Collation of Data and Genetic Parameter Estimation in Different Experimental Canadian Beef Cattle Populations Measured for Feed Efficiency

J.J. Crowley[1], P. Stothard[1], J. Basarab[1,2,3], S.P. Miller[1,4,5], C. Li[1,3], Z. Wang[1], G. Plastow[1], M. de Pauw[1], S.M. Moore[5] and D.B. Lu[1]

[1]University of Alberta, AB, Canada, [2]Alberta Agriculture and Rural Development, [3]Agriculture & Agri-Food Canada, [4]AgResearch, Invermay, New Zealand, [5]University of Queensland, Centre for Animal Science, QAAFI, Australia

ABSTRACT: Improvement in feed efficiency (FE) can contribute to a large increase in profitability of a beef production system but its measurement requires considerable expense and time. Hence, there is significant merit in combining existing FE databases for further genetic analyses. Four experimental datasets were collated from the University of Alberta, University of Guelph, Alberta Agriculture and Rural Development, and Agriculture and Agri-Food Canada which summed to 7317 FE records after edits. Residual feed intake (RFI) and residual intake and gain (RIG) were calculated across the entire dataset as measures of FE. (Co)variance components were estimated between datasets. Heritability of RFI across all datasets was 0.41 and varied from 0.29 to 0.48 within dataset. Genetic correlations between datasets for RFI ranged between 0.77 and 0.86 indicating that it is appropriate to pool data from the aforementioned datasets. Similarly, genetic correlations for RIG ranged from 0.75 to 0.85.

Keywords: beef cattle; feed efficiency; pooling data

Introduction

Feed intake is one of the largest variable costs in beef production and so an improvement in feed efficiency contributes to increased profitability. However, both the economic cost of measuring feed intake and the length of time it takes to build a dataset of any relevance prohibits large scale analysis of the trait and subsequent selection for its improvement. Furthermore, in the age of genomics, phenotypic data becomes more valuable and owing to the expense of data recording in the case of feed efficiency, genomics can play a large role in utilizing all phenotypes available. Many institutions and research groups pursue the analysis of feed efficiency and so there are datasets of moderate size in existence that can benefit from amalgamation. Previous studies in both dairy and beef (Banos et al., 2011; De Haas et al., 2012; Berry et al., 2013; Bolormaa et al., 2013) have investigated the viability of collating datasets resting in smaller repositories into one larger dataset for subsequent genomic analysis and breeding value estimation. De Haas (2012) showed that accuracy of genomic selection improved after collating data on DMI from several research populations.

More specific to this study, the Canadian Cattle Genome Project was set up to pursue the collection of whole genome sequence data on the Canadian beef population by targeting influential animals within the major breeds. Concurrent to this, one of the phenotypes of interest in the project is feed efficiency and the whole genome sequence information of the animals with this valuable phenotype. For a dataset of feed efficiency phenotypes in this project it was proposed to collate data from 4 separate research populations in Canada (3 in Alberta [AB]; 1 in Ontario [ON]). The objective of this study was to assess the appropriateness and viability of collating information from different sources in order to create a phenotypic database that will lend itself to future genetic and genomic analyses.

Materials and Methods

Data. The aforementioned 4 data sources were as follows: (1) an Angus (AN)/Charolais (CH) research herd previously located at OneFour research substation at Lethbridge, AB, now at the University of Alberta research station, Kinsella, AB (ANC); (2) the hybrid research herd of the University of Alberta, (KIN); (3) data emanating from the Phenomic Gap project at Lacombe Research Centre, AB (PG); and (4) data from University of Guelph's Elora Beef Research Centre, ON (UoG). Data from all sources summed to a total of 8740 animals. Detailed descriptions of the populations, treatments and data edits can be found in Mao et al. (2013), Durunna et al. (2012), Akanno et al. (2014) and Lu et al. (2013), respectively.

Briefly, ANC were purebred AN and CH totaling 1644 steers and heifers. Animals were performance tested post weaning for approximately 120 days between the years 2005 and 2012 where feed intake (FI) was measured daily, bodyweight (BW) measured every other week and ultrasound fat measures (BFAT) taken every 28 days. In the KIN dataset, animals were 923 crossbred steers, the breed composition of which was AN, CH, dairy and beef synthetics and hybrids originating on the Kinsella ranch (Nkrumah et al., 2004). Performance test periods were approximately 120 days between 2004 and 2009. Feed intake was measured daily, BW every other week and BFAT at the start and end of test. In the PG dataset, 4453 bulls, steers and heifers comprised of AN, CH, Hereford (HE), Gelbvieh (GV), Limousin (LI), Red Angus (AR), Shorthorn (SS) and Simmental (SM) breed fractions, and TX and M4 strains from Beefbooster Inc. Test periods varied from 76 to 112 days following adjustment periods of 28 to 36 days from 2003 to 2013. Feed intake was measured daily, BW measured on two consecutive days at the start and end of the test period and at approximately 28-d intervals throughout the test period and BFAT was measured at the start and end of test. Finally, the UoG dataset was comprised of AN, CH, GV, LM, Piedmontese (PI) and SM breed fractions. Data was collected on bulls, steers and heifers between 2000 and 2011. Performance test

periods averaged 111 days after adaption periods of 28 to 36 days. Bodyweight was recorded at 28d intervals, DMI daily and BFAT once at the start and once at the end of test.

Trait definitions and data edits. In the ANC, KIN and PG datasets, FI was collected daily using the GrowSafe system (GrowSafe Systems Ltd., Airdrie, Alberta, Canada). In the UoG dataset, feed intake was collected using either the Insentec system (Insentec, B.V. Marknesse, Netherlands) or Calan-Broadbent feeding doors (American Calan, Northwood, NH). Average FI and dry matter intake (DMI) was subsequently calculated on a per animal basis. In all datasets, ADG was estimated by regressing BW on day of test and subsequently mid-test metabolic liveweight ($BW^{0.75}$) was estimated from this regression. The BFAT measure used in the calculation of residual feed intake (RFI) was final ultrasound BFAT measured at the end of test. Data for this study was received from the multiple sources in a single record/animal format i.e. average DMI and ADG already computed.

While data was received in a relatively clean format from each source, collating this many records from different origins necessitated further data edits. For this study, where feed efficiency is the trait of focus, it was deemed necessary to include BFAT in the equation estimating RFI to accurately represent feed efficiency. To this end, a large number of observations (n=950) collected in the earlier years were omitted in the absence of this phenotype. Furthermore, animals with a missing observation of any of the traits or model effects of interest were deleted (n=144) as were animals older than 450d at the start of test (n=139) and any record that was greater than 3 SD from the mean estimated within dataset of any or all of ADG, DMI, $BW^{0.75}$ and BFAT (n=138). Finally, only animals in a contemporary (CG) with 5 or more records were retained (n=7317).

After data edits, RFI was assumed to represent the residuals from a multiple regression model regressing DMI on ADG, BFAT and $BW^{0.75}$ with a CG effect included (defined as dataset, test year, group). Subsequently, RIG was defined as RG - RFI, both standardized to a variance of 1, where RG = residual gain, the residuals from a multiple regression model regressing ADG on DMI, BFAT and $BW^{0.75}$ with CG included.

Genotypes. DNA was available on 7002 animals with feed efficiency information. Genotyping was performed on DNA extracted from blood or tissue, using either version 1 or 2 of the Illumina BovineSNP50 BeadChip. SNP with spurious position and/or less than 95% call rates were excluded, leaving a final number of 42204 SNP on 29 Bos Taurus autosomes (BTA) being used in this study. The number of SNP varied among chromosomes, with BTA1 having the largest number of SNP (2778), and BTA27 having the fewest (762).

Model and (co)variance components. Least square means for feed efficiency and its component traits were estimated among datasets. Fixed effects in the model included sex (bull, steer or heifer), breed (classed as primary, secondary breed), test year, dataset, age at the start of test, a quadratic effect of age and a sex by age interaction. Using ASReml software (Gilmour, 2010), a linear animal model was used to estimate (co)variance components. Fixed effects were similar to those previously explained with the explicit omission of test year and data source and the inclusion of CG and animal as random effects. Pedigree recording was sparse in the majority of the data and there were negligible pedigree links between datasets. Hence, relationships between animals were accounted for by fitting a genomic relationship matrix which was generated using method 1 of VanRaden's approach (VanRaden, 2008) with a slight modification. Briefly, animals from within a dataset were put into subgroups based on their sire's breed. Each subgroup had their own vector P being $2(p_i - 0.5)$, where p_i was the frequency of allele B at locus *i* within that subgroup. Matrix M had animals on the rows and genotypes in the columns, coded -1, 0, 1 for zero, 1 and 2 copies of allele B. Matrix Z was created by subtracting each row of M with its corresponding P. Then $G = \frac{zz'}{\frac{2}{N}\sum_{j=1}^{a} p_{ij}(1-p_{ij})n_j}$ where *N* was the total number of animals from all different datasets, n_j the number of animals in the j^{th} subgroup, *a* the number of subgroups. Genetic correlations were estimated for RFI and RG between different datasets to assess the viability of pooling data and to identify if a genotype by environment interaction exists.

Results and Discussion

Differences in means for each trait existed across at least 3 data sources per trait except for RFI and RIG (Table 1). No difference in mean existed between datasets for both feed efficiency traits due to their definition. While no difference in means existed for RFI across datasets,

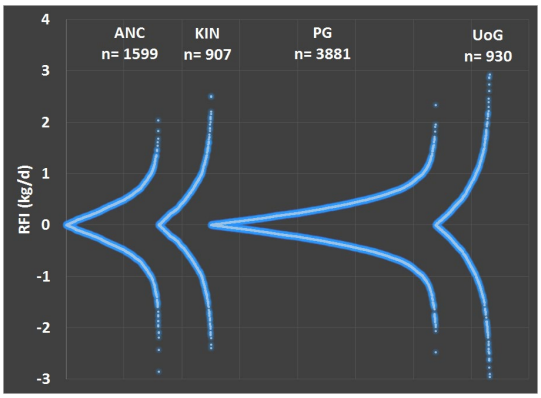

Figure 1 shows slight difference in the phenotypic variation within each dataset. The UoG dataset shows the greatest variation in RFI but is not surprising given the means of the component traits. The heritability for RFI estimated in the collated dataset was 0.41 (Table 2) and ranged from 0.29 (UoG) to 0.48 (ANC) when estimated within dataset. While RFI is represented by the residual term from a regression and can contain both actual differences in feed efficiency and other error terms, the heritability of RFI across datasets was consistent with the literature (Berry and Crowley,

2013) which gives confidence to correct modelling of the trait. Genetic correlations between datasets for RFI ranged from 0.77 to 0.86 and from 0.75 to 0.85 for RIG (Table 3). These correlations are of a magnitude where one is confident to say that a pooling of these separate datasets is appropriate in order to increase dataset size to pursue further genome wide association studies and genomic predictions. It is not surprising that the correlations between datasets are high; while raised in different locations under different protocols, the post weaning test period protocol is well established minimizing any environmental variance.

Conclusion

Genetic correlations between datasets and similar proportions of genetic variation indicate that it is indeed suitable to collate the separate datasets investigated. This will facilitate future genetic analysis studies for feed efficiency and has set the foundation for an ever expanding phenotypic database. Also, it opens the door for international collaboration as there are no doubt many similar but separate datasets in different countries.

Literature Cited

Akanno, E., Plastow, G., Lu D. et al. (2014) J. Anim, Sci. (submitted)

Banos, G., Coffey, M. P., Veerkamp, R. F. et al. (2011). Animal 6:1040–1048.

Berry, D. P., Coffey, M. P., Pryce, J. E. et al. (2014). J. Dairy Sci., (submitted).

Berry, D. P. and Crowley, J. J. (2012). J. Anim. Sci. 90:109-115

Bolormaa, S., Pryce, J. E., Kemper, K. et al. (2013). J. Anim. Sci. 91:3088-3104.

de Haas, Y., Calus, M. P. L., R. F. Veerkamp et al. (2012). J. Dairy Sci., 95 :6103–6112.

Durunna, O., Plastow, G,. Mujibi, F. D. N et al. (2011). J. Anim. Sci.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., and Thompson, R. (2009) ASReml user guide release 3.0.

Lu, D., Miller, S., Sargolzaei, M. et al. (2013) J. Anim. Sci. 91:3612-3633

Mao, F., Chen, L., Vinsky, M. et al. (2013) J. Anim. Sci. 91:2067-2076

Nkrumah, J. D., Sherman, E. L., Li C. et al. (2007) J. Anim. Sci. 85(12): 3170–3181

VanRaden, P. M. (2008). J. Dairy Sci. 91:4414-4423

Acknowledgements

Genome Canada, Genome Alberta, Alberta Livestock and Meat Agency Ltd., Alberta Innovates Bio Solutions, Alberta Innovates Technology Futures, Alberta Agriculture and Rural Development, Western Economic Development, Agriculture and Agri-Food Canada, Beefbooster Inc., Ontario Ministry of Agriculture, Food and Rural Affairs, the Ontario Cattlemen's Association, the Ontario Farm Innovation Program, the Agriculture Adaptation Council, Ontario Research and Development Program, MITACS Accelerate, and Beef Improvement Opportunities are gratefully acknowledged for funding the project.

Table 1. Least square means of performance and feed efficiency traits[1] among different data sources

	Mean	ANC	KIN	PG	UoG	SE[2]
n	-	1599	907	3881	930	-
Start age, d	299	312[a]	301[b]	297[bc]	284[c]	2.82
DMI, kg/d	9.22	9.31[a]	9.98[b]	8.72[c]	10.39[d]	0.07
ADG, kg/d	1.46	1.40[a]	1.62[b]	1.33[c]	1.96[d]	0.02
BW, kg	430	430[a]	454[b]	420[c]	457[b]	3.32
BFAT, mm	8.03	9.46[a]	6.24[b]	6.13[b]	14.66[c]	0.23
RFI, kg/d	0	0.10	-0.02	-0.02	-0.06	0.05
RIG	0	-0.28	-0.04	-0.01	0.34	0.11

Within a row, means without a common superscript differ ($P < 0.05$)

[1]DMI= average dry matter intake, ADG= average daily gain, BW = mid-test bodyweight, BFAT = final ultrasound backfat, RFI = residual feed intake and RIG = residual intake and gain

[2]ANC = Angus/Charolais dataset; KIN = Kinsella research station dataset; PG = Phenomic Gap dataset; UoG = University of Guelph dataset

[3]Pooled standard error

Table 2. Heritabilities with standard errors in parenthesis and genetic SD for residual feed intake (RFI) and residual intake and gain (RIG)

Data Source		RFI	RG
Total, n=7317	h^2	0.41 (0.07)	0.35 (0.08)
	σ^2_a	0.26	0.59
ANC, n=1599	h^2	0.48 (0.09)	0.40 (0.11)
	σ^2_a	0.28	0.60
KIN, n=907	h^2	0.35 (0.11)	0.29 (0.13)
	σ^2_a	0.27	0.58
PG, n=3881	h^2	0.47 (0.06)	0.41 (0.07)
	σ^2_a	0.24	0.59
UoG, n= 930	h^2	0.29 (0.10)	0.24 (0.14)
	σ^2_a	0.27	0.59

Table 3. Genetic correlations (standard errors in parenthesis) between different populations[1] for residual feed intake (RFI; above the diagonal) and residual intake and gain (RIG; below the diagonal)

	ANC	KIN	PG	UoG
ANC	1	0.80 (0.16)	0.86 (0.14)	0.79 (0.20)
KIN	0.78 (0.17)	1	0.83 (0.16)	0.77 (0.19)
PG	0.85 (0.16)	0.82 (0.16)	1	0.83 (0.16)
UoG	0.78 (0.19)	0.75 (0.20)	0.8 (0.16)	1

[1]ANC = Angus/Charolais dataset; KIN = Kinsella research station dataset; PG = Phenomic Gap dataset; UoG = University of Guelph dataset