Proc. Assoc. Advmt. Anim. Breed. Genet. 22:153-156

INCREASING THE ACCURACY OF GENOMIC PREDICTION FOR RESIDUAL FEED INTAKE IN DAIRY CATTLE BY USING SNPs ASSOCIATED WITH RFI IN BEEF CATTLE

M. Khansefid[1,2], J.E. Pryce[2,3] and M.E. Goddard[1,2]

[1] Department of Agriculture and Food Systems, The University of Melbourne, Melbourne, Victoria 3010, Australia
[2] Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Victoria 3083, Australia
[3] School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia

SUMMARY

Residual feed intake (RFI) is a measure of the efficiency of animals in converting feed to products. Improving RFI in dairy cattle can reduce the costs of raising heifers and producing milk. However, calculating RFI requires expensive equipment to measure the feed intake for each individual. Since April 2015, a "Feed Saved" breeding value has been available in Australia that combines RFI with maintenance requirements. However, the size of the reference population used for genomic prediction of RFI is comparatively and consequently the accuracy of predictions are modest. To improve the prediction accuracy for RFI, the current reference population consisting of 843 heifers and 236 Australian cows and 954 European cows (357 British and 597 Dutch) was extended by including RFI measurements of 206 Australian cows. Furthermore, information from markers which were associated with RFI in 4,772 beef cattle ($p < 0.001$) was used to construct a genomic relationship matrix (GRM) for dairy cattle. We also compared the use of imputed whole genome sequence (WGS) data with 800K SNP-chip genotypes and 2 methods of calculating a GRM described by VanRaden (2008) and Yang *et al.* (2010). The use of the SNPs from the 800K SNP-chip which were associated with RFI in beef cattle improved the accuracy of genomic estimated breeding values (GEBVs) in dairy cattle. However, the use of imputed WGS data did not improve prediction accuracy, especially when the Yang *et al.* method of calculating the GRM was used. The Yang *et al.* method gives extra weight to rare alleles and these SNPs have low imputation accuracy. So it is likely that errors in imputation affect the results when using WGS and this effect is magnified when Yang *et al.* is used to construct the GRM. The best model tested was the GRM built using the Yang et *al.* method with SNP-chip genotypes and when extra weight was given to the SNPs associated with RFI in beef cattle. The accuracy of GEBVs for RFI in the best model for heifers and cows were 0.67 and 0.46, respectively.

INTRODUCTION

The efficiency of dairy cattle in utilizing feed to grow and produce milk is one of the main factors influencing the profitability of production (Berry and Crowley 2013). Residual feed intake (RFI) is one of the criteria for measuring feed efficiency. RFI is the difference between actual and predicted feed intake for each individual (Koch *et al.* 1963) which has high to moderate heritability in growing heifers and low to moderate heritability in milking cows (Berry and Pryce 2014). Hence, improving the efficiency of animals in converting feed to products is feasible by selecting and breeding cattle which need less feed than average to gain the same weight, or produce the same amount of milk. However, RFI is expensive to measure because it requires precise measurements of individual feed consumption, weight gain, and also milk yield and its components in milking cows and this has limited direct selection for feed efficiency in dairy cattle (Beever and Doyle 2007). Moreover, due to the polygenic architecture of feed efficiency, it is hard to find major genes influencing RFI. Genomic selection, using single nucleotide polymorphisms

(SNP) genotypes to estimate breeding values without measuring feed intake on selection candidates, could overcome this limitation (Meuwissen *et al.* 2001). However, genomic selection still requires a genotyped reference population with phenotypes for RFI. The limited size of reference populations, especially for RFI in milking cows, results in modest prediction accuracies. Since April 2015 the "Feed Saved" breeding value has been available in Australia and is also part of the national selection index (Pryce *et al.* 2015). Feed Saved includes genomic breeding values for RFI in heifers and cows and maintenance requirements in lactating cows and uses genomic predictions of RFI using a reference population of Australian cows and heifers and European cows. The aim of this research is to increase the accuracy of genomic estimated breeding values (GEBVs) for RFI in Australian heifers and milking cows through 1) including more animals in the reference population, 2) RFI information from non-dairy breeds and 3) using whole genome sequence (WGS) instead of SNP-chip genotypes.

MATERIALS AND METHODS

Animals and RFI measurements. The RFI measurements used for this study were from 843 Australian heifers and 440 Australian cows (139 animals had RFI measurements as heifers and cows), 954 European cows (357 British and 597 Dutch) and 4,772 beef cattle (Khansefid *et al.* 2014; Pryce *et al.* 2015). In this study, the RFI measurements in Australian cows were recalculated after including 206 new animals to the model described by Pryce *et al.* (2015).

Genotypes. The Australian heifers had 800K (Illumina HD Bovine SNP chip) genotypes (Pryce *et al.* 2012) and the rest of the dairy cattle had 50K (Illumina BovineSNP50K) genotypes which were imputed to HD (Pryce *et al.* 2014). For beef cattle, the SNP genotypes were either from HD or imputed from lower density (7K, 10K or 50K) to 800K (Khansefid *et al.* 2014). Moreover, for all datasets, the SNP-chip genotypes were imputed to WGS genotypes using FImpute (Sargolzaei *et al.* 2014) and RUN4 of 1000Bulls project as the reference.

Genome-wide association study (GWAS). The GWAS was conducted using beef cattle data to find associations between each SNP and RFI measurements using the model described by Khansefid *et al.* (2014) but using WGS genotypes in addition to SNP-chip genotypes.

Genomic relationship matrix (GRM). The GRMs were constructed using 2 methods (Yang *et al.* (2010) and VanRaden (2008)) for SNP-chip ($GRM_{SNP\text{-}chip}$) and WGS genotypes (GRM_{WGS}). Separate GRMs were also calculated from the SNP-chip ($GRM_{SNP\text{-}chip}$*) and WGS genotypes (GRM_{WGS}*) using the SNPs that were associated with RFI in beef cattle (p <0.001).

Statistical model. RFI measurements for heifers, Australian cows and European cows were considered to be 3 separate traits and were fitted in a multi-trait model (Equation 1) to calculate GEBVs using ASReml (Gilmour *et al.* 2009), where **y** is a T×1 vector consisting of RFI measurements on 1 or more of the 3 traits for each animals, **Z** is an incidence matrix associating observations to animals and traits, **g** contains the breeding values for each of 3 traits for all animals distributed as N(0, **G**⊗**K**), **G** is the genomic relationship matrix and **K** is a matrix of additive genetic variances and covariances between RFI in 3 datasets and **e** is a vector of residual terms.

$$\mathbf{y}=\mathbf{Z}\mathbf{g}+\mathbf{e} \quad [1]$$

To give extra weights to the SNPs associated with RFI in beef cattle, the average of $GRM_{SNP\text{-}chip}$ and $GRM_{SNP\text{-}chip}$* (i.e. $GRM_{SNP\text{-}chip\ \&\ SNP\text{-}chip*}$) and also the average of GRM_{WGS} and GRM_{WGS}* (i.e. $GRM_{WGS\ \&\ WGS*}$) were calculated and fitted in equation 1.

Accuracy of genomic prediction. The accuracy of genomic predictions was calculated with a 5 fold cross-validation strategy. The dataset was divided into 5 subsets, 4 of the subsets were used as a reference population and the 5th subset was used as a validation sample. The animals in the 5 subsets were selected randomly except paternal half sibs were always placed in the same subset. Then, the GEBVs of validation animals, whose phenotypes were not included in the analysis, were

estimated by genomic BLUP. The accuracy of each validation set was calculated as the correlation between GEBVs and RFI phenotypes divided by the square root of estimated heritability (h^2 for RFI in heifers and Australian cows were estimated 0.33 and 0.26, respectively) and the average across 5 validation sets was reported as the accuracy of prediction.

RESULTS AND DISCUSSION

Genotypes. In SNP-chip genotypes, 569,179 SNPs were in common between the datasets and had minor allele frequency (MAF) greater than 0.001. In WGS genotypes, 24,352,503 SNPs had MAF > 0.001.

GWAS. Among the common SNPs that had MAF > 0.001, 1,739 SNPs in SNP-chip genotypes and 60,646 SNPs in WGS genotypes were significantly associated with RFI in beef cattle ($p < 0.001$). So, about 0.3% of the SNPs were associated with RFI in beef cattle ($p < 0.001$) for both the SNP-chip or WGS genotypes.

Accuracy of genomic prediction. The accuracies of genomic predictions using different GRMs in Equation 1, are shown in Table 1. Substituting $GRM_{SNP\text{-}chip}$ with GRM_{WGS} did not improve the prediction accuracies. The accuracies were actually reduced when the Yang *et al.* (2010) method was used in GRM_{WGS} construction. So, the assumption of Yang *et al.* (2010) that rare alleles are more informative for WGS data seems to be incorrect. In WGS data there are many more SNPs with low MAF distributed across genome than in SNP-chip markers. Moreover, the accuracy of imputation is lower for rare alleles and therefore giving extra weight to these SNPs could reduce the accuracy of genomic predictions.

When the SNP-chip genotypes were used to construct the GRMs, there was no noticeable difference between the GRMs according to Yang *et al.* (2010) or VanRaden (2008). However, using WGS genotypes to make the GRMs, there was a slight superiority in constructing the GRMs according to VanRaden (2008).

Giving extra weight to the SNPs associated with RFI in beef cattle by using $GRM_{SNP\text{-}chip\ \&\ SNP\text{-}chip*}$ improved the accuracy of predictions in heifers and Australian cows. However, using $GRM_{WGS\ \&\ WGS*}$ in the model did not improve the prediction accuracy. When using WGS, the SNPs associated with RFI ($p < 0.001$) in beef cattle were distributed across the genome and included many low MAF SNPs. The rate of false positive associations seems to be higher for SNPs with low MAF because the imputation has more errors for these SNPs.. This problem could potentially be solved if a more stringent p-value was used to choose the SNPs for GRM_{WGS}*, however due to polygenic architecture of RFI, some SNPs with small effects would also be excluded.

Table 1. The accuracy of RFI predictions for Australian heifers and cows using different GRMs constructed according to Yang *et al.* (2010) and VanRaden (2008) (in parenthesis)

GRM	Accuracy of RFI prediction for heifers	Accuracy of RFI prediction for cows
$GRM_{SNP\text{-}chip}$	0.57 (0.56)	0.34 (0.35)
$GRM_{SNP\text{-}chip\ \&\ SNP\text{-}chip*}$	0.67 (0.65)	0.46 (0.41)
GRM_{WGS}	0.52 (0.55)	0.31 (0.39)
$GRM_{WGS\ \&\ WGS*}$	0.50 (0.53)	0.32 (0.34)

$GRM_{SNP\text{-}chip}$ is constructed from 569,179 SNP-chip genotypes.
$GRM_{SNP\text{-}chip\ \&\ SNP\text{-}chip*}$ is the average of $GRM_{SNP\text{-}chip}$ and the GRM built from 1,739 SNPs in SNP-chip genotypes which were significantly associated with RFI in beef cattle ($p < 0.001$).
GRM_{WGS} is constructed from 24,352,503 WGS genotypes.
$GRM_{WGS\ \&\ WGS*}$ is the average of GRM_{WGS} and the GRM built from 60,646 SNPs in WGS genotypes which were significantly associated with RFI in beef cattle ($p < 0.001$).

The best model tested was the GRM built using the Yang et *al.* method with SNP-chip genotypes and when extra weight was given to the SNPs associated with RFI in beef cattle. The accuracy of GEBVs for RFI in the best model for heifers and cows were 0.67 and 0.46, respectively.

CONCLUSION

This study shows that giving extra weight to SNPs that were associated with RFI in beef cattle increased the accuracy of GEBVs in dairy cattle. However, using imputed WGS data instead of 800K SNP-chip genotypes did not improve the prediction accuracy of genomic BLUP especially when the Yang *et al.* (2010) method was used to build GRM. The poor performance of WGS could be due to imputation errors and the use of BLUP rather than a non-linear method of calculating GEBVs. So, in order to benefit from using WGS genotypes, we need to: 1) use more accurate imputation, or direct genotyping of sequence variants, 2) find suitable statistical models such as Bayesian models, which allow a large proportion of SNPs to have zero effects and 3) use knowledge about the functionality of sequence variants. However, the current solution is to use SNP chip genotypes and to give extra weight in the GRM to the SNPs associated with RFI in beef cattle.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Wageningen University and Scotland's Rural University College for data contribution, and also Dairy Futures CRC and DairyBio for funding this research.

REFERENCES

Beever, D.E. and Doyle, P.T. (2007) *Anim. Prod. Sci.* **47**: 645.

Berry, D.P. and Crowley, J.J. (2013) *J. Anim. Sci.* **91**: 1594.

Berry, D.P. and Pryce, J.E. (2014) *10th World Congr. Genet. Appl. Livest. Prod., Vancouver, Canada. Am. Soc. Anim. Sci., Champaign, IL.*

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Thompson, R. and Butler, D. (2009) *VSN International Ltd, Hemel Hempstead, UK.*

Khansefid, M., Pryce, J.E., Bolormaa, S., Miller, S.P., Wang, Z., Li, C. and Goddard, M.E. (2014) *J. Anim. Sci.* **92**: 3270.

Koch, R.M., Swiger, L.A., Chambers, D. and Gregory, K.E. (1963) *J. Anim. Sci.* **22**: 486.

Meuwissen, T.H., Hayes B.J. and Goddard M. E. (2001) *Genetics* **157**: 1819.

Pryce, J.E., Gonzalez-Recio, O., Nieuwhof, G., Wales, W.J., Coffey, M.P., Hayes, B.J. and Goddard, M.E. (2015) *J. Dairy Sci.* **98**: 7340.

Pryce, J.E., Johnston, J., Hayes, B.J., Sahana, G., Weigel, K.A., McParland, S., Spurlock, D., Krattenmacher, N., Spelman, R.J., Wall, E. and Calus, M.P.L. (2014). *J. Dairy Sci.* **97**: 1799.

Pryce, J.E., Arias, J., Bowman, P.J., Davis, S.R., Macdonald, K.A., Waghorn, G.C., Wales, W.J., Williams, Y.J., Spelman, R.J. and Hayes, B.J. (2012) *J. Dairy Sci.* **95**: 2108.

Sargolzaei, M., Chesnais, J.P. and Schenkel, F.S. (2014) *BMC genomics* **15**: 1.

VanRaden, P.M. (2008) *J. Dairy Sci.* **91**: 4414.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E. and Visscher P.M. (2010) *Nat. Genet.* **42**: 565.