Accuracy of genomic predictions of residual feed intake in Hereford with Uruguayan and Canadian training populations

O. Ravagnolo[1], I. Aguilar[1] , J.J. Crowley[23] , M.I. Pravia[4], M. Lema[1], F.L. Macedo[4], S. Scott[5], E.A. Navajas[4,1]

[1] Instituto Nacional de Investigación Agropecuaria. Programa Nacional de Producción de Carne y Lana, Canelones, Uruguay
oravagnolo@inia.org.uy (Corresponding Author)
[2]Canadian Beef Breeds Council, Calgary, Alberta, Canada
[3]Livestock Gentec, University of Alberta, Edmonton, Canada
[4] Instituto Nacional de Investigación Agropecuaria, Unidad de Biotecnología, Canelones, Uruguay
[5] Canadian Hereford Association, Calgary, Alberta, Canada

Summary

Dataset from Canadian and Uruguayan training populations were joined to analyse improvement of predictability for RFI. Three training populations where defined, URY (only data from Uruguay, 731), CAN (only data from Canada, 1168) and TOTAL (joint dataset, 1899). Genealogical information from the two countries was merged based on the international identification and cross reference list, with the pedigree file resulting in 17289 animals. Heritability estimates for RFI from the joined data was 0.26 ± 0.07, with σ_g^2=0.093 and σ_e^2 = 0.262. Four different panels were imputed to Illumina 50k.Accuracy of genomic predictions was assessed by cross validation with Random and Kmeans clustering methods. Average BIFacc (1-sqrt(PEV/ σ^2_a) of the genomic predictions obtained were very low with the lowest being URY (0.024 and 0.056 respectively),and the highest TOTAL (0.068 and 0.105 respectively). As was expected, random accuracies were higher than Kmeans accuracies. Accuracies (r(adjPhen,DGV)/h) increased proportionally with the size of the training population from 0.017 and 0.163 for Kmeans and random approach respectively with URY, to 0.245 and 0.304 respectively with CAN. Smaller accuracy obtained in TOTAL was due to URY animals, where bulls and steers showed very different results, for bulls from 0.17 to 0.21 and for steers from 0.14 to 0.033. A benefit for joining training populations was observed, but mostly for the smallest population. Difference within URY emphasises need to ensure relatedness within the training population and most of all, with the population for which predictions are aimed.
Keywords: genomic selection, feed efficiency, training population, beef cattle, accuracy

Introduction

The demands for livestock products are increasing, and beef production seems not to be an exception. This implies a challenge to beef production that has to increase productivity without increasing area or environmental footprint (a finite commodity), increasing costs (competing in disadvantage with chicken and pigs) or lowering product quality (its main advantage). Genetic improvement of feed efficiency is relevant given the potential of lowering feeding costs (main production cost in most production systems) and seems possible with the advent of new genomic tools (Hayes *et al.*, 2013). Given the relevance of this

challenge for Uruguay, a large national project with the main goal of enhancing competitiveness of the Uruguayan beef industry by implementing genomic tools to genetically improve feed efficiency has been in place since 2014 (Navajas *et al*., 2014), enabling collection of individual feed intake in Hereford bulls and steers. Similarly, since 2014, the Canadian Hereford Association has been performing a similar task. With the close ties and ready exchange of genetics between the two countries, a joint genetic/genomic evaluation was obvious next step.

Merging data from different origins (experimental stations, commercial herds, countries) for feed efficiency has been one way of increasing prediction accuracy of conventional (EPD) and genomic expected progeny differences (GEPD) for difficult and costly to measure traits as individual feed intake in cattle (Berry *et al*., 2014,).

In 2016, Canada and Uruguay joined their feed intake information with the objective of implementing an international genetic evaluation for feed efficiency and investigating the impact of integrating both countries training populations on the prediction of genomic EPD, with the advantage that both countries used same trait definition and measurement protocols. Traditional EPD for Residual Feed Intake (RFI) with the use of information for both countries were published in 2016.

The objective of present study was to compare the accuracy of genomic predictions for RFI based on national and bi-national training populations.

Material and methods

Animals and phenotypes

Individual feed intake was recorded in post weaning tests of 70 days, after 28 days of acclimatization to diet and feeding system, using an automated feeding system (Growsafe) following Beef Improvement Federation guidelines (BIF 2006) as described by Pravia *et al*. (2015).

Canadian dataset comprised of 1168 bull records coming from 104 herds in 8 tests carried out from 2012 to 2016, from 309 sires. Uruguayan data included records of 498 bulls and 233 steers coming from 45 studs and 3 commercial steer producers in 10 tests carried out from 2014 to 2016, from 256 sires. Each stud had to provide at least 3 bulls from 2 different sires in order to preserve contemporary group definition by the Pan American genetic evaluation run by ABRI.

Three training populations where defined, URY (only data from Uruguay, 731), CAN (only data from Canada, 1168) and TOTAL (joint dataset, 1899).

Trait definition and model

Performance test datasets from both countries (described in Table 1) were pooled and the following model was used to analyse dry matter intake (DMI) with the residuals taken to represent RFI;

$$DMI_{ijk} = CG_j + b_1*ADG_i + b_2 * MWt_i + b_3\ BFat_i + e_{ijk} \quad (1)$$

Where DMI was calculated as the arithmetic mean of daily dry matter intake for each animal from valid days on test (kg).

CG was defined by Trial + PEN (j=1-62)

ADG was average daily gain calculated by regression using all weights (kg/day)

MWt was the metabolic weight defined as mid test weight ^0.75 (kg)

Bfat was the subcutaneous fat depth measured at the end of test by ultrasound (mm).
b_1, b_2 and b_3 were the partial regression coefficients for each trait on DMI

Table 1. Description of datasets from each country

Trait	CAN		URY	
	Mean	SD[1]	Mean	SD
Start Age, d	275.93	28.64	290.70	28.713
DMI, kg/d	8.556	0.901	8.930	1.360
ADG, kg/d	1.398	0.489	1.428	0.495
MWt, kg	91.647	7.589	72.919	8.747
BFat, mm	4.817	1.873	3.611	1.110
RFI[2], kg/d	-0.003	0.556	-0.001	0.6041

[1] SD: standard deviation
[2]RFI obtained from Model 1

Estimation of heritability

Variance components were estimated for RFI using the following model:

$$RFI = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (2)$$

Where RFI is the residuals resulting from model 1, **X** is the incidence matrices for fixed effects **β**, **Z** is the matrix linking animals to phenotypes, **u** is the vector of random animal effects, and **e** is the vector of random residual effects.

Fixed effects were age of the animal in days at the start of the test (171-396), age of dam as a class variable (1-5, 5 corresponding to unknown) and contemporary group defined by Trial+pen (1-62).

It was assumed that $\mathbf{u} \sim N(0,\mathbf{A}\sigma^2_g)$ where A is the additive relationship matrix and σ^2_a is the genetic additive variance, and $\mathbf{e} \sim N(0,\mathbf{I}\sigma^2_e)$ where **I** is the corresponding identity matrix and σ^2_e the residual variance.

AGE was the age of the animal in days at the start of the test (171-396).
Aod was age of the dam as a class variable (j=1-5, class 5 was unknown age of dam)
CG was defined by Trial+PEN (k=1-62)
Animal was the random effect of animal l (l=1 - 2463)

Genealogical information from the two countries was merged based on the international identification and cross reference list. The pedigree file included five generations of ancestors of the animals with RFI, with a total of 17289 animals. Estimation of variance component and heritability was carried out using AIREMLF90 (Misztal *et al*., 2002). Resulting variance components were used as input parameters for the genomic predictions.

Genotypes

Table 2 provides information of genotypes used for the analysis. All panels were imputed to the 50k SNP panel (BovineSNP50, Illumina San Diego, CA) using Fimpute (Sargolzaei *et al*, 2014) with available pedigree and genotypic information. All genotypes with call rate>=0.85 and MAF>=0.01 were used, only autosomes were considered, resulting

in 45383 SNP for the analyses.

Table 2. Description of genomic data

Trait	SNP[1]	CAN	URY
Illumina BovineSNP50v2 BeadChip 50k	52,890	329	0
Illumina BovineHD BeadChip 700k	47,813	0	731
GeneSeek Genomic Profiler HD 80k	27,735	525	0
GeneSeek Genomic Profiler HDv2 140k	40,288	314	0

[1] Number of common SNPs with Illumina 50k

Genomic predictions: description of model and methodology

Direct genomic values (DGV) predictions were obtained using GBLUP as implemented in BLUPF90 (Misztal *et al*., 2016).

$$\mathbf{RFI} = \mathbf{X}\beta + \mathbf{Zg} + \mathbf{e} \quad (3)$$

Where RFI is as used in **Model 2**, **X** is the incidence matrix for fixed effects (β) also as defined in **Model 2**, **Z** is the matrix linking animals to phenotypes, **g** is a vector of genomic breeding values, and **e** is a vector of residual effects. It was assumed that $\mathbf{g} \sim N(0,\mathbf{G}\sigma^2_g)$ where **G** is the genomic relationship matrix (VanRaden 2008) and σ^2_g is the genetic additive variance, and that $\mathbf{e} \sim N(0,\mathbf{I}\sigma^2_e)$ where **I** is the corresponding identity matrix and σ^2_e the residual variance.

Prediction ability

Accuracy of genomic predictions was assessed by cross validation, by dividing the datasets into four groups. Three of the groups were used to predict the fourth and this was repeated until all four datasets had been predicted. For the clustering two methods were used:

1. Random, where dataset was randomly divided into the four groups, this process was repeated 30 times.
2. Kmeans, as described by Saatchi *et al.* (2011), where the clustering maximized the genomic relationship within each set and minimized it across the other sets.

Both analyses were carried out for three scenarios, using only CAN, only URY or using TOTAL datasets.

Accuracy was defined as r(adjPhen,DGV)/h where adjPhen is the phenotype adjusted by fixed effects and h the square root of the heritability (Legarra *et al.*, 2008) . Phenotypes were adjusted using a BLUP analysis using Model 2.

BIF accuracy

BIF accuracies were calculated using prediction error variance (PEV) from the inverse of the mixed model equations, BIFacc= 1-sqrt(PEV/ σ^2_a. Only accuracies for animals in the predicted set were reported.

Results and Discussion

Heritability estimates for RFI from Model 2 were 0.26 ± 0.07, with σ_g^2=0.093 and σ_e^2 = 0.262, somewhat lower than the pooled heritability obtained by Berry and Crowley (2013), but within the range of most in their study studies (pooled heritability for RFI 0.33, range from 0.07 to 0.62). Estimated variances from Model 2 were used in all further analysis.

Average BIFacc of the genomic predictions obtained when no phenotype was used were very low for the three training populations, with the lowest being URY, the smallest dataset, and the highest TOTAL, the largest dataset. In all three cases, as was expected, random accuracies were higher than Kmeans accuracies (Table 3).

Table 3. BIF and cross validation accuracy for three training populations

Training Population	N	BIF accuracy[1]		Cross validation accuracy[2]				
		Kmeans	Random	Kmeans	Random			
					All	Can[3]	Ub[4]	Ust[5]
URY	731	0.024	0.056	0.017	0.163		0.17	0.14
CAN	1168	0.043	0.096	0.245	0.304	0.304		
TOTAL	1899	0.068	0.105	0.171	0.243	0.298	0.210	0.033

[1] 1-sqrt(PEV/ σ^2_a).

[2] r(adjPhen,DGV)/h where adjPhen is the phenotype adjusted by fixed effects

[3,4,5] Can are bulls from CAN dataset, Ub are bulls and Ust are steers respectively from URY dataset .

When comparing CAN and URY results, predictability increased proportionally with the size of the training population where accuracies increased from 0.017 and 0.163 for Kmeans and random approach respectively with URY, to 0.245 and 0.304 respectively with CAN.

As was observed for BIFAcc, in general, Kmeans scenario predictions were lower than the random scenario predictions due to the design of the training populations used for the prediction.

The relationship between size and accuracy was not observed for TOTAL results, which had lower accuracies, of 0.243 for random scenario, than CAN, in spite of having a larger population size (1899 vs 1168, Table 3). When analysing results (TOTAL) for animals within each country, CAN animals (bulls) had an accuracy of 0.298, very similar to the one obtained when only the CAN population was used as training population. So the smaller accuracy obtained in TOTAL was due to URY animals, where bulls and steers show very different results. In the case of URY, bulls accuracies increase from 0.17 to 0.21 (Ub for URY vs Ub for TOTAL), while for steers, accuracies decreases from 0.14 to 0.033 (Us for URY vs Us for TOTAL). We hypothesise that the URY steer population is less related to the bulls of CAN and URY, and that the accuracy of 0.14 for steers when using only URY as training population is due to them representing a large proportion of the training population (32%) compared to TOTAL where they are a small proportion (12 %). The smaller relatedness of steers might also be an explanation for the large difference observed between Kmeans and random approach only observed for URY.

Conclusion

Although predictabilities were in general low, the magnitudes were moderate with the larger bi-national training population, with more than 1000 records, confirming once again the relevance of continuing to increase training populations for RFI.

A benefit for joining training populations was observed, but the benefit was mostly for the smallest population, while for the larger population, average accuracy of predicted animals increased slightly, thus yielding a relatively small benefit.

Difference within URY emphasises need to ensure relatedness within the training population and most of all, with the population for which predictions are aimed.

List of References

Berry, D.P and Crowley, JJ. CELL BIOLOGY SYMPOSIUM: Genetics of feed efficiency in dairy and beef cattle. J. Anim. Sci. 2013. 91:1594-1613.

Berry DP, Coffey MP, Pryce JE, De Haas Y, Løvendahl P, Krattenmacher N, Crowley JJ, Wang Z, Spurlock D, Weigel K, Macdonald K. International genetic evaluations for feed intake in dairy cattle through the collation of data from multiple sources. Journal of Dairy Science. 2014 Jun 30;97(6):3894-905.

Crowley JJ, McGee M, Kenny DA, Crews DH, Evans RD, Berry DP. Phenotypic and genetic parameters for different measures of feed efficiency in different breeds of Irish performance-tested beef bulls. Journal of Animal Science. 2010 Mar 1;88(3):885-94.

Hayes, B.J, H.A. Lewin, M.E. Goddard, 2013. The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. Trends Genet; 29: 206-214

Legarra, A., C. Robert-Granie, E. Manfredi, and J.-M. Elsen. 2008. Performance of genomic selection in mice. Genetics 180:611–618.

Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH. BLUPF90 and related programs (BGF90). InProceedings of the 7th world congress on genetics applied to livestock production 2002 Aug 19 (Vol. 28, No. 07).

Misztal, I., S. Tsuruta, D. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2016. Manual for BLUPF90 family of programs. Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

Navajas EA, Pravia MI, Lema M, Clariget J, Aguilar I, Ravagnolo O, Brito G, Peraza P, Rizza MD, Montossi F. Genetic improvement of feed efficiency and carcass and meat quality of hereford cattle by genomics. Archivos Latinoamericanos de Producción Animal. 2014;22(5):13-6.

Pravia, M.I., G. Ciappesoni, I. Aguilar, O. Ravagnolo, M.Lema, EA Navajas. Analysis of feed intake and feed efficiency in Hereford Bulls and steers. 52a Reunião Anual da Sociedade Brasileira de Zootecnia. 2015. Belo Horizonte – MG, Jul 19-23 de Jul.

Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. BMC genomics. 2014 Jun 17;15(1):478.

Saatchi M, McClure MC, McKay SD, Rolf MM, Kim J, Decker JE, Taxis TM, Chapple RH, Ramey HR, Northcutt SL, Bauck S. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. Genetics Selection Evolution. 2011 Nov 28;43(1):40.

VanRaden, P. 2008. Efficient methods to compute genomic predictions. J Dairy Sci. 91:4414–4423.