Genomic Prediction of Genetic Residual Feed Intake Integrating a Novel Energy Sink for Change in Body Reserves.

R.B. Stephansen*[1]*, J. Lassen*[1,2]*, V.M. Thorup*[3]*, B.G. Poulsen*[1]*, J. Jensen*[1]*,
***G. Sahana*[1]*, and O.F. Christensen*[1]**
[1] *Center for Quantitative Genetics and Genomics, Aarhus University, C. F. Møllers Allé 3, 8000 Aarhus C, Denmark*
[2] *VikingGenetics, Ebeltoftvej 16, 8960 Randers SO, Denmark*
[3] *Department of Animal and Veterinary Sciences, Aarhus University, Blichers Allé 20, 8830 Tjele, Denmark*
Corresponding author: rasmus.stephansen@qgg.au.dk

Abstract

Traditionally, a two-step modeling approach of residual feed intake (**RFI**) is incorporated into the Feed Saved index at dairy cattle genetic evaluation centers. Challenges have been identified in the 1st step on handling fixed effects in the statistical model and dealing with missing phenotypes. This could be solved using a multi-variate modelling approach for genetic RFI (**gRFI**). Most existing RFI models use changes in body weight, and therefore, likely inadequately account for changes in body reserves because energy density differs between mobilization and deposition, and between adipose and muscle tissue. Alternatively, energy balance can be estimated from body reserve changes (**EB_{body}**). Therefore, this study aimed to explore a genomic evaluation of gRFI in Nordic primiparous cows using EB_{body} as energy sink for changes in body reserves. Weekly records were collected from 2,029 Jersey (**JER**) cows, 3,178 Red Dairy Cattle (**RDC**) cows, and 4,661 Holstein (**HOL**) cows. For JER and RDC, the feed intake data was obtained with the Cattle Feed InTake system (**CFIT**, VikingGenetics, Denmark). For HOL, feed intake data was collected from CFIT farms and a research farm (857 cows and 25,547 weekly records). The genotyping rate for cows with data were 92% for JER and RDC, and 81% for HOL. The gRFI model was a random regression multi-variate model with 2nd order Legendre polynomials for additive genetic and permanent environmental effects. The gRFI model was validated with an across-herd cross-validation scheme using the Legarra Reverter method and reporting bias, dispersion and correlation terms. Breeding values were predicted using the single-step approach for both genotyped and non-genotyped animals. The bias was close to 0 for all breeds. The dispersion coefficients were found in an acceptable range at 0.92 (DMI) and 0.87 (gRFI) for HOL and 0.96 (DMI) and 0.85(gRFI) for RDC, while overdispersion was observed for JER (DMI:0.75, gRFI:0.69). Correlations between genomic breeding values, estimated with whole and partial phenotypic information, were moderately high for all breeds (DMI: 0.51-0.68, gRFI: 0.46-0.59). In conclusion, it was possible to construct a genomic gRFI model for all three Nordic dairy cattle breeds and integrate EB_{body} as an energy sink indicator. We observed promising validation metrics for HOL and RDC, but JER models need further refinement. The results demonstrate selection for gRFI is expected to provide genetic gain of feed efficiency in dairy cattle.

Key words: feed efficiency, Feed Saved, multi-variate modelling, Nordic dairy cattle

Introduction

Improving feed efficiency through genetics poses an important part of enhancing economic viability and environmental sustainability in dairy cattle farming (VandeHaar et al., 2016). Several genetic evaluation centers have integrated the "Feed Saved" index, as selection criteria for feed efficiency in the national breeding goals. A significant component of this index lies in the residual feed intake (**RFI**) part, which traditionally is modelled in a two-step process (Tempelman and Lu, 2020). Initially, a precorrection step generates a model-based residual for feed intake, serving as the phenotype for subsequent genetic evaluation. However, challenges arise concerning the handling of fixed effects and missing records within this initial step. To address these challenges, Tempelman and Lu (2020) proposed the genetic multi-variate approach to RFI (**gRFI**) based on the work by Kennedy et al. (1993). This model has not been tested within Nordic breeds.

Most existing RFI models address body reserve management using changes in body weight (**ΔBW**). However, this approach likely suffers deficiency because of significant variations in energy density between mobilization and deposition, as well as among different tissue types (adipose and muscle). An alternative is outlined by Thorup et al. (2018), who proposed to estimate energy balance from changes in body reserves (**EB_{body}**) by employing energy-specific coefficients tailored to tissue types and energy status. However, the effect of EB_{body} has yet to be investigated for RFI models.

Genomic prediction offers implementation of traits that have relatively few records due to expensive recording schemes (e.g. feed efficiency). Studies have demonstrated the feasibility of genomic prediction for dry matter intake (Berry et al., 2014, De Haas et al., 2015). As a limited number of records are available, the traditional forward prediction outlined in Mäntysaari et al. (2010) were not feasible for validation of genomic predictions. Alternatively, the Legarra-Reverter cross-validation method (Legarra and Reverter, 2018), using whole and partial datasets seems attractive. However, limited literature exists on genomic validation of gRFI.

This study aimed to explore the ability to establish a genomic evaluation of gRFI and perform herd cross-validation, using Nordic primiparous cows and incorporating the EB_{body} as energy sink trait for body reserve management.

Materials and Methods

The modelling of the multi-variate gRFI model is based on weekly means of dry matter intake (**DMI**), energy corrected milk (**ECM**), and body weight (**BW**) records for each individual cow. The phenotyping systems were the Cattle Feed InTake (**CFIT**) system installed on 19 commercial Danish farms and research data from the Danish Cattle Research Center (**DCRC**) at AU-Foulum. A detailed description of the 3D camera based CFIT system is outlined in Lassen et al. (2023) and for DCRC in Li et al. (2017) and Stephansen et al. (2023).

Feed intake data

The data compromised repeated records from one to 45 weeks in milk of 3,873 HOL cows with 161K weekly CFIT DMI records (2,564 primiparous), 2,068 JER cows with 93K weekly CFIT DMI records (1,505 primiparous), 3,235 RDC cows with 139K weekly CFIT DMI records (2,006 primiparous) and 878 HOL cows with 50K weekly DCRC DMI records from the Roughage Intake Control System (Insentec B.V., Marknesse, the Netherlands) (835 primiparous). A detailed description of the data and quality control can be found in Stephansen et al. (2024).

Energy balance from body reserves

We adapted the estimation method of EB_{body}, using frequent BW measurements from Thorup et al. (2013) as:

$$EB_{body}, MJ/day = z \times \Delta BL + y \times \Delta BP_{std}$$

where EB_{body} is the energy balance phenotype calculated from frequent BW measurements and expressed in changes of mega joule per day, z is the energy coefficient for lipid, being 39.6 MJ/kg mobilized and 56 MJ/kg deposited adipose tissue, y is the energy coefficient for protein, being 13.5 MJ/kg mobilized and 50 MJ/kg deposited muscle tissue, ΔBL is the change in body lipid and ΔBP_{std} is the predicted change in body protein outlined in Thorup et al.

(2013). Details on the modelling of EB_{body} can be found in Thorup et al. (2018) and in context of this data in Stephansen et al. (2024).

Pedigree and Genotypes

Breed-specific pedigrees used from the Danish cattle database and underwent a pruning process using the DMU trace software (Madsen, 2012) for cows with data. The pruned pedigrees consisted of 18,432 HOL animals, 7,294 JER animals, and 12,423 RDC animals. Phantom parent groups were assigned to animals with missing parents using combinations of sex (Male or Female), breed (breed in analysis or other breeds), country (HOL: Nordic, EU, North America & rest; JER+RDC: Nordic & rest), and birth year classes (HOL: <2000, 2000-2010, & >2010; JER+RDC: before and after 2000).

Imputed genotypes were provided by Nordic Cattle Genetic Evaluation (Skejby, Denmark). Most animals were genotyped with 50k Illumina Bovine SNP50 or imputed from the LD chip panels. The imputation was done by SEGES Innovation (Skejby, Denmark) and part of the routine genetic evaluations in Nordic Genetic Cattle Evaluation (Skejby, Denmark). For Holstein 46,342 single nucleotide polymorphisms (**SNP**) were available, which were 41,897 SNPs for Jersey and 46,914 SNPs for RDC. Genotypes from animals born before 2000 were omitted because genotypic information on distantly related animals contribute little to accuracy of prediction in focal animals, and because including genomic information across multiple generations can promote prediction bias.

To calculate the relationship matrix encompassing genotyped and non-genotyped cows for a ssGBLUP analysis, we calculated the inverse of **H** as (Aguilar et al., 2010, Christensen and Lund, 2010):

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & (\omega\mathbf{G} + (1-\omega)\mathbf{A}_{22})^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix},$$

where $\mathbf{A}^{-1}$ is the inverse of the pedigree relationship matrix, **G** the genomic relationship matrix, ω is the relative weight of the polygenetic effect (ω=0.8), $\mathbf{A}_{22}$ is the part of the pedigree relationship matrix with genotyped animals, and $\mathbf{A}_{22}^{-1}$ is the inverse of $\mathbf{A}_{22}$. The genomic relationship matrix was calculated according to VanRaden (2008) using method 1 and the *invgmatrix* software by Su and Madsen (2011).

Statistical model

Variance components for the multi-variate model were estimated using a Gibbs sampler in the RJMC module in DMU version 5.5 Madsen and Jensen (2013). For variance component estimation, the pedigree-based relationship matrix was used for following multi-variate model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Mh} + \mathbf{Za} + \mathbf{Wpe} + \mathbf{e},$$

where **y** is the vector of phenotypes with sub-vectors for DMI, ECM, BW and EB_{body} in the different weeks of lactation; **b** is the fixed effects year x experimental diet at DCRC or version of CFIT system, a fourth order Legendre polynomial fixed regression on weeks in milk and nested within herd, and a second order Legendre polynomial fixed regression on age at calving; **h** is the vector of random effects for herd × year × test-week (record date); **a** is the vector of random regressions for random additive genetic effect of cows with sub-vectors for each of the traits; **pe** is the vector of random regressions for random permanent environmental effects of cows with sub-vectors for each of the traits. Weekly means were modelled across traits from one to 45 weeks in milk by a second Legendre polynomials (intercept, linear, quadratic) for both **a** and **pe**; **e** is the vector of random residual effects with sub-vectors of all traits included in the analysis. **X** is the design matrix for fixed effects, **M** is the design matrix for herd × year × test-week random effects and **Z** and **W** are the design matrices with covariable matrices containing

Legendre polynomial coefficients corresponding to week of lactation. Details on post-model processing of variance components to derive heritability, additive genetic correlations, and genetic regressors can be found in Stephansen et al. (2024).

Genomic herd cross-validation

We aimed to perform genomic validation by herd. Thereby, we assessed the expected value of genomic breeding values (**GEBV**) in herds which do not have the CFIT system. The estimated variance components of the multi-variate model and $\mathbf{H}^{-1}$ were applied to ssGBLUP models to estimate GEBVs using the DMU5 module with the preconditioned conjugate gradient computation method.

To set up the different datasets for the herd cross-validation, we first formed a whole dataset containing all phenotypic information, that was used to estimate GEBVs ($\mathbf{GEBV_{whole}}$). Hereafter, we formed three partial datasets for HOL and two for JER and RDC. In each of the partial datasets, we omitted all phenotypic information for 1-3 herds, and a herd could only appear as validation herd in one partial dataset. Assigning herds to be validation herds in the different partial datasets were done such that a group of validation herds consisted of herds that were geographically close and approximately 1,000 validation cows. These partial datasets (7 in total across breeds) were used to predict GEBVs ($\mathbf{GEBV_{partial}}$). A few of the CFIT herds were not used as validation herds and the DCRC herd were always included in the training population for HOL to avoid backward predictions in time. Using only validation animals, we created following linear model to assess herd cross-validation metrics according to Legarra and Reverter (2018):

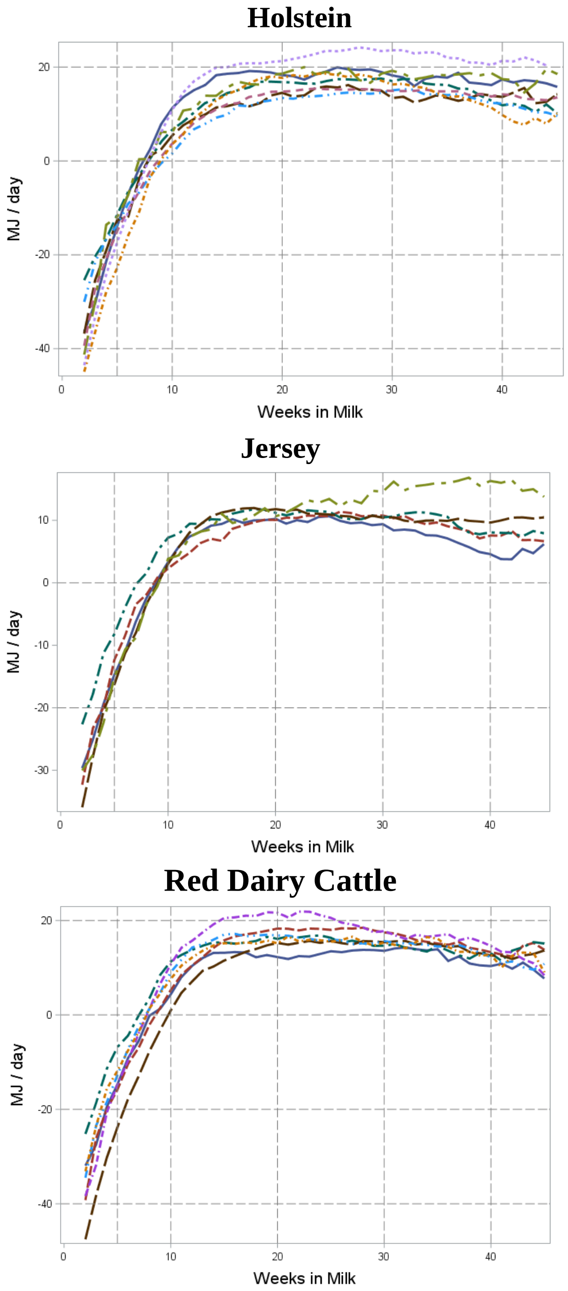
$$\mathrm{GEBV_{whole}} = \mu_{w,p} + ß_{w,p} \times \mathrm{GEBV_{partial}} + \varepsilon$$

where $\mathrm{GEBV_{whole}}$ was the GEBVs of validation animals with full phenotypic information, $\mu_{w,p}$ was the intercept (bias term), $ß_{w,p}$ was the slope (dispersion term), $\mathrm{GEBV_{partial}}$ was the GEBVs of validation animals with no phenotypic information and ε was the residual. From the linear model we also reported the correlation ($\boldsymbol{\rho_{w,p}}$) for the lactation-sum GEBVs of DMI and gRFI. Detailed information on how lactation-sum results were calculated can be found in Stephansen et al. (2024).

Results & Discussion

Figure 1 presents the average phenotypic level of $\mathrm{EB_{body}}$ through first lactation. For all breeds, the cows undergo a period of negative energy balance in early lactation, which becomes positive between 5-10 weeks in milk. These phenotypic results of $\mathrm{EB_{body}}$ in terms of level and pattern throughout first lactation are in line with the findings in Holstein and Jersey with experimental data (Thorup et al., 2018).

Variance component from the tested gRFI model can be found in Stephansen et al. (2024). Genomic validation results, using the Legarra-Reverter method, are presented in Table 1. To the authors' best knowledge, no studies have conducted by-herd cross-validation of GEBVs for feed efficiency traits in dairy cattle. We observed limited bias for DMI and gRFI in all breeds, comparing $\mu_{w,p}$ to the lactation-sum additive variance level. Acceptable $ß_{w,p}$ values was found for HOL and RDC, but overdispersion was observed in JER. Further research is needed for JER on the observed overdispersion, when more data is collected. Moderately high $\rho_{w,p}$ were found across breeds and highest for DMI (0.51-0.68) compared to gRFI (0.46-0.59). The pattern across breeds shows the highest cross-validation correlations were obtained for HOL, the breed with most cows, while lowest for JER, the breed with



Holstein

Jersey

Red Dairy Cattle

Figure 1. Lactation curves of energy balance calculated from changes in body reserves as MJ/day. Color and line pattern represents different herds.

smallest number of cows. The results suggest that genomic predictions of gRFI in herds with no phenotypic information can provide reliable GEBVs that can be used to generate genetic gain for feed efficiency.

This study aimed to validate the effect of GEBVs for gRFI in herds with no phenotypic information. However, it could be emphasized that not all phenotypic information would be missing, such as phenotypic information on milk production from a test-day recording scheme and BW records from some herds with milking robots or other measuring techniques (Lidauer et al., 2019). Future research should aim to investigate the effect of not having all phenotypic information missing in a herd cross-validation study, but as well validate the effect of missing information at different life stages, such as very young animals before first mating and only phenotypic information in very early lactation used for extension of the lactation. This can potentially be valuable information for management and breeding decisions on dairy farms, but as well for the breeding companies.

Table 1: Results from Genomic Legarra-Reverter validation using a herd cross-validation scheme for primiparous Nordic breeds. HOL = Holstein, JER = Jersey, RDC = Red Dairy Cattle, DMI = Dry Matter Intake, gRFI = genetic Residual Feed Intake, $\mu_{w,p}$ = intercept (bias term), ß$_{w,p}$ = slope of regression (dispersion term), $\rho_{w,p}$ = correlation between genomic breeding values with whole and partial phenotypic information for validation animals.

			Estimates	
Trait		HOL	JER	RDC
DMI	$\mu_{w,p}$	-0.36	3.64	-1.15
	ß$_{w,p}$	0.92	0.75	0.96
	$\rho_{w,p}$	0.68	0.51	0.66
gRFI	$\mu_{w,p}$	0.34	1.69	-2.00
	ß$_{w,p}$	0.87	0.69	0.85
	$\rho_{w,p}$	0.59	0.46	0.54

Conclusions

We aimed to evaluate gRFI genomically and test the feasibility of incorporating a novel energy sink trait for changes in body energy for Nordic primiparous cows, using data from the 3D camera-based system CFIT and the DCRC research herd. The genomic validation results show limited bias, and acceptable dispersion of predicted breeding values for HOL and RDC. However, overdispersion of predicted breeding values was observed in JER. Correlations between GEBVs from whole and partial datasets of validation cows, shows moderately high (0.46-0.59). These results show that selecting for gRFI GEBVs are expected to provide genetic gain of feed efficiency in other dairy herds.

Acknowledgments

The authors gratefully acknowledge the financial support provided by the CFIT project (9090-00083B, Innovation Fund Denmark). We thank the Nordic Genetic Cattle Evaluation for providing the genotypes for this project. Here we also thank Trine Andersen and Ulrik Sander Nielsen from SEGES Innovation (Skejby, Denmark) for their assistance on genotypes and phenotypes. We also thank the participating farmers who allowed the CFIT installation in their facilities. A special thanks to Per Madsen, QGG (AU, Denmark) for his feedback on the random regression modelling. Also, a special thanks to Nicolas Charles Friggens, PEGASE (INRAe, France) for his guidance in developing the EB_{body} phenotype for this study.

References

Aguilar, I., I. Misztal, D. Johnson, A. Legarra, S. Tsuruta, and T. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93(2):743-752. https://doi.org/10.3168/jds.2009-2730

Berry, D. P., M. P. Coffey, J. Pryce, Y. De Haas, P. Løvendahl, N. Krattenmacher, J. Crowley, Z. Wang, D. Spurlock, and K. Weigel. 2014. International genetic evaluations for feed intake in dairy cattle through the collation of data from multiple sources. *J. Dairy Sci.* 97(6):3894-3905. https://doi.org/10.3168/jds.2013-7548

Christensen, O. F. and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. GSE 42(1):2. https://doi.org/10.1186/1297-9686-42-2

De Haas, Y., J. Pryce, M. Calus, E. Wall, D. Berry, P. Løvendahl, N. Krattenmacher, F. Miglior, K. Weigel, and D. Spurlock. 2015. Genomic prediction of dry matter intake in dairy cattle from an international data set consisting of research herds in Europe, North America, and Australasia. *J. Dairy Sci.* 98(9):6522-6534. https://doi.org/10.3168/jds.2014-9257

Kennedy, B. W., J. H. Van der Werf, and T. H. Meuwissen. 1993. Genetic and statistical properties of residual feed intake. *J. Anim. Sci.* 71(12):3239-3250. https://doi.org/10.2527/1993.71123239x

Lassen, J., J. R. Thomasen, and S. Borchersen. 2023. Repeatabilities of individual measure of feed intake and body weight on in-house commercial dairy cattle using a 3D camera system. *J. Dairy Sci.* 106(12):9105-9114. https://doi.org/10.3168/jds.2022-23177

Legarra, A. and A. Reverter. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Gen.Sel.Evol.* 50:1-18. https://doi.org/10.1186/s12711-018-0426-6

Li, B., B. Berglund, W. F. Fikse, J. Lassen, M. H. Lidauer, P. Mäntysaari, and P. Løvendahl. 2017. Neglect of lactation stage leads to naive assessment of residual feed intake in dairy cattle. *J. Dairy Sci.* 100(11):9076-9084. https://doi.org/10.3168/jds.2017-12775

Lidauer, M. H., A. M. Leino, R. S. Stephansen, J. Pösö, U. N. Nielsen, F. W. Fikse, and G. P. Aamand. 2019. Genetic Evaluation for Maintenance – Towards Genomic Breeding Values for Saved Feed in Nordic Dairy Cattle. Pages 21-55 in Proc. *Interbull Bulletin*, NO. 55, Cincinnati, Ohio, USA.

Madsen, P. 2012. DmuTrace. in A program for extracting the pedigree for a subset of animals from a larger pedigree. Version. 2 ed, Center for Quantitative Genetics and Genomics.

Madsen, P. and J. Jensen. 2013. DMU A Package for Analysing Multivariate Mixed Models. Version 6, release 5.2. Accessed 15th June 2017, 2017. http://dmu.agrsci.dk/DMU/Doc/Current/dmuv6_guide.5.2.pdf

Mäntysaari, E., Z. Liu, and P. VanRaden. 2010. Interbull validation test for genomic evaluations. *Interbull bulletin* (41):17-17.

Stephansen, R. B., P. Martin, C. I. V. Manzanilla-Pech, B. Gredler-Grandl, G. Sahana, P. Madsen, K. Weigel, R. J. Tempelman, F. Peñagaricano, K. L. Parker Gaddis, H. M. White, J. E. P. Santos, J. E. Koltes, F. Schenkel, D. Hailemariam, G. Plastow, E. Abdalla, M. J. VandeHaar, R. F. Veerkamp, C. Baes, and J. Lassen. 2023. Novel genetic parameters for genetic residual feed intake in dairy cattle using time series data from multiple parities and countries in North America and Europe. *J. Dairy Sci.* 106(12):9078-9094. https://doi.org/10.3168/jds.2023-23330

Stephansen, R. B., J. Lassen, V. M. Thorup, B. G. Poulsen, J. Jensen, G. Sahana, and O. F. Christensen. 2024. Genetic Modeling of Productive Lifetime Feed Efficiency Integrating a Novel Sub-Model for Body Reserve Management in Nordic Dairy Cattle. *J. Dairy Sci.*, in review.

Su, G. and P. Madsen. 2011. User's Guide for Gmatrix. A program for computing Genomic relationship matrix. University of Aarhus, Tjele, Denmark.

Tempelman, R. and Y. Lu. 2020. Symposium review: Genetic relationships between different measures of feed efficiency and the implications for dairy cattle selection indexes. *J. Dairy Sci.* 103(6):5327-5345. https://doi.org/10.3168/jds.2019-17781

Thorup, V. M., S. Højsgaard, M. R. Weisbjerg, and N. C. Friggens. 2013. Energy balance of individual cows can be estimated in real-time on farm using frequent liveweight measures even in the absence of body condition score. *Animal* 7(10):1631-1639. https://doi.org/10.1017/S1751731113001237

Thorup, V. M., M. G. Chagunda, A. Fischer, M. R. Weisbjerg, and N. C. Friggens. 2018. Robustness and sensitivity of a blueprint for on-farm estimation of dairy cow energy balance. *J. Dairy Sci.* 101(7):6002-6018. https://doi.org/10.3168/jds.2017-14290

VandeHaar, M. J., L. E. Armentano, K. Weigel, D. M. Spurlock, R. J. Tempelman, and R. F. Veerkamp. 2016. Harnessing the genetics of the modern dairy cow to continue improvements in feed efficiency. *J. Dairy Sci.* 99(6):4941-4954. https://doi.org/10.3168/jds.2015-10352

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91(11):4414-4423. https://doi.org/10.3168/jds.2007-0980