Genomic evaluation for two-way crossbred performance in cattle

Q. Mei[1,2*], H. Liu[3], T. Xiang[1] and O. F. Christensen[2]

[1]Huazhong Agricultural University, No. 1 Shizishan Avenue, Hongshan District, 430070 Wuhan, P.R. China; [2]Aarhus University, Blichers Alle' 20, 8830 Tjele, Denmark; [3]SEGES, Agro Food Park 15, 8200 Aarhus N, Denmark; [*]quanshun@qgg.au.dk

Abstract

Dairy cattle production systems are mostly based on using purebreds, but recently crossbreeding has received increased interest. For genetic evaluation including crossbreds, ideally, models accounting for breed specific effects should perform better compared with simple models. A novel single-step best linear unbiased prediction method provides an opportunity to combine purebred and crossbred information when breed specific effects are accounted for. In this study, we applied this single-step method to evaluate two-way crossbred performance for average daily gain in crossbred beef on dairy. Results showed that the single-step method, which accounted for breed specific effects, had a higher predictive ability. We conclude that this method is useful in the genetic evaluation for crossbred performance in cattle production system.

Introduction

Crossbreeding is common practice in many livestock production systems, in particular for pigs and poultry. For dairy and beef cattle, the production systems are mostly based on using purebreds, but recently crossbreeding between dairy breed cows and beef breed bulls has received increased interest for a number of reasons (Berry 2021). In particular, the use of x-sorted dairy semen and herd specific strategies for the necessary number of replacement heifers has provided the option that some dairy cows in a herd can be inseminated with beef semen, and furthermore, meat production from crosses between dairy cows and beef bulls has lower environmental footprint than meat production from beef cattle.

Genomic selection (Meuwissen et al. 2001) is another potentially important factor to make the cattle meat production more efficient and with reduced environmental impact. Genomic evaluation in cattle has primarily been done within pure breeds, but ideally for crossbreeding with dairy cows, genomic breeding values on beef bulls should be estimated for such crossbred performance based on phenotypic records on crossbreds. Genomic models that split genomic effect of crossbreds into breed specific terms and account for the breed of origin of alleles (BOA) should in principle perform better than simpler models that do not take BOA into account, but this has not always been seen (Sevillano et al. 2017).

Commonly in genomic evaluation, for many reasons, not all relevant individuals are genotyped. For this purpose, so-called single-step genomic best linear unbiased prediction (ssGBLUP) was introduced (Legarra et al. 2014). Xiang et al. (2016) investigated an extension of ssGBLUP for crossbred performance in a terminal two-breed crossbreeding system for pigs, and concluded that models using marker genotypes of crossbreds improved accuracies of prediction for crossbred performances (Xiang et al. 2016). The approach investigated in that paper was based on BOA and so-called breed specific partial relationship matrices that combine genomic and pedigree-based relationships (Christensen et al. 2014).

The objective of this paper is to investigate predictive performance of a ssGBLUP model based on BOA compared to the usual ssGBLUP model that did not account for BOA for crossbred beef on dairy cattle.

Materials & Methods

Data

All data sets were provided by SEGES cattle. In this study, 2,329 two-way crossbred calves with purebred Belgian Blue beef (BBL) sires and purebred Holstein dairy (HOL) dams were on test for 1 month. These calves had an average age of 210 ± 36 days in the beginning of test, and 245 ± 32 days in the end of test. During this period, feed intake was recorded for each animal, and body weight of each animal was recorded at both start and end of test period. Average daily gain (ADG) of each animal within this period was calculated as the increase in body weight divided by number of days between the two recordings. After data editing, 1,749 crossbred calves with ADG were retained.

Pedigree for crossbred animals was traced back three generations, and there were 13,612 animals in the pedigree. Among these animals, 36 BBL and 138 HOL were genotyped with either EuroG 10k Bead Chip or EuroG MD Bead Chip, and 1,137 crossbreds were genotyped with EuroG MD Bead Chip. In purebred animals, imputation from EuroG 10k Bead Chip to EuroG MD Bead Chip was done with software Beagle 5.2 (Browning and Browning 2007). Quality controls of genomic data with software Plink were as follows: it was first confirmed that no individuals had call-rate less than 90%; following that, SNPs with call-rate less than 90% were removed; SNPs with a minor allele frequency less than 0.05 were removed; SNPs deviating strongly from Hardy Weinberg equilibrium within breed ($p < 10\text{-}7$) were also removed (Purcell et al. 2007). After quality control, 1,311 individuals and 45,288 SNPs were retained. The retained genotype data was phased with software Beagle 5.2 (Browning and Browning 2007).

Allele tracing

The detection of BOA in crossbred animals is necessary for constructing breed specific partial relationship matrix. In this study, for detecting BOA we used the recently developed AllOr algorithm (Eiríksson et al. 2021), which is based on the comparison of haplotypes in overlapping windows.

Statistical Model

Two ssGBLUP models for genomic prediction were evaluated. The first model is the usual ssGBLUP model (M_1):

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e} \quad (1)$$

where $\mathbf{y}$ is a vector of phenotypic records for ADG; $\mathbf{Xb}$ refers to the fixed effects including herd-year-season, sex, and covariate of body weight at start test; $\mathbf{u}$ is a vector of random additive genetic effects for each animal, and $\mathbf{Z}$ is the corresponding incidence matrix; $\mathbf{e}$ is a vector of random residual effects. The random effects follow normal distributions of $\mathbf{u} \sim N(\mathbf{0}, \mathbf{H}\sigma_a^2)$, and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$, where $\mathbf{H}$ is the combined pedigree-based and marker-based relationship matrix, which can be constructed as previously reported (Legarra et al. 2014); $\mathbf{I}$ is the identity matrix; σ_a^2 is the additive genetic variance; σ_e^2 is the residual variance.

The second model is the partial ssGBLUP model (M_2):

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z_A u_A} + \mathbf{Z_B u_B} + \mathbf{e} \quad (2)$$

where $\mathbf{y}$, $\mathbf{Xb}$, and $\mathbf{e}$ are as in Equation (1); $\mathbf{u_A}$ is a vector of random additive genetic effects from BBL, $\mathbf{u_B}$ is a vector of random additive genetic effects from HOL; $\mathbf{Z_A}$ and $\mathbf{Z_B}$ are the

corresponding incidence matrices. The random effects follow normal distributions of $\mathbf{u_A} \sim N(\mathbf{0}, \mathbf{H_A}\sigma_A^2)$, and $\mathbf{u_B} \sim N(\mathbf{0}, \mathbf{H_B}\sigma_B^2)$, where $\mathbf{H_A}$ and $\mathbf{H_B}$ are combined pedigree-based and marker-based breed specific partial relationship matrices, which can be constructed as previously reported (Christensen et al. 2014); σ_A^2 and σ_B^2 were additive genetic variances for crossbred performances of BBL and HOL, respectively. The above analyses were carried out using restricted maximum likelihood (REML) algorithm and best linear unbiased prediction (BLUP) in the software DMU (Madsen and Jensen 2013).

In order to reveal the differences in predictive abilities between models M_1 and M_2, data were split into a training dataset and a validation dataset by a cut-off date of October 20, 2020. The numbers of animals in the training and validation datasets were 1,399 and 350, respectively. Predictive abilities were calculated as the correlation between predicted total genetic values ($\mathbf{g}$) and the corrected phenotypes ($\mathbf{Y_c}$) in the validation dataset. Vector $\mathbf{Y_c}$ was calculated based on model M_1 with full dataset. In model M_1, $\mathbf{g}$ was equal to the additive genetic effects ($\mathbf{g} = \mathbf{u}$); in model M2, $\mathbf{g}$ was equal to the sum of additive genetic effects from BBL and HOL ($\mathbf{g} = \mathbf{u_A} + \mathbf{u_B}$). Furthermore, dispersion bias of genomic predictions for each model was assessed as the regression coefficient of $\mathbf{Y_c}$ on $\mathbf{g}$, with an expected result of 1.

Results

Table 1. Estimates of variance components and heritabilities in different models.

Model	σ_a^2	σ_A^2	σ_B^2	σ_e^2	h_a^2	h_{AB}^{2*}
M_1	0.010			0.088	0.105	
M_2		0.011	0.013	0.086		0.122

*: h_{AB}^2=heritabilities of crossbred animals $(0.5(\sigma_A^2 + \sigma_B^2)/[0.5(\sigma_A^2 + \sigma_B^2) + \sigma_e^2])$.

Table 2. Predictive abilities for crossbred animals in the validation dataset in different models

Model	$r(\mathbf{Y_c}, \mathbf{g})$ [1]	$b(\mathbf{Y_c}, \mathbf{g})$ [2]
M_1	0.052	0.733
M_2	0.059	0.774

1: $r(\mathbf{Y_c}, \mathbf{g})$ = Pearson correlation coefficients between $\mathbf{Y_c}$ and $\mathbf{g}$.
2: $b(\mathbf{Y_c}, \mathbf{g})$ = Regression coefficients of $\mathbf{Y_c}$ on $\mathbf{g}$.

Variance component estimation

Estimates of variance components and heritabilities are shown in Table 1. In model M_2, the additive genetic variances from BBL (σ_A^2) and HOL (σ_B^2) were equal to 0.011 and 0.013, and the additive genetic variance of crossbred animals was equal to 0.012 (half of the sum of σ_A^2 and σ_B^2), which was a little bit higher than in model M_1. The heritability in models M_1 and M_2 was 0.105 and 0.122, respectively.

Predictive abilities

Predictive abilities of models M_1 and M_2 are shown in Table 2. The Pearson correlation coefficients between the corrected phenotypes and total genetic values of crossbred animals in model M_2 was 0.059, being a little bit higher than in model M_1 where it was 0.052. In terms of regression coefficients of corrected phenotypes on the total genetic values, model M_1 deviated more from 1 compared with model M_2.

Discussion

In this study, the estimated heritability of ADG was low to moderate, which was a little bit lower than previously reported (Freetly et al. 2020). One potential reason could be that ADG in this study was calculated in a shorter time period compared to previous studies. Furthermore, Ahlberg et al. (2018) pointed out that during different periods, the correlations for each shortened test duration was different. In addition, the smaller dataset could also be another possible reason.

Results of predictive abilities indicated that partial ssGBLUP model accounting for breed-specific effects performs better than the usual ssGBLUP model in terms of accuracy and dispersion bias. This phenomenon has also been observed in other species (Xiang et al. 2016). The result presented here only focused on the prediction of total genetic values in crossbreds. In future study, we plan to investigate accuracy of estimated breeding value of BBL sires for crossbred performance when mated to a Holstein cow.

References

Ahlberg, C. M., K. Allwardt, A. Broocks, K. Bruno, L. McPhillips, et al. (2018) Journal of Animal Science 96(8):3043-3054. https://doi.org/10.1093/jas/sky209

Berry, D. (2021) Journal of dairy science 104(4):3789-3819. https://doi.org/10.3168/jds.2020-19519

Browning, S. R. and B. L. Browning (2007) The American Journal of Human Genetics 81(5):1084-1097. https://doi.org/10.1086/521987

Christensen, O. F., P. Madsen, B. Nielsen and G. Su (2014) Genetics Selection Evolution 46(1):1-9. https://doi.org/10.1186/1297-9686-46-23

Eiríksson, J. H., E. Karaman, G. Su and O. F. Christensen (2021) Genetics Selection Evolution 53(1):1-13. https://doi.org/10.1186/s12711-021-00678-3

Freetly, H. C., L. A. Kuehn, R. M. Thallman and W. M. Snelling (2020) Journal of Animal Science 98(1):skz394. https://doi.org/10.1093/jas/skz394

Legarra, A., O. F. Christensen, I. Aguilar and I. Misztal (2014) Livestock Science 166:54-65. https://doi.org/10.1016/j.livsci.2014.04.029

Madsen, P. and J. Jensen (2013) Center for Quantitative Genetics and Genomics, Dept of Molecular Biology and Genetics, University of Aarhus, Research Centre Foulum, Box 50:8830.

Meuwissen, T. H. E., B. J. Hayes and M. E. Goddard (2001) Genetics 157(4):1819-1829. https://doi.org/10.1093/genetics/157.4.1819

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, et al. (2007) Am J Hum Genet 81(3):559-575. https://doi.org/10.1086/519795

Sevillano, C. A., J. Vandenplas, J. W. M. Bastiaansen, R. Bergsma and M. P. L. Calus (2017) Genet Sel Evol 49(1):75. https://doi.org/10.1186/s12711-017-0350-1

Xiang, T., B. Nielsen, G. Su, A. Legarra and O. Christensen (2016) Journal of Animal Science 94(3):936-948. https://doi.org/10.2527/jas.2015-9930