PREDICTION OF RESIDUAL FEED INTAKE FROM GENOME AND METAGENOME PROFILES IN FIRST LACTATION HOLSTEIN-FRIESIAN DAIRY CATTLE

M. Wang[1], J. E. Pryce[1,2], Keith Savin[1] and B. J. Hayes[1,2]

[1]Biosciences Research Division, Department of Economic Development, Agribio, 5 Ring Road, Bundoora, VIC 3086, Australia
[2]La Trobe University, Bundoora, VIC 3086, Australia

SUMMARY

Traits such as feed efficiency in dairy cattle are likely to be influenced by the genome of the host and the composition and abundance of microbiomes in the rumen. Here we describe an integrative approach that utilizes both genomic (SNP) and rumen microbiome data to predict future residual feed intake (RFI). The approach was tested in a small sample, of 28 Australian Holstein-Friesian dairy cattle that had 30K SNP genomic predictions for RFI and rumen microbiome profiles. The genomic and microbiome profile predictions were combined using a linear regression model. Results are very preliminary due to the small size of the data set, however, the prediction accuracy in cross validation was maximized when both SNP and rumen microbiome profiles were used (r=0.57; 95% CI: 0.33:0.72). These results, while promising, should be repeated in a larger data set.

INTRODUCTION

Feed efficiency is a key economic trait for livestock species, including dairy cattle. One measure of feed efficiency is residual feed intake, which is the approximate difference between the actual feed intake and estimated feed intake based on a regression model that takes into account energy costs for body maintenance and production over a defined production period (Connor, 2014). Macdonald *et al.* (2014) and Pryce *et al.* (2014) both demonstrated that genomic estimated breeding values (GEBV) for RFI could be derived which predict residual feed intake (RFI) with moderate accuracy. In addition to the cow's own genome, the profile of the rumen microbiome (species composition and abundance) has been shown to be associated with some traits, particularly methane emissions (Ross *et al.* 2013a; Kittelmann *et al.* 2014). So an obvious question is, can we improve predictions of future RFI phenotypes by integrating genomic predictions from SNP genotypes with rumen microbiome profiles. This seems promising, as integration of genomic, transcriptomes, proteomics and metabolomics information has already returned high accuracy in predicting type 2 diabetes (Chen *et al.* 2012). The objectives of this study were to investigate: (1) can rumen microbiome profiles be used to predict RFI for dairy cattle? (2) can the accuracy of prediction be increased by integrating using GEBV and rumen microbiome profiles?

MATERIALS AND METHODS

The dataset included 28 first parity Australian Holstein-Friesian dairy cows which were born in 2 different years at the Ellinbank research station, Victoria, Australia. Fifteen out of 28 cattle were born in July to September 2008, referred to here as FCE1 animals. The rest, referred as FCE2 animals, were born in July to September 2009. Rumen samples and dry matter intake data were collected during 1st lactation, which was in February 2011, at the age of 938 ± 12 day for FCE1 cattle, and in November 2011, at the age of 812 ± 18 day for FCE2 cattle, respectively. All animals were fed similar diets, which constituted predominantly of alfalfa hay pressed into cubes. In lactating cows the diet was supplemented with crushed wheat. Feed was always available *ad libitum*. RFI phenotypes were calculated by regressing DMI on fixed effects and liveweight and

growth in heifers and DMI on fixed effects, liveweight and production in lactating cows as described by Macdonald *et al.* (2014).

To calculate GEBVs for the 28 animals, the reference dataset comprised a total of 815 Australian growing heifers of which 74 also had RFI measurements in first lactation (Macdonald *et al*. 2014). The genotype data described by de Haas *et al.* (2012) that comprised 30,949 SNP were used to construct the genomic relationship matrix using the Yang *et al*. (2010) method. The analysis, using G-REML, was performed using ASReml software (Gilmour *et al*. 2009). A bivariate model similar to that derived from Pryce *et al.* (2015) was fitted, so that the covariance between growing heifer and cow RFI could be estimated. The model used was:

$$y_T = X_T b_T + Z_T g_T + e_T$$

Where y_T was the $2xn$ matrix of observations on all traits, X_T was the incidence matrix for fixed effects, b_T was the matrix of solution of fixed effects, Z_T was an incidence matrix mapping records to animals, g_T was the corresponding genomic breeding values for animals with genotypes for all traits, and e_T was a $2xn$ matrix of residual terms. The g_T was assumed to be distributed as $N(0, G \otimes K)$, where G was the animal by animal genomic relationship matrix and K was a $2x2$ matrix of additive genetic variances between heifers and cows. Then $V(e_T) = R \otimes I$, where R was a $2x2$ matrix of error variances and I was an nxn identify matrix.

Twenty-eight microbiome samples were extracted using the PowerMaxSoil DNA Isolation kit (MoBio) and sequenced on the HiSeq 2000 (Illumina) as per Ross *et al.* (2013b). Raw sequencing reads were trimmed from 5'-end and retained for downstream analysis if the 5'-end reached a maximum of 3 bases whose phred quality score were <15; the average remaining read quality was ≥20; and remaining read length was ≥50bp. This resulted in more than 268 million reads from all samples passed filtering. Trimmed reads were subsequently aligned to reference library using Bowtie2 (version 2.2.2; Langmead and Salzberg 2012). The reference library was composed of assembled rumen microbiome contigs from 3 smaller collections of sequences (Hess *et al.* 2011; Ross *et al.* 2012; Ross *et al.* 2013b). Contigs from the 3 sources were concatenated and sequences <250bp were removed. An overall alignment rate of 17.36% from all animals was attained and the distribution of sequences aligning to reference contigs was plotted in Figure 1.

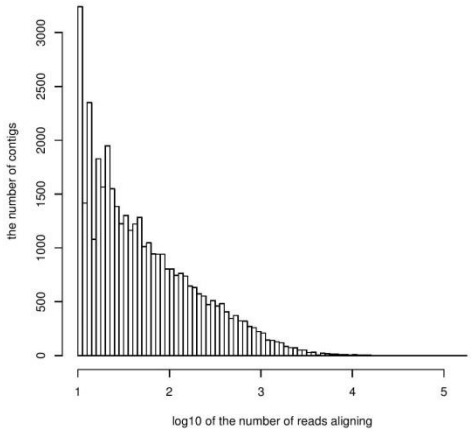

Figure 1. A histogram of read distribution. The majority of contigs had 10 to 100 reads aligned to. Contigs that had less than 10 reads aligning were removed from analysis.

Rumen microbiome profile prediction (RMP) for RFI was performed in the free R statistical software (version 3.1.2; The R Foundation for Statistical Computing; http://www.r-project.org/)

and package rrBLUP (Endelman *et al.* 2011) was used. A metagenomics relationship matrix calculated as per Ross *et al.* (2013a) was fitted into best linear regression model (BLUP) and validated using two-fold cross-validation (CV), where FCE1 and FCE2 were either training or validation sets and an alternative procedure, called leave-one-out where we sequentially removed just one animal from the dataset to estimate its genomic breeding value using the remaining data. Animals being predicted were always omitted from training set. Integrative (genomic and metagenomics) prediction was performed in R statistical software. Twenty-eight measured RFI values, GEBVs for RFI and RMP were fitted into a linear regression model. The coefficients in the output were multiplied with GEBV and RMP respectively to calculate the integrative predicted RFI. Accuracy was assessed by Pearson's correlation, 'r', that is, the correlation between the measured values with predicted values. Ninety-five percent confidence interval (CI) was calculated via bootstrapping with 10,000 replicates. Coding scripts are available upon request.

RESULTS AND DISCUSSION

The accuracy of genomic prediction was 0.33 (95% CI: 0.07:0.59; Table 1). A non-zero accuracy was observed for RFI calculated using rumen microbiome profile prediction under leave-one-out CV (r=0.49; 95% CI: 0.2:0.67; Table 1), but the accuracy of rumen microbiome profile prediction under two-fold CV was much lower (r= 0.08; 95% CI: -0.39:0.34; Table 1).

When both the cow's genome and rumen microbiome information were used for predicting RFI, the accuracies were the highest in both two-fold (r=0.38; 95% CI: 0.05:0.65; Table 1) and leave-one-out (r=0.57; 95% CI: 0.33:0.72; Table 1) testings.

Table 1 accuracy comparison among genomic, metagenomics and integrative predictions

Sequence source	CV[2] method	Correlation	95% CI[#]	Significant
Cow's Genome	Not available	0.33	(0.07, 0.59)	Y
Rumen microbiome	Two-fold	0.08	(-0.39, 0.34)	N
	Leave-one-out	0.49	(0.2, 0.67)	Y
Integration[1]	Two-fold	0.38	(0.05, 0.65)	Y
	Leave-one-out	0.57	(0.33, 0.72)	Y

[1]Integration: both cow's genome and rumen microbiome information were used.
[2]CV: cross validation.
[#]95% Confidence interval of the Pearson's correlation coefficient r based on 10,000 bootstraps.

Our results showed two main findings: firstly, rumen microbiome profiles may be able to predict RFI in some circumstances; secondly, integrating genomic and metagenomics information can increase prediction accuracy. The idea of integrating genetic information has already been realised in human research (Chen *et al.* 2012), but to our knowledge this study is the first to apply it to predict RFI in livestock. Four main elements affect the performance of prediction from rumen microbiome profiles: the number of samples in the study, size of the reference library, diversity of reference library and sequence depth (Ross *et al.* 2012). Even though the number of samples involved in our study was small, by updating the rumen reference library and maintaining a sequence depth of a minimum of 3 million reads, we obtained a similar accuracy as that in Ross *et al.* (2013a) study. We saw a growth of overall alignment rate as compared with Ross *et al.* (2013a). This could continually be improved by adding internationally collaborative references such as the Hungate1000 database (Nordberg *et al.* 2014). Currently rumen samples are still relatively hard to obtain; therefore a wider mining of ruminant metagenomics sequencing data will rely on technical improvements on sample collection.

In conclusion, microbiome information appears to be useful in predicting RFI of the same host animals. Prediction accuracy could be increased when both cow's genome and rumen microbiome

profiles are used together, though given the small samples size used here, the analysis needs to be repeated in a larger data set.

REFERENCES

Chen R., Mias G.I., Li-Pook-Than J., Jiang L., Lam H.Y., *et al* (2012) *Cell* **148**: 1293.

Connor E.E. (2014) *Animal*. **8**:1.

de Haas Y., Calus M. P., Veerkamp R.F., Wall E., Coffey M.P., *et al* (2012) *J. Dairy Sci.* **95**: 6103.

Endelman J.B. (2011) *Plant Genome* **4**: 250.

Gilmour A.R., Gogel B.J., Cullis B.R. and Thompson R. (2009) 'ASReml User Guide' Release 3.0, VSN International Ltd.

Kittelmann S., Pinares-Patino C.S., Seedorf H., Kirk M.R., Ganesh S., *et al* (2014) *PloS one* **9**: e103171.

Langmead B., Salzberg S. (2012) *Nat. Methods* **9**:357.

Macdonald K.A., Pryce J.E., Spelman R.J., Davis S.R., Wales W.J., *et al* (2014) *J. Dairy Sci.* **97**: 1427.

Nordberg H., Cantor M., Dusheyko S., Hua S., Poliakov A., *et al* (2014) *Nucleic Acids Res.* **42**: D26.

Pryce J.E., Gonzalez-Recio O., Nieuwhof G., Wales W., Coffey M.P., *et al* (2015) *J. Dairy Sci.* (submitted).

Pryce J.E., Wales W.J., de Haas Y., Veerkamp R.F. and Hayes B.J. (2014) *Animal.* **8**(1): 1.

Ross E.M., Moate P.J., Marett L.C., Cocks B.G. and Hayes B.J. (2013a) *PloS one* **8**(9): e73056.

Ross E.M., Moate P.J., Marett L., Cocks B.G. and Hayes B.J. (2013b) *J. Dairy Sci.* **96**(9): 6030.

Ross, E.M., Moate P.J., Bath C.R., Davidson S.E., Sawbridge T.I., *et al* (2012) *BMC Genetics* **13**: 53.

Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., *et al* (2010) *Nat. Genetics* **42**:565.