

RESEARCH ARTICLE

Open Access



Genomic evaluation for two-way crossbred performance in cattle

Quanshun Mei^{1,2}, Huiming Liu³, Shuhong Zhao¹, Tao Xiang^{1*} and Ole F Christensen^{2*} 

Abstract

Background Dairy cattle production systems are mostly based on purebreds, but recently the use of crossbreeding has received increased interest. For genetic evaluations including crossbreds, several methods based on single-step genomic best linear unbiased prediction (ssGBLUP) have been proposed, including metafounder ssGBLUP (MF-ssGBLUP) and breed-specific ssGBLUP (BS-ssGBLUP). Ideally, models that account for breed effects should perform better than simple models, but knowledge on the performance of these methods is lacking for two-way crossbred cattle. In addition, the differences in the estimates of genetic parameters (such as the genetic variance component and heritability) between these methods have rarely been investigated. Therefore, the aims of this study were to (1) compare the estimates of genetic parameters for average daily gain (ADG) and feed conversion ratio (FCR) between these methods; and (2) evaluate the impact of these methods on the predictive ability for crossbred performance.

Methods Bivariate models using standard ssGBLUP, MF-ssGBLUP and BS-ssGBLUP for the genetic evaluation of ADG and FCR were investigated. To measure the predictive ability of these three methods, we estimated four estimators, bias, dispersion, population accuracy and ratio of population accuracies, using the linear regression (LR) method.

Results The results show that, for both ADG and FCR, the heritabilities were low with the three methods. For FCR, the differences in the estimated genetic parameters were small between the three methods, while for ADG, those estimated with BS-ssGBLUP deviated largely from those estimated with the other two methods. Bias and dispersion were similar across the three methods. Population accuracies for both ADG and FCR were always higher with MF-ssGBLUP than with ssGBLUP, while with BS-ssGBLUP the population accuracy was highest for FCR and lowest for ADG.

Conclusions Our results indicate that in the genetic evaluation for crossbred performance in a two-way crossbred cattle production system, the predictive ability of MF-ssGBLUP and BS-ssGBLUP is greater than that of ssGBLUP, when the estimated variance components are consistent across the three methods. Compared with BS-ssGBLUP, MF-ssGBLUP is more robust in its superiority over ssGBLUP.

Background

Crossbreeding is commonly used in many livestock production systems [1], especially for pig and poultry. For dairy and beef cattle, production systems are mostly based on purebreds, but recently the use of crossbreeding between dairy breed cows and beef breed bulls has received increased interest for a number of reasons [1]. In particular, meat production from crossbreds between dairy cows and beef bulls has a lower environmental footprint than that from beef cattle [2]. Furthermore, because the improved reproductive performance of dairy cows

*Correspondence:

Tao Xiang

tao.xiang@mail.hzau.edu.cn

Ole F Christensen

olef.christensen@qgg.au.dk

¹ Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education & Key Laboratory of Swine Genetics and Breeding of Ministry of Agriculture, Huazhong Agricultural University, Wuhan 430070, China

² Center for Quantitative Genetics and Genomics, Aarhus University, C. F. Møllers Allé 3, 8000 Aarhus C, Denmark

³ SEGES Cattle, Agrofood Park 15, 8200 Aarhus N, Denmark



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

reduces the need for replacement heifers, some dairy cows in a herd can be inseminated with beef semen.

For livestock production systems that use crossbreds, although the breeding goal is to improve crossbred performance, selection usually takes place in the purebred lines [3], which is sub-optimal since the genetic performances of purebred (PB) and crossbred (CB) animals differ [4–6]. By reviewing the existing literature on the genetic correlations between the performances of purebred and crossbred pigs, Wientjes and Calus [7] found an average genetic correlation of 0.6, while based on the review of 14 studies on broilers and layers, Calus et al. [8] found an average genetic correlation of 0.71. These results indicate that it is meaningful to select for CB performance as well as for PB performance in crossbred production systems.

Since 2010, single-step genomic best linear unbiased prediction (ssGBLUP) has been used as a standard genomic selection (GS) method in the pig industry, and has shown a high predictive ability for both genotyped and non-genotyped animals [9–11]. However, when crossbred information is considered, ssGBLUP does not fit well due to the existence of genetic differences between breeds (i.e., allele frequency, linkage disequilibrium, and gametic phase) [12]. An alternative ssGBLUP, called breed-specific ssGBLUP (BS-ssGBLUP) that integrates purebred and crossbred information, was proposed by Christensen et al. [13] based on multiple breed-specific relationship matrices [14]. Xiang et al. [5] applied this method on real pig data and validated its superiority over ssGBLUP. However, other studies have not confirmed this superiority [15, 16]. Another method called meta-founder ssGBLUP (MF-ssGBLUP) has been developed by Legarra et al. [17] to model genetic differences between breeds. The differences in the estimates of genetic parameters and predictive ability between these three methods have been investigated using pig data [5, 18] and simulated data [19], but not with data on crossbreds between dairy and beef cattle, thus more research is needed. In addition, to date, ssGBLUP and MF-ssGBLUP have been successfully used in crossbred cattle to estimate genetic parameters [20, 21], but not BS-ssGBLUP.

Thus, the objectives of this study were to (1) compare the estimates of genetic parameters in crossbred beef and

dairy cattle for average daily gain (ADG) and feed conversion ratio (FCR) with ssGBLUP, MF-ssGBLUP and BS-ssGBLUP; and (2) evaluate the impact of these methods on the predictive ability for crossbred performance.

Methods

Data

All datasets were provided by SEGES Innovation Cattle and Nordic Cattle Genetic Evaluation. In this study, 4089 two-way crossbred calves (BH) with purebred Belgian Blue beef (BBL) sires and purebred Holstein dairy (HOL) dams were on test for about 1 month. During this period, feed intake was recorded for each animal, and body weight of each animal was recorded at both the start and end of the test period. ADG (kg/d) and FCR (kg/kg) of each animal within this period were calculated as the increase in body weight divided by number of days and the average daily feed intake divided by average daily gain, respectively. After data editing of feed intake and body weight records (see Additional file 1: Fig. S1), 2592 crossbred calves were retained, with ADG available for all the calves and FCR for 2306 calves. The birth dates of these calves ranged from June 1 2019 to December 1 2021. These 2592 crossbred animals originate from 67 sires and 2419 dams, with an average number of progeny per sire of 38.7 and per dam of 1.1, and the average size of paternal half-sibling families was 37.6. The average age of these calves was 207 days (standard deviation (SD)=34 days) at the beginning of the test, and 243 days (SD=33 days) at the end of the test. Descriptive statistics of the phenotypes are in Table 1.

Pedigree for the crossbred animals was traced back three generations, and included 30,643 animals with 846 BBL, 25,709 HOL and 4088 BH. Among these animals, 43 BBL and 882 HOL were genotyped with the EuroG 10K Bead chip, and 39 BBL, 1590 HOL, and 1780 BH were genotyped with the Eurogenomics 75K custom SNP chip. Among the parents of the BH, 52 BBL and 319 HOL were genotyped. For all genotyped animals, the procedure of filling-in missing genotypes and imputation from the EuroG 10K Bead chip to the Eurogenomics 75K custom SNP chip was done with the Beagle 5.2 software [22]. Quality control of the genomic data was done using the Plink software as follows [23]: first, we checked

Table 1 Descriptive statistics

Trait	Mean (SD)	Min	Max	Number of animals with phenotype	Number of animals with both phenotype and genotype	Number of animals in partial dataset (cut-off date, April 1 2021)	Number of animals in partial dataset (cut-off date, May 1 2021)
ADG (kg)	1.411 (0.345)	0.077	2.857	2592	1258	1944	2016
FCR	4.984 (1.751)	0.436	19.663	2306	1115	1786	1855

ADG: average daily gain; FCR: feed conversion ratio

that no individuals had a call-rate lower than 90%; then, SNPs with a call-rate lower than 90%, SNPs with a minor allele frequency lower than 0.01, and SNPs that deviated strongly from the Hardy–Weinberg equilibrium within breed ($p < 10^{-7}$) were removed. Finally, 4329 animals (81 BBL and 2468 HOL, and 1780 BH) and 48,777 SNPs were retained after quality control and imputation. The retained genotype data were phased with the Beagle 5.2 software [22].

Statistical models

A bivariate animal model was used to estimate genetic parameters and breeding values for ADG and FCR. To construct the single-step relationship matrices, three methods, standard ssGBLUP, MF-ssGBLUP, and BS-ssGBLUP, were incorporated in the bivariate model, as follows.

Standard ssGBLUP

With the aim of extending the marker-based relationship matrices to the non-genotyped animals, Legarra et al. [9] and Christensen and Lund [10] developed ssGBLUP.

The statistical bivariate model for ssGBLUP is:

$$y = Xb + Zu + e, \tag{1}$$

where y is the vector of phenotypic records for ADG and FCR in crossbred calves; b is the vector of fixed effects including the effects of sex, pen (during the experiment), herd-year-season (year and season of the testing period), and covariate of the weight at the start of the test for ADG and FCR; u is the vector of random additive genetic effects for ADG and FCR; e is the vector of the random residual error for ADG and FCR; X and Z are the corresponding incidence matrices.

It is assumed that the random effects follow normal distributions, i.e. $u \sim N(0, \sum_u \otimes H)$ and $e \sim N(0, \sum_e \otimes I)$, where H is the combined pedigree-based and marker-based relationship matrix presented below; I is the corresponding identity matrix; \sum_u is the genetic (co)variance matrix, \sum_e is the residual (co)variance matrix, and \otimes denotes the Kronecker product. The (co)variance matrices are as follows:

$$\sum_u = \begin{bmatrix} \sigma_{u_{ADG}}^2 & \sigma_{u_{ADG}u_{FCR}} \\ \text{sym} & \sigma_{u_{FCR}}^2 \end{bmatrix},$$

$$\text{and } \sum_e = \begin{bmatrix} \sigma_{e_{ADG}}^2 & \sigma_{e_{ADG}e_{FCR}} \\ \text{sym} & \sigma_{e_{FCR}}^2 \end{bmatrix},$$

where $\sigma_{u_{ADG}}^2$ is the additive genetic variance of ADG, $\sigma_{u_{FCR}}^2$ is the additive genetic variance of FCR, $\sigma_{u_{ADG}u_{FCR}}$ is the additive genetic covariance between ADG and FCR; $\sigma_{e_{ADG}}^2$ is the residual variance of ADG, $\sigma_{e_{FCR}}^2$ is the residual variance of FCR, and $\sigma_{e_{ADG}e_{FCR}}$ is the residual covariance between ADG and FCR.

The combined pedigree-based and marker-based relationship matrix H is defined as [9, 10]:

$$H = \begin{bmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} + A_{12}A_{22}^{-1}GA_{22}^{-1}A_{21} & A_{12}A_{22}^{-1}G \\ GA_{22}^{-1}A_{21} & (1 - \omega)G + \omega A_{22} \end{bmatrix},$$

where A is the pedigree relationship matrix, G is the genomic realized relationship matrix; subscripts 1 and 2 stand for non-genotyped and genotyped animals, respectively; ω is interpreted as the relative weight on the polygenic effect, which is set as 0.05 in this study as commonly done [24, 25].

The relationship matrix G was constructed as [26]:

$$G = \frac{ZZ'}{\sum_{i=1}^m 2p_iq_i},$$

where m is the number of SNPs, p_i is the frequency of allele A at marker i and $q_i = 1 - p_i$; Z is the incidence matrix with elements of $2 - 2p_i$, $1 - 2p_i$ and $-2p_i$ for AA , Aa , and aa , respectively. Matrix G was adjusted to be compatible with matrix A as described by Christensen et al. [27].

MF-ssGBLUP

To account for allele frequency in the base population and compatibility between the pedigree and genomic additive relationship matrices, Legarra et al. [17] developed a new method named MF-ssGBLUP, based on developments described in Christensen [27].

The statistical bivariate model for MF-ssGBLUP is:

$$y = Xb + Zu + e, \tag{2}$$

where y , b , u , e , X and Z are as defined above.

The difference between Eq. (1) and Eq. (2) is the definition of the additive genetic relationship matrix. For Eq. (2), it is assumed that the random effects follow normal distributions, i.e. $u \sim N(0, \sum_u \otimes H_{MF})$, where H_{MF} is the combined pedigree-based and marker-based meta-founder relationship matrix, and \sum_u contains the genetic variance and covariance parameters.

The matrix H_{MF} is defined as:

$$\mathbf{H}_{MF} = \begin{bmatrix} \mathbf{A}_{11}^{\Gamma} - \mathbf{A}_{12}^{\Gamma} \mathbf{A}_{22}^{\Gamma^{-1}} \mathbf{A}_{21}^{\Gamma} + \mathbf{A}_{12}^{\Gamma} \mathbf{A}_{22}^{\Gamma^{-1}} \mathbf{G}^{0.5} \mathbf{A}_{22}^{\Gamma^{-1}} \mathbf{A}_{21}^{\Gamma} & \mathbf{A}_{12}^{\Gamma} \mathbf{A}_{22}^{\Gamma^{-1}} \mathbf{G}^{0.5} \\ \mathbf{G}^{0.5} \mathbf{A}_{22}^{\Gamma^{-1}} \mathbf{A}_{21}^{\Gamma} & (1 - \omega) \mathbf{G}^{0.5} + \omega \mathbf{A}_{22}^{\Gamma^{-1}} \end{bmatrix},$$

where ω is as defined above; \mathbf{A}^{Γ} is the pedigree relationship matrix with metafounders, $\mathbf{G}^{0.5}$ is the genomic realized relationship matrix with allele frequencies equal to 0.5; subscripts 1 and 2 stand for non-genotyped and genotyped animals, respectively.

The construction of the relationship matrix \mathbf{A}^{Γ} is based on the estimated metafounder relationship matrix, $\mathbf{\Gamma}$ [17], which represents the within- and across-population relationship matrix and is expressed as follows:

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_B & \gamma_{B,H} \\ \text{sym} & \gamma_H \end{bmatrix},$$

where γ_B is the metafounder relationship for BBL; γ_H is the metafounder relationship for HOL; $\gamma_{B,H}$ is the across-metafounder relationship between the BBL and HOL populations. The generalized least squares method was used to estimate $\mathbf{\Gamma}$ as described by Garcia-Baccino et al. [28].

The relationship matrix $\mathbf{G}^{0.5}$ is constructed as:

$$\mathbf{G}^{0.5} = \frac{\mathbf{Z}\mathbf{Z}'}{s},$$

where \mathbf{Z} is the incidence matrix with elements of 1, 0 and -1 for AA, Aa, and aa, respectively; $s = m/2$.

The genetic variance and covariance parameters from MF-ssGBLUP were estimated under the assumption that founders are related, while in other models usually unrelated founders are assumed for the genetic variance. To be comparable with estimates from other models that estimate genetic variance for unrelated founders, such as standard ssGBLUP and BS-ssGBLUP in our study, we multiplied the estimates of the genetic parameters estimated with MF-ssGBLUP by $1 + \frac{\text{diag}(\mathbf{\Gamma})}{2} - \bar{\mathbf{\Gamma}}$, following the suggestion of Legarra et al. [17].

BS-ssGBLUP

BS-ssGBLUP assumes that the substitution effects of breed-specific alleles differ between breeds. This method was developed by Christensen et al. [13] based on previous studies [14, 29].

The statistical bivariate model for BS-ssGBLUP is:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_B \mathbf{u}_B + \mathbf{Z}_H \mathbf{u}_H + \mathbf{e}, \tag{3}$$

where \mathbf{y} , \mathbf{b} , \mathbf{e} , and \mathbf{X} are as defined above; \mathbf{u}_B is the vector of random additive genetic effects from BBL for ADG and FCR, \mathbf{u}_H is the vector of random additive genetic effects from HOL for ADG and FCR; \mathbf{Z}_B and \mathbf{Z}_H are the corresponding incidence matrices.

It is assumed that the random effects follow normal distributions, i.e. $\mathbf{u}_B \sim N(\mathbf{0}, \sum_{\mathbf{u}_B} \otimes \mathbf{H}_B)$ and $\mathbf{u}_H \sim N(\mathbf{0}, \sum_{\mathbf{u}_H} \otimes \mathbf{H}_H)$, where \mathbf{H}_B and \mathbf{H}_H are combined pedigree-based and marker-based breed specific partial relationship matrices for BBL and HOL; $\sum_{\mathbf{u}_B}$ is the genetic (co)variance matrix for BBL, $\sum_{\mathbf{u}_H}$ is the genetic (co)variance matrix for HOL. The (co)variance matrices are as follows:

$$\sum_{\mathbf{u}_B} = \begin{bmatrix} \sigma_{\mathbf{u}_{B\text{ADG}}}^2 & \sigma_{\mathbf{u}_{B\text{ADG}}\mathbf{u}_{B\text{FCR}}} \\ \text{sym} & \sigma_{\mathbf{u}_{B\text{FCR}}}^2 \end{bmatrix},$$

$$\text{and } \sum_{\mathbf{u}_H} = \begin{bmatrix} \sigma_{\mathbf{u}_{H\text{ADG}}}^2 & \sigma_{\mathbf{u}_{H\text{ADG}}\mathbf{u}_{H\text{FCR}}} \\ \text{sym} & \sigma_{\mathbf{u}_{H\text{FCR}}}^2 \end{bmatrix},$$

where $\sigma_{\mathbf{u}_{B\text{ADG}}}^2$ is the additive genetic variance of ADG from BBL, $\sigma_{\mathbf{u}_{B\text{FCR}}}^2$ is the additive genetic variance of FCR from BBL, $\sigma_{\mathbf{u}_{B\text{ADG}}\mathbf{u}_{B\text{FCR}}}$ is the additive genetic covariance between ADG and FCR from BBL; $\sigma_{\mathbf{u}_{H\text{ADG}}}^2$ is the additive genetic variance of ADG from HOL, $\sigma_{\mathbf{u}_{H\text{FCR}}}^2$ is the additive genetic variance of FCR from breed HOL, $\sigma_{\mathbf{u}_{H\text{ADG}}\mathbf{u}_{H\text{FCR}}}$ is the additive genetic covariance between ADG and FCR from HOL.

The breed-specific matrix \mathbf{H}_B is defined as:

$$\mathbf{H}_B = \begin{bmatrix} \mathbf{A}_{11}^{(B)} - \mathbf{A}_{12}^{(B)} \mathbf{A}_{22}^{(B)^{-1}} \mathbf{A}_{21}^{(B)} + \mathbf{A}_{12}^{(B)} \mathbf{A}_{22}^{(B)^{-1}} \mathbf{G}^{(B)} \mathbf{A}_{22}^{(B)^{-1}} \mathbf{A}_{21}^{(B)} & \mathbf{A}_{12}^{(B)} \mathbf{A}_{22}^{(B)^{-1}} \mathbf{G}^{(B)} \\ \mathbf{G}^{(B)} \mathbf{A}_{22}^{(B)^{-1}} \mathbf{A}_{21}^{(B)} & (1 - \omega) \mathbf{G}^{(B)} + \omega \mathbf{A}_{22}^{(B)^{-1}} \end{bmatrix},$$

where ω is as defined above; $\mathbf{A}^{(B)}$ is the breed-specific pedigree relationship matrix from BBL, $\mathbf{G}^{(B)}$ is the breed-specific genomic realized relationship matrix from BBL; subscripts 1 and 2 stand for non-genotyped and genotyped animals, respectively.

The breed-specific pedigree relationship matrix $\mathbf{A}^{(B)}$ was previously described by García-Cortés and Toro [14]. Matrix $\mathbf{G}^{(B)}$ is split into submatrices with indices denoting genotyped BBL and crossbred animals as follows:

$$\mathbf{G}^{(B)} = \begin{bmatrix} \mathbf{G}_{B,B}^{(B)} & \mathbf{G}_{B,BH}^{(B)} \\ \text{sym} & \mathbf{G}_{BH,BH}^{(B)} \end{bmatrix},$$

with these submatrices being defined as:

$$\mathbf{G}_{B,B}^{(B)} = \frac{(\mathbf{M}_B - 2\mathbf{p}_B\mathbf{1})(\mathbf{M}_B - 2\mathbf{p}_B\mathbf{1})'}{2\mathbf{p}_B'(\mathbf{1} - \mathbf{p}_B)},$$

$$\mathbf{G}_{B,BH}^{(B)} = \frac{(\mathbf{M}_B - 2\mathbf{p}_B\mathbf{1})(\mathbf{Q}_B - \mathbf{p}_B\mathbf{1})'}{2\mathbf{p}_B'(\mathbf{1} - \mathbf{p}_B)},$$

$$\text{and } \mathbf{G}_{BH,BH}^{(B)} = \frac{(\mathbf{Q}_B - \mathbf{p}_B\mathbf{1})(\mathbf{Q}_B - \mathbf{p}_B\mathbf{1})'}{2\mathbf{p}_B'(\mathbf{1} - \mathbf{p}_B)},$$

where \mathbf{M}_B and \mathbf{Q}_B contain the breed BBL specific allele contents of the reference allele for BBL (coded as 0, 1, or 2) and for BH (coded as 0 or 1), respectively, for which tracing of the breed of origin of alleles is required; $\mathbf{1}$ is a vector of 1s; and \mathbf{p}_B is the vector of BBL specific allele frequencies. Finally, matrix $\mathbf{G}^{(B)}$ is adjusted to be compatible with matrix $\mathbf{A}^{(B)}$, as described by Christensen et al. [13]. The definition of the breed-specific matrix \mathbf{H}_H is similar to the definition of the \mathbf{H}_B matrix.

Tracing the breed of origin of alleles in F1 crosses is expected to be very accurate [30], and was conducted on the phased genotypes, separately, for each chromosome per individual. Among the 1780 genotyped crossbred animals, 1447 crossbred animals had 47 genotyped sires, whereas for the 333 remaining crossbred animals, none of the parents were genotyped. When the sire (or dam) was genotyped, four comparisons between crossbred and purebred phased alleles were made. For each comparison, when a crossbred allele differed from the corresponding purebred allele, it was counted as a difference. The chromosome with the smallest number of differences was assigned to the breed of the parent. When neither of the parents was genotyped, for each non-overlapping sliding window of 50 consecutive SNPs, comparisons between the two crossbred segments of phased alleles and segments of phased alleles in the reference panel were made for each breed. For

each of the two crossbred segments, the number of copies was counted for each breed, and the segment was considered to originate from the breed with the largest number of copies. Finally, each crossbred chromosome was assigned to the breed from which the majority of its segments originated. This procedure is the same as in Xiang et al. [5].

For tracing of alleles and the construction of the breed-specific matrices \mathbf{H}_B and \mathbf{H}_H in BS-ssGBLUP, we developed an R package named cBar2, which has been uploaded on github (<https://github.com/TXiang-lab/cBar2>).

Estimation of genetic parameters in the above bivariate models with the three methods was carried out using the restricted maximum likelihood (REML) algorithm in the software DMU [31] via the wrapper of R package blupADC [32].

The heritability and genetic correlation estimates and their standard errors in ssGBLUP and MF-ssGBLUP were calculated as described by Falconer [33] and Mrode [34]. For BS-ssGBLUP, the heritability estimates for ADG and FCR were calculated as $h_{ADG}^2 = \frac{\sigma_{u_{BHADG}}^2}{\sigma_{u_{BHADG}}^2 + \sigma_{e_{ADG}}^2}$ and $h_{FCR}^2 = \frac{\sigma_{u_{BHFCR}}^2}{\sigma_{u_{BHFCR}}^2 + \sigma_{e_{FCR}}^2}$, where $\sigma_{u_{BHADG}}^2$ and $\sigma_{u_{BHFCR}}^2$ are defined as $0.5(\sigma_{u_{BADG}}^2 + \sigma_{u_{HADG}}^2)$ and $0.5(\sigma_{u_{BFCR}}^2 + \sigma_{u_{HFCR}}^2)$. The standard errors of the heritabilities, $\sigma_{(h_{ADG}^2)}$ and $\sigma_{(h_{FCR}^2)}$, were obtained by the deltaMethod implemented in the R package msm [35]. The genetic correlation between ADG and FCR was calculated as $r = \frac{0.5(\sigma_{u_{BADG}u_{BFCR}} + \sigma_{u_{HADG}u_{HFCR}})}{\sqrt{\sigma_{u_{BHADG}}^2 * \sigma_{u_{BHFCR}}^2}}$, and its standard error was also obtained by the deltaMethod.

Model-based reliability

For ssGBLUP, the model-based reliability was calculated as follows:

$$\text{Rel}_i = \mathbf{1} - \frac{\text{PEV}_{i,i}}{\mathbf{H}_{i,i}\sigma_u^2},$$

where \mathbf{H} is as defined previously; Rel_i is the reliability of the individual i , σ_u^2 is the additive genetic variance estimated with BS-ssGBLUP, PEV is the prediction error (co) variance matrix, which can be obtained by inverting the coefficient matrix of Henderson's mixed model equations corresponding to the model used [34].

Both MF-ssGBLUP and BS-ssGBLUP can model the genetic difference between breeds, and the individual model-based reliability can be calculated within each breed [13, 36]. For MF-ssGBLUP, the model-based reliability was calculated as described in Bermann et al. [36].

For individual i , the reliability within each metafounder is calculated as:

$$Rel_i^{mf} = 1 - \frac{PEV(\mathbf{u}_i) + PEV(\mathbf{u}_{mf}) - 2PEV(\mathbf{u}_i, \mathbf{u}_{mf})}{\left(\mathbf{H}_{ii}^{(\Gamma)} + \mathbf{H}_{mf,mf}^{(\Gamma)} - 2\mathbf{H}_{i,mf}^{(\Gamma)}\right)\sigma_{u(\Gamma)}^2},$$

where \mathbf{u} , PEV and \mathbf{H}^Γ are as defined previously; Rel_i^{mf} is the reliability of individual i within metafounder mf ; and $\sigma_{u(\Gamma)}^2$ is the additive genetic variance estimated with MF-ssGBLUP.

For BS-ssGBLUP, the model-based reliability is calculated as follows:

$$Rel_i = 1 - \frac{PEV_{i,i}}{\mathbf{H}_{A,i}\sigma_u^2},$$

where \mathbf{H}_A and PEV are as defined previously; Rel_i is the reliability of the individual i , and σ_u^2 is the additive genetic variance estimated with BS-ssGBLUP.

In this study, we investigated the model-based reliabilities of BBL sires having offspring with phenotypes.

Estimators of the LR method

In this study, four estimators, i.e. bias ($\hat{\Delta}$), dispersion (\hat{b}), population accuracy (\hat{acc}) and ratio of accuracies ($\hat{\rho}$) were estimated with the LR method [37] and were used to evaluate the impact of each of the three methods (ssGBLUP, MF-ssGBLUP, and BS-ssGBLUP) on the estimated breeding values (EBV) for crossbred performance, since the LR method has proven to show better analytical properties than the ordinary cross-validation method [37–39]. EBV of focal individuals were denoted as $\hat{\mathbf{u}}_p$ and $\hat{\mathbf{u}}_w$ based on the partial and the whole dataset, respectively. The partial dataset was defined as the set of crossbred animals which were born before a specified cut-off date (we used two cut-off dates in this study, April 1 2021 and May 1 2021), and focal individuals were those born after the specific cut-off date. The number of individuals

in the partial datasets for each trait are in Table 1. For BS-ssGBLUP, EBV are equal to the sum of \mathbf{u}_H and \mathbf{u}_B . The estimators are summarized below.

Bias

The bias estimator $\hat{\Delta}$ is defined as the difference between the mean of EBV based on the partial dataset and the mean of EBV based on the whole dataset, i.e. $\hat{\Delta} = \overline{\hat{\mathbf{u}}_p} - \overline{\hat{\mathbf{u}}_w}$.

In absence of bias, the expected value of this estimator is 0.

Dispersion

The dispersion estimator is defined as the slope of the regression of $\hat{\mathbf{u}}_w$ on $\hat{\mathbf{u}}_p$, which is equal to $\hat{b} = \frac{Cov(\hat{\mathbf{u}}_w, \hat{\mathbf{u}}_p)}{Var(\hat{\mathbf{u}}_p)}$.

The expected value of this estimator is 1 under the assumption that $\hat{\mathbf{u}}_p$ has no dispersion bias, while $\hat{b} < 1$ indicates over-dispersion, and $\hat{b} > 1$ indicates under-dispersion of $\hat{\mathbf{u}}_p$.

Population accuracy

The population accuracy of focal individuals based on the partial dataset can be calculated as $\hat{acc} = \sqrt{\frac{Cov(\hat{\mathbf{u}}_w, \hat{\mathbf{u}}_p)}{(1+\bar{F}-2\bar{f})\sigma_{u,\infty}^2}}$, where \bar{F} is the average inbreeding coefficient of focal individuals, $2\bar{f}$ is the average relationship between focal individuals, and $\sigma_{u,\infty}^2$ is the estimated genetic variance with a partial dataset (assuming that the focal individuals are not under selection in the partial dataset).

Ratio of population accuracies

The ratio of population accuracies estimator is defined as the Pearson correlation between $\hat{\mathbf{u}}_w$ and $\hat{\mathbf{u}}_p$, which is equal to $\hat{\rho} = \frac{Cov(\hat{\mathbf{u}}_w, \hat{\mathbf{u}}_p)}{\sqrt{Var(\hat{\mathbf{u}}_w)Var(\hat{\mathbf{u}}_p)}}$. This is an estimator for $\frac{acc_p}{acc_w}$, where acc_p is the population accuracy based on the partial dataset, and acc_w is the population accuracy based on the whole dataset.

Table 2 Estimates of variance components and their standard error (SE) obtained with three methods

Trait	Method	σ_u^2 (SE)	σ_B^2 (SE)	σ_H^2 (SE)	σ_e^2 (SE)	r_g (SE)	h^2 (SE)
ADG	ssGBLUP	0.008(0.004)	–	–	0.090(0.004)	– 0.531(0.239)	0.082(0.041)
	MF-ssGBLUP	0.007(0.004)	–	–	0.085(0.004)	– 0.515(0.251)	0.076(0.041)
	BS-ssGBLUP	–	0.004(0.003)	0.023(0.009)	0.083(0.005)	– 0.620(0.197)	0.140(0.050)
FCR	ssGBLUP	0.189(0.093)	–	–	2.191(0.099)	– 0.531(0.239)	0.079(0.038)
	MF-ssGBLUP	0.192(0.096)	–	–	2.199(0.097)	– 0.515(0.251)	0.080(0.039)
	BS-ssGBLUP	–	0.227(0.119)	0.106(0.201)	2.215(0.123)	– 0.620(0.197)	0.070(0.048)

σ_u^2 : additive genetic variance; σ_B^2 : additive genetic variance for breed BBL; σ_H^2 : additive genetic variance for breed HOL; σ_e^2 : residual variance; r_g : genetic correlation between ADG and FCR; h^2 : heritability, $h^2 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$ for ssGBLUP and MF-ssGBLUP, $h^2 = \frac{0.5(\sigma_B^2 + \sigma_H^2)}{0.5(\sigma_B^2 + \sigma_H^2) + \sigma_e^2}$ for BS-ssGBLUP

MF-ssGBLUP: Metafounder ssGBLUP; BS-ssGBLUP: Breed-specific ssGBLUP; ADG: average daily gain; FCR: feed conversion ratio

Table 3 Mean model-based reliability of purebred bulls for their crossbred performance

Trait	Method	All	Genotyped	Non-genotyped
ADG	ssGBLUP	0.243	0.266	0.188
	MF-ssGBLUP	0.323	0.354	0.251
	BS-ssGBLUP	0.221	0.243	0.170
FCR	ssGBLUP	0.261	0.285	0.206
	MF-ssGBLUP	0.348	0.380	0.274
	BS-ssGBLUP	0.348	0.374	0.287

MF-ssGBLUP: Metafounder ssGBLUP; BS-ssGBLUP: Breed-specific ssGBLUP; ADG: average daily gain; FCR: feed conversion ratio

Table 4 Bias ($\hat{\Delta}$), dispersion (\hat{b}), population accuracy (\hat{acc}) and ratio of population accuracies ($\hat{\rho}$) of EBV for focal individuals (cut-off date, April 1, 2021) obtained with three methods

Trait	Method	$\hat{\Delta}$	\hat{b}	\hat{acc}	$\hat{\rho}$
ADG	ssGBLUP	0.009	1.035	0.267	0.714
	MF-ssGBLUP	0.011	1.030	0.273	0.729
	BS-ssGBLUP	0.006	1.015	0.239	0.520
FCR	ssGBLUP	-0.036	0.782	0.210	0.691
	MF-ssGBLUP	0.003	0.786	0.213	0.699
	BS-ssGBLUP	0.001	0.787	0.257	0.737

MF-ssGBLUP: Metafounder ssGBLUP; BS-ssGBLUP: Breed-specific ssGBLUP; ADG: average daily gain; FCR: feed conversion ratio

Results

Genetic parameters

Estimated variance components and heritabilities for ADG and FCR and the estimated genetic correlations between ADG and FCR are in Table 2. The genetic variances and covariance obtained with MF-ssGBLUP were scaled for comparison with those of the other two methods. For MF-ssGBLUP, the metafounder relationship coefficients γ_B , γ_{BH} and γ_H were estimated to be 0.702, 0.570, and 0.672, respectively.

The estimated variance components for ssGBLUP and MF-ssGBLUP were similar. The estimates of the heritabilities for ssGBLUP and MF-ssGBLUP were also similar for both ADG (0.082 and 0.076) and FCR (0.079 and 0.080). However, for BS-ssGBLUP, ADG had a heritability estimate of 0.140, which differed from the estimate obtained with the other two methods. The genetic correlation between ADG and FCR was negative and moderate to high for all methods, i.e. -0.531(0.239), -0.515(0.251), and -0.620(0.197) for ssGBLUP, MF-ssGBLUP and BS-ssGBLUP, respectively.

Model-based reliability

Table 3 shows the mean model-based reliabilities of purebred sires for their crossbred performance for ssGBLUP,

MF-ssGBLUP and BS-ssGBLUP. Model-based reliabilities were computed for sires having offspring with phenotypes, and are presented as an average of all sires, an average of genotyped sires, and an average of non-genotyped sires. Among these 67 sires, only 47 have been genotyped. On average, genotyped sires had higher reliabilities than non-genotyped sires, regardless of which method was used. For ADG, MF-ssGBLUP had the highest model-based reliability (0.323), and BS-ssGBLUP had the lowest model-based reliability (0.221), while for FCR, BS-ssGBLUP and MF-ssGBLUP had the highest model-based reliability (0.348), and ssGBLUP had the lowest model-based reliability (0.261). For both traits, MF-ssGBLUP always had a higher model-based reliability than ssGBLUP.

Predictive ability

Four estimators ($\hat{\Delta}$, \hat{b} , \hat{acc} and $\hat{\rho}$) in the LR method were used to evaluate the predictive ability of ssGBLUP, MF-ssGBLUP, and BS-ssGBLUP for two focal sets of individuals. The results for the different datasets of individuals are in Table 4 for those with a cut-off date at April 1 2021 and in Additional file 2: Table S1 for those with a cut-off date at May 1 2021. The results differed slightly between datasets, but the conclusions were similar. Therefore, in the remainder of the paper, we focus only on the results in Table 4.

As shown in Table 4, the differences between $\hat{\Delta}$ across the three methods were small. For all methods, the values of $\hat{\Delta}$ were close to the expected value (equal to 0) for both traits, while the values of \hat{b} were close to the expected value (equal to 1) for ADG and deviated from the expected value for FCR. For ADG, population accuracy (\hat{acc}) was highest (0.273) with MF-ssGBLUP, and lowest (0.239) with BS-ssGBLUP, while for FCR, it was highest (0.257) with BS-ssGBLUP, and lowest (0.210) with ssGBLUP. For both traits, population accuracy was higher with MF-ssGBLUP than with ssGBLUP. The ratios of population accuracies based on the partial and whole datasets were for ssGBLUP, MF-ssGBLUP and BS-ssGBLUP, respectively, 0.714, 0.729, and 0.520 for ADG, and 0.691, 0.699, and 0.737 for FCR.

Discussion

In this work, first we compared the estimates of genetic parameters for ADG and FCR obtained with ssGBLUP, MF-ssGBLUP and BS-ssGBLUP. In general, variance components and heritability estimates for FCR did not differ considerably between methods, while for ADG, those estimated with BS-ssGBLUP deviated largely from those estimated with ssGBLUP and MF-ssGBLUP. Then, we evaluated the impact of these methods on the predictive ability for crossbred performance. For both traits,

the estimators ($\widehat{\Delta}$, \widehat{b} , and \widehat{acc}) in the LR method showed that the predictive ability of MF-ssGBLUP was always superior to that of ssGBLUP, whereas the comparison of the predictive ability of BS-ssGBLUP with the other two methods showed no consistent result.

Genetic parameters

Variance components, heritabilities and genetic correlations obtained with ssGBLUP and with MF-ssGBLUP were similar for both ADG and FCR. This observation is in line with previous studies [18, 40]. However, for BS-ssGBLUP, the estimated genetic parameters for FCR were similar to those with the other two methods, while the result was opposite for ADG. As shown in Table 2, the additive genetic variance for FCR in the sire breed and dam breed was 0.227 and 0.106, respectively, while for ADG, it was 0.004, and 0.023, respectively. Our results are not consistent with those reported by Poulsen et al. [19] on simulated data, who found that the estimated variance components from the three methods were similar, with those from MF-ssGBLUP being closer to those from BS-ssGBLUP than those from ssGBLUP. One possible reason for this difference may be the lack of sufficient information in our dataset to distinguish the additive genetic variances between the sire breed and the dam breed in BS-ssGBLUP. To date, few studies have examined whether there are differences in the variance components, heritabilities and genetic correlations between these three methods, and further investigation is needed.

In our study, ADG and FCR were lowly heritable with heritability estimates ranging from 0.081 to 0.153 for ADG, and from 0.080 to 0.084 for FCR. A few studies have reported similarly low values [41–43], but in general, ADG and FCR are considered as moderately to highly heritable traits [44, 45]. Our results could be due to the short testing period used. In general, ADG and FCR are normally collected over longer test periods (3–6 months) [44, 45] than the one-month test period in our study. Furthermore, Ahlberg et al. [46] pointed out that during different periods, the phenotypic correlations for each shortened test duration differed. Although the heritability estimates for ADG and FCR are lower than those reported in previous studies, the moderate to high negative genetic correlation between ADG and FCR is in agreement with other studies [44, 45].

Model-based reliabilities

In terms of model-based reliability with MF-ssGBLUP, the usual definition (expressed as $1 - \frac{PEV_{ii}}{H_{ii}^{(r)} \sigma_u^2(r)}$) is inappropriate for metafounder relationships, as pointed out by Bermann et al. [36], since it would underestimate

reliabilities. To account for this, Bermann et al. [36] proposed a new method where reliabilities are calculated from contrasts to a reference metafounder. By applying this method with MF-ssGBLUP in our study, the reliabilities of purebred sires increased by almost 30%, compared with the usual definition (results not shown). In our study, there were two metafounders, one representing BBL and the other HOL. Each individual would have two reliabilities corresponding to BBL and HOL. For BS-ssGBLUP, there were also two reliabilities based on two breed-specific relationship matrices.

Within each method, for ADG and FCR, the reliabilities for the genotyped sire group were always larger than for the non-genotyped sire group. This result is in line with previous studies [5, 18]. In terms of reliabilities across methods, as expected, MF-ssGBLUP always had higher reliabilities than ssGBLUP. However, for BS-ssGBLUP, the results were not consistent, i.e. for FCR the reliabilities from BS-ssGBLUP were similar to those from MF-ssGBLUP, but for ADG they were the lowest among the three methods. This could be due to the fact that the genetic parameters estimated for ADG with BS-ssGBLUP deviated a lot from the estimated parameters with the other two methods, but also to the small sample size for the sires.

Predictive ability

In this study, four estimators, $\widehat{\Delta}$, \widehat{b} , \widehat{acc} and $\widehat{\rho}$, in the LR method [37] were used to evaluate the predictive ability of ssGBLUP, BS-ssGBLUP and MF-ssGBLUP. Table 4 shows that for $\widehat{\Delta}$ and \widehat{b} , there are little differences between these three methods.

The difference between \widehat{b} and its expected value showed that the EBV of FCR were over-dispersed, and that their deviation from the expected value were larger than for ADG. Mäntysaari et al. [47] have suggested that over-dispersion of EBV may be due to strong selection. In terms of $\widehat{\rho}$, Legarra and Reverter [37] pointed out that it is an estimator of change in population accuracy, but not a measure of population accuracy. Its reciprocal minus 1 can be interpreted as the relative increase of population accuracy from partial to whole information. For example, a value of 0.699 for the $\widehat{\rho}$ of FCR with MF-ssGBLUP means that the corresponding increase in population accuracy from the partial to the whole dataset is 43.1%.

As expected, MF-ssGBLUP always had a slightly higher population accuracy than ssGBLUP. In a multiple-breed beef cattle population, Junqueira et al. [20] and Kluska et al. [21] found that, compared to ssGBLUP, MF-ssGBLUP decreased bias in genomic evaluations. The same result has also been found for crossbred pigs [18]. However, with BS-ssGBLUP, opposite \widehat{acc} values were obtained for ADG and FCR, which is similar to

the model-based reliabilities that also showed opposite results for ADG and FCR with BS-ssGBLUP. As already mentioned, one possible reason is that the estimated genetic variance with BS-ssGBLUP for ADG deviated a lot from the estimated parameters with the other two methods. For FCR, for which the estimated genetic variance components were similar across the three methods, both BS-ssGBLUP and MF-ssGBLUP had a better predictive ability than ssGBLUP, which is in line with a previous study [19]. In addition, we found that for FCR, BS-ssGBLUP had a better predictive ability than MF-ssGBLUP, which was not consistent with the results of Poulsen et al. [19] who reported similar predictive abilities for BS-ssGBLUP and MF-ssGBLUP. A possible reason for the conflicting results observed in our study may be that the metafounder relationship matrix Γ could be accurately estimated in the simulated dataset in Poulsen et al. [19], whereas in our case the estimates of Γ maybe inaccurate, and could be biased because of the small number of genotyped animals, as is the case for BBL. Inaccurate estimates of Γ may affect the performance of MF-ssGBLUP. Moreover, missing genotypes were imputed based on a combination of different SNP panels (EuroG 10k Bead chip and Eurogenomics 75K custom SNP chip), which could make the estimation of Γ even less accurate. We have also investigated the predictive ability of pedigree BLUP and metafounder pedigree BLUP methods (see Additional file 2: Table S1) and found that these two methods had a higher estimated population accuracy than ssGBLUP, MF-ssGBLUP and BS-ssGBLUP, but also that the estimated genetic variances were much smaller. These are puzzling results, which show that it is necessary to better understand how the estimation of the population accuracy in the LR method performs with imprecisely estimated parameters.

In terms of allele tracing, errors in detecting the breed of origin of alleles can affect a model's predictive ability especially for a distantly-related crossbred population [15, 30, 48]. In our study, only few such errors were expected since all the alleles on one chromosome should originate from the same breed (either the sire breed or the dam breed) [30]. We also tested the accuracy of allele tracing in a simulated two-way crossbred population, and this was equal to 100% (results not shown). However, in more complicated situations (three-way, four-way, and rotational crossbred populations), our method is not suitable, and a more advanced method for tracing the breed origin of alleles is needed [30, 49].

Overall, MF-ssGBLUP and BS-ssGBLUP had a better predictive ability than ssGBLUP, when the estimated

variance components were consistent across the methods. However, more research with larger datasets is needed for investigating the differences between these methods.

Conclusions

Our results reveal that, for FCR, there are little differences in the estimated genetic parameters of a bivariate model among the ssGBLUP, MF-ssGBLUP, and BS-ssGBLUP methods. However, for ADG, the estimated genetic parameters obtained with BS-ssGBLUP showed a large deviation compared to those with ssGBLUP and MF-ssGBLUP. The values of four estimators implemented in the LR method showed that, for the genetic evaluation for crossbred performance in a two-way crossbred cattle production system, MF-ssGBLUP and BS-ssGBLUP had a better predictive ability than ssGBLUP, when the estimated variance components were consistent across the three methods. In general, compared with BS-ssGBLUP, MF-ssGBLUP is more robust in its superiority over ssGBLUP.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12711-023-00792-4>.

Additional file 1: Figure S1. Details of phenotype editing.

Additional file 2: Table S1. Bias ($\hat{\Delta}$), dispersion (\hat{b}), population accuracy (\hat{aCC}), ratio of population accuracies ($\hat{\rho}$) of EBV, average inbreeding coefficient (F) and average relationship (2f) for focal individuals (cut off date, May 1, 2021) and the estimated additive genetic variance in partial dataset ($\sigma_{u,\infty}^2$) and whole dataset (σ_u^2) with different method.

Acknowledgements

Project partners in the FutureBeefCross project are acknowledged for conducting the data collection. Anders Fogh and Line Hjortø (SEGES Innovation Cattle) are acknowledged for providing explanations about the experiment and doing quality control of the data. Jette Odgaard Villemoes is acknowledged for English editing of the manuscript.

Author contributions

QM performed data analysis and wrote the manuscript. OFC, TX, SZ and HL supervised and assisted at all stages of the study, including the writing of the manuscript. All authors read and approved the final manuscript.

Funding

SZ acknowledges funding from the National Key R&D Program of China (2019YFE0115400); OFC and HL acknowledge funding from the FutureBeef-Cross project supported by the Green Development and Demonstration Programme (GUDP) from the Danish Ministry of Food, Agriculture and Fisheries (J. nr. 34009-18-1434); QM acknowledges funding from the China Scholarship Council (CSC) Scholarship.

Availability of data and materials

The phenotypic data is owned by partners of the FutureBeefCross project. The pedigree and genomic data are property of Nordic Cattle Genetic Evaluation Ltd (NAV, Aarhus, Denmark) and Viking Genetics (Randers, Denmark). None of this data is for public distribution.

Declarations

Ethics approval and consent to participate

Data recording and sample collection were conducted following Danish laws of management and welfare procedures for animal production.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 30 June 2022 Accepted: 8 March 2023

Published online: 17 March 2023

References

- Berry D. Invited review: beef-on-dairy—the generation of crossbred beef x dairy cattle. *J Dairy Sci.* 2021;104:3789–819.
- de Vries M, van Middelaar CE, de Boer IJM. Comparing environmental impacts of beef production systems: a review of life cycle assessments. *Livest Sci.* 2015;178:279–88.
- Stock J, Bennewitz J, Hinrichs D, Wellmann R. A review of genomic models for the analysis of livestock crossbred data. *Front Genet.* 2020;11:568.
- Bedere N, Berghof TV, Peeters K, Pinaud-van der Laan M-H, Visscher J, David I, et al. Using egg production longitudinal recording to study the genetic background of resilience in purebred and crossbred laying hens. *Genet Sel Evol.* 2022;54:26.
- Xiang T, Nielsen B, Su G, Legarra A, Christensen OF. Application of single-step genomic evaluation for crossbred performance in pig. *J Anim Sci.* 2016;94:936–48.
- Dekkers JC. Marker-assisted selection for commercial crossbred performance. *J Anim Sci.* 2007;85:2104–14.
- Wientjes YCJ, Calus MPL. Board invited review: the purebred-crossbred correlation in pigs: a review of theory, estimates, and implications. *J Anim Sci.* 2017;95:3467–78.
- Calus M, Bos J, Duenk P, Wientjes Y, editors. The purebred-crossbred correlation in broilers and layers: a review. In *Proceedings of the 71th Annual Meeting of the European Federation of Animal Science: 1–4 December 2020; virtual meeting: 2020*.
- Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. *J Dairy Sci.* 2009;92:4656–63.
- Christensen OF, Lund MS. Genomic prediction when some animals are not genotyped. *Genet Sel Evol.* 2010;42:2.
- Legarra A, Christensen OF, Aguilar I, Misztal I. Single step, a general approach for genomic selection. *Livest Sci.* 2014;166:54–65.
- Lourenco DAL, Tsuruta S, Fragomeni BO, Chen CY, Herring WO, Misztal I. Crossbreed evaluations in single-step genomic best linear unbiased predictor using adjusted realized relationship matrices. *J Anim Sci.* 2016;94:909–19.
- Christensen OF, Madsen P, Nielsen B, Su G. Genomic evaluation of both purebred and crossbred performances. *Genet Sel Evol.* 2014;46:23.
- García-Cortés LA, Toro MÁ. Multibreed analysis by splitting the breeding values. *Genet Sel Evol.* 2006;38:601–15.
- Ibáñez-Escriche N, Fernando RL, Toosi A, Dekkers JC. Genomic selection of purebreds for crossbred performance. *Genet Sel Evol.* 2009;41:12.
- Sevillano CA, Vandenplas J, Bastiaansen JWM, Bergsma R, Calus MPL. Genomic evaluation for a three-way crossbreeding system considering breed-of-origin of alleles. *Genet Sel Evol.* 2017;49:75.
- Legarra A, Christensen OF, Vitezica ZG, Aguilar I, Misztal I. Ancestral relationships using metafounders: finite ancestral populations and across population relationships. *Genetics.* 2015;200:455–68.
- Xiang T, Christensen OF, Legarra A. Genomic evaluation for crossbred performance in a single-step approach with metafounders. *J Anim Sci.* 2017;95:1472–80.
- Poulsen BG, Ostensen T, Nielsen B, Christensen OF. Predictive performances of animal models using different multibreed relationship matrices in systems with rotational crossbreeding. *Genet Sel Evol.* 2022;54:25.
- Junqueira VS, Lopes PS, Lourenco D, Silva FFE, Cardoso FF. Applying the metafounders approach for genomic evaluation in a multibreed beef cattle population. *Front Genet.* 2020;11: 556399.
- Kluska S, Masuda Y, Ferraz JBS, Tsuruta S, Eler JP, Baldi F, et al. Metafounders may reduce bias in composite cattle genomic predictions. *Front Genet.* 2021;12: 678587.
- Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet.* 2007;81:1084–97.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75.
- van Grevenhof EM, Vandenplas J, Calus MP. Genomic prediction for crossbred performance using metafounders. *J Anim Sci.* 2019;97:548–58.
- Macedo FL, Christensen OF, Astruc J-M, Aguilar I, Masuda Y, Legarra A. Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. *Genet Sel Evol.* 2020;52:47.
- VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci.* 2008;91:4414–23.
- Christensen OF. Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. *Genet Sel Evol.* 2012;44:37.
- García-Baccino CA, Legarra A, Christensen OF, Misztal I, Pocrnic I, Vitezica ZG, et al. Metafounders are related to Fst fixation indices and reduce bias in single-step genomic evaluations. *Genet Sel Evol.* 2017;49:34.
- Wei M, van der Werf JH. Maximizing genetic response in crossbreds using both purebred and crossbred information. *Anim Prod.* 1994;59:401–13.
- Eiriksson JH, Karaman E, Su G, Christensen OF. Breed of origin of alleles and genomic predictions for crossbred dairy cows. *Genet Sel Evol.* 2021;53:84.
- Madsen P, Jensen J. A user's guide to DMU. A package for analysing multivariate mixed models. Version 6, release 5.2. University of Aarhus: Center for Quantitative Genetics and Genomics. 2013.
- Mei Q, Fu C, Li J, Zhao S, Xiang T. blupADC: An R package and shiny toolkit for comprehensive genetic data analysis in animal and plant breeding. *bioRxiv.* 2021. <https://doi.org/10.1101/2021.09.09.459557>.
- Falconer D. Introduction to quantitative genetics. Harlow: Pearson Education Limited; 1996.
- Mrode RA. Linear models for the prediction of animal breeding values. Wallingford: CABI Publishing; 2014.
- Jackson C. Multi-state models for panel data: the msm package for R. *J Stat Softw.* 2011;38:1–28.
- Bermann M, Aguilar I, Lourenco D, Misztal I, Legarra A. Reliabilities of estimated breeding values in models with metafounders. *Genet Sel Evol.* 2023;55:6.
- Legarra A, Reverter A. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet Sel Evol.* 2019;50:53.
- Macedo FL, Reverter A, Legarra A. Behavior of the linear regression method to estimate bias and accuracies with correct and incorrect genetic evaluation models. *J Dairy Sci.* 2020;103:529–44.
- Bermann M, Legarra A, Hollifield MK, Masuda Y, Lourenco D, Misztal I. Validation of single-step GBLUP genomic predictions from threshold models using the linear regression method: an application in chicken mortality. *J Anim Breed Genet.* 2021;138:4–13.
- Fu C, Ostensen T, Christensen OF, Xiang T. Single-step genomic evaluation with metafounders for feed conversion ratio and average daily gain in Danish Landrace and Yorkshire pigs. *Genet Sel Evol.* 2021;53:79.
- Inoue K, Kobayashi M, Shoji N, Kato K. Genetic parameters for fatty acid composition and feed efficiency traits in Japanese Black cattle. *Animal.* 2011;5:987–94.
- Rolf MM, Taylor JF, Schnabel RD, McKay SD, McClure MC, Northcutt SL, et al. Genome-wide association analysis for feed efficiency in Angus cattle. *Anim Genet.* 2012;43:367–74.
- Martin P, Taussat S, Vinet A, Krauss D, Maupetit D, Renand G. Genetic parameters and genome-wide association study regarding feed efficiency and slaughter traits in Charolais cows. *J Anim Sci.* 2019;97:3684–98.
- Polizel GHG, Grigoletto L, Carvalho ME, Junior PR, Ferraz JBS, de Almeida Santana MH. Genetic correlations and heritability estimates for dry matter

intake, weight gain and feed efficiency of Nelore cattle in feedlot. *Livest Sci.* 2018;214:209–10.

45. Torres-Vázquez JA, van der Werf JH, Clark SA. Genetic and phenotypic associations of feed efficiency with growth and carcass traits in Australian Angus cattle. *J Anim Sci.* 2018;96:4521–31.
46. Ahlberg CM, Allwardt K, Broocks A, Bruno K, McPhillips L, Taylor A, et al. Test duration for water intake, ADG, and DMI in beef cattle. *J Anim Sci.* 2018;96:3043–54.
47. Mäntysaari E, Koivula M, Strandén I. Symposium review: single-step genomic evaluations in dairy cattle. *J Dairy Sci.* 2020;103:5314–26.
48. Lopes MS, Bovenhuis H, Hidalgo AM, Van Arendonk JA, Knol EF, Bastiaansen JW. Genomic selection for crossbred performance accounting for breed-specific effects. *Genet Sel Evol.* 2017;49:51.
49. Vandenplas J, Calus MP, Sevillano CA, Windig JJ, Bastiaansen JW. Assigning breed origin to alleles in crossbred animals. *Genet Sel Evol.* 2016;48:61.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

