Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Accuracy of Molecular Breeding Values for Production and Efficiency Traits of Canadian Crossbred Beef Cattle Using a Cross-Validation Approach

E. C. Akanno[1], G. S. Plastow[1], C. Li[1, 2], S. P. Miller[1, 3] and J. A. Basarab[1, 4]
[1]Livestock Gentec, University of Alberta, Edmonton, AB, [2]Lacombe Research Centre, Agriculture and Agri-Food Canada, Lacombe, AB, [3]Centre for Genetic Improvement of Livestock, University of Guelph, ON, [4]Lacombe Research Centre, Alberta Agriculture and Rural Development, Lacombe, AB, Canada

ABSTRACT: A total of 4,218 animals genotyped for 50 K SNPs were used to evaluate accuracy of molecular breeding values (MBVs) of Canadian crossbred beef cattle traits. A cross-validation approach that involves random sampling of all animals to form 5-groups was applied. One group was dedicated as a validation set and the remaining four groups were combined to form a reference set. Genomic best linear unbiased prediction method was applied for training and prediction of MBVs in the validation set. Estimated breeding values (EBV) were used as the phenotype and a bivariate animal model was applied to each trait to estimate the genetic correlation between MBV and EBV. Accuracy of MBVs ranged from 0.05 to 0.68 for the studied traits. The results suggest that genomic predictions can be produced for crossbred beef cattle with greater efficiency for selection and management especially in younger animals.
Keywords: feed efficiency; growth; SNP

Introduction

The application of molecular breeding values (MBVs) as a selection tool for genetic improvement of beef production traits is currently receiving attention in the beef cattle industry. Accuracy of MBVs for important traits of beef cattle remains a major factor influencing a breeder's decision to adopt a genome-based selection method. Previous research (Meuwissen et al. (2001); Habier et al. (2007)) has shown that trait heritability, linkage disequilibrium between markers and underlying quantitative trait loci, genetic relationship among individuals and the applied statistical method are among the factors that affects accuracy of MBVs realizable for a given trait. In addition, the multibreed and crossbred structure of beef cattle populations can create divergence between individuals in the reference and implementation populations and thus, may lower accuracy of selection (Kachman et al. (2013)). A cross-validation approach can be used in conjunction with a statistical method to exploit the relationship among individuals in order to improve accuracy (Luan et al. (2009)). Using a cross-validation procedure makes the assumption that phenotypes related to the validation animals are omitted when marker effects are trained exclusively in the reference set. Furthermore, a genome-based selection method has the advantage that MBV can be estimated early in life of the animal and thus, may improve accuracy of selection for younger animals. Therefore, the objective of this study was to evaluate accuracy of MBVs for production and efficiency traits of Canadian crossbred beef cattle using a cross-validation approach.

Materials and Methods

Data: A total of 4,218 animals that were part of the Phenomic Gap project as previously described by Akanno et al. (2014) (Submitted to J. Anim. Sci.) were genotyped with BovineSNP50 BeadChip (Illumina, San Diego, CA) at Delta Genomics (Edmonton, AB). Genotypes at a particular locus were filtered from further analysis based on selection criteria of at least 95% animal call rate and 0.1 minor allele frequency, resulting in a total of 49,531 SNPs left for analysis. A few animals had missing genotypes and these were imputed using Beagle (Browning and Browning (2009)). Pedigree record was available and consisted of 7,490 individuals, 449 sires and 2,942 dams. Only four animals were inbred with average inbreeding of 0.25. All animals used in this study were spring born between 2002 and 2012 and had phenotypes for post-weaning traits. The analyzed traits included average daily gain (ADG), dry matter intake (DMI), mid-test metabolic weight (MMWT), feed conversion ratio (FCR), residual feed intake (RFI), residual BW gain (RG) and residual intake and BW gain (RIG). The method described by Berry and Crowley (2012) were used to calculate RFI, RG and RIG for all animals.

Models: A traditional animal model was used to estimate breeding values (EBVs) for all animals using their trait phenotypes as response variables. Fixed effects in the model were contemporary group (sex, herd of origin, birth year, diet and management) and covariate for age at the end of test. The EBVs were not adjusted for breed in order to allow for across breed comparison with the genome-based model. The predicted EBVs from the animal model were used as response variables in the genome-based analysis. The reliability (R^2) of the EBVs was estimated as $1-(SE^2/\sigma^2_g)$, where SE and σ^2_g are the standard error of EBV and genetic variance estimated from animal model using ASreml (Gilmour et al. (2009)). Heritability of the traits based on a single trait animal model was also reported. The genome-based model for computing MBV was:

$$\mathbf{y} = \mu + \mathbf{Wg} + \mathbf{e}$$

where $\mathbf{y}$ is a vector of the phenotypes (EBV for each trait) in the reference set, μ is the overall mean, $\mathbf{W}$ is the incidence matrix for the random genomic effect, $\mathbf{g}$ is a vector

containing a random genomic effect for each animal distributed as ~ N (0, $\mathbf{G}\sigma^2_g$), where **G** is a marker based relationship matrix, which was built according to VanRaden (2008) based on all available SNPs (n = 49,531) and **e** is a vector of random residual term. As can be seen in the model design, SNP effects were not estimated for each single SNP, but the genomic relationship matrix was used to model a genomic effect for each individual. It was thus possible to predict MBV for animals in the validation set given their genotypes for each replicate. All analysis was performed using the gebv software (Sargolzaei et al. (2009)).

Cross-validation procedure: A mutually exclusive random sampling of all animals from the whole data set was performed to form 5-groups replicated 5 times with an average of 834 animals per group. In each analysis within a replicate, one group was dedicated as validation set while the remaining four groups were combined to form the reference set. The EBVs of the animals in the validation set were assumed to be unknown thus, it resulted in every animal having MBVs that were predicted without using its own EBV, allowing their EBV to be used for validation. The whole procedure was repeated 5 times.

Validation on younger animals: In another scenario, animals born from 2002 to 2010 were used as the reference set (n = 3,017) while the later born (2011-2012) animals were dedicated as the validation set (n = 1,201). Here, a new EBV was computed for the older animals using an animal model excluding phenotypes for the younger animals, thus parent average (PA) and their R^2 were generated for the younger animals in the validation group.

Validation analyses: For each of the traits and replicates in the cross-validation part, a two-trait animal model was applied using the MBVs and EBVs for all animals to estimate variance and covariance components. The model included only the mean for the MBV and EBV, respectively. The purpose of fitting this model was to estimate the genetic correlation between the MBV and the EBV which was taken as the expected accuracy of the selection method. The extent of prediction bias was analyzed by comparing the regression of EBVs on MBVs. Similarly, a two-trait animal model was applied in the validation for younger animals to estimate genetic correlation between predicted MBVs and the EBVs. The EBV of the younger animals that included their own phenotype was available from the initial analysis that was performed using an animal model. In addition, the accuracy of the PA was computed as the square root of the average reliability. The efficiency of selection on PA versus selection on MBV in younger animals was determined for each trait as the ratio of two accuracies:

$$\text{Efficiency} = \frac{\text{Accuracy of MBV}}{\text{Accuracy of PA}} = \frac{r_{g\,(MBV,EBV)}}{\sqrt{R^2_{PA}}}$$

Results and Discussion

Summary statistics and heritability: Mean performance and heritability of production and efficiency traits of Canadian crossbred beef cattle are summarized in Table 1. The results showed considerable phenotypic and genetic variation for the studied traits typical of beef cattle. The means and standard deviations of traits were in agreement with previous studies (Arthur et al. (2001); Berry and Crowley (2012)). Furthermore, the heritability estimates observed for DMI and growth traits in the validation population were consistent with results from Arthur et al. (2001) and Schenkel et al. (2004) while estimates of heritability for efficiency traits were slightly lower than the values reported by Berry and Crowley (2012). The reliability of EBV predicted for the studied animals were low to moderate for most traits and is largely based on the trait heritability and some animals that are sires themselves with progeny in the data.

Table 1. Number of records used (N), raw mean, standard deviation (SD), heritability (h^2) ± standard error (SE), and average reliability (R^2) of estimated breeding values for production and efficiency traits of Canadian crossbred beef cattle.

Traits[1]	N	Mean	SD	h^2±SE	R^2
ADG	4180	1.40	0.423	0.22±0.04	0.31
DMI	4129	8.99	1.98	0.41±0.04	0.45
MMWT	4145	93.26	13.29	0.68±0.04	0.70
FCR	4134	6.80	2.10	0.09±0.03	0.15
RFI	4123	0.00	0.556	0.21±0.03	0.32
RG	4123	-0.00	0.181	0.08±0.03	0.14
RIG	4123	-0.00	1.714	0.14±0.03	0.24

[1]ADG = Average daily gain in kg/d; DMI = Dry matter intake in kg DM/d; MMWT = Mid-test metabolic weight in kg; FCR = Feed conversion ratio; RFI = Residual feed intake in kg DM/d; RG = Residual BW gain in kg/d; RIG = Residual intake & BW gain

Accuracy of genomic predictions: Table 2 shows the estimates of heritabilities for MBVs and EBVs, respectively, genetic correlations between MBV and EBVs which was used as phenotype in the genome-based model and the coefficient for regressing EBV on MBV. Heritability of breeding values less than one reflect effects of missing pedigree information, sample identification and independent errors as well as possible contamination of MBV by non-additive genetic component while training in crossbreds. Estimates of genetic correlations reflect the accuracy of MBV pooled across replicates. The regression coefficient for all traits is expected to be equal to one where values greater or lower than one reflects an under or over estimation of MBVs. The accuracy of MBVs predicted from field data depends on the accuracy of EBV used as response variable. The EBVs are likely to be influenced by systematic effects such as breed composition, which was not included in the model, and random errors associated with estimating the EBV, however, the level of accuracy observed from this study suggest utility of MBV as a selection tool in genetic

improvement and marker-assisted management programs for crossbred beef cattle.

Table 2. Heritabilities (h^2) ± standard error (SE) of molecular breeding values (MBV) and estimated breeding values (EBV), genetic correlation between MBV and EBV and regression coefficients ($\hat{\beta}$) of EBV on MBV.

Traits[1]	h^2_{MBV}±SE	h^2_{EBV}±SE	$r_{g\,(MBV,EBV)}$±SE	$\hat{\beta}$
ADG	0.92±0.02	0.97±0.03	0.34±0.03	1.32
DMI	0.90±0.02	0.92±0.03	0.48±0.03	1.21
MMWT	0.80±0.02	0.82±0.03	0.68±0.03	1.00
FCR	0.92±0.01	0.99±0.03	0.27±0.03	1.70
RFI	0.94±0.02	0.97±0.03	0.37±0.02	1.30
RG	0.94±0.01	0.99±0.02	0.31±0.02	1.87
RIG	0.94±0.01	0.98±0.03	0.34±0.02	1.44

[1]ADG = Average daily gain in kg/d; DMI = Dry matter intake in kg DM/d; MMWT = Mid-test metabolic weight in kg; FCR = Feed conversion ratio; RFI = Residual feed intake in kg DM/d; RG = Residual BW gain in kg/d; RIG = Residual intake & BW gain

Accuracy of selection in younger animals: The accuracy of selection methods in younger animals without phenotype is given in Table 3. The genetic correlation between MBVs and corresponding EBVs were marginally greater than the accuracy of PA for a few traits. In reality, as practiced in dairy cattle, the MBV can be blended with the PA in order to improve accuracy of selection. However, the results of the present study showed greater accuracy of selection when using MBV to select for growth and feed efficiency than PA. Therefore, genetic selection can be carried out early in younger animals using MBVs in the absence of pedigree information.

Table 3. Genetic correlation between molecular breeding value[1] (MBV) and estimated breeding value[2] (EBV), square root of average reliability of parent average (PA) and efficiency of selection on MBV versus PA.

Trait[3]	$r_{g\,(MBV,EBV)}$	$\sqrt{R^2_{PA}}$	Efficiency
ADG	0.31	0.18	1.72
DMI	0.30	0.22	1.36
MMWT	0.19	0.27	0.70
FCR	0.10	0.13	0.77
RFI	0.26	0.19	1.37
RG	0.05	0.11	0.45
RIG	0.23	0.17	1.35

[1]Molecular breeding value predicted for later born (2011-2012) after training in early born (2002-2010) animals.
[2]Estimated breeding value computed for later born (2011-2012) from initial analysis that included their own phenotypes
[3]ADG = Average daily gain in kg/d; DMI = Dry matter intake in kg DM/d; MMWT = Mid-test metabolic weight in kg; FCR = Feed conversion ratio; RFI = Residual feed intake in kg DM/d; RG = Residual BW gain in kg/d; RIG = Residual intake & BW gain

Conclusion

There are considerable benefits for utility of MBVs in genetic improvement of production and efficiency traits of Canadian crossbred beef cattle. Moderate to high accuracy of MBVs as achieved in this study using a cross-validation approach demonstrates potential of genomic predictions for selection and marker-assisted management in admixed and crossbred populations. Genomic prediction provides an alternative tool for selection in younger animals especially in the absence of pedigree information. Accuracy of MBVs that is greater than 25% is expected for most production traits.

Acknowledgement

The work described here was supported by the following funding sources: Alberta Livestock and Meat Agency Ltd., Alberta Innovates Bio Solutions, Alberta Agriculture and Rural Development, Agriculture and Agri-Food Canada, the Beef Cattle Research Council Beef Cluster, Genome Canada, and Genome Alberta. The following groups also participated in the work: Beefbooster Inc. and Deseret Ranches of Alberta.

Literature Cited

Arthur, P. F., Archer, J. A., Johnson, D. J. et al. (2001). J. Anim. Sci., 79:2805-2811

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

Browning, B. L., and Browning, S. R. (2009). Am. J. Hum. Genet., 84:210-223

Gilmour, A.R., Gogel, B.J., Cullis, B.R. et al. (2009). ASReml User Guide Release 3.0.

Habier, D., Fernando, R. L., Dekkers, J. C. M. (2007). Genetics, 177:2389

Kachman, S. D. Spangler, M. L., Bennett, G. L., et al. (2013). Genet. Sel. Evol., 43:17

Luan, T., Woolliams, J. A., Lien, S. et al. (2009). Genetics, 183:1119-1126.

Meuwissen, T. H. E., Hayes, B. J., Goddard, M. E. (2001). Genetics., 157:1819-1829

Sargolzaei, M., Schenkel, F. S., VanRaden, P. M. (2009). Technical report to the Dairy Cattle Breeding and Genetics Committee. University of Guelph, October 7

Schenkel, F. S., Miller, S. P., and Wilton, J. W. (2004). Can. J. Anim. Sci., 84:177-185

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