Explaining Genetic Variation in Energy Balance Using High Density SNP Information

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

Introduction

Accounting for energy balance (EB) in selection programs has been complicated, since measuring feed intake in progeny testing schemes is not practical. Currently, much attention has been placed on the implementation of genomic selection in animal breeding schemes. 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 (GEBV) 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.

Material and methods

Available data. Data on 613 Holstein-Friesian heifers born between 1990 and 1997 were collected during the first 15 wk of lactation. All cows were fed ad libitum. Live weight, feed intake, and milk yield were measured on 565 animals. Feed intake was recorded daily using automated feed intake units. Live weight and milk yield were 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 to give an overall EB phenotype. More comprehensive details on the data used can be found in Veerkamp et al. (2000). Raw EB phenotypes were pre-adjusted for year-season of calving and age at calving (linear, quadratic) using ASReml (Gilmour et al. (2006)), since their inclusion was not feasible in the final model due to software limitations. The residuals from this analysis were used as the EB phenotypes for the prediction of the breeding values.

[*] Animal Breeding and Genomics Centre, Wageningen UR Livestock Research,. PO Box 65. NL-8200 AB Lelystad
[†] Melbourne School of land and Environment, The University of Melbourne, Parkville, VIC. 3001, Australia
[§] The Cooperative Research Centre for Beef Genetic Technologies, University of New England, Armidale, NSW 2351, Australia
[‡] Department of Primary Industries, 1 Park Drive, Bundoora, VIC. 3083, Australia

From the 613 heifers, 588 had known pedigree and these were genotyped using the Illumina 50K SNP panel (54,001 SNP in total). The quality control criteria for selecting the final set of SNP were; a call rate of over 90%, a GenCall score >0.2 and a GenTrain score >0.55, 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. After all editing steps, in total, 43,011 SNP and 548 animals were retained. Of these 548 animals, 527 had phenotypes for EB.

Statistical analyses. Two models using Gibbs sampling were applied to estimate additive breeding values. One model included the available SNP information, resulting in a DGV for each animal. The second model used was a simple additive polygenic model, where the EBV calculated by this model were the estimated polygenic effect for each animal. Both models were run for 10,000 iterations to ensure convergence with the first 1000 iterations used as burn in. A 10-fold cross validation approach was carried out, such that the data was randomly partitioned into 10 subsets. Each subset was retained once as the validation dataset and the remaining 9 sets were used to predict the GEBV of those animals in the validation set. The model described in Calus et al. (2008) was used to predict the GEBV. The GEBV were calculated as the sum of the estimated SNP effects and the polygenic effect. The same data subsets and approach were used with a simple polygenic model excluding the SNP information for comparison. The GEBV were assessed using accuracy $r_{y\hat{g}}$ of the predicted GEBV ($\hat{g}$) when compared with the phenotypes (y) and thus the $r^2_{y\hat{g}}$. The accuracy of selection ($r_{g\hat{g}}$) when comparing the true breeding values (g) and GEBV has been reported to be a function of the heritability, the number of phenotypic records and the number of effective QTL (Daetwyler et al. (2008)). This function was adapted for use with the accuracy when comparing phenotypes and GEBV.

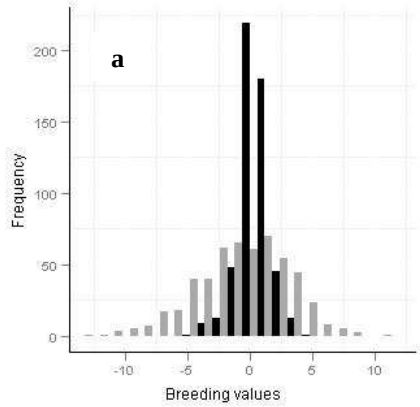
Results and discussion

The model including the SNP information yielded an overall accuracy of 0.294 and thus an $r^2_{y\hat{g}}$ of 0.086, when comparing the phenotypes and GEBV in the combined validation sets (Table 1). For the model excluding the SNP information with only the polygenic effect an overall accuracy of 0.211 and $r^2_{y\hat{g}}$ of 0.044 was found.

Table 1: Accuracies (r_{ij}) and reliabilities (r^2_{ij}) for direct genomic values and estimated breeding values where y is the phenotype (energy balance), g is true breeding value and $\hat{g}$ is the predicted breeding value

	$r_{y\hat{g}}$	$r_{g\hat{g}}$	$r^2_{y\hat{g}}$	$r^2_{g\hat{g}}$
Direct Genomic Value (DGV)	0.294	0.516	0.086	0.265
Estimated Breeding Values (EBV)	0.211	0.370	0.044	0.135

The calculated reliability ($r^2_{g\bar{g}}$ =0.265) of the model including SNP information was double that of the EBV produced by the polygenic model ($r^2_{g\bar{g}}$ =0.135). This implied that the model including SNP information explained twice as much variation than the polygenic model, which is illustrated also by the range of the breeding values (see Figure 1a). 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 can 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 of 0.325 calculated with this data set was consistent with results of other studies (Huttmann et al. (2009); Veerkamp (1998)).



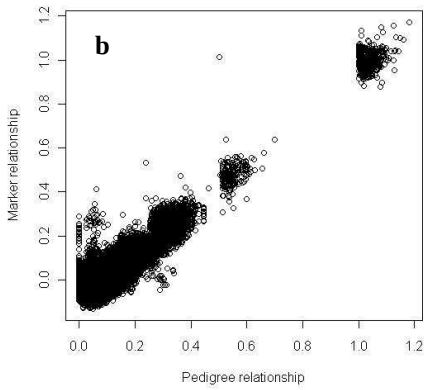


Figure 1a: Histogram of direct genomic values (DGV) and estimated breeding values (EBV), (■) represents the EBV predicted by the polygenic model and (■) represents the DGV predicted by the model including the SNP information
Figure 1b: Comparison of coefficients of the polygenic model (pedigree relationship) and coefficients of the model including SNP information (marker relationship)

The pedigree check step for data quality control proved a very effectual 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 that is not possible using non-Mendelian checking. Figure 1b 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 of 1 but are recorded as full sibs in the pedigree.

The model using the SNP information 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. Due to the small number of records and large number of SNP, the power of the association study to identify QTL was

very low. There was no SNP with a high enough posterior probability to be confident that it was linked to a QTL. Despite being unable to conclusively establish QTL associated with EB, results of the study allowed the 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 non-production traits means that potentially numerous genes and pathways could be involved in the variation observed in EB. 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 EB and would provide a better idea of the effective number of QTL. The ability to select and include EB in selection indexes may indirectly increase the genetic gain for fertility traits. Veerkamp et al. (2000) reported genetic correlations between EB and fertility (i.e. interval between calving and start of luteal activity) of -0.60. This moderate to high genetic correlation implies that genetic gain for EB should also result in improved fertility.

Conclusions

The use of SNP information to predict DGV is shown to explain variation between the EB of animals, confirming the genetic background of EB. The use of SNP information showed an increase in the accuracy of prediction 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 non-production traits (e.g. fertility).

Acknowledgements

The authors acknowledge funding for the study from the Dutch Dairy Board (PZ) and Senter Novem. 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. 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

Calus, M.P.L., Meuwissen, T.H.E., de Roos, A.P.W. *et al.* (2008). *Genetics*, 178:553-561.

Daetwyler, H.D., Villanueva, B., and Woolliams, J.A. (2008). *PLOS ONE*, 3:e3395.

Gilmour A.R., Gogel, B.J., Cullis, B.R. *et al.* (2006). *ASREML User Guide.*

Hayes, B.J., Bowman, P.J., Chamberlain, A.J. *et al.* (2009). *J. Dairy Sci.*, 92:433-443.

Huttmann, H., Stamer, E., Junge, W. *et al.* (2009). *Animal*, 3:181-188.

Veerkamp, R.F. (1998). *J. Dairy Sci.*, 81:1109-1119.

Veerkamp, R.F., Oldenbroek, J.K., Van der Gaast, H.J. *et al.* (2000). *J. Dairy Sci.*, 83:577-583.