Gene by environment interactions for production traits in Australian dairy cattle

M. Lillehammer,^{*1} B. J. Hayes,[†] T. H. E. Meuwissen,^{*} and M. E. Goddard^{††}

^{*}Norwegian University of Life Sciences, Department of Animal and Aquacultural Sciences, 1432 Ås, Norway

[†]Biosciences Research Division, Department of Primary Industries Victoria, 1 Park Drive, Bundoora 3083, Australia

^{††}University of Melbourne, Parkville VIC 3010, Australia

ABSTRACT

Dairy farming is carried out under a wide range of production environments, including large variations in the level of feeding. Although reranking of dairy sires based on the level of feeding of their daughters has been reported, detecting the genetic mutations that cause this genotype by environment interaction has not been previously attempted. In our experiment to find genetic markers for such mutations, we selected 388 Holstein bulls from the Australian dairy bull population and genotyped them for 9,919 single nucleotide polymorphism (SNP) markers. Production data, consisting of first-lactation test-day records for milk yield, fat yield, protein yield, protein percentage, and fat percentage, from the daughters of the genotyped bulls were used to estimate the effect of each SNP, which was modeled as a regression on herd mean test-day yield, where herd mean test-day yield is a descriptor of the environment. Data were analyzed with 4 models; in 2 models, daughter records were analyzed directly, with and without taking sire relationships into account. With the other 2 models, sire reaction norms for each trait were calculated and marker effects on the sire reaction norms were estimated with and without taking sire relationships into account. The results showed that using daughter records directly and accounting for sire relationships was necessary to obtain high power and to limit the number of false positives. With this approach, SNP with significant effects were found for all traits. Log transformation of the data did not affect the power of gene detection. The significant markers were categorized according to their joint effects on production and environmental sensitivity. Potential gene candidates and application of the markers is discussed. About one-third of the significant markers affect intercept and slope in opposite directions, and some of these facilitate marker-assisted selection for robustness.

Key words: gene by environment interaction, genome scan, selection for robustness, dense single nucleotide polymorphism marker genotyping

INTRODUCTION

Genetic variation in response to environmental changes has been reported for milk production from dairy cattle, particularly when the “environment” is the level of feeding of the herd (Hayes et al., 2003). This genetic variation has been characterized at the between-sire level. Knowledge about how genotypes or single genes are affected by environmental changes would allow more accurate predictions of the performance of a sire’s daughters given the level of feeding or the production system in which his daughters are to be milked.

Random regression models have been used to describe genotype by environment interactions (Kolmodin et al., 2002; Calus and Veerkamp, 2003; Fikse et al., 2003) and gene by environment interactions (Lillehammer et al., 2007b) in dairy cattle. Gene by environment interactions can be included in the QTL search by modeling the gene effect as a function of a continuous environmental variable (Lillehammer et al., 2007a). For production traits, production level has been suggested as an environmental variable that is able to capture a complex environment consisting of several components into one variable (Kolmodin et al., 2002). Production level has been described as herd-year average (Kolmodin et al., 2004), herd-year effect (Lillehammer et al., 2007b), herd average milk peak yield (Fikse et al., 2003), herd-year-season effect (Calus et al., 2002), and herd-test-day average (Hayes et al., 2003). The benefit of using herd-test-days to estimate production level, and using test-day records instead of lactation records, is that variation within cows can be included and more accurate estimates of sire breeding values for environmental sensitivity are obtained (Hayes et al., 2003).

In dairy cattle, DNA samples from sires used in artificial insemination are often easily available. Genotypes from sires are therefore used, even for analyses of

Received June 25, 2008.

Accepted February 11, 2009.

¹Corresponding author: marie.lillehammer@nofima.no

milk production traits. Records from daughters of the genotyped sires are used to describe sire performance. Often, records from all the daughters of a genotyped sire are averaged and this average is used to describe sire performance. When environmental interactions are included in the model, every single daughter record is used instead of an average of daughters by the same sire, to capture the range of performance across different herd-test-day production levels (Lillehammer et al., 2007a). The disadvantage of using single daughter records directly in the analysis of gene by environment effects is that it substantially increases the number of records and thereby the computational time.

To reduce computational time, a 2-step approach can be used, where the effects of each sire on the intercept and slope of the random regression are estimated. These sire effects can then be used as dependent variables in the QTL detection model, giving a much smaller data set, and decrease the computational time. However, this approach builds on the assumption that the sire term is able to capture all the effects of the single markers when the markers themselves are not included in the model, which may not be the case. Another approach is to ignore relationships between sires. This may speed up the estimation process, but could give more false positives, because related individuals will have many markers in common (Emahazion et al., 2001).

We can speculate a priori on the value of genes affecting environmental sensitivity that may be found. Genes that affect both general production and environmental sensitivity of production in the same direction are expected to be common, because of the positive genetic correlation between the production traits and their environmental sensitivities (Calus et al., 2002). However, these genes might be difficult to detect because of the similarity between their effect and the polygenic effect; that is, because too many genes have similar effects (Lillehammer et al., 2007a). Genes that affect general production and environmental sensitivity of production in opposite directions are probably less common, but easier to detect (Lillehammer et al., 2007a). These genes might be of great interest because they facilitate joint selection for high production and low environmental sensitivity (i.e., selection for robustness). Some genes might also affect environmental sensitivity without affecting general production. They might seem well suited to select for robustness, but genes that have no effect in the average environment will have a negative effect on production in half the production environments, assuming that reaction norms for single genes are linear. Therefore, the most promising candidates for marker-assisted selection for robustness would have a negative correlation between average production and environmental sensitivity, and the allele that decreases

environmental sensitivity should have positive effects on production, not only in the average environment, but over most of the environmental range.

The aim of this study was to conduct a genome-wide association study for interactions between single genes and production level in Australian dairy cattle. This was done by searching for markers that affect production traits to a different extent on herd-test-days with different production levels. The same data were analyzed with 4 different models to see if a simple model could provide the same results but with less computational demands than the full model. The results indicate that genes affecting environmental interactions can be detected and categorized by our approach. Possible candidate genes for selection for robustness are presented.

MATERIALS AND METHODS

Three hundred eighty Holstein bulls were selected from the Australian dairy bull population for genotyping. The bulls selected were those with extreme large and small EBV for the Australian selection index, $ASI = (3.8 \times \text{protein}) + (0.9 \times \text{fat}) - (0.048 \times \text{milk})$, and the records for the subcomponents are based on performance of the bulls daughters (a progeny test).

Genotyping was performed by Affymetrix (San Diego, CA) using 9,919 SNP markers. The SNP were mapped to Bovine Genome assembly 3.1 by BLAST search (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>); 9,037 of the markers were polymorphic in the 380 bulls.

First-lactation test-day records on milk yield, protein yield, fat yield, fat percentage, and protein percentage from all the daughters of the genotyped sires, as well as other first-lactation cows belonging to the same herd-test-days, were extracted from the Australian Dairy Herd Improvement Scheme (ADHIS) database. All data were checked for a valid sire ID and acceptable ranges for the production traits and the age of cow. Only records from herd-test-days with at least 12 tested first-lactation cows were used. The data set included 5,108,391 test-day records on 851,076 cows from 141,244 different herd-test-days.

All data analyses were performed by using the AS-REML software (Gilmour et al., 2002). Four different analyses were done and their results compared. Single daughter records were analyzed with and without taking sire relationships into account. Sire effects were also analyzed with and without taking sire relationships into account.

Precorrection Model

Before the analyses, data were precorrected. The precorrection model was common to all analyses. The

parameters in the model are from Hayes et al. (2003), but herd-test-day is fitted as a random effect to give more precise predictions of the herd-test-day solutions (Strabel and Szwachkowski, 1999):

$$y_{ijkl} = \mu + YS_i + \sum_{n=1}^3 A_n x_n + \sum_{n=1}^8 D_n Z_n + HTD_j + S_k + e_{ijkl},$$

where y_{ijkl} is milk yield, protein yield, fat yield, protein percentage, fat percentage, $\ln(\text{milk yield})$, $\ln(\text{protein yield})$, $\ln(\text{fat yield})$, $\ln(\text{protein percentage})$, or $\ln(\text{fat percentage})$ from the i th year-season of calving, j th herd-test-day, k th sire, and l th cow in her first lactation; μ is the overall mean; YS_i is the effect of the i th year-season of calving; x_n is the n th-order orthogonal polynomial corresponding to age on day of test; A_n is a fixed regression coefficient of the trait on age at test; Z_n is the n th-order polynomial corresponding to days in milk at test; D_n is a fixed regression coefficient of the trait on days in milk; HTD_j is the random effect of the j th herd-test-day; S_k is the random effect of the k th sire; and e_{ijkl} is the residual effect.

The data were corrected for the fixed effects of year-season of calving, age of the cow, and days in milk, and for the random effect of herd-test-day. The corrected yield and the random herd-test-day effects were used in the subsequent analyses, where the herd-test-day effects described the environment. The environment of each trait was therefore distributed around a mean of zero. Only daughters of the genotyped sires were included in subsequent analyses, and this data set included 593,902 records on 95,954 cows from 141,244 herd-test-days. Records from daughters of sires with missing genotypes for a particular SNP were deleted when the effect of that SNP was analyzed, so the actual number of records used varied between SNP.

Full Model Analysis

In the full model, the corrected yield was used as the dependent variable. The model included both a fixed regression of the genotype in each position and random regressions of sire and cow on herd-test-day solution obtained from the pre-correction model:

$$cy_{ijklp} = \mu + Y_i + b \times htdsol_j + \sum_{m=0}^1 G_{lm} htdsol_j^m + \sum_{m=0}^1 S_{km} htdsol_j^m + \sum_{m=0}^1 P_{pm} htdsol_j^m + e_{ijklp},$$

where cy_{ijklp} is the corrected trait value from the i th year of calving, j th herd-test-day, k th sire, l th genotype, and p th cow obtained from the previous step; μ is the overall mean; Y_i is the effect of the i th year of calving;

$htdsol_j$ is the j th herd-test-day effect estimated in the previous step; b is a fixed regression coefficient on the herd-test-day effect; G_{lm} is the fixed regression coefficient of the l th genotype on the herd-test-day effect; S_{km} is the random regression coefficient of the k th sire on the herd-test-day effect;

$$Var(S) = A \otimes \begin{bmatrix} \sigma_{s0}^2 & \sigma_{s0s1} \\ \sigma_{s0s1} & \sigma_{s1}^2 \end{bmatrix},$$

where A is the relationship matrix among the sires, σ_{s0}^2 is the sire variance of intercept, σ_{s1}^2 is the sire variance of slope, and σ_{s0s1}^2 is the covariance between sire intercept and sire slope; P_{pm} is the random regression coefficient of the p th cow on the herd-test-day effect, and

$$Var(P) = I \otimes \begin{bmatrix} \sigma_{p0}^2 & \sigma_{p0p1} \\ \sigma_{p0p1} & \sigma_{p1}^2 \end{bmatrix},$$

σ_{p0}^2 is the permanent environmental variance of intercept, σ_{p1}^2 is the permanent environmental variance of slope, and σ_{p0p1}^2 is the covariance between permanent environmental intercept and permanent environmental slope; e_{ijklp} is the residual effect; and $Var(e) = \sigma_e^2$. Thus, the expected value of cy_{ijklp} is

$$E(cy_{ijklp}) = \mu + Y_i + b \times htdsol_j + \sum_{j=0}^1 G_{lm} htdsol_j^m,$$

and its variance is

$$Var(cy) = \begin{bmatrix} \mathbf{H}^0 & \mathbf{H}^1 \end{bmatrix} Var(S) \begin{bmatrix} \mathbf{H}^0 & \mathbf{H}^1 \end{bmatrix}' + \begin{bmatrix} \mathbf{H}_p^0 & \mathbf{H}_p^1 \end{bmatrix} Var(P) \begin{bmatrix} \mathbf{H}_p^0 & \mathbf{H}_p^1 \end{bmatrix}' + I\sigma_e^2,$$

where element (ik) of matrix $\mathbf{H}^0$ ($\mathbf{H}^1$) equals 1 ($htdsol_j$) if the i th record was produced by a daughter of the k th sire and zero otherwise; and element (ik) of matrix $\mathbf{H}_p^0$ ($\mathbf{H}_p^1$) equals 1 ($htdsol_j$) if the i th record was produced by the k th daughter, and zero otherwise.

Ignoring Pedigree Information

The same model was used as in the full model analysis, except that $Var(S) = I \otimes \begin{bmatrix} \sigma_{s0}^2 & \sigma_{s0s1} \\ \sigma_{s0s1} & \sigma_{s1}^2 \end{bmatrix}$, ignoring relationships between the sires.

The Two-Step Procedure – Step 1

For the 2-step procedure, the precorrected records were used to predict sire effects for intercept and slope with the model

$$cy_{ijklp} = \mu + Y_i + b \times htdsol_j + \sum_{m=0}^1 S_{km} htdsol_{jm} + \sum_{m=0}^1 P_{pm} htdsol_{jm} + e_{ijklp},$$

where all parameters are defined as above. Each sire got 2 predicted effects, 1 for intercept (S_{k0}) and 1 for slope (S_{k1}) for each trait. These predicted sire effects were used in the subsequent SNP analyses, where effects on intercept and slope were tested separately in univariate analyses.

Step 2: Including Pedigree Information

Pedigree information was included in the second step of the procedure, attempting to distinguish between differences between sires that can be described by the relationship matrix and differences that can be described by the SNP under investigation: $S_{lkm} = \mu_m + S0_k + G0_l + e_{lkm}$, where S_{lkm} is the estimated effect for intercept ($m = 0$) or slope ($m = 1$) (analyzed separately) for the k th sire with genotype l , whereas $G0_l$ is the fixed effect of the l th genotype; $S0_k$ is the random effect of sire k , and other parameters are as defined above; $Var(S0) = A\sigma_s^2$, where A is the relationship matrix among the sires and σ_s^2 is the sire variance.

Step 2: Ignoring Pedigree Information

When pedigree information was ignored, genotype was the only effect in addition to μ and residual in the second step of the analysis: $S_{lkm} = \mu_m + G0_l + e_{lkm}$, where parameters are as defined above.

Results from the 3 simpler analyses were compared with the full analysis with regard to number of significant SNP found for the different traits. For each trait, the number of significant SNP when a log-transformation of the data was used was compared with the number of significant SNP when the data were analyzed without any transformations. Two tables of results were made, one with a nominal significance threshold of $P < 0.0001$ and one with a more relaxed nominal significance threshold of $P < 0.001$. In both cases the F -threshold was obtained from the F -table and degrees of freedom assumed to be 1 and infinite because of the large number of records analyzed.

Single nucleotide polymorphisms significant ($P < 0.001$) for one or more traits were categorized after their joint effect on general production and environmental sensitivity. Each category was presented and its suitability for selection was discussed. All the SNP significant ($P < 0.001$) for one or more traits were clustered according to their positions. Those SNP located <1,000 bp apart were considered to belong to the same cluster. Being a part of such a cluster was regarded as additional evidence for a SNP effect to be real. Clusters that facilitated selection for robustness were described in more detail.

RESULTS

The numbers of detected SNP with the different models at different nominal significance levels are summarized in Table 1 ($P < 0.0001$) and Table 2 ($P < 0.001$). One hundred one highly significant ($P < 0.0001$) SNP were found to affect environmental sensitivity of one of the investigated traits. Relaxing the significance threshold gave 225 SNP affecting environmental sensitivity of one or more traits. In total, the most stringent significance level gave 157 positive tests (marker-trait combinations), and 18 false positives were expected, giving a false discovery rate of 0.11. Relaxing the threshold gave 610 significant marker-trait combinations. With 181 expected false positives, that gives a false discovery rate of 0.30. Regardless of whether daughter records or sire breeding values were used as the dependent variable (and regardless of the chosen significance level), ignoring pedigree information resulted in a large increase in the number of SNP significant for intercept of all traits. This is a result of a large number of false positives due to population structure. For the number of SNP detected to affect slope, the effect of excluding pedigree information was less consistent. Using sire effect as the dependent variable resulted in fewer significant SNP for both significance levels, indicating that the 2-step procedure is less powerful.

Log-transformation is sometimes used as a correction for heterogeneous genetic variance, and is assumed to eliminate some interaction caused by scaling effects (Hayes et al., 2003). The correlations between intercept and slope for sires (genetic) and cows (permanent environment) are reported in Table 3 and show that log-transformation reduces these correlations substantially for most of the traits. This shows that the log-transformation succeeds in removing scaling effects caused by heterogeneous genetic variance across the environmental scale. When pedigree information was excluded, the log-transformation caused a reduction of the number

Table 1. Number of significant ($P < 0.0001$) SNP detected for each trait with each model

Trait	Daughter records				Sire breeding values			
	With pedigree		Without pedigree		With pedigree		Without pedigree	
	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope
Milk yield	3	1	80	26	1	0	4	0
Ln ¹ milk yield	5	4	131	3	1	1	10	1
Protein yield	8	27	307	8	0	1	38	10
Ln protein yield	9	45	365	2	0	1	40	1
Protein percentage	1	1	44	4	0	3	10	10
Ln protein percentage	4	0	57	4	1	2	9	8
Fat yield	11	8	135	5	0	0	11	4
Ln fat yield	7	14	181	2	0	0	15	0
Fat percentage	3	1	9	7	0	0	3	0
Ln fat percentage	5	0	16	8	0	0	4	0

¹Log-transformation.

of detected SNP affecting slope in most traits for both significance levels. When pedigree information was included, the effect of log-transformation was less consistent. The genetic correlation between slope and intercept also remained positive after log-transformation. This result is because the log-transformation does not remove all scaling effects of the percentage traits.

The SNP with significant effects on slope or intercept can be divided into 2 categories: those that affect intercept and slope in the same direction and those that affect intercept and slope in opposite directions. Figure 1 shows the difference between the categories: Figure 1a shows a gene with positively correlated effects on intercept and slope, whereas Figure 1b shows a gene with negatively correlated effects on intercept and slope. For all traits, SNP belonging to both categories were found. About two-thirds of the detected SNP affected intercept and slope in the same direction, whereas one-third of the significant SNP showed opposite effects on slope and intercept. Single nucleotide polymorphisms that af-

fect intercept and slope in the same direction will cause an increase in environmental sensitivity if used to select for increased production. To reduce environmental sensitivity, without decreasing production in general, genes with negative correlation between general production (intercept) and environmental sensitivity (slope) are needed. The detected genes with negative correlated effects on intercept and slope can be further divided into 3 subclasses due to their suitability as selection candidates for robustness.

Some of the detected genes have very small effects on intercept, and the effects of the alleles are equal close to the average environment. An example of such a gene is shown in Figure 2. These genes are less suitable for selection because the positive allele in low-yielding environments is negative in high-yielding environments. Examples of such genes were found for all traits.

Some of the detected genes have very small effects on environmental sensitivity. An example of a gene affecting general production only is shown in Figure 3. These

Table 2. Number of significant ($P < 0.001$) SNP detected for each trait with each model

Trait	Daughter records				Sire breeding values			
	With pedigree		Without pedigree		With pedigree		Without pedigree	
	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope
Milk yield	23	10	257	67	4	2	18	5
Ln ¹ milk yield	22	27	328	19	25	1	43	2
Protein yield	43	58	607	60	2	1	156	69
Ln protein yield	28	69	710	8	2	2	170	2
Protein percentage	27	23	107	23	9	11	29	23
Ln protein percentage	24	12	134	31	9	12	32	20
Fat yield	66	42	360	20	4	4	52	24
Ln fat yield	16	33	451	7	5	0	76	4
Fat percentage	23	31	61	47	5	3	11	6
Ln fat percentage	28	5	71	29	5	6	11	4

¹Log-transformation.

Table 3. Genetic and permanent environmental correlations¹ between slope and intercept for the investigated traits

Trait	With pedigree				Without pedigree			
	Sire correlation ²		Cow correlation ³		Sire correlation ²		Cow correlation ³	
	Direct	Ln ⁴ -transform	Direct	Ln-transform	Direct	Ln-transform	Direct	Ln-transform
Milk yield	0.55	−0.08	0.93	−0.10	0.65	0.01	0.77	−0.05
Protein yield	0.80	0.11	0.93	−0.07	0.72	0.09	0.65	−0.04
Protein percentage	0.81	0.70	0.42	−0.08	0.65	0.52	0.16	−0.05
Fat yield	0.88	0.51	0.98	−0.05	0.84	0.40	0.74	−0.02
Fat percentage	0.75	0.49	0.42	0.10	0.58	0.28	0.19	0.02

¹Standard error ≈ 0.05 for all the correlations.

²The sire correlation is the genetic correlation between slope and intercept.

³The cow correlation is the permanent environmental correlation between slope and intercept.

⁴Log-transformation.

genes might be useful for selection for production, but less suitable to change environmental sensitivity, because they have little effect on that trait. Examples of such genes were found for all traits.

The remaining SNP within the category of genes with negative correlated effects on intercept and slope

are the most attractive selection candidates for robustness. For these genes, the allele positive for production in most or all environments has a negative effect on slope. They can be used to select for robustness by increasing production most in low-yielding environments. Examples of such genes were found for all traits.

Forty-two clusters were found, consisting of 2 to 6 SNP significant for one or more traits, where a cluster is defined as significant SNP located <1,000 bp apart. Single nucleotide polymorphisms within a cluster tended to be significant for the same trait(s) and belong to the same category. Twenty-eight of the detected clusters had a positive correlation between the effect on general production and the effect on environmental sensitivity. Among the clusters that had a negative correlation between intercept and slope, 6 affected slope only by having a crossing point close to the average environment. Two clusters with a negative correlation between intercept and slope seemed to have no effect on slope. The remaining 7 clusters seemed to have a negatively correlated effect on intercept and slope, and could be selection candidates for robustness. Four of these are described in more detail below.

A cluster of 2 SNP located on *Bos taurus* autosome (BTA) 8 shows a pleiotropic effect on milk and protein yield. As shown in Figure 4, the positive allele of this gene increases both milk and protein yield while reducing the environmental sensitivities for the same traits. The linkage disequilibrium (LD) between these 2 markers, measured as r^2 , was 0.987. A high LD is expected for SNP located this close to each other, and confirms that these 2 SNP should show similar effects.

Bos taurus autosome 14 was found to contain 3 clusters of significant SNP, 2 of which could be selection candidates for robustness of fat percentage. The 2 candidates consisted of 2 and 3 SNP, respectively; their effects are shown in Figure 5. The LD between the SNP in the first cluster was 0.923, whereas the LD between the SNP in the second cluster ranged from 0.912 to 1.

Figure 1. a) Single nucleotide polymorphisms with positively correlated effects on intercept and slope of a trait. The allele that increases production also increases environmental sensitivity; b) SNP with negatively correlated effects on intercept and slope of a trait. The allele that increases production decreases environmental sensitivity. Such genes may facilitate marker-assisted selection for robustness.

Figure 2. An example of an SNP affecting slope only. These genes do not facilitate marker-assisted selection if animals are bred for a broad range of environments because the positive allele in one environment is negative in other environments.

The LD between the first and the second cluster was 0.03

Another possible selection candidate for robustness of milk yield is located on BTA 20. This cluster consists of 2 SNP, and the effect of the gene is shown in Figure 6. The positive allele increases milk yield while decreasing environmental sensitivity for milk yield. The LD between the 2 SNP in this cluster was 1.

DISCUSSION

The aim of this study was to describe interactions between single genes and the production environment for production traits in Australian dairy cattle. In addition to the full model, simpler models were applied to test whether computational time could be reduced. Reducing the computational time by calculating sire effects for intercept and slope and searching for SNP affecting the sire effects resulted in decreased power for all traits, compared with searching for SNP that directly affect each daughter record (Tables 1 and 2). This indicates that some of the variance generated by

Figure 3. An example of an SNP affecting intercept only. Such genes facilitate marker-assisted selection to change general production without affecting environmental sensitivity.

Figure 4. The allelic effects of the alleles of a gene detected on *Bos taurus* autosome 8 with environmentally dependent effect on milk and protein yield.

the SNP is modeled elsewhere other than in the sire variance in the first model in the 2-step analysis. Scans for QTL that do not consider environmental interactions usually use such a procedure; either by calculating daughter yield deviations (DYD; e.g., Ashwell et al., 2004) or by estimating sire breeding values (e.g., Olsen et al., 2002). However, this seems inadequate when sire effects have to be predicted both on intercept and slope, where there is extra information from the performance of daughters across the range of the environmental descriptor. Szyda et al. (2008) found that reliability of DYD in a random regression test-day model was dependent on the number of test-days per daughter as well as the number of daughters per son. The average number of test-days per daughter in our study was 6.19, whereas Szyda et al. (2008) found that a minimum of 6 was needed to get reliable DYD. Szyda et al. (2008) calculated DYD expressed as regression coefficients on days in milk, which is similar to calculating DYD as regression coefficients on an environmental variable. One difference could be that a cow with 6 test-days would usually have the test-days distributed over a large part of a 305-d lactation, but over a small range of environments, as environmental conditions are

Figure 5. The allelic effects of the alleles of 2 clusters detected on *Bos taurus* autosome 14 with environmentally dependent effects on fat percentage. The solid lines represent the effects of the alleles of the first cluster, while the dotted lines represent the effects of the alleles of the second cluster.

assumed to be less variable within herds. This might increase the minimum number of test-day observations needed per daughter to obtain reliable DYD.

As related sires share several markers, similarities in daughter records because of relationships can be captured by several SNP in the model when the SNP are tested one by one. Ignoring pedigree information was therefore expected to cause an increased number of significant SNP. Ignoring pedigree information resulted in an increased number of SNP affecting intercept, but the effect of ignoring pedigree was inconsistent for SNP affecting slope (Tables 1 and 2). Ignoring pedigree information seemed to decrease the power of detecting SNP affecting slope. Whether the number of detections is increased or decreased when pedigree is ignored depends on whether the number of SNP missed because of reduced power is greater or less than the number of extra false positives. This is confirmed within the traits where the model that ignores pedigree detects more SNP than the model that includes pedigree information. Even for these traits, several SNP detected when pedigree information is included are not detected when pedigree information is ignored.

Robustness is the ability of an animal to adapt well to different environments. Traits such as health, fertility, and other fitness traits will affect robustness. In this study, only production traits are measured, and maintaining production over a wide range of environments is used as a measurement of adaptability. In theory, an animal can achieve less variable production by 2 strategies: producing less in good environments or producing more in poor environments (or both). For practical purposes, the second strategy is the advantageous one, because we ideally want the animals to be robust without decreasing production. Increased robustness is obtained by selecting for alleles with a positive effect on the trait and a negative correlation between production and environmental sensitivity.

As the environment is described by the predicted environmental effect on production, the average environmental sensitivity in the population is, by definition, 1. The gene effects are modeled as deviations from the average, so an allele with a negative slope will cause an environmental sensitivity <1 , whereas an allele with a positive slope will cause the animal to have an environmental sensitivity >1 . For biological reasons, it is unlikely that an animal can have a negative environmental sensitivity, because that would imply that the animal produces less if the environment is improved. An allele with a negative slope will therefore cause the environmental sensitivity to approach zero and thereby increase robustness, defined as less change in production across environments.

Because general production and environmental sensitivity are correlated traits, selection for one of them will also change the other one. So far, selection has been performed on general production, ignoring possible interactions with environment. This has caused a correlated increase in environmental sensitivity (Kolmodin et al., 2003). High environmental sensitivity can be beneficial if the environment is highly controllable. Environmental improvements will then have large effects on production. In uncontrollable environments, however, highly sensitive animals will fluctuate more in their production and produce less under poor conditions. Animals with lower environmental sensitivity (robust animals) produce more consistently across environments and are well suited for changing environments, variable environments, or export to different environments.

Genes with a negative correlation between general production and environmental sensitivity facilitate selection for robustness of production across the range of feeding levels. However, genes belonging to the 3 different categories of genes with negative correlations between slope and intercept, described in the Results section, have different applications in selection. Genes

Figure 6. The allelic effects of the alleles of a gene detected on *Bos taurus* autosome 20 with environmentally dependent effect on milk yield.

with almost no effect on intercept, illustrated by Figure 2, are generally poor selection candidates. Selection for one of the alleles will cause a negative response in about half of the environments. These genes could be seen as candidates when breeding for specialized animals, designed to produce in a specific environment. Breeding animals for one specific environment would require the environment to be highly controllable, and is therefore usually not recommendable.

The genetic correlation between general production and environmental sensitivity was positive for all traits and varied from 0.55 to 0.88 (Table 3). This suggests that most genes should affect intercept and slope in the same direction, and that a smaller fraction of the genes affect only production or only environmental sensitivity. These expectations are in general confirmed by the results both for single SNP and for clusters. In both cases about two-thirds of the detected genes had a positive correlation between general production and environmental sensitivity and about half of the remaining genes had no effect on either intercept or slope. The method of detection used here has been shown to give a detection bias, because it has a larger power to detect genes with reranking of alleles between environments than genes where the allelic effects only differs in scale (Lillehammer et al., 2007a). This will cause a small detection frequency of genes that have a small effect in extremely poor environments and increase in effect over the environmental scale. Log-transformation is used to decrease the detection frequency of such "scaling" genes, and could increase the detection bias. Therefore, the true fraction of genes that have a positive correlation between general production and environmental sensitivity is probably greater than two-thirds.

Two categories of possible selection candidates were found, one that affects production in general without affecting environmental sensitivity (illustrated by Figure 3) and one that affects production and environmental sensitivity in opposite directions. Genes similar to the example in Figure 3 facilitate selection for greater production while keeping environmental sensitivity unchanged, whereas genes that affect production and environmental sensitivity in opposite directions facilitate selection for robustness. Which selection strategy is desirable depends on how the current environmental sensitivity in the population is, compared with optimal environmental sensitivity.

Optimal environmental sensitivity depends on variability and predictability of the environment. Environmental sensitivity will always be present in a population, in the sense that improved environment will cause improved production. Robust animals will react less to changes in the environment, causing environmental improvements to give a smaller increase

in production, but at the same time, environmental disturbances will also have less effect on production. The optimal environmental sensitivity would therefore be greater in high-producing environments than in low-producing environments. In this study, reaction norms are assumed to be linear, causing environmental sensitivity to be constant over the environmental range. Extending the model to include nonlinear reaction norms might lead to the discovery of genes sensitive to improvements in the environment, but less sensitive to deterioration of the environment (compared with an average environment). However, such a model extension could also cause computational time to be increased and result in increased standard errors of the parameter estimates and convergence problems. When linear reaction norms are assumed, decisions about the desired environmental sensitivity have to be constant for the range of environments considered. Animals that are too robust might be of ethical concern, as they continue their high production even under conditions in which they should have decreased production as a response to a stressful environment (Kolmodin et al., 2003). However, to breed for robustness to avoid a further increase in environmental sensitivity or even to obtain a decrease in environmental sensitivity seems worthwhile in uncontrolled and variable environments. In Australia, uncontrollable environmental conditions, such as the weather, vary between years, and in some years amount and quality of the food might be reduced. In more controlled environments, where the animals are kept indoors and fed mainly on concentrates, robustness would be less important. The genes that affect production without affecting environmental sensitivity (Figure 2) facilitate selection for production in environments where the current environmental sensitivity in the population is assumed to be optimal.

One reason why QTL are reported in this study that have not been reported earlier could be that some QTL are only detectable when environmental interaction is taken into account. The presented selection candidates for robustness were significant for intercept, and should therefore be detectable also without taking environmental interaction into account, but because they have the highest effect in low-yielding environments, they would be hard to detect in studies performed in high-yielding environments. Gene by environment interaction is one possible cause of inconsistency between QTL mapping studies, such that QTL detected in one environment may not be detected in another, or if detected, may have different effects.

When using a dense SNP marker map, it is expected that several SNP positions can be significant because they are in LD with the same gene. Moreover, after clustering, there might be several clusters in LD with

the same gene, as we only identified SNP closer than 1,000 bp apart to belong to the same cluster. An example could be the 2 clusters on BTA14 found to affect fat percentage. They are located 12 Mbp apart, but have similar effects (Figure 5). One indication as to whether these positions are in LD with the same or with different genes is the LD between the 2 clusters, which was found to be as low as 0.03. This indicates that more than one gene having environmentally dependent effects on fat percentage segregate on this chromosome. Several genes with environmentally dependent effects for milk yield have previously been found close to each other on BTA6 (Lillehammer et al., 2008).

Four candidates especially suited to select for robustness under variable environmental conditions are presented in this article. The presented candidates were confirmed by more than one significant SNP within a region of 1,000 bp. They facilitate selection for robustness for one or more traits and all represent opportunities to increase level of production while reducing environmental sensitivity. Hence, improved robustness does not have to come at the expense of reduced productivity.

In addition to robustness of production over an environmental range, other traits are of great importance when breeding for animals adaptable to a wide range of environments. Traits such as disease resistance, fertility traits, and survival traits are important to include in a sustainable animal breeding program, aiming for robust animals. To quantify the importance of the detected genes in breeding programs, a possible effect of the genes on other traits needs to be investigated. If high production in a poor environment comes at the expense of fertility, for example, breeding for increased production in bad environments could cause a loss of profitability. For the detected genes to facilitate selection for robustness, the increased production in poor environments should not come at the expense of an increased stress on the animals. For instance, cows might achieve increased production in a poor environment without high stress because of a low feed requirement for maintenance.

CONCLUSIONS

When using random regression to search for gene by environment interactions affecting dairy production traits, daughter records should be used directly, rather than being transformed to sire records, to increase power. Log-transformation of the phenotypes did not increase the power of gene detection. About 1 in 3 of the detected associations had a negative correlation between intercept and slope, and, given their patterns

of effects, some of these are good candidates to accelerate selection for robustness of production traits. Two-thirds of the detected associations had a positive correlation between intercept and slope and would increase environmental sensitivity if used to increase overall production.

REFERENCES

- Ashwell, M. S., D. W. Heyen, T. S. Sonstegard, C. P. Van Tassell, Y. Da, P. M. VanRaden, M. Ron, J. I. Weller, and H. A. Lewin. 2004. Detection of quantitative trait loci affecting milk production health and reproductive traits in Holstein cattle. *J. Dairy Sci.* 87:468–475.
- Calus, M. P. L., A. F. Groen, and G. de Jong. 2002. Genotype $\times$ environment interaction for protein yield in Dutch dairy cattle as quantified by different models. *J. Dairy Sci.* 85:3115–3123.
- Calus, M. P. L., and 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.
- Emahazion, T., L. Feuk, M. Jobs, S. L. Sawyer, D. Fredman, D. St Clair, J. A. Prince, and A. J. Brookes. 2001. SNP association studies in Alzheimer's disease highlight problems for complex disease analysis. *Trends Genet.* 17:407–413.
- Fikse, W. F., R. Rekaya, and K. A. Weigel. 2003. Genotype $\times$ environment interaction for milk production in Guernsey cattle. *J. Dairy Sci.* 86:1821–1827.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2002. ASReml User Guide Release 1.0. VSN International Ltd., Hemel Hempstead, UK.
- Hayes, B. J., M. Carrick, P. Bowman, and M. E. Goddard. 2003. Genotype $\times$ 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, B. Danell, and H. Jorjani. 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. A Anim. Sci.* 54:139–151.
- Kolmodin, R., E. Strandberg, H. Jorjani, and B. Danell. 2003. Selection in presence of a genotype by environment interaction: Response in environmental sensitivity. *Anim. Sci.* 76:375–385.
- Kolmodin, R., E. Strandberg, P. Madsen, J. Jensen, and H. Jorjani. 2002. Genotype by environment interaction in Nordic dairy cattle studied by use of reaction norms. *Acta Agric. Scand. A Anim. Sci.* 52:11–24.
- Lillehammer, M., M. Árnýasi, S. Lien, H. G. Olsen, E. Sehested, J. Ødegård, and T. H. E. Meuwissen. 2007b. 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, S. Lien, and T. H. E. Meuwissen. 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. 2007a. Random regression models for detection of gene by environment interaction. *Genet. Sel. Evol.* 39:105–121.
- Olsen, H. G., L. Gomez-Raya, D. I. Våge, I. Olsaker, H. Klungland, M. Svendsen, T. Adnøy, A. Sabry, G. Klemetsdal, N. Schulman, W. Krämer, G. Thaller, K. Rønningen, and S. Lien. 2002. A genome scan for quantitative trait loci affecting milk production in Norwegian dairy cattle. *J. Dairy Sci.* 85:3124–3130.
- Strabel, T., and T. Szwaczkowski. 1999. The use of test day models with small size of contemporary groups. *J. Anim. Breed. Genet.* 116:379–386.
- Szyda, J., E. Ptak, J. Komisarek, and A. Zarnecki. 2008. Practical application of daughter yield deviations in dairy cattle breeding. *J. Appl. Genet.* 49:183–191.