GSE **G**enetics
Selection
Evolution

Open Access

Genomic evaluation for a three-way crossbreeding system considering breed-of-origin of alleles

Claudia A. Sevillano[1,2*], Jeremie Vandenplas[1], John W. M. Bastiaansen[1], Rob Bergsma[2] and Mario P. L. Calus[1]

Abstract

Background: Genomic prediction of purebred animals for crossbred performance can be based on a model that estimates effects of single nucleotide polymorphisms (SNPs) in purebreds on crossbred performance. For crossbred performance, SNP effects might be breed-specific due to differences between breeds in allele frequencies and linkage disequilibrium patterns between SNPs and quantitative trait loci. Accurately tracing the breed-of-origin of alleles (BOA) in three-way crosses is possible with a recently developed procedure called BOA. A model that accounts for breed-specific SNP effects (BOA model), has never been tested empirically on a three-way crossbreeding scheme. Therefore, the objectives of this study were to evaluate the estimates of variance components and the predictive accuracy of the BOA model compared to models in which SNP effects for crossbred performance were assumed to be the same across breeds, using either breed-specific allele frequencies (G_A model) or allele frequencies averaged across breeds (G_B model). In this study, we used data from purebred and three-way crossbred pigs on average daily gain (ADG), back fat thickness (BF), and loin depth (LD).

Results: Estimates of variance components for crossbred performance from the BOA model were mostly similar to estimates from models G_A and G_B. Heritabilities for crossbred performance ranged from 0.24 to 0.46 between traits. Genetic correlations between purebred and crossbred performance (r_{pc}) across breeds ranged from 0.30 to 0.62 for ADG and from 0.53 to 0.74 for BF and LD. For ADG, prediction accuracies of the BOA model were higher than those of the G_A and G_B models, with significantly higher accuracies only for one maternal breed. For BF and LD, prediction accuracies of models G_A and G_B were higher than those of the BOA model, with no significant differences. Across all traits, models G_A and G_B yielded similar predictions.

Conclusions: The BOA model yielded a higher prediction accuracy for ADG in one maternal breed, which had the lowest r_{pc} (0.30). Using the BOA model was especially relevant for traits with a low r_{pc}. In all other cases, the use of crossbred information in models G_A and G_B, does not jeopardize predictions and these models are more easily implemented than the BOA model.

Background

Genomic selection (GS) is more accurate than pedigree-based selection, and thus was developed for purebred (PB) populations of many farm species [1–4]. However, many production systems use crossbreeding schemes to produce crossbred (CB) individuals for commercial production. Crossbreeding in plants is common practice in many crops, such as maize. Crossbreeding in animals is common practice for pigs and poultry, and, in cattle, the use of crosses or composite breeds contributes largely to the beef and dairy industry. If selection is based on the performance measured on PB individuals, the rate of genetic change observed in CB individuals may be reduced because of differences in additive variance between PB and CB individuals, and because the genetic correlation between performance in PB and CB individuals (r_{pc}) is lower than 1 [5, 6]. With r_{pc} values of 0.7 or lower, using

*Correspondence: claudia.sevillanodelaguila@wur.nl
[1] Wageningen University & Research Animal Breeding and Genomics, 6700 AH Wageningen, The Netherlands
Full list of author information is available at the end of the article

Sevillano *et al. Genet Sel Evol* (2017) 49:75

Page 2 of 14

only PB performance was predicted to yield considerably less genetic progress in CB performance compared to using performance of both PB and CB [7, 8]. In pigs, r_{pc} lower than 0.7 were reported for daily gain, daily feed intake, feed conversion ratio and residual feed intake [9–11], and also in poultry for egg number [12], and in cattle for weight-related traits [13]. In maize, the correlation between PB and CB performance for grain yield (GY) is lower than that for grain dry matter content (GDMC), and it was observed that models that do not include CB information failed to predict the performance of CB for GY but for GDMC yielded a high prediction accuracy [14].

With GS, training with CB information is facilitated because GS eliminates the disadvantages of having to record pedigree data on CB individuals [7]. Moreover, GS using CB information could benefit from models that estimate the effects on CB performance of markers that segregate within the parental breeds, as suggested by Dekkers [7], Ibáñez-Escriche et al. [15], Kinghorn et al. [16] and Christensen et al. [17, 18] in the context of animal breeding, and by Schrag et al. [14] in the context of hybrid performance in maize.

A commonly used GS model, known as genomic best linear unbiased prediction (GBLUP) [19], replaces the pedigree-based relationship matrix by a genomic relationship matrix. The values in the genomic relationship matrix are a function of allele content and allele frequencies [20]. Consequently, the genomic relationship matrix is built under the assumption that all individuals belong to the same population, with the same average allele contents. Moreover, GBLUP implicitly assumes a single value for the linkage disequilibrium between a single nucleotide polymorphism (SNP) and a quantitative trait locus (QTL). When individuals originate from different populations, as in the crossbreeding context, these assumptions are violated because allele frequencies and the linkage disequilibrium patterns across the genome differ between breeds [21–23]. Models that account for breed-specific allele frequencies were tested with simulated and real data and showed no improvement in prediction accuracies [24–26]. Models that, in addition to including breed-specific allele frequencies, also account for breed-specific SNP effects did outperform models in which SNP effects were assumed to be the same across breeds. However, these results were only observed in simulation studies under some conditions (i.e., low SNP density, large training data size, and low breed relatedness) and where breed-of-origin of alleles was assumed to be known without error [15, 27]. With real data from a two-way crossbreeding scheme, Xiang et al. [28] and Lopes et al. [29] reached different conclusions. When using a model that accounted for breed-specific SNP effects compared to a model in which SNP effects were assumed to be the same

across breeds, Xiang et al. [28] found improved prediction accuracies and reduced bias of prediction, whereas, Lopes et al. [29] found similar prediction accuracies between the two models. The benefit of a two-way CB is that tracing the breed-of-origin of alleles is relatively straightforward. However, many crossbreeding schemes are based on a three-way cross, for which tracing the breed-of-origin of alleles is considerably more complicated [30]. Recently we have developed a procedure that enables breed-of-origin assignment (BOA) of alleles in three-way CB animals [31]. BOA allows empirical testing of the model that accounts for breed-specific SNP effects in real data. Therefore, the objectives of this study were to evaluate the estimates of variance components and the accuracy of a model that accounts for breed-specific SNP effects using information from both PB and three-way CB pigs for average daily gain (ADG), back fat thickness (BF), and loin depth (LD).

Methods
Data
The pig data consisted of three PB populations: Synthetic boar (S), Landrace (LR), and Large White (LW), and a three-way CB population: S (LR × LW) or S (LW × LR), produced by crossing the above-mentioned PB populations. The numbers of available genotypes and phenotypes per trait and per population are in Table 1. All pigs were genotyped using one of the three following SNP panels: Illumina PorcineSNP60.v2 BeadChip (60 K. v2), Illumina PorcineSNP60 BeadChip (60 K), or Illumina PorcineSNP10 BeadChip (10 K). Pigs genotyped with the 60 K or 10 K chips were imputed to the 60 K. v2 panel using FImpute Version 2.2 software [32]. SNP quality control and imputation were applied on the same dataset in a previous study [31], in which more details are provided. The final SNP set for subsequent analyses consisted of 52,164 SNPs. Phenotypes for ADG (g/day), BF (mm), and LD (mm), were measured for most of the PB and CB pigs. ADG for PB was calculated as the difference of on-test body weight measured on average at 60 days

Table 1 Number of genotypes and phenotypes available for each trait and population

Population	Genotypes	ADG	BF	LD
S	2733	2575	2616	2595
LR	4148	2333	3605	2386
LW	7103	5294	6769	5469
CB	1706	1675	1676	1681
Total	15,690	11,877	14,666	12,131

S Synthetic boar, *LR* Landrace, *LW* Large White, *CB* three-way crossbred pigs

ADG average daily gain, *BF* back fat thickness, *LD* loin depth

Sevillano *et al. Genet Sel Evol* (2017) 49:75

Page 3 of 14

of age and off-test body weight measured on average at 173 days of age. ADG for CB was calculated as the difference of on-test body weight measured on average at 70 days of age and body weight at the end of the finishing period, which was on average 120 kg. BF and LD for PB were measured on average at 173 days of age using an ultrasound instrument, while BF and LD for CB were measured on the carcass after slaughter using a probe, named "capteur gras maigre" (CGM; Sydel, France). For all phenotyped pigs, four generations of pedigree information were included.

Analyses

GBLUP model with breed-specific partial relationship matrices (BOA model)

To account for the breed-specific effect of SNPs, the following 4-trait animal model with three breed-specific partial relationship matrices ($\mathbf{G^{(S)}}$, $\mathbf{G^{(LR)}}$ and $\mathbf{G^{(LW)}}$) was fitted (BOA model):

$$\mathbf{y_S} = \mathbf{X_S b_S} + \mathbf{W_S u_S} + \mathbf{Z_S a_S} + \mathbf{e_S},$$

$$\mathbf{y_{LR}} = \mathbf{X_{LR} b_{LR}} + \mathbf{W_{LR} u_{LR}} + \mathbf{Z_{LR} a_{LR}} + \mathbf{e_{LR}},$$

$$\mathbf{y_{LW}} = \mathbf{X_{LW} b_{LW}} + \mathbf{W_{LW} u_{LW}} + \mathbf{Z_{LW} a_{LW}} + \mathbf{e_{LW}},$$

$$\begin{aligned}\mathbf{y_{CB}} = {} & \mathbf{X_{CB} b_{CB}} + \mathbf{W_{CB} u_{CB}} + \mathbf{Z_{CB} g_{CB}^{(S)}} \\ & + \mathbf{Z_{CB} g_{CB}^{(LR)}} + \mathbf{Z_{CB} g_{CB}^{(LW)}} + \mathbf{e_{CB}},\end{aligned}$$

where $\mathbf{y_S}$, $\mathbf{y_{LR}}$, $\mathbf{y_{LW}}$, and $\mathbf{y_{CB}}$ are the vectors of the phenotypes for S, LR, LW, and CB pigs, respectively; $\mathbf{b_S}$, $\mathbf{b_{LR}}$, $\mathbf{b_{LW}}$, and $\mathbf{b_{CB}}$ represent the vectors of fixed effects (listed in Table 2) and $\mathbf{X_S}$, $\mathbf{X_{LR}}$, $\mathbf{X_{LW}}$, and $\mathbf{X_{CB}}$ are the respective incidence matrices relating pig records to fixed effects; $\mathbf{u_S}$, $\mathbf{u_{LR}}$, $\mathbf{u_{LW}}$, and $\mathbf{u_{CB}}$ represent the vectors of random common litter effects, and $\mathbf{W_S}$, $\mathbf{W_{LR}}$, $\mathbf{W_{LW}}$, and $\mathbf{W_{CB}}$ are the respective incidence matrices relating pig records to litter effects; $\mathbf{a_S}$, $\mathbf{a_{LR}}$, and $\mathbf{a_{LW}}$, are the vectors of additive genetic effects in PB, $\mathbf{g_{CB}^{(S)}}$, $\mathbf{g_{CB}^{(LR)}}$, and $\mathbf{g_{CB}^{(LW)}}$ are the vectors of the additive genetic effect of PB gametes in CB, and $\mathbf{Z_S}$, $\mathbf{Z_{LR}}$, $\mathbf{Z_{LW}}$, and $\mathbf{Z_{CB}}$ are the respective incidence matrices. Because each model was run for each trait and only pigs with phenotypes were included, $\mathbf{Z}$ incidence matrices relating pig records to additive genetic effects were identity matrices when variance components were estimated. Finally, $\mathbf{e_S}$, $\mathbf{e_{LR}}$, $\mathbf{e_{LW}}$, and $\mathbf{e_{CB}}$ represent the vectors of random residual effects. The variance–covariance of the common litter effect and residual effect were:

$$\mathrm{Var}\begin{bmatrix} \mathbf{u_S} \\ \mathbf{u_{LR}} \\ \mathbf{u_{LW}} \\ \mathbf{u_{CB}} \end{bmatrix} = \begin{bmatrix} \sigma_{u_S}^2 & 0 & 0 & 0 \\ 0 & \sigma_{u_{LR}}^2 & 0 & 0 \\ 0 & 0 & \sigma_{u_{LW}}^2 & 0 \\ 0 & 0 & 0 & \sigma_{u_{CB}}^2 \end{bmatrix} \otimes \mathbf{I},$$

Table 2 Fixed effects used in the GBLUP models for average daily gain (ADG), back fat thickness (BF), and loin depth (LD), for purebred (PB) (i.e. S, LR, LW) and three-way crossbred (CB) pigs

Trait	Population	Fixed effects
ADG	PB	farm_breed_sex + b_a × birth weight
	CB	trial + farm_sex + b_a × birth weight
BF, LD	PB	farm_breed_sex + b_b × off_test BW
	CB	trial + farm_sex + b_c × hot carcass weight

b_a, b_b, b_c, are regression coefficients for birth weight, off-test BW, and hot carcass weight, respectively

and

$$\mathrm{Var}\begin{bmatrix} \mathbf{e_S} \\ \mathbf{e_{LR}} \\ \mathbf{e_{LW}} \\ \mathbf{e_{CB}} \end{bmatrix} = \begin{bmatrix} \sigma_{e_S}^2 & 0 & 0 & 0 \\ 0 & \sigma_{e_{LR}}^2 & 0 & 0 \\ 0 & 0 & \sigma_{e_{LW}}^2 & 0 \\ 0 & 0 & 0 & \sigma_{e_{CB}}^2 \end{bmatrix} \otimes \mathbf{I}.$$

The variance–covariance of additive genetic effect for breed S origin was:

$$\begin{aligned}\mathrm{Var}\begin{bmatrix} \mathbf{a_S} \\ \mathbf{a_{CB}^{(S)}} \\ \mathbf{g_S} \\ \mathbf{g_{CB}^{(S)}} \end{bmatrix} &= \begin{bmatrix} \sigma_{a_S}^2 & \sigma_{a_S,g_S} \\ \sigma_{g_S,a_S} & \sigma_{g_S}^2 \end{bmatrix} \otimes \mathbf{G^{(S)}} \\[2mm] &= \begin{bmatrix} \sigma_{a_S}^2 & \sigma_{a_S,g_S} \\ \sigma_{g_S,a_S} & \sigma_{g_S}^2 \end{bmatrix} \otimes \begin{bmatrix} \mathbf{G_{S,S}} & \mathbf{G_{S,CB}^{(S)}} \\ \mathbf{G_{CB,S}^{(S)}} & \mathbf{G_{CB,CB}^{(S)}} \end{bmatrix},\end{aligned}$$

where S pigs have additive effects (i.e. breeding values), $\mathbf{a_S}$ for PB performance and $\mathbf{a_{CB}^{(S)}}$ for CB performance. The CB pigs have additive effects from the breed S gametes, $\mathbf{g_{CB}^{(S)}}$ for CB performance and $\mathbf{g_S}$ for PB performance. This last effect, $\mathbf{g_S}$, is an artificial random vector that is added to be able to define the variance–covariance of additive genetic effects with the above Kronecker product, but does not have practical relevance. The matrix $\mathbf{G^{(S)}}$ is a breed-specific partial relationships matrix for breed S which contains four blocks, one for within S pigs ($\mathbf{G_{S,S}}$), two for S with CB pigs ($\mathbf{G_{S,CB}^{(S)}}$ and $\mathbf{G_{CB,S}^{(S)}}$), and one for within CB pigs ($\mathbf{G_{CB,CB}^{(S)}}$).

The variance–covariance structures for the origin of breeds LR and LW are defined similarly, and the three variance–covariance structures are assumed independent, i.e. no covariances are considered between S, LR, and LW effects [18]. There are six genetic variance components, two for each breed-of-origin, and three covariance components, one for each breed-of-origin. To construct the three breed-specific partial relationship matrices, $\mathbf{G^{(S)}}$, $\mathbf{G^{(LR)}}$, and $\mathbf{G^{(LW)}}$, we used the breed-of-origin of phased alleles in CB pigs. Then, the breed-specific

Sevillano *et al. Genet Sel Evol (2017) 49:75*

Page 4 of 14

partial relationship submatrices are defined as, e.g. breed S origin:

$$\mathbf{G_{S,S}} = \left(\mathbf{M^S} - 2\mathbf{1p^{S'}}\right)\mathbf{D^S}\left(\mathbf{M^S} - 2\mathbf{1p^{S'}}\right)'/N,$$

$$\mathbf{G_{S,CB}} = \left(\mathbf{M^S} - 2\mathbf{1p^{S'}}\right)\mathbf{D^S}\left(\mathbf{M^{CB}} - \mathbf{1p^{S'}}\right)'/N,$$

$$\mathbf{G_{CB,CB}} = \left(\mathbf{M^{CB}} - \mathbf{1p^{S'}}\right)\mathbf{D^S}\left(\mathbf{M^{CB}} - \mathbf{1p^{S'}}\right)'/N,$$

where $\mathbf{M^S}$ is a matrix containing breed-specific allele content for breed S pigs (coded as 0, 1, or 2), $\mathbf{M^{CB}}$ is a matrix containing breed-specific allele content for CB pigs (coded as 0, or 1), alleles that were not assigned a breed-of-origin were set to missing, $\mathbf{p^S}$ is the vector of breed S specific frequencies of the counted allele (p_j^s). p_j^s was calculated across S and CB pigs by counting the occurrences of alleles originating from the S breed and coded as 1, across the S breed and in CB, divided by the total number of S alleles in the S breed and CB on locus j. $\mathbf{D^S}$ is diagonal with $D_{jj}^S = \frac{1}{2p_j^S\left(1 - p_j^S\right)}$. N is the number of SNPs.

The breed-specific partial relationship submatrices $\mathbf{G^{(LR)}}$ and $\mathbf{G^{(LW)}}$ are defined similarly to $\mathbf{G^{(S)}}$. However, the entries of the $\mathbf{M^{CB}}$ matrix containing the breed-specific allele content for CB pigs are set to a missing value if the origin of the allele corresponds to the other maternal line, and effectively does not contribute to the breed-specific partial relationship matrix.

Assigning breed-of-origin to alleles in crossbreds

To infer the breed-of-origin of the alleles in CB pigs, we used the BOA approach that was developed by Vandenplas et al. [31]. It consists of three steps: (1) phasing the haplotypes of both PB and CB pigs with AlphaPhase1.1 software [33], (2) determining the unique haplotypes among the PB, and (3) assigning the breed-of-origin for each allele carried on the haplotypes of CB. This approach was applied to the same dataset in a previous study [31]. On average, 95.2% of the alleles of the three-way CB pigs were assigned a breed-of-origin. These alleles with their assigned breed-of-origin were used to build the breed-specific partial relationship matrices. Alleles that were not assigned a breed-of-origin were set to missing, and effectively did not contribute to any of the breed-specific partial relationship matrices.

GBLUP model with the genomic relationship matrix

For comparison to the BOA model, the following 4-trait animal model was fitted (G model):

$$\mathbf{y_S} = \mathbf{X_S b_S} + \mathbf{W_S u_S} + \mathbf{Z_S a_S} + \mathbf{e_S},$$

$$\mathbf{y_{LR}} = \mathbf{X_{LR} b_{LR}} + \mathbf{W_{LR} u_{LR}} + \mathbf{Z_{LR} a_{LR}} + \mathbf{e_{LR}},$$

$$\mathbf{y_{LW}} = \mathbf{X_{LW} b_{LW}} + \mathbf{W_{LW} u_{LW}} + \mathbf{Z_{LW} a_{LW}} + \mathbf{e_{LW}},$$

$$\mathbf{y_{CB}} = \mathbf{X_{CB} b_{CB}} + \mathbf{W_{CB} u_{CB}} + \mathbf{Z_{CB} a_{CB}} + \mathbf{e_{CB}},$$

where vectors and matrices are defined as in the BOA model, with the only difference being that the additive genetic effect in CB pigs was defined only by one vector, $\mathbf{a_{CB}}$. Therefore, the variance–covariance matrix of genetic effects was:

$$\mathrm{Var}\begin{bmatrix} \mathbf{a_S} \\ \mathbf{a_{LR}} \\ \mathbf{a_{LW}} \\ \mathbf{a_{CB}} \end{bmatrix} = \begin{bmatrix} \sigma_{a_S}^2 & \sigma_{a_S,a_{LR}} & \sigma_{a_S,a_{LW}} & \sigma_{a_S,a_{CB}} \\ \sigma_{a_S,a_{LR}} & \sigma_{a_{LR}}^2 & \sigma_{a_{LR},a_{LW}} & \sigma_{a_{LR},a_{CB}} \\ \sigma_{a_S,a_{LW}} & \sigma_{a_{LR},a_{LW}} & \sigma_{a_{LW}}^2 & \sigma_{a_{LW},a_{CB}} \\ \sigma_{a_S,a_{CB}} & \sigma_{a_{LR},a_{CB}} & \sigma_{a_{LW},a_{CB}} & \sigma_{a_{CB}}^2 \end{bmatrix} \otimes \mathbf{G}.$$

This model was implemented using two different genomic relationship matrices ($\mathbf{G}$) as explained in the next sections.

Genomic relationship matrix using allele frequencies across all genotyped pigs ($\mathbf{G_A}$ matrix)

The $\mathbf{G_A}$ matrix was constructed using the second method in VanRaden [20]:

$$\mathbf{G_A} = \left(\mathbf{M} - 2\mathbf{1p'}\right)\mathbf{D}\left(\mathbf{M} - 2\mathbf{1p'}\right)'/N,$$

where $\mathbf{M}$ is a matrix containing SNP genotypes for each pig (coded as 0, 1, or 2), $\mathbf{p}$ is the vector of the frequencies of the counted allele (p_j), calculated across the genotyped population, $\mathbf{D}$ is diagonal with $D_{jj} = \frac{1}{p_j(1 - p_j)}$, and N is the number of SNPs.

Genomic relationship matrix using breed-specific allele frequencies ($\mathbf{G_B}$ matrix)

To account for population structure, we also used a genomic relationship matrix based on genotypes centered and scaled by breed-specific allele frequencies ($\mathbf{G_B}$):

$$\mathbf{G_B} = \left(\mathbf{M} - 2\mathbf{1p'_B}\right)\mathbf{D^B}\left(\mathbf{M} - 2\mathbf{1p'_B}\right)'/N,$$

where each $\mathbf{p_B}$ is the vector of the frequencies of the counted allele (p_{Bj}). p_{Bj} was obtained by summing the contribution of each pure breed j and the weighted contribution of the CB. The weight was 0.5 for S, and 0.25 for LR and LW. $\mathbf{D^B}$ is diagonal with $D_{jj}^B = \frac{1}{p(1 - p_{Bj})}$.

Estimation of variance components and BLUP

Implementation of the aforementioned GBLUP models required estimates for all variance components involved. Variance components were estimated for each of the three models using the ASReml software [34]. Instead of one 4-trait multivariate model, three bivariate models

Sevillano *et al. Genet Sel Evol* (2017) 49:75

Page 5 of 14

were fitted to overcome workspace memory limitation of the software. Each analysis included PB of one of the three breeds and all CB. As a consequence, genetic co-variances between breeds were not estimated. For the BOA model, these genetic co-variances are not considered and thus are effectively equal to 0. For the other two models, we also assumed that these co-variances were not significant, and therefore, we set them to 0 in the subsequent BLUP analyses. Variance components of the bivariate models were combined to obtain the full variance–covariance matrices for the 4-trait model. The variance–covariance matrices were combined by averaging the three CB variance components estimated in each of the bivariate models. If necessary, the combined variance–covariance matrices were bended to make them positive definite [35]. Bending changed the variance–covariance components on average by 7.5% (0.3 to 18.5%). BLUP for the three models were obtained using the MiXBLUP software [36].

Cross-validation

The accuracy of EBV of PB pigs for CB performance from the three models was evaluated as the average accuracy obtained from fourfold cross-validation. Because of different degrees of relationship between PB and CB, genotyped S, LR, or LW pigs were first divided into four mutually exclusive clusters, using the K-means clustering method applied to a dissimilarity matrix computed from elements of the $\mathbf{G_A}$ matrix [37]. Then, each CB pig was assigned to the PB cluster with the closest relationship based on the $\mathbf{G_A}$ matrix. For the maternal breed LW, the CB pigs were not very evenly distributed across the clusters, with one cluster including most of the CB. Therefore, for this breed, the cluster with the largest number of CB pigs was randomly split into four groups and each of those groups was joined with one of the other clusters.

In each training analysis, the data excluded PB and CB pigs from one fold to train on the remaining three folds to predict EBV for CB performance of the excluded PB pigs (validation set). This resulted in every PB pig having EBV for CB performance that were obtained without using performance of the most closely-related CB pigs for training. Thus, the information coming from the most closely-related CB pigs could be used for validation. The number of pigs in the validation and training sets for each of the folds of the cross-validation and for each trait are in Tables 3, 4 and 5 for S, LR, and LW, respectively.

Validation set

The PB pigs cannot have an own performance for CB performance, and also in our data, they do not have

Table 3 Cross-validation strategy for crossbred performance of Synthetic boar (S)

Fold	Training		Validation		
	S	CB	S	CB	CB-extra*
ADG					
1	2115	1535	460	140	199
2	2119	1341	456	334	268
3	1895	605	680	1070	297
4	1596	1544	979	131	145
BF					
1	2132	1536	484	140	188
2	2144	1344	472	332	246
3	1932	604	684	1072	289
4	1640	1544	976	132	145
LD					
1	2128	1541	467	140	200
2	2132	1348	463	333	272
3	1921	605	674	1076	299
4	1604	1549	991	132	145

Numbers of individuals for Synthetic boar (S), three-way crossbred (CB) and extra three-way crossbred pigs (CB-extra) in the training and validation sets per trait, i.e. average daily gain (ADG), back fat thickness (BF), and loin depth (LD)

* Three-way crossbred pigs with only phenotypic information, and no genotyping

Table 4 Cross-validation strategy for crossbred performance of Landrace (LR)

Groups	Training		Validation		
	LR	CB	LR	CB	CB-extra*
ADG					
1	1584	1564	748	111	456
2	1825	1523	507	152	465
3	1762	1531	570	144	456
4	1825	407	507	1268	471
BF					
1	2829	1565	775	111	463
2	2492	1523	1112	153	472
3	3002	1532	602	144	463
4	2489	408	1115	1268	478
LD					
1	1631	1570	754	111	463
2	1891	1528	494	153	472
3	1823	1537	562	144	463
4	1810	408	575	1273	478

Numbers of individuals for Landrace (LR), three-way crossbred (CB), and extra three-way crossbred pigs (CB-extra) in the training and validation sets per trait, i.e. average daily gain (ADG), back fat thickness (BF), and loin depth (LD)

* Three-way crossbred pigs with only phenotypic information, no genotyped

Sevillano *et al. Genet Sel Evol* (2017) 49:75

Page 6 of 14

Table 5 Cross-validation strategy for crossbred performance of Large White (LW)

Groups	Training		Validation		
	LR	CB	LR	CB	CB-extra*
ADG					
1	3628	1193	1666	482	468
2	3612	1269	1682	406	468
3	4008	1111	1286	564	468
4	4634	1452	660	223	468
BF					
1	4870	1191	1899	485	475
2	4954	1271	1815	405	475
3	4381	1113	2388	563	475
4	6102	1453	667	223	475
LD					
1	3759	1196	1710	485	475
2	3678	1275	1791	406	475
3	4162	1114	1307	567	475
4	4808	1458	661	223	475

Numbers of individuals for Large White (LW), three-way crossbred (CB), and extra three-way crossbred pigs (CB-extra) in the training and validation sets per trait, i.e. average daily gain (ADG), back fat thickness (BF), and loin depth (LD)

* Three-way crossbred pigs with only phenotypic information, no genotyped

large offspring groups, which would allow to compute a phenotype as average offspring performance. Therefore, we calculated deregressed proofs (DRP) for PB pigs within the validation sets to validate the predictions of our models. For this, first we obtained EBV from the G model with a pedigree-based relationship matrix. This resulted in an EBV for CB performance for each PB pig. The EBV were estimated based on performance of the CB pigs assigned to each of the validation folds (Tables 3, 4, and 5 for S, LR, and LW, respectively). Phenotype information was also available for an additional 501 CB pigs (CB-extra) that were not genotyped. These records were used in each of the four validation folds (Tables 3, 4, 5 for S, LR, and LW, respectively). Within each validation fold, the EBV of PB pigs for CB performance were then deregressed according to Calus et al. [38]. The deregression involved removal of all effects of relatives in the same validation set, and correction for regression to the mean, to obtain a more accurate estimate of the expected phenotype. In addition, a weighting factor (*w*) was estimated for each DRP value based on the reliability of the calculated DRP. These *w* are the effective record contributions [39], and reflect the amount of information in the DRP contributed by the animal itself, correcting for any information of the relatives that contributed to its EBV before deregression.
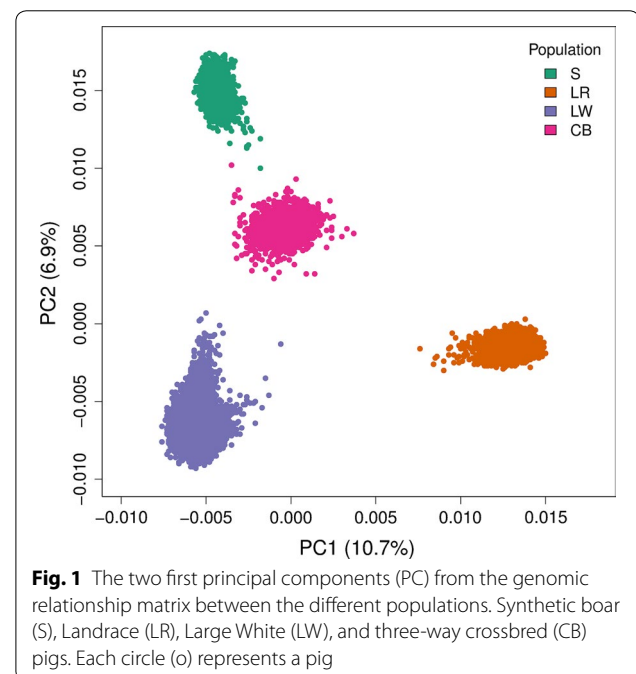
Predictive ability

Accuracies of the BOA and G models were calculated as the weighted correlation between the DRP and the EBV of PB pigs for CB performance, where the weighting factor *w* was used to account for differences in the amount of available information on relatives to estimate DRP. The standard error (SE) of the correlations were approximated as $(1 - r^2)/\sqrt{N}$, were r is the accuracy of the model, and N is the number of validation animals [40].

Results

Genotyped population and relationship matrices

The three breeds, S, LR, and LW, were clearly different populations as shown in Fig. 1 based on the first two principal components of the $\mathbf{G_A}$ matrix. The CB population appeared intermediate among the PB populations. The divergence among the three populations estimated with Weir and Cockerham's F_{ST} [41], were equal to 0.17 between S and LR, 0.12 between S and LW, and 0.14 between LW and LR, which indicated that they are distantly-related breeds.

The relationships between breeds, calculated with the $\mathbf{G_A}$ matrix were mainly negative (Table 6), with average relationships between breeds ranging from − 0.13 to − 0.07. When using the $\mathbf{G_B}$ matrix, the average relationships between all breeds are zero by definition. When using breed-specific partial relationship matrices ($\mathbf{G^{(S)}}$, $\mathbf{G^{(LR)}}$ and $\mathbf{G^{(LW)}}$), only the relationships based on common alleles originating from the same breed were considered and, consequently no relationships were estimated

Fig. 1 The two first principal components (PC) from the genomic relationship matrix between the different populations. Synthetic boar (S), Landrace (LR), Large White (LW), and three-way crossbred (CB) pigs. Each circle (o) represents a pig

Sevillano *et al. Genet Sel Evol* (2017) 49:75

Page 7 of 14

Table 6 Descriptive statistics for relationship between populations based on different genomic relationship matrices

Relationship between	Matrix[a]	Mean	Median	Min	Max	SD
S-LR	G_A	− 0.13	− 0.13	− 0.22	0.00	0.02
	G_B	0.00	0.00	− 0.09	0.09	0.02
S-LW	G_A	− 0.07	− 0.07	− 0.18	0.12	0.02
	G_B	0.00	0.00	− 0.11	0.11	0.02
LR-LW	G_A	− 0.13	− 0.13	− 0.23	0.16	0.02
	G_B	0.00	0.00	− 0.13	0.23	0.02
CB (diagonal)	$G^{(S)}$	0.49	0.49	0.40	0.80	0.04
	$G^{(LR)}$	0.23	0.23	0.02	0.40	0.04
	$G^{(LW)}$	0.23	0.23	0.07	0.39	0.04
	G_A	0.96	0.95	0.88	1.07	0.03
	G_B	0.94	0.93	0.86	1.08	0.03

[a] $G^{(S)}$ = partial relationship matrix for breed Synthetic boar (S); $G^{(LR)}$ = partial relationship matrix for breed Landrace (LR); $G^{(LW)}$ = partial relationship matrix for breed Large White (LW); G_A = genomic relationship matrix by allele frequencies obtained across the genotyped population; G_B = genomic relationship matrix by breed-specific allele frequencies

between breeds. For CB pigs, the diagonal elements of the G_A and G_B matrices had an average of 0.96 and 0.94, respectively. For the $G^{(S)}$, $G^{(LR)}$ and $G^{(LW)}$ matrices, as they are partial relationship matrices, the diagonal elements for CB pigs had averages of 0.49, 0.32, and 0.30 for $G^{(S)}$, $G^{(LR)}$ and $G^{(LW)}$, respectively. These averages are close to the expected values, i.e. 0.50 for the S breed and 0.25 for the LR and LW breeds.

Variance components, heritabilities, and genetic correlations

Estimated variance components for ADG, BF, and LD using the BOA model with the $G^{(S)}$, $G^{(LR)}$ and $G^{(LW)}$ matrices, the G model with the G_A matrix (G_A model), and the G model with G_B matrix (G_B model) are in Table 7. The standard errors of the estimated variance components in Table 7 are provided in Additional file 1: Table S1. Regardless of the model and trait, the PB additive genetic variance estimated for the maternal breeds, i.e. LR and LW, were very similar. For the maternal breeds, CB additive genetic variance was larger than PB additive genetic variance for all traits. For the paternal breed, the opposite was observed, i.e. CB additive genetic variance was smaller than PB additive genetic variance, for all traits except BF. Estimates of CB heritability tended to be higher than estimates of PB heritability for all traits except LD.

A comparison between models showed that PB and CB additive genetic variances for the maternal breeds were similar between the G_A and G_B models. For the paternal breed S, compared to the G_B model, the G_A model estimated a larger PB additive genetic variance, and smaller CB additive genetic variance. Estimated PB additive genetic variances with the BOA model were similar to

those obtained with the G_A or G_B models and the estimated CB additive genetic variances with the BOA model, on average across the three breeds, were larger than those obtained with the G_A or G_B models. The estimates of PB and CB heritability were similar across models, while estimates obtained with the BOA model tended to be slightly lower and those with the G_B model tended to be slightly higher than with the G_A model. The genetic correlations for traits between PB and CB pigs estimated with the BOA model were generally similar to those of the G_A and G_B models, except for the genetic correlation between LR and CB pigs for ADG that was much higher than that estimated with the G_A and G_B models. The genetic correlations between PB and CB pigs estimated with the G_A or G_B models were similar. In general, the SE of PB additive genetic variances and heritabilities were similar across models, although the SE of the three CB additive genetic variances estimated with the BOA model were much larger than the SE of the single CB additive genetic variance estimated with the G_A or G_B models. The SE of the estimated genetic correlations were relatively large, ranging from 0.10 to 0.29, across all models and traits.

For the BOA model, the CB variance for litter effect was about three times larger than that obtained with the G_A or G_B models. Estimates of the CB residual variance were also slightly larger when using the BOA model compared to the G_A and G_B models. Estimates of PB variance for litter and residual effects by the G_A and G_B models were similar among breeds. Estimates of CB variance for litter and residual effects by the G_A and G_B models were similar among the maternal breeds, while for breed S, the CB variance for litter and residual effects was lower with the G_A model than with the G_B model. In summary,

Sevillano *et al. Genet Sel Evol* (2017) 49:75

Page 8 of 14

Table 7 Additive genetic variance (σ_a^2), litter variance (σ_u^2), residual variance (σ_e^2), and heritabilities for each breed for PB and CB performance, and genetic correlation between purebred and CB pigs (r_{PC}), estimated for each trait using the BOA[a], G_A^b, and G_B^c models

Model	Breed	$\sigma_{a_{PB}}^2$	$\sigma_{u_{PB}}^2$	$\sigma_{e_{PB}}^2$	h_{PB}^2	$\sigma_{a_{CB}}^2$	$\sigma_{u_{CB}}^2$ *	$\sigma_{e_{CB}}^2$ *	h_{CB}^2	r_{pc}
ADG										
BOA	S	2699	2925	6124	0.23	2316	853	4192	0.34**	0.50
	LR	2165	2291	3778	0.26	3566				0.62
	LW	2123	1595	4602	0.26	2258				0.57
G_A	S	3386	2850	6068	0.28	2053*	258	3576	0.35	0.52
	LR	2461	2282	3718	0.29					0.31
	LW	2336	1563	4595	0.28					0.61
G_B	S	2775	2846	6082	0.24	2261*	262	3592	0.37	0.52
	LR	2248	2287	3703	0.27					0.30
	LW	2154	1640	4568	0.26					0.59
BF										
BOA	S	0.82	0.55	1.27	0.31	1.90	0.88	3.96	0.38**	0.74
	LR	1.09	0.60	1.73	0.32	3.74				0.67
	LW	1.33	0.86	1.67	0.34	4.16				0.58
G_A	S	1.18	0.55	1.26	0.40	2.18*	0.33	3.32	0.37	0.73
	LR	1.38	0.59	1.71	0.38					0.72
	LW	1.57	0.85	1.64	0.39					0.65
G_B	S	0.98	0.54	1.26	0.35	2.40*	0.34	3.34	0.39	0.69
	LR	1.26	0.59	1.70	0.35					0.70
	LW	1.44	0.85	1.64	0.37					0.62
LD										
BOA	S	10.59	6.00	8.43	0.42	11.59	3.20	31.45	0.24**	0.53
	LR	5.72	3.00	6.65	0.37	7.23				0.58
	LW	6.04	3.55	6.93	0.37	12.86				0.53
G_A	S	12.78	5.93	8.41	0.47	9.05*	0.11	28.89	0.24	0.57
	LR	6.58	2.98	6.60	0.41					0.57
	LW	6.82	3.56	6.89	0.40					0.68
G_B	S	10.58	5.87	8.33	0.43	10.00*	0.05	28.89	0.26	0.55
	LR	5.82	2.97	6.57	0.38					0.56
	LW	6.09	3.55	6.86	0.37					0.62

S Synthetic boar, *LR* Landrace, *LW* Large White, *CB* three-way crossbred pigs

ADG average daily gain, *BF* back fat thickness, *LD* loin depth

[a] BOA model, model with breed-specific relationship matrices

[b] G_A model, model with genomic relationship matrix by allele frequencies obtained across the genotyped population

[c] G_B model, model with genomic relationship matrix by breed-specific allele frequencies

* Average from the three bivariate models

** $(0.5\sigma_{a_S}^2 + 0.25\sigma_{a_{LR}}^2 + 0.25\sigma_{a_{LW}}^2)/(0.5\sigma_{a_S}^2 + 0.25\sigma_{a_{LR}}^2 + 0.25\sigma_{a_{LW}}^2 + \sigma_{u_{CB}}^2 * + \sigma_{e_{CB}}^2 *)$

estimated variance components were mostly similar across models, apart from the CB litter variance that was considerably larger with the BOA model compared to the other two models.
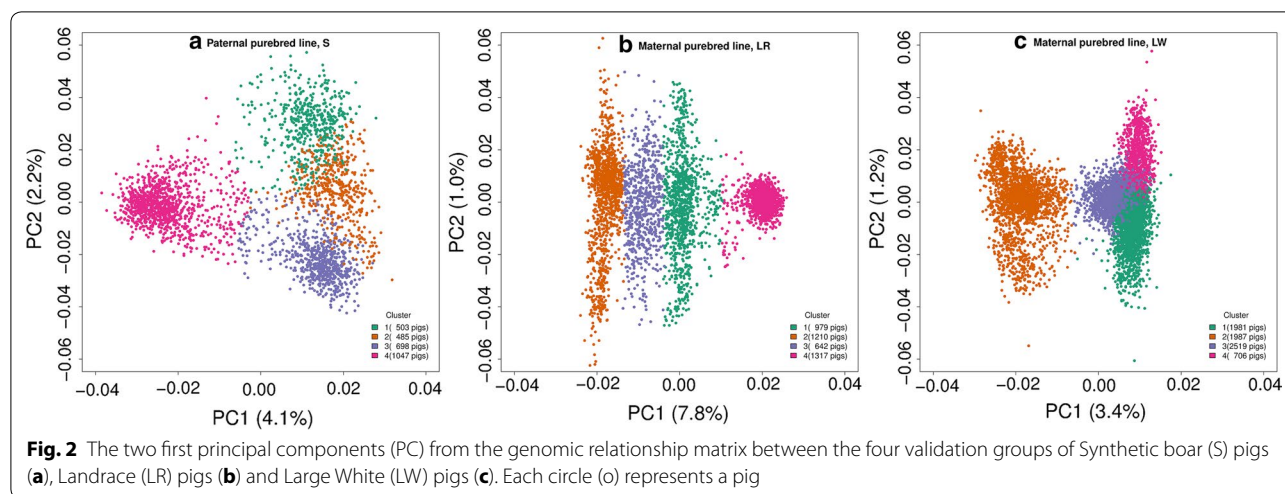
Predicting breeding values of PB pigs for CB performance with different models

For each breed S, LR, and LW, four validation groups were formed to perform the 4-fold cross-validation.

Figure 2 represents the first two principal components from the **G_A** matrix and shows that the grouping for the cross-validation was done correctly. The first two principal components explained 6.3% of the variability among S pigs, 8.8% among LR pigs and 4.65% among LW pigs.

Accuracies of the three models for the estimated breeding values of S pigs for CB performance are in Table 8. For ADG, the BOA model yielded slightly better accuracies than the G_A and G_B models. The opposite was

Sevillano *et al. Genet Sel Evol* (2017) 49:75

Page 9 of 14

Fig. 2 The two first principal components (PC) from the genomic relationship matrix between the four validation groups of Synthetic boar (S) pigs (**a**), Landrace (LR) pigs (**b**) and Large White (LW) pigs (**c**). Each circle (o) represents a pig

Table 8 Accuracies* of BOA[a], G_A^b and G_B^c models calculated for each of the four folds of cross-validation for estimating breeding values of the paternal breed Synthetic boar pigs for crossbred performance for each trait, and average weighting factor (*w*) of the calculated DRP per validation fold

Folds	w	BOA	G_A	G_B
ADG				
1	0.49	0.055	0.055	0.057
2	0.12	0.128	0.111	0.094
3	0.21	0.170	0.156	0.152
4	0.07	0.063	0.084	0.082
Mean		0.104	0.102	0.096
BF				
1	0.31	0.168	0.168	0.162
2	0.39	0.201	0.157	0.159
3	0.52	0.191	0.294	0.280
4	0.25	0.150	0.179	0.177
Mean		0.178	0.199	0.195
LD				
1	0.55	0.204	0.234	0.236
2	0.67	0.212	0.209	0.207
3	0.88	0.127	0.140	0.134
4	0.45	0.088	0.135	0.142
Mean		0.158	0.179	0.180

ADG average daily gain, *BF* back fat thickness, *LD* loin depth

* Accuracies measured as weighted correlation between DRP and EBV of S pigs for crossbred performance. Approximate standard errors SE, computed as $(1-r^2)/N \sqrt{(1-r^2)/N}$, were equal to 0.023 to 0.024 for the mean accuracies across the folds, for all combinations of traits and methods.

[a] BOA model, model with breed-specific relationship matrices

[b] G_A model, model with genomic relationship matrix by allele frequencies obtained across the genotyped population

[c] $\mathbf{G_B}$ model, model with genomic relationship matrix by breed-specific allele frequencies

observed for BF and LD, where the G_A and G_B models yielded slightly better accuracies than the BOA model. Accuracies of the three models for the estimated breeding values of LR pigs for CB performance are in Table 9. For ADG, the BOA model yielded higher accuracies than the G_A and G_B models. For BF and LD, there was no difference in accuracies between the three models. Accuracies of the three models for the estimated breeding values of LW pigs for CB performance are in Table 10. The trait ADG is not included, because the reliabilities of the EBV of LW pigs within the validation groups for CB performance for this trait were too low to be used for proper validation. Similar to the results for the LR breed, there was no difference in accuracies between the three models for the traits BF and LD. In general, accuracies from models G_A and G_B were similar.

Discussion

Properties of the relationship matrices

Genomic relationships within and across populations are defined differently depending on how the genetic covariance between individuals is calculated. Using across-breed allele frequencies when the correlations of allele frequencies between breeds differ from 1, could lead to genomic relationships between animals of different breeds that are on average negative [26], as observed for the $\mathbf{G_A}$ matrix. This was not the case for the $\mathbf{G_B}$ matrix, in which the genomic relationships between animals of different breeds was on average 0, as expected for distantly-related breeds.

Diagonal elements (D) from a pedigree-based relationship matrix have a value of 1 when there is no inbreeding. Because a genomic relationship matrix is built to resemble a pedigree-based relationship matrix and the current genotyped population is considered the base population

Sevillano *et al. Genet Sel Evol* (2017) 49:75

Page 10 of 14

Table 9 Accuracies* of BOA[a], G$_A$[b], and G$_B$[c] models calculated for each of the four folds of cross-validation for estimating breeding values of the maternal breed Landrace pigs for crossbred performance for each trait, and weighting factor (*w*) of the calculated DRP

Folds	*w*	BOA	G$_A$	G$_B$
ADG				
1	0.20	0.133	0.106	0.099
2	0.23	0.190	0.095	0.111
3	0.21	0.159	0.106	0.106
4	0.22	0.094	0.007	0.014
Mean		0.144	0.079	0.083
BF				
1	0.09	0.185	0.169	0.171
2	0.07	0.186	0.210	0.199
3	0.10	0.223	0.216	0.215
4	0.09	0.144	0.149	0.141
Mean		0.184	0.186	0.181
LD				
1	0.43	0.224	0.206	0.203
2	0.47	0.085	0.107	0.107
3	0.45	0.239	0.232	0.228
4	0.47	0.170	0.208	0.207
Mean		0.179	0.188	0.186

ADG average daily gain, *BF* back fat thickness, *LD* loin depth

* Accuracies measured as weighted correlation between DRP and EBV of LR pigs for crossbred performance. Approximate standard errors SE, computed as $(1-r^2)/N \sqrt{(1-r^2)/N}$, were equal to 0.023 to 0.024 for the mean accuracies across the folds, for all combinations of traits and methods.

[a] BOA model, model with breed-specific relationship matrices

[b] G$_A$ model, model with genomic relationship matrix by allele frequencies obtained across the genotyped population

[c] G$_B$ model, model with genomic relationship matrix by breed-specific allele frequencies

Table 10 Accuracies* of BOA[a], G$_A$[b], and G$_B$[c] models calculated for each of the four folds of cross-validation for estimating breeding values of the maternal breed Large White pigs for crossbred performance for each trait, and weighting factor (w) of the calculated DRP

Folds	*w*	BOA	G$_A$	G$_B$
BF				
1	0.21	0.217	0.221	0.216
2	0.13	0.095	0.094	0.089
3	0.28	0.190	0.175	0.170
4	0.23	0.219	0.242	0.243
Mean		0.180	0.183	0.180
LD				
1	0.62	0.235	0.234	0.232
2	0.38	0.103	0.126	0.126
3	0.74	0.226	0.229	0.228
4	0.64	0.297	0.318	0.318
Mean		0.215	0.227	0.226

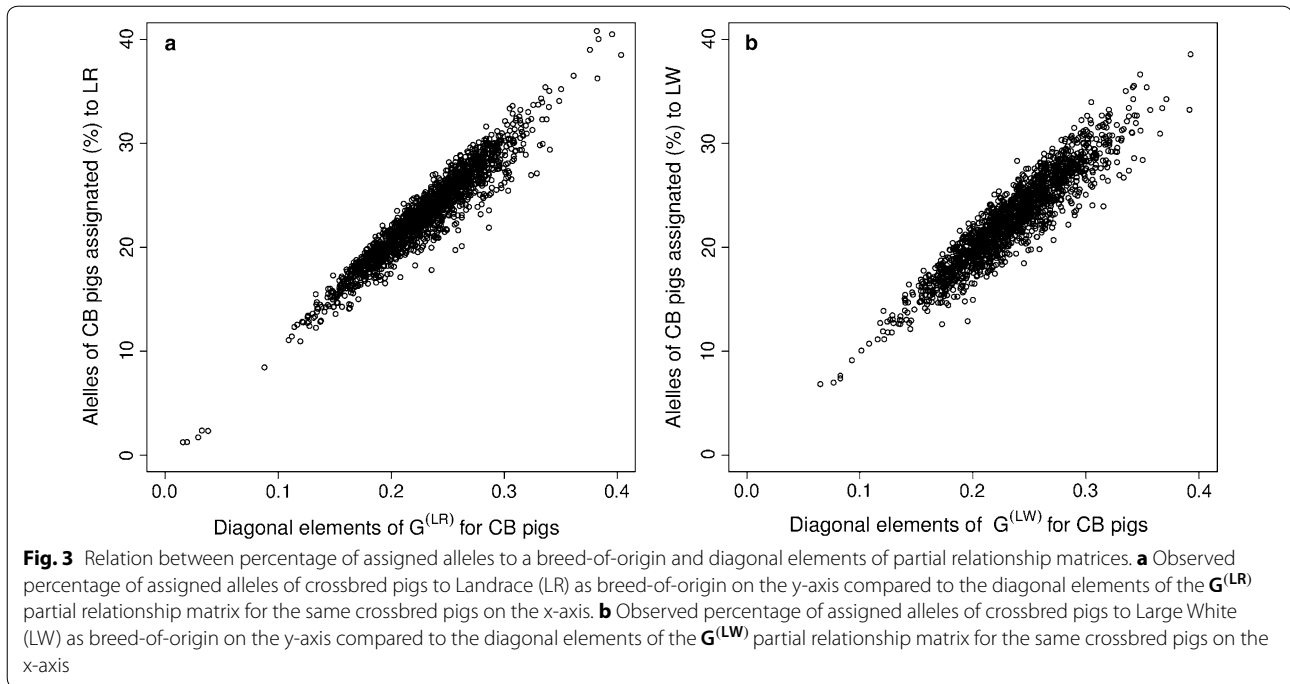*ADG* average daily gain, *BF* back fat thickness, *LD* loin depth

* Accuracies measured as weighted correlation between DRP and EBV of LR pigs for crossbred performance. Approximate standard errors SE, computed as $(1-r^2)/N \sqrt{(1-r^2)/N}$, were equal to 0.023 to 0.024 for the mean accuracies across the folds, for all combinations of traits and methods.

[a] BOA model, model with breed-specific relationship matrices

[b] G$_A$ model, model with genomic relationship matrix by allele frequencies obtained across the genotyped population

[c] G$_B$ model, model with genomic relationship matrix by breed-specific allele frequencies

[20], the average D from a genomic relationship matrix is expected to be 1, as we observed for the **G$_A$** and **G$_B$** matrices. To calculate the partial relationship matrices, **G$^{(S)}$**, **G$^{(LR)}$** and **G$^{(LW)}$**, the D for CB pigs were expected to be 0.5 for **G$^{(S)}$**, and 0.25 for **G$^{(LR)}$** and **G$^{(LW)}$**, expressing the proportion of the genome in CB pigs contributed by each breed S, LR, and LW, respectively. Using all 52,164 SNPs, Fig. 3 shows how the diagonal elements among CB pigs from the **G$^{(LR)}$**, and **G$^{(LW)}$** matrices increased as the percentage of alleles of CB pigs assigned to the respective maternal breed as breed-of-origin increased.

Variance components across models

Estimating variance components for the 4-trait multivariate models was not possible due to workspace memory limitation when trying to run the full BOA model with the three partial relationship matrices or the G models with the relationship matrices containing the four populations.

Therefore, for the G models, the construction of a full variance–covariance matrix based on sub-models was required, in this case three bivariate models. This procedure of constructing a full variance–covariance matrix is often used in genetic evaluation [35]. The combined variance–covariance matrices in the G$_A$ and G$_B$ models for BF were considerably bended (variance components changed up to 10.9%) and this may have affected the results. The combined variance–covariance matrix in the G$_A$ model for LD was also bended, but, in this case, the components changed only up to 2.5%. For ADG, no bending of the variance–covariance matrix was required for any of the models. An advantage of the BOA model, since variance–covariance matrices are by breed, is that it allows the estimates of the CB additive genetic variance contributed by the different parental breeds to differ. With the G$_A$ and G$_B$ models, these differences cannot be observed because there is only one estimate for CB additive genetic variance across the three breeds. A disadvantage of the BOA model is that estimates must be based on half the information (for the paternal breed) or on a quarter of the information (for the maternal breeds) compared to estimates from the G$_A$ or G$_B$ models. Therefore, the SE of CB additive genetic variances estimated with the BOA model were

Sevillano *et al. Genet Sel Evol* (2017) 49:75

Page 11 of 14

Fig. 3 Relation between percentage of assigned alleles to a breed-of-origin and diagonal elements of partial relationship matrices. **a** Observed percentage of assigned alleles of crossbred pigs to Landrace (LR) as breed-of-origin on the y-axis compared to the diagonal elements of the $\mathbf{G}^{(LR)}$ partial relationship matrix for the same crossbred pigs on the x-axis. **b** Observed percentage of assigned alleles of crossbred pigs to Large White (LW) as breed-of-origin on the y-axis compared to the diagonal elements of the $\mathbf{G}^{(LW)}$ partial relationship matrix for the same crossbred pigs on the x-axis

much larger than the SE of CB additive genetic variances estimated with the G_A and G_B models. With the BOA model, we could observe that estimates of CB additive genetic variance differed between the three breeds for all traits. This means that r_{pc} should also be interpreted separately by breed. The estimates of r_{pc} differed slightly across models. In theory, the CB additive variance components estimated with the BOA model comprises the variation observed in CB pigs due only to the alleles coming from the analyzed breed. Therefore, differences in r_{pc} estimated with the G_A or G_B model rather than the BOA model were expected. For instance, for ADG, the r_{pc} estimated with the BOA model for S and LW were slightly smaller than those estimated with the G_A and G_B models. However, the r_{pc} estimated with the BOA model for LR was twice as high compared to that of the other two models. One explanation is that a large part of the CB additive variance can come mainly from variation observed among the alleles originating from a specific breed and this is not captured when all alleles are assumed to have the same origin.

In the literature, r_{pc} for production traits have been calculated from pedigree information only [6, 42] and vary greatly, but on average they are higher than our estimates, probably because the breeds were different or the estimates were an average across different breeds. In general, the investigated traits showed a moderate r_{pc} indicating that using CB information together with PB information in the reference population might be beneficial for selection of PB pigs for CB performance. Using

CB information is expected to be most important for combinations of trait and breed for which r_{pc} is low, for instance for ADG in breed LR.

From the estimates of the BOA model, we observed that CB litter effect and residual variance were much larger than those obtained with the G_A or G_B models. Because the genotypes of only one breed at a time were used in the bivariate BOA model, the litter and residual effect variance in the BOA model likely absorbed the variance coming from the genetic relationships from the breeds that were absent in the model. To investigate the impact of these possibly inflated litter and residual variances, we tried to correct this by setting the CB litter effect and residual variance of the BOA model equal to the average estimates from the G_A and G_B models. Using these new variance estimates did not affect the accuracies of the BOA model compared to the G_A and G_B models (results not shown).

Predictive ability across models

The three breeds used in this study are distantly related and correlations between breed-specific allele frequencies were low: 0.31 for breeds S and LR, 0.54 for breeds S and LW, and 0.39 for breeds LR and LW. However, taking population structure into account by accounting for different allele frequencies in the three different breeds (G_B model) did not improve the accuracy for predicting EBV compared with using allele frequencies obtained across genotyped populations (G_A model). In a study with CB

Sevillano *et al. Genet Sel Evol* (2017) 49:75

Page 12 of 14

sheep, Moghaddar et al. [25] reported limited impact on prediction accuracy when adjusting for breed-specific allele frequency, also when differences in allele frequencies between breeds were large. Makgahlela et al. [24] and Lourenco et al. [26] also observed no advantage of using breed-specific allele frequencies for constructing the relationship matrix, even when this led to observable changes in the coefficients of the relationship matrix. Although correlations between breed-specific allele frequencies were low, correlations between these breed-specific allele frequencies and the across-breed frequency were relatively high, simply because the breed-specific allele frequencies are included in the across-breed allele frequency. In our study, the correlations between the breed-specific allele frequencies and the across-breed frequency were equal to 0.74, 0.68, and 0.89, for breeds S, LR, and LW, respectively. The correlation between breed LW allele frequency and the across-breed frequency was higher than the others, because the LW breed has the largest number of pigs (Table 1), therefore, it has a larger contribution to the across-breed allele frequencies across breeds. The correlation between crossbred allele frequency and the across-breed frequency was equal to 0.93. Therefore, using breed-specific or across-breed frequencies in the calculation of the relationship coefficient between a PB and CB pig will have little effect on predicted EBV of PB for CB performance.

In the G_A and G_B models, genetic co-variances between breeds were assumed to be zero. To test if this was a correct assumption, covariances between PB lines were also estimated by fitting three additional bivariate models (one for each pair of PB) for the trait ADG using the G_A model. Variance components of the six bivariate models were combined to obtain the full variance–covariance matrices for the 4-trait model. This combination was performed by averaging the three variance components estimated for each population, i.e. S, LR, LW and CB. In this case, it was not necessary to bend the combined variance–covariance matrix to make it positive definite. The genetic correlations between PB performance for ADG were 0.13 ($\pm$ 0.24) between S and LR, 0.39 ($\pm$ 0.14) between S and LW, and 0.36 ($\pm$ 0.16) between LR and LW. These estimates were in line with estimated values of 0.23 and 0.30 between a Danish Landrace and Danish Yorkshire population [43]. Moreover, for breeds S and LR, the value of zero was within one SD. Accuracies of the G_A model taking into account the covariance between PB for estimating breeding values of S pigs for CB performance, were similar to prediction accuracy of the G_A model assuming the covariances between PB to be zero (Table 11). This was expected because relationships between pigs from different breeds were low and showed very little variation (Table 6). Therefore, the G_A

Table 11 Accuracies* of the $G_A{}^a$ model assuming zero covariance between purebreds and with covariances calculated between purebreds ($G_{A-covariancePB}$), for each of the four folds of cross-validation for estimating breeding values of the paternal breed Synthetic boar pigs for crossbred performance for average daily gain (ADG)

Folds	$G_A{}^a$	$G_{A-covariancePB}$
1	0.055	0.054
2	0.111	0.110
3	0.156	0.171
4	0.084	0.085
Mean	0.102	0.105

* Accuracies measured as weighted correlation between DRP and EBV of S pigs for crossbred performance

a G_A model, model with genomic relationship matrix by allele frequencies obtained across the genotyped population. Results are the same as in Table 8

model assuming the covariances between PB to be zero are not expected to affect accuracies, even when genetic correlations between PB are moderate.

The BOA model assumes that relationships between PB are zero, and thus also effectively assumes that the covariances between PB are zero. A study from Xiang et al. [43] compares the BOA approach in a single-step model against a single-step model with metafounders, where the last model defines relationships between the pedigree base populations across breeds but also takes genomic relationships across breeds into account. Taken together their conclusions that both models perform similarly and our findings, these results suggest that considering genomic relationships and covariances between PB lines has limited relevance in models for predicting crossbred performance for pig crossbreeding programs.

Compared to the G_A and G_B models, taking population structure into account by using breed-specific partial relationships as in the BOA model, including breed-specific allele frequencies, had some impact on the accuracy of EBV. The BOA model had a positive impact for traits with a low r_{pc} as for ADG in breed LR (0.30). BF and LD showed higher r_{pc} (0.55 to 0.73), and accuracies of the BOA model for these traits was similar to those of the G_A or G_B models. Comparing PB lines, somewhat higher accuracies could have been expected for the S line, because the sire line contributes 50% of the genome of the CB, while the dam lines contribute only 25%. Thus, the sire line will have a larger variance in genomic relationships with the CB pigs used for training, which is expected to yield higher accuracies [15]. Nevertheless, in our study, accuracies were very comparable across the sire and dam lines. The BOA model was previously tested on simulated data [15, 27], and on real data but for a two-breed cross scheme [28, 29]. These studies also compared the BOA model to models similar

Sevillano *et al. Genet Sel Evol* (2017) 49:75

Page 13 of 14

to G_A and G_B. Ibáñez-Escriche et al. [15] used a simulated population of two-way and three-way CB, for a trait with a heritability of 0.3. They observed that the prediction accuracy of EBV of PB pigs for CB performance with the G_A model was often equal or higher compared to that with the BOA model. The superiority of the BOA model was only observed when PB populations were distant or unrelated, and SNP density was low. Similarly, Esfandyari et al. [27] tested the BOA model with a simulated two-way CB population for a trait with a heritability of 0.3 and a r_{pc} of 0.78. They observed a higher response to selection in CB animals when the BOA model was used compared to the G_A model, but, again, only when PB populations were distantly related. Vandenplas et al. [44] predicted the average reliability of EBV for CB performance obtained from the G_B and BOA models using simulated PB and two-way CB data and different heritabilities (0.20, 0.40, and 0.95), r_{pc} (0.30 and 0.70), and population relatedness. In their study, average reliabilities of the BOA model were always lower than those of the G_B model. The difference in reliabilities between the BOA and G_B models also increased with increasing heritability, r_{pc} and with the population relatedness. Using real data of two-way CB, Xiang et al. [28] and Lopes et al. [29] tested the BOA approach. Xiang et al. [28] used a single-step model with a trait that had a CB heritability of 0.10 and r_{pc} of 0.59 and 0.73 between each breed. They obtained up to 13% higher accuracy for EBV of PB pigs for CB performance considering breed-specific SNP effects. Lopes et al. [29] tested the BOA approach with two traits that had a CB heritability of 0.14 and 0.37, respectively and r_{pc} higher than 0.88. They obtained similar prediction accuracies with the BOA approach than with a model that did not account for breed-specific SNP effects in CB animals. The results from these studies indicate that breeding values are better estimated with the BOA model for traits with a low heritability and low r_{pc}. In our study, CB and PB heritabilities were higher than 0.22, which may have limited the positive impact of the BOA model. Therefore, already considering distantly-related breeds, the BOA model seems to outperform the G_A and G_B models for predicting breeding values of PB animals for CB performance, only when the r_{pc} and heritabilities of the analysed trait are low.

Conclusions

A positive impact of the BOA model was observed for ADG in breed LR, which showed a low r_{pc} (0.30). Results from the literature and from our study suggest that, in cases where traits have a combination of low r_{pc} and low heritabilities, and breeds are distantly related, the use of the BOA model is justified. In other cases, using CB information in a model that does not account for breed-specific SNP effects in CB animals, such as the G_A and G_B

models, does not seem to jeopardize predictions and may be preferred because it can be more easily implemented than the BOA model.

Additional file

Additional file 1: Table S1. Standard errors of additive genetic variance (σ_a^2), litter variance (σ_u^2), residual variance (σ_e^2), and heritabilities for each breed for PB and CB performance, and genetic correlations between purebred and CB pigs (r_{pc}), estimated for each trait using the BOA[a], G_A^b, and G_B^c models. Description: Standard errors of the estimated variance components, heritabilities, and genetic correlations shown in Table 7.

Authors' contributions
CAS prepared the data, conducted the analyses, prepared figures and tables, and wrote the first draft of the manuscript. JV was involved in the construction and evaluation of the models and in the discussion of analysis issues. JWMB participated in the design of the study and coordination. RB participated in the design of the study, and was involved in the interpretation of the phenotypic information. MPLC participated in the design of the study and coordination, was involved in the construction and evaluation of the models, and the discussion of analysis issues. All authors read and approved the final manuscript.

Author details
[1] Wageningen University & Research Animal Breeding and Genomics, 6700 AH Wageningen, The Netherlands. [2] Topigs Norsvin Research Center, 6640 AA Beuningen, The Netherlands.

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

References

1. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: genomic selection in dairy cattle: progress and challenges. J Dairy Sci. 2009;92:433–43.
2. Forni S, Aguilar I, Misztal I, Deeb N. Genomic relationships and biases in the evaluation of sow litter size. In: Proceedings of the 9th world congress on genetics applied to livestock production: 1–6 August 2010; Liepzig. 2010.

Sevillano *et al. Genet Sel Evol* (2017) 49:75

Page 14 of 14

3. Jannink JL, Lorenz AJ, Iwata H. Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics. 2010;9:166–77.

4. Wolc A, Stricker C, Arango J, Settar P, Fulton JE, O'Sullivan NP, et al. Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. Genet Sel Evol. 2011;43:5.

5. Wei M, van der Steen HAM. Comparison of reciprocal recurrent selection with pure-line selection systems in animal breeding (a review). Anim Breed Abstr. 1991;59:281–98.

6. Brandt H, Täubert H. Parameter estimates for purebred and crossbred performances in pigs. J Anim Breed Genet. 1998;115:97–104.

7. Dekkers JCM. Marker-assisted selection for commercial crossbred performance. J Anim Sci. 2007;85:2104–14.

8. Van Grevenhof IE, Van der Werf JH. Design of reference populations for genomic selection in crossbreeding programs. Genet Sel Evol. 2015;47:14.

9. Lutaaya E, Misztal I, Mabry JW, Short T, Timm HH, Holzbauer R. Genetic parameter estimates from joint evaluation of purebreds and crossbreds in swine using the crossbred model. J Anim Sci. 2001;9:3002–7.

10. Nakavisut S, Crump R, Suarez M, Graser HU. Genetic correlations between the performance of purebred and crossbred pigs. Proc Assoc Adv Anim Breed Genet. 2005;16:99–102.

11. Knap P, Wang L. Pig breeding for improved feed efficiency. In: Patience JF, editor. Feed efficiency in swine. Wageningen: Wageningen Academic Publishers; 2012. p. 167–81.

12. Wei M, van der Werf J. Genetic correlation and heritabilities for purebred and crossbred performance in poultry egg production traits. J Anim Sci. 1995;73:2220–6.

13. Newman S, Reverter A, Johnston DJ. Purebred-crossbred performance and genetic evaluation of postweaning growth and carcass traits in Bos indicus × Bos taurus crosses in Australia. J Anim Sci. 2002;80:1801–8.

14. Schrag TA, Möhring J, Maurer HP, Dhillon BS, Melchinger AE, Piepho HP, et al. Molecular marker-based prediction of hybrid performance in maize using unbalanced data from multiple experiments with factorial crosses. Theor Appl Genet. 2009;118:741–51.

15. Ibáñēz-Escriche N, Fernando RL, Toosi A, Dekkers JC. Genomic selection of purebreds for crossbred performance. Genet Sel Evol. 2009;41:12.

16. Kinghorn B, Hickey J, Van Der Werf J. Reciprocal recurrent genomic selection for total genetic merit in crossbred individuals. In: Proceedings of the 9th world congress on genetics applied to livestock production: 1–6 August 2010; Leipzig. 2010.

17. Christensen OF, Madsen P, Nielsen B, Su G. Genomic evaluation of both purebred and crossbred performances. Genet Sel Evol. 2014;46:23.

18. Christensen OF, Legarra A, Lund MS, Su G. Genetic evaluation for three-way crossbreeding. Genet Sel Evol. 2015;47:98.

19. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157:1819–29.

20. VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008;91:4414–23.

21. de Roos A, Hayes BJ, Spelman RJ, Goddard ME. Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. Genetics. 2008;179:1503–12.

22. Veroneze R, Bastiaansen JWM, Knol EF, Guimarães SE, Silva FF, Harlizius B, et al. Linkage disequilibrium patterns and persistence of phase in purebred and crossbred pig (*Sus scrofa*) populations. BMC Genet. 2014;15:126.

23. Makgahlela ML, Strandén I, Nielsen US, Sillanpää MJ, Mäntysaari EA. The estimation of genomic relationships using breedwise allele frequencies among animals in multibreed populations. J Dairy Sci. 2013;96:5364–75.

24. Makgahlela ML, Strandén I, Nielsen US, Sillanpää MJ, Mäntysaari EA. Using the unified relationship matrix adjusted by breed-wise allele frequencies in genomic evaluation of a multibreed population. J Dairy Sci. 2014;97:1117–27.

25. Moghaddar N, Swan AA, van der Werf JH. Comparing genomic prediction accuracy from purebred, crossbred and combined purebred and crossbred reference populations in sheep. Genet Sel Evol. 2014;46:58.

26. Lourenco DA, 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.

27. Esfandyari H, Sørensen AC, Bijma P. A crossbred reference population can improve the response to genomic selection for crossbred performance. Genet Sel Evol. 2015;47:76.

28. 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.

29. Lopes MS, Bovenhuis H, Hidalgo AM, Arendonk JA, Knol EF, Bastiaansen JW. Genomic selection for crossbred performance accounting for breed-specific effects. Genet Sel Evol. 2017;49:51.

30. Vandenplas J, Calus MPL, Sevillano CA, Windig JJ, Bastiaansen JWM. Assigning breed origin to alleles in crossbred animals. Genet Sel Evol. 2016;48:61.

31. Sevillano CA, Vandenplas J, Bastiaansen JWM, Calus MPL. Empirical determination of breed-of-origin of alleles in three-way crossbred pigs. Genet Sel Evol. 2016;48:55.

32. Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 2014;15:478.

33. Hickey JM, Kinghorn BP, Tier B, Wilson JF, Dunstan N, van der Werf JH. A combined long-range phasing and long haplotype imputation method to impute phase for SNP genotypes. Genet Sel Evol. 2011;43:12.

34. Gilmour A, Gogel B, Cullis B, Thompson R. ASReml user guide release 3.0. Hemel Hempstead: VSN International Ltd.; 2009.

35. Jorjani H, Klei L, Emanuelson U. A simple method for weighted bending of genetic (co) variance matrices. J Dairy Sci. 2003;86:677–9.

36. Ten Napel J, Calus MPL, Lidauer M, Stranden I, Mäntysaari E, Mulder H, et al. MiXBLUP, the Mixed-model Best Linear Unbiased Prediction software for PCs for large genetic evaluation systems. Version 2.0. Wageningen. 2016.

37. Saatchi M, McClure MC, McKay SD, Rolf MM, Kim J, Decker JE, et al. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. Genet Sel Evol. 2011;43:40.

38. Calus MP, Vandenplas J, Ten Napel J, Veerkamp RF. Validation of simultaneous deregression of cow and bull breeding values and derivation of appropriate weights. J Dairy Sci. 2016;99:6403–19.

39. Přibyl J, Madsen P, Bauer J, Přibylová J, Šimečková M, Vostrý L, et al. Contribution of domestic production records, Interbull estimated breeding values, and single nucleotide polymorphism genetic markers to the single-step genomic evaluation of milk production. J Dairy Sci. 2013;96:1865–73.

40. Stuart A, Ord K. Kendall's advanced theory of statistics, vol. 1. 6th ed. London: Hodder Education; 1994.

41. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984;38:1358–70.

42. Zumbach B, Misztal I, Tsuruta S, Holl J, Herring W, Long T. Genetic correlations between two strains of Durocs and crossbreds from differing production environments for slaughter traits. J Anim Sci. 2007;85:901–8.

43. 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.

44. Vandenplas J, Windig JJ. Calus MP Prediction of the reliability of genomic breeding values for crossbred performance. Genet Sel Evol. 2017;49:43.