

RESEARCH

Open Access

# Genomic evaluation of both purebred and crossbred performances

Ole F Christensen<sup>1\*</sup>, Per Madsen<sup>1</sup>, Bjarne Nielsen<sup>2</sup> and Guosheng Su<sup>1</sup>

## Abstract

**Background:** For a two-breed crossbreeding system, Wei and van der Werf presented a model for genetic evaluation using information from both purebred and crossbred animals. The model provides breeding values for both purebred and crossbred performances. Genomic evaluation incorporates marker genotypes into a genetic evaluation system. Among popular methods are the so-called single-step methods, in which marker genotypes are incorporated into a traditional animal model by using a combined relationship matrix that extends the marker-based relationship matrix to non-genotyped animals. However, a single-step method for genomic evaluation of both purebred and crossbred performances has not been developed yet.

**Results:** An extension of the Wei and van der Werf model that incorporates genomic information is presented. The extension consists of four steps: (1) the Wei van der Werf model is reformulated using two partial relationship matrices for the two breeds; (2) marker-based partial relationship matrices are constructed; (3) marker-based partial relationship matrices are adjusted to be compatible to pedigree-based partial relationship matrices and (4) combined partial relationship matrices are constructed using information from both pedigree and marker genotypes. The extension of the Wei van der Werf model can be implemented using software that allows inverse covariance matrices in sparse format as input.

**Conclusions:** A method for genomic evaluation of both purebred and crossbred performances was developed for a two-breed crossbreeding system. The method allows information from crossbred animals to be incorporated in a coherent manner for such crossbreeding systems.

## Background

Production systems based on crossbreeding are predominant in pig and chicken breeding and take advantage of the increased performance of crossbred animals compared to purebred animals. For a two-breed crossbreeding system, Wei and van der Werf (Appendix 2 in [1]) presented a model for genetic evaluation using information from both purebred and crossbred animals. The model provides estimated breeding values for purebred (mating with own breed) and crossbred (mating with the other breed) performances that are different but correlated. The model is particularly attractive since it can fit a breeding goal that includes both purebred and crossbred performances (see

Jiang and Groen [2]). This model is the starting point of our paper.

Genomic selection [3] has offered a new paradigm for livestock breeding and has been successfully applied for selection within purebred populations [4-6]. Moreover, genomic selection also offers greater opportunities for incorporating information from crossbreds and selecting for crossbred performance [7-9]. Genomic selection of purebreds for crossbred performance was proposed by Ibánñez-Escriche et al. [7] that used phenotypes on crossbreds only, and a genomic model with breed of origin specific allele substitution effects. The resulting breeding values for purebred animals were for crossbred performance. Although the study included genomic data, it was less sophisticated than the Wei and van der Werf model [1] since each animal had only one breeding value and phenotype recordings in purebreds were not used. In addition, it assumed that all relevant animals were

\*Correspondence: OleF.Christensen@agrsci.dk

<sup>1</sup>Aarhus University, Department of Molecular Biology and Genetics, Center for Quantitative Genetics and Genomics, Blichers Allé 20, P.O. BOX 50, DK-8830 Tjele, Denmark

Full list of author information is available at the end of the article

genotyped, which would not be a very likely scenario in practice.

In cases in which not all animals are genotyped, the so-called single-step methods [10-12] provide a coherent approach for genomic evaluation. These methods incorporate marker genotypes into a traditional animal model [13] by using a combined relationship matrix that extends the marker-based relationship matrix of VanRaden [14] to non-genotyped animals, and they have been shown to perform well for genomic evaluation of dairy cattle [11,15], pigs [16,17] and chickens [18]. Misztal et al. [19] provided an extension with “unknown-parent groups” to allow for different populations, but using such an approach on data from both purebred and crossbred animals would assume equal genetic variances in the two breeds and in the crossbreds, and also that breeding values for purebred and crossbred performances are the same. A single-step method for genomic evaluation of both purebred and crossbred performances has not been developed yet.

In a genomic model, when crossbred animals are genotyped, it is natural to split the additive genetic effect of crossbreds into breed of origin specific genetic components, as in Ibánñez-Escriche et al. [7]. Each of these components is a partial genetic effect, in the sense that only breed-specific alleles are used. This use of the terminology “partial genetic effect” is consistent with the model of Garcia-Cortes and Toro [20], in which for multi-breed analysis the additive genetic value is split into several independent parts depending on their genetic origin, with the variance-covariance structure of each part being determined by a partial relationship matrix (constructed from pedigree). A partial relationship matrix is a relationship matrix that describes relationships only according to genetic origin. From this point of view, partial relationship matrices are key when constructing a single-step method for both purebred and crossbred performances. However, the Wei and van der Werf model is not formulated using partial relationship matrices, and it therefore needs to be reformulated for the purpose of incorporating genomic information.

The aim of this paper is to present an extension of the Wei van der Werf model that incorporates genomic information. The extension consists of four steps: (1) the Wei van der Werf model is reformulated using two partial relationship matrices [20] for the two breeds; (2) marker-based partial relationship matrices similar to VanRaden [14] are constructed; (3) marker-based partial relationship matrices are adjusted to be compatible to pedigree-based partial relationship matrices, similar to Christensen et al. [17] and (4) combined partial relationship matrices are constructed using information from both pedigree and marker genotypes, similar to the combined relationship matrix of Legarra et al., Aguilar et al. and Christensen and Lund [10-12]. This extension of the Wei and van der Werf

model can be implemented using software that allows inverse covariance matrices in sparse format as input.

## Methods

### The Wei and van der Werf model

Here, the Wei van der Werf model (Appendix 2 in [1]) is presented. The two breeds are named  $\mathcal{A}$  and  $\mathcal{B}$ , and it is assumed that all crossbred animals  $\mathcal{AB}$  have known purebred parents. The number of animals in the pedigree is  $n_{\mathcal{A}}$  and  $n_{\mathcal{B}}$  for breed  $\mathcal{A}$  and breed  $\mathcal{B}$ , respectively, and the number of crossbred animals is  $n_{\mathcal{AB}}$ . The model for the phenotypes is a trivariate model

$$\begin{aligned} \mathbf{y}_{\mathcal{A}} &= \mathbf{X}_{\mathcal{A}}\boldsymbol{\beta}_{\mathcal{A}} + \mathbf{Z}_{\mathcal{A}}\mathbf{a}_{\mathcal{A}} + \mathbf{e}_{\mathcal{A}}, \\ \mathbf{y}_{\mathcal{B}} &= \mathbf{X}_{\mathcal{B}}\boldsymbol{\beta}_{\mathcal{B}} + \mathbf{Z}_{\mathcal{B}}\mathbf{a}_{\mathcal{B}} + \mathbf{e}_{\mathcal{B}}, \\ \mathbf{y}_{\mathcal{AB}} &= \mathbf{X}_{\mathcal{AB}}\boldsymbol{\beta}_{\mathcal{AB}} + \mathbf{c}_{\mathcal{AB}} + \mathbf{e}_{\mathcal{AB}}, \end{aligned} \quad (1)$$

where the vectors  $\mathbf{y}_{\mathcal{A}}$ ,  $\mathbf{y}_{\mathcal{B}}$  and  $\mathbf{y}_{\mathcal{AB}}$  contain phenotypes on the breed  $\mathcal{A}$ , breed  $\mathcal{B}$  and crossbred  $\mathcal{AB}$  animals, respectively, and for the three breed groups  $\mathcal{A}$ ,  $\mathcal{B}$  and  $\mathcal{AB}$ , the vectors  $\mathbf{X}_{\mathcal{A}}\boldsymbol{\beta}_{\mathcal{A}}$ ,  $\mathbf{X}_{\mathcal{B}}\boldsymbol{\beta}_{\mathcal{B}}$  and  $\mathbf{X}_{\mathcal{AB}}\boldsymbol{\beta}_{\mathcal{AB}}$  contain fixed effects, and  $\mathbf{e}_{\mathcal{A}} \sim N(\mathbf{0}, \sigma_{\mathcal{A}}^2 \mathbf{I})$ ,  $\mathbf{e}_{\mathcal{B}} \sim N(\mathbf{0}, \sigma_{\mathcal{B}}^2 \mathbf{I})$  and  $\mathbf{e}_{\mathcal{AB}} \sim N(\mathbf{0}, \sigma_{\mathcal{AB}}^2 \mathbf{I})$  are the residual error vectors. The  $n_{\mathcal{A}}$ -dimensional vector  $\mathbf{a}_{\mathcal{A}}$  contains breeding values for purebred performance for breed  $\mathcal{A}$  animals (mating within breed  $\mathcal{A}$ ), and matrix  $\mathbf{Z}_{\mathcal{A}}$  is an incidence matrix assigning breeding values to records. Vector  $\mathbf{a}_{\mathcal{B}}$  and matrix  $\mathbf{Z}_{\mathcal{B}}$  are defined similarly for breed  $\mathcal{B}$ . Finally, the  $n_{\mathcal{AB}}$ -dimensional vector  $\mathbf{c}_{\mathcal{AB}}$  contains the additive genetic effects for crossbred animals, and these are related to the vectors of breeding values for purebred animals for crossbred performance (mating with the other breed) as follows

$$\mathbf{c}_{\mathcal{AB}} = 0.5(\tilde{\mathbf{Z}}_{\mathcal{AB},\mathcal{A}}\mathbf{c}_{\mathcal{A}} + \tilde{\mathbf{Z}}_{\mathcal{AB},\mathcal{B}}\mathbf{c}_{\mathcal{B}}) + \boldsymbol{\Phi}_{\mathcal{AB}}, \quad (2)$$

where the matrices  $\tilde{\mathbf{Z}}_{\mathcal{AB},\mathcal{A}}$  and  $\tilde{\mathbf{Z}}_{\mathcal{AB},\mathcal{B}}$  assign purebred parents to crossbred offspring,  $\mathbf{c}_{\mathcal{A}}$  is an  $n_{\mathcal{A}}$ -dimensional vector containing breeding values for crossbred performance for breed  $\mathcal{A}$  animals (mating with breed  $\mathcal{B}$  animals),  $\mathbf{c}_{\mathcal{B}}$  is an  $n_{\mathcal{B}}$ -dimensional vector containing breeding values for crossbred performance for breed  $\mathcal{B}$  animals (mating with breed  $\mathcal{A}$  animals), and the vector  $\boldsymbol{\Phi}_{\mathcal{AB}}$  contains the Mendelian sampling effects.

The genetic covariances are described by

$$\begin{aligned} \text{Var} \begin{bmatrix} \mathbf{a}_{\mathcal{A}} \\ \mathbf{c}_{\mathcal{A}} \end{bmatrix} &= \boldsymbol{\Sigma}^{(\mathcal{A})} \otimes \mathbf{A}_{\mathcal{A}}, \\ \text{Var} \begin{bmatrix} \mathbf{a}_{\mathcal{B}} \\ \mathbf{c}_{\mathcal{B}} \end{bmatrix} &= \boldsymbol{\Sigma}^{(\mathcal{B})} \otimes \mathbf{A}_{\mathcal{B}}, \\ \text{Var} [\boldsymbol{\Phi}_{\mathcal{AB}}] &= \mathbf{D}_{\mathcal{AB}}, \end{aligned}$$

and the three vectors are independent. The matrices  $\mathbf{A}_{\mathcal{A}}$  and  $\mathbf{A}_{\mathcal{B}}$  are the additive relationship matrices for breed

$\mathcal{A}$  and breed  $\mathcal{B}$ , respectively,  $\otimes$  denotes the Kronecker product, and

$$\Sigma^{(A)} = \begin{bmatrix} \Sigma_{11}^{(A)} & \Sigma_{12}^{(A)} \\ \Sigma_{21}^{(A)} & \Sigma_{22}^{(A)} \end{bmatrix},$$

$$\Sigma^{(B)} = \begin{bmatrix} \Sigma_{11}^{(B)} & \Sigma_{12}^{(B)} \\ \Sigma_{21}^{(B)} & \Sigma_{22}^{(B)} \end{bmatrix}$$

are the  $2 \times 2$  variance-covariance matrices containing the genetic variances for purebred breeding values and crossbred breeding values, and the covariance between the two, for breed  $\mathcal{A}$  and breed  $\mathcal{B}$ , respectively. The variance-covariance matrix of the Mendelian sampling term is a diagonal matrix  $\mathbf{D}_{AB}$  with elements

$$\begin{aligned} (D_{AB})_{ii} &= \text{Var}(\mathbf{c}_{ABi}) - \frac{\text{Var}(\mathbf{c}_{Af(i)}) + \text{Var}(\mathbf{c}_{Bm(i)})}{4} \\ &= (\Sigma_{22}^{(A)} + \Sigma_{22}^{(B)})/2 \\ &\quad - (\Sigma_{22}^{(A)}(A_A)_{f(i)f(i)} + \Sigma_{22}^{(B)}(A_B)_{m(i)m(i)})/4 \\ &= \Sigma_{22}^{(A)}(1/2 - (A_A)_{f(i)f(i)}/4) \\ &\quad + \Sigma_{22}^{(B)}(1/2 - (A_B)_{m(i)m(i)}/4), \end{aligned} \quad (3)$$

where for crossbred animal  $i$ ,  $f(i)$  denotes the breed  $\mathcal{A}$  parent and  $m(i)$  denotes the breed  $\mathcal{B}$  parent.

The Wei and van der Werf model is an additive genetic model in the sense that the breeding values for purebred performance,  $\mathbf{a}_A$ ,  $\mathbf{a}_B$ , are additive genetic effects, and the breeding values for crossbred performance,  $\mathbf{c}_A$ ,  $\mathbf{c}_B$ , in combination with the genetic effects  $\mathbf{c}_{AB}$  are also additive genetic effects. The model therefore does not contain dominance genetic effects explicitly. In practice, such an additive genetic model may also partly capture dominant gene actions and other non-additive gene actions [21]. The fact that genetic correlations between purebred and crossbred performances are different from one would be due to the presence of dominant gene actions in combination with different allele frequencies in the two breeds [22], in addition to genetic effects being different in different environments. In addition, the model captures the general level of heterosis in crossbred animals since it has a separate fixed mean effect for crossbred animals.

Wei and van der Werf [1] made an alternative formulation of the model. The term  $\mathbf{c}_{AB}$  is not of interest for genetic evaluation when crossbred animals are not used for breeding, and Wei and van der Werf reformulated the model using  $\epsilon_{AB} = \Phi_{AB} + \mathbf{e}_{AB}$  as the residual error term for the crossbred phenotypes and thereby expressed the model as a reduced model using only the terms  $\mathbf{a}_A$ ,  $\mathbf{c}_A$ ,  $\mathbf{a}_B$  and  $\mathbf{c}_B$  with breeding values for purebred animals. Note that due to different levels of inbreeding of parents (see formula (3)), the term  $\epsilon_{AB}$  has heterogeneous variance, and assuming a constant variance is an approximation. The reduced model can be implemented using software

that handle multi-trait genetic models. For the purpose of this paper, observed marker genotypes on crossbred animals provide information on the Mendelian sampling term  $\Phi_{AB}$ , and the absorption of this term into the residual error term is therefore not well-suited. For this reason we do not follow the reduced model in this paper.

Finally, the special case where  $\Sigma_{22}^{(A)} = \Sigma_{22}^{(B)} = \Sigma_{22}$  can be formulated as

$$\text{Var} \begin{bmatrix} \mathbf{a}_A \\ \star \\ \star \\ \star \\ \mathbf{a}_B \\ \star \\ \star \\ \mathbf{c}_A \\ \mathbf{c}_B \\ \mathbf{c}_{AB} \end{bmatrix} = \begin{bmatrix} \Sigma_{11}^{(A)} & 0 & \Sigma_{12}^{(A)} \\ 0 & \Sigma_{11}^{(B)} & \Sigma_{12}^{(B)} \\ \Sigma_{21}^{(A)} & \Sigma_{21}^{(B)} & \Sigma_{22} \end{bmatrix} \otimes \mathbf{A},$$

where  $\star$  denotes artificial random vectors such that the genetic variance-covariance matrix can be expressed using a Kronecker product, and  $\mathbf{A}$  is the usual additive relationship matrix for all animals. This can therefore be implemented using a combined pedigree across all animals. We will return to this special case in the Discussion section.

### Reformulated model

Here, the Wei and van der Werf [1] model is reformulated using breed-specific partial relationship matrices, as in Garcia-Cortes and Toro [20]. Partial relationship matrices describe relationships according to genetic origin.

The starting point of the reformulation is the Mendelian sampling term for the crossbred animals in formula (2). This term can be split into breed of origin effects,  $\Phi_{AB} = \Phi_{AB}^{(A)} + \Phi_{AB}^{(B)}$ , where  $\Phi_{AB}^{(A)}$  and  $\Phi_{AB}^{(B)}$  are independent. Formula (3) can be formulated in matrix notation as  $\mathbf{D}_{AB} = \text{Var}(\Phi_{AB}^{(A)}) + \text{Var}(\Phi_{AB}^{(B)})$ , where  $\text{Var}(\Phi_{AB}^{(A)}) = \Sigma_{22}^{(A)}(0.5\mathbf{I}_{n_{AB}} - 0.25\text{diag}(\tilde{\mathbf{Z}}_{AB,A}\mathbf{A}_A\tilde{\mathbf{Z}}_{AB,A}^T))$  and  $\text{Var}(\Phi_{AB}^{(B)}) = \Sigma_{22}^{(B)}(0.5\mathbf{I}_{n_{AB}} - 0.25\text{diag}(\tilde{\mathbf{Z}}_{AB,B}\mathbf{A}_B\tilde{\mathbf{Z}}_{AB,B}^T))$ , with  $\mathbf{I}_{n_{AB}}$  being an identity matrix of size  $n_{AB}$ , and  $\text{diag}(\tilde{\mathbf{Z}}_{AB,A}\mathbf{A}_A\tilde{\mathbf{Z}}_{AB,A}^T)$  and  $\text{diag}(\tilde{\mathbf{Z}}_{AB,B}\mathbf{A}_B\tilde{\mathbf{Z}}_{AB,B}^T)$  denoting diagonal matrices containing diagonal elements in matrices  $\tilde{\mathbf{Z}}_{AB,A}\mathbf{A}_A\tilde{\mathbf{Z}}_{AB,A}^T$  and  $\tilde{\mathbf{Z}}_{AB,B}\mathbf{A}_B\tilde{\mathbf{Z}}_{AB,B}^T$ , respectively. In other words, we decompose the Mendelian sampling term into Mendelian sampling terms for the  $\mathcal{A}$  and  $\mathcal{B}$  gametes. The additive genetic effect for crossbred animals in formula (2) can then be expressed as

$$\mathbf{c}_{AB} = \mathbf{c}_{AB}^{(A)} + \mathbf{c}_{AB}^{(B)},$$

where  $\mathbf{c}_{AB}^{(A)} = 0.5\tilde{\mathbf{Z}}_{AB,A}\mathbf{c}_A + \Phi_{AB}^{(A)}$ ,  $\mathbf{c}_{AB}^{(B)} = 0.5\tilde{\mathbf{Z}}_{AB,B}\mathbf{c}_B + \Phi_{AB}^{(B)}$ , and  $\mathbf{c}_{AB}^{(A)}$  and  $\mathbf{c}_{AB}^{(B)}$  are independent, i.e. the genetic effects for crossbred animals is split into two breed of

origin genetic effects. Thus, the model equation system (1) can be written as:

$$\begin{aligned} \mathbf{y}_A &= \mathbf{X}_A \boldsymbol{\beta}_A + \mathbf{Z}_A \mathbf{a}_A + \mathbf{e}_A, \\ \mathbf{y}_B &= \mathbf{X}_B \boldsymbol{\beta}_B + \mathbf{Z}_B \mathbf{a}_B + \mathbf{e}_B, \\ \mathbf{y}_{AB} &= \mathbf{X}_{AB} \boldsymbol{\beta}_{AB} + \mathbf{c}_{AB}^{(A)} + \mathbf{c}_{AB}^{(B)} + \mathbf{e}_{AB}. \end{aligned} \quad (4)$$

Focusing on breed  $\mathcal{A}$ , the variance-covariance matrix of the genetic effects  $\mathbf{c}_{AB}^{(A)}$  becomes

$$\begin{aligned} \text{Var}(\mathbf{c}_{AB}^{(A)}) &= 0.25 \Sigma_{22}^{(A)} \tilde{\mathbf{Z}}_{AB,A} \mathbf{A}_A \tilde{\mathbf{Z}}_{AB,A}^T \\ &+ \Sigma_{22}^{(A)} (0.5 \mathbf{I}_{n_{AB}} - 0.25 \text{diag}(\tilde{\mathbf{Z}}_{AB,A} \mathbf{A}_A \tilde{\mathbf{Z}}_{AB,A}^T)) \\ &= \Sigma_{22}^{(A)} \mathbf{A}_{AB}^{(A)}, \end{aligned}$$

where  $\mathbf{A}_{AB}^{(A)}$  is a matrix with elements  $(\mathbf{A}_{AB}^{(A)})_{ii} = 0.5$  and  $(\mathbf{A}_{AB}^{(A)})_{i,i'} = 0.25(\tilde{\mathbf{Z}}_{AB,A} \mathbf{A}_A \tilde{\mathbf{Z}}_{AB,A}^T)_{i,i'}$  when  $i \neq i'$ , and the covariance matrix between  $\mathbf{c}_{AB}^{(A)}$  and  $\mathbf{c}_A$  becomes

$$\text{Cov}(\mathbf{c}_{AB}^{(A)}, \mathbf{c}_A) = 0.5 \Sigma_{22}^{(A)} \tilde{\mathbf{Z}}_{AB,A} \mathbf{A}_A.$$

Therefore, the variance-covariance matrix of breed  $\mathcal{A}$  specific genetic effects for crossbred performance equals

$$\text{Var} \begin{bmatrix} \mathbf{c}_A \\ \mathbf{c}_{AB}^{(A)} \end{bmatrix} = \Sigma_{22}^{(A)} \mathbf{A}^{(A)},$$

where the symmetric  $(n_A + n_{AB})$ -dimensional matrix

$$\mathbf{A}^{(A)} = \begin{bmatrix} \mathbf{A}_A & 0.5 \mathbf{A}_A \tilde{\mathbf{Z}}_{AB,A}^T \\ 0.5 \tilde{\mathbf{Z}}_{AB,A} \mathbf{A}_A & \mathbf{A}_{AB}^{(A)} \end{bmatrix}$$

is the breed  $\mathcal{A}$  specific partial relationship matrix in Garcia-Cortes and Toro [20] (see below).

Similarly, the variance-covariance matrix of breed  $\mathcal{B}$  specific genetic effects for crossbred performance equals

$$\text{Var} \begin{bmatrix} \mathbf{c}_B \\ \mathbf{c}_{AB}^{(B)} \end{bmatrix} = \Sigma_{22}^{(B)} \mathbf{A}^{(B)},$$

where

$$\mathbf{A}^{(B)} = \begin{bmatrix} \mathbf{A}_B & 0.5 \mathbf{A}_B \tilde{\mathbf{Z}}_{AB,B}^T \\ 0.5 \tilde{\mathbf{Z}}_{AB,B} \mathbf{A}_B & \mathbf{A}_{AB}^{(B)} \end{bmatrix}$$

is the breed  $\mathcal{B}$  specific partial relationship matrix (see below).

Garcia-Cortes and Toro [20] presented a partition of the variance-covariance matrix of additive genetic values into breed-specific and breed-segregation terms, where each term is a scaled partial relationship matrix. The partial relationship matrices are constructed using recursive formulas similar to usual recursive formulas for the additive relationship matrix [23]. For the two-breed terminal crossbreeding system, the partition results in breed  $\mathcal{A}$  and breed  $\mathcal{B}$  specific partial relationship matrices, but no breed-segregation partial relationship matrices; we refer to Garcia-Cortes and Toro [20] for the general case. The

recursive formulas for the breed  $\mathcal{A}$  specific partial relationships are:

$$\begin{aligned} A_{ii}^{(A)} &= f_i^A + A_{f(i)m(i)}^{(A)}/2, \\ A_{i'i'}^{(A)} &= (A_{f(i)i'}^{(A)} + A_{m(i)i'}^{(A)})/2, \end{aligned}$$

where  $f(i)$  and  $m(i)$  are the two parents of animal  $i$ , animal  $i'$  is not a descendant of  $i$ , and  $f_i^A$  is the breed  $\mathcal{A}$  proportion of individual  $i$  (equal to 1 for purebred  $\mathcal{A}$  animals, 0 for purebred  $\mathcal{B}$  animals and 0.5 for crossbred animals). To insure that partial relationship matrices are invertible, Munilla-Leguizamón and Cantet [24] suggested to redefine the partial relationship matrices such that only elements that are non-null by breed origin were included, i.e. for the breed  $\mathcal{A}$  specific partial relationships shown here, the elements related to purebred  $\mathcal{B}$  animals are excluded. In this paper, we followed that suggestion, and it is not difficult to check that the matrix in (5) is indeed the breed  $\mathcal{A}$  specific partial relationship matrix. Using matrix formulation, the breed  $\mathcal{A}$  specific partial relationship matrix is  $\mathbf{A}^{(A)} = \mathbf{T} \mathbf{D} \mathbf{T}^T$  where  $\mathbf{D}$  is a diagonal matrix with elements  $D_{ii} = 1 - (A_{f(i)f(i)}^{(A)} + A_{m(i)m(i)}^{(A)})/4$  when animal  $i$  is breed  $\mathcal{A}$ , and  $D_{ii} = 0.5 - A_{f(i)f(i)}^{(A)}/4$  when animal  $i$  is crossbred with breed  $\mathcal{A}$  parent  $f(i)$ . For matrix  $\mathbf{T}$ , the inverse matrix  $\mathbf{T}^{-1}$  is a lower triangular matrix with diagonal elements equal to 1 and in the lower diagonal, the only non-zero elements are  $-0.5$  for offspring parent elements. An example with a small pedigree is in Table 1, and the corresponding partial relationship matrices are in Tables 2 and 3.

The reformulation of the Wei and van der Werf model is completed by introducing two artificial random vectors  $\mathbf{a}_{AB}^{(A)}$  and  $\mathbf{a}_{AB}^{(B)}$  such that genetic variance-covariance matrices can be presented using Kronecker products. For breed  $\mathcal{A}$ , the genetic covariances are described by

$$\text{Var} \begin{bmatrix} \mathbf{a}_A \\ \mathbf{a}_{AB}^{(A)} \\ \mathbf{c}_A \\ \mathbf{c}_{AB}^{(A)} \end{bmatrix} = \Sigma^{(A)} \otimes \mathbf{A}^{(A)},$$

**Table 1 Example pedigree**

Id	Father	Mother	Breed group
1	0	0	$\mathcal{A}$
2	0	0	$\mathcal{A}$
3	1	2	$\mathcal{A}$
4	0	0	$\mathcal{B}$
5	0	0	$\mathcal{B}$
6	4	5	$\mathcal{B}$
7	2	6	$\mathcal{AB}$
8	3	6	$\mathcal{AB}$

**Table 2 Breed  $\mathcal{A}$  specific partial relationship matrix  $\mathbf{A}^{(\mathcal{A})}$  for the pedigree in Table 1**

Id	1	2	3	7	8
1	1				
2	0	1			
3	1/2	1/2	1		
7	0	1/2	1/4	1/2	
8	1/4	1/4	1/2	1/8	1/2

and similar for breed  $\mathcal{B}$ , the genetic covariances are described by

$$\text{Var} \begin{bmatrix} \mathbf{a}_{\mathcal{B}} \\ \mathbf{a}_{\mathcal{A}\mathcal{B}}^{(\mathcal{B})} \\ \mathbf{c}_{\mathcal{B}} \\ \mathbf{c}_{\mathcal{A}\mathcal{B}}^{(\mathcal{B})} \end{bmatrix} = \boldsymbol{\Sigma}^{(\mathcal{B})} \otimes \mathbf{A}^{(\mathcal{B})}.$$

Implementing the model requires inverses of the two partial relationship matrices. The inverse of a partial relationship matrix  $(\mathbf{A}^{(\mathcal{A})})^{-1}$  can be expressed by the usual formula

$$(\mathbf{A}^{(\mathcal{A})})^{-1} = (\mathbf{T}^{-1})^T \mathbf{D}^{-1} \mathbf{T}^{-1},$$

and the usual methods for computing the diagonal elements of the partial relationship matrix and the inverse partial relationship matrix in sparse format [25,26] can be applied.

The model is a trivariate model with breed  $\mathcal{A}$  and  $\mathcal{B}$  specific genetic effects for both purebred and crossbred performances, and can be implemented using a software package for multivariate mixed models that either explicitly can construct inverses of partial relationship matrices from pedigree or alternatively can use inverse covariance matrices in sparse format as input (e.g., DMU [http://dmu.agrsci.dk], WOMBAT [http://didgeridoo.une.edu.au/km/wombat.php], ASReml [http://www.vsnl.co.uk/software/asreml], blupf90 [http://nce.ads.uga.edu/wiki/doku.php]), MiX99 [http://www.mtt.fi/BGE/Software/MiX99]).

**Table 3 Breed  $\mathcal{B}$  specific partial relationship matrix  $\mathbf{A}^{(\mathcal{B})}$  for the pedigree in Table 1**

Id	4	5	6	7	8
4	1				
5	0	1			
6	1/2	1/2	1		
7	1/4	1/4	1/2	1/2	
8	1/4	1/4	1/2	1/4	1/2

Extending the model to incorporate genomic information requires the construction of two combined breed-specific partial relationship matrices expressed as inverse matrices, and for this purpose, marker-based breed-specific partial relationship matrices need to be constructed, and marker-based and pedigree-based partial relationship matrices need to be made compatible. These are the topics of the following subsections.

### Marker-based partial relationship matrix

Here, a marker-based breed  $\mathcal{A}$  specific partial relationship matrix is constructed. The assumption here is that the marker genotypes for crossbred animals are phased such that it is known which allele originated from breed  $\mathcal{A}$  and which allele originated from breed  $\mathcal{B}$ . The marker genotype matrix  $\mathbf{m}^{\mathcal{A}}$  for purebred  $\mathcal{A}$  animals has elements  $m_{ij}^{\mathcal{A}} = -1, 0$  or  $1$  if SNP  $j$  of individual  $i$  is 11, 12, or 22, respectively. For crossbred animals, the breed  $\mathcal{A}$  marker allele matrix  $\mathbf{q}^{\mathcal{A}}$  has elements  $q_{ij}^{\mathcal{A}} = -0.5$  or  $0.5$  if loci  $j$  of individual  $i$  has breed  $\mathcal{A}$  allele 1 or 2, respectively.

Constructing a marker-based breed-specific partial relationship matrix similar to the marker-based relationship matrix of VanRaden [14] is done by using the breed-specific alleles only. The marker-based breed  $\mathcal{A}$  specific partial relationship matrix  $\mathbf{G}^{(\mathcal{A})}$  is divided into submatrices with indices denoting genotyped breed  $\mathcal{A}$  and crossbred animals,

$$\mathbf{G}^{(\mathcal{A})} = \begin{bmatrix} \mathbf{G}_{\mathcal{A},\mathcal{A}}^{(\mathcal{A})} & \mathbf{G}_{\mathcal{A},\mathcal{A}\mathcal{B}}^{(\mathcal{A})} \\ \mathbf{G}_{\mathcal{A}\mathcal{B},\mathcal{A}}^{(\mathcal{A})} & \mathbf{G}_{\mathcal{A}\mathcal{B},\mathcal{A}\mathcal{B}}^{(\mathcal{A})} \end{bmatrix},$$

which are defined as

$$\begin{aligned} \mathbf{G}_{\mathcal{A},\mathcal{A}}^{(\mathcal{A})} &= \frac{(\mathbf{m}^{\mathcal{A}} - (2\bar{\rho}^{\mathcal{A}} - 1)\mathbf{1}^T)(\mathbf{m}^{\mathcal{A}} - (2\bar{\rho}^{\mathcal{A}} - 1)\mathbf{1}^T)^T}{s^{\mathcal{A}}}, \\ \mathbf{G}_{\mathcal{A},\mathcal{A}\mathcal{B}}^{(\mathcal{A})} &= \frac{(\mathbf{m}^{\mathcal{A}} - (2\bar{\rho}^{\mathcal{A}} - 1)\mathbf{1}^T)(\mathbf{q}^{\mathcal{A}} - (\bar{\rho}^{\mathcal{A}} - \frac{1}{2})\mathbf{1}^T)^T}{s^{\mathcal{A}}}, \\ \mathbf{G}_{\mathcal{A}\mathcal{B},\mathcal{A}\mathcal{B}}^{(\mathcal{A})} &= \frac{(\mathbf{q}^{\mathcal{A}} - (\bar{\rho}^{\mathcal{A}} - \frac{1}{2})\mathbf{1}^T)(\mathbf{q}^{\mathcal{A}} - (\bar{\rho}^{\mathcal{A}} - \frac{1}{2})\mathbf{1}^T)^T}{s^{\mathcal{A}}}, \end{aligned} \quad (5)$$

where the vector  $\bar{\rho}^{\mathcal{A}}$  contains estimated breed  $\mathcal{A}$  specific allele frequencies based on marker genotypes for purebred animals and breed  $\mathcal{A}$  specific marker alleles for crossbred animals, and  $s^{\mathcal{A}}$  is a scaling parameter. The scaling parameter  $s^{\mathcal{A}}$  is unspecified here since we adjust the marker-based partial relationship matrix to make it compatible with the pedigree-based partial relationship matrices, similar to Christensen et al. [17] (see below).

The marker-based breed  $\mathcal{B}$  specific partial relationship matrix  $\mathbf{G}^{(\mathcal{B})}$  is constructed similarly. Matrices  $\mathbf{G}^{(\mathcal{A})}$  and  $\mathbf{G}^{(\mathcal{B})}$  correspond to two different covariance structures, while matrix  $\mathbf{G}^{(\mathcal{A}\mathcal{B})}$  does not exist. For crossbred animals that are genotyped, the genetic effect is the sum of two effects, with variance-covariance matrices proportional to  $\mathbf{G}_{\mathcal{A},\mathcal{A}\mathcal{B}}^{(\mathcal{A})}$  and  $\mathbf{G}_{\mathcal{A}\mathcal{B},\mathcal{A}\mathcal{B}}^{(\mathcal{B})}$ , respectively. Since a genetic effect

with a marker-based relationship matrix can be equivalently formulated as a sum of allele substitution effects, the genetic effect for crossbred animal  $i$  therefore equals

$$c_{AB,i} = c_{AB,i}^{(A)} + c_{AB,i}^{(B)}$$

$$= \sum_{j=1}^p (q_{ij}^A - (\rho_j^A - 1/2))\alpha_j^A + (q_{ij}^B - (\rho_j^B - 1/2))\alpha_j^B,$$

where  $\alpha_j^A, \alpha_j^B$  are independent breed of origin specific substitution effects for SNP  $j = 1 \dots, p$ . The model for crossbred animals is therefore as described by Ibánñez-Escriche et al. [7].

### Compatibility of marker-based and pedigree-based partial relationship matrices

Marker-based and pedigree-based partial relationship matrices must be compatible [17,27,28]. In order to achieve this, either the marker-based partial relationship matrix or the pedigree-based partial relationship matrix must be adjusted [28]. Here, we show how to adjust the breed  $\mathcal{A}$  specific marker-based partial relationship matrix,  $\mathbf{G}^{(A)}$ , to the breed  $\mathcal{A}$  specific pedigree-based partial relationship matrix for the subset of genotyped animals,  $\mathbf{A}_{11}^{(A)}$ , similar to Christensen et al [17]. The adjustment is of the form

$$\mathbf{G}_a^{(A)} = \mathbf{G}^{(A)}\beta + \left[ \begin{array}{c} \mathbf{1}_{n_1}(\mathbf{1}_{n_1})^T \quad \frac{1}{2}\mathbf{1}_{n_1}(\mathbf{1}_{n_2})^T \\ \frac{1}{2}\mathbf{1}_{n_2}(\mathbf{1}_{n_1})^T \quad \frac{1}{4}\mathbf{1}_{n_2}(\mathbf{1}_{n_2})^T \end{array} \right] \alpha$$

$$= \mathbf{G}^{(A)}\beta + \mathbf{K}\alpha,$$

with submatrices corresponding to purebred genotyped and crossbred genotyped animals,  $\mathbf{1}$  denoting a vector of ones (with sub-index denoting the dimension:  $n_1$  is equal to the number of genotyped purebred animals and  $n_2$  to the number of genotyped crossbred animals); matrix  $\mathbf{K}$  being implicitly defined; and  $\alpha$  and  $\beta$  are parameters that need to be estimated. The form of the adjustment above is explained in Appendix A. According to Christensen et al. [17], the parameters  $\alpha$  and  $\beta$  can be determined by solving a system of two equations

$$\bar{A}_{11}^{(A)} = \bar{G}^{(A)}\beta + \bar{K}\alpha,$$

$$\overline{dA}_{11}^{(A)} = \overline{dG}^{(A)}\beta + \overline{dK}\alpha,$$

where  $\bar{A}_{11}^{(A)}$ ,  $\bar{G}^{(A)}$  and  $\bar{K}$  denote averages of all elements of the matrices  $\mathbf{A}_{11}^{(A)}$ ,  $\mathbf{G}^{(A)}$  and  $\mathbf{K}$ , and  $\overline{dA}_{11}^{(A)}$ , respectively, and  $\overline{dA}_{11}^{(A)}$ ,  $\overline{dG}^{(A)}$  and  $\overline{dK}$  denote averages of diagonal elements of the three matrices, respectively. Based on  $\bar{K} = (n_1 + n_2/2)^2 / (n_1 + n_2)^2$  and  $\overline{dK} = (n_1 + n_2/4) / (n_1 + n_2)$ , the resulting parameter estimates become

$$\beta = \frac{\bar{A}_{11}^{(A)} - \overline{dA}_{11}^{(A)} \frac{(n_1 + n_2/2)^2}{(n_1 + n_2)(n_1 + n_2/4)}}{\bar{G}^{(A)} - \overline{dG}^{(A)} \frac{(n_1 + n_2/2)^2}{(n_1 + n_2)(n_1 + n_2/4)}},$$

$$\alpha = \frac{(\overline{dA}_{11}^{(A)} - \overline{dG}^{(A)}\beta)(n_1 + n_2)}{(n_1 + n_2/4)}.$$

Note that parameter  $\beta$  is completely confounded with the scaling parameter  $s^A$  in (5), and the choice of the scaling parameter is therefore irrelevant.

### Combined pedigree-based and marker-based partial relationship matrix

The combined partial relationship matrix ( $\mathbf{H}^{(A)}$ ) can be constructed similar to the combined relationship matrix for purebred animals [10-12]. On the inverse scale, the elements of the matrix for non-genotyped animals do not depend on the marker genotypes. Therefore,

$$(\mathbf{H}^{(A)})^{-1} = \left[ \begin{array}{ccc} (\mathbf{G}_\omega^{(A)})^{-1} - (\mathbf{A}_{11}^{(A)})^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & & \mathbf{0} \end{array} \right] + (\mathbf{A}^{(A)})^{-1}, \quad (6)$$

with  $\mathbf{G}_\omega^{(A)} = (1 - \omega)\mathbf{G}_a^{(A)} + \omega\mathbf{A}_{11}^{(A)}$ . Parameter  $\omega$  is the fraction of genetic variance not captured by the marker genotypes, and in practice should be chosen to maximize accuracy and minimize bias of the resulting estimated breeding values [17].

Computation of the submatrix  $\mathbf{A}_{11}^{(A)}$  follows the Colleau algorithm [29,30], which is based on the decomposition  $\mathbf{A}^{(A)} = \mathbf{TDT}^T$  shown in a previous subsection. The essential idea is to compute the  $i$ th column of  $\mathbf{A}_{11}^{(A)}$  by computing  $\mathbf{A}^{(A)}\mathbf{e}_i$ , where  $\mathbf{e}_i$  is a vector with element  $i$  equal to 1 and all other elements equal to 0, based on Misztal et al. [30]. The algorithm consists of computing consecutively  $\mathbf{r} = \mathbf{T}^T\mathbf{e}_i$  by solving the sparse system  $(\mathbf{T}^{-1})^T\mathbf{r} = \mathbf{e}_i$  for  $\mathbf{r}$ ,  $\mathbf{t} = \mathbf{Dr}$ , and finally  $\mathbf{A}^{(A)}\mathbf{e}_i = \mathbf{T}^T\mathbf{t}$  by solving the sparse system  $(\mathbf{T}^{-1})^T(\mathbf{A}^{(A)}\mathbf{e}_i) = \mathbf{r}$  for  $\mathbf{A}^{(A)}\mathbf{e}_i$ .

In summary, computations for creating  $(\mathbf{H}^{(A)})^{-1}$  are straightforward. First, matrices  $(\mathbf{A}^{(A)})^{-1}$ ,  $\mathbf{G}^{(A)}$  and  $\mathbf{A}_{11}^{(A)}$  are computed, then  $\mathbf{G}^{(A)}$  is adjusted,  $\mathbf{G}_\omega^{(A)} = (1 - \omega)\mathbf{G}_a^{(A)} + \omega\mathbf{A}_{11}^{(A)}$  is computed, and finally matrices  $\mathbf{A}_{11}^{(A)}$  and  $\mathbf{G}_\omega^{(A)}$  are inverted. The sparse inverse matrices  $(\mathbf{H}^{(A)})^{-1}$  and  $(\mathbf{H}^{(B)})^{-1}$  are used as input when implementing the extension of the Wei and van der Werf model.

### Discussion

This paper demonstrates how to incorporate marker genotypes into the Wei and van der Werf model for genetic evaluation using both purebred and crossbred information. The approach builds on using partial relationship matrices, and assumes that the marker genotypes of crossbreds can be phased such that the breed of origin

of alleles is known. Many different algorithms for phasing have been developed [31,32], and it has been shown that the accuracy of phasing depends among others on size of the sample and relatedness of animals within the sample.

An alternative to using combined partial relationship matrices would be to specify one combined relationship matrix across all animals in the three breed groups  $\mathcal{A}$ ,  $\mathcal{B}$  and  $\mathcal{AB}$ . As mentioned in the Methods section, this is actually a special case of the model where  $\Sigma_{22}^{(A)} = \Sigma_{22}^{(B)}$ . With this approach, only one marker-based relationship matrix would have to be created and there would be no need to know the breed of origin of alleles. However, the adjustment of the marker-based relationship matrix to be compatible to the pedigree-based relationship matrix becomes more complicated when both breeds are considered at the same time and, as mentioned, this model is less sophisticated than the model developed in this paper.

More complicated crossbreeding systems with three breeds (mating crossbred  $\mathcal{AB}$  animals with purebred  $\mathcal{C}$  animals) or four breeds (mating breeds  $\mathcal{A}$  and  $\mathcal{B}$ , mating breeds  $\mathcal{C}$  and  $\mathcal{D}$ , and finally mating the two groups of crossbred animals  $\mathcal{AB}$  with  $\mathcal{CD}$ ) are typically used in pig and chicken production. The three-breed crossbreeding system was studied using pig data by Ibanez et al. [33], using the Garcia-Cortes and Toro [20] decomposition of the relationship matrix, and assuming breeding values for purebred and crossbred performances were identical. However, an extension of the Wei and van der Werf model to the three-breed crossbreeding system can be formulated as follows (only the vector containing genetic effects for terminal crossbreds is shown),

$$\mathbf{c}_{(AB)C} = \mathbf{c}_{(AB)C}^{(A)} + \mathbf{c}_{(AB)C}^{(B)} + \mathbf{c}_{(AB)C}^{(C)} + \mathbf{c}_{(AB)C}^{(AB)},$$

where the genetic terms  $\mathbf{c}_{(AB)C}^{(A)}$ ,  $\mathbf{c}_{(AB)C}^{(B)}$  and  $\mathbf{c}_{(AB)C}^{(C)}$  are related to the vectors containing breeding values for  $(AB)C$  crossbred performance for purebred animals,  $\mathbf{c}_A$ ,  $\mathbf{c}_B$  and  $\mathbf{c}_C$ , respectively, by partial relationships. The genetic term  $\mathbf{c}_{(AB)C}^{(AB)}$  is a breed-segregation term that is independent of the other genetic terms, and has variance-covariance matrix  $\Sigma_{(AB)C}^{(AB)} \mathbf{I}_{n_{(AB)C}}/2$ , where  $\Sigma_{(AB)C}^{(AB)}$  is a parameter. Thus, the genetic parameter  $\Sigma_{(AB)C}^{(AB)}$  and the error variance parameter  $\sigma_{(AB)C}^2$  are not both identifiable,  $\mathbf{c}_{(AB)C}^{(AB)}$  can be incorporated into the residual error, and the three breed crossbreeding model can be formulated using three breed-specific partial relationship matrices. Extending the three-breed crossbreeding model to include observed marker genotypes is currently being investigated.

The model presented in this paper is an additive genetic model (in the sense that it considers and estimates substitution effects), but in practice it may capture both additive

gene actions and partly dominant gene actions. Using purebred pig data, Su et al. [34] showed that when an additive genomic model was extended to explicitly incorporate dominance genomic effects, improved accuracies of predictions of both total genetic values and breeding values were obtained. Using simulated data from crossbred animals, Zeng et al. [9] showed that an increased response to selection was obtained with a genomic model with dominance genetic effects compared to an additive genomic model. Lo et al. [35] extended the Wei and van der Werf model to include dominance genetic effects and this model has been used in several studies on real data [36,37]. The formulation of that extension is based on extending the reduced model of Wei and van der Werf (see the Methods section) by incorporating a dominance genetic effect for the purebred phenotypes and a full-sib family effect for the crossbred phenotypes. Similar to the reduced model, this model formulation does not directly contain individual genetic effects for crossbred animals and is, therefore, not well-suited for incorporating genomic information on crossbred animals. A marker-based dominance relationship matrix was proposed by Su et al. [34], but this would need to be extended to a combined dominance relationship matrix, and further extended to a crossbreeding system. Extending the model in this paper to contain dominance genetic effects would be an interesting topic for future research.

## Conclusions

A method for genomic evaluation of both purebred and crossbred performances was developed for a two-breed crossbreeding system. The method allows information from crossbred animals to be incorporated in a coherent manner for such crossbreeding systems.

## Appendix A

In this appendix, we present the explanation behind the adjustment of the marker-based partial relationship matrix. Marker-based relationships, with allele frequencies equal to the observed ones, reflect relationships relative to the genotyped animals, whereas pedigree-based relationships are relative to the base population of the pedigree. The idea behind the adjustment of the marker-based partial relationship matrix is to translate relationships to become relative to the base population of the pedigree, instead of being relative to the given set of animals, as suggested by Powell et al. [38], and which is also the idea behind the adjustment in Christensen et al. [17].

For a given set of animals (purebred and crossbred) and a given breed, let us assume that breed-specific gametes are randomly assigned to animals (purebred animals receive two gametes each, and crossbred animals

receive one gamete each), and let  $\alpha$  be equal to twice the gametic relationship coefficient. The partial relationship matrix for these animals,  $\mathbf{A}^p$ , has entries

$$\begin{cases} 1 + \alpha/2 & i = i' \text{ purebred,} \\ \alpha & i \neq i' \text{ purebreds,} \\ \alpha/2 & i \text{ purebred, } i' \text{ crossbred,} \\ 1/2 & i = i' \text{ crossbred,} \\ \alpha/4 & i \neq i' \text{ crossbreds,} \end{cases}$$

and can therefore be written as

$$\mathbf{A}^p = \left(1 - \frac{\alpha}{2}\right) \begin{bmatrix} \mathbf{I}_{n_1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\mathbf{I}_{n_2} \end{bmatrix} + \begin{bmatrix} \alpha \mathbf{1}_{n_1} (\mathbf{1}_{n_1})^T & \frac{\alpha}{2} \mathbf{1}_{n_1} (\mathbf{1}_{n_2})^T \\ \frac{\alpha}{2} \mathbf{1}_{n_2} (\mathbf{1}_{n_1})^T & \frac{\alpha}{4} \mathbf{1}_{n_2} (\mathbf{1}_{n_2})^T \end{bmatrix}, \quad (7)$$

with submatrices corresponding to purebred and crossbred animals,  $\mathbf{1}$  being a vector of ones and  $\mathbf{I}$  the identity matrix (with sub-indices denoting the dimension:  $n_1$  equal to number of purebred animals and  $n_2$  number of crossbred animals). The matrix

$$\begin{bmatrix} \mathbf{I}_{n_1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2}\mathbf{I}_{n_2} \end{bmatrix},$$

would be a partial relationship matrix when gametes are unrelated ( $\alpha = 0$ ), and therefore the partial relationship matrix relative to the given set of animals. Hence, the formula (7) shows how relationships relative to the given set of animals are related to relationships relative to the base population of the pedigree. Therefore, it provides a formula to translate a marker-based relationship matrix (with allele frequencies being the observed ones) to have the same base population as the pedigree-based relationship matrix. As in Christensen et al. [17], we substitute  $\beta$  for  $1 - \alpha/2$  to incorporate the scaling parameter  $s^A$  in (5).

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

OFC initiated the study, derived the methods and wrote the paper. PM, BN and GS contributed to the conception of the study and provided input to the writing of the paper. All authors have read and approved the final manuscript.

#### Acknowledgements

The work was performed in a project funded through the Green Development and Demonstration Programme (grant no. 3405-11-0279) by the Danish Ministry of Food, Agriculture and Fisheries, the Pig Research Centre and Aarhus University. Comments from two anonymous reviewers greatly helped to improve the clarity of the presentation.

#### Author details

<sup>1</sup>Aarhus University, Department of Molecular Biology and Genetics, Center for Quantitative Genetics and Genomics, Blichers Allé 20, P.O. BOX 50, DK-8830 Tjele, Denmark. <sup>2</sup>Danish Agriculture and Food Council, Pig Research Centre, Breeding and Genetics, Axeltorv 3, DK-1609 Copenhagen V, Denmark.

Received: 5 July 2013 Accepted: 24 February 2014

Published: 25 March 2014

#### References

- Wei M, van der Werf JHJ: **Maximizing genetic response in crossbreds using both purebred and crossbred information.** *Anim Prod* 1994, **59**:401–413.
- Jiang X, Groen AF: **Combined crossbred and purebred selection for reproduction traits in a broiler dam line.** *J Anim Breed Genet* 1999, **116**:111–125.
- Meuwissen THE, Hayes BJ, Goddard ME: **Prediction of total genetic value using genome-wide dense marker maps.** *Genetics* 2001, **157**:1819–1829.
- Loberg A, Dürr JW: **Interbull survey on the use of genomic information.** *Interbull Bull* 2009, **39**:3–14.
- Brune T: **Stand und Perspektiven der Genomischen Selection beim Schwein.** *Fakultät Agrarwissenschaften und Landschaftsarchitektur, Hochschule Osnabrück* 2011. [Bachelorarbeit].
- Fulton JE: **Genomic selection for poultry breeding.** *Anim Front* 2012, **2**:30–36.
- Ibáñez-Escriche N, Fernando RL, Toosi A, Dekkers JCM: **Genomic selection of purebreds for crossbred performance.** *Genet Sel Evol* 2009, **41**:12.
- Kinghorn BP, Hickey JM, van der Werf JHJ: **Reciprocal recurrent genomic selection for total genetic merit in crossbred individuals.** In *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production*. Edited by German Society of Animal Science. Leipzig; 2010. paper 0036. [<http://www.kongressband.de/wcgalp2010/assets/pdf/0036.pdf>]
- Zeng J, Toosi A, Fernando RL, Dekkers JCM, Garrick DJ: **Genomic selection of purebred animals for crossbred performance in the presence of dominant gene action.** *Genet Sel Evol* 2013, **45**:11.
- Legarra A, Aguilar I, Misztal I: **A relationship matrix including full pedigree and genomic information.** *J Dairy Sci* 2009, **92**:4656–4663.
- Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ: **Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluations of Holstein final score.** *J Dairy Sci* 2010, **93**:743–752.
- Christensen OF, Lund MS: **Genomic prediction when some animals are not genotyped.** *Genet Sel Evol* 2010, **42**:2.
- Henderson CR: **Sire evaluations and genetic trends.** *J Anim Sci* 1973, **Symposium**:10–41.
- VanRaden PM: **Efficient methods to compute genomic predictions.** *J Dairy Sci* 2008, **91**:4414–4423.
- Gao H, Christensen OF, Madsen P, Nielsen US, Zhang Y, Lund MS, Su G: **Comparison on genomic predictions using GBLUP models and two single-step blending methods with different relationship matrices in the Nordic Holstein population.** *Genet Sel Evol* 2012, **44**:8.
- Forni S, Aguilar I, Misztal I: **Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information.** *Genet Sel Evol* 2011, **43**:1.
- Christensen OF, Madsen P, Nielsen B, Ostersen T, Su G: **Single-step methods for genomic evaluation in pigs.** *Animal* 2012, **6**:1565–1571.
- Chen CY, Misztal I, Aguilar I, Tsuruta S, Meuwissen THE, Aggrey SE, Wind T, Muir WM: **Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one-step: an example using broiler chickens.** *J Anim Sci* 2010, **89**:23–28.
- Misztal I, Vitezica ZG, Legarra A, Aguilar I, Swan AA: **Unknown-parent groups in single-step genomic evaluation.** *J Anim Breed Genet* 2013, **130**:252–258.
- García-Cortés LA, Toro MA: **Multibreed analysis by splitting the breeding values.** *Genet Sel Evol* 2006, **38**:601–615.
- Hill WG, Goddard ME, Visscher PM: **Data and theory point to mainly additive genetic variance for complex traits.** *PLoS Genet* 2008, **4**:e1000008.
- Wei M, van der Werf JHJ, Brascamp EW: **Relationship between purebred and crossbred parameters: II Genetic correlation between purebred and crossbred performance under the model with two loci.** *J Anim Breed Genet* 1991, **108**:262–269.
- Henderson CR: **A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values.** *Biometrics* 1976, **32**:69–83.



24. Munilla Leguizamón S, Cantet RJC: **Equivalence of multibreed animal models and hierarchical Bayes analysis for maternally influenced traits.** *Genet Sel Evol* 2010, **42**:20.
25. Quaas RL: **Computing the diagonal elements and inverse of a large numerator relationship matrix.** *Biometrics* 1976, **32**:949–953.
26. Meuwissen THE, Luo Z: **Computing inbreeding coefficients in large populations.** *Genet Sel Evol* 1992, **24**:305–313.
27. Vitezica ZG, Aguilar I, Misztal I, Legarra A: **Bias in genomic predictions for populations under selection.** *Genet Res* 2011, **93**:357–366.
28. Christensen OF: **Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation.** *Genet Sel Evol* 2012, **44**:37.
29. Colleau JJ: **An indirect approach to the extensive calculation of relationship coefficients.** *Genet Sel Evol* 2002, **34**:409–421.
30. Misztal I, Legarra A, Aguilar I: **Computing procedures for genetic evaluation including phenotypic, full pedigree and genomic information.** *J Dairy Sci* 2009, **92**:4648–4655.
31. Hickey JM, Kinghorn BP, Tier B, Wilson JF, Dunstan N, van der Werf JHJ: **A combined long-range phasing and long haplotype imputation method to impute phase for SNP genotypes.** *Genet Sel Evol* 2011, **43**:12.
32. Browning SR, Browning BL: **Haplotype phasing: existing methods and new developments.** *Nat Rev Genet* 2011, **12**:703–714.
33. Ibánñez-Escriche N, Reixach J, Lleonart N, Noguera JL: **Genetic evaluation combining purebred and crossbred data in a pig breeding scheme.** *J Anim Sci* 2011, **89**:3881–3889.
34. Su G, Christensen OF, Ostersen T, Henryon M, Lund MS: **Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers.** *PLoS ONE* 2012, **7**:e45293.
35. Lo LL, Fernando RL, Grossman M: **Genetic evaluation by BLUP in two-breed terminal crossbreeding systems under dominance.** *J Anim Sci* 1997, **75**:2877–2884.
36. 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, **79**:3002–3007.
37. Lutaaya E, Misztal I, Mabry JW, Short T, Timm HH, Holzbauer R: **Joint evaluation of purebreds and crossbreds in swine.** *J Anim Sci* 2002, **80**:2263–2266.
38. Powell JE, Visscher PM, Goddard ME: **Reconciling the analysis of IBD and IBS in complex trait studies.** *Nat Rev Genet* 2010, **11**:800–805.

doi:10.1186/1297-9686-46-23

**Cite this article as:** Christensen *et al.*: Genomic evaluation of both purebred and crossbred performances. *Genetics Selection Evolution* 2014 **46**:23.

Submit your next manuscript to BioMed Central  
and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

