

# Using Genome-Wide Association Analysis to Characterize Environmental Sensitivity of Milk Traits in Dairy Cattle

Melanie Streit,\* Robin Wellmann,\* Friedrich Reinhardt,<sup>†</sup> Georg Thaller,<sup>‡</sup> Hans-Peter Piepho,<sup>§</sup> and Jörn Bennewitz\*<sup>1</sup>

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

**ABSTRACT** Genotype-by-environment interaction (GxE) has been widely reported in dairy cattle. One way to analyze 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 single-nucleotide polymorphisms (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 approximately 13 million daughter records by the use of linear reaction norm models. The daughters were offspring from 2297 sires. Sires were genotyped for 54k SNPs. The environment was defined as the average milk energy yield performance of the herds at the time during which the daughter observations were recorded. The sire estimates were used as observations in a genome-wide association analysis, using 1797 sires. Significant SNPs were confirmed in an independent validation set (500 sires of the same population). 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 that affect 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.

## KEYWORDS

environmental sensitivity genotype-by-environment interaction association analysis dairy cattle

Breeding cattle for milking traits relies on the use of daughter records for an 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 has resulted in a substantial genetic gain for milking traits during 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, which questions 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

Copyright © 2013 Streit *et al.*

doi: 10.1534/g3.113.006536

Manuscript received March 25, 2013; accepted for publication April 25, 2013

This is an open-access article distributed under the terms of the Creative Commons Attribution Unported License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supporting information is available online at <http://www.g3journal.org/lookup/suppl/doi:10.1534/g3.113.006536/-/DC1>

<sup>1</sup>Corresponding author: Garbenstraße 17, 70599 Hohenheim, Germany. E-mail: j.bennewitz@uni-hohenheim.de

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 that include the average herd production level as a continuous environmental descriptor are widely used to study GxE (Calus *et al.* 2002; Kolmodin *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 frequently are fitted with the use of 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 that it be placed 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 frequently has been 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 nonfluctuating environments because sensitive individuals are able to benefit from these environmental conditions. In poor, fluctuating, or unforeseeable environments, robust individuals are desired, if the robustness does not come at the expense of a decrease 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 by using approximately 10,000 single-nucleotide polymorphisms (SNPs) in the Australian dairy cattle population to find significant SNPs affecting GP and ES. Several SNPs were significant, and approximately 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 been analyzed frequently (*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), which would allow partitioning of removable by data transformation and nonremovable 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 by the use of first-order

■ **Table 1** Sire variance components of the random regression analyses

Trait	$\sigma_{S_0}^2$	$\sigma_{S_1}^2$	$\rho_{S_0S_1}$
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) <sup>a</sup>	9.50 (< 0.01)	0.11 (< 0.01)	0.61
ln(fat yield) <sup>a</sup>	12.70 (< 0.01)	0.13 (< 0.01)	0.73
ln(milk yield) <sup>a</sup>	10.55 (< 0.01)	0.14 (< 0.01)	0.68

Standard errors are shown in parentheses.  $\sigma_{S_0}^2$  ( $\sigma_{S_1}^2$ ) denotes the intercept (slope) sire variance, with correlation  $\rho_{S_0S_1}$ .

<sup>a</sup> Values are multiplied by 10,000.

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. To remove GxE causing scaling effects, the observations were log-transformed in some analyses.

## MATERIALS AND METHODS

### Data and data editing

In total 2356 progeny-tested German Holstein sires were genotyped with the Illumina BovineSNP50 BeadChip, which contains a total of 54,001 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 2297 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 also were excluded. This data filtering was performed with 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).

Approximately 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 totaled approximately 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 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.*,  $energy\ yield = 0.802 \times milk\ yield + 38.4 \times fat\ yield + 23.6 \times protein\ yield$ , where the yields are measured in kilograms (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 SD 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

■ **Table 2** Number of discovered and validated SNPs for intercept and slope for the traits on the observed scale

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

The FDR  $q$ -values (FDR) of the significant SNP with the largest error probability ( $P \approx 0.001$ ) in the discovery dataset are shown. SNP, single-nucleotide polymorphism; FDR, false-discovery rate.

shown in the Supporting Information, Figure S1. 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 Vereinigte Informationssysteme Tierhaltung (2013) for a detailed description. For additional information please see File S2, File S3, and File S4.

### Statistical analysis

In a previous study we applied a second-order sire model, which gave an improved fit compared with 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 first-order sire model in the present study. The following random regression model was applied in the first step for all three milk yield traits:

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

where  $cy_{ijk}$  is the corrected yield of daughter  $i$  of sire  $j$  at herd test day  $k$ ,  $\mu$  is the overall mean,  $ht dsme_k$  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  $\text{var} \begin{bmatrix} s_0 \\ s_1 \end{bmatrix} = I \otimes \begin{bmatrix} \sigma_{s_0}^2 & \sigma_{s_0 s_1} \\ \sigma_{s_0 s_1} & \sigma_{s_1}^2 \end{bmatrix}$ , and that of the daughter effects is  $\text{var} \begin{bmatrix} d_0 \\ d_1 \end{bmatrix} = I \otimes \begin{bmatrix} \sigma_{d_0}^2 & \sigma_{d_0 d_1} \\ \sigma_{d_0 d_1} & \sigma_{d_1}^2 \end{bmatrix}$ , with  $I$  being the identity matrix. The estimated sire effects were used as observations in an association analysis [see model (2)]. In contrast to classical sire models, the relationship among sires was ignored, which 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).

To model heterogeneous residual variance across the environments, the observations were ordered according to the environmental descriptor and grouped into 10 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.

The whole data set was randomly split into a discovery data set ( $n = 1797$  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:

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

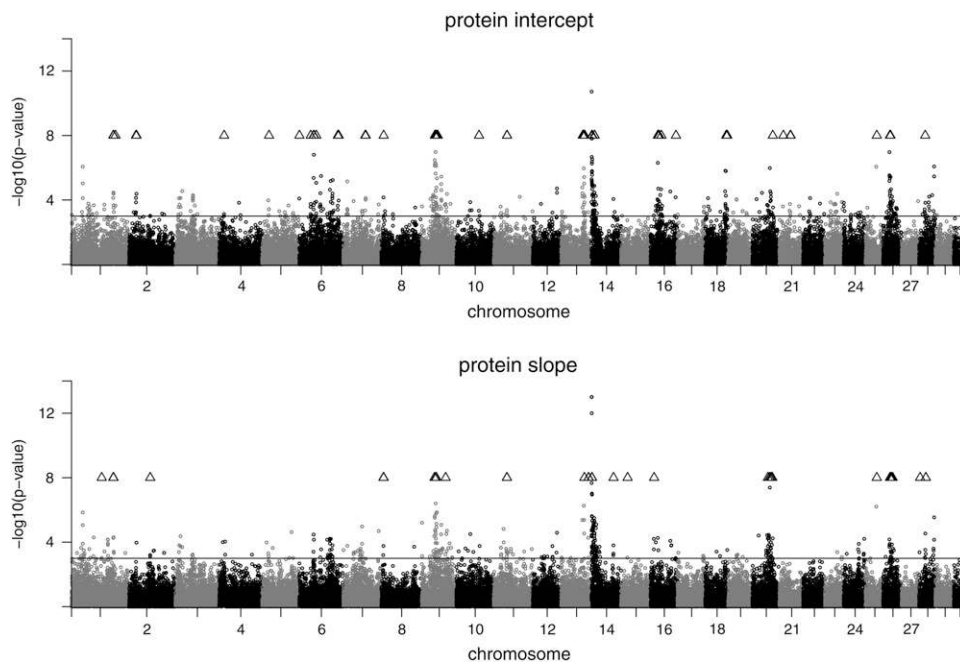
where  $\hat{s}_{jm}$  is the estimated random sire effect for GP ( $m = 0$ ) and ES ( $m = 1$ ). These estimates were taken from the results of model (1). The model was applied for the two traits ( $m = 0$  for GP and  $m = 1$  for ES) separately. The effect of the SNP  $k$  was modeled as a regression on the number of copies of the allele with the greater frequency ( $x = 0, 1, \text{ or } 2$ ), with  $b_{km}$  being the regression coefficient. To control for the population structure, we fitted a random sire effect with the covariance structure  $A\sigma_{sm}^2$ , where  $A$  is the numerator relationship matrix calculated from high-quality pedigree information and  $\sigma_{sm}^2$  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 of  $P < 0.001$  as significant. To judge how many false-positive results were among the significant associations we applied the false-discovery rate (FDR) technique. We calculated for each association test an FDR  $q$ -value by using the software QVALUE (Storey and Tibshirani 2003). The FDR  $q$ -value of the significant SNP with the lowest test statistic ( $P \approx 0.001$ ) provided an estimate of the proportion of false-positive results 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  $P < 0.01$  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).

■ **Table 3** Number of discovered and validated SNPs for intercept and slope for the traits on the log-scale

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

The FDR  $q$ -values (FDR) of the significant SNP with the largest error probability ( $p \approx 0.001$ ) in the discovery dataset are shown. SNP, single-nucleotide polymorphism; FDR, false-discovery rate.

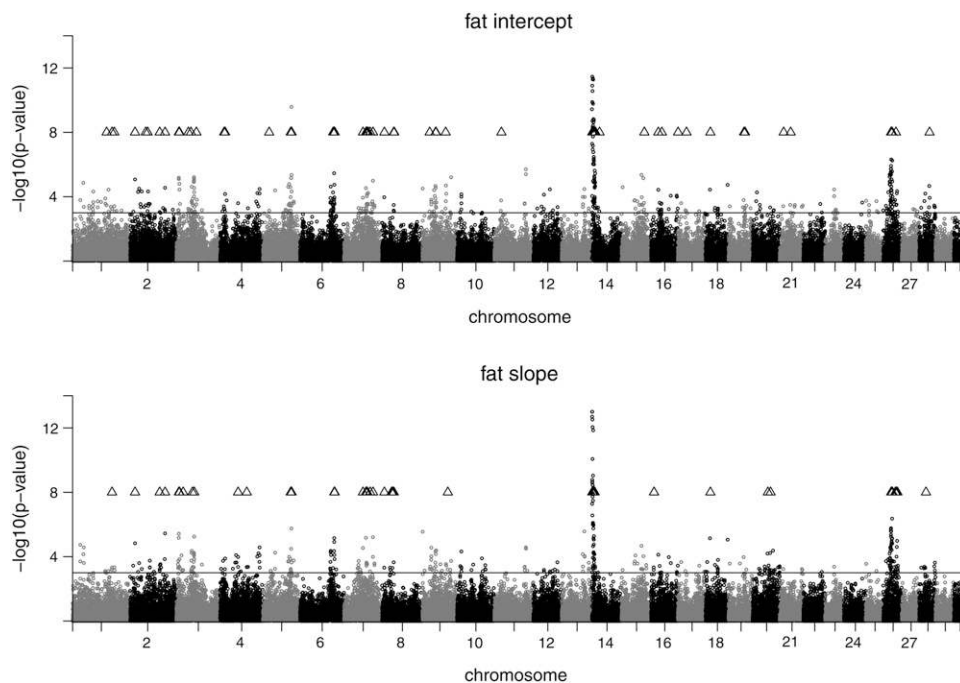


**Figure 1** Test statistic profile of SNP effects for protein yield intercept (top) and protein yield slope (bottom) in the discovery data set. The nominal significance level ( $P < 0.001$ ) is indicated by a solid line. Positions of validated SNPs are indicated by a triangle.

To identify SNPs that not only cause scaling effects within the environmental range considered in our study, we also applied the models 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).

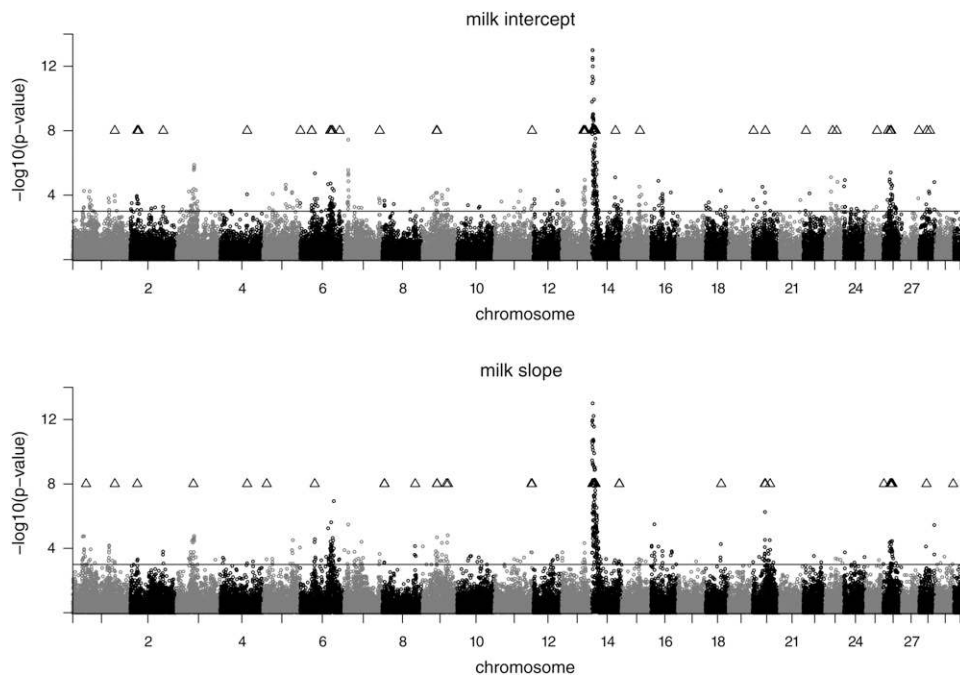
## RESULTS

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 analyzed 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.





**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 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 approximately 7–9% of these are false-positive results. 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 greater.

The plots of the test statistic along the chromosomes are shown in Figure 1, Figure 2, and Figure 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.

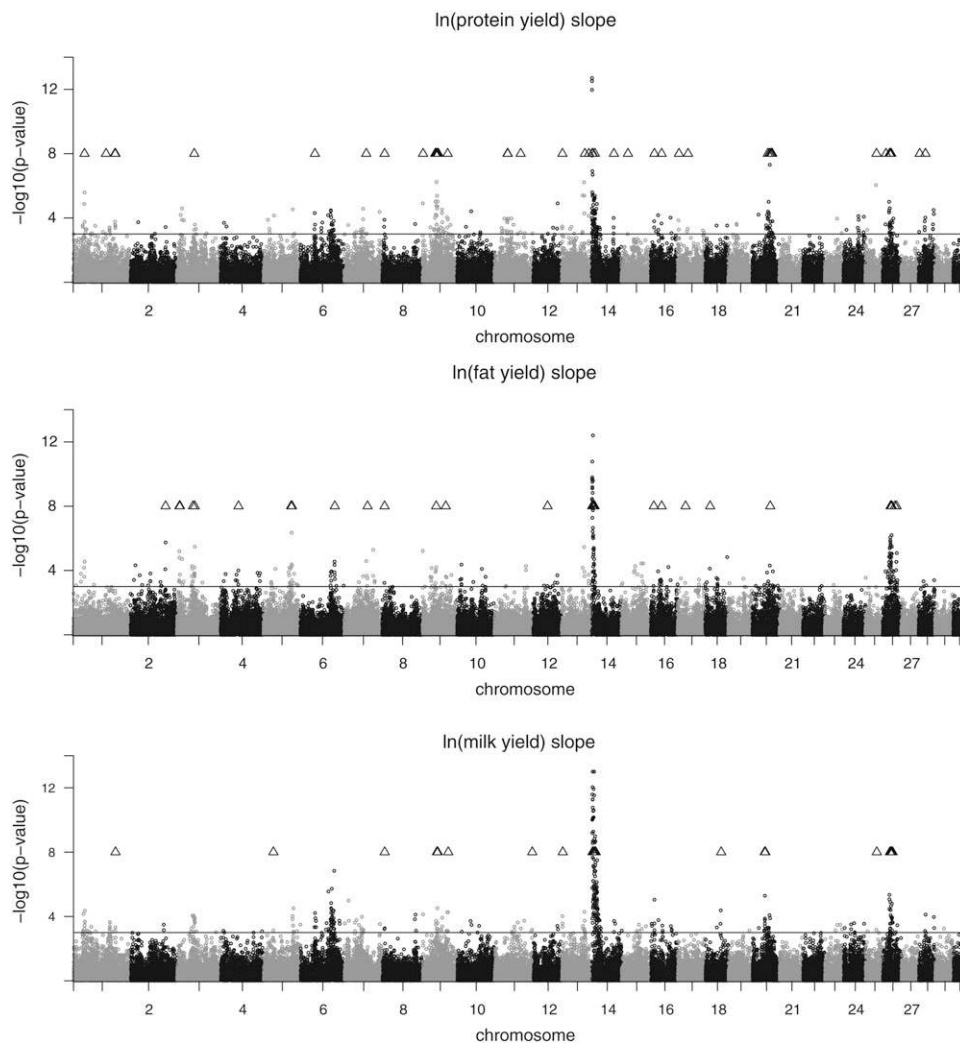
In Figure 5 the estimates of the validation set are shown for SNPs that were either significant for intercept, or for slope, or for both. The slope effect of the allele that increases the intercept is shown. It can be seen that every validated SNP affects both traits in the same direction,

and the correlation between the solutions is highly positive. This was less pronounced if the data were log-transformed (Figure 6). For In (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,  $\sigma$ ) were observed for milk yield, with 11 (4) SNPs showing an intercept (slope) effect larger than  $0.3\sigma$ . 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 Table S1; estimates for the log-transformed observations are presented in Table S2.

## 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, Table S1). Given the FDR  $q$ -values of the SNPs in the discovery set (Table 2 and Table 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 whether 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 quantitative trait



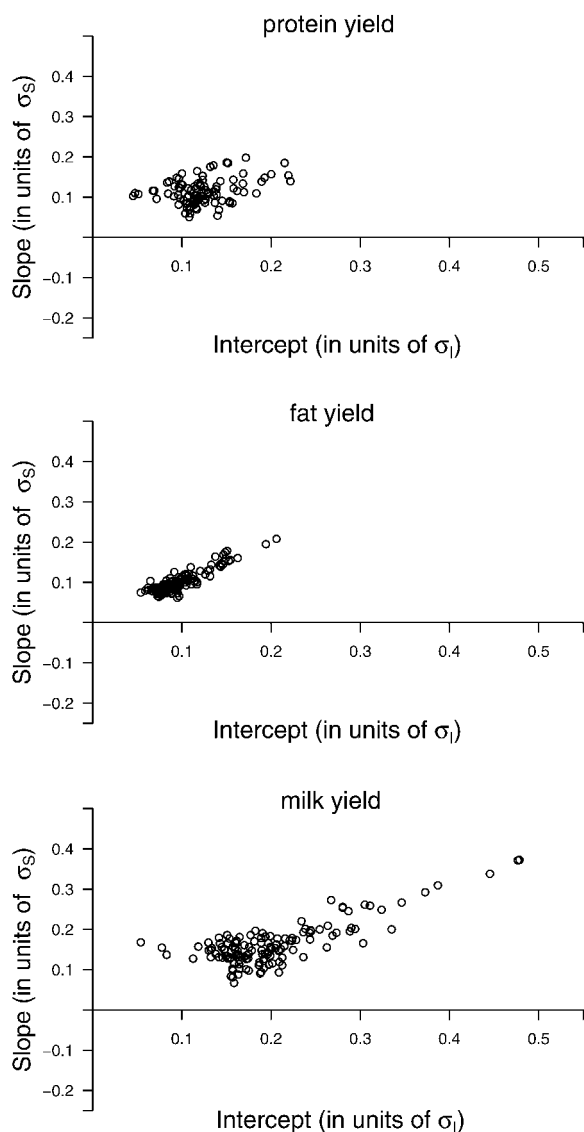
**Figure 4** Test statistic profile of SNP effects for ln(protein yield) slope (top), ln(fat yield) slope (middle), and ln(milk yield) slope (bottom) in the discovery data set. The nominal significance level ( $P < 0.001$ ) is indicated by a solid line. Positions of validated SNPs are indicated by a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

loci (QTL) regions using multimarker and pedigree information and uses these probabilities for QTL fine-mapping (Meuwissen *et al.* 2002). This method is, however, computationally demanding and needs greater marker densities. Another multimarker 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 aforementioned approaches 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 by 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 that segregate in the German Holstein population. This is most obvious on BTA14, where 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 harbors 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. In addition, 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).

Some validated SNPs for ES are in chromosomal regions similar to those found in other dairy cattle populations harboring 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



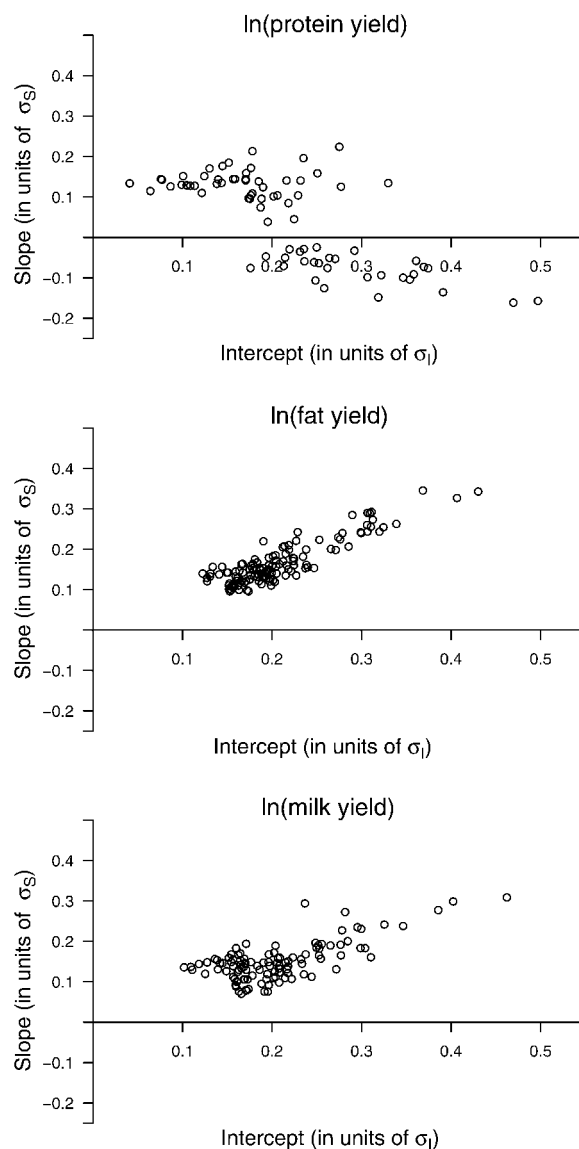
**Figure 5** Estimated SNP effects for the traits on the observed scale. The term  $\sigma_S$  ( $\sigma_I$ ) denotes the sire intercept (slope) SD. Each SNP was validated within the population either for intercept, slope or both. Estimates were taken from the validation set.

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). To remove these scaling effects, a log-transformation was applied. The results from the association analysis applied to the log-transformed data revealed SNPs that were not removable by this kind of transformation. These validated SNPs are of special interest because they point to chromosomal regions harboring genes with an effect on ES that 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  $\ln(\text{protein yield})$  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, Table S2). Hence, these effects were completely removable by the log-transfor-

mation. 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. Because 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.

As described in the *Introduction*, breeding for robustness for both milk production and health traits is an issue in dairy cattle. In this



**Figure 6** Estimated SNP effects for the traits on the log-scale. The term  $\sigma_S$  ( $\sigma_I$ ) denotes the sire intercept (slope) SD. Each SNP was validated within the population either for intercept, slope or both. Estimates were taken from the validation set.

study only milk production traits were considered. On the basis of 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.

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

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

## LITERATURE CITED

- 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*. American Seed Trade Association, Washington, DC.
- Bennowitz, 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.
- Bennowitz, 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. Moiso, 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, and 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'Connell, C. P. van Tassell, T. S. Sonstergard *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, and 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, and 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, 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 *et al.*, 2009 A validated genome wide association study to breed cattle adapted to an environment altered by climate change. *PLoS ONE* 4: e6676.
- Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman, and 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: e1001139.
- 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, and C. Gondro. Springer Science+Business Media B. V., Dordrecht, The Netherlands.
- Knap, P. W., 2005 Breeding robust pigs. *Aust. J. Exp. Agric.* 45: 763–773.
- 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.
- Kolmodin, R., E. Strandberg, P. Madsen, J. Jensen, and H. Jorjani, 2002 Genotype by environment interaction in nordic dairy cattle studied using reaction norms. *Acta Agric. Scand. Sect. Anim. Sci.* 52: 11–24.
- 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.
- 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., M. Árnýasi, 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, and T. H. E. Meuwissen, 2009a Reducing the bias of estimates of genotype by environment interactions in random regression sire models. *Genet. Sel. Evol.* 41: 30.
- Lillehammer, M., B. J. Hayes, T. H. E. Meuwissen, and 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, Sunderland, MA.
- 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: e5350 10.1371/journal.pone.0005350.
- Meuwissen, T. H. E., B. J. Hayes, and 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, and 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, and 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 Milchenenergiegehaltes beim Schwarzbunten Milchrind. *J. Anim. Physiol. Anim. Nutr. (Berl.)* 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. USA* 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.
- 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.
- van Tienderen, P. H., and H. P. Koelewijn, 1994 Selection on reaction norms, genetic correlations and constraints. *Genet. Res.* 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.
- Vereinigte Informationssysteme Tierhaltung United Data System Verden w. V. Description of the national animal genetic evaluation system for German Holstein dairy cattle. Available at: <http://www.vit.de/index.php?id=publikationen-zws>. Accessed: May 10, 2013.
- 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: e40711.
- Weikard, R., C. Kühn, T. Goldammer, G. Freyer, and 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.* 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. USA* 99: 9300–9305.

Communicating editor: B. S. Yandell