Predicting energy balance for dairy cows using high-density single nucleotide polymorphism information

K. L. Verbyla,*†‡§¹ M. P. L. Calus,* H. A. Mulder,* Y. de Haas,* and R. F. Veerkamp*

*Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, Lelystad 8200AB, the Netherlands

†Melbourne School of Land and Environment, The University of Melbourne, Parkville, Victoria 3001, Australia

‡Department of Primary Industries, 1 Park Drive, Bundoorra, Victoria 3083, Australia

§The Cooperative Research Centre for Beef Genetic Technologies, University of New England, Armidale, NSW 2351, Australia

ABSTRACT

The objective of this study was to investigate the genetic basis of energy balance (EB) and the potential use of genomic selection to enable EB to be incorporated into selection programs. Energy balance provides an essential link between production and nonproduction traits because both depend on a common source of energy. A small number (527) of Dutch Holstein-Friesian heifers with phenotypes for EB were genotyped. Direct genomic values were predicted for these heifers using a model that included the genotypic information. A polygenic model was also applied to predict estimated breeding values using only pedigree information. A 10-fold cross-validation approach was employed to assess the accuracies of the 2 sets of predicted breeding values by correlating them with phenotypes. Because of the small number of phenotypes, accuracies were relatively low (0.29 for the direct genomic values and 0.21 for the estimated breeding values), where the maximum possible accuracy was the square root of heritability (0.57). Despite this, the genomic model produced breeding values with reliability double that of the breeding values produced by the polygenic model. To increase the accuracy of the genomic breeding values and make it possible to select for EB, measurement and recording of EB would need to improve. The study suggests that it may be possible to select for minimally recorded traits; for instance, those measured on experimental farms, using genomic selection. Overall, the study demonstrated that genomic selection could be used to select for EB, confirming its genetic background.

Key words: energy balance, genomic selection, dairy cow, genetic variation

INTRODUCTION

Due to declining calving performance and conception rates at first service (Royal et al., 2000) many countries have introduced measures of fertility into national selection indices to address declining fertility rates in dairy cattle (Miglior et al., 2005). One explanation for these declining rates is the difference between energy intake and energy usage that occurs during early lactation. This difference is defined as energy balance (EB). Energy balance provides an essential link between production and nonproduction traits because both depend on a common source of energy. This energy must be partitioned efficiently to maintain production levels as well as the animal's ability to remain healthy and fertile. Severe negative energy balance (NEB) during early lactation has been cited as an underlying cause of the negative relationship of health and fertility with production (Butler and Smith, 1989; Jorritsma et al., 2003; Pryce et al., 2004).

Recently, the major focus had been on trying to overcome the NEB problem by modifying the diet during the dry period (Dewhurst et al., 2000; Agenäs et al., 2003; McNamara et al., 2003; Garnsworthy et al., 2008a,b). Other suggested approaches to overcome NEB include varying the length of the dry period (Watters et al., 2009) and the frequency of milking (McNamara et al., 2008). However, estimates of genetic parameters suggest that EB is not only a consequence of a poor match between nutrition and production, but is also genetically induced (Veerkamp, 1998; Veerkamp et al., 2003; Coffey et al., 2004; Friggens et al., 2007). Veerkamp (1998) reviewed the results of different studies that reported genetic correlations for a variety of energy measures and milk yield, with values ranging from -0.05 to -0.91 and heritabilities for energy traits that ranged from 0.19 to 0.69. Coffey et al. (2004) demonstrated that distinct genetic lines responded differently to a range of diets and differed in the time taken to return to positive EB. Similarly, Friggens et al. (2007) concluded that variability among animals on a stable nutritional diet could not be accounted for by

Received November 20, 2009.

Accepted February 5, 2010.

¹Corresponding author: klara.verbyla@dpi.vic.gov.au

environmental factors and indicated a genetic basis for EB. Thus, an alternative to management approaches may be to select animals that are genetically predisposed to maintain a better EB.

Accounting for EB in selection programs is complicated, because measuring feed intake in progeny testing schemes is not practical. Currently, much attention has been placed on the implementation of genomic selection. Genomic selection uses genomic information to predict and select animals based on their direct genomic values (DGV), predicted directly from SNP information, or their genomically enhanced breeding values, which are calculated by blending the DGV with conventional proofs. Genomic prediction simultaneously estimates the marker effects and creates an equation to predict DGV for genotyped selection candidates, including (young) animals that do not have phenotypic records. The recent implementation of genomic selection has been shown to increase both selection accuracy and genetic gain over traditional selection methods (Hayes et al., 2009).

In this study, we examined whether genomic prediction could be used to estimate DGV for EB using a small Dutch experimental farm data set. Our objective was to demonstrate the genetic basis of EB and the potential use of genomic selection to facilitate inclusion of EB in selection programs.

MATERIALS AND METHODS

Data

Data on 613 Holstein-Friesian heifers born between 1990 and 1997 were collected during the first 15 wk of lactation; 450 cows participated in the breeding program of CRV (Arnhem, the Netherlands) and 163 cows originated from the experimental farm (t'Gen, the Netherlands). All animals were housed together on a single farm under the same environmental and management influences. All cows were fed ad libitum. Live weight, feed intake, and milk yield were measured on 565 of the animals. Milk samples were taken on a fixed day of the week for measurement of fat, protein, and lactose yields. Feed intake was recorded daily using automated feed intake units. Live weight was recorded once a week. Energy balance (MJ/d) was calculated using the method described in Veerkamp et al. (2000) as the difference between energy intake and the calculated energy requirements for milk, fat, and protein yields, and maintenance costs as a function of live weight. Energy balance values across wk 2 to 15 were averaged, where possible, to give an overall EB phenotype. Comprehensive details on the data used can be found

in Veerkamp et al. (2000). Raw EB phenotypes were preadjusted for year-season of calving and age at calving (linear, quadratic) using ASReml (Gilmour et al., 2006), because their inclusion was not feasible in the final model because of software limitations. The residuals from this analysis were used as the EB phenotypes for the prediction of breeding values.

In total, 588 of the 613 heifers had known pedigrees and these were genotyped using the Illumina 50K SNP panel (54,001 SNP in total; Illumina, San Diego, CA). The quality control criteria for selecting the final set of SNP were a call rate of >90%, a GenCall score >0.2, and a GenTrain score >0.55 (Illumina descriptive statistics relating to genotype quality), a minor allele frequency of >2.5%, and a lack of deviation from Hardy-Weinberg equilibrium, $\chi^2 < 600$ (Wiggans et al., 2009). Animals with greater than 5% missing SNP genotypes were removed. Non-Mendelian error checks identified genotypes of daughters that were inconsistent with their dams. A further, more comprehensive pedigree check was performed by comparing the coefficients of the additive genetic relationship matrix and the genomic relationship matrix (G matrix) calculated via the first method described in VanRaden (2008). This enabled inconsistencies between recorded half and full siblings to be examined. Animals with many inconsistencies between the pedigree and G matrix were removed. After all editing steps, 43,011 SNP and 548 animals were retained. Of these 548 animals, 527 had phenotypes for EB.

Statistical Analysis

Models. Two models using Gibbs sampling were applied to estimate additive breeding values. One model included the available SNP information. This model used stochastic search variable selection (SSVS; George and McCulloch, 1993), which introduces an indicator variable I_j that determines whether SNP j has a large significant effect or whether the effect is insignificant and is therefore scaled back toward zero. The indicator variable for each locus j has a Bernoulli prior distribution:

$$I_j \sim \text{Bernoulli}(p).$$

The prior probability p is chosen to reflect the information available on how many QTL affect the trait of interest. It can be quantified as the number of SNP expected to be linked to a QTL divided by the total number of SNP. For a complex trait such as EB, it was assumed that about 1% of the SNP were linked to a QTL ($p = 0.01$).

The SNP model can be expressed as follows:

$$\mathbf{y} = \mathbf{1}_n \mu + \sum_{j=1}^m \left(\mathbf{X}_j \left(\mathbf{q}_j v_j \right) \right) + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where $\mathbf{y}$ is the vector of phenotypes of the trait being analyzed for all n individuals, $\mathbf{1}_n$ is a vector of ones of length n , μ is the mean, m is the number of SNP markers, $\mathbf{X}_j$ is the $(n \times k)$ design matrix containing the information on the possible k alleles at the j th marker for all individuals (where $x_{jk} = 0, 1, 2$, having 0, 1, or 2 copies of the k th allele, respectively), $\mathbf{q}_j$ is the vector $(k \times 1)$ containing the effects of all k possible alleles at locus j where q_{jk} are drawn from a standard normal distribution $N(0,1)$, v_j is the standard deviation of the allelic effects at locus j and is dependent on whether the locus effect is considered significant using the indicator variable, $\mathbf{u}$ is the vector of random additive polygenic effects of length n ($\mathbf{Z}$ is the associated design matrix) and is assumed to be normally distributed, $\mathbf{u} \sim N(0, \sigma_u^2 \mathbf{A})$, where $\mathbf{A}$ is the pedigree-derived additive genetic relationship matrix, and $\mathbf{e}$ is the residual error also assumed to be normally distributed, $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$, where $\mathbf{I}$ is the $n \times n$ identity matrix.

Note that the allele substitution effect of a locus j can be calculated from the estimated effects as $a_j = (q_{j1} - q_{j2})v_j$, where q_{j1} (q_{j2}) is the effect of allele 1 (2) at locus j . For the full specification of the priors used and an alternative formulation of the model, see Calus et al. (2008) and Meuwissen and Goddard (2004). The DGV were calculated as the sum of estimated SNP effects and the polygenic effect:

$$DGV = \sum_{j=1}^p \left(\mathbf{X}_{ij} \left(\hat{q}_j v_j \right) \right) + \hat{u}_i.$$

The second model used was a simple additive polygenic model: $\mathbf{y} = \mathbf{1}_n \mu + \mathbf{Z}\mathbf{u} + \mathbf{e}$, where the EBV calculated by this model were the estimated polygenic effects for each animal ($EBV = \hat{u}_i$). Both models were run for 10,000 iterations to ensure convergence, with the first 1,000 iterations used as burn in.

Validation. Because of the small size of the data set, a 10-fold cross validation approach was carried out to assess the accuracy of predicted breeding values. The data set was randomly partitioned into 10 subsets each containing 10% of the data. Each subset was retained once as the validation data set and the remaining 9 became the reference sets. Results from the reference sets

were then used to predict breeding values of animals in the validation set. Accordingly, each animal appeared only once in a validation set and had only one predicted DGV.

The DGV and EBV were assessed using accuracy, $\mathbf{r}_{y\hat{g}}$, defined as the Pearson correlation of the predicted breeding values (DGV or EBV) ($\hat{g}$) and the phenotypes (y). The maximum achievable accuracy due to the correlation being between phenotypes and predicted breeding values was equal to the square root of the heritability of the phenotypes. The observed heritability for EB was estimated by fitting a model with year-season and age at calving (linear and quadratic regression) as the fixed effects and a random animal effect (a). The random animal effect was assumed normally distributed, $a \sim N(0, \sigma_a^2 \mathbf{G})$, where σ_a^2 was the additive genetic variance and $\mathbf{G}$ was the genomic relationship matrix calculated via the first method described in VanRaden (2008). Deriving the heritability this way has been shown to produce estimates much closer to the true value than using the pedigree-based relationship matrix (Hayes and Goddard, 2008).

Because no daughter yield deviations (**DYD**) or reliable breeding values were available, the predicted breeding values (DGV and EBV) were compared with phenotypes. Most studies estimating accuracies of DGV use DYD or reliable EBV predicted for proven bulls and consequently report accuracies of selection ($\mathbf{r}_{y\hat{g}}$) and reliabilities ($\mathbf{r}_{\hat{g}\hat{g}}^2$) that compare DGV and the closest estimate of the true breeding values (g). Thus, for these studies the accuracy of selection was calculated (Daetwyler et al., 2008; Goddard, 2009) as

$$\mathbf{r}_{y\hat{g}} = \sqrt{\frac{\lambda h^2}{\lambda h^2 + 1}},$$

$$\text{and } \lambda = \frac{n_p}{n_G}, \quad [1]$$

where h^2 is the observed heritability, n_p is the number of phenotypic records, and n_G is the number of effective QTL or chromosome segments. This function can be used when the accuracy is calculated using the correlation between the predicted DGV and phenotypes ($\mathbf{r}_{y\hat{g}}$). Falconer and Mackay (1996) state that $\mathbf{r}_{\hat{g}\hat{g}} = \sigma_{\hat{g}} / \sigma_g$. The accuracy between DGV and phenotypes can be similarly expressed as $\mathbf{r}_{y\hat{g}} = \sigma_{\hat{g}} / \sigma_y$. And $\mathbf{r}_{y\hat{g}}$ can be denoted as:

$$\mathbf{r}_{y\hat{g}} = \frac{\sigma_{\hat{g}}}{\sigma_g} \times \frac{\sigma_g}{\sigma_y},$$

which can be rewritten as $\mathbf{r}_{y\hat{g}} = \mathbf{r}_{g\hat{g}} \times \sqrt{h^2}$, and when combined with [1] gives

$$\mathbf{r}_{y\hat{g}} = \sqrt{\frac{\lambda h^4}{\lambda h^2 + 1}}. \quad [2]$$

Hence, $\mathbf{r}_{y\hat{g}}$ can also be transformed into $\mathbf{r}_{g\hat{g}}$. The accuracy, $\mathbf{r}_{y\hat{g}}$, was used to calculate the number of QTL affecting EB and the number of records needed to improve the accuracy of the predicted DGV.

RESULTS

The pedigree check step for data quality control proved a very effective additional measure to identify any animal that had an incorrectly recorded pedigree or where an animal may have been misidentified. It allowed checking of half-sibling and full-sibling relationships, which is not possible using non-Mendelian checking. Figure 1 effectively illustrates the additional information contained in the SNP data about the relatedness of the animals. This is most obviously shown by the monozygotic twins that have a marker relationship

Figure 1. Comparison of the coefficients of the additive relationship matrix (pedigree relationship) and the coefficients of the genomic relationship matrix (markers relationship).

Table 1. Accuracies and reliabilities¹ of direct genomic values (DGV) and EBV

Model ²	$\mathbf{r}_{y\hat{g}}$	$\mathbf{r}_{g\hat{g}}$	$\mathbf{r}_{y\hat{g}}^2$	$\mathbf{r}_{g\hat{g}}^2$
DGV	0.294	0.516	0.086	0.265
EBV	0.211	0.370	0.044	0.135

¹ $\mathbf{r}_{y\hat{g}}$ = Pearson correlation between the predicted breeding values ($\hat{g}$) and the phenotypes (y); $\mathbf{r}_{g\hat{g}}$ = accuracy of selection [comparing the predicted breeding values ($\hat{g}$) and the true breeding values (g)]; $\mathbf{r}_{y\hat{g}}^2$ = reliability of the predicted phenotypes; and $\mathbf{r}_{g\hat{g}}^2$ = reliability of the predicted breeding values.

²DGV was predicted using the model that included both the SNP and polygenic effects, and EBV was predicted using the model that included only the polygenic effect.

of 1 (because of identical DNA) but are recorded as full sibs in the pedigree. The negative marker relationships are due to the method used to calculate the G matrix, which ideally uses the allele frequencies that were present in the base population (VanRaden, 2008). However, because the frequencies in the base population were unknown, the G matrix was calculated using the allele frequencies in the available highly selected population resulting in negative marker relationships.

The accuracies ($\mathbf{r}_{y\hat{g}}$) of predicting phenotypes for the 2 models and the $\mathbf{r}_{y\hat{g}}^2$ are shown in Table 1. Transformed values, using [1] to give the accuracies of selection ($\mathbf{r}_{g\hat{g}}$) and reliabilities ($\mathbf{r}_{g\hat{g}}^2$), are also shown. The model that included the SNP information yielded an overall accuracy of 0.29, which was higher than the overall accuracy of 0.21 produced by the polygenic model.

The calculated reliability ($\mathbf{r}_{g\hat{g}}^2$) of the DGV is double that of the EBV produced by the polygenic model. This implies that the DGV explained twice as much variation as the EBV, which is also illustrated by the range of breeding values (see Figure 2). The predicted DGV and EBV were positively correlated with a value of 0.70.

The heritability for EB was estimated separately, as described earlier, with a moderate value of 0.325 (SE = 0.12). This value was then used with the accuracy $\mathbf{r}_{y\hat{g}}$ and number of phenotypic records to predict the number of effective QTL for EB. A total of 472 effective QTL were predicted. Figure 3 shows a plot of $\mathbf{r}_{y\hat{g}}^2$ ($\mathbf{r}_{g\hat{g}}^2$ is provided for comparison on the second y-axis) against the number of effective QTL for differing heritabilities where the number of records was kept constant at the available number of 527. This shows the effect that the number of effective QTL would have on the expected accuracies and reliabilities of the DGV. It demonstrated

Figure 2. Histogram of direct genomic value (DGV) and EBV; black bars represent the EBV predicted by the polygenic model, and gray bars represent the DGV predicted by the model including the SNP information.

that the greater the number of QTL affecting the trait, the lower the expected accuracy and reliability. This is because of a lack of information available in the limited number of phenotypes to be able to accurately estimate large numbers of QTL effects. Figure 3 also illustrates that this reduction in reliability, as the number of effective

Figure 3. Accuracy of prediction versus the number of effective QTL, where the number of records is fixed to the number used in this study (527); $r^2(y, \hat{g})$ is the squared correlation between the phenotypes and the predicted direct genomic values (DGV, characterized in the text as $\mathbf{r}_{y\hat{g}}^2$); $r^2(g, \hat{g})$ is the estimated reliability between the true breeding value and the predicted DGV (characterized in the text as $\mathbf{r}_{g\hat{g}}^2$).

tive QTL increases, is more gradual for higher heritabilities.

The number of total records needed to improve the accuracy was also investigated and results are shown in Figure 4. The heritability was set at the observed value for EB (0.325). It is evident from Figure 4 that the number of effective QTL has a significant effect on the number of records needed to improve the accuracy. The greater the number of effective QTL, the larger the number of phenotypic records required to reach higher accuracies and reliabilities. We predicted that 5,818 records with phenotype and genotype information would be needed for an $\mathbf{r}_{y\hat{g}}^2$ of 0.24 ($\mathbf{r}_{g\hat{g}}^2$ of 0.80) for EB with the predicted 472 effective QTL.

DISCUSSION

The objective of this study was to demonstrate the genetic basis of EB and show that it could be incorporated into selection programs using genomic selection based on a limited reference population. Energy balance is a minimally recorded trait and consequently only a small number of phenotypic records was available. Despite the limitation on available data, genomic prediction was able to produce accuracies greater than a traditional polygenic model. Thus, the results indicated that EB could be estimated using genomic prediction. The low accuracy gained can be explained as a direct result of the small number of phenotypic records and the moderate heritability found for this trait. The heritability calculated with this data set was consistent with results of other studies (Veerkamp,

Figure 4. Accuracy of prediction versus the number of records for a fixed heritability of 0.325; $r^2(y, \hat{g})$ is the squared correlation between the phenotypes and the predicted direct genomic values (DGV, characterized in the text as $\mathbf{r}_{y\hat{g}}^2$); $r^2(g, \hat{g})$ is the estimated reliability between the true breeding value and the predicted DGV (characterized in the text as $\mathbf{r}_{g\hat{g}}^2$).

1998; Huttman et al., 2009). To consider including EB in breeding schemes, higher accuracies than found here would be necessary. This increase in accuracy could be facilitated through an increase in the heritability of the trait or an increase in the number of phenotypic records. One way to increase the heritability would be to standardize the environmental conditions to reduce nongenetic differences between animals, but this may be difficult to do in practice. An alternative approach to increase the heritability of phenotypes would be to use deregressed breeding values or DYD of proven bulls as phenotypes, based on EB records of many daughters. This allows for an increase in the accuracy while keeping the number of genotyped animals constant. This scenario would not lead to any additional genotyping costs because most bulls may already be genotyped as part of reference populations for other breeding goal traits. Note, however, that this approach may still be more costly because of the (much) higher number of recorded EB phenotypes that would be needed. An increase in the number of available records would also allow for an increase in the accuracy of predicted DGV as indicated in other studies (Hayes and Goddard, 2008; Goddard, 2009). The required increase could occur only if the measurement and recording of EB improved.

Because of infrequent recording of EB, a seemingly obvious solution would be to immediately select for a widely recorded trait such as BCS to reduce NEB indirectly. The problem with using BCS is that after the first 60 DIM, the genetic correlations between EB and BCS decrease markedly (Huttman et al., 2009). However, until the recording of EB increases to useful levels, BCS does provide a viable option to attempt to select animals with a better EB. In the future having both EB and BCS phenotypes available should allow for the best prediction of energy partitioning and utilization.

The model used to predict the DGV could also be used for whole-genome association studies. Thus, the produced posterior probabilities of SNP were examined to see if there were any significant associations with EB. Because of the small number of records and large number of SNP, the power of the association study to identify QTL was very low and this was evident. There was no SNP with a high enough posterior probability to be confident that it was linked to a QTL. The prior for the expected number of QTL affecting EB varied but results were consistently low (results not presented). Although the posterior probabilities were low, one SNP had 10-fold higher posterior probabilities than all the other SNP in all analyses regardless of the prior probability used. This SNP is located on BTA21 and is in extremely close proximity to, and appears in association with, the nuclear receptor subfamily 2, group F, member 2 (NR2F2), otherwise known as chicken ovalbumin up-

stream promoter transcription factor II (COUP-TFII); COUP-TFII has been previously reported as playing an essential role in regulating adipogenesis, glucose homeostasis and energy metabolism (Xu et al., 2008; Li et al., 2009). It has also been reported as regulating growth hormone receptor 1A promoter activity (Xu et al., 2004), mediating progesterone, and controlling estrogen levels and thus involved in reproduction (Klinge et al., 1997; Nakshatri et al., 2000; Takamoto et al., 2005; Kurihara et al., 2007; Petit et al., 2007). Although the results of this association study are not conclusive and further validation is required, COUP-TFII appears to be a good candidate gene for EB.

Despite being unable to establish QTL conclusively associated with EB, results of the study allowed an estimation of the number of effective QTL influencing EB. Given the nature and complexity of EB, the number of predicted effective QTL (472) was plausible. The relationships with both production and nonproduction traits mean that numerous genes and pathways could be involved in the variation observed in EB. Previous whole-genome association studies of residual feed intake and other traits related to EB in beef cattle identify between 4 and 120 QTL affecting the traits studied (Barendse et al., 2007; Sherman et al., 2009). These values are significantly lower than the predicted 472, but reflect the power of the studies to detect significant QTL and the number of SNP (which were 2,194 and 8,786 respectively), rather than the true number of effective QTL. An increase in the number of phenotypic records would also allow genome-wide association studies for EB in dairy cattle to identify possible candidate genes affecting the trait and would provide a better idea of the effective number of QTL.

The ability to select and include EB in selection indices may indirectly increase the genetic gain for fertility traits. The interval between calving and start of luteal activity (**C-LA**) has been demonstrated to be an indicator of fertility during later lactation (Darwash et al., 1999; van der Lende et al., 2004; Petersson et al., 2007). Veerkamp et al. (2000) reported genetic correlations between EB and C-LA of -0.60 (and -0.49 for C-LA adjusted for milk, fat, and protein). A moderate to high genetic correlation similar to what was previously reported would mean that genetic gain for EB should also result in improved fertility. For example, if a bull had 25 daughters, the accuracy of selection for the bull's EBV would be 0.40 for fertility (assuming a heritability of 0.03), whereas the accuracy of selection for the bull's EBV would be 0.83 for EB. Thus, given a genetic correlation of -0.5 , the accuracy of selection for fertility using EB would be 0.41. Consequently, for bulls with this number of daughters or fewer, selection using EB would result in greater genetic gain for fertil-

ity compared with selecting for fertility itself. However, as the number of offspring per bull increases beyond 25, the benefit of using EB rather than fertility is lost, such that selection for fertility itself will produce better genetic gains. Thus, the use of EB in selection indices, in addition to fertility, may prove beneficial and result in increased genetic gain for fertility.

In addition to the possible benefits of improved fertility, EB could be used with feed intake data to select animals for feed efficiency (Veerkamp, 1998) or to reduce methane emission (Hegarty et al., 2007). Improving feed efficiency could be economically desirable because feed costs contribute the greatest proportion to production costs (Simm et al., 1994). However, feed efficiency data alone cannot distinguish whether the energy is used for production or maintenance. This may result in selection of animals with low intake and high yield that, consequently, have problems related to high NEB. Thus, NEB and improved feed efficiency (or intake) data should be considered simultaneously to effectively reduce the feeding costs while not having detrimental effects on the animals' health and fertility.

Many other traits including several fertility and reproduction traits such as milk progesterone profiles and milk quality traits are difficult to record. Accounting for these traits, like EB, in selection has been complicated, because measuring them in progeny testing schemes is not practical. This study demonstrates that it is possible for such traits with similar heritabilities and expected number of QTL to produce DGV with accuracies >0.8 when there are more than approximately 2,600 (2,581 predicted for EB) phenotypic records available for use as the reference population. This demonstrates that it is possible to select for these traits using genomic selection by combining data from experimental and nucleus herds, where individually there are a limited numbers of raw phenotypic records.

The statistical approaches used in this study are generally accepted as appropriate for genomic prediction. Genomic prediction is often performed using a 2-step procedure in which the input phenotypes are precorrected so that the model predicting the DGV includes only the mean, polygenic, and SNP effects. Because precorrection will always introduce a new source of error, our preference would be to include all fixed effects in the models used to predict the breeding values. There is ongoing development of the genomic prediction program used, to allow the inclusion of multiple discrete and continuous fixed effects in a single model.

CONCLUSIONS

The use of SNP information to predict DGV is shown to explain variation among the EB of animals,

confirming the genetic background of EB. The use of SNP information showed an increase in the accuracy of selection for EB over the simple polygenic model. However, the extent of recording would need to be improved to increase the accuracy. In the future, selection for EB could be performed using genomic selection, which could provide a valuable tool in finding a balance between production and nonproduction traits.

ACKNOWLEDGMENTS

The authors acknowledge funding for the study from the Dutch Dairy Board (PZ; Zoetermeer, the Netherlands) and Senter Novem (The Hague, the Netherlands). The study was also performed as part of the RobustMilk and SABRETRAIN projects. The RobustMilk project is financially supported by the European Commission under the Seventh Research Framework Programme, Grant Agreement KBBE-211708. The first author was funded by the SABRETRAIN Marie Curie Host Fellowships for Early Stage Research Training, as part of the 6th Framework Programme of the European Commission. This publication represents the views of the authors, not the European Commission, and the Commission is not liable for any use that may be made of the information

REFERENCES

- Agenäs, S., E. Burstedt, and K. Holtenius. 2003. Effects of feeding intensity during the dry period. 1. Feed intake, body weight, and milk production. *J. Dairy Sci.* 86:870–882.
- Barendse, W., A. Reverter, R. J. Bunch, B. E. Harrison, W. Barris, and M. B. Thomas. 2007. A validated whole-genome association study of efficient food conversion in cattle. *Genetics* 176:1893–1905.
- Butler, W. R., and R. D. Smith. 1989. Interrelationships between energy-balance and postpartum reproduction function in dairy cattle. *J. Dairy Sci.* 72:767–783.
- Calus, M. P. L., T. H. E. Meuwissen, A. P. W. de Roos, and R. F. Veerkamp. 2008. Accuracy of genomic selection using different methods to define haplotypes. *Genetics* 178:553–561.
- Coffey, M. P., G. Simm, J. D. Oldham, W. G. Hill, and S. Brotherstone. 2004. Genotype and diet effects on energy balance in the first three lactations of dairy cows. *J. Dairy Sci.* 87:4318–4326.
- Daetwyler, H. D., B. Villanueva, and J. A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* 3:e3395.
- Darwash, A. O., G. E. Lamming, and J. A. Woolliams. 1999. The potential for identifying heritable endocrine parameters associated with fertility in post-partum dairy cows. *Anim. Sci.* 68:333–347.
- Dewhurst, R. J., J. M. Moorby, M. S. Dhanoa, R. T. Evans, and W. J. Fisher. 2000. Effects of altering energy and protein supply to dairy cows during the dry period. 1. Intake, body condition, and milk production. *J. Dairy Sci.* 83:1782–1794.
- Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction to Quantitative Genetics*. 4th ed. Longmans Green, Harlow, UK.
- Friggens, N. C., P. Berg, P. Theilgaard, I. R. Korsgaard, K. L. Ingvarsen, P. Lovendahl, and J. Jensen. 2007. Breed and parity effects on energy balance profiles through lactation: Evidence of genetically driven body energy change. *J. Dairy Sci.* 90:5291–5305.

- Garnsworthy, P. C., A. Lock, G. E. Mann, K. D. Sinclair, and R. Webb. 2008a. Nutrition, metabolism, and fertility in dairy cows: 1. Dietary energy source and ovarian function. *J. Dairy Sci.* 91:3814–3823.
- Garnsworthy, P. C., A. Lock, G. E. Mann, K. D. Sinclair, and R. Webb. 2008b. Nutrition, metabolism, and fertility in dairy cows: 2. Dietary fatty acids and ovarian function. *J. Dairy Sci.* 91:3824–3833.
- George, E. I., and R. E. McCulloch. 1993. Variable selection via Gibbs sampling. *J. Am. Stat. Assoc.* 88:881–889.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2006. ASREML Program User Manual. 2nd ed. VSN International Ltd., Hemel Hempstead, UK.
- Goddard, M. 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica* 136:245–257.
- Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92:433–443.
- Hayes, B. J., and M. E. Goddard. 2008. Technical note: Prediction of breeding values using marker-derived relationship matrices. *J. Anim. Sci.* 86:2089–2092.
- Hegarty, R. S., J. P. Goopy, R. M. Herd, and B. McCorkell. 2007. Cattle selected for lower residual feed intake have reduced daily methane production. *J. Anim. Sci.* 85:1479–1486.
- Huttman, H., E. Stamer, W. Junge, G. Thaller, and E. Kalm. 2009. Analysis of feed intake and energy balance of high-yielding first lactating Holstein cows with fixed and random regression models. *Animal* 3:181–188.
- Jorritsma, R., T. Wensing, T. A. M. Kruip, P. Vos, and J. Noordhuizen. 2003. Metabolic changes in early lactation and impaired reproductive performance in dairy cows. *Vet. Res.* 34:11–26.
- Klinge, C. M., B. F. Silver, M. D. Driscoll, G. Sathya, R. A. Bambara, and R. Hilf. 1997. Chicken ovalbumin upstream promoter transcription factor interacts with estrogen receptor, binds to estrogen response elements and half-sites, and inhibits estrogen-induced gene expression. *J. Biol. Chem.* 272:31465–31474.
- Kurihara, I., D. K. Lee, F. G. Petit, J. Jeong, K. Lee, J. P. Lydon, F. J. DeMayo, M. J. Tsai, and S. Y. Tsai. 2007. COUP-TFII mediates progesterone regulation of uterine implantation by controlling ER activity. *PLoS Genet.* 3:1053–1064.
- Li, L. P., X. Xie, J. Qin, G. S. Jeha, P. K. Saha, J. Yan, C. M. Haueter, L. Chan, S. Y. Tsai, and M. J. Tsai. 2009. The nuclear orphan receptor COUP-TFII plays an essential role in adipogenesis, glucose homeostasis, and energy metabolism. *Cell Metab.* 9:77–87.
- McNamara, S., J. J. Murphy, F. P. O'Mara, M. Rath, and J. F. Mee. 2008. Effect of milking frequency in early lactation on energy metabolism, milk production and reproductive performance of dairy cows. *Livest. Sci.* 117:70–78.
- McNamara, S., J. J. Murphy, M. Rath, and F. P. O'Mara. 2003. Effects of different transition diets on energy balance, blood metabolites and reproductive performance in dairy cows. *Livest. Prod. Sci.* 84:195–206.
- Meuwissen, T. H. E., and M. E. Goddard. 2004. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genet. Sel. Evol.* 36:261–279.
- Miglior, F., B. L. Muir, and B. J. Van Doormaal. 2005. Selection indices in Holstein cattle of various countries. *J. Dairy Sci.* 88:1255–1263.
- Nakshatri, H., M. S. Mendonca, P. Bhat-Nakshatri, N. M. Patel, R. J. Goulet, and K. Cornetta. 2000. The orphan receptor COUP-TFII regulates G2/M progression of breast cancer cells by modulating the expression/activity of p21(WAF1/CIP1), cyclin D1, and cdk2. *Biochem. Biophys. Res. Commun.* 270:1144–1153.
- Petersson, K. J., B. Berglund, E. Strandberg, H. Gustafsson, A. P. F. Flint, J. A. Woolliams, and M. D. Royal. 2007. Genetic analysis of postpartum measures of luteal activity in dairy cows. *J. Dairy Sci.* 90:427–434.
- Petit, F. G., S. P. Jamin, I. Kurihara, R. R. Behringer, F. J. DeMayo, M. J. Tsai, and S. Y. Tsai. 2007. Deletion of the orphan nuclear receptor COUP-TFII in uterus leads to placental deficiency. *Proc. Natl. Acad. Sci. USA* 104:6293–6298.
- Pryce, J. E., M. D. Royal, P. C. Garnsworthy, and I. L. Mao. 2004. Fertility in the high-producing dairy cow. *Livest. Prod. Sci.* 86:125–135.
- Royal, M. D., A. O. Darwash, A. P. E. Flint, R. Webb, J. A. Woolliams, and G. E. Lamming. 2000. Declining fertility in dairy cattle: Changes in traditional and endocrine parameters of fertility. *Anim. Sci.* 70:487–501.
- Sherman, E. L., J. D. Nkrumah, C. Li, R. Bartusiak, B. Murdoch, and S. S. Moore. 2009. Fine mapping quantitative trait loci for feed intake and feed efficiency in beef cattle. *J. Anim. Sci.* 87:37–45.
- Simm, G., R. F. Veerkamp, and P. Persaud. 1994. The economic performance of dairy cows of different predicted genetic merit for milk solids production. *Anim. Prod.* 58:313–320.
- Takamoto, N., I. Kurihara, K. Lee, F. J. DeMayo, M. J. Tsai, and S. Y. Tsai. 2005. Haploinsufficiency of chicken ovalbumin upstream promoter transcription factor II in female reproduction. *Mol. Endocrinol.* 19:2299–2308.
- van der Lende, T., L. Kaal, R. M. G. Roelofs, R. F. Veerkamp, C. Schrooten, and H. Bovenhuis. 2004. Infrequent milk progesterone measurements in daughters enable bull selection for cow fertility. *J. Dairy Sci.* 87:3953–3957.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423.
- Veerkamp, R. F. 1998. Selection for economic efficiency of dairy cattle using information on live weight and feed intake: A review. *J. Dairy Sci.* 81:1109–1119.
- Veerkamp, R. F., B. Beerda, and T. van der Lende. 2003. Effects of genetic selection for milk yield on energy balance, levels of hormones, and metabolites in lactating cattle, and possible links to reduced fertility. *Livest. Prod. Sci.* 83:257–275.
- Veerkamp, R. F., J. K. Oldenbroek, H. J. Van Der Gaast, and J. H. J. Van Der Werf. 2000. Genetic correlation between days until start of luteal activity and milk yield, energy balance, and live weights. *J. Dairy Sci.* 83:577–583.
- Watters, R. D., M. C. Wiltbank, J. N. Guenther, A. E. Brickner, R. Rastani, P. M. Fricke, and R. R. Grummer. 2009. Effect of dry period length on reproduction during the subsequent lactation. *J. Dairy Sci.* 92:3081–3090.
- Wiggans, G. R., T. S. Sonstegard, P. M. VanRaden, L. K. Matukumalli, R. D. Schnabel, J. F. Taylor, F. S. Schenkel, and C. P. Van Tassell. 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. *J. Dairy Sci.* 92:3431–3436.
- Xu, Q., N. Walther, and H. Jiang. 2004. Chicken ovalbumin upstream promoter transcription factor II (COUP-TFII) and hepatocyte nuclear factor 4 gamma (HNF-4 gamma) and HNF-4 alpha regulate the bovine growth hormone receptor 1A promoter through a common DNA element. *J. Mol. Endocrinol.* 32:947–961.
- Xu, Z., S. Yu, C. H. Hsu, J. Eguchi, and E. D. Rosen. 2008. The orphan nuclear receptor chicken ovalbumin upstream promoter-transcription factor II is a critical regulator of adipogenesis. *Proc. Natl. Acad. Sci. USA* 105:2421–2426.