

## Phenotypic and genetic associations between feed efficiency and FTIR milk-spectra

H.O. Toledo-Alvarado<sup>1\*</sup>, R.J. Tempelman<sup>2</sup>, M. Lopez-Cruz<sup>3</sup>, M.J. VandeHaar<sup>2</sup>, J.E.P. Santos<sup>4</sup>, F. Peñagaricano<sup>5</sup>, P. Khanal<sup>2</sup> and G. de los Campos<sup>6,7,8</sup>

<sup>1</sup> Department of Genetics and Biostatistics, National Autonomous University of Mexico, Ciudad Universitaria, P.O. Box 04510, Mexico City, Mexico; <sup>2</sup> Department of Animal Science, Michigan State University, Michigan, P.O. Box 48824, East Lansing, USA; <sup>3</sup> Department of Plant, Soil and Microbial Sciences, Michigan State University, Michigan, P.O. Box 48824, East Lansing, USA; <sup>4</sup> Department of Animal Sciences, University of Florida, Florida, P.O. Box 32608, Gainesville, USA; <sup>5</sup> Department of Animal and Dairy Sciences, University of Wisconsin-Madison, Wisconsin, P.O. Box 53706, Madison, USA; <sup>6</sup> Department of Epidemiology and Biostatistics, Michigan State University, Michigan, P.O. Box 48824, East Lansing, USA; <sup>7</sup> Department of Statistics and Probability, Michigan State University, Michigan, P.O. Box 48824, East Lansing, USA; <sup>8</sup> Institute for Quantitative Health Science and Engineering, Michigan State University, Michigan, P.O. Box 48824, East Lansing, USA; \*[h.toledo.a@fmvz.unam.mx](mailto:h.toledo.a@fmvz.unam.mx)

### Abstract

Data involving 9,226 records from 483 US Holstein cows were used to study the genetic relationship between feed efficiency, measured as Residual Feed Intake (RFI) and the Fourier transform infrared (FTIR) spectra of milk samples. Estimates of heritabilities, genetic and phenotypic correlations, as well as the accuracy of an indirect selection using spectra were calculated. The prediction accuracy of RFI was also evaluated using different methodologies. Genotypes were used to fit univariate and bivariate (each individual wavenumber and RFI) GBLUP analyses. A genetic relationship between RFI and wavenumbers associated with fat, protein and urea were observed. Notable differences were detected in the patterns of phenotypic versus genetic correlations. The maximum indirect selection accuracy estimated using a particular wavenumber was 0.11. No notable differences in the accuracy of genetic prediction were observed between different methodologies that use the full spectra for RFI prediction, and the accuracies were generally higher (~0.15) than using the individual wavenumbers.

### Introduction

Feed efficiency is an economically and environmentally relevant trait; however, measuring it in commercial dairy farms is not feasible. Therefore, there is an increased interest in using milk-spectroscopy to predict and use feed efficiency to improve farming management and genetic programs. Feed efficiency can be measured indirectly through the residual feed intake (RFI) of cows in experimental conditions; these data can be used to calibrate equations to predict RFI from phenotypes recorded in commercial conditions such as milk-spectra. The absorbance of any given Fourier transform infrared (FTIR)-spectra wavelength is a proxy phenotype that is potentially affected by genetic and environmental factors with a given heritability (Bittante and Cecchinato, 2013). Usually, competing prediction methods (e.g., partial least squares, LASSO, Bayesian methods) are compared based on the (phenotypic) correlation between predicted RFI and measured RFI in testing data sets. Such phenotypic correlations are affected by both genetic and environmental effects affecting RFI and the measured traits. From a genetic perspective, the accuracy of indirect selection may be a more appropriate measure of performance. Therefore, the objective of this study was to quantify the genetic prediction accuracy of individual wavelengths and of commonly used prediction models as predictors of genetic merit for RFI.

## Materials & Methods

Data from several experimental trials conducted at the University of Florida (7,205 records from 311 cows) and Michigan State University (2,021 records from 172 cows) were used in this study. Cows were randomly assigned to individual feeding gates used to evaluate daily feed intake. Cows had ad libitum access to their respective diets, which were fed as total mixed ration (TMR). Dry matter intake (DMI) for individual cows was based on the amount of TMR offered minus the amount refused multiplied by the dried matter content of the TMR in a particular week. Data collected each week included the body weights, milk yields and milk composition, and the FTIR spectral data associated with milk samples. All milk samples were analysed using Bentley FTS instruments with 899 spectral points from wavenumber 649 to 3998  $\text{cm}^{-1}$ . Animals were genotyped on various available chips and imputed. The genotypes were edited for quality control and 78,964 SNP were included in the analysis. A better description of the data can be found in Gaddis et al. (2021).

**RFI Calculation.** An energy sink model (1) was used to estimate RFI using the net energy for lactation ( $\text{NE}_L$ ) as determined in the NRC (2001). The weekly metabolic body weight (MBW) was calculated as the weekly average  $\text{BW}^{0.75}$ . The weekly body weight changes  $\Delta\text{BW}$  was calculated using a linear regression of actual or projected BW on days from the beginning to the end of a week (Berry and Crowley, 2013). Thus, the model to measure residual feed intake included the fixed effects of parity (primiparous or multiparous) interacting with the specific fifth-order Legendre polynomial regressions on DIM, and linear regressions on  $\text{NE}_L$ , MBW and the  $\Delta\text{BW}$ . The random effects included the effects of experiment-specific rations and test weeks. The estimated residuals ( $\hat{\epsilon}$ ) from this model were determined to be the weekly RFI.

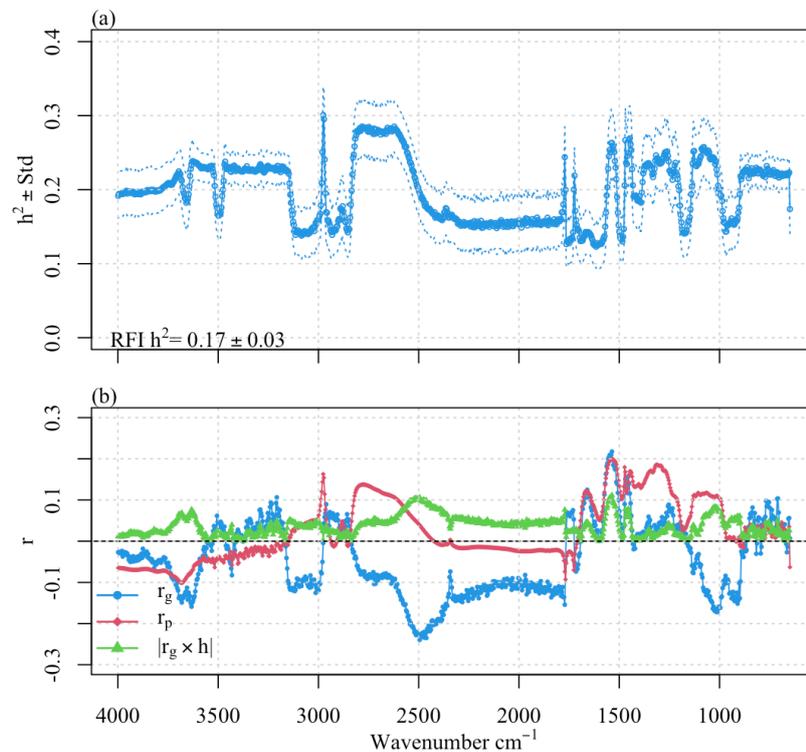
**Prediction accuracy of individual wavenumbers.** The accuracy of indirect selection of an indicator trait (a wavenumber in our case) is the product of the square-root of the heritability of the indicator trait ( $h_I^2$ ) times the genetic correlation between the measured trait (wavenumber) and the selection objective (RFI,  $r_{g_y, g_I}$ ), that is  $\text{Acc}(g_{y_i}, I_i) = r_{g_y, g_I} \sqrt{h_I^2}$  (Falconer and Mackay 1996), where  $I$  stands for each wavenumber. Therefore, we first fitted a sequence of 899 bivariate GBLUP models using RFI and one wavenumber (adjusted by the same effects included in the model to estimate the RFI) as the two traits. From these analyses we estimated  $h_I^2$ ,  $r_{g_y, g_I}$ , and  $\text{Acc}(g_{y_i}, I_i)$ . All the analyses were performed using the Multitrait() function of the BGLR R-package (Pérez and de los Campos, 2014), with genomic relationship matrices derived from genotypes, including the random effects of the animal and the permanent environment.

**Genetic prediction accuracy of prediction equations.** To evaluate the genetic prediction accuracy that can be achieved when combining information from all the spectrum we calibrated prediction equations for RFI (with the 899 available wavelengths as predictors) using Partial Least Squares (PLS with 10 components) and five Bayesian methods which differed on the prior used. To evaluate accuracy, and following Lopez-Cruz et al. (2020), we partitioned the data into a training and a testing set. The testing set was formed by randomly selecting 50 cows and assigning the phenotypes of those cows to the testing set. The remaining data was used for model training. We then calibrated the prediction equations using the training data. We then applied the resulting prediction equations to the milk spectra data of the testing set to obtain a prediction of RFI phenotypes. Finally, we estimated the genetic prediction accuracy of each method by fitting bivariate GBLUP models with RFI and the predicted RFIs as traits, including

the random effects of the animal and the permanent environment. To account for uncertainty due to sampling of training and testing sets we conducted 100 random training-testing partitions and applied the procedure above described to each of the partitions; from this, we present results averaged over 100 training-testing partitions.

## Results

The estimated heritability for the RFI was  $0.17 \pm 0.03$ , while for the spectra the estimated heritabilities (Figure 1.a) ranged between 0.12 and 0.30. The maximum heritability was found at wavenumber  $2,976 \text{ cm}^{-1}$ ; in this region, Fat-B [absorbance by the carbon–hydrogen stretch (C–H)] is usually filtered (Kaylegian et al., 2006). Genetic and phenotypic correlations varied with spectra region (Figure 1.b) with estimated genetic correlations ranging between -0.24 and 0.22. From wavenumber  $1,167$  to  $1,689 \text{ cm}^{-1}$ , positive genetic and phenotypic correlations were estimated, this region is associated with the prediction of protein ( $1,546 \text{ cm}^{-1}$ ) (Kaylegian et al., 2006), urea ( $1,469 \text{ cm}^{-1}$ ), fat ( $1,460 \text{ cm}^{-1}$ ) (Hansen, 1998), and acetone ( $1,238 \text{ cm}^{-1}$ ) (Hansen, 1999). The maximum value observed for the accuracy of the index was 0.11 at the wavenumber  $1,536 \text{ cm}^{-1}$ , note that protein is usually filtered from  $1,546 \text{ cm}^{-1}$  to  $1,492 \text{ cm}^{-1}$  (Kaylegian et al., 2006).



**Figure 1. (a) Heritability ( $h^2$ ) of FTIR wavelengths; (b) Genetic correlation ( $r_g$ ), phenotypic correlation ( $r_p$ ), and index accuracy  $|r_g \times h|$  between each wavelength and RFI.**

The heritabilities of predicted phenotypes across different methodologies (Table 1) were similar to the estimated heritability of RFI. No major differences were found between the different methodologies for using milk-spectra in the prediction of RFI (ranging from 0.14 to 0.16). However, the Bayesian methods (A and Lasso) showed a genetic correlation and accuracy slightly higher than the PLS traditional methodology. An important point to note is that the accuracy was higher when using the entire spectrum compared to the analysis of the individual wavelengths.

**Table 1. Heritability ( $h^2$ ) of predicted Residual Feed Intake, genetic correlation ( $r_g$ ) and genetic prediction accuracy between predicted and observed RFI using different methodologies.**

Method	$h^2 \pm \text{Std}$	$r_g \pm \text{Std}$	Accuracy $\pm \text{Std}$
PLS	$0.15 \pm 0.04$	$0.35 \pm 0.19$	$0.14 \pm 0.08$
Bayes A	$0.16 \pm 0.05$	$0.38 \pm 0.19$	$0.16 \pm 0.09$
Bayes B	$0.19 \pm 0.06$	$0.34 \pm 0.20$	$0.15 \pm 0.09$
Bayes C	$0.18 \pm 0.06$	$0.31 \pm 0.20$	$0.14 \pm 0.09$
Bayes Lasso	$0.16 \pm 0.05$	$0.38 \pm 0.19$	$0.15 \pm 0.09$
Bayes Ridge Regression	$0.16 \pm 0.04$	$0.37 \pm 0.19$	$0.15 \pm 0.08$

## Discussion

The heritabilities for the spectrum were comparable with other studies. Bittante and Cecchinato (2013) and Rovere et al. (2019) estimated the highest heritabilities from 1,582 to 930  $\text{cm}^{-1}$ ; whereas we observed several other regions with heritabilities above 0.25 and associated with fat, protein and urea (1,238 to 2,976  $\text{cm}^{-1}$ ). Regions of the spectra had modest  $r_g$  with RFI, this, together with the  $h^2$  of individual wavenumbers, led to individual band accuracies of indirect selection ranging from 0 to 0.1. The wavenumbers most predictive of genetic merit for RFI were located in the fingerprint and in the mid-infrared region (from 3,048 to 1,701  $\text{cm}^{-1}$ ). When we used all milk-spectra data into prediction equations for RFI we found higher accuracies of indirect selection, with Bayes A achieving the highest accuracy. Since the accuracy of direct selection is  $\sqrt{h_{RFI}^2}$ , this is a relative efficiency of indirect selection of more than 1/3<sup>rd</sup> (0.36 using  $0.15/\sqrt{0.17}$ ). This information could be used to derive predictions of RFI at large scale that could be considered as indicator traits for selection to improve feed efficiency.

## References

- Berry, D.P., and Crowley J.J. (2013) *J. Anim. Sci.* 91(4):1594-1613. <https://doi.org/10.2527/jas.2012-5862>
- Bittante G., and Cecchinato A. (2013) *J. Dairy. Sci.* 96(9):5991-6006. <https://doi.org/10.3168/jds.2013-6583>
- Falconer D.S., and Mackay T.F.C. (1996) *Introduction to Quantitative Genetics*. 4th ed. Pearson Education.
- Gaddis K., VanRaden P., Tempelman R., Weigel K., White H., *et al.* (2021) *Interbull Bulletin* (56), 147-152.
- Hansen P.W. (1998) *Milchwissenschaft* 53(5):251-255.
- Hansen P.W. (1999) *J. Dairy Sci.* 82(9):2005-2010. [https://doi.org/10.3168/jds.S0022-0302\(99\)75437-8](https://doi.org/10.3168/jds.S0022-0302(99)75437-8).
- Kaylegian K.E., Houghton G.E., Lynch J.M., Fleming J.R., and Barbano D.M. (2006) *J. Dairy Sci.* 89(8):2817-2832. [https://doi.org/10.3168/jds.s0022-0302\(06\)72555-3](https://doi.org/10.3168/jds.s0022-0302(06)72555-3)
- Lopez-Cruz, M., Olson, E., Rovere, G., Crossa, J., Dreisigacker, S., *et al.* (2020). *Sci. Rep.* 10(1), 1-12. <https://doi.org/10.1038/s41598-020-65011-2>
- National Research Council (NRC) (2001). *Nutrient Requirements of Dairy Cattle*. Natl. Acad. Press, Washington DC, USA.
- Pérez P., and de los Campos G. (2014) *Genetics* 198(2):483-495. <https://doi.org/10.1534/genetics.114.164442>
- Rovere G., de los Campos G., Tempelman R.J., Vazquez A.I., Miglior F., *et al.* (2019) *J. Dairy Sci.* 102(2): 1354-1363. <https://doi.org/10.3168/jds.2018-15109>