, the interval containing these yields is mapped to a smaller interval than an interval of the same size containing small yields. Thus, the transformation causes large yields to decrease in variance more drastically than small yields.

This causes positive slopes of the regression lines for large yields to decrease more than positive slopes of regression lines for small yields.

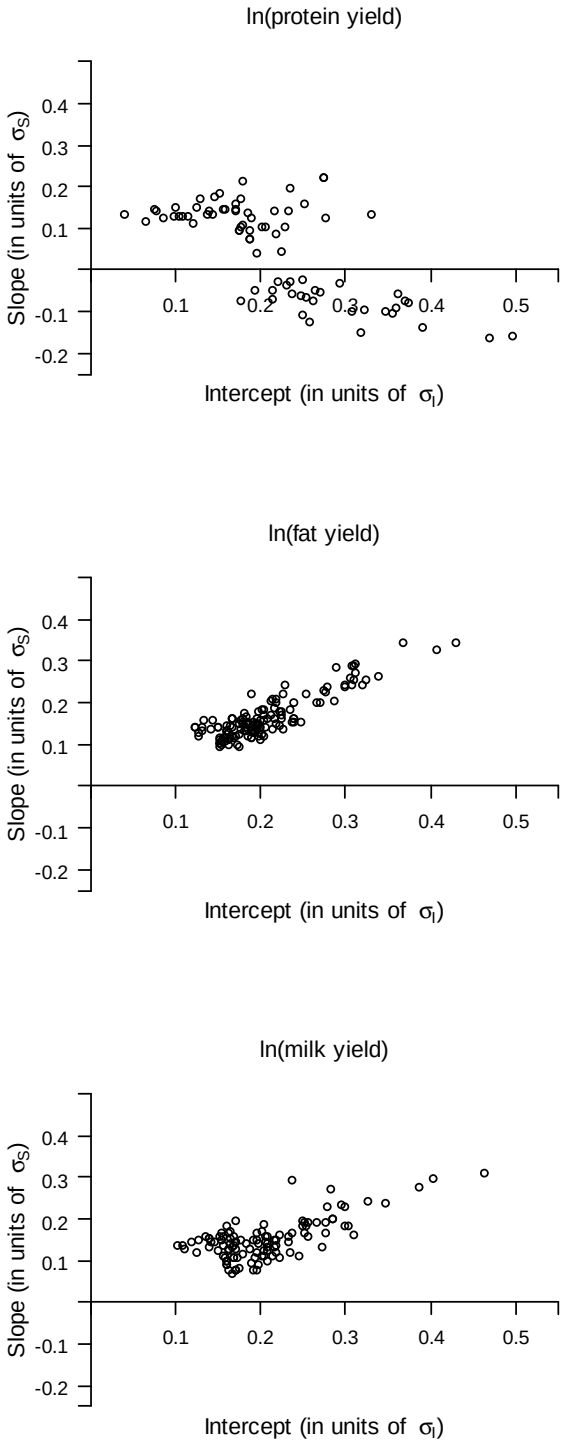

Figure 6 Estimated SNP effects for the traits on the log-scale. The term σ_S (σ_i) denotes the sire intercept (slope) standard deviation. Each SNP was validated within the population either for intercept, slope or both. Estimates were taken from the validation set.

As described in the introduction, breeding for robustness for both milk production and health traits is an issue in dairy cattle. In this study only milk production traits were considered. Based on our results, it seems that simultaneously breeding for an increase in milk GP and a decrease in ES by applying marker-assisted selection is difficult, because no SNPs showed opposite directions of the effects. Genomic selection can be seen as marker assisted selection on a genome-wide scale. It is currently implemented in many dairy cattle populations (Goddard and Hayes 2009). Improving ES by genomic selection should be possible by considering ES as an additional trait and by estimating genomic assisted breeding values for this trait. A reference population for the estimation of marker effects is needed. Existing reference populations mainly built by progeny-tested bulls can also be used for ES, provided that the daughters are distributed over a wide range of environments. As done in this study, the daughter records can then be used for the estimation of sire effects for ES, which in turn, can be used to estimate marker effects. The most appropriate method for this estimation depends on the genetic architecture of the trait, i.e. on the number of genes affecting the trait and on the distribution of the effect size (Hayes *et al.* 2010). The current study shows that for the estimation of marker effects for ES a model should be used that is tailored to traits affected by many genes with small effects and few with large effects.

CONCLUSIONS

We presented GxE for milk traits resulting in substantial scaling effects. Many SNP clusters affecting GP and ES could be identified and validated. The effects of some SNPs for ES were not removable by a data transformation, indicating that these are not solely scaling effects. The positions of these clusters were often found in well-known candidate regions affecting milk traits. No validated SNP showed effects for ES and GP in opposite directions. We showed that ES of milk traits is a typical quantitative trait controlled by many genes with small and few genes with large effects.

ACKNOWLEDGMENTS

MS was supported by a grant from the Deutsche Forschungsgemeinschaft, DFG. The authors thank Chris Baes (Swiss College of Agriculture, Zollikofen, Switzerland) for language corrections. The manuscript has benefited from the critical and helpful comments of two anonymous reviewers.

REFERENCES

Beavis, W. D., 1994 The power and deceit of QTL experiments: lessons from comparative QTL studies, pp. 250–266 in 49th Annual Corn and Sorghum Industry Research Conference, edited by ASTA, Washington, DC.

Bennewitz, J., N. Reinsch, S. Paul, C. Looft, B. Kaupe et al., 2004a The DGAT1 K232A mutation is not solely responsible for the milk production quantitative trait locus on the bovine chromosome 14. J. Dairy Sci. 87: 431–442.

Bennewitz, J., N. Reinsch, V. Guiard, S. Fritz, H. Thomsen et al., 2004b Multiple quantitative trait loci mapping with cofactors and application of alternative variants of the false discovery rate in an enlarged granddaughter Design. Genetics 168: 1019–1027.

Blott, S., J.-J. Kim, S. Moisio, A. Schmidt-Küntzel, A. Cornet et al., 2003 Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. Genetics 163: 253–266.

Calus, M. P. L., A. F. Groen, G. de Jong, 2002 Genotype x environment interaction for protein yield in dutch dairy cattle as quantified by different models. J. Dairy Sci. 85: 3115–3123.

Cole, J. B., P. M. vanRaden, J. R. O'Conell, C. P. van Tassell, T. S. Sonstegard et al. 2009 Distribution and location of genetic effects for dairy traits. J. Dairy Sci. 92: 2931–2946.

de Jong, G. 1995 Phenotypic plasticity as a product of selection in a variable environment. Am. Nat., 145: 493–512.

de Jong, G., and P. Bijma, 2002 Selection and phenotypic plasticity in evolutionary biology and animal breeding. Livest. Prod. Sci. 78: 195–214.

Dekkers, J. C. M., and F. Hospital, 2002 The use of molecular genetics in the improvement of agricultural populations. Nat. Rev. Genet. 3:22–32.

Dutilleul, P., and C. Potvin, 1995 Among-environment heteroscedasticity and genetic autocorrelation: implications for the study of phenotypic plasticity. Genetics 139: 1815–1829.

Fikse, W. F., R. Rekaya, K. A. Weigel, 2003 Assessment of environmental descriptors for studying genotype by environment interaction. Livest. Prod. Sci. 82: 223–231.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, R. Thompson, 2009 ASReml User Guide 3.0. VSN International Ltd, Hemel Hempstead, UK.

Goddard, M. E., and B. J. Hayes, 2009 Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10: 381–391.

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford et al., 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: 222–231.

Hayes, B. J., M. Carrick, P. Bowman, M. E. Goddard, 2003 Genotype x environment interaction for milk production of daughters of australian dairy sires from test-day records. J. Dairy Sci. 86: 3736–3744.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, K. Savin, C. P. van Tassell et al., 2009 A validated genome wide association study to breed cattle adapted to an environment altered by climate change. PLoS ONE 4(8): e6676. doi:10.1371/journal.pone.0006676.

Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman, M. E. Goddard, 2010 Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in holstein cattle as contrasting model traits. PLoS Genet 6(9): e1001139. doi:10.1371/journal.pgen1001139.

James, J. W., 2009 Genotype by environment interaction in farm animals, pp. 151–167 in Adaptation and Fitness in Animal Populations. Evolutionary and Breeding Perspectives on Genetic Resource Managemant edited by J. van der Werf, H.-U. Graser, R. Frankham, C. Gondro, Springer Science+Business Media B. V.

Knap P. W., 2005 Breeding robust pigs. Aust. J. Exp. Agr. 45:763–773.

König, S., G. Dietl, I. Raeder, H. H. Swalve, 2005 Genetic relationships for dairy performance between large-scale and small-scale farm conditions. J. Dairy. Sci. 88: 4087–4096.

Kolmodin, R., E. Strandberg, P. Madsen, J. Jensen, H. Jorjani, 2002 Genotype by environment interaction in nordic dairy cattle studied using reaction norms. Acta Agric. Scand. Sect. A Anim. Sci. 52: 11–24.

Kolmodin, R., and P. Bijma, 2004 Response to mass selection when the genotype by environment interaction is modelled as a linear reaction norm. Genet. Sel. Evol. 36: 435–454.

Kühn, C., G. Freyer, R. Weikard, T. Goldammer, M. Schwerin, 1999 Detection of QTL for milk production traits in cattle by application of a specifically developed marker map of BTA6. Anim. Genet. 30: 333–340.

Lillehammer, M., M. Árnyasi, S. Lien, H. G. Olsen, E. Sehested et al., 2007 A genome scan for quantitative trait locus by environment interactions for production traits. J. Dairy Sci. 90: 3482–3489.

Lillehammer, M., M. E. Goddard, H. Nilsen, E. Sehested, H. G. Olsen et al., 2008 Quantitative trait locus-by-environment interaction for milk yield traits on bos taurus autosome 6. Genetics 179: 1539–1546.

Lillehammer, M., J. Ødegård, T. H. E. Meuwissen, 2009a Reducing the bias of estimates of genotype by environment interactions in random regression sire models. Gen. Sel. Evol. 41:30. doi:10.1186/1297-9686-41-30.

Lillehammer, M., B. J. Hayes, T. H. E. Meuwissen, M. E. Goddard, 2009b Gene by environment interactions for production traits in australian dairy cattle. J. Dairy Sci. 92: 4008–4017.

Lynch, M., and B. Walsh, 1998 Genetics and Analysis of Quantitative Traits. Sinauer Associates Inc Publishers, Sunderland, MA, USA.

Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan et al., 2009 Development and characterization of a high density SNP genotyping assay for cattle. PLoS ONE 4(4): e5350. doi:10.1371/journal.pone.0005350.

Meuwissen, T. H. E., B. J. Hayes, M. E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157: 1819–1829.

Meuwissen, T. H. E., A. Karlsen, S. Lien, I. Olsaker, M. E. Goddard, 2002 Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. Genetics. 161: 373–379.

Muir, W., N. E. Nyquist, S. Xu, 1992 Alternative partitioning of the genotype-by-environment interaction. Theor. Appl. Genet. 84: 193–200.

Nostitz, B., and H. Mielke, 1995 Vergleich verschiedener Methoden der Bestimmung des Milchenergiegehaltes beim Schwarzbunten Milchrind. J. Anim. Physiol. a. Anim. Nutr. 73: 9–18.

Olsen, H., B. Hayes, M. Kent, T. Nome, M. Svendsen et al. 2011 Genome wide association mapping in Norwegian Red cattle identifies quantitative trait loci for fertility and milk production on BTA12. Anim. Genet. 42: 466 –474.

Prinzenberg, E.–M., C. Weimann, H. Brandt, J. Bennewitz, E. Kalm et al., 2003 Polymorphism of the bovine CSN1S1 promoter: linkage mapping, intragenic haplotypes, and effects on milk production traits. J. Dairy Sci. 86: 2696–2705.

Pryce, J. E., S. Bolormaa, A. J. Chamberlain, P. J. Bowman, K. Savin et al., 2010 A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. J. Dairy Sci. 93: 3331–3345.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira et al., 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81: 559–575.

Qanbari, S., E. C. G. Pimentel, J. Tetens, G. Thaller, P. Lichtner et al., 2010 The pattern of linkage disequilibrium in German Holstein cattle. Anim. Genet. 41: 346–356.

Scheet, P., and M. Stephens, 2006 A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet. 78: 629–644.

Storey, J. D., and R. Tibshirani, 2003 Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. 100: 9440–9445.

Strandberg, E., S. Brotherstone, E. Wall, M. P. Coffey, 2009 Genotype by environment interaction for first-lactation female fertility traits in UK dairy cattle. J. Dairy Sci. 92: 3437–3446.

Streit, M., F. Reinhardt, G. Thaller, J. Bennewitz, 2012 Reaction norms and genotype-by-environment interaction in the German Holstein dairy cattle. J. Anim. Breed. Genet. 129: 380–389.

van Tienderen P. H., and H. P. Koelewijn, 1994 Selection on reaction norms, genetic correlations and constraints. Genet. Res. Camb. 64: 115–125.

Veerkamp, R. F., H. A. Mulder, M. P. L. Calus, J. J. Windig, and J. ten Napel, 2009 Statistical genetics to improve robustness of dairy cows. Proc. Assoc. Advmt. Anim. Breed. Genet. 18: 406–413.

VIT United Data System Verden w.V. Description of the national animal genetic evaluation system for German Holstein dairy cattle. www.vit.de.

Wang, X., C. Wurmser, H. Pausch, S. Jung, F. Reinhardt et al., 2012 Identification and dissection of four major QTL affecting milk fat content in the german Holstein-Friesian population. PLoS ONE 7(7): e40711. doi:10.1371/journal.pone.0040711.

Weikard, R., C. Kühn, T. Goldammer, G. Freyer, M. Schwerin, 2005 The bovine PPARGC1A gene: molecular characterization and association of an SNP with variation of milk fat synthesis. Physiol. Genomics 21: 1–13.

Wellmann, R., and J. Bennewitz, 2011 The contribution of dominance to the understanding of quantitative genetic variation. Genet. Res. Camb. 93:139–154.

Winter, A., W. Krämer, F. A. O. Werner, S. Kollers, S. Kata et al., 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. 99: 9300–9305.

APPENDIX

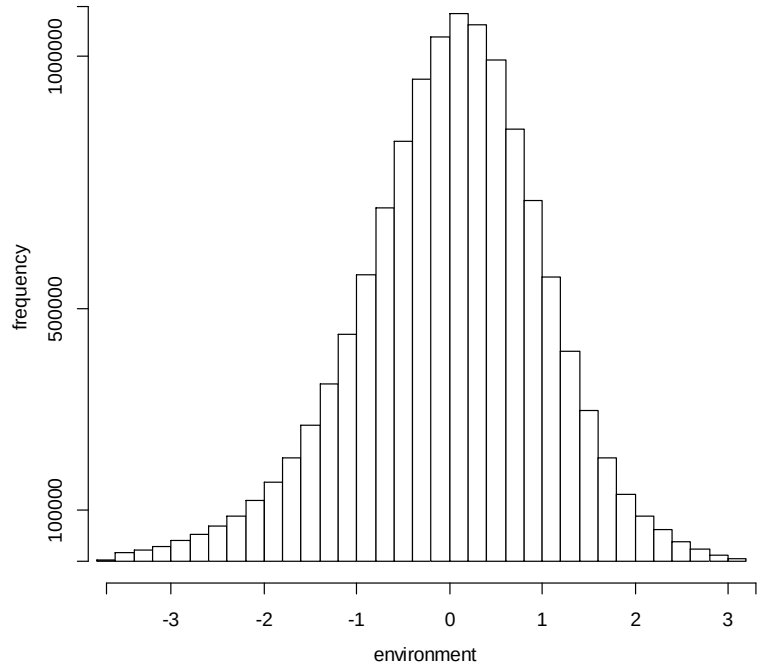

Figure S1 Histogram of the environmental descriptor milk energy yield.

Table S1 Validated SNPs with chromosome (BTA), position in base pairs (bp), F-values and effects for intercept and slope. Effect estimates were taken from the validation set. Validated SNPs are indicated in bold type F-values and effect estimates.

			F-values in discovery dataset		Effects (in σ)	
SNP name	BTA	bp	Intercept	Slope	Intercept	Slope
Protein yield						
ARS-BFGL-NGS-54077	1	87338372	7.09	**10.96**	0.094	**0.148**
ARS-BFGL-BAC-7205	1	120983738	**15.75**	**16.35**	**0.117**	**0.104**
ARS-BFGL-NGS-99492	1	121607486	**16.95**	**16.54**	**0.125**	**0.126**
ARS-BFGL-BAC-13578	1	121811393	**13.78**	7.48	**0.121**	0.102
ARS-BFGL-NGS-98257	1	127680108	**13.65**	4.44	**0.109**	0.075
ARS-BFGL-NGS-86079	2	19126180	**15.75**	6.79	**0.136**	0.114
Hapmap53232-rs29020795	2	19202356	**14.56**	9.90	**0.130**	0.109
Hapmap43615-BTA-54400	2	19255900	**11.87**	7.53	**0.139**	0.106
Hapmap60669-rs29018484	2	20687353	**16.88**	15.06	**0.222**	0.140
BTA-101354-no-rs	2	58818593	10.96	**11.72**	0.067	**0.116**
Hapmap28102-BTA-152636	2	58842740	10.84	**11.37**	0.069	**0.116**
ARS-BFGL-NGS-113152	4	14644971	**10.88**	8.87	**0.143**	0.139
ARS-BFGL-NGS-68464	5	18395406	**12.71**	10.07	**0.145**	0.090
Hapmap33079-BTA-163567	6	1936	**15.61**	5.94	**0.142**	0.068
Hapmap54442-rs29025673	6	31579806	**14.14**	8.19	**0.102**	0.092
Hapmap27701-BTC-050761	6	40183166	**12.68**	8.34	**0.126**	0.094
Hapmap31819-BTA-156590	6	40421547	**13.80**	9.11	**0.126**	0.094
Hapmap23201-BTC-072836	6	40655229	**21.27**	17.26	**0.113**	0.095
Hapmap32946-BTC-046820	6	41129701	**27.71**	15.88	**0.137**	0.086

ARS-BFGL-NGS-39570	6	46320087	**14.72**	3.26	**0.169**	0.159
ARS-BFGL-NGS-43679	6	109680595	**13.67**	4.00	**0.140**	0.054
BTB-00281303	6	111612203	**15.07**	2.41	**0.153**	0.090
ARS-BFGL-NGS-113181	7	62800839	**12.32**	10.78	**0.129**	0.108
ARS-BFGL-NGS-113819	7	63609102	**15.25**	6.29	**0.153**	0.089
ARS-BFGL-NGS-109819	7	63664393	**15.72**	5.91	**0.154**	0.086
ARS-BFGL-NGS-4062	8	5505897	6.17	**11.71**	0.116	**0.122**
Hapmap44053-BTA-28733	8	6369477	**15.90**	**14.16**	**0.118**	**0.133**
BTB-00389006	9	39062436	16.19	**16.89**	0.097	**0.124**
BTA-06997-rs29021351	9	40153426	**24.71**	22.35	**0.220**	0.154
BTA-83317-no-rs	9	40393986	**26.22**	21.89	**0.114**	0.081
BTA-83528-no-rs	9	41691114	**25.23**	18.48	**0.170**	0.112
ARS-BFGL-NGS-37982	9	44200288	**22.50**	15.00	**0.105**	0.084
ARS-BFGL-NGS-52530	9	44230587	**25.27**	**25.92**	**0.215**	**0.185**
ARS-BFGL-NGS-103934	9	44255942	**17.63**	17.33	**0.162**	0.115
Hapmap57331-rs29009884	9	45457232	14.35	**12.31**	0.072	**0.096**
ARS-BFGL-NGS-75844	9	45497534	14.31	**12.54**	0.072	**0.096**
ARS-BFGL-NGS-39444	9	45590254	14.31	**12.54**	0.072	**0.096**
BTA-10828-no-rs	9	46600974	24.59	**23.15**	0.096	**0.105**
BTA-83605-no-rs	9	48362593	**13.56**	10.43	**0.116**	0.103
BTB-00391835	9	52160813	**11.20**	5.70	**0.157**	0.085
ARS-BFGL-NGS-25071	9	75512134	7.97	**11.90**	0.051	**0.107**
ARS-BFGL-NGS-80176	10	64041482	**12.30**	9.49	**0.121**	0.129
ARS-BFGL-NGS-88689	11	29901111	7.53	**12.41**	0.123	**0.143**
ARS-BFGL-NGS-87426	11	30070765	**11.73**	**12.74**	**0.114**	**0.127**
ARS-BFGL-NGS-21332	11	30108643	8.61	**12.87**	0.120	**0.132**
ARS-BFGL-NGS-118724	11	30366110	**12.45**	10.44	**0.107**	0.121
ARS-BFGL-NGS-112015	13	63235775	**13.16**	2.09	**0.108**	0.050
Hapmap54034-rs29026486	13	65183781	**11.97**	4.07	**0.106**	0.073
ARS-BFGL-NGS-63777	13	67075815	**21.53**	**17.56**	**0.200**	**0.157**
ARS-BFGL-NGS-103635	13	67816926	**11.00**	1.02	**0.104**	0.059
ARS-BFGL-NGS-52851	13	77352863	11.87	**16.39**	0.048	**0.110**
Hapmap38308-BTA-33903	13	77756882	4.13	**11.00**	0.045	**0.102**
Hapmap30383-BTC-005848	14	76704	**45.65**	**31.66**	**0.168**	**0.134**
BTA-34956-no-rs	14	101474	**20.91**	7.85	**0.129**	0.112
ARS-BFGL-NGS-57820	14	236533	**72.24**	**56.71**	**0.193**	**0.148**
ARS-BFGL-NGS-34135	14	260342	**33.61**	18.35	**0.114**	0.095
ARS-BFGL-NGS-94706	14	281534	**32.30**	17.16	**0.118**	0.093
ARS-BFGL-NGS-4939	14	443936	**76.22**	**55.93**	**0.189**	**0.138**
ARS-BFGL-NGS-107379	14	679601	**61.01**	**51.65**	**0.158**	**0.122**
Hapmap25384-BTC-001997	14	835055	**27.13**	16.99	**0.117**	0.070
Hapmap24715-BTC-001973	14	856890	**22.88**	13.75	**0.116**	0.073
ARS-BFGL-NGS-103064	14	1193335	**23.15**	9.03	**0.109**	0.060
Hapmap25486-BTC-072553	14	1285036	**19.13**	8.43	**0.115**	0.083
ARS-BFGL-BAC-20965	14	5225005	**14.28**	8.91	**0.136**	0.104
Hapmap23851-BTC-048718	14	5387835	**17.21**	16.92	**0.126**	0.095
BTB-01988444	14	56535371	3.95	**11.67**	0.092	**0.126**
ARS-BFGL-NGS-100131	15	21041772	9.75	**11.06**	0.099	**0.132**
BTB-01698088	16	9738423	11.05	**16.13**	0.086	**0.139**
ARS-BFGL-NGS-56645	16	23920210	**15.37**	5.10	**0.117**	0.092
BTA-38367-no-rs	16	26379579	**12.89**	9.03	**0.096**	0.082
ARS-BFGL-NGS-41039	16	27619045	**14.91**	5.09	**0.183**	0.109
ARS-BFGL-NGS-38023	16	33318455	**18.15**	9.91	**0.107**	0.104
ARS-BFGL-NGS-26559	16	33367687	**11.43**	6.70	**0.124**	0.112
ARS-BFGL-NGS-59645	16	73117625	**15.94**	10.48	**0.124**	0.092
BTA-97501-no-rs	18	57095120	**16.33**	8.05	**0.108**	0.069

ARS-BFGL-NGS-15837	18	62533851	**11.11**	4.42	**0.110**	0.118
ARS-BFGL-BAC-33672	20	44108671	10.68	**12.05**	0.152	**0.185**
BTA-17135-no-rs	20	49557519	11.88	**16.26**	0.132	**0.175**
BTA-41516-no-rs	20	49593049	10.63	**15.64**	0.117	**0.165**
BTB-01251603	20	50418473	9.19	**12.88**	0.172	**0.198**
ARS-BFGL-BAC-34291	20	54837742	8.66	**13.62**	0.123	**0.152**
Hapmap53927-rs29025287	20	56464296	8.94	**13.45**	0.135	**0.179**
BTB-00787949	20	56494412	5.89	**12.27**	0.150	**0.186**
ARS-BFGL-BAC-36842	20	58820614	6.51	**14.46**	0.097	**0.146**
ARS-BFGL-NGS-38258	20	60185080	**11.42**	7.68	**0.138**	0.119
BTA-12959-no-rs	21	10922512	**13.46**	9.60	**0.114**	0.104
ARS-BFGL-NGS-101900	21	30314497	**14.12**	4.85	**0.106**	0.099
ARS-BFGL-NGS-110044	21	30892171	**11.81**	2.78	**0.158**	0.143
ARS-BFGL-NGS-55374	25	28795160	**24.40**	**25.01**	**0.215**	**0.185**
ARS-BFGL-NGS-2464	26	18709176	**20.40**	10.13	**0.098**	0.097
ARS-BFGL-NGS-71584	26	18863914	28.46	**15.95**	0.091	**0.102**
ARS-BFGL-NGS-113339	26	19918015	**16.35**	7.34	**0.105**	0.083
BTA-62184-no-rs	26	20014035	**17.65**	9.58	**0.118**	0.099
BTA-60778-no-rs	26	20090833	**21.14**	**14.75**	**0.127**	**0.117**
BTB-00932332	26	22551770	15.48	**11.32**	0.101	**0.130**
ARS-BFGL-NGS-107403	26	23470277	18.11	**12.42**	0.098	**0.122**
Hapmap28763-BTA-162328	26	26472420	12.31	**12.37**	0.085	**0.109**
BTB-01622498	28	1436040	3.42	**12.49**	0.100	**0.158**
ARS-BFGL-NGS-119076	28	15361777	**12.02**	8.12	**0.126**	0.086
ARS-BFGL-NGS-43501	28	17344828	14.42	**17.58**	0.083	**0.136**
ARS-BFGL-NGS-118693	28	17493199	10.68	**13.14**	0.139	**0.127**
Fat yield						
Hapmap38956-BTA-43309	1	98853038	**12.91**	10.82	**0.101**	0.114
BTB-01562245	1	113644592	**11.25**	**11.71**	**0.149**	**0.175**
BTA-104132-no-rs	1	113669783	**11.95**	**12.23**	**0.151**	**0.178**
ARS-BFGL-BAC-7205	1	120983738	**12.64**	12.33	**0.079**	0.072
ARS-BFGL-NGS-88388	2	13718482	**19.97**	**18.86**	**0.076**	**0.081**
ARS-BFGL-NGS-112315	2	40959609	**12.86**	5.12	**0.093**	0.091
ARS-BFGL-NGS-44416	2	48557519	**11.31**	9.45	**0.089**	0.082
ARS-BFGL-NGS-8503	2	87880854	**11.07**	**11.84**	**0.117**	**0.095**
BTB-00108243	2	112019423	**17.65**	**21.62**	**0.137**	**0.164**
BTB-01678000	3	6985014	**20.43**	20.37	**0.077**	0.068
BTB-01678060	3	7009487	**20.08**	**21.54**	**0.092**	**0.091**
ARS-BFGL-NGS-5956	3	8558755	**18.73**	**14.50**	**0.092**	**0.093**
ARS-BFGL-NGS-112616	3	8598511	**17.12**	**12.48**	**0.085**	**0.085**
ARS-BFGL-NGS-102139	3	24018818	9.79	**11.56**	0.067	**0.078**
Hapmap50814-BTA-89905	3	41486200	**11.97**	9.50	**0.095**	0.079
BTB-01851577	3	50093239	**11.69**	6.77	**0.089**	0.072
INRA-648	3	54347354	13.39	**11.31**	0.147	**0.168**
Hapmap43441-BTA-103289	3	61621627	17.59	**20.74**	0.068	**0.083**
BTA-68164-no-rs	3	68215262	**13.11**	8.02	**0.095**	0.062
BTB-00131847	3	68241075	**13.16**	8.09	**0.097**	0.066
BTA-54952-no-rs	4	11830564	**11.96**	9.08	**0.089**	0.090
ARS-BFGL-NGS-20815	4	15021946	**15.91**	14.68	**0.083**	0.076
ARS-BFGL-NGS-117196	4	53206872	9.84	**12.44**	0.073	**0.085**
ARS-BFGL-NGS-110647	4	77565085	9.95	**12.41**	0.081	**0.093**
Hapmap39895-BTA-15668	5	15392995	**11.42**	7.32	**0.093**	0.081
ARS-BFGL-NGS-108617	5	98082173	**15.12**	**11.57**	**0.107**	**0.110**
ARS-BFGL-NGS-95906	5	100351926	**20.33**	**17.38**	**0.090**	**0.091**
Hapmap53294-rs29016908	5	101090418	**40.43**	**22.97**	**0.086**	**0.095**

Hapmap60021-ss46526426	5	101979581	**21.19**	**13.64**	**0.095**	**0.088**
BTB-00270281	6	95770023	**16.38**	12.12	**0.095**	0.073
Hapmap58150-rs29020620	6	96724594	**11.78**	8.53	**0.076**	0.070
BTB-01700063	6	99086447	**21.66**	**16.75**	**0.114**	**0.102**
Hapmap53916-rs29021982	6	99581269	**18.46**	**15.67**	**0.094**	**0.094**
Hapmap48078-BTA-77495	6	99827767	**13.22**	**14.54**	**0.076**	**0.079**
ARS-BFGL-NGS-14880	7	53879989	**14.95**	**17.58**	**0.115**	**0.104**
ARS-BFGL-NGS-113181	7	62800839	**17.55**	**20.39**	**0.077**	**0.080**
BTB-02035459	7	63196194	**11.52**	8.67	**0.073**	0.073
BTB-01219396	7	63221359	**11.52**	8.67	**0.073**	0.073
ARS-BFGL-NGS-113819	7	63609102	12.18	**13.22**	0.071	**0.081**
ARS-BFGL-NGS-109819	7	63664393	12.02	**12.76**	0.073	**0.083**
BTA-12616-no-rs	7	64712171	**12.81**	8.91	**0.074**	0.063
ARS-BFGL-NGS-65419	7	66102696	**15.93**	**13.18**	**0.092**	**0.094**
ARS-BFGL-NGS-12863	7	68960712	**16.23**	7.43	**0.105**	0.092
ARS-BFGL-NGS-23091	7	69342623	**13.06**	6.26	**0.110**	0.095
BTB-01222854	7	74587278	**13.21**	**11.20**	**0.105**	**0.104**
BTB-01321253	7	83625073	**19.55**	**20.56**	**0.083**	**0.077**
Hapmap44053-BTA-28733	8	6369477	**15.07**	**12.22**	**0.096**	**0.113**
BTA-102639-no-rs	8	29764842	7.33	**11.04**	0.083	**0.105**
ARS-BFGL-NGS-68597	8	32188298	9.59	**12.15**	0.144	**0.139**
BTB-01184997	8	36171902	**11.48**	10.62	**0.097**	0.092
Hapmap31805-BTA-154485	8	36232703	**12.95**	**13.66**	**0.078**	**0.084**
Hapmap41758-BTA-116042	8	37197678	**11.23**	7.67	**0.105**	0.105
BTB-00384442	9	22407445	**11.25**	8.57	**0.094**	0.079
BTA-83317-no-rs	9	40393986	**18.16**	16.64	**0.087**	0.077
BTA-83528-no-rs	9	41691114	**17.66**	15.90	**0.132**	0.115
Hapmap34441-BES9_Contig154_536	9	73154820	**11.99**	10.47	**0.084**	0.091
UA-IFASA-2589	9	82175488	9.65	**11.75**	0.072	**0.086**
Hapmap30370-BTA-99862	11	14499391	**14.20**	8.56	**0.104**	0.100
Hapmap29758-BTC-003619	14	5261	**39.73**	**33.19**	**0.112**	**0.109**
Hapmap30381-BTC-005750	14	50873	**59.46**	**54.84**	**0.129**	**0.129**
Hapmap30383-BTC-005848	14	76704	**65.31**	**55.73**	**0.132**	**0.132**
BTA-34956-no-rs	14	101474	**29.89**	**29.89**	**0.106**	**0.104**
ARS-BFGL-NGS-57820	14	236533	**147.55**	**121.82**	**0.194**	**0.195**
ARS-BFGL-NGS-34135	14	260342	**79.17**	**71.47**	**0.146**	**0.144**
ARS-BFGL-NGS-94706	14	281534	**75.70**	**68.83**	**0.143**	**0.141**
ARS-BFGL-NGS-4939	14	443936	**159.02**	**132.41**	**0.206**	**0.208**
ARS-BFGL-NGS-71749	14	596340	**41.77**	**36.70**	**0.144**	**0.147**
ARS-BFGL-NGS-107379	14	679601	**101.31**	**78.31**	**0.163**	**0.160**
ARS-BFGL-NGS-18365	14	741868	**46.48**	**33.09**	**0.084**	**0.094**
Hapmap30922-BTC-002021	14	763332	**49.05**	**35.93**	**0.087**	**0.101**
Hapmap25384-BTC-001997	14	835055	**63.54**	**55.55**	**0.103**	**0.119**
Hapmap24715-BTC-001973	14	856890	**60.80**	**54.16**	**0.105**	**0.120**
BTA-35941-no-rs	14	894253	**97.39**	**78.45**	**0.155**	**0.156**
ARS-BFGL-NGS-101653	14	931163	**41.64**	**36.14**	**0.133**	**0.144**
ARS-BFGL-NGS-26520	14	996983	**69.35**	**59.88**	**0.149**	**0.159**
UA-IFASA-6878	14	1044040	**36.43**	26.64	**0.074**	0.071
ARS-BFGL-NGS-22866	14	1131951	**44.94**	**42.65**	**0.126**	**0.119**
ARS-BFGL-NGS-103064	14	1193335	**61.80**	**58.85**	**0.110**	**0.138**
ARS-BFGL-NGS-3122	14	1264232	**33.56**	**31.89**	**0.092**	**0.126**
Hapmap25486-BTC-072553	14	1285036	**34.57**	**35.23**	**0.108**	**0.121**
Hapmap30646-BTC-002054	14	1461084	**68.94**	**51.83**	**0.142**	**0.142**
Hapmap30086-BTC-002066	14	1490177	**110.60**	**85.69**	**0.147**	**0.152**
Hapmap30374-BTC-002159	14	1546590	**89.52**	**73.07**	**0.153**	**0.154**
ARS-BFGL-NGS-74378	14	1889209	**48.29**	**35.14**	**0.099**	**0.107**

ARS-BFGL-NGS-117542	14	1913107	**28.30**	**24.11**	**0.103**	**0.105**
UA-IFASA-9288	14	2201869	**48.59**	**32.70**	**0.089**	**0.103**
Hapmap24777-BTC-064977	14	2261622	**14.06**	**12.67**	**0.079**	**0.095**
Hapmap32970-BTC-064990	14	2288509	**31.99**	**24.55**	**0.080**	**0.095**
Hapmap24986-BTC-065021	14	2313594	**31.99**	**24.55**	**0.080**	**0.095**
ARS-BFGL-NGS-22111	14	2347218	19.48	**18.07**	0.059	**0.079**
UA-IFASA-7269	14	2370255	19.48	**18.07**	0.059	**0.079**
Hapmap26072-BTC-065132	14	2391825	24.28	**24.38**	0.070	**0.087**
ARS-BFGL-NGS-113575	14	2484498	**35.10**	**30.78**	**0.085**	**0.097**
ARS-BFGL-NGS-118081	14	2511264	**41.29**	**37.93**	**0.113**	**0.116**
ARS-BFGL-NGS-56327	14	2580413	**60.32**	**50.89**	**0.105**	**0.116**
ARS-BFGL-NGS-100480	14	2607582	**75.52**	**63.36**	**0.121**	**0.128**
UA-IFASA-5306	14	2711614	**48.24**	**32.60**	**0.099**	**0.110**
Hapmap27703-BTC-053907	14	2826072	27.55	**23.02**	0.063	**0.081**
Hapmap22692-BTC-068210	14	3018725	**36.91**	**23.96**	**0.080**	**0.086**
Hapmap23302-BTC-052123	14	3099634	**36.05**	**20.28**	**0.090**	**0.096**
UA-IFASA-6329	14	3465238	**25.41**	**16.40**	**0.094**	**0.083**
ARS-BFGL-NGS-3571	14	3587017	**26.21**	**18.53**	**0.099**	**0.101**
ARS-BFGL-NGS-110563	14	3799229	**25.68**	**15.99**	**0.103**	**0.105**
Hapmap32262-BTC-066621	14	3834070	**13.39**	8.84	**0.087**	0.083
ARS-BFGL-NGS-115947	14	3865963	**34.05**	**19.86**	**0.107**	**0.109**
Hapmap30988-BTC-056315	14	4693900	**20.44**	**15.88**	**0.097**	**0.091**
ARS-BFGL-NGS-110894	14	5282437	**15.13**	7.87	**0.094**	0.101
UA-IFASA-6647	14	5808643	**21.59**	12.18	**0.072**	0.073
ARS-BFGL-NGS-102953	14	5867265	**19.06**	9.11	**0.076**	0.078
ARS-BFGL-NGS-37911	14	14274020	**11.29**	9.61	**0.090**	0.096
ARS-BFGL-NGS-16622	15	64781292	**11.71**	9.29	**0.089**	0.085
BTA-37324-no-rs	15	64802082	**10.94**	8.43	**0.091**	0.083
Hapmap54310-rs29012181	16	8166691	9.72	**12.29**	0.093	**0.092**
Hapmap43402-BTA-91283	16	8249735	8.15	**11.65**	0.086	**0.088**
ARS-BFGL-NGS-56645	16	23920210	**11.80**	8.21	**0.082**	0.073
ARS-BFGL-NGS-26559	16	33367687	**11.77**	10.10	**0.089**	0.094
ARS-BFGL-NGS-93660	16	77541689	**12.02**	12.76	**0.088**	0.073
ARS-BFGL-NGS-18128	17	22770146	**11.24**	8.71	**0.096**	0.100
ARS-BFGL-NGS-91287	18	10052638	**17.13**	**20.25**	**0.117**	**0.101**
ARS-BFGL-NGS-111247	19	43146804	**14.32**	15.09	**0.075**	0.072
ARS-BFGL-NGS-24479	19	45901285	**11.09**	9.17	**0.092**	0.087
ARS-BFGL-NGS-113693	19	45926259	**12.27**	10.29	**0.092**	0.087
BTB-00783355	20	43550938	10.23	**16.06**	0.087	**0.111**
Hapmap48608-BTA-111028	20	52535573	12.29	**16.19**	0.076	**0.088**
BTA-12959-no-rs	21	10922512	**12.61**	11.88	**0.074**	0.071
ARS-BFGL-NGS-101900	21	30314497	**12.99**	6.17	**0.073**	0.066
ARS-BFGL-NGS-39397	26	21166268	**12.98**	10.82	**0.078**	0.081
Hapmap46411-BTA-15820	26	21404446	**14.10**	**12.01**	**0.094**	**0.103**
Hapmap31825-BTA-158647	26	21476707	**15.45**	**11.71**	**0.073**	**0.079**
ARS-BFGL-NGS-110077	26	21729361	**13.86**	**14.13**	**0.086**	**0.086**
ARS-BFGL-NGS-18603	26	21853286	21.66	**22.77**	0.067	**0.078**
ARS-BFGL-NGS-116481	26	22411701	**22.81**	**22.25**	**0.077**	**0.079**
Hapmap24832-BTA-138805	26	22449570	**23.72**	**23.06**	**0.077**	**0.079**
ARS-BFGL-NGS-6259	26	22492302	**21.27**	**20.72**	**0.077**	**0.079**
BTB-00932332	26	22551770	22.91	**21.24**	0.073	**0.084**
ARS-BFGL-NGS-1092	26	24837303	**17.02**	**19.39**	**0.096**	**0.093**
UA-IFASA-4715	26	25330026	16.24	**20.09**	0.063	**0.081**
ARS-BFGL-NGS-38386	26	32836475	7.14	**11.37**	0.065	**0.103**
ARS-BFGL-NGS-105944	26	34196569	**11.04**	**11.12**	**0.106**	**0.109**
ARS-BFGL-NGS-53731	26	37801879	13.09	**15.19**	0.062	**0.087**

ARS-USMARC-Parent-EF034086-no-rs	26	38309860	16.12	**17.87**	0.054	**0.075**
Hapmap35000-BES9_Contig272_944	26	38309861	16.12	**17.87**	0.054	**0.075**
ARS-BFGL-NGS-60822	28	17252660	16.04	**12.16**	0.066	**0.079**
UA-IFASA-6208	28	27379038	**12.33**	8.31	**0.081**	0.081
Milk yield						
ARS-BFGL-NGS-23720	1	38486179	4.91	**10.88**	0.146	**0.130**
ARS-BFGL-BAC-13578	1	121811393	**13.51**	**11.19**	**0.189**	**0.179**
ARS-BFGL-NGS-86079	2	19126180	**14.12**	9.49	**0.194**	0.182
Hapmap53232-rs29020795	2	19202356	**12.91**	**11.87**	**0.170**	**0.161**
Hapmap60669-rs29018484	2	20687353	**13.93**	12.24	**0.324**	0.249
ARS-BFGL-NGS-23786	2	22727721	**12.86**	8.56	**0.170**	0.132
UA-IFASA-4555	2	103573252	**10.91**	7.81	**0.347**	0.267
ARS-BFGL-NGS-119065	3	56970477	8.93	**12.42**	0.280	**0.256**
ARS-BFGL-NGS-75548	4	78246158	**15.43**	**11.18**	**0.446**	**0.338**
BTA-74090-no-rs	5	7569039	9.55	**12.21**	0.178	**0.147**
Hapmap33079-BTA-163567	6	1936	**13.03**	9.90	**0.175**	0.098
ARS-BFGL-NGS-117147	6	31235325	**10.90**	3.52	**0.199**	0.135
BTA-75680-no-rs	6	31470687	**11.66**	6.98	**0.157**	0.102
Hapmap54442-rs29025673	6	31579806	**12.51**	8.27	**0.157**	0.102
Hapmap31819-BTA-156590	6	40421547	10.20	**11.60**	0.165	**0.124**
UA-IFASA-2111	6	85289127	**14.59**	15.14	**0.192**	0.122
Hapmap25708-BTC-043671	6	88263655	**11.16**	11.06	**0.213**	0.111
Hapmap40845-BTA-97263	6	92788189	**11.94**	7.87	**0.196**	0.143
BTB-01428914	6	93850918	**12.79**	8.07	**0.157**	0.098
BTB-00281303	6	111612203	**10.87**	3.17	**0.209**	0.093
Hapmap53417-rs29014877	7	106039868	**13.08**	11.74	**0.164**	0.114
ARS-BFGL-NGS-4062	8	5505897	6.84	**11.25**	0.199	**0.183**
Hapmap44053-BTA-28733	8	6369477	13.69	**11.43**	0.151	**0.166**
BTA-82314-no-rs	8	97201103	9.03	**15.81**	0.162	**0.172**
ARS-BFGL-NGS-52530	9	44230587	**12.72**	10.89	**0.273**	0.192
Hapmap49359-BTA-88301	9	45435722	8.60	**11.76**	0.147	**0.149**
ARS-BFGL-NGS-75844	9	45497534	7.87	**11.04**	0.133	**0.150**
ARS-BFGL-NGS-39444	9	45590254	7.87	**11.04**	0.133	**0.150**
BTA-10828-no-rs	9	46600974	**15.86**	**18.25**	**0.160**	**0.150**
ARS-BFGL-NGS-25071	9	75512134	9.95	**13.07**	0.083	**0.137**
ARS-BFGL-NGS-62628	9	82136926	10.06	**11.58**	0.149	**0.133**
ARS-BFGL-NGS-105675	11	104891662	11.47	**14.14**	0.145	**0.156**
ARS-BFGL-NGS-78549	11	106534689	**12.38**	**14.06**	**0.234**	**0.220**
ARS-BFGL-NGS-56157	13	63208626	**13.95**	5.96	**0.163**	0.088
Hapmap54034-rs29026486	13	65183781	**14.87**	4.39	**0.188**	0.090
ARS-BFGL-NGS-103635	13	67816926	**13.47**	4.35	**0.157**	0.081
Hapmap44949-BTA-33430	13	67920496	**11.12**	5.80	**0.209**	0.119
Hapmap29758-BTC-003619	14	5261	**41.37**	**45.38**	**0.175**	**0.156**
Hapmap30381-BTC-005750	14	50873	**46.72**	**39.87**	**0.280**	**0.254**
Hapmap30383-BTC-005848	14	76704	**130.39**	**113.46**	**0.373**	**0.292**
BTA-34956-no-rs	14	101474	**69.39**	**51.08**	**0.287**	**0.246**
ARS-BFGL-NGS-57820	14	236533	**218.33**	**207.08**	**0.478**	**0.372**
ARS-BFGL-NGS-34135	14	260342	**111.86**	**96.34**	**0.305**	**0.261**
ARS-BFGL-NGS-94706	14	281534	**109.45**	**91.11**	**0.311**	**0.259**
ARS-BFGL-NGS-4939	14	443936	**226.90**	**210.30**	**0.477**	**0.371**
ARS-BFGL-NGS-71749	14	596340	**32.17**	24.98	**0.205**	0.153
ARS-BFGL-NGS-107379	14	679601	**165.98**	**159.61**	**0.387**	**0.310**
ARS-BFGL-NGS-18365	14	741868	**36.88**	44.68	**0.202**	0.115
Hapmap30922-BTC-002021	14	763332	**27.70**	33.83	**0.187**	0.111

Hapmap25384-BTC-001997	14	835055	**86.29**	**71.08**	**0.294**	**0.201**
Hapmap24715-BTC-001973	14	856890	**76.90**	**63.03**	**0.290**	**0.203**
BTA-35941-no-rs	14	894253	**81.51**	**91.67**	**0.245**	**0.196**
ARS-BFGL-NGS-101653	14	931163	**48.62**	**43.61**	**0.228**	**0.173**
ARS-BFGL-NGS-26520	14	996983	**55.92**	**51.37**	**0.211**	**0.151**
UA-IFASA-6878	14	1044040	**82.94**	**84.88**	**0.254**	**0.200**
ARS-BFGL-NGS-22866	14	1131951	**53.32**	**49.99**	**0.190**	**0.190**
ARS-BFGL-NGS-103064	14	1193335	**78.12**	**56.27**	**0.288**	**0.196**
ARS-BFGL-NGS-3122	14	1264232	**56.10**	**39.10**	**0.262**	**0.155**
Hapmap25486-BTC-072553	14	1285036	**54.16**	**34.14**	**0.269**	**0.185**
Hapmap30646-BTC-002054	14	1461084	**61.98**	**70.96**	**0.216**	**0.177**
Hapmap30086-BTC-002066	14	1490177	**77.59**	**89.06**	**0.244**	**0.198**
Hapmap30374-BTC-002159	14	1546590	**84.37**	**91.65**	**0.236**	**0.193**
ARS-BFGL-NGS-74378	14	1889209	**51.60**	**62.95**	**0.264**	**0.209**
ARS-BFGL-NGS-117542	14	1913107	**37.52**	**38.66**	**0.239**	**0.201**
UA-IFASA-9288	14	2201869	**34.84**	**42.80**	**0.209**	**0.147**
Hapmap32970-BTC-064990	14	2288509	**19.04**	19.18	**0.165**	0.102
Hapmap24986-BTC-065021	14	2313594	**19.04**	19.18	**0.165**	0.102
Hapmap26072-BTC-065132	14	2391825	**37.91**	**27.05**	**0.221**	**0.173**
Hapmap26527-BTC-005059	14	2418618	**26.38**	**22.86**	**0.175**	**0.140**
ARS-BFGL-NGS-113575	14	2484498	**15.20**	**18.65**	**0.173**	**0.136**
ARS-BFGL-NGS-118081	14	2511264	**27.04**	**26.10**	**0.189**	**0.167**
ARS-BFGL-NGS-56327	14	2580413	**37.86**	**38.51**	**0.224**	**0.179**
ARS-BFGL-NGS-100480	14	2607582	**47.61**	**45.61**	**0.243**	**0.192**
ARS-BFGL-NGS-42263	14	2681399	**23.44**	**18.82**	**0.197**	**0.165**
UA-IFASA-5306	14	2711614	**36.83**	**45.80**	**0.223**	**0.171**
ARS-BFGL-NGS-54400	14	2736946	**21.12**	**16.78**	**0.206**	**0.163**
Hapmap22783-BTC-068255	14	2989274	**25.42**	**32.59**	**0.207**	**0.178**
Hapmap22692-BTC-068210	14	3018725	**34.59**	**45.38**	**0.199**	**0.156**
Hapmap23302-BTC-052123	14	3099634	**36.42**	**52.77**	**0.206**	**0.161**
Hapmap25217-BTC-067767	14	3189311	**35.96**	24.53	**0.189**	0.094
UA-IFASA-6329	14	3465238	**25.58**	**29.83**	**0.166**	**0.133**
ARS-BFGL-NGS-56339	14	3498808	**15.52**	**20.81**	**0.160**	**0.155**
UA-IFASA-8927	14	3640095	**18.48**	**18.50**	**0.154**	**0.130**
Hapmap32262-BTC-066621	14	3834070	**28.92**	**27.90**	**0.193**	**0.140**
Hapmap30091-BTC-005211	14	3940999	**26.82**	**23.11**	**0.225**	**0.149**
ARS-BFGL-BAC-24839	14	3993201	**22.12**	**26.61**	**0.164**	**0.137**
ARS-BFGL-BAC-24804	14	4157676	**42.02**	**59.28**	**0.163**	**0.136**
Hapmap51646-BTA-86764	14	4302230	**33.30**	41.88	**0.186**	0.111
ARS-BFGL-NGS-112858	14	4956374	**25.89**	**37.11**	**0.199**	**0.155**
Hapmap51078-BTA-87682	14	5064062	**15.80**	**12.92**	**0.192**	**0.127**
ARS-BFGL-NGS-55227	14	5085415	**19.63**	**33.44**	**0.163**	**0.145**
Hapmap32236-BTC-049785	14	5139497	19.37	**34.36**	0.142	**0.131**
ARS-BFGL-BAC-20965	14	5225005	**23.50**	**18.29**	**0.244**	**0.175**
Hapmap33635-BTC-049051	14	5318260	**16.46**	5.75	**0.192**	0.103
Hapmap27091-BTC-048823	14	5356987	**29.54**	27.13	**0.197**	0.113
Hapmap23851-BTC-048718	14	5387835	**28.66**	**24.81**	**0.236**	**0.131**
Hapmap32234-BTC-048199	14	5640337	**32.52**	**37.23**	**0.209**	**0.148**
Hapmap26283-BTC-048098	14	5696728	15.08	**30.24**	0.151	**0.128**
Hapmap32948-BTC-047992	14	5839289	**22.95**	**32.42**	**0.190**	**0.138**
Hapmap24756-BTC-047876	14	5915294	**20.39**	**19.47**	**0.215**	**0.158**
Hapmap25716-BTC-047850	14	5937549	**16.11**	**19.08**	**0.196**	**0.139**
Hapmap23799-BTC-047701	14	6044245	**11.39**	12.82	**0.212**	0.130
ARS-BFGL-BAC-8730	14	6252100	**30.95**	21.19	**0.187**	0.118
Hapmap53312-rs29018332	14	60576872	**20.09**	6.88	**0.159**	0.067
BTA-35525-no-rs	14	72400486	9.43	**11.65**	0.078	**0.154**

ARS-BFGL-BAC-23729	14	73127978	4.90	**11.73**	0.054	**0.168**
Hapmap43128-BTA-105550	15	52541506	**12.54**	7.43	**0.155**	0.084
Hapmap59019-rs29021918	18	41938135	10.02	**16.37**	0.118	**0.157**
ARS-BFGL-NGS-17557	20	1290829	**13.98**	11.50	**0.172**	0.127
Hapmap39811-BTA-122745	20	35432864	10.43	**13.05**	0.161	**0.165**
BTB-01888575	20	35457272	8.48	**11.86**	0.177	**0.185**
BTA-50244-no-rs	20	36466784	6.63	**12.12**	0.130	**0.167**
BTA-50402-no-rs	20	36667999	7.94	**15.52**	0.141	**0.179**
Hapmap26466-BTA-160199	20	36746234	8.57	**13.82**	0.151	**0.185**
BTA-50386-no-rs	20	36837402	6.13	**13.57**	0.134	**0.154**
BTA-50376-no-rs	20	36915967	**15.87**	**25.25**	**0.160**	**0.170**
Hapmap57531-rs29013890	20	36955574	**13.12**	**17.52**	**0.181**	**0.170**
Hapmap54884-rs29017180	20	50564369	8.80	**17.44**	0.174	**0.128**
ARS-BFGL-NGS-37182	22	5301596	**10.98**	10.56	**0.267**	0.273
ARS-BFGL-NGS-80066	23	20578210	**11.18**	7.83	**0.335**	0.200
Hapmap43294-BTA-56514	23	32759583	**12.29**	8.13	**0.303**	0.165
ARS-BFGL-NGS-55374	25	28795160	**12.07**	10.48	**0.273**	0.192
BTB-02094179	26	467952	7.37	**10.98**	0.183	**0.196**
ARS-BFGL-NGS-2127	26	13563198	**13.90**	8.58	**0.158**	0.115
ARS-BFGL-NGS-2464	26	18709176	**13.00**	8.73	**0.150**	0.134
ARS-BFGL-NGS-77668	26	18760372	19.50	**16.74**	0.132	**0.131**
ARS-BFGL-NGS-23064	26	18788121	18.84	**15.64**	0.132	**0.131**
ARS-BFGL-NGS-71584	26	18863914	19.45	**14.43**	0.113	**0.127**
BTB-00930720	26	21323659	**12.27**	**10.90**	**0.148**	**0.151**
Hapmap31825-BTA-158647	26	21476707	**14.00**	**11.18**	**0.160**	**0.135**
BTB-00931481	26	21631982	21.41	**17.08**	0.157	**0.157**
ARS-BFGL-NGS-18603	26	21853286	13.82	**14.96**	0.138	**0.139**
BTB-00932332	26	22551770	15.47	**14.23**	0.153	**0.177**
ARS-BFGL-NGS-107403	26	23470277	18.24	**16.31**	0.165	**0.181**
ARS-BFGL-NGS-119314	26	25634039	11.67	**14.68**	0.130	**0.148**
ARS-BFGL-NGS-109460	27	46280579	**14.33**	3.91	**0.172**	0.101
ARS-BFGL-NGS-116386	28	18996324	7.27	**15.71**	0.189	**0.148**
BTB-02080610	28	19938671	**12.70**	9.76	**0.201**	0.151
Hapmap58649-rs29011010	28	28185115	**15.70**	9.11	**0.154**	0.127
ARS-BFGL-NGS-116154	29	49020982	8.01	**10.91**	0.142	**0.165**

Table S2 Validated SNPs with chromosome (BTA), position in base pairs (bp), F-values and effects for intercept and slope. Effect estimates were taken from the validation set. Validated SNPs are indicated in bold type F-values and effect estimates. Results from the analysis of the log-transformed data set.

			F-values in discovery dataset		Effects (in σ)	
SNP name	BTA	bp	Intercept	Slope	Intercept	Slope
ln(Protein)						
BTB-00014850	1	33047886	24.41	**22.24**	0.037	**0.120**
BTA-17929-no-rs	1	94762863	14.03	**10.87**	0.130	**0.157**
ARS-BFGL-BAC-7205	1	120983738	**15.50**	**13.42**	**0.190**	**0.124**
ARS-BFGL-NGS-99492	1	121607486	16.79	**14.24**	0.159	**0.144**
BTB-01116994	1	122957014	13.08	**12.97**	0.130	**0.171**
ARS-BFGL-NGS-5956	3	8558755	**11.21**	6.20	**0.225**	0.045
Hapmap22861-BTA-141421	3	60025569	10.41	**12.01**	0.093	**0.089**
Hapmap39895-BTA-15668	5	15392995	**14.67**	12.25	**0.219**	0.085
ARS-BFGL-NGS-68464	5	18395406	**12.53**	9.43	**0.229**	0.104
Hapmap23201-BTC-072836	6	40655229	21.27	**16.53**	0.122	**0.110**

Hapmap58150-rs29020620	6	96724594	**20.75**	7.37	**0.188**	0.096
BTB-00281303	6	111612203	**15.31**	2.36	**0.206**	0.104
ARS-BFGL-NGS-14880	7	53879989	**12.33**	17.75	**0.277**	0.125
ARS-BFGL-NGS-113181	7	62800839	**12.44**	**10.95**	**0.186**	**0.085**
Hapmap44053-BTA-28733	8	6369477	**16.59**	**14.72**	**0.232**	**0.141**
Hapmap31805-BTA-154485	8	36232703	**11.49**	4.12	**0.187**	0.074
Hapmap57000-rs29011304	9	26662	13.22	**19.26**	0.112	**0.109**
BTB-00389006	9	39062436	16.14	**15.30**	0.144	**0.134**
BTA-83317-no-rs	9	40393986	**26.21**	**19.61**	**0.212**	**0.119**
BTA-83528-no-rs	9	41691114	**25.03**	16.16	**0.330**	0.134
BTA-59878-no-rs	9	44095822	28.62	**15.86**	0.140	**0.100**
ARS-BFGL-NGS-52530	9	44230587	**25.18**	**25.13**	**0.275**	**0.224**
ARS-BFGL-NGS-103934	9	44255942	17.70	**18.51**	0.224	**0.138**
BTA-10828-no-rs	9	46600974	24.96	**21.33**	0.109	**0.128**
Hapmap24524-BTA-107865	9	47934344	10.86	**17.83**	0.064	**0.105**
ARS-BFGL-NGS-51043	9	51493578	7.19	**11.78**	0.004	**0.091**
BTB-00391835	9	52160813	**11.19**	5.86	**0.202**	0.101
ARS-BFGL-NGS-114465	9	79835130	7.21	**12.68**	0.032	**0.099**
UA-IFASA-2589	9	82175488	**16.53**	14.78	**0.175**	0.095
ARS-BFGL-NGS-88689	11	29901111	7.51	**14.05**	0.176	**0.171**
ARS-BFGL-NGS-87426	11	30070765	11.64	**14.05**	0.140	**0.143**
ARS-BFGL-NGS-21332	11	30108643	8.57	**14.57**	0.171	**0.159**
ARS-BFGL-NGS-118724	11	30366110	12.31	**11.69**	0.171	**0.145**
Hapmap45971-BTA-102151	11	71170530	15.89	**11.36**	0.071	**0.107**
ARS-BFGL-BAC-12483	13	1310816	9.44	**12.89**	0.185	**0.138**
Hapmap45253-BTA-15908	13	1498184	**11.23**	8.74	**0.250**	0.159
ARS-BFGL-NGS-63777	13	67075815	21.41	**16.44**	0.253	**0.162**
ARS-BFGL-NGS-52851	13	77352863	11.80	**15.82**	0.114	**0.128**
Hapmap29758-BTC-003619	14	5261	**14.75**	11.14	**0.271**	-0.053
Hapmap30383-BTC-005848	14	76704	**45.90**	**32.63**	**0.319**	**-0.148**
BTA-34956-no-rs	14	101474	**21.25**	8.68	**0.258**	-0.126
ARS-BFGL-NGS-57820	14	236533	**72.35**	**54.42**	**0.470**	**-0.162**
ARS-BFGL-NGS-34135	14	260342	**33.86**	17.84	**0.354**	-0.104
ARS-BFGL-NGS-94706	14	281534	**32.66**	16.59	**0.346**	-0.100
ARS-BFGL-NGS-4939	14	443936	**76.53**	**54.67**	**0.497**	**-0.157**
ARS-BFGL-NGS-107379	14	679601	**61.30**	**51.49**	**0.391**	**-0.136**
Hapmap25384-BTC-001997	14	835055	**27.04**	14.45	**0.247**	-0.061
Hapmap24715-BTC-001973	14	856890	**22.96**	11.30	**0.253**	-0.064
BTA-35941-no-rs	14	894253	**25.36**	22.23	**0.375**	-0.076
ARS-BFGL-NGS-101653	14	931163	**20.57**	14.60	**0.322**	-0.094
ARS-BFGL-NGS-26520	14	996983	**20.29**	13.50	**0.361**	-0.058
UA-IFASA-6878	14	1044040	**32.09**	28.23	**0.176**	-0.076
ARS-BFGL-NGS-22866	14	1131951	**20.13**	10.87	**0.307**	-0.098
ARS-BFGL-NGS-103064	14	1193335	**22.91**	7.53	**0.264**	-0.050
ARS-BFGL-NGS-3122	14	1264232	**19.19**	6.75	**0.220**	-0.029
Hapmap25486-BTC-072553	14	1285036	**19.18**	7.04	**0.262**	-0.076
Hapmap30646-BTC-002054	14	1461084	**25.97**	**27.24**	**0.346**	**0.130**
Hapmap30086-BTC-002066	14	1490177	**22.30**	21.27	**0.358**	-0.091
Hapmap30374-BTC-002159	14	1546590	**26.64**	21.65	**0.370**	-0.073
ARS-BFGL-NGS-74378	14	1889209	**18.11**	16.86	**0.236**	-0.059
ARS-BFGL-NGS-117542	14	1913107	**16.54**	12.41	**0.249**	-0.107
ARS-BFGL-NGS-100480	14	2607582	**11.61**	6.51	**0.292**	-0.033
UA-IFASA-5306	14	2711614	**11.40**	11.87	**0.236**	-0.029
Hapmap22692-BTC-068210	14	3018725	**12.60**	13.43	**0.193**	-0.047
Hapmap23302-BTC-052123	14	3099634	**14.23**	18.84	**0.215**	-0.050
UA-IFASA-6329	14	3465238	**10.89**	16.37	**0.231**	-0.036
ARS-BFGL-NGS-110563	14	3799229	**12.33**	14.60	**0.250**	-0.024
Hapmap32262-BTC-066621	14	3834070	**14.68**	9.51	**0.213**	-0.071
Hapmap32236-BTC-049785	14	5139497	12.66	**20.47**	0.142	**0.111**

BTB-01988444	14	56535371	3.79	**10.99**	0.077	**0.143**
ARS-BFGL-NGS-100131	15	21041772	9.65	**10.88**	0.076	**0.145**
BTB-01698088	16	9738423	11.30	**14.27**	0.156	**0.144**
ARS-BFGL-NGS-56645	16	23920210	**15.64**	5.92	**0.199**	0.060
ARS-BFGL-NGS-38023	16	33318455	18.62	**13.72**	0.105	**0.128**
ARS-BFGL-NGS-26559	16	33367687	**11.76**	10.47	**0.216**	0.141
ARS-BFGL-NGS-59645	16	73117625	**16.15**	10.33	**0.235**	0.086
ARS-BFGL-NGS-21141	17	3125444	5.55	**14.51**	0.092	**0.105**
Hapmap58542-rs29018933	17	29430592	4.76	**12.24**	0.071	**0.087**
ARS-BFGL-BAC-33672	20	44108671	10.55	**12.67**	0.235	**0.196**
BTA-17135-no-rs	20	49557519	11.68	**16.41**	0.154	**0.168**
BTB-01251603	20	50418473	8.95	**11.30**	0.178	**0.214**
Hapmap54884-rs29017180	20	50564369	4.84	**12.50**	0.059	**0.087**
ARS-BFGL-BAC-34291	20	54837742	8.44	**14.24**	0.145	**0.177**
Hapmap53927-rs29025287	20	56464296	8.63	**11.17**	0.152	**0.185**
ARS-BFGL-BAC-36842	20	58820614	6.52	**11.22**	0.125	**0.151**
BTA-12959-no-rs	21	10922512	**13.78**	9.40	**0.178**	0.109
ARS-BFGL-NGS-101900	21	30314497	**14.16**	5.32	**0.176**	0.104
ARS-BFGL-NGS-110044	21	30892171	**11.75**	3.37	**0.196**	0.050
ARS-BFGL-NGS-55374	25	28795160	**24.35**	**24.33**	**0.275**	**0.224**
ARS-BFGL-NGS-4066	26	9059815	10.32	**13.50**	0.064	**0.115**
ARS-BFGL-NGS-2464	26	18709176	20.66	**13.07**	0.087	**0.126**
ARS-BFGL-NGS-77668	26	18760372	22.13	**17.22**	0.098	**0.104**
ARS-BFGL-NGS-23064	26	18788121	21.97	**16.74**	0.098	**0.103**
ARS-BFGL-NGS-71584	26	18863914	28.62	**19.66**	0.041	**0.134**
BTA-62184-no-rs	26	20014035	17.96	**11.22**	0.078	**0.083**
BTA-60778-no-rs	26	20090833	21.08	**17.89**	0.100	**0.151**
ARS-BFGL-NGS-25126	26	20165024	14.10	**13.56**	0.049	**0.089**
ARS-BFGL-NGS-116902	26	20191815	14.47	**13.97**	0.063	**0.090**
Hapmap31825-BTA-158647	26	21476707	**13.12**	7.25	**0.174**	0.097
ARS-BFGL-NGS-18603	26	21853286	14.86	**13.52**	0.160	**0.095**
ARS-BFGL-NGS-116481	26	22411701	10.02	**11.08**	0.184	**0.084**
Hapmap24832-BTA-138805	26	22449570	10.14	**11.13**	0.184	**0.084**
BTB-00932332	26	22551770	15.62	**12.20**	0.171	**0.140**
ARS-BFGL-NGS-107403	26	23470277	18.20	**12.99**	0.138	**0.132**
BTB-01622498	28	1436040	3.55	**11.44**	0.130	**0.127**
ARS-BFGL-NGS-43501	28	17344828	14.40	**14.28**	0.099	**0.130**
ARS-BFGL-NGS-118693	28	17493199	10.91	**12.75**	0.071	**0.100**
UA-IFASA-6208	28	27379038	**13.36**	7.50	**0.195**	0.038
BTA-99382-no-rs	28	41568645	**12.46**	12.19	**0.187**	0.074
ln(Fat)						
Hapmap38956-BTA-43309	1	98853038	**13.07**	10.74	**0.212**	0.206
BTB-01562245	1	113644592	**11.37**	9.54	**0.306**	0.290
BTA-104132-no-rs	1	113669783	**12.14**	9.85	**0.311**	0.293
ARS-BFGL-BAC-7205	1	120983738	**12.75**	9.11	**0.165**	0.117
ARS-BFGL-NGS-88388	2	13718482	**20.06**	16.60	**0.162**	0.130
ARS-BFGL-NGS-112315	2	40959609	**12.90**	3.87	**0.194**	0.128
ARS-BFGL-NGS-44416	2	48557519	**11.37**	7.58	**0.185**	0.121
BTB-00108243	2	112019423	**17.58**	**22.93**	**0.290**	**0.285**
BTB-01678000	3	6985014	**20.47**	18.76	**0.166**	0.121
BTB-01678060	3	7009487	**19.85**	**20.49**	**0.196**	**0.160**
ARS-BFGL-NGS-5956	3	8558755	**18.50**	**14.08**	**0.195**	**0.147**
ARS-BFGL-NGS-112616	3	8598511	**16.94**	**12.10**	**0.180**	**0.133**
Hapmap50814-BTA-89905	3	41486200	**11.61**	7.40	**0.197**	0.120
INRA-648	3	54347354	**13.24**	**12.30**	**0.309**	**0.288**
Hapmap43441-BTA-103289	3	61621627	17.33	**21.78**	0.141	**0.137**
BTA-68164-no-rs	3	68215262	**12.86**	6.92	**0.199**	0.110
BTB-00131847	3	68241075	**12.93**	7.45	**0.203**	0.119

BTA-54952-no-rs	4	11830564	**11.77**	9.97	**0.184**	0.139
ARS-BFGL-NGS-20815	4	15021946	**16.20**	14.19	**0.171**	0.121
ARS-BFGL-NGS-117196	4	53206872	9.67	**11.42**	0.150	**0.143**
BTB-01252613	5	1851325	**10.90**	7.03	**0.182**	0.150
Hapmap39895-BTA-15668	5	15392995	**10.89**	6.92	**0.189**	0.131
ARS-BFGL-NGS-108617	5	98082173	**15.13**	**11.12**	**0.224**	**0.171**
ARS-BFGL-NGS-95906	5	100351926	**20.56**	**16.92**	**0.192**	**0.133**
Hapmap53294-rs29016908	5	101090418	**39.46**	**25.68**	**0.178**	**0.153**
Hapmap60021-ss46526426	5	101979581	**21.63**	**16.47**	**0.200**	**0.142**
BTB-00270281	6	95770023	**16.41**	9.91	**0.201**	0.123
Hapmap58150-rs29020620	6	96724594	**11.60**	5.12	**0.163**	0.100
BTB-01700063	6	99086447	**21.30**	14.69	**0.237**	0.153
Hapmap53916-rs29021982	6	99581269	**18.02**	**13.96**	**0.196**	**0.150**
Hapmap48078-BTA-77495	6	99827767	**13.14**	12.88	**0.158**	0.116
ARS-BFGL-NGS-14880	7	53879989	**14.54**	13.67	**0.240**	0.155
ARS-BFGL-NGS-113181	7	62800839	**17.30**	17.77	**0.161**	0.122
BTB-02035459	7	63196194	**11.49**	7.58	**0.152**	0.102
BTB-01219396	7	63221359	**11.49**	7.58	**0.152**	0.102
BTA-12616-no-rs	7	64712171	**12.78**	8.69	**0.154**	0.100
ARS-BFGL-NGS-65419	7	66102696	**16.31**	**13.08**	**0.191**	**0.130**
ARS-BFGL-NGS-12863	7	68960712	**16.46**	7.71	**0.215**	0.135
ARS-BFGL-NGS-23091	7	69342623	**13.36**	6.96	**0.227**	0.136
BTB-01222854	7	74587278	**13.33**	9.64	**0.218**	0.151
BTA-23130-no-rs	7	76690557	**11.05**	5.66	**0.174**	0.096
BTB-01321253	7	83625073	**20.03**	20.87	**0.175**	0.126
Hapmap44053-BTA-28733	8	6369477	**15.72**	**11.82**	**0.201**	**0.183**
BTB-01184997	8	36171902	**11.71**	9.07	**0.205**	0.141
Hapmap31805-BTA-154485	8	36232703	**12.94**	9.44	**0.162**	0.110
Hapmap41758-BTA-116042	8	37197678	**11.21**	6.40	**0.224**	0.160
BTB-00384442	9	22407445	**10.97**	7.61	**0.198**	0.127
BTA-83317-no-rs	9	40393986	**18.23**	14.82	**0.183**	0.131
BTA-83528-no-rs	9	41691114	**18.04**	**15.22**	**0.286**	**0.207**
ARS-BFGL-NGS-52530	9	44230587	**13.85**	10.44	**0.238**	0.161
Hapmap34441-BES9_Contig154_536	9	73154820	**12.06**	**11.10**	**0.175**	**0.151**
Hapmap50263-BTA-122214	10	70455224	**11.11**	9.21	**0.271**	0.198
Hapmap30370-BTA-99862	11	14499391	**13.90**	7.21	**0.213**	0.170
Hapmap33349-BTA-127624	12	47348606	10.61	**11.04**	0.176	**0.137**
Hapmap29758-BTC-003619	14	5261	**39.84**	**35.22**	**0.234**	**0.181**
Hapmap30381-BTC-005750	14	50873	**59.74**	**58.91**	**0.274**	**0.229**
Hapmap30383-BTC-005848	14	76704	**64.76**	**61.82**	**0.276**	**0.224**
BTA-34956-no-rs	14	101474	**29.60**	**33.07**	**0.223**	**0.179**
ARS-BFGL-NGS-57820	14	236533	**146.22**	**135.30**	**0.407**	**0.327**
ARS-BFGL-NGS-34135	14	260342	**78.42**	**79.35**	**0.307**	**0.244**
ARS-BFGL-NGS-94706	14	281534	**74.85**	**76.21**	**0.300**	**0.240**
ARS-BFGL-NGS-4939	14	443936	**157.26**	**144.90**	**0.430**	**0.343**
ARS-BFGL-NGS-71749	14	596340	**41.92**	**40.96**	**0.306**	**0.260**
ARS-BFGL-NGS-107379	14	679601	**100.53**	**86.20**	**0.339**	**0.263**
ARS-BFGL-NGS-18365	14	741868	**46.78**	**38.21**	**0.176**	**0.159**
Hapmap30922-BTC-002021	14	763332	**49.33**	**41.41**	**0.181**	**0.175**
Hapmap25384-BTC-001997	14	835055	**64.12**	**62.10**	**0.214**	**0.208**
Hapmap24715-BTC-001973	14	856890	**61.25**	**60.82**	**0.219**	**0.210**
BTA-35941-no-rs	14	894253	**97.77**	**85.15**	**0.324**	**0.255**
ARS-BFGL-NGS-101653	14	931163	**41.53**	**38.61**	**0.279**	**0.240**
ARS-BFGL-NGS-26520	14	996983	**69.06**	**65.29**	**0.313**	**0.274**
UA-IFASA-6878	14	1044040	**35.95**	29.85	**0.152**	0.115
ARS-BFGL-NGS-22866	14	1131951	**44.65**	**45.92**	**0.266**	**0.201**
ARS-BFGL-NGS-103064	14	1193335	**62.36**	**64.75**	**0.229**	**0.242**
ARS-BFGL-NGS-3122	14	1264232	**34.28**	**34.89**	**0.190**	**0.220**
Hapmap25486-BTC-072553	14	1285036	**34.91**	**40.10**	**0.227**	**0.221**

Hapmap30646-BTC-002054	14	1461084	**68.84**	**57.75**	**0.299**	**0.242**
Hapmap30086-BTC-002066	14	1490177	**110.42**	**94.52**	**0.310**	**0.255**
Hapmap30374-BTC-002159	14	1546590	**89.68**	**76.49**	**0.320**	**0.243**
ARS-BFGL-NGS-74378	14	1889209	**48.22**	**36.95**	**0.205**	**0.172**
ARS-BFGL-NGS-117542	14	1913107	**28.04**	**26.03**	**0.215**	**0.189**
UA-IFASA-9288	14	2201869	**48.16**	**32.53**	**0.183**	**0.167**
Hapmap24777-BTC-064977	14	2261622	**13.82**	**12.31**	**0.166**	**0.163**
Hapmap32970-BTC-064990	14	2288509	**32.04**	**27.00**	**0.167**	**0.164**
Hapmap24986-BTC-065021	14	2313594	**32.04**	**27.00**	**0.167**	**0.164**
ARS-BFGL-NGS-22111	14	2347218	18.98	**18.88**	0.122	**0.140**
UA-IFASA-7269	14	2370255	18.98	**18.88**	0.122	**0.140**
Hapmap26072-BTC-065132	14	2391825	23.93	**24.54**	0.144	**0.156**
ARS-BFGL-NGS-113575	14	2484498	**34.94**	**31.97**	**0.178**	**0.164**
ARS-BFGL-NGS-118081	14	2511264	**41.44**	**40.47**	**0.238**	**0.199**
ARS-BFGL-NGS-56327	14	2580413	**60.00**	**53.41**	**0.219**	**0.199**
ARS-BFGL-NGS-100480	14	2607582	**74.83**	**67.79**	**0.253**	**0.223**
UA-IFASA-5306	14	2711614	**47.61**	**32.91**	**0.204**	**0.185**
Hapmap27703-BTC-053907	14	2826072	26.82	**24.14**	0.131	**0.133**
Hapmap22692-BTC-068210	14	3018725	**36.74**	**24.94**	**0.167**	**0.142**
Hapmap23302-BTC-052123	14	3099634	**35.66**	**21.05**	**0.186**	**0.154**
UA-IFASA-6329	14	3465238	**25.62**	**19.44**	**0.200**	**0.159**
ARS-BFGL-NGS-3571	14	3587017	**26.26**	**20.08**	**0.208**	**0.163**
ARS-BFGL-NGS-110563	14	3799229	**25.21**	**17.13**	**0.216**	**0.161**
Hapmap32262-BTC-066621	14	3834070	**13.02**	8.96	**0.185**	0.139
ARS-BFGL-NGS-115947	14	3865963	**33.68**	**21.03**	**0.225**	**0.177**
Hapmap51646-BTA-86764	14	4302230	23.90	**16.80**	0.128	**0.129**
Hapmap30988-BTC-056315	14	4693900	**20.21**	**18.28**	**0.205**	**0.158**
ARS-BFGL-NGS-110894	14	5282437	**14.60**	9.05	**0.197**	0.149
UA-IFASA-6647	14	5808643	**21.04**	13.97	**0.152**	0.111
ARS-BFGL-NGS-102953	14	5867265	**18.86**	10.66	**0.157**	0.114
ARS-BFGL-NGS-16622	15	64781292	**11.60**	9.80	**0.191**	0.153
Hapmap54310-rs29012181	16	8166691	9.54	**10.99**	0.195	**0.138**
ARS-BFGL-NGS-56645	16	23920210	**11.94**	8.17	**0.172**	0.098
ARS-BFGL-NGS-26559	16	33367687	**11.84**	**12.32**	**0.187**	**0.148**
BTA-16056-no-rs	17	20423826	11.69	**11.86**	0.368	**0.345**
ARS-BFGL-NGS-91287	18	10052638	**17.05**	**15.68**	**0.247**	**0.154**
ARS-BFGL-NGS-111247	19	43146804	**14.75**	12.96	**0.158**	0.109
ARS-BFGL-NGS-24479	19	45901285	**11.53**	6.99	**0.188**	0.114
ARS-BFGL-NGS-113693	19	45926259	**12.69**	7.75	**0.188**	0.114
Hapmap48608-BTA-111028	20	52535573	12.23	**16.55**	0.159	**0.144**
BTA-12959-no-rs	21	10922512	**12.62**	9.98	**0.154**	0.108
ARS-BFGL-NGS-101900	21	30314497	**12.91**	7.10	**0.153**	0.096
ARS-BFGL-NGS-55374	25	28795160	**14.21**	10.68	**0.238**	0.161
ARS-BFGL-NGS-39397	26	21166268	**13.43**	**12.13**	**0.163**	**0.147**
Hapmap46411-BTA-15820	26	21404446	**14.57**	**12.21**	**0.197**	**0.179**
Hapmap31825-BTA-158647	26	21476707	**15.83**	**12.66**	**0.151**	**0.143**
ARS-BFGL-NGS-110077	26	21729361	**14.14**	**16.15**	**0.180**	**0.149**
ARS-BFGL-NGS-116481	26	22411701	**22.85**	**23.56**	**0.159**	**0.133**
Hapmap24832-BTA-138805	26	22449570	**23.76**	**24.42**	**0.159**	**0.133**
ARS-BFGL-NGS-6259	26	22492302	**21.29**	**22.06**	**0.159**	**0.133**
ARS-BFGL-NGS-1092	26	24837303	**16.65**	**19.74**	**0.195**	**0.146**
UA-IFASA-4715	26	25330026	16.09	**18.89**	0.128	**0.121**
ARS-BFGL-NGS-38386	26	32836475	7.32	**11.01**	0.134	**0.157**
ARS-BFGL-NGS-105944	26	34196569	**11.54**	8.78	**0.221**	0.146
ARS-BFGL-NGS-53731	26	37801879	13.47	**15.29**	0.131	**0.140**
UA-IFASA-6208	28	27379038	**12.29**	8.03	**0.169**	0.117
ln(Milk)						
ARS-BFGL-BAC-13578	1	121811393	**14.00**	**12.21**	**0.197**	**0.168**

ARS-BFGL-NGS-86079	2	19126180	**13.56**	7.17	**0.203**	0.172
Hapmap53232-rs29020795	2	19202356	**12.37**	8.78	**0.177**	0.149
Hapmap60669-rs29018484	2	20687353	**14.57**	10.61	**0.346**	0.238
ARS-BFGL-NGS-75548	4	78246158	**15.29**	8.67	**0.462**	0.309
ARS-BFGL-NGS-115922	5	29921450	10.99	**11.43**	0.278	**0.228**
BTA-75680-no-rs	6	31470687	**11.54**	6.24	**0.161**	0.101
Hapmap54442-rs29025673	6	31579806	**12.50**	7.42	**0.161**	0.101
UA-IFASA-2111	6	85289127	**14.17**	11.85	**0.195**	0.119
Hapmap25708-BTC-043671	6	88263655	**11.20**	10.56	**0.215**	0.109
Hapmap40845-BTA-97263	6	92788189	**11.64**	5.88	**0.203**	0.125
BTB-01428914	6	93850918	**12.73**	7.77	**0.159**	0.092
Hapmap53417-rs29014877	7	106039868	**12.83**	10.48	**0.168**	0.106
ARS-BFGL-NGS-4062	8	5505897	7.21	**11.81**	0.207	**0.159**
Hapmap31053-BTA-111664	9	27739862	**14.86**	10.09	**0.197**	0.139
BTA-59878-no-rs	9	44095822	**11.31**	8.53	**0.165**	0.142
ARS-BFGL-NGS-52530	9	44230587	**12.96**	**12.03**	**0.285**	**0.201**
ARS-BFGL-NGS-38561	9	45373250	8.38	**11.34**	0.140	**0.131**
BTA-10828-no-rs	9	46600974	**16.48**	**17.49**	**0.169**	**0.157**
Hapmap24524-BTA-107865	9	47934344	8.70	**13.36**	0.126	**0.120**
ARS-BFGL-NGS-62628	9	82136926	10.01	**11.28**	0.157	**0.139**
ARS-BFGL-NGS-78549	11	106534689	**12.86**	**11.55**	**0.250**	**0.185**
ARS-BFGL-BAC-12483	13	1310816	14.05	**15.25**	0.157	**0.156**
BTA-15911-no-rs	13	1371524	12.94	**12.38**	0.170	**0.146**
ARS-BFGL-NGS-56157	13	63208626	**13.37**	5.20	**0.172**	0.079
Hapmap54034-rs29026486	13	65183781	**14.01**	4.09	**0.192**	0.076
ARS-BFGL-NGS-103635	13	67816926	**12.80**	4.96	**0.163**	0.076
Hapmap29758-BTC-003619	14	5261	**42.73**	**42.44**	**0.184**	**0.139**
Hapmap30381-BTC-005750	14	50873	**47.35**	**38.59**	**0.295**	**0.235**
Hapmap30383-BTC-005848	14	76704	**132.62**	**116.70**	**0.386**	**0.277**
BTA-34956-no-rs	14	101474	**69.69**	**49.64**	**0.300**	**0.231**
ARS-BFGL-NGS-94706	14	281534	**110.80**	**85.22**	**0.326**	**0.241**
ARS-BFGL-NGS-107379	14	679601	**170.53**	**162.75**	**0.402**	**0.299**
ARS-BFGL-NGS-18365	14	741868	**38.59**	45.89	**0.204**	0.112
Hapmap30922-BTC-002021	14	763332	**28.82**	33.10	**0.188**	0.096
Hapmap25384-BTC-001997	14	835055	**86.77**	**65.63**	**0.304**	**0.183**
Hapmap24715-BTC-001973	14	856890	**77.38**	**57.27**	**0.299**	**0.183**
BTA-35941-no-rs	14	894253	**84.21**	**90.38**	**0.255**	**0.194**
ARS-BFGL-NGS-101653	14	931163	**49.12**	**42.64**	**0.238**	**0.168**
ARS-BFGL-NGS-26520	14	996983	**56.71**	**48.24**	**0.218**	**0.139**
UA-IFASA-6878	14	1044040	**85.46**	**86.53**	**0.265**	**0.190**
ARS-BFGL-NGS-22866	14	1131951	**54.37**	**48.14**	**0.204**	**0.189**
ARS-BFGL-NGS-3122	14	1264232	**55.37**	**35.73**	**0.272**	**0.131**
Hapmap25486-BTC-072553	14	1285036	**53.65**	**31.62**	**0.277**	**0.165**
Hapmap30646-BTC-002054	14	1461084	**63.96**	**69.53**	**0.223**	**0.160**
Hapmap30086-BTC-002066	14	1490177	**80.04**	**85.10**	**0.253**	**0.182**
Hapmap30374-BTC-002159	14	1546590	**86.76**	**91.59**	**0.248**	**0.196**
ARS-BFGL-NGS-74378	14	1889209	**53.11**	**61.21**	**0.276**	**0.191**
ARS-BFGL-NGS-117542	14	1913107	**37.97**	**36.27**	**0.252**	**0.190**
UA-IFASA-9288	14	2201869	**36.57**	42.92	**0.218**	0.122
Hapmap32970-BTC-064990	14	2288509	**19.61**	18.23	**0.171**	0.079
Hapmap24986-BTC-065021	14	2313594	**19.61**	18.23	**0.171**	0.079
Hapmap26527-BTC-005059	14	2418618	**27.21**	**25.01**	**0.186**	**0.130**
ARS-BFGL-NGS-113575	14	2484498	**15.74**	18.23	**0.178**	0.115
ARS-BFGL-NGS-118081	14	2511264	**27.25**	**25.05**	**0.196**	**0.149**
ARS-BFGL-NGS-56327	14	2580413	**38.75**	**39.00**	**0.233**	**0.156**
ARS-BFGL-NGS-100480	14	2607582	**48.59**	**44.93**	**0.252**	**0.165**
ARS-BFGL-NGS-42263	14	2681399	**24.01**	**17.93**	**0.209**	**0.160**
UA-IFASA-5306	14	2711614	**38.68**	**45.11**	**0.233**	**0.145**
ARS-BFGL-NGS-54400	14	2736946	**21.70**	**15.31**	**0.218**	**0.151**

Hapmap22692-BTC-068210	14	3018725	**36.15**	**43.09**	**0.208**	**0.140**
Hapmap23302-BTC-052123	14	3099634	**38.58**	**51.15**	**0.215**	**0.148**
Hapmap25217-BTC-067767	14	3189311	**36.24**	21.15	**0.195**	0.076
UA-IFASA-6329	14	3465238	**26.01**	28.86	**0.170**	0.106
ARS-BFGL-NGS-56339	14	3498808	**16.13**	**20.80**	**0.166**	**0.134**
UA-IFASA-8927	14	3640095	**18.63**	16.26	**0.157**	0.112
Hapmap30091-BTC-005211	14	3940999	**27.46**	21.04	**0.236**	0.119
ARS-BFGL-BAC-24839	14	3993201	**22.93**	22.70	**0.173**	0.106
ARS-BFGL-NGS-112858	14	4956374	**27.34**	35.86	**0.208**	0.133
Hapmap51078-BTA-87682	14	5064062	**16.06**	12.13	**0.202**	0.110
ARS-BFGL-NGS-55227	14	5085415	**21.02**	33.40	**0.171**	0.130
Hapmap32236-BTC-049785	14	5139497	20.69	**34.26**	0.149	**0.126**
ARS-BFGL-BAC-20965	14	5225005	**23.51**	**18.01**	**0.255**	**0.157**
Hapmap33635-BTC-049051	14	5318260	**15.80**	5.33	**0.197**	0.092
Hapmap27091-BTC-048823	14	5356987	**30.86**	26.34	**0.208**	0.098
Hapmap23851-BTC-048718	14	5387835	**29.32**	22.85	**0.244**	0.113
Hapmap32234-BTC-048199	14	5640337	**33.67**	**36.21**	**0.217**	**0.131**
Hapmap26283-BTC-048098	14	5696728	**16.56**	27.75	**0.158**	0.107
Hapmap25716-BTC-047850	14	5937549	**17.19**	**17.52**	**0.208**	**0.123**
Hapmap23799-BTC-047701	14	6044245	**12.10**	11.67	**0.222**	0.108
ARS-BFGL-BAC-8730	14	6252100	**31.45**	21.44	**0.194**	0.106
Hapmap53312-rs29018332	14	60576872	**20.04**	6.97	**0.166**	0.070
Hapmap43128-BTA-105550	15	52541506	**12.39**	6.76	**0.160**	0.088
Hapmap59019-rs29021918	18	41938135	10.34	**16.90**	0.128	**0.148**
Hapmap39811-BTA-122745	20	35432864	10.75	**11.29**	0.170	**0.145**
BTA-50235-no-rs	20	35883921	7.85	**11.52**	0.237	**0.294**
BTA-50402-no-rs	20	36667999	8.05	**14.59**	0.152	**0.160**
Hapmap26466-BTA-160199	20	36746234	8.46	**13.28**	0.163	**0.166**
BTA-50376-no-rs	20	36915967	**15.91**	**20.90**	**0.168**	**0.137**
Hapmap57531-rs29013890	20	36955574	**13.03**	**14.21**	**0.190**	**0.148**
ARS-BFGL-NGS-37182	22	5301596	**11.30**	10.22	**0.282**	0.273
Hapmap43294-BTA-56514	23	32759583	**11.76**	7.34	**0.311**	0.161
ARS-BFGL-NGS-55374	25	28795160	**12.33**	**11.61**	**0.285**	**0.201**
ARS-BFGL-NGS-2127	26	13563198	**13.62**	10.34	**0.163**	0.130
ARS-BFGL-NGS-2464	26	18709176	**12.92**	**10.90**	**0.154**	**0.149**
ARS-BFGL-NGS-77668	26	18760372	19.84	**21.11**	0.139	**0.154**
ARS-BFGL-NGS-23064	26	18788121	19.17	**19.93**	0.139	**0.154**
ARS-BFGL-NGS-71584	26	18863914	19.57	**17.19**	0.119	**0.144**
BTB-00930720	26	21323659	**12.52**	**11.21**	**0.154**	**0.168**
Hapmap31825-BTA-158647	26	21476707	**13.93**	**11.98**	**0.165**	**0.154**
BTB-00931481	26	21631982	21.92	**19.19**	0.165	**0.170**
ARS-BFGL-NGS-18603	26	21853286	14.66	**16.99**	0.145	**0.147**
ARS-BFGL-NGS-114149	26	22137070	5.96	**10.88**	0.109	**0.137**
BTB-00932332	26	22551770	15.99	**15.53**	0.160	**0.184**
ARS-BFGL-NGS-107403	26	23470277	18.82	**18.04**	0.171	**0.194**
BTA-60935-no-rs	26	23985824	17.32	**15.01**	0.142	**0.146**
ARS-BFGL-NGS-119314	26	25634039	12.11	**14.59**	0.136	**0.157**
BTB-00935537	26	26325360	17.75	**14.89**	0.102	**0.135**
Hapmap28763-BTA-162328	26	26472420	9.67	**11.44**	0.111	**0.129**
ARS-BFGL-NGS-109460	27	46280579	**13.48**	2.64	**0.174**	0.083
BTB-02080610	28	19938671	**12.76**	10.74	**0.206**	0.146
Hapmap58649-rs29011010	28	28185115	**15.98**	8.69	**0.162**	0.125

CHAPTER FOUR

Genome-wide Association Analysis to Identify Gene by Environment Interaction for Milk Protein Yield and Level of Somatic Cell Score as Environmental Descriptor in German Holsteins

MELANIE STREIT[*], FRIEDRICH REINHARDT[†], GEORG THALLER[‡], and JÖRN BENNEWITZ[*]

[*]Institute of Animal Husbandry and Breeding,
University of Hohenheim,
Stuttgart, Germany
[†]Vereinigte Informationssysteme Tierhaltung w.V. (VIT),
Verden, Germany
[‡]Institute of Animal Breeding and Husbandry,
Christian-Albrechts-University,
Kiel, Germany

published in:
Journal of Dairy Science (2013) 96: 7318 – 7324
http://dx.doi.org/10.3168/jds.2013-7133

GxE WITH SCS AS ENVIRONMENTAL DESCRIPTOR

Genome-wide Association Analysis to Identify Gene by Environment Interaction for Milk Protein Yield and Level of Somatic Cell Score as Environmental Descriptor in German Holsteins

M. Streit[*], F. Reinhardt[†], G. Thaller[‡], and J. Bennewitz[*]

[*]Institute of Animal Husbandry and Breeding, University of Hohenheim, Stuttgart, Germany
[†]Vereinigte Informationssysteme Tierhaltung w.V. (vit), Verden, Germany
[‡]Institute of Animal Breeding and Husbandry, Christian-Albrechts-University, Kiel, Germany

Corresponding author:
Jörn Bennewitz
Garbenstraße 17
70599 Hohenheim
Germany
phone: +49 711-459 23570
fax: +49 711-459 23101
E-mail: j.bennewitz@uni-hohenheim.de

INTERPRETIVE SUMMARY

Genotype-by-environment interaction was frequently studied in dairy cattle breeding using reaction norm models. Often the environment is modelled by average herd milk production levels, but the level of udder health and hygiene might also be an important environmental factor. In this article we report the results from a large scale genome-wide association analysis for the intercept and slope of milk protein reaction norms when using the average herd test day solution for somatic cell score as an environmental descriptor. We were able to detect and confirm several SNP cluster affecting intercept and slope. The results may have implications for breeding robust cows.

ABSTRACT

Genotype-by-environment interaction (GxE) has been widely reported in dairy cattle. If the environment can be measured on a continuous scale reaction norms can be applied to study GxE. The average herd milk production level was frequently used as an environmental descriptor, which is mainly influenced by the level of feeding or the feeding regime. Another important environmental factor is the level of udder health and hygiene, for which the average herd level of somatic cell count might be a descriptor. In the present study we conducted a genome-wide association analysis to identify SNPs that affect intercept and slope of milk protein yield reaction norms when using the average herd test day solution for somatic cell score as an environmental descriptor. Sire estimates for intercept and slope of the reaction norms were calculated from around 12 million daughter records, using linear reaction norm models. Sires were genotyped for 54k SNPs. The sire estimates were used as observations in the association analysis, using 1,797 sires. Significant SNPs were confirmed in an independent validation set consisting of 500 sires. A known major gene affecting protein yield was included as a covariable in the statistical model. A number of 60 (21) SNPs was confirmed for intercept with a $p \leq 0.01$ ($p \leq 0.001$) in the validation set. These figures were 28 and 11 for slope. Most, but not all, SNPs affecting slope affected also intercept. A comparison with an earlier study revealed that SNPs affecting slope where in general also significant for slope when the environment was modelled by the average herd milk production level, although both environmental descriptors were only low correlated.

Key Words : genotype-by-environment interaction, somatic cell score, association analysis, reaction norm

INTRODUCTION

The term genotype x environment interaction (GxE) refers to differences in response of genotypes to changes in the environment (Lynch and Walsh, 1998). The environment where dairy farming is practised varies considerably and consequently GxE were investigated in several studies, see e.g. König et al. (2005) and Strandberg et al. (2009) and references therein. If the environment can be measured on a continuous scale, reaction norm models are often used to study GxE (e.g. Kolmodin et al., 2002; Hayes et al., 2003; Strandberg et al., 2009). In a reaction norm model, the phenotype is modelled as a function of the environment, where the phenotype is produced. The slope of the reaction norm is a measure of the environmental sensitivity. Individuals with steep (flat) slope are called environmental sensitive (robust) individuals. A non-zero variance of the slope indicates the presence of GxE.

The choice of an appropriate environmental descriptor is of fundamental importance. Frequently, the average herd production level of the trait under consideration is used as an environmental descriptor, because it combines many unobservable environmental factors affecting the phenotype. In addition, this can be estimated with high precision for many dairy herds, provided that herd size is not too small. Probably the most important environmental factor is the level of feeding, which is captured by affecting average herd milk production (Hayes et al., 2003).

Another important environmental factor for dairy cattle is the level of hygiene and of udder health. This is not recorded routinely in Germany, although there is a trend towards implementing special recording schemes on contract herds. However, the level of somatic cell count (SCC) is routinely collected on many dairy farms for management and breeding purposes. Barkema et al. (1999) and, in a recent review, Dufour et al. (2011) pointed out significant relationships of udder health management practise and herd somatic cell count (SCC). Because somatic cell score (SCS) is included in the panel of traits for which routinely genetic evaluations are performed in Germany, reliable average herd SCS levels are available that can be used as environmental descriptor in GxE analysis.

Considering environmental sensitivity or, equivalently, robustness is an increasing issue in dairy cattle breeding (Veerkamp et al., 2009; Hayes et al., 2009). One way to do this is to apply marker assisted or genomic selection. In both breeding schemes some knowledge of the genes affecting environmental sensitivity is needed. Therefore, Lillehammer et al. (2009) and

Hayes et al. (2009) extended classical genome wide association analysis (GWAS) towards considering GxE effects. They estimated for each SNP an effect for intercept and for slope and found environmental sensitive and robust SNPs. The latter one were significant for the intercept but not for the slope. In addition, Lillehammer et al. (2009) mapped a higher number of QTL if GxE was taken into account. Hence, although the model becomes more complex and additional parameters have to be estimated, considering GxE might result in an increase in statistical power to map QTL.

In a previous study we applied higher order reaction norm random regression models and found highly significant GxE effects in German Holsteins for both environmental descriptors average herd milk production level and average herd SCS (Streit et al., 2012). Both descriptors were slightly negatively correlated (-0.18). In a subsequent study we conducted a GWAS to identify SNPs for milk production traits affecting slope and intercept of the reaction norms in German Holsteins (Streit et al., 2013). We used the average herd milk energy yield as a continuous environmental descriptor and applied similar statistical models proposed by Lillehammer et al. (2009) and Hayes et al. (2009). Numerous SNPs could be identified that were significantly associated with intercept and slope.

The aim of the present study was to conduct a large scale GWAS for milk protein yield with the average herd somatic cell score as an environmental descriptor. We applied a three-step procedure. In the first step, estimates for intercept and slope of sire reaction norms were calculated using first-order random regression sire models. These estimates were used in a second step as observations in an association analysis. In the third step, significant SNP associations were confirmed in an independent validation set of the same population.

MATERIALS AND METHODS

Data and data editing

In total 2,356 progeny tested German Holstein sires were genotyped with the Illumina BovineSNP50 BeadChip (Illumina, San Diego, CA; Matukumalli et al., 2009). The sires were born between 1983 and 2003. Data filtering was done using PLINK (Purcell et al., 2007) using the following criteria. Individuals with more than 10% missing marker genotypes were removed (59 individuals). An SNP was excluded if it had a minor allele frequency less than 3%, a call rate less than 90%, a significant deviation from the Hardy-Weinberg-equilibrium (p<0.001), or if the position on the genome was unknown. SNPs on the sex chromosome were

also excluded. A total of 41,349 SNPs remained in the data set. Sporadic missing genotypes were imputed using fastPHASE (Scheet and Stephens, 2006). The average linkage disequilibrium between pairwise SNPs at distances of < 25 kb was $r^2 = 0.3$ (Qanbari et al., 2010).

Around 12 million first lactation test day records for milk protein yield from daughters of the sires were used. The number of daughters per bull ranged from 50 to 74,842 and totalled around 1.3 million. Test day records were corrected for the fixed effects herd test day, days in milk, age at calving, calving season and the random permanent environment effect. These correction factors were obtained from the routine animal genetic evaluation. Only daughters with at least seven observations per year were considered. The environmental descriptors were herd test day solutions for somatic cell score, which was obtained from routine animal evaluation. Observations in extreme and rare environments were discarded (around two percent of the observations). These restrictions ensured that there were enough observations and variation of the environmental descriptor to apply reaction norms within individual cows and that the results were not affected by (unreliable) extreme and rare environments. The environmental descriptor was rescaled to have a mean of zero and a standard deviation of one.

Statistical analysis

The following reaction norm random regression model was applied in the first step:

$$cy_{ijk} = \mu + b * htdsscs_k + \sum_{m=0}^{1} s_{jm} * htdsscs_k^m + \sum_{m=0}^{1} d_{ijm} * htdsscs_k^m + e_{ijk}, \quad (1)$$

where cy_{ijk} is the corrected protein yield of daughter i of sire j at herd test day k, μ is the overall mean, $htdsscs_k$ is the herd test day solution for SCS at herd test day k with the fixed regression coefficient b, s_{jm} is the random sire effect of sire j of order m, d_{ijm} the random daughter effect of daughter i of sire j of order m, and e is the random residual. The covariance structure of the sire regression effects is $Var\begin{bmatrix} s_0 \\ s_1 \end{bmatrix} = A \otimes \begin{bmatrix} \sigma^2_{s_0} & \sigma_{s_0 s_1} \\ \sigma_{s_0 s_1} & \sigma^2_{s_1} \end{bmatrix}$, and of the daughter effects is $Var\begin{bmatrix} d_0 \\ d_1 \end{bmatrix} = I \otimes \begin{bmatrix} \sigma^2_{d_0} & \sigma_{d_0 d_1} \\ \sigma_{d_0 d_1} & \sigma^2_{d_1} \end{bmatrix}$, with A (I) being the numerator relationship (identity) matrix. The model was fitted using ASReml 3.0 (Gilmour et al., 2009). The estimated sire effects for slope and intercept of the protein yield reaction norms (estimated in model 1) were treated as two different traits and were used as observations in association analysis (model 2).

The whole data set was randomly split into a discovery data set (n = 1,797 bulls) and a validation data set (n = 500 bulls). In the second step of the statistical analysis, we performed genome-wide association analyses using the discovery data set. It is known that *DGAT1* segregates in this population and shows a substantial effect also for protein yield (Thaller et al., 2003). The *DGAT1* K232A-substitution is not included in the Illumina BovineSNP50 BeadChip, but the SNP ARS-BFGL-NGS-4939 is in nearly complete linkage disequilibrium with this substitution in the German Holsteins (r^2 = 0.998, Wang et al., 2012). Therefore, we included this marker as a covariate in the statistical model for the association analysis. The following mixed linear model was applied for each marker in turn:

$$\hat{s}_{jt} = \mu_t + b_t * z_j + sire_{jt} + b_{kt} * x_{jk} + e_{jt}, \quad (2)$$

where $\hat{s}_{jt}$ is the estimated sire effect for trait t (t being intercept and slope, respectively). The term z denotes the number of copies of the allele with the higher frequency of SNP ARS-BFGL-NGS-4939 (z = 0, 1, or 2) and b_t is the regression coefficient. The effect of each SNP k was modelled similarly as a regression on the number of copies of the allele with the higher frequency (x = 0, 1, or 2), with b_{kt} being the regression coefficient. In order to control the population structure, we fitted a random sire effect with the covariance structure $A\sigma^2_{st}$, where σ^2_{st} is a variance attributable to the sires. This model was applied for each SNP k in turn, resulting in 41,348 association tests per trait. We declared each SNP with a pointwise error probability below p<0.001 as significant. In order to judge how many false positives were among the significant associations we applied the false discovery rate (FDR) technique. We calculated for each association test an FDR q-value using the software QVALUE (Storey and Tibshirani, 2003). The FDR q-value of the significant SNP with the lowest test statistic (p≈0.001) provided an estimate of the proportion of false positives among the significant associations.

In the third step, we confirmed significant SNP associations within the same population in the validation set. The same statistical model was applied, but only to significant SNPs. We declared an SNP as confirmed if the p-value in the validation set was either p<0.001 (stringent) or p<0.01 (less stringent significance level) and the signs of the effects were the same in both sets. The less stringent significance criterion was used in addition, because less multiple testing was performed, and the stringent significance level would reduce the power to confirm SNPs. A similar strategy was applied by Pryce et al. (2010).

RESULTS

The estimated variance components are shown in Table 1. The daughter variance components are larger than the sire variance components. The standard errors are small for all estimated components. The correlation between intercept and slope is negative for both the sire and the daughter. Table 2 summarises the results from the association analysis. The FDR analysis revealed that around 8% of the significant associations were false positives for intercept. Less SNPs were significant for slope and the FDR was slightly higher. As expected, for both traits around 60% less SNPs could be confirmed using the stringent validation compared to the less stringent validation.

Table 1. Variance components (σ^2) and correlations (ρ) of the random regression analyses. Standard errors are shown in parentheses.

Variance components and correlations	estimates
$\sigma^2_{s_0}$, $\sigma^2_{s_1}$, $\sigma_{s_0 s_1}$, and $\rho_{s_0 s_1}$	845.80 (25.77), 1.04 (0.16), -11.75 (1.76), and -0.4
$\sigma^2_{d_0}$, $\sigma^2_{d_1}$, $\sigma_{d_0 d_1}$, $\rho_{d_0 d_1}$, and σ^2_e	1746.38 (2.95), 25.36 (1.22), -43.10 (1.26), -0.2, and 5167.07 (2.28)

The plots of the test statistic along the chromosomes are shown in Figure 1. Chromosomal positions of validated SNPs (less stringent validation) are indicated by a triangle symbol. Significant SNPs were found on many chromosomes. Highest significance was observed for all three traits for the SNP being in near complete linkage disequilibrium with *DGAT1* K232A-substitution on BTA14. No other SNP on this chromosome was significant. Promising SNP clusters affecting intercept were identified on BTA1, 6, 7, 9, 13, 16, 18, 26, and 28. Not all clusters were also significant for slope, see e.g. the SNP clusters on BTA6, BTA13, and BTA16.

The SNPs that could be validated (stringent validation) for at least one trait are shown in Table 3. The sign of the effects were in the opposite direction for slope and intercept. Again, it is observable that not all SNPs affecting intercept were also significant for slope. In contrast, some SNPs on BTA11, 16, 18, and 21 were validated for slope, but not significant for intercept. These SNPs would have been missed without modelling GxE in the analysis. All

SNPs that could be validated for at least one trait (lees stringent validation) are shown in the Supplemental table.

Table 2. Number of discovered and validated SNPs (stringent and less stringent validation, $p \leq 0.001$ and $p \leq 0.01$, respectively) for intercept(protein yield) and slope(protein yield). The FDR *q*-values (FDR) of the significant SNP with the largest error probability ($p \approx 0.001$) in the discovery dataset are shown.

Trait	Discovery dataset ($p \leq 0.001$)	FDR	Validation dataset ($p \leq 0.01$)	Validation dataset ($p \leq 0.001$)
Intercept(protein yield)	407	0.08	60	21
Slope(protein yield)	261	0.12	28	11

DISCUSSION

In the present study numerous SNPs were identified that affected intercept and slope of sire protein reaction norms when the environment was modelled as average herd test day SCS levels. Compared to the intercept variance, the variance of the slopes was small (Table 1). However, it is important to note that the slope variance depends on the range of the environmental values and a wider range yields a larger slope variance. In a previous study we found this variance, and therefore the presence of GxE with this environmental descriptor, to be highly significant (Streit et al., 2012). The correlation between intercept and slope depends on where the intersection point of the reaction norm model is placed. As recommended by Kolmodin and Bijma (2004), we placed it in the average environment. In this case the intercept estimate can be interpreted as an estimate for average or general production and the slope as an estimate for the environmental sensitivity. The negative correlation between intercept and slope under this condition means that with a decrease in the average herd SCS level (i.e. in an 'improved' environment) the genotype value increases. This trend was frequently observed also with other environmental descriptors, e.g. average herd milk production (Kolmodin et al., 2002; Lillehammer et al., 2009). Hence, with an ongoing selection on general production there will be substantial correlated selection response also for environmental sensitivity.

Many SNPs were identified and confirmed that are involved in intercept and slope. Some interesting SNP or SNP clusters are located next to candidate genes for milk protein segregating in the Germans Holsteins, e.g. on BTA6 (Kühn et al., 1999; Bennewitz et al.,

2004a), BTA5, and BTA20 (Wang et al., 2012). Interestingly, no other SNP on BTA14 next to the one being in nearly complete linkage disequilibrium with *DGAT1* K232A-substitution was significantly associated with intercept and slope. Hence, this SNP explained the whole major QTL for milk traits frequently reported in this population (Thaller et al., 2003; Bennewitz et al., 2004b). The significant SNPs solely for slope (Table 3) demonstrate the advantages of considering GxE. For these SNPs the allele that is superior in one half of the environment is inferior in the second half and both alleles show a similar effect in the average environment and, hence, they are not significant for intercept.

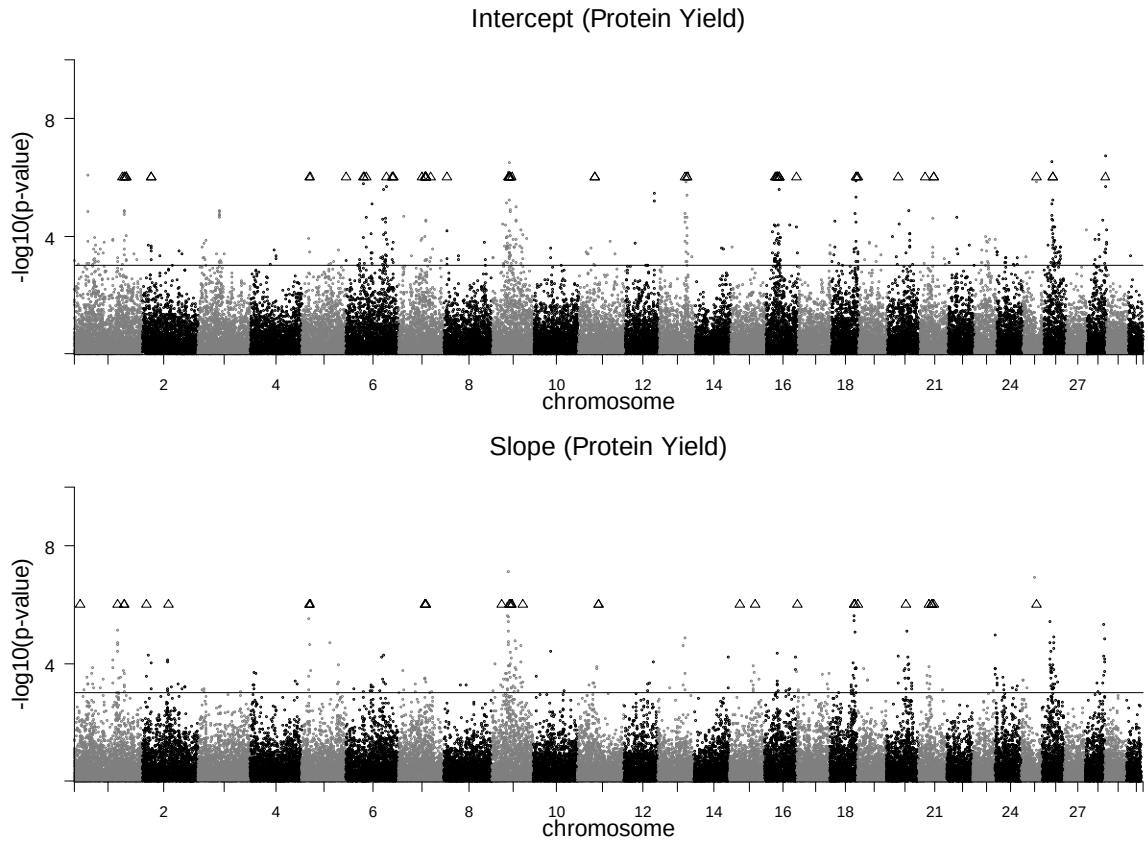

Figure 1. Test statistic profile of SNP effects for intercept and slope in the discovery data set. The nominal significance level (p<0.001) is indicated by a solid line. Positions of validated SNPs (less stringent validation) are indicated by a triangle. The test statistic of SNP ARS-BFGL-NGS-4939 on BTA14 is 12, both times for intercept (protein yield) and slope (protein yield) and is not shown in the figure.

Many SNPs affecting slope in this study affected the slope also when average herd milk production was used as environment (Streit et al., 2013), although the correlation between two environmental descriptors was low and slightly negative. The sign of the effects were almost

always such, that in an 'improved' environment (higher average milk production or lower SCS) the effect of the SNP increased (Table 3 and Streit et al., 2013). Hence, for robustness breeding purposes for both, a fluctuating milk production environment and a fluctuating SCS environment, a similar set of SNPs is important. Exceptions are SNPs on BTA11, 16, 18, and 21 that were significant for slope but not for intercept (Table 3). These were not significant for slope in our previous study, where the environment was modelled as the average herd milk production (Streit et al., 2013). Hence, these SNPs are not involved in general production and are environmental sensitive only to the SCS environment.

Lillehammer et al. (2009) discussed the use of intercept and slope SNPs for breeding of robust animals. They suggested considering SNPs with effects for intercept and slope in opposite directions, i.e. SNPs with a smaller effect in an improved environment, because these SNPs are not in line with the polygenic correlation between intercept and slope. This class of SNPs was not identified in this study. Therefore, if robustness for milk traits in a fluctuating SCS environment is desired, single marker assisted selection can not be recommended. Instead, genomic selection considering important SNPs simultaneously seems to be more promising.

Several environmental descriptors were used in reaction norm models in dairy cattle (Fikse et al., 2003; Strandberg et al., 2009). In this study average herd SCS levels obtained from routine genetic animal evaluation were used as an indicator trait for udder health and hygiene conditions on the farms at the time where the trait yield was recorded. Herd test day observations were used, which offered the possibility to capture also the within cow variance (Hayes et al., 2003). Indeed, the daughter variance was substantial (Table 1). It modelled the additive genetic variance not included in the sire effects and some within cow variability. In addition, it was possible to use the variance of the environmental descriptor within herd among the test days. This might be especially useful for the average herd test day SCS environmental descriptor, because it can be expected that the SCS fluctuation within herds is larger compared to milk traits, e.g. due to temporary herd mastitis infections. Mastitis can be due to different classes of pathogens, e.g. environmental associated like *Escherichia coli* or contagious associated like *Streptococcus agalactiae* and *Sterptococcus aureus*. Mastitis due to the first class of pathogens results in general in a lower somatic cell count compared to the second class (Barkema et al., 1998). Hence, average herd SCS levels as a proxy to describe the herd level of udder health and hygiene have limits, as SCS has as a breeding goal to improve mastitis resistance.

Table 3. SNPs with a successful stringent validation ($p \leq 0.001$ in the validation set) for at least one trait with chromosome (BTA), position in base pairs (bp), F-values and sign of the effct of the allele with the higher frequency for the three traits. The trait for which the validation was successful is indicated in bold type F-values.

SNP	BTA	bp	F-values (sign of effect)			
			Intercept (protein yield)		Slope (protein yield)	
ARS-BFGL-BAC-7205	1	120983738	19.03	(-)	**14.11**	(+)
ARS-BFGL-NGS-99492	1	121607486	**18.91**	(-)	**13.66**	(+)
BTA-114011-no-rs	1	125911737	**11.76**	(+)	1.48	(-)
BTB-00056059	1	126163013	**11.51**	(-)	3.27	(+)
ARS-BFGL-NGS-68464	5	18395406	**12.88**	(-)	**18.06**	(+)
Hapmap33079-BTA-163567	6	1936	**11.58**	(+)	6.07	(-)
BTA-110673-no-rs	6	111383112	**11.36**	(-)	6.73	(+)
BTB-00281303	6	111612203	**14.34**	(-)	7.89	(+)
ARS-BFGL-NGS-113181	7	62800839	**15.09**	(+)	**12.83**	(-)
ARS-BFGL-NGS-113819	7	63609102	**17.37**	(+)	8.77	(-)
ARS-BFGL-NGS-109819	7	63664393	**17.65**	(+)	10.14	(-)
BTB-01880776	7	64095706	**11.68**	(+)	7.98	(-)
ARS-BFGL-NGS-52530	9	44230587	**24.26**	(-)	29.15	(+)
BTB-00391835	9	52160813	**11.14**	(-)	**11.57**	(+)
ARS-BFGL-NGS-113322	11	38523074	8.07	(+)	**14.43**	(-)
BTA-93012-no-rs	11	38544853	8.03	(+)	**14.66**	(-)
ARS-BFGL-NGS-63777	13	67075815	**21.32**	(+)	11.37	(-)
ARS-BFGL-NGS-4939*	14	443936	**78.92**	(+)	**65.82**	(-)
ARS-BFGL-NGS-113877	16	29456497	**12.24**	(-)	10.66	(+)
Hapmap38953-BTA-38562	16	29487613	**12.24**	(-)	10.66	(+)
Hapmap50594-BTA-121054	16	29757246	**11.34**	(+)	4.24	(-)
ARS-BFGL-NGS-26559	16	33367687	**14.47**	(-)	12.04	(+)
ARS-BFGL-NGS-59645	16	73117625	**16.58**	(+)	10.28	(-)
BTB-02013769	16	74080339	5.84	(+)	**14.35**	(-)
ARS-BFGL-NGS-35499	18	50131636	5.47	(+)	**11.14**	(-)
ARS-BFGL-NGS-6001	21	25932661	7.83	(-)	**14.69**	(+)
Hapmap52867-rs29023496	21	26401500	7.02	(-)	**12.11**	(+)
ARS-BFGL-NGS-110044	21	30892171	**12.64**	(+)	7.59	(-)
ARS-BFGL-NGS-55374	25	28795160	**23.44**	(-)	28.25	(+)

* this SNP is in near complete linkage disequilibrium with *DGAT1* K232A-substitution (Wang et al., 2012)

CONCLUSIONS

GxE for protein yield and average herd SCS as environmental descriptor were detected using a first order sire reaction norm model. Many SNP or SNP clusters could be identified that affected intercept and slope of the reaction norm. The number of SNPs affecting intercept was larger. Some significant SNPs affected only slope. Considering GxE improved the statistical power to map SNP involved in protein yield variation. A comparison with an earlier study

revealed that SNPs affecting slope where in general also significant for slope when the environment was modelled by the average herd milk production level, although both environmental descriptors were only low correlated. An across breed analysis in combination with a higher marker density is desired to validate the effects also in another population and to fine map the underlying causal mutation.

ACKNOWLEDGMENTS

MS was supported by a grant from the Deutsche Forschungsgemeinschaft, DFG.

REFERENCES

Barkema, H. W., Y. H. Schukken, T. J. G. M. Lam, M. L. Beiboer, H. Wilmink, G. Benedictus, and A. Brand. 1998. Incidence of Clinical Mastitis in Dairy Herds Grouped in Three Categories by Bulk Milk Somatic Cell Counts. J. Dairy Sci. 81:411–419.

Barkema, H. W., J. D. van der Ploeg, Y. H. Schukken, T. J. G. M. Lam, G. Benedictus, and A. Brand. 1999. Management Style and Its Associations with Bulk Milk Somatic Cell Count and Incidence Rate of Clinical Mastitis. J. Dairy Sci. 82 :1655–1663.

Bennewitz, J., N. Reinsch, V. Guiard, S. Fritz, H. Thomsen, C. Looft, C. Kühn, M. Schwerin, C. Weimann, G. Erhardt, F. Reinhardt, R : Reents, D. Boichard, and E. Kalm. 2004a. Multiple Quantitative Trait Loci Mapping With Cofactors and Application of Alternative Variants of the False Discovery Rate in an Enlarged Granddaughter Design. Genetics 168:1019–1027.

Bennewitz, J., N. Reinsch, S. Paul, C. Looft, B. Kaupe, C. Weimann, G. Erhardt, G. Thaller, C. Kühn, M. Schwerin, H. Thomsen, F. Reinhardt, R. Reents, and E. Kalm. 2004b. The DGAT1 K232A Mutation Is Not Solely Responsible for the Milk Production Quantitative Trait Locus on the Bovine Chromosome 14. J. Dariy Sci. 87:431–442.

Dufour, S., A. Fréchette, H. W. Barkema, A. Mussell, and D. T. Scholl. 2011. Invited review : Effect of udder health management practices on herd somatic cell count. J. Dairy Sci. 94:563–579. doi:10.3168/jds.2010-3715.

Fikse, W. F., R. Rekaya, and K. A. Weigel. 2003. Assessment of environmental descriptors for studying genotype by environment interaction. Livest. Prod. Sci. 82:223–231. doi:10.1016/S0301-6226(03)00009-5.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2009. ASReml User Guide 3.0. VSN International Ltd., Hemel Hempstead, HP1 1ES, UK.

Hayes, B. J., M. Carrick, P. Bowman, and M. E. Goddard. 2003. Genotype x Environment Interaction for Milk Production of Daughters of Australian Dairy Sires from Test-Day Records. J. Dairy Sci. 86:3736–3744.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, K. Savin, C. P. van Tassell, T. S. Sonstegard, M. E. Goddard. 2009. A validated Genome Wide Association Study to Breed Cattle Adapted to an Environment Altered by Climate Change. PLoS ONE: 4:e6676.

König, S., G. Dietl, I. Raeder, and H. H. Swalve. 2005. Genetic Relationships for Dairy Performance Between Large-Scale and Small-Scale Farm Conditions. J. Dairy Sci. 88:4087–4096.

Kolmodin, R. E. Strandberg, P. Madsen, J. Jensen, and H. Jorjani. 2002. Genotype by Environment Interaction in Nordic Dairy Cattle Studied Using Readction Norms. Acta Agric. Scand., Sect. A, Animal Sci. 52:11–24.

Kolmodin, R., and P. Bijma. 2004. Response to mass selection when the genotype by environment interaction is modelled as a linear reaction norm. Genet. Sel. Evol. 36:435–454.

Kühn, C., G. Freyer, R. Weikard, T. Goldammer, and M. Schwerin. 1999. Detection of QTL for milk production traits in cattle by application of a specifically developed marker map of BTA6. Anim. Genet. 30:333–340.

Lillehammer, M., J. Ødegård, and T. H. E. Meuwissen. 2009. Reducing the bias of estimates of genotype by environment interactions in random regression sire models. Genet. Sel. Evol. 41:30. doi:10.1186/1297-9686-41-30.

Lynch, M., and B. Walsh. 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associates Inc Publishers, Sunderland, MA, USA.

Matukumalli, L., K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, M. P. Heaton, J. O'Connell, S. S. Moore, T. P. L. Smith, T. S. Sonstegard, and C. P. van Tassell. 2009. Development and Characterization of a High Densitiy SNP Genotyping Assay for Cattle. PLoS ONE 4:e5350.

Pryce, J. E., S. Bolormaa, A. J. Chamberlain, P. J. Bowman, K. Savin, M. E. Goddard, and B. J. Hayes. 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. J. Dairy Sci. 93:3331–3345. doi:10.3168/jds.2009-2893.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet. 81:559–575.

Qanbari, S., E. C. G. Pimentel, J. Tetens, G. Thaller, P. Lichtner, A. R. Sharifi, and H. Simianer. 2010. The pattern of linkare disequilibrium in German Holstein cattle. Anim. Genet. 41:346 – 356.

Scheet, P., and M. Stephens. 2006. A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data: Applications to Inferring Missing Genotypes and Haplotypic Phase. Am. J. Hum. Genet. 78:629–644.

Storey, J. D., and R. Tibshirani. 2003. Statistical significance for genomewide studies. PNAS 100:9440–9445.

Strandberg, E., S. Brotherstone, E. Wall, and M. P. Coffey. 2009. Genotype by environment interaction for first-lactation female fertility traits in UK dairy cattle. J. Dairy Sci. 92:3437–3446. doi:10.3168/jds.2008-1844.

Streit, M., F. Reinhardt, G. Thaller, and J. Bennewitz. 2012. Reaction norms and genotype-by-environment interaction in the German Holstein dairy cattle. J. Anim. Breed. Genet. 129:380–389. doi:10.1111/j.1439-0388.2012.00999.x.

Streit, M., R. Wellmann, F. Reindardt, G. Thaller, H. P. Piepho, and J. Bennewitz. 2013. Using genome-wide association analysis to characterize environmental sensitivity of milk traits in dairy cattle. G3. doi: 10.1534/g3.113.006536.

Thaller, G., W. Krämer, A. Winter, B. Kaupe, G. Erhardt, and R. Fries. 2003. Effects of DGAT1 variants on milk production traits in German cattle breeds. J. Anim. Sci. 81:1911–1918.

Veerkamp, R. F., H. A. Mulder, M. P. L. Calus, J. J. Windig, and J. ten Napel. 2009. Statistical genetics to improve robustness of dairy cows. Proc. Assoc. Advmt. Anim. Breed. Genet. 18:406–413.

Wang, X., C. Wurmser, H. Pausch, S. Jung, F. Reindardt, J. Tetens, G. Thaller, and R. Fries. 2012. Identification and Dissection of Four Major QTL Affecting Milk Fat Content in the German Holstein-Friesian Population. PLoS ONE 7:e40711.

APPENDIX

Table S1 Validated SNPs with chromosome (BTA), position in base pairs (bp), F-values and effects for intercept and slope. Effect estimates were taken from the validation set. Validated SNPs are indicated in bold type F-values and effect estimates. SNPs with n. c. are not converged.

			F-values in discovery dataset		Effects (in σ)	
SNP name	BTA	bp	Intercept	Slope	Intercept	Slope
Protein yield						
BTB-00046888	1	104729910	11.97	**20.24**	0.139	**-0.073**
BTB-01744054	1	116869213	**11.08**	8.18	**-0.191**	0.094
ARS-BFGL-BAC-7205	1	120983738	**19.03**	**14.11**	**-0.192**	**0.102**
ARS-BFGL-NGS-99492	1	121607486	**18.91**	**13.66**	**-0.192**	**0.099**
BTA-114011-no-rs	1	125911737	**11.76**	1.48	**0.230**	-0.064
BTB-00056059	1	126163013	**11.51**	3.27	**-0.238**	0.109
ARS-BFGL-NGS-98257	1	127680108	**15.28**	10.53	**0.170**	-0.097
BTB-01929922	2	8730899	9.09	**13.34**	0.294	**-0.129**
ARS-BFGL-NGS-86079	2	19126180	**13.44**	8.15	**0.187**	-0.077
Hapmap53232-rs29020795	2	19202356	**12.97**	11.47	**0.185**	-0.071
ARS-BFGL-NGS-81155	5	15353413	8.27	**13.26**	0.192	**-0.089**
Hapmap39895-BTA-15668	5	15392995	**14.82**	**22.03**	**0.164**	**-0.098**
ARS-BFGL-NGS-68464	5	18395406	**12.88**	**18.06**	**-0.241**	**0.099**
Hapmap33079-BTA-163567	6	1936	**11.58**	6.07	**0.201**	-0.084
Hapmap23201-BTC-072836	6	40655229	**15.00**	10.19	**-0.156**	0.067
Hapmap32946-BTC-046820	6	41129701	**23.19**	11.02	**0.193**	-0.067
ARS-BFGL-NGS-39570	6	46320087	**13.56**	4.10	**0.259**	-0.088
Hapmap58150-rs29020620	6	96724594	**22.75**	12.20	**-0.169**	0.060
ARS-BFGL-NGS-43679	6	109680595	**11.67**	8.67	**-0.215**	n. c.
BTA-110673-no-rs	6	111383112	**11.36**	6.73	**-0.272**	n. c.
BTB-00281303	6	111612203	**14.34**	7.89	**-0.237**	0.109
ARS-BFGL-NGS-14880	7	53879989	**14.53**	8.42	**-0.213**	0.055
ARS-BFGL-NGS-113181	7	62800839	**15.09**	**12.83**	**0.202**	**-0.100**
BTB-02035459	7	63196194	**11.60**	**12.98**	**0.164**	**-0.077**
BTB-01219396	7	63221359	**11.60**	**12.98**	**0.164**	**-0.077**
ARS-BFGL-NGS-113819	7	63609102	**17.37**	8.77	**0.220**	-0.088
ARS-BFGL-NGS-109819	7	63664393	**17.65**	10.14	**0.214**	-0.095
BTB-01880776	7	64095706	**11.68**	7.98	**0.268**	-0.109
Hapmap51654-BTA-90094	7	75414240	**11.99**	10.88	**-0.179**	0.076
Hapmap44053-BTA-28733	8	6369477	**16.02**	9.29	**0.193**	-0.071
Hapmap52337-rs29022325	9	23298098	9.83	**12.02**	0.173	**-0.078**
BTA-06997-rs29021351	9	40153426	**24.52**	24.26	**-0.312**	0.127
BTA-83528-no-rs	9	41691114	**24.60**	18.46	**-0.241**	0.092
ARS-BFGL-NGS-37982	9	44200288	**20.64**	21.52	**0.155**	-0.076
ARS-BFGL-NGS-52530	9	44230587	**24.26**	**29.15**	**-0.306**	**0.136**
ARS-BFGL-NGS-103934	9	44255942	**16.01**	18.00	**-0.231**	n. c.
Hapmap34923-BES9_Contig458_891	9	48193041	13.81	**13.21**	0.139	**-0.106**
BTA-83605-no-rs	9	48362593	**12.38**	15.20	**-0.157**	0.077
BTB-01347039	9	51688446	3.57	**11.59**	-0.161	**0.104**
BTB-00391835	9	52160813	**11.14**	**11.57**	**-0.217**	**0.120**
UA-IFASA-2589	9	82175488	15.49	**12.47**	-0.142	**0.087**

ARS-BFGL-NGS-87426	11	30070765	**11.03**	9.95	**-0.175**	0.053
ARS-BFGL-NGS-118724	11	30366110	**10.87**	10.56	**-0.176**	0.070
ARS-BFGL-NGS-113322	11	38523074	8.07	**14.43**	0.245	**-0.137**
BTA-93012-no-rs	11	38544853	8.03	**14.66**	0.244	**-0.139**
ARS-BFGL-NGS-112015	13	63235775	**14.37**	6.43	**0.165**	-0.065
ARS-BFGL-NGS-63777	13	67075815	**21.32**	11.37	**0.302**	-0.109
ARS-BFGL-NGS-103635	13	67816926	**10.88**	2.58	**0.160**	-0.056
ARS-BFGL-NGS-111222	15	57601444	9.10	**11.17**	0.137	**-0.092**
ARS-BFGL-NGS-56645	16	23920210	**16.88**	9.95	**0.183**	-0.080
ARS-BFGL-NGS-41039	16	27619045	**14.56**	5.69	**0.249**	-0.019
ARS-BFGL-NGS-113877	16	29456497	**12.24**	10.66	**-0.221**	0.103
Hapmap38953-BTA-38562	16	29487613	**12.24**	10.66	**-0.221**	0.103
Hapmap50594-BTA-121054	16	29757246	**11.34**	4.24	**0.214**	-0.078
ARS-BFGL-BAC-35294	16	32829224	**11.48**	12.56	**-0.171**	0.074
ARS-BFGL-NGS-38023	16	33318455	**22.31**	16.69	**0.162**	-0.065
ARS-BFGL-NGS-26559	16	33367687	**14.47**	12.04	**-0.194**	0.070
Hapmap42928-BTA-38715	16	33890490	**11.06**	7.58	**0.165**	-0.067
ARS-BFGL-NGS-59645	16	73117625	**16.58**	10.28	**0.216**	-0.082
BTB-02013769	16	74080339	5.84	**14.35**	0.174	**-0.114**
ARS-BFGL-NGS-35499	18	50131636	5.47	**11.14**	0.206	**-0.124**
ARS-BFGL-BAC-36979	18	52926135	6.35	**10.93**	0.104	**-0.085**
ARS-BFGL-NGS-7458	18	54989995	**18.67**	21.69	**0.153**	-0.062
BTA-97501-no-rs	18	57095120	**17.32**	14.37	**0.161**	n. c.
ARS-BFGL-NGS-15837	18	62533851	**12.51**	14.48	**0.157**	n. c.
ARS-BFGL-NGS-38846	18	62814660	10.58	**14.62**	0.129	**-0.088**
ARS-BFGL-NGS-78095	20	22630732	**11.06**	5.12	**-0.173**	0.091
BTB-00783355	20	43550938	12.72	**14.13**	-0.130	**0.105**
BTA-12959-no-rs	21	10922512	**11.14**	6.46	**0.178**	-0.081
ARS-BFGL-NGS-73507	21	20412682	7.63	**13.36**	0.078	**-0.105**
ARS-BFGL-NGS-6001	21	25932661	7.83	**14.69**	-0.214	**0.136**
Hapmap52867-rs29023496	21	26401500	7.02	**12.11**	-0.201	**0.131**
ARS-BFGL-NGS-101900	21	30314497	**17.87**	10.16	**0.180**	-0.119
ARS-BFGL-NGS-110044	21	30892171	**12.64**	7.59	**0.236**	n. c.
ARS-BFGL-NGS-55374	25	28795160	**23.44**	**28.25**	**-0.306**	**0.136**
BTA-62184-no-rs	26	20014035	**16.55**	13.00	**0.161**	-0.063
BTA-60778-no-rs	26	20090833	**20.71**	17.07	**0.183**	n. c.
BTA-99382-no-rs	28	41568645	**12.82**	12.25	**-0.147**	0.064

GENERAL DISCUSSION

The study was divided into two parts: in the first chapter, putative interaction effects between a major gene and the polygenic term for 5 milk traits were investigated. In the second part (chapter two to four) the aim was to identify genes which influence genotype by environment interaction. All studies were done for the German Holstein population.

Major gene by polygene effects

In chapter one, interaction effects between a major gene (DGAT1 K232A) and a polygenic term were analysed. Only one major gene was considered and proof of principles was done. It would be desirable to enlarge these analyses to other genes to detect more interactions. With respect to interaction, only additive x additive interactions were considered, because daughter observations of genotyped sires were used, but daughters were not genotyped. For further analyses, it would be desirable to consider all forms of interaction (additive x dominant, dominant x additive and dominant x dominant), but genotypes and observations have to be available for the same animals to do so. This means cows need to have observations as well as genotypes. As more and more cows are being genotyped worldwide, this should soon be possible.

Genotype by environment interaction

The ability to alter phenotype in response to changes in the environment is called environmental sensitivity (ES) and is often used in animal breeding literature (Falconer and Mackay, 1996). A difference in ES results in genotype by environment interaction (GxE), which means that changes in the environment lead to differences in genotype responses. Effects of GxE can appear in two different ways: first, it can cause scaling effects, when the difference between the phenotypes of two genotypes changes in magnitude between environments, but not in prefix. Secondly, it can cause re-ranking effects, when these difference cause changes in prefix. Re-ranking complicates animal breeding, because the best allele in one environment is not the best in all environments. But re-ranking is rare, scaling appears more often (Calus et al., 2002; Kolmodin et al., 2002) whereas each scaling effects becomes a re-ranking effect if the environment under consideration varies enough. In our analyses, we found large scaling-effects, but only few and small re-ranking effects.

Reaction norm vs. multiple trait model

According to Lynch and Walsh (1998), there are two different possibilities to study GxE: use of a multiple trait or a reaction norm model. If the distribution of the environment is discrete, a multiple trait model is the logical choice. In this case, the phenotypes in different environments are treated as different traits and the genetic correlation is calculated. If there is a low correlation, the traits are different and they are influenced by different genes. If there is a high correlation, nearly the same genes influence the traits. If the distribution of the environment is continuous, this model has the disadvantages that there is a loss of information because of building classes and time and capacity for calculation is very high. In this case, a reaction norm model would be the logical choice. Here, a term for GxE is included into a traditional quantitative model. A reaction norm describes the performance of a genotype as a function of a gradually changing environment (Lynch and Walsh, 1998). The first derivative of the reaction norm, the slope, is defined as ES (de Jong, 1995). In our studies, we used a reaction norm model, because the distributions of the environmental descriptors were continuous. The entire environmental variance was used and thereby the accuracy increased. In chapter two, models with higher order were chosen. In the results of this chapter, it was shown that the reaction norms are nearly linear. Because higher order models are more complex to calculate, we used a linear model for the following studies in agreement with Kolmodin et al. (2002) and Calus and Veerkamp (2003) (chapter three and four).

To map SNPs, a genome-wide association analysis with a two-step-method was used. Lillehammer et al. (2009) used a one-step-method but because we had about 13 million observations, this would be to complex to calculate. So, we decided to calculate breeding values with a reaction norm model in the first step and we did a genome-wide association analysis in the second step.

Data

Our data were first lactation test day records corrected for the fixed effects herd test day, days in milk, age at calving and calving season. Around two percent of the observations in extreme and rare environments were discarded to have enough variation in each environment. Because of these corrections, the estimated GxE effects are probably somewhat smaller. The advantage of test day records as opposed to lactation records is that within-cow variation can be used. The disadvantage or difficulty is the mass of observations. Our cows had at least seven observations in the first lactation and each sire had a minimum of 50 daughters, which resulted in nearly 13 million observations. In order to identify SNPs which not only cause

scaling effects within the environmental range considered in our study, we log-transformed the observations in chapter three (Hayes et al., 2003; Lillehammer et al., 2009).

Environmental descriptor

The choice of an environmental descriptor is probably the most important point when doing GxE-analyses. All components which are non-genetic can be seen as environment, i.e. temperature, herd size, … (Falconer and Mackay, 1996). GxE appears because of two reasons: some genes can be expressed only in some specific environments and sometimes a change in regulation of genes depends on the environment. That's a reason why different studies with different environmental descriptors can detect different SNPs. For statistical analyses, it is important to have a lot of observations. In dairy cattle breeding, artificial insemination is used in a wide field of production environments and a lot of observations for each sire are available with a continuous distribution. Production level is used very often as an environmental descriptor (Calus et al., 2002; Kolmodin et al., 2002). The advantage is a continuous distribution. However, it is more a description of the reaction on the environment than an environmental descriptor. According to Calus et al. (2004) GxE effects can be under-estimated, because the correlation between true environmental parameter and environmental descriptor is less than one. Use of a true environmental descriptor would be better, but in this case, often fewer observations are available and recording is more expensive. A solution could be not to use only one environmental descriptor, but to combine some descriptors to a new one (Lillehammer et al., 2009). In our analyses, we used different environmental descriptors. In chapter two, herd test day solutions (htds) for the corresponding trait under observation were used. The advantages of htds are that they summarize a complex environment, they are easily available and between-cow and within-cow variation can be used (Hayes et al., 2003). The disadvantage is that the explanatory variable contains partly the same information as the dependent variable. Additionally, a new environmental descriptor called htds milk energy yield, which is a combination of htds protein, htds fat and htds milk, was implemented. As it can be seen in chapter two, the correlations between the new htds milk energy yield and the htds for each trait were high; therefore only this descriptor was used in chapter three. Htds milk energy yield can be seen as an indicator for level of feeding and summarizes the single htds in an appropriate way. In chapter four, htds SCS (somatic cell score) was used as an indicator for level of hygiene. SCS as an environmental descriptor can be interesting, because it can be expected that the SCS fluctuation within herds is larger compared to milk traits, e.g. due to temporary herd mastitis infections. Mastitis can be

environmentally associated or contagiously associated. In chapter three and four, the same models were used. As environmental descriptor, htds milk energy yield was used in chapter three and in chapter four htds SCS was used. For some chromosomes, SNPs could be validated with both environmental descriptors. For htds milk energy yield, SNPs were found on BTA 8, 9, 14, 26, 27, and 28. With htds SCS, SNPs were found on BTA 5, 7, 9, 18, and 21. There are some SNPs which are only significant in reference to a special environment. Additionally, a lot of validated SNPs are found for both environmental descriptors, even if the correlation between the descriptors is low. There are some SNPs which react to environmental changes in general.

Validation

The widespread use of a limited number of sires in dairy cattle leads to high degree of linkage disequilibrium (LD), even of SNPs that are separated by several hundred kbp. Additionally, there are a lot of related animals in the studies. If this is not considered in genome-wide association analyses, this leads to a lot of false positives. A good solution can be to validate found SNPs in other populations, because there is a low probability that SNP will be found twice in different populations as false positive. This is called an across-breed-validation (Hayes et al., 2003; Pryce et al., 2010). We have not done this validation in chapters three and four, because we did not have data of other populations. Instead, we split our dataset, which consisted of 2,297 sires into a discovery dataset (n = 1,797 sires) and an independent smaller validation dataset (n = 500 sires). It can be expected that some of our detected and validated loci are false positives. Otherwise, a lot of our validated SNPs are in well-known regions on the chromosomes, where genes that affect milk production can be found. In case of the population under study in this thesis (German Holstein), German Simmental could be a possible population for an across-breed validation.

Consider GxE in dairy cattle breeding?

Today, GxE is not considered in dairy cattle breeding in Germany. To include it will be very complex, because for testing it is important to have observations of daughters of each sire in a wide range of environments. Additionally, it would be difficult to use in practice. In our studies in chapter two, we found only few and small re-ranking effects for milk traits and SCS. Hence, it is not necessary to include GxE in breeding value estimation for the German Holstein population. A reason for this could be that environmental differences in Germany are

comparable small and the environmental range under observation is too small to show large re-ranking effects.

In chapters three and four, we did genome-wide association analyses to find SNPs which influence ES. To improve ES, the optimum would be to use SNPs with a high level of GP (intercept) and a flat or even negative ES (slope). In our studies, we did not find such SNPs. Improvement of ES is only possible at cost of GP.

Future research

For future researches, it will be of interest to use a higher marker density (e. g. use of a 700 K SNP chip), to have more data and to possibly find some SNPs which influence GP and ES in a different way. Another point could be the descriptor of the environment. If data of e. g. feeding regime, hygienic status, geographical region, … would be available for a lot of cows, GxE could be calculated more accurately. For validation, it will be good to use an across-breed-approach. For data of German Holstein, German Simmental could be a possible population, because they live in the same environment and can therefore be used for validation. Additionally, other genomic methods for genome-wide association analyses can be used, e. g. Bayesian methods (Meuwissen et al., 2001), which account simultaneously for LD among markers and not only for LD between marker and QTL.

REFERENCES

Calus, M. P. L., A. F. Groen, G. de Jong. 2002. Genotype x Environment Interaction for Protein Yield in Dutch Dairy Cattle as Quantified by Different Models. J. Dairy Sci. 85: 3115 – 3123.

Calus, M. P. L., R. F. Veerkamp. 2003. Estimation of Environmental Sensitivity of Genetic Merit for Milk Production Traits Using a Random Regression Model. J. Dairy Sci. 86: 3756 – 3764.

Calus, M. P. L., P. Bijma, R. F. Veerkamp. 2004. Effects of data structure on the estimation of covariance functions to describe genotype by environment interactions in a reaction norm model. Genet. Sel. Evol. 36: 489 – 507.

Falconer, D. S., T. F. C. Mackay. 1996. Introduction to quantitative genetics. 4th ed. Longman Scientific and Technical, New York, NY.

Hayes, B. J., M. Carrick, P. Bowman, M. E. Goddard. 2003. Genotype x Environment Interaction for Milk Production of Daughters of Australian Dairy Sires from Test-Day Records. J. Dairy Sci. 86: 3736 – 3744.

Kolmodin, R., E. Strandberg, P. Madsen, J. Jensen, H. Jorjani. 2002. Genotype by Environment Interaction in Nordic Dairy Cattle Studied Using Reaction Norms. Acta Agric. Scand., Sect. A, Animal Sci. 52: 11 – 24.

Lillehammer, M., B. J. Hayes, T. H. E. Meuwissen, M. E. Goddard. 2009. Gene by environment interactions for production traits in Australian dairy cattle. J. Dairy Sci. 92: 4008 – 4017.

Lynch, M., B. Walsh. 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associates Inc Publishers, Sunderland, MA, USA.

Meuwissen, T. H. E., B. J. Hayes, M. E. Goddard. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genet. 157: 1819 – 1829.

Pryce, J. E., S. Bolormaa, A. J. Chamberlain, P. J. Bowman, K. Savin, M. E. Goddard, B. J. Hayesl. 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. J. Dairy Sci. 93: 3331 – 3345.

GENERAL SUMMARY (ENGLISH)

The aim of this thesis was to analyze the influence of DGAT1 and to find SNPs which influence environmental sensitivity (ES). ES is the first derivative of reaction norm models, which are used to analyze genotype-by-environment interactions (GxE). All analyses were done for German Holstein dairy cattle.

Putative interaction effects between DGAT1 K232A mutation and the polygenic terms (all genes except DGAT1) were investigated in **chapter one**. This was done for five milk production traits (milk yield, protein yield, fat yield, protein percentage and fat percentage) in the German Holstein dairy cattle population. Therefore, mixed models are used. The test for interaction relied on the comparison of polygenic variance components depending on the sire's genotypes at DGAT1 K232A. Found substitution effects were highly significant for all traits. Significant interactions between DGAT1 K232A and the polygenic term were found for milk fat and protein percentage. These interactions could be used in breeding schemes. Depending on the DGAT1 K232A genotypes of the sample, in which the sire will be used, three polygenic breeding values of a sire can be calculated. Because the genotypes of the samples are often unknown and usually heterogeneous, this is not a practical approach. Rank correlations between the three polygenic EBVs were always above 0.95, which suggested very little re-ranking.

GxE were studied in **chapter two**. For this, reaction norm random regression sire models were used in the German Holstein dairy cattle population. Around 2300 sires with a minimum of 50 daughters per sire and at minimum seven first-lactation test day observations per daughter were analyzed. As traits, corrected test day records for milk yield, protein yield, fat yield and somatic cell score (SCS) were used. As environmental descriptors, we used herd test day solutions (htds) for milk traits, milk energy yield or SCS. Second-order orthogonal polynomial regressions were applied to the sire effects. Results showed significant slope variances of the reaction norms, which caused a non-constant additive genetic variance across the environmental ranges considered, which pointed to the presence of minor GxE effects. When the environment improved, the additive genetic variance increased, meaning higher (lower) htds for milk traits (SCS). This was also influenced by pure scaling effects, because the non-genetic variance increased in an improved environment and the heritability was less influenced by the environment. For the environmental ranges considered in this study, GxE

effects caused very little re-ranking of the sires. To obtain unbiased genetic parameters, it was important to model heterogeneous residual variances.

A large genome-wide association analysis was conducted in **chapter three** to identify SNPs that affect general production (GP) and environmental sensitivity (ES) of milk traits. Around 13 million daughter records were used to calculate sire estimates for GP and ES with help of linear reaction norm models. Daughters were offspring from 2297 sires. The sires were genotyped with a 54k SNP chip. As environmental descriptor, the average milk energy yield performance of the herds at the time where the daughter observations were recorded was used. The sire estimates were used as observations in genome-wide association analyses using 1797 sires. With help of an independent validation set (500 sires of the same population), significant SNPs were confirmed. To separate GxE scaling and other GxE effects, the observations were log-transformed. GxE effects could be found with help of reaction norm models and numerous significant SNPs could be validated for GP and ES, whereas many SNPs affecting GP also affected ES. ES of milk traits is a typical quantitative trait, which is controlled by many genes with small effects and few genes with larger effect. Effects of some SNPs for ES were not removable by log-transformation of observations, indicating that these are not solely scaling effects. Positions of founded clusters were often in well-known candidate regions affecting milk traits. No SNPs, which show effects for GP and ES in opposite directions could be found.

Environmental descriptor in GxE analyses is often modelled by average herd milk production levels. Another possibility could be the level of hygiene and udder health. In **chapter four**, the same models were used as in chapter three. A genome-wide association analysis was done using htds for SCS as an environmental descriptor. With help of this, several SNP clusters affecting intercept and slope could be detected and confirmed. Many SNPs or clusters affecting intercept and slope could be identified, but in total, the number of SNPs affecting intercept was larger. The same SNPs could be detected and validated with and without considering GxE in reaction norm models. Some SNPs affecting only slope were found. For slope, nearly the same SNPs could be found with SCS as an environmental descriptor as presented in chapter three, although both environmental descriptors were only slightly correlated.

GENERAL SUMMARY (GERMAN)

Ziel dieser Arbeit war es den Einfluss von DGAT1 zu untersuchen und SNPs zu finden, die die Umweltsensitivität (ES) beeinflussen. ES ist die erste Ableitung eines Reaktionsnormmodells und wird genutzt, um Genotyp-Umwelt-Interaktionen (GxE) zu analysieren. Alle Untersuchungen wurden anhand von Daten der Rasse Deutsche Holstein durchgeführt.

Mögliche Interaktionseffekte zwischen der DGAT1 K232A Mutation und einem polygenen Term (alle Gene außer DGAT1) wurden in **Kapitel eins** untersucht. Dies geschah für die fünf Milchproduktionsmerkmale Milchmenge, Proteinmenge, Fettmenge, Proteinanteil und Fettanteil in der deutschen Holstein Population. Hierzu wurden gemischte Modelle genutzt. Der Test auf Interaktion beruhte auf dem Vergleich der polygenen Varianzkomponenten in Abhängigkeit des Genotyps des Bullen an dem Gen DGAT1 K232A. Die gefundenen Substitutionseffekte waren hoch signifikant für alle Merkmale. Signifikante Interaktionen zwischen DGAT1 K232A und dem polygenen Term konnten für Fett- und Proteingehalt gefunden werden. Diese Interaktionen können in Zuchtprogrammen genutzt werden. Abhängig vom DGAT1 K232A Genotyp der Kühe, an die ein Bulle angepaart werden soll, können drei polygene Zuchtwerte eines Bullen berechnet werden. Da die Genotypen der Kühe oft unbekannt und normalerweise heterogen sind, ist dies allerdings keine praxisnahe Vorgehensweise. Die berechneten Rangkorrelationen zwischen den drei polygenen Zuchtwerten waren immer größer 0.95, was bedeutet, dass sehr wenige Rangierungseffekte aufgetreten sind.

GxE wurden mit zufälligen Reaktionsnormregressionsvatermodellen innerhalb der deutschen Holstein-Population in **Kapitel zwei** untersucht. Die Daten von ca. 2300 Bullen mit mindestens 50 Töchtern pro Bulle und mindestens sieben Testtagsbeobachtungen pro Tochter innerhalb der ersten Laktation wurden analysiert. Die betrachteten Merkmale waren korrigierte Testtagsbeobachtungen für Milchmenge, Proteinmenge, Fettmenge und Zellzahl (SCS). Als Umweltparameter wurden Herdentesttagseffekte für die Milchmerkmale, den Milchenergiegehalt und SCS hinzugezogen. Orthogonale Polynomregressionen zweiter Ordnung wurden für die Vatereffekte betrachtet. Die Ergebnisse zeigten signifikante Varianzen der Steigung der Reaktionsnormen, was eine nicht konstante additiv genetische Varianz innerhalb des gewählten Umweltbereiches bedingt. Das deutet wiederum auf das

Vorkommen kleiner GxE-Effekte hin. Verbessert sich die Umwelt, steigt die additiv genetische Varianz, was bedeutet, das höhere (niedrigere) Herdentesttagseffekte für Milchmerkmale (SCS) auftreten. Begründet sind diese Veränderungen durch Skaleneffekte, da die nichtgenetische Varianz in einer besseren Umwelt ansteigt und dadurch die Heritabilität von der Umwelt weniger beeinflusst wird. Für den ausgewählten Umweltbereich dieser Studie erklären GxE nur wenige Rangierungseffekte der Bullen. Um unabhängige genetische Parameter zu erhalten, war es wichtig die Restvarianz heterogen zu modellieren.

Eine genomweite Assoziationsanalyse (GWAS) wurde in **Kapitel drei** durchgeführt, um SNPs zu finden, die Produktionsniveau (GP) und ES der Milchmerkmale beeinflussen. Ca. 13 Millionen Beobachtungen der Töchter wurden genutzt, um Schätzer für GP und ES der Bullen mit Hilfe eines linearen Reaktionsnormmodells zu schätzen. Die Töchter waren von 2297 verschiedenen Bullen, die mit einem 54k SNP Chip genotypisiert wurden. Als Umweltparameter wurde die durchschnittliche Milchenergiemenge der Herden betrachtet, die zeitgleich zu den Töchterbeobachtungen gemessen wurde. Die Schätzwerte von 1797 Bullen wurden anschließend in einer GWAS untersucht. Mit Hilfe eines unabhängigen Validierungsdatensatzes (500 Bullen derselben Population), konnten signifikante SNPs bestätigt werden. Um bei den GxE reine Skaleneffekt von sonstigen Effekten trennen zu können, wurden die Daten logarithmiert. GxE-Effekte konnten mit Hilfe der Reaktionsnormmodelle gefunden werden und zahlreiche signifikante SNPs für GP und ES validiert werden, wobei viele SNPs sowohl GP als auch ES beeinflussen. ES der Milchmerkmale ist ein typisches quantitatives Merkmal, das von vielen Genen mit kleinen Effekten und wenigen Genen mit großen Effekten beeinflusst wird. Die Effekte einiger SNPs für ES konnten durch das Logarithmieren nicht entfernt werden. Das zeigt, dass diese Effekte nicht nur Skaleneffekte sind. Die Positionen der gefundenen Cluster sind oft in bekannten Kandidatengenregionen für Milchmerkmale. Es konnten keine SNPs gefunden werden, die GP und ES in unterschiedlichen Richtungen beeinflussen.

Der Umweltparameter in GxE-Analysen wird oft als durchschnittliches Milchproduktionsniveaus der Herde modelliert. Eine andere Möglichkeit können das Hygieneniveau und die Eutergesundheit sein. In **Kapitel vier** wurden die gleichen Modelle wie in Kapitel drei genutzt. Bei der anschließenden GWAS wurden Herdentesttagseffekte für SCS als Umweltparameter genutzt. Dadurch konnten viele SNP-Cluster entdeckt und bestätigt werden, die das Produktionsniveau und die Steigung beeinflussen. Weiterhin konnten viele

Cluster identifiziert werden. Wobei die Anzahl der SNPs, die das Produktionsniveau beeinflussen höher war. Auch wenn die Umweltvariable nicht betrachtet wurde, konnten für das Produktionsniveau die gleichen SNPs gefunden werden. Aber es wurden auch einige SNPs gefunden, die nur die Steigung beeinflussen. Dies waren fast die gleichen SNPs, wie die in Kapitel drei gefundenen SNPs (Umweltparameter Milchenergiemenge).

DANKSAGUNG

An erster Stelle möchte ich mich bei Herrn Prof. Dr. Jörn Bennwitz für die Überlassung des Themas und die Unterstützung und Betreuung dieser Arbeit sehr herzlich bedanken.

Für die Übernahme des Koreferates bedanke ich mich bei Herrn Prof. Dr. G. Thaller.

Mein Dank gilt ebenfalls Chris Baes, die meine Arbeit sprachlich unterstützt hat.

Allen Kollegen und Freunden, die mich während meiner Promotion begleitet haben, ein großes Dankeschön.

Einen ganz herzlichen Dank meinem Freund und meiner Familie.

Mein Dank gilt ebenfalls der DFG (Deutsche Forschungsgemeinschaft) für die finanzielle Unterstützung.

EIDESSTATTLICHE ERKLÄRUNG

Ich versichere an Eides statt, dass ich die vorliegende Arbeit selbständig verfasst habe. Alle statistischen Analysen, welche auch in den Vorabveröffentlichungen aufgeführt worden sind, wurden von mir selbst durchgeführt. Alle Stellen, die wörtlich oder dem Sinn nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche kenntlich gemacht. Ich versichere außerdem, dass ich keine andere als die angegebene Literatur verwendet habe. Diese Versicherung bezieht sich auch auf alle in der Arbeit enthaltenen Zeichnungen, Skizzen, bildlichen Darstellungen und dergleichen. Die Hilfe einer kommerziellen Promotionsvermittlung oder –beratung wurde nicht in Anspruch genommen.

Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Mondorf, 11. Juni 2013

Melanie Streit

CURRICULUM VITAE

persönliche Angaben	
Name	Melanie Elisabeth Streit
Geburtsdatum	18.05.1984
Geburtsort	Saarlouis
Nationalität	deutsch
Ausbildung	
2010 – 2013	Universität Hohenheim Promotion
2008 – 2010	Universität Hohenheim Studium der Agrarwissenschaften (Fachrichtung Tierwissenschaften) mit Abschluss zum Master of Science
2003 – 2007	FH Bingen Studium der Landwirtschaft mit Abschluss zum Dipl. Ing. (FH) Landwirtschaft
1994 – 2003	Gymnasium am Stefansberg, Merzig
1990 – 1994	Grund- und Hauptschule Hilbringen
Praxiserfahrung	
ab 07/2013	LKV Bayern e.V.
10/2008 – 03/2010	Universität Hohenheim Studentische Hilfskraft am Institut für Tierproduktion in den Tropen und Subtropen, Fachgebiet Tierhaltung und Tierzüchtung in den Tropen und Subtropen
08/2008 – 09/2008	Milcherzeugerberatungsring Pfalz-Hunsrück e. V., Kusel Praktikum
09/2007 – 03/2008	Lemmer Fullwood GmbH, Lohmar Assistentin im Marketing
03/2006 – 08/2006	Gut Behl, Plön, Milchviehbetrieb, Praktikum
08/2004 – 03/2005	Laiterie Freiwald, Freistroff (Frankreich) Milcherzeugerberatung, Praktikum
07/2003 – 09/2003	Lehr- und Versuchsanstalt Hofgut Neumühle, Münchweiler a. d. Alsenz (Rinder, Schweine, Schafe, Damtiere), Praktikum