

RESEARCH ARTICLE

Open Access



Genetic evaluation for three-way crossbreeding

Ole F. Christensen^{1*}, Andres Legarra², Mogens S. Lund¹ and Guosheng Su¹

Abstract

Background: Commercial pig producers generally use a terminal crossbreeding system with three breeds. Many pig breeding organisations have started to use genomic selection for which genetic evaluation is often done by applying single-step methods for which the pedigree-based additive genetic relationship matrix is replaced by a combined relationship matrix based on both marker genotypes and pedigree. Genomic selection is implemented for purebreds, but it also offers opportunities for incorporating information from crossbreds and selecting for crossbred performance. However, models for genetic evaluation for the three-way crossbreeding system have not been developed.

Results: Four-variate models for three-way terminal crossbreeding are presented in which the first three variables contain the records for the three pure breeds and the fourth variable contains the records for the three-way crossbreds. For purebred animals, the models provide breeding values for both purebred and crossbred performances. Heterogeneity of genetic architecture between breeds and genotype by environment interactions are modelled through genetic correlations between these breeding values. Specification of the additive genetic relationships is essential for these models and can be defined either within populations or across populations. Based on these two types of additive genetic relationships, both pedigree-based, marker-based and combined relationships based on both pedigree and marker information are presented. All these models for three-way crossbreeding can be formulated using Kronecker matrix products and therefore fitted using Henderson's mixed model equations and standard animal breeding software.

Conclusions: Models for genetic evaluation in the three-way crossbreeding system are presented. They provide estimated breeding values for both purebred and crossbred performances, and can use pedigree-based or marker-based relationships, or combined relationships based on both pedigree and marker information. This provides a framework that allows information from three-way crossbred animals to be incorporated into a genetic evaluation system.

Background

Commercial pig producers generally use a terminal crossbreeding system with three breeds. In this system, F1 sows from two maternal breeds are mated to purebred boars from a breed that has high-level production traits (growth, leanness, feed efficiency) to produce pigs for slaughter. Commonly, boar lines in Europe are Duroc and Pietrain and sows are crosses between Large White and Landrace. Genetic evaluation is usually done within

each of these breeds based on recorded phenotypes on purebred animals. However, ideally genetic evaluation of purebreds should incorporate phenotypes of interest recorded on crossbreds, and breeding values for performance in the three-way cross should be estimated.

Many pig breeding organisations have started to use genomic selection [1], for which genetic evaluation is often done by applying single-step methods [2–4] to handle the fact that only a fraction of the animals are genotyped. Here, the pedigree-based additive genetic relationship matrix is replaced by a combined relationship matrix based on both marker genotypes and pedigree. Genomic selection is implemented for purebreds, but it also offers opportunities for incorporating

*Correspondence: olef.christensen@mbg.au.dk

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

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

information from crossbreds and selecting for crossbred performance [5–7].

For two-way terminal crossbreeding (two breeds named \mathcal{A} and \mathcal{B} , and all crossbred animals \mathcal{AB} have known purebred parents), Wei and van der Werf [8] proposed the following trivariate model:

$$\begin{aligned} \mathbf{y}_A &= \mathbf{X}_A \boldsymbol{\beta}_A + \mathbf{Z}_A \mathbf{a}_A + \mathbf{e}_A, \\ \mathbf{y}_B &= \mathbf{X}_B \boldsymbol{\beta}_B + \mathbf{Z}_B \mathbf{a}_B + \mathbf{e}_B, \\ \mathbf{y}_{AB} &= \mathbf{X}_{AB} \boldsymbol{\beta}_{AB} + \mathbf{g}_{AB} + \mathbf{e}_{AB}, \end{aligned} \quad (1)$$

where the vectors \mathbf{y}_A , \mathbf{y}_B and \mathbf{y}_{AB} contain phenotypes on animals from breeds \mathcal{A} and \mathcal{B} and from the cross \mathcal{AB} , respectively, and for the three populations \mathcal{A} , \mathcal{B} and \mathcal{AB} , the vectors $\boldsymbol{\beta}_A$, $\boldsymbol{\beta}_B$ and $\boldsymbol{\beta}_{AB}$ contain fixed effects (note that intercepts should always be included!), and $\mathbf{e}_A \sim N(\mathbf{0}, \sigma_{e,A}^2 \mathbf{I})$, $\mathbf{e}_B \sim N(\mathbf{0}, \sigma_{e,B}^2 \mathbf{I})$ and $\mathbf{e}_{AB} \sim N(\mathbf{0}, \sigma_{e,AB}^2 \mathbf{I})$ are the residual error vectors. The vectors \mathbf{a}_A and \mathbf{a}_B contain breeding values for purebred performance (mating within breed) for breeds \mathcal{A} and \mathcal{B} , respectively, and the vector of genetic values on the crossbreds, \mathbf{g}_{AB} , is related to the vectors of breeding values on purebred animals for crossbred performance (mating with the other breed), \mathbf{g}_A and \mathbf{g}_B , by additive pedigree-based relationships (throughout this paper, additive genetic effects for purebred performance and for crossbred performance are denoted by \mathbf{a} and \mathbf{g} , respectively). Each animal has then two breeding values (one related to mating within breed, e.g. \mathbf{a}_A , and another related to mating to another breed to produce the cross, e.g. \mathbf{g}_A) and these are correlated. Genetic correlations less than 1 are due to the presence of non-additive gene action in combination with different allele frequencies in the two breeds [9, 10], but also to genotype by environment interactions. The model also assumes different genetic variances in the two pure breeds, which is often the case in practice. Christensen et al. [11] reformulated the model using partial relationship matrices (see below) and constructed those from a combination of marker genotypes and pedigree in such a way that it could be fitted by using standard animal breeding software, i.e. a single-step method was developed.

The aim of this work was to develop models for three-way terminal crosses that handle both pedigree-based and marker-based relationships, as well as combined relationship matrices based on both pedigree and marker genotypes. As indicated above, an essential part of the model is the specification of relationships such that the model can be fitted by using standard animal breeding software.

Methods

We present a specific scenario with records on all three pure breeds and on three-way production pigs, but not on two-way crossbred sows, having in mind production traits such as daily gain, leanness or feed efficiency.

However, since we will specify relationships across all five populations, it is straightforward to generalise to other scenarios with records.

The model for this three-way terminal crossbreeding system is in principle a straightforward generalisation of the Wei and van der Werf model [8] to the following four-variate model:

$$\begin{aligned} \mathbf{y}_A &= \mathbf{X}_A \boldsymbol{\beta}_A + \mathbf{Z}_A \mathbf{a}_A + \mathbf{e}_A, \\ \mathbf{y}_B &= \mathbf{X}_B \boldsymbol{\beta}_B + \mathbf{Z}_B \mathbf{a}_B + \mathbf{e}_B, \\ \mathbf{y}_C &= \mathbf{X}_C \boldsymbol{\beta}_C + \mathbf{Z}_C \mathbf{a}_C + \mathbf{e}_C, \\ \mathbf{y}_{C(AB)} &= \mathbf{X}_{C(AB)} \boldsymbol{\beta}_{C(AB)} + \mathbf{g}_{C(AB)} + \mathbf{e}_{C(AB)}, \end{aligned} \quad (2)$$

where notation is defined as for Eq. (1), and it is assumed that all $C(AB)$ animals have a purebred C father and crossbred AB mother, and that these AB animals all have purebred parents. Breed C animals have two breeding values that are correlated, \mathbf{a}_C for purebred performance (mating within breed) and \mathbf{g}_C for $C(AB)$ crossbred performance (mating between a male and a AB crossbred female). Breed A animals also have two breeding values, \mathbf{a}_A for purebred performance (mating within breed) and \mathbf{g}_A for crossbred $C(AB)$ performance (mating with a breed B animal whose female AB crossbred offspring is mated with a breed C male). Finally, breed B animals have two breeding values \mathbf{a}_B and \mathbf{g}_B , defined similarly to the breeding values for breed A . For each breed, association between breeding values for purebred and crossbred performances is determined by a 2×2 genetic variance-covariance matrix. An essential part of the model is the specification of the additive relationships between genetic values for crossbred performance on crossbred animals and purebred animals, and in particular marker-based versions of these relationships such that pedigree-based and marker-based relationships are consistent. These relationships should also be specified in such a way that the model can be formulated using Kronecker products, allowing the model to be fitted by using Henderson's mixed model equations and standard animal breeding software. Additive relationships are relationships between gene substitution effects and these can be defined either within populations or across populations [12]. These two approaches will be called "partial genetic" and "common genetic" approaches in the following.

Lo et al. [13] derived the following recursive formulas for the variance and covariance of genotypic values for animals composed of multiple breeds under an additive model. Let the genotypic value of individual i be g_i , then the additive variance is:

$$\begin{aligned} \text{Var}(g_i) &= \sum_b f_i^b \sigma_{g,b}^2 + \text{Cov}(g_{f(i)}, g_{m(i)}) \\ &+ 2 \sum_b \sum_{b' > b} (f_{f(i)}^b f_{f(i)}^{b'} + f_{m(i)}^b f_{m(i)}^{b'}) \sigma_{g,b,b'}^2 \end{aligned} \quad (3)$$

where b and b' denote breeds, f_i^b is the breed b content of individual i , $\sigma_{g,b}^2$ is the breed b genetic variance, $g_{f(i)}$ and $g_{m(i)}$ are the additive genetic values of parents $f(i)$ and $m(i)$, respectively, and $\sigma_{g,b,b'}^2$ is the breed b and breed b' segregation genetic variance. The additive covariance between genotypic values of individuals i and i' is:

$$\text{Cov}(g_i, g_{i'}) = (\text{Cov}(g_{f(i)}, g_{f(i')}) + \text{Cov}(g_{m(i)}, g_{m(i')}))/2, \tag{4}$$

when $i' \neq i$ is not a descendant of i .

García-Cortés and Toro [14] showed that Eqs. (3) and (4) could be expressed as (using matrix notation):

$$\text{Var}(\mathbf{g}) = \sum_b \sigma_b^2 \mathbf{A}^b + \sum_b \sum_{b' > b} \sigma_{b,b'}^2 \mathbf{A}^{b,b'}$$

where the \mathbf{A}^b and $\mathbf{A}^{b,b'}$ matrices are separately defined using recursions, and that this provides a partition of the vector of genotypic values into:

$$\mathbf{g} = \sum_b \mathbf{g}^b + \sum_b \sum_{b' > b} \mathbf{g}^{b,b'}$$

where all the \mathbf{g}^b , $\mathbf{g}^{b,b'}$ vectors are independent, $\text{Var}(\mathbf{g}^b) = \sigma_b^2 \mathbf{A}^b$ and $\text{Var}(\mathbf{g}^{b,b'}) = \sigma_{b,b'}^2 \mathbf{A}^{b,b'}$. They termed matrix \mathbf{A}^b as the breed b specific partial relationship matrix and matrix $\mathbf{A}^{b,b'}$ the breed b and breed b' segregation partial relationship matrix. The vectors \mathbf{g}^b and $\mathbf{g}^{b,b'}$ depend on genetic origin, such that \mathbf{g}^b is the breed b specific partial genetic vector, and $\mathbf{g}^{b,b'}$ is the breed b and breed b' segregation partial genetic vector. Matrices \mathbf{A}^b and $\mathbf{A}^{b,b'}$ have sparse inverses that can be computed using the usual methods for the additive relationship matrix (see [14]). In this paper, the approach using a partition of the genetic effects into independent terms is named partial genetic approach.

Legarra et al. [15] proposed that pedigree relationships should be specified across all animals, and that for base animals in the pedigree, the pedigree-based relationships within and across breeds and inbreeding should be estimated from observed marker genotypes. This approach is contradictory to the García-Cortés and Toro [14] approach described above, since it violates the assumption of independence of the \mathbf{g}^b and $\mathbf{g}^{b,b'}$ vectors. The approach in which relationships are specified across breeds is named common genetic approach.

First, partial genetic and common genetic approaches for constructing pedigree-based relationships are presented, then the corresponding two different ways of constructing marker-based relationships are presented, and finally the genetic variances and covariances in model (2) are shown for the two approaches. Detailed derivations are in the "Appendix".

Additive genetic model for crossbred $\mathcal{C}(\mathcal{AB})$ performance: partial genetic approach

For the three-way crossbreeding system, the decomposition of the additive genetic effects by García-Cortés and Toro [14] is as follows. For a $\mathcal{C}(\mathcal{AB})$ crossbred animal,

$$g_i = g_{\mathcal{C}(\mathcal{AB}),i}^{\mathcal{C}} + g_{\mathcal{C}(\mathcal{AB}),i}^{\mathcal{A}} + g_{\mathcal{C}(\mathcal{AB}),i}^{\mathcal{B}} + g_{\mathcal{C}(\mathcal{AB}),i}^{\mathcal{AB}}$$

where terms $\mathbf{g}_{\mathcal{C}(\mathcal{AB})}^{\mathcal{C}}$, $\mathbf{g}_{\mathcal{C}(\mathcal{AB})}^{\mathcal{A}}$, $\mathbf{g}_{\mathcal{C}(\mathcal{AB})}^{\mathcal{B}}$ are breed of origin specific partial genetic effects and $\mathbf{g}_{\mathcal{C}(\mathcal{AB})}^{\mathcal{AB}}$ is a breed-segregation term. For a \mathcal{AB} crossbred sow,

$$g_i = g_{\mathcal{AB},i}^{\mathcal{A}} + g_{\mathcal{AB},i}^{\mathcal{B}}$$

with terms $\mathbf{g}_{\mathcal{AB}}^{\mathcal{A}}$ and $\mathbf{g}_{\mathcal{AB}}^{\mathcal{B}}$ being breed of origin partial genetic effects. Finally, for purebred animals, the three vectors of breeding values for crossbred $\mathcal{C}(\mathcal{AB})$ performance, $\mathbf{g}_{\mathcal{A}}$, $\mathbf{g}_{\mathcal{B}}$ and $\mathbf{g}_{\mathcal{C}}$, are defined as being equal to the genotypic values.

In this way, a breed-specific partial genetic effect is defined for all animals containing the specific breed, and a breed-segregation partial genetic effect is defined for crossbred $\mathcal{C}(\mathcal{AB})$ animals. Assuming that base individuals in the three breeds are not related across breeds implies that:

$$\mathbf{g}^{\mathcal{A}} = \begin{bmatrix} \mathbf{g}^{\mathcal{A}} \\ \mathbf{g}_{\mathcal{AB}}^{\mathcal{A}} \\ \mathbf{g}_{\mathcal{C}(\mathcal{AB})}^{\mathcal{A}} \end{bmatrix}, \mathbf{g}^{\mathcal{B}} = \begin{bmatrix} \mathbf{g}^{\mathcal{B}} \\ \mathbf{g}_{\mathcal{AB}}^{\mathcal{B}} \\ \mathbf{g}_{\mathcal{C}(\mathcal{AB})}^{\mathcal{B}} \end{bmatrix}, \mathbf{g}^{\mathcal{C}} = \begin{bmatrix} \mathbf{g}^{\mathcal{C}} \\ \mathbf{g}_{\mathcal{C}(\mathcal{AB})}^{\mathcal{C}} \end{bmatrix}$$

are independent. In addition, for a crossbred $\mathcal{C}(\mathcal{AB})$ individual the fact that it inherits either a breed \mathcal{A} or \mathcal{B} allele is independent of what particular alleles the \mathcal{AB} mother has and what alleles all other \mathcal{AB} individuals have, and hence $\mathbf{g}_{\mathcal{C}(\mathcal{AB})}^{\mathcal{AB}}$ is independent of the vectors above.

The variance-covariance matrices of the partial genetic effects become (García-Cortés and Toro [14]):

$$\text{Var} \begin{bmatrix} \mathbf{g}^{\mathcal{A}} \\ \mathbf{g}_{\mathcal{AB}}^{\mathcal{A}} \\ \mathbf{g}_{\mathcal{C}(\mathcal{AB})}^{\mathcal{A}} \end{bmatrix} = \sigma_{g,\mathcal{A}}^2 \mathbf{A}^{\mathcal{A}}, \text{Var} \begin{bmatrix} \mathbf{g}^{\mathcal{B}} \\ \mathbf{g}_{\mathcal{AB}}^{\mathcal{B}} \\ \mathbf{g}_{\mathcal{C}(\mathcal{AB})}^{\mathcal{B}} \end{bmatrix} = \sigma_{g,\mathcal{B}}^2 \mathbf{A}^{\mathcal{B}},$$

$$\text{Var} \begin{bmatrix} \mathbf{g}^{\mathcal{C}} \\ \mathbf{g}_{\mathcal{C}(\mathcal{AB})}^{\mathcal{C}} \end{bmatrix} = \sigma_{g,\mathcal{C}}^2 \mathbf{A}^{\mathcal{C}}, \text{Var}[\mathbf{g}_{\mathcal{C}(\mathcal{AB})}^{\mathcal{AB}}] = \sigma_{g,\mathcal{AB}}^2 \mathbf{A}^{\mathcal{AB}},$$

where the breed-specific partial relationship matrices are defined by the recursive formulas:

$$A_{ii}^b = f_i^b + A_{f(i)m(i)}^b/2,$$

$$A_{ii'}^b = (A_{f(i)i'}^b + A_{m(i)i'}^b)/2,$$

for breed $b = \mathcal{A}, \mathcal{B}, \mathcal{C}$, with f_i^b denoting the breed b proportion, and the breed-segregation partial relationship matrix is defined by the recursive formulas:

$$A_{ii}^{AB} = 2(f_{f(i)}^A f_{f(i)}^B + f_{m(i)}^A f_{m(i)}^B) + A_{f(i)m(i)}^{AB}/2,$$

$$A_{i i'}^{AB} = (A_{f(i) i'}^{AB} + A_{m(i) i'}^{AB})/2,$$

where in both cases non-contributing animals are not included in the resulting matrices. We immediately see that $\mathcal{C}(\mathcal{AB})$ animals are the only animals contributing to matrix \mathbf{A}^{AB} , and since $f_{m(i)}^A = f_{m(i)}^B = 1/2$ for these animals, the matrix is a diagonal matrix with diagonal elements equal to $2 \times 1/2 \times 1/2 = 1/2$. This specification of additive relationships using partial relationship matrices is equivalent to the specification in Eqs. (3) and (4).

To illustrate the different partial relationship matrices, we analysed the small pedigree in Table 1. Tables 2, 3, 4 and 5 show the partial relationship matrices for this example.

Wei and van der Werf [8] presented a reduced form of the two-way crossbreeding model (1) in which the Mendelian sampling term of the genetic effect on crossbred animals was included in the residual error term. A reduced model can also be formulated for the three-way crossbreeding model by expressing:

$$g_i = 0.5g_{f(i)} + 0.5g_{m(i)} + \Phi_{\mathcal{C}(\mathcal{AB}),i}$$

$$= 0.5g_{\mathcal{C},f(i)} + 0.5g_{\mathcal{AB},m(i)}^A + 0.5g_{\mathcal{AB},m(i)}^B + \Phi_{\mathcal{C}(\mathcal{AB}),i}$$

Table 1 Example pedigree

id	Father	Mother	Population
1	0	0	\mathcal{A}
2	0	0	\mathcal{A}
3	0	0	\mathcal{B}
4	0	0	\mathcal{C}
5	1	2	\mathcal{A}
6	3	5	\mathcal{AB}
7	4	6	$\mathcal{C}(\mathcal{AB})$
8	4	6	$\mathcal{C}(\mathcal{AB})$

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

id	1	2	5	6	7	8
1	1	0	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
2		1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
5			1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$
6				1	$\frac{1}{4}$	$\frac{1}{4}$
7					1	$\frac{1}{8}$
8						1

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

id	3	6	7	8
3	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$
6		1	$\frac{1}{4}$	$\frac{1}{4}$
7			1	$\frac{1}{8}$
8				1

Table 4 Breed \mathcal{C} specific partial relationship matrix $\mathbf{A}^{\mathcal{C}}$ for the pedigree in Table 1

id	4	7	8
4	1	$\frac{1}{2}$	$\frac{1}{2}$
7		1	$\frac{1}{4}$
8			1

Table 5 Breed \mathcal{AB} segregation partial relationship matrix $\mathbf{A}^{\mathcal{AB}}$ for the pedigree in Table 1

id	9	10
9	$\frac{1}{2}$	0
10		$\frac{1}{2}$

for a $\mathcal{C}(\mathcal{AB})$ crossbred animal i , where $\Phi_{\mathcal{C}(\mathcal{AB}),i}$ is the Mendelian sampling term. The Mendelian sampling terms are independent among the $\mathcal{C}(\mathcal{AB})$ crossbred animals, and by making the approximation that father $f(i)$ is not inbred and since mother $m(i)$ is not inbred, the variance is constant. In this way, the Mendelian sampling error term can be included into the residual error term $\mathbf{e}_{\mathcal{C}(\mathcal{AB})}$ in model (2), and the model can be formulated using three breed-specific partial relationship matrices defined on the $\mathcal{A}, \mathcal{B}, \mathcal{C}$ and \mathcal{AB} animals. However, as explained in Christensen et al. [11], such a reduced model cannot be extended to incorporate marker genotypes since these provide information about the Mendelian sampling term. Therefore, we did not pursue the reduced form of the model any further.

Note that model (2) with relationships as presented here is the most obvious generalisation of the Wei and van der Werf model in Eq. (1) from two to three breeds since base individuals are assumed unrelated. Without a formulation using partial relationship matrices, it would be difficult to estimate parameters in this model using standard animal breeding software.

Additive genetic model for crossbred $\mathcal{C}(\mathcal{AB})$ performance: common genetic approach

In the previous subsection, base animals were assumed to be unrelated. An alternative proposed by Legarra et al.

[15] is to assume that base animals are related and inbred within breeds and related between breeds with relationships determined by:

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_A & \gamma_{A,B} & \gamma_{A,C} \\ \gamma_{A,B} & \gamma_B & \gamma_{B,C} \\ \gamma_{A,C} & \gamma_{B,C} & \gamma_C \end{bmatrix}.$$

This means that among the base animals, the variance-covariance of genetic effects is as follows. The variance-covariance within breed is defined by:

$$\text{Var}(g_i) = \sigma_g^2(1 + \gamma_b/2),$$

for an individual in breed b , and

$$\text{Cov}(g_i, g_{i'}) = \sigma_g^2 \gamma_b,$$

for two individuals in breed b , i.e. base animals are inbred with coefficient $\gamma_b/2$ and related with relationship coefficient γ_b . Furthermore,

$$\text{Cov}(g_i, g_{i'}) = \sigma_g^2 \gamma_{b,b'},$$

for two individuals in different breeds b and b' , i.e. base animals in different breeds are related. Therefore, a joint relationship matrix is specified among all base animals, and by applying the usual recursive definition:

$$A(\mathbf{\Gamma})_{ii} = 1 + A(\mathbf{\Gamma})_{f(i)m(i)}/2,$$

$$A(\mathbf{\Gamma})_{i'i'} = (A(\mathbf{\Gamma})_{f(i)i'} + A(\mathbf{\Gamma})_{m(i)i'})/2,$$

an additive relationship matrix $\mathbf{A}(\mathbf{\Gamma})$ is defined across all animals with relationships among the three base populations \mathcal{A} , \mathcal{B} and \mathcal{C} defined by matrix $\mathbf{\Gamma}$. The variance-covariance of genetic effects is therefore determined by

$$\text{Var} \begin{bmatrix} \mathbf{g}_A \\ \mathbf{g}_B \\ \mathbf{g}_C \\ \mathbf{g}_{AB} \\ \mathbf{g}_{C(AB)} \end{bmatrix} = \sigma_g^2 \mathbf{A}(\mathbf{\Gamma}). \tag{5}$$

Table 6 shows the common relationship matrix for the pedigree in Table 1.

Legarra et al. [15] suggested a framework where individuals in the base population of the pedigree are related because they originate from overlapping ancestral populations with a finite size, and they termed each of these ancestral populations as a meta-founder to be included in the pedigree. Here, $\mathcal{A}, \mathcal{B}, \mathcal{C}$ are meta-founders, and each base individual in the pedigree has a meta-founder, which is both its parents; see example in Table 7. When extending the pedigree and the matrix $\mathbf{A}(\mathbf{\Gamma})$ with these meta-founders, Legarra et al. [15] showed that the algorithms for computing the sparse inverse matrix $\mathbf{A}(\mathbf{\Gamma})^{-1}$ directly as in Henderson [16] and submatrices of $\mathbf{A}(\mathbf{\Gamma})$ by the Colleau algorithm [17] are as usual.

The parameter σ_g^2 in Eq. (5) does not correspond to the usual genetic variance which is the variance among unrelated individuals in the base population. As explained in Legarra et al. [15], $\sigma_g^2(1 - \gamma_b/2)$ corresponds to the variance among unrelated breed b animals, and therefore the genetic variances for crossbred $\mathcal{C}(\mathcal{A}\mathcal{B})$ performance are $\sigma_g^2(1 - \gamma_A/2)$, $\sigma_g^2(1 - \gamma_B/2)$ and $\sigma_g^2(1 - \gamma_C/2)$, corresponding to $\sigma_{g,A}^2$, $\sigma_{g,B}^2$ and $\sigma_{g,C}^2$ in the previous section, respectively. In addition, Legarra et al. [15] explained that the breed-segregation variance is $\sigma_g^2((\gamma_A + \gamma_B)/2 - \gamma_{A,B})/4$, which corresponds to $\sigma_{g,AB}^2$ in the previous section.

Genomic model for crossbred $\mathcal{C}(\mathcal{A}\mathcal{B})$ performance: partial genetic approach

Marker-based partial relationship matrices are constructed by tracing breed of origin of alleles and defining relationships according to breed of origin. Assume that breed of origin of alleles can be determined for all animals and define breed-specific allele content matrices as: matrix \mathbf{m}^b with entries 0, 1, 2 for purebred b animals, matrices \mathbf{z}^A and \mathbf{z}^B with entries 0, 1 for paternal

Table 6 Common relationship matrix $\mathbf{A}(\mathbf{\Gamma})$ for the pedigree in Table 1

id	1	2	3	4	5	6	7	8
1	$1 + \frac{\gamma_A}{2}$	γ_A	γ_{AB}	γ_{AC}	$\frac{1}{2} + \frac{3\gamma_A}{4}$	$\frac{1}{4} + \frac{3\gamma_A}{8} + \frac{\gamma_{AB}}{2}$	$\frac{1}{8} + \frac{3\gamma_A}{16} + \frac{\gamma_{AB}}{4} + \frac{\gamma_{AC}}{2}$	$\frac{1}{8} + \frac{3\gamma_A}{16} + \frac{\gamma_{AB}}{4} + \frac{\gamma_{AC}}{2}$
2		$1 + \frac{\gamma_A}{2}$	γ_{AB}	γ_{AC}	$\frac{1}{2} + \frac{3\gamma_A}{4}$	$\frac{1}{4} + \frac{3\gamma_A}{8} + \frac{\gamma_{AB}}{2}$	$\frac{1}{8} + \frac{3\gamma_A}{16} + \frac{\gamma_{AB}}{4} + \frac{\gamma_{AC}}{2}$	$\frac{1}{8} + \frac{3\gamma_A}{16} + \frac{\gamma_{AB}}{4} + \frac{\gamma_{AC}}{2}$
3			$1 + \frac{\gamma_B}{2}$	γ_{BC}	γ_{AB}	$\frac{1}{2} + \frac{\gamma_B}{4} + \frac{\gamma_{AB}}{2}$	$\frac{1}{4} + \frac{\gamma_B}{8} + \frac{\gamma_{AB}}{4} + \frac{\gamma_{AC}}{2}$	$\frac{1}{4} + \frac{\gamma_B}{8} + \frac{\gamma_{AB}}{4} + \frac{\gamma_{AC}}{2}$
4				$1 + \frac{\gamma_C}{2}$	γ_{AC}	$\frac{\gamma_{AC} + \gamma_{BC}}{2}$	$\frac{1}{2} + \frac{\gamma_C + \gamma_{AC} + \gamma_{BC}}{4}$	$\frac{1}{2} + \frac{\gamma_C + \gamma_{AC} + \gamma_{BC}}{4}$
5					$1 + \frac{\gamma_A}{2}$	$\frac{1}{2} + \frac{\gamma_A}{4} + \frac{\gamma_{AB}}{2}$	$\frac{1}{4} + \frac{\gamma_A}{8} + \frac{\gamma_{AB}}{4} + \frac{\gamma_{AC}}{2}$	$\frac{1}{4} + \frac{\gamma_A}{8} + \frac{\gamma_{AB}}{4} + \frac{\gamma_{AC}}{2}$
6						$1 + \frac{\gamma_{AB}}{2}$	$\frac{1}{2} + \frac{\gamma_{AB} + \gamma_{AC} + \gamma_{BC}}{4}$	$\frac{1}{2} + \frac{\gamma_{AB} + \gamma_{AC} + \gamma_{BC}}{4}$
7							$1 + \frac{\gamma_{AC} + \gamma_{BC}}{4}$	$\frac{1}{2} + \frac{\gamma_C + \gamma_{AB}}{8} + \frac{\gamma_{AC} + \gamma_{BC}}{4}$
8								$1 + \frac{\gamma_{AC} + \gamma_{BC}}{4}$

Table 7 Pedigree in Table 1 with metafounders

id	father	mother
\mathcal{A}	-	-
\mathcal{B}	-	-
\mathcal{C}	-	-
1	\mathcal{A}	\mathcal{A}
2	\mathcal{A}	\mathcal{A}
3	\mathcal{B}	\mathcal{B}
4	\mathcal{C}	\mathcal{C}
5	1	2
6	3	5
7	4	6
8	4	6

and maternal alleles, respectively, for crossbred \mathcal{AB} animals, matrix $\mathbf{z}^{\mathcal{C}}$ with entries 0, 1 for paternal allele of crossbred $\mathcal{C}(\mathcal{AB})$ animals, and finally matrices $\mathbf{z}_p^{\mathcal{A}}$ and $\mathbf{z}_p^{\mathcal{B}}$ with entries 0, 1, respectively, for crossbred $\mathcal{C}(\mathcal{AB})$ animals when the breed-specific allele is inherited and zero otherwise. This means that breed of origin of each allele needs to be traced, usually by a phasing software [18].

Marker-based breed-specific partial relationship matrices are constructed as follows (details can be found in the “Appendix”). For breed \mathcal{A} , 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 \mathcal{AB} animals,

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

which are defined as

$$\begin{aligned} \mathbf{G}_{\mathcal{A},\mathcal{A}}^{\mathcal{A}} &= \frac{(\mathbf{m}^{\mathcal{A}} - 2\mathbf{p}^{\mathcal{A}}\mathbf{1}^{\mathcal{T}})(\mathbf{m}^{\mathcal{A}} - 2\mathbf{p}^{\mathcal{A}}\mathbf{1}^{\mathcal{T}})^{\mathcal{T}}}{s^{\mathcal{A}}}, \\ \mathbf{G}_{\mathcal{A},\mathcal{AB}}^{\mathcal{A}} &= \frac{(\mathbf{m}^{\mathcal{A}} - 2\mathbf{p}^{\mathcal{A}}\mathbf{1}^{\mathcal{T}})(\mathbf{z}^{\mathcal{A}} - \mathbf{p}^{\mathcal{A}}\mathbf{1}^{\mathcal{T}})^{\mathcal{T}}}{s^{\mathcal{A}}}, \\ \mathbf{G}_{\mathcal{A},\mathcal{C}(\mathcal{AB})}^{\mathcal{A}} &= \frac{(\mathbf{m}^{\mathcal{A}} - 2\mathbf{p}^{\mathcal{A}}\mathbf{1}^{\mathcal{T}})(\mathbf{z}_p^{\mathcal{A}} - \mathbf{p}_p^{\mathcal{A}})^{\mathcal{T}}}{s^{\mathcal{A}}}, \\ \mathbf{G}_{\mathcal{AB},\mathcal{AB}}^{\mathcal{A}} &= \frac{(\mathbf{z}^{\mathcal{A}} - \mathbf{p}^{\mathcal{A}}\mathbf{1}^{\mathcal{T}})(\mathbf{z}^{\mathcal{A}} - \mathbf{p}^{\mathcal{A}}\mathbf{1}^{\mathcal{T}})^{\mathcal{T}}}{s^{\mathcal{A}}}, \\ \mathbf{G}_{\mathcal{AB},\mathcal{C}(\mathcal{AB})}^{\mathcal{A}} &= \frac{(\mathbf{z}^{\mathcal{A}} - \mathbf{p}^{\mathcal{A}}\mathbf{1}^{\mathcal{T}})(\mathbf{z}_p^{\mathcal{A}} - \mathbf{p}_p^{\mathcal{A}})^{\mathcal{T}}}{s^{\mathcal{A}}}, \\ \mathbf{G}_{\mathcal{C}(\mathcal{AB}),\mathcal{C}(\mathcal{AB})}^{\mathcal{A}} &= \frac{(\mathbf{z}_p^{\mathcal{A}} - \mathbf{p}_p^{\mathcal{A}})(\mathbf{z}_p^{\mathcal{A}} - \mathbf{p}_p^{\mathcal{A}})^{\mathcal{T}}}{s^{\mathcal{A}}}, \end{aligned}$$

where the vector $\mathbf{p}^{\mathcal{A}}$ contains breed \mathcal{A} specific allele frequencies, matrix $\mathbf{p}_p^{\mathcal{A}}$ has elements (i, j) equal to $p_j^{\mathcal{A}}$ when the crossbred $\mathcal{C}(\mathcal{AB})$ individual i inherited an \mathcal{A} specific

allele and zero otherwise, and $s^{\mathcal{A}}$ is a scaling parameter. The marker-based breed \mathcal{B} specific partial relationship matrix $\mathbf{G}^{\mathcal{B}}$ is defined similarly to $\mathbf{G}^{\mathcal{A}}$, and the marker-based breed \mathcal{C} specific partial relationship matrix is

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

where submatrices are defined as

$$\begin{aligned} \mathbf{G}_{\mathcal{C},\mathcal{C}}^{\mathcal{C}} &= \frac{(\mathbf{m}^{\mathcal{C}} - 2\mathbf{p}^{\mathcal{C}}\mathbf{1}^{\mathcal{T}})(\mathbf{m}^{\mathcal{C}} - 2\mathbf{p}^{\mathcal{C}}\mathbf{1}^{\mathcal{T}})^{\mathcal{T}}}{s^{\mathcal{C}}}, \\ \mathbf{G}_{\mathcal{C},\mathcal{C}(\mathcal{AB})}^{\mathcal{C}} &= \frac{(\mathbf{m}^{\mathcal{C}} - 2\mathbf{p}^{\mathcal{C}}\mathbf{1}^{\mathcal{T}})(\mathbf{z}^{\mathcal{C}} - \mathbf{p}^{\mathcal{A}}\mathbf{1}^{\mathcal{T}})^{\mathcal{T}}}{s^{\mathcal{C}}}, \\ \mathbf{G}_{\mathcal{C}(\mathcal{AB}),\mathcal{C}(\mathcal{AB})}^{\mathcal{C}} &= \frac{(\mathbf{z}^{\mathcal{C}} - \mathbf{p}^{\mathcal{C}}\mathbf{1}^{\mathcal{T}})(\mathbf{z}^{\mathcal{C}} - \mathbf{p}^{\mathcal{C}}\mathbf{1}^{\mathcal{T}})^{\mathcal{T}}}{s^{\mathcal{C}}}, \end{aligned}$$

where the vector $\mathbf{p}^{\mathcal{C}}$ contains estimated breed \mathcal{C} specific allele frequencies and $s^{\mathcal{C}}$ is a scaling parameter.

The breed-segregation partial relationship matrix is defined as:

$$G_{i,i'}^{\mathcal{AB}} = \sum_j r_{d_i^j} r_{d_{i'}^j} / (2n),$$

where $r_{d_i^j} = 1$ when $d_i^j \in \mathcal{A}$ and $r_{d_i^j} = -1$ when $d_i^j \in \mathcal{B}$, $r_{d_{i'}^j}$ is defined similarly to $r_{d_i^j}$, and n is the number of markers. Note that diagonal elements of $\mathbf{G}^{\mathcal{AB}}$ equal diagonal elements of $\mathbf{A}^{\mathcal{AB}}$ (i.e. 1/2). Off-diagonal elements of $\mathbf{G}^{\mathcal{AB}}$ measure whether pairs of individuals share more alleles from a particular parental breed (\mathcal{A} or \mathcal{B}) than expected. Expectations of off-diagonal elements $\mathbf{G}^{\mathcal{AB}}$ equal off-diagonal elements of $\mathbf{A}^{\mathcal{AB}}$ (i.e. 0).

Relationship matrices that combine pedigree and marker information [2, 4] can then be constructed. Below, indices 1 and 2 in submatrices denote non-genotyped and genotyped animals, respectively. The breed $b = \mathcal{A}, \mathcal{B}, \mathcal{C}$ specific combined relationship matrices are given by their sparse inverses

$$(\mathbf{H}^b)^{-1} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\mathbf{G}^b)^{-1} - (\mathbf{A}_{22}^b)^{-1} \end{bmatrix} + (\mathbf{A}^b)^{-1}, \quad (6)$$

for $b = \mathcal{A}, \mathcal{B}, \mathcal{C}$, and because $\mathbf{A}^{\mathcal{AB}} = \mathbf{I}/2$ the breed-segregation combined relationship matrix is

$$\mathbf{H}^{\mathcal{AB}} = \begin{bmatrix} \mathbf{I}/2 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{\mathcal{AB}} \end{bmatrix}. \quad (7)$$

Matrices $(\mathbf{A}^{\mathcal{A}})^{-1}$, $(\mathbf{A}^{\mathcal{B}})^{-1}$ and $(\mathbf{A}^{\mathcal{C}})^{-1}$ can be computed directly in sparse format and matrices $\mathbf{A}_{22}^{\mathcal{A}}$, $\mathbf{A}_{22}^{\mathcal{B}}$ and $\mathbf{A}_{22}^{\mathcal{C}}$ can be computed by the Colleau algorithm [17]; see Christensen et al. [11].

The breed-specific partial marker-based relationship matrices above require estimates of breed-specific allele

frequencies. Such estimates can be obtained from marker genotypes of purebred animals and breed-specific marker alleles for crossbred animals. Furthermore, there is a need to adjust these matrices to be compatible with partial pedigree relationship matrices similar to Christensen et al. [11, 19], i.e. $\mathbf{G}_a^b = \mathbf{G}^b \beta_b + \alpha_b \mathbf{J}^b$ where α_b and β_b are parameters and \mathbf{J}^b is a matrix with entries $J_{i,i'}^b = f_i^b f_{i'}^b$. The scaling parameters s^b in marker-based relationship matrices \mathbf{G}^b , $b=A, B, C$ are unspecified above, since the compatibility adjustment involves a scaling parameter β_b for each breed, and therefore s^b can be arbitrary. On the other hand, matrix \mathbf{G}^{AB} does not need an adjustment.

Finally, to incorporate the fact that marker genotypes only capture a fraction of the genetic effects, the partial marker-based relationship matrices \mathbf{G}^b , $b \in A, B, C$ and \mathbf{G}^{AB} above may be replaced by matrices $\mathbf{G}_\omega^b = \mathbf{G}^b(1 - \omega) + \mathbf{A}^b \omega$, $b \in A, B, C$ and $\mathbf{G}^{AB}(1 - \omega) + \mathbf{A}^{AB} \omega$, respectively, where ω is the fraction of genetic variance not captured by marker genotypes [4].

Genomic model for crossbred $\mathcal{C}(AB)$ performance: common genetic approach

The marker-based relationship matrix is constructed as usual across all genotyped animals:

$$\mathbf{G} = \frac{(\mathbf{m} - \mathbf{1}\mathbf{1}^T)(\mathbf{m} - \mathbf{1}\mathbf{1}^T)^T}{s},$$

where \mathbf{m} is the gene content matrix with entries 0, 1, 2 and s is scaling parameter. As in Christensen [20] and Legarra et al. [15], we chose common allele frequencies, i.e. $p_j = 0.5$, and then determine the parameters in matrix $\mathbf{\Gamma}$ and parameter s such that the pedigree-based and marker-based relationship matrices are compatible. Parameters in matrix $\mathbf{\Gamma}$ and scaling parameter s can be estimated by matching $\mathbf{A}(\mathbf{\Gamma})$ and \mathbf{G} for purebred individuals; see Legarra et al. [15]. For example, if genotyping is done in each of the three pure breeds then the following system of equations can be used to determine the parameters:

$$\begin{aligned} \bar{G}_{A,B}/s &= \gamma_{A,B}, \quad \bar{G}_{A,C}/s = \gamma_{A,C}, \quad \bar{G}_{B,C}/s = \gamma_{B,C}, \\ \bar{G}_{A,A}/s &= \bar{A}_{A,A}(1 - \gamma_A/2) + \gamma_A, \\ \bar{G}_{B,B}/s &= \bar{A}_{B,B}(1 - \gamma_B/2) + \gamma_B, \\ \bar{G}/s &= (\text{diag}(\bar{A}_{A,A})(1 - \gamma_A/2) + \gamma_A)/3 \\ &\quad + (\text{diag}(\bar{A}_{B,B})(1 - \gamma_B/2) + \gamma_B)/3 \\ &\quad + (\text{diag}(\bar{A}_{C,C})(1 - \gamma_C/2) + \gamma_C)/3, \end{aligned}$$

where $\bar{G}_{A,A}$, $\bar{G}_{A,B}$, $\bar{G}_{A,C}$, $\bar{G}_{B,B}$, $\bar{G}_{B,C}$ and $\bar{G}_{C,C}$ denote averages of elements in submatrices of \mathbf{G} , $\bar{G} = (\bar{G}_{A,A} + \bar{G}_{B,B} + \bar{G}_{C,C})/3$, $\bar{A}_{A,A}$, $\bar{A}_{A,B}$, $\bar{A}_{A,C}$, $\bar{A}_{B,B}$, $\bar{A}_{B,C}$ and $\bar{A}_{C,C}$ denote averages of elements in submatrices of \mathbf{A}_{22} , and $\text{diag}(\bar{A}_{A,A})$, $\text{diag}(\bar{A}_{B,B})$, $\text{diag}(\bar{A}_{C,C})$ denote averages of diagonal elements in submatrices of \mathbf{A}_{22} .

This is a linear system of 7 equations with 7 parameters $\gamma_A, \gamma_B, \gamma_C, \gamma_{A,B}, \gamma_{A,C}, \gamma_{B,C}$ and $1/s$ and can therefore be solved directly to obtain estimates.

The relationship matrix that combines pedigree and marker information becomes

$$\mathbf{H}(\mathbf{\Gamma})^{-1} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - (\mathbf{A}(\mathbf{\Gamma})_{22})^{-1} \end{bmatrix} + \mathbf{A}(\mathbf{\Gamma})^{-1}. \quad (8)$$

Finally, similar to the previous section, the marker-based relationship matrices \mathbf{G} above may be replaced by $\mathbf{G}_\omega = \mathbf{G}(1 - \omega) + \mathbf{A}(\mathbf{\Gamma})\omega$ where ω is the fraction of genetic variance that is not captured by marker genotypes.

Genetic models for both purebred and crossbred $\mathcal{C}(AB)$ performances

In the previous sections, partial genetic and common genetic models for additive genetic effects for crossbred $\mathcal{C}(AB)$ performance were presented, and in both cases genomic versions of the models and combined relationship matrices were shown. Now, we show how the genetic variances and covariances for the model in Eq. (2) look like in the two cases.

For the partial genetic case, the vector of genetic effects on crossbred $\mathcal{C}(AB)$ individuals equals $\mathbf{g}_{\mathcal{C}(AB)}^C + \mathbf{g}_{\mathcal{C}(AB)}^A + \mathbf{g}_{\mathcal{C}(AB)}^B + \mathbf{g}_{\mathcal{C}(AB)}^{AB}$ and based on $\mathbf{g}_{\mathcal{C}(AB)}^A$, $\mathbf{g}_{\mathcal{C}(AB)}^B$ and $\mathbf{g}_{\mathcal{C}(AB)}^C$, breed-specific partial relationships define the breeding values for crossbred $\mathcal{C}(AB)$ performance on purebred animals, \mathbf{g}_A , \mathbf{g}_B and \mathbf{g}_C , respectively. Combining these effects with the breeding values for purebred performances, \mathbf{a}_A , \mathbf{a}_B and \mathbf{a}_C , the variance-covariance of genetic effects is determined by

$$\begin{aligned} \text{Var} \begin{bmatrix} \mathbf{a}_A \\ \star \\ \star \\ \text{-----} \\ \mathbf{g}_A \\ \mathbf{g}_{AB}^A \\ \mathbf{g}_{\mathcal{C}(AB)}^A \end{bmatrix} &= \Sigma_A \otimes \mathbf{H}^A, \\ \text{Var} \begin{bmatrix} \mathbf{a}_B \\ \star \\ \star \\ \text{-----} \\ \mathbf{g}_B \\ \mathbf{g}_{AB}^B \\ \mathbf{g}_{\mathcal{C}(AB)}^B \end{bmatrix} &= \Sigma_B \otimes \mathbf{H}^B, \\ \text{Var} \begin{bmatrix} \mathbf{a}_C \\ \star \\ \star \\ \text{-----} \\ \mathbf{g}_C \\ \mathbf{g}_{\mathcal{C}(AB)}^C \end{bmatrix} &= \Sigma_C \otimes \mathbf{H}^C, \\ \text{Var} [\mathbf{g}_{\mathcal{C}(AB)}^{AB}] &= \sigma_{g,AB}^2 \mathbf{H}^{AB}, \end{aligned}$$

with the four vectors being independent. Here, \otimes denotes the Kronecker product, \star denotes artificial random vectors such that the genetic variance-covariance matrices can be expressed using Kronecker products and matrices

$$\Sigma_b = \begin{bmatrix} \sigma_{a,b}^2 & \sigma_{ag,b} \\ \sigma_{ag,b} & \sigma_{g,b}^2 \end{bmatrix},$$

for $b = A, B, C$, are the 2×2 variance-covariance matrices containing the genetic variances for purebred breeding values and crossbred breeding values, and the covariance between them. Thus, using partial relationship matrices provides a formulation of the model in Eq. (2) using Kronecker products, such that parameters can be estimated and breeding values predicted using standard animal breeding software. In this model, there are 10 genetic parameters and $2(n_A + n_B + n_C + n_{AB}) + 3n_{C(AB)}$ genetic values where $n_{\mathcal{X}}$ is the number of individuals in population \mathcal{X} .

For the common genetic case, all individuals are related, and breeding values for crossbred $C(AB)$ performance on purebred animals, \mathbf{g}_A , \mathbf{g}_B and \mathbf{g}_C , are defined by additive relationships to the genetic effects on crossbreds, $\mathbf{g}_{C(AB)}$. Combining these effects with the breeding values for purebred performances, \mathbf{a}_A , \mathbf{a}_B and \mathbf{a}_C , the variance-covariance of genetic effects equals:

$$\text{Var} \begin{pmatrix} \mathbf{a}_A \\ \star \\ \star \\ \star \\ \star \\ \text{-----} \\ \star \\ \mathbf{a}_B \\ \star \\ \star \\ \star \\ \text{-----} \\ \star \\ \star \\ \mathbf{a}_C \\ \star \\ \star \\ \text{-----} \\ \mathbf{g}_A \\ \mathbf{g}_B \\ \mathbf{g}_C \\ \mathbf{g}_{AB} \\ \mathbf{g}_{C(AB)} \end{pmatrix} = \Sigma \otimes \mathbf{H}(\Gamma),$$

where \star denotes artificial random vectors and Σ is the 4×4 genetic variance-covariance matrix:

$$\Sigma = \begin{bmatrix} \sigma_{a,A}^2 & \sigma_{a,A,B} & \sigma_{a,A,C} & \sigma_{ag,A} \\ \sigma_{a,A,B} & \sigma_{a,B}^2 & \sigma_{a,B,C} & \sigma_{ag,B} \\ \sigma_{a,A,C} & \sigma_{a,B,C} & \sigma_{a,C}^2 & \sigma_{ag,C} \\ \sigma_{ag,A} & \sigma_{ag,B} & \sigma_{ag,C} & \sigma_g^2 \end{bmatrix}.$$

The formulation of the model in Eq. (2) using Kronecker products implies that parameters can be estimated and breeding values predicted using standard animal breeding software. This model contains 10 genetic parameters and $2(n_A + n_B + n_C) + n_{AB} + n_{C(AB)}$ genetic values, and in addition, 6 parameters in matrix Γ .

In the common genetic case, there are three parameters $\sigma_{a,A,B}$, $\sigma_{a,A,C}$ and $\sigma_{a,B,C}$ which are genetic covariances between purebred performances, and these parameters are not present in the partial genetic case. The reason is that they would not be identifiable since there is no specification of the relationships across breeds in the partial genetic case. In the common genetic case, the identifiability of $\sigma_{a,A,B}$, $\sigma_{a,A,C}$ and $\sigma_{a,B,C}$ relies on the genomic relationships between pairs of animals in different breeds. In the partial genetic case, there are four genetic parameters for crossbred performance, $\sigma_{g,A}^2$, $\sigma_{g,B}^2$, $\sigma_{g,C}^2$ and $\sigma_{g,AB}^2$ that scale each of the four partial relationship matrices, whereas in the common genetic case there is only one such parameter σ_g^2 . As explained in a previous section, there is a correspondence between these parameters via the parameters in matrix Γ as follows: $\sigma_{g,b}^2 = \sigma_g^2(1 - \gamma_b/2)$, $b = A, B, C$, $\sigma_{g,AB}^2 = \sigma_g^2((\gamma_A + \gamma_B)/2 - \gamma_{A,B})/4$. However, note that there is a difference between estimating $\sigma_{g,A}^2$, $\sigma_{g,B}^2$, $\sigma_{g,C}^2$ and $\sigma_{g,AB}^2$ from phenotypes as in the partial genetic case, and determining these from a general σ_g^2 and parameters in Γ , which are estimated based on marker genotypes as in the common genetic case.

Discussion

For three-way crossbreeding, we presented models based on both pedigree-based, marker-based and combined relationships. Using combined relationship matrices results in a model for genetic evaluation where both pedigree and marker genotypes are used simultaneously for genetic evaluation, i.e. a single-step method for genomic evaluation. This paper provides the models and mathematical formulas, but a numerical implementation is needed before the methods are ready for use in practice. Such methods make it possible to incorporate phenotypes and genotypes on crossbreds into an existing genetic evaluation system, assuming that such a system is based on a single-step method.

The models for three-way crossbreeding investigated in this paper were four-variate models where each variable was measured in a specific population, A , B , C or $C(AB)$. The main scenario that we have in mind is a scenario where the four variables represent the same biological trait measured in four different genetic backgrounds and possibly different environments, but in principle the four variables could also be different biological traits. An extension of the model to a situation where multiple

biological traits are measured in each of the four populations is in principle straightforward since the additive relationship matrices are the same, although in practice it may require the estimation of a very large number of genetic parameters. Extending the approaches to other types of models that are implemented in standard animal breeding software, like threshold models, models with indirect genetic effects, models for test-day records, etc. is also in principle straightforward. Finally, modifying the models to other scenarios with data recording, for example with records on \mathcal{AB} individuals or no records on one of the pure breeds, is also straightforward. In general, designing data recording for these complicated models is an issue, and for example to obtain precise estimates of the genetic correlation parameters, it would be important that the relationships between crossbred animals with records and purebred animals with records are close.

Two types of approaches for constructing additive relationships were presented, based on different assumptions about allele substitution effects of causal loci or SNPs. In the partial genetic approach, allele substitution effects of SNPs were assumed independent between breeds, whereas in the common genetic approach, they were assumed to be the same in different breeds. The partial genetic approach requires that alleles are traced according to breed of origin, which is feasible in some scenarios but may be difficult with sufficient accuracy in others. In particular, when crossbred $\mathcal{C}(\mathcal{AB})$ animals are genotyped, a reasonable requirement is that breed \mathcal{C} fathers are also genotyped which would make the tracing of the breed \mathcal{C} paternal allele feasible, but the tracing of the breed of origin (\mathcal{A} or \mathcal{B}) of the maternal allele may be more uncertain and depend on whether \mathcal{AB} mothers are genotyped (may not be due to logistical issues), maternal grandfathers are genotyped and maternal grandmothers are genotyped (may be difficult to obtain if these are from multiplier herds). An advantage of the common genetic approach is that the marker-based relationship matrix is easier to construct because tracing the breed of origin of alleles is not required, but a disadvantage may be the computational burden of using a larger relationship matrix. In addition, parameters in matrix Γ need to be estimated and the sensitivity of genetic evaluation to these estimates is unknown. Future research using simulated and real data is needed to clarify the differences between the two approaches.

Other terminal crossbreeding systems are of interest in pig production. Models for two-way crossbreeding are relevant for sow-traits measured on animals from breed \mathcal{A} and \mathcal{B} and cross \mathcal{AB} , and such models were presented in Christensen et al. [11] using partial genetic relationship matrices. An alternative to this partial genetic approach would be to use the common genetic approach

presented here. The four-way crossbreeding system where crossbred \mathcal{CD} sires are mated to \mathcal{AB} dams to produce $(\mathcal{CD})(\mathcal{AB})$ pigs for slaughter, is also used in pig production. The approaches in this paper can be extended to such a system, and the resulting model would be a five-variate model. Using the partial genetic approach, there would be four breed-specific partial relationship matrices and two breed-segregation partial relationship matrices, and the corresponding model for purebred and crossbred performances would contain 14 genetic parameters, whereas using the common genetic approach, the model for purebred and crossbred performances would contain 15 genetic parameters.

Many papers have reported genetic correlations between purebred and crossbred performances [21–26]. The reported estimated correlations ranged from 0.38 to 0.946, depending on trait and on differences in the environment, and in general with relatively high standard error on the estimates. The higher the genetic correlation, the less gain there will be by including crossbred data into the genetic evaluation system. All these results are from two-way crosses, and the authors are not aware of publications based on data from three-way crossbreeding where data in purebred and crossbred populations are considered to be different traits. The models presented in this paper should be useful to investigate such data from three-way crossbreeding.

Conclusion

Models for genetic evaluation in the three-way crossbreeding system are presented. These models provide estimated breeding values for both purebred and crossbred performances, and can use pedigree-based or marker-based relationships, or combined relationships based on both pedigree and marker information. This provides a framework that allows information from three-way crossbred animals to be incorporated into a genetic evaluation system.

Authors' contributions

OFC conceived the study and derived the formulas with help from AL. OFC took the lead in writing the manuscript, and AL, GS and MSL helped with the writing and with discussions. All authors read and approved the final manuscript.

Author details

¹ Department of Molecular Biology and Genetics, Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Allé 20, P.O. BOX 50, 8830 Tjele, Denmark. ² INRA, UMR 1388 GenPhySE, BP52627, 31326 Castanet Tolosan, France.

Acknowledgements

The work was performed in a project funded through the Green Development and Demonstration Programme (grant no. 34009-12-0540) by the Danish Ministry of Food, Agriculture and Fisheries, the Pig Research Centre and Aarhus University. AL acknowledges financing from INRA SelGen metaprogram in projects X-Gen, EpiSel, SelHet and SelDir, and is grateful to the genotoul bioinformatics platform Toulouse Midi-Pyrenees for providing computing

storage resources. Comments and corrections from reviewers and editors are also acknowledged.

Competing interests

The authors declare that they have no competing interests.

Appendix

Here, explicit and detailed derivations of the additive relationships across purebred and crossbred animals related to the $C(AB)$ crossbreeding system are presented.

In the derivation, both partial genetic and common genetic approaches for the variance-covariance of genetic effects for crossbred $C(AB)$ performance are inspired by the derivation in Lo et al. [13] of formulas (3) and (4). Lo et al. [13] based their derivation on the genotypic value expressed as a sum over loci of effects of paternal and maternal alleles:

$$g_i = \mu + \sum_j (\alpha_{s_i^j} + \alpha_{d_i^j}),$$

where $\alpha_{s_i^j}$ and $\alpha_{d_i^j}$ are the additive effects of the paternal and maternal alleles, respectively, and these effects depend on the breed of origin b of the alleles j , such that $E[\alpha_{s_i^j} | s_i \in b] = \epsilon_j^b$ and $E[\alpha_{d_i^j} | d_i \in b] = \epsilon_j^b$, where the expectation is taken across all individuals in breed b and symbol \in is used to denote breed of origin of an allele. The term ϵ_j^b is the mean additive effect and this is different between breeds due to different allele frequencies in different breeds. To be explicit, $\epsilon_j^b = p_j^b a_j + (1 - p_j^b)(-a_j) = (p_j^b - 1/2)2a_j$ where p_j^b is the allele frequency in breed b and a_j is the additive effect of the j -th allele. Above, both expectations and allele frequencies refer to the base populations, and it is assumed that in each base population, alleles are assigned randomly to individuals. It is further assumed that effects for different loci are independent. Here, we introduce the notation $\alpha_{s_i^j}^b = \alpha_{s_i^j} - \epsilon_j^b$ when $s_i^j \in b$ and $\alpha_{d_i^j}^b = \alpha_{d_i^j} - \epsilon_j^b$ when $d_i^j \in b$, such that the expectations of the α^b s are equal to 0.

First, pedigree-based additive genetic relationships are derived using the partial genetic approach and common genetic approach, respectively, and second the corresponding marker-based relationships are derived.

Additive genetic model for crossbred $C(AB)$ performance: partial genetic approach

Contrary to Lo et al. [13] and Garcia-Cortes and Toro [14], in the derivation presented here, we first split the genotypic values according to breeds of origin instead of computing the variances and covariances and then splitting them. The reason for this is for similarity with the derivation of the corresponding genomic model using the partial genetic approach that appears in a following subsection.

For the $C(AB)$ crossbred animals studied here, the paternal allele is always breed C and the maternal allele is either breed A or B with equal probability, and therefore the genotypic value becomes:

$$\begin{aligned} g_i &= \mu + \sum_j \alpha_{s_i^j} + \sum_{j:d_i^j \in A} \alpha_{d_i^j} + \sum_{j:d_i^j \in B} \alpha_{d_i^j} \\ &= \tilde{\mu}_{C(AB)} + \sum_j \alpha_{s_i^j}^C + \sum_{j:d_i^j \in A} \alpha_{d_i^j}^A + \sum_{j:d_i^j \in B} \alpha_{d_i^j}^B \\ &\quad + 0.5 \sum_j r_{d_i^j} (\epsilon_j^A - \epsilon_j^B) \\ &= \tilde{\mu}_{C(AB)} + g_{C(AB),i}^C + g_{C(AB),i}^A + g_{C(AB),i}^B + g_{C(AB),i}^{AB}, \end{aligned}$$

where $\tilde{\mu}_{C(AB)} = \mu + \sum_j \epsilon_j^C + 0.5(\sum_j \epsilon_j^A + \sum_j \epsilon_j^B)$, $r_{d_i^j} = 1$ when $d_i^j \in A$ and $r_{d_i^j} = -1$ when $d_i^j \in B$, and terms $g_{C(AB),i}^C$, $g_{C(AB),i}^A$, $g_{C(AB),i}^B$ and $g_{C(AB),i}^{AB}$ are defined implicitly. In this way, the genotypic value has been split into partial genetic effects, where the terms $g_{C(AB),i}^C$, $g_{C(AB),i}^A$, $g_{C(AB),i}^B$ are breed of origin specific and $g_{C(AB),i}^{AB}$ is a breed-segregation term.

For the AB crossbred sows (where for simplicity of notation it is assumed that their fathers are breed A and mothers breed B),

$$\begin{aligned} g_i &= \mu + \sum_j \alpha_{s_i^j} + \sum_j \alpha_{d_i^j} = \tilde{\mu}_{AB} + \sum_j \alpha_{s_i^j}^A + \sum_j \alpha_{d_i^j}^B \\ &= \tilde{\mu}_{AB} + g_{AB,i}^A + g_{AB,i}^B, \end{aligned}$$

with $\tilde{\mu}_{AB} = \mu + \sum_j \epsilon_j^A + \sum_j \epsilon_j^B$, and terms $g_{AB,i}^A$ and $g_{AB,i}^B$ defined implicitly.

For a purebred animal of breed b

$$\begin{aligned} g_i &= \mu + \sum_j \alpha_{s_i^j} + \sum_j \alpha_{d_i^j} = \tilde{\mu}_b + \sum_j \alpha_{s_i^j}^b + \sum_j \alpha_{d_i^j}^b \\ &= \tilde{\mu}_b + g_{b,i}, \end{aligned}$$

where $\tilde{\mu}_b = \mu + 2 \sum_j \epsilon_j^b$ and $g_{b,i} = \sum_j \alpha_{s_i^j}^b + \sum_j \alpha_{d_i^j}^b$ is the breeding value for crossbred $C(AB)$ performance. From this, the three vectors of breeding values for crossbred $C(AB)$ performance, \mathbf{g}_A , \mathbf{g}_B and \mathbf{g}_C , are defined.

In this way, a breed-specific partial genetic effect has been defined for all animals containing the specific breed, and a breed-segregation partial genetic effect has been defined for crossbred $C(AB)$ animals. The resulting variance-covariance matrices are as shown in the “Methods” section.

Note that the different means $\tilde{\mu}_{C(AB)}$, $\tilde{\mu}_{AB}$, $\tilde{\mu}_A$, $\tilde{\mu}_B$ and $\tilde{\mu}_C$ should strictly speaking be included into the genetic effects and breeding values, but they have been omitted here, since the genetic values are for performance in a

specific $\mathcal{C}(AB)$ cross and these means cannot be inferred from data.

The variance of the breed-segregation term for a $\mathcal{C}(AB)$ animal equals

$$\begin{aligned} \text{Var}(g_{\mathcal{C}(AB),i}^{AB}) &= \text{Var}\left(0.5 \sum_j r_{d_i^j} (\epsilon_j^A - \epsilon_j^B)\right) \\ &= \sum_j (\epsilon_j^A - \epsilon_j^B)^2 / 4 = \sigma_{g,AB}^2 / 2, \end{aligned}$$

where $\sigma_{g,AB}^2 = \sum_j (\epsilon_j^A - \epsilon_j^B)^2 / 2$ is the breed-segregation variance, i.e. the additional genetic variance in an F2 cross compared to an F1 cross. The assumption that has been used here is that the ϵ_j^A s and ϵ_j^B s are fixed constants.

Additive genetic model for crossbred $\mathcal{C}(AB)$ performance: common genetic approach

Using the notation defined above, the genetic value for crossbred performance of purebred b animal i equals

$$g_i = \mu + \sum_{j=1}^N \alpha_{s_i^j}^b + \sum_{j=1}^N \alpha_{d_i^j}^b = \mu + \sum_{j=1}^N (\alpha_{s_i^j}^b + \alpha_{d_i^j}^b + 2\epsilon_j^b),$$

where the $\alpha_{s_i^j}^b$ s and $\alpha_{d_i^j}^b$ s are independent between breeds, and N is the number of genes. Compared to the partial genetics approach, here we assume that both the α^b s and the ϵ_j^b s are random variables. The randomness in the α^b s is because different animals inherit different alleles, but the randomness in ϵ_j^b has a different origin. Note that the differences in ϵ_j^b between breeds is due to differences in allele frequencies, and as explained previously, $\epsilon_j^b = p_j^b a_j + (1 - p_j^b)(-a_j) = (p_j^b - 1/2)2a_j$, where p_j^b is the allele frequency and a_j is the allelic effect. Assigning prior distributions with expectations 1/2 to the allele frequencies as in Christensen [20] corresponds to assigning a prior distribution with expectation 0 and variance proportional to $(a_j)^2$ to ϵ_j^b for all loci. Furthermore, assuming that prior distributions for allele frequencies are correlated between breeds implies that covariances of the ϵ_j^b s become proportional to $(a_j)^2$.

Thus, we assume that the ϵ_j^b s are random variables that are independent of the $\alpha_{s_i^j}^b$ s and $\alpha_{d_i^j}^b$ s and have a mean of 0. Furthermore, we assume that the ϵ_j^b s are correlated between breeds,

$$\text{Var} \begin{bmatrix} \epsilon_j^A \\ \epsilon_j^B \\ \epsilon_j^C \end{bmatrix} = \begin{bmatrix} \tau_{\epsilon,A}^2 & \tau_{\epsilon,A,B} & \tau_{\epsilon,A,C} \\ \tau_{\epsilon,A,B} & \tau_{\epsilon,B}^2 & \tau_{\epsilon,A,C} \\ \tau_{\epsilon,A,C} & \tau_{\epsilon,A,B} & \tau_{\epsilon,C}^2 \end{bmatrix} (a_j)^2.$$

The variable $\alpha_{s_i^j}^b$ when $s_i^j \in b$ is a random variable with mean 0, and since it may be expressed as $\alpha_{s_i^j}^b = (z_j - p_j^b)a_j$ where $z_j = 0, 1$ for paternal allele being 1 and 2, respectively, we see that the variance

of $\alpha_{s_i^j}^b$ is proportional to $(a_j)^2$. Similarly, $\alpha_{d_i^j}^b$ is a random variable with mean 0 and variance proportional to $(a_j)^2$. For animals in the base population of the pedigree, the variances of $\alpha_{s_i^j}^b$ and $\alpha_{d_i^j}^b$ equal $\tau_{\alpha,b}^2 (a_j)^2$ with $\tau_{\alpha,b}^2 = E[\text{Var}(z_j - p_j^b) | p_j^b] = E[p_j^b(1 - p_j^b)] = E[p_j^b] - (\text{Var}[p_j^b] + E[p_j^b]^2) = 1/4 - \tau_{\epsilon,b}^2$ where expectation is taken with respect to the prior distribution of p_j^b , and the mutual covariances between $\alpha_{s_i^j}^b$, $\alpha_{s_{i'}^j}^b$, $\alpha_{d_i^j}^b$ and $\alpha_{d_{i'}^j}^b$ are all zero when $i \neq i'$.

Therefore, in the base population of the pedigree, the variances and covariances of the genetic values $g_i = \sum_j (\alpha_{s_i^j}^b + \alpha_{d_i^j}^b + 2\epsilon_j^b)$ become

$$\text{Var}(g_i) = (2\tau_{\alpha,b}^2 + 4\tau_{\epsilon,b}^2) \sum_j (a_j)^2,$$

for an individual in breed b ,

$$\text{Cov}(g_i, g_{i'}) = 4\tau_{\epsilon,b}^2 \sum_j (a_j)^2,$$

for two different individuals in breed b , and finally

$$\text{Cov}(g_i, g_{i'}) = 4\tau_{\epsilon,b,b'} \sum_j (a_j)^2,$$

for two individuals in different breeds b, b' .

Defining new parameters $\sigma_g^2 = \sum_j (a_j)^2 (2\tau_{\alpha,b}^2 + 2\tau_{\epsilon,b}^2)$ and $\gamma_b = 4\tau_{\epsilon,b}^2 / (2\tau_{\alpha,b}^2 + 2\tau_{\epsilon,b}^2)$, we see that the elements of the variance-covariance matrix within breed is defined by:

$$\text{Var}(g_i) = \sigma_g^2 (1 + \gamma_b / 2),$$

for an individual in breed b , and

$$\text{Cov}(g_i, g_{i'}) = \sigma_g^2 \gamma_b,$$

for two individuals in breed b , i.e. animals in the base population are inbred with coefficient $\gamma_b/2$ and related with relationship coefficient γ_b . Note that from $\tau_{\alpha,b}^2 + \tau_{\epsilon,b}^2 = 1/4$, we see that $\sigma_g^2 = \sum_j (a_j)^2 / 2$ does not depend on breed and $\gamma_b = 8\tau_{\epsilon,b}^2$. Furthermore, define $\gamma_{b,b'} = 8\tau_{\epsilon,b,b'}$, then

$$\text{Cov}(g_i, g_{i'}) = \sigma_g^2 \gamma_{b,b'},$$

for two individuals in different breeds b, b' , i.e. base animals in different breeds are related. Therefore, a joint relationship matrix is specified among all base animals, and by applying the usual recursive definition, an additive relationship matrix is defined across all animals as shown in the "Methods" section.

Note that the breed-segregation term disappears here, since the differences in the ϵ^A and ϵ^B terms are

incorporated into the Γ matrix. However, breed-segregation variance is still present under this model, which is illustrated by

$$\begin{aligned}\text{Var}(g_{C(AB),i}^{AB}) &= \text{Var}(0.5 \sum_j r_{d_i^j} (\epsilon_j^A - \epsilon_j^B)) \\ &= \sum_{j=1}^n E((\epsilon_j^A - \epsilon_j^B)^2)/4 \\ &= (\tau_{\epsilon,A}^2 + \tau_{\epsilon,B}^2 - 2\tau_{\epsilon,A,B}) \sum_j (a_j)^2/4 \\ &= (\gamma_A + \gamma_B - 2\gamma_{A,B})\sigma_g^2/16 \\ &= \sigma_{g,AB}^2/2,\end{aligned}$$

where $\sigma_{g,AB}^2 = \sigma_g^2((\gamma_A + \gamma_B)/2 - \gamma_{A,B})/4$ is the breed-segregation variance, i.e. the additional genetic variance in an F2 cross compared to an F1 cross; see Legarra et al. (2015) [15].

Genomic model for crossbred $C(AB)$ performance: partial genetic approach

Based on pedigree relationships, the randomness comes from not knowing which alleles are inherited, but when having marker genotypes this is actually known. Here, the randomness comes from assigning distributions on effects. In addition, causal loci are replaced by SNPs that are in linkage disequilibrium with causal loci, but for shortness of notation we will use the same notation for the effects of SNPs as that used for the causal loci.

Now, for SNP j

$$\alpha_{s_i^b}^b = (z_{s_i^b} - p_j^b)\beta_j^b, \quad \alpha_{d_i^b}^b = (z_{d_i^b} - p_j^b)\beta_j^b,$$

where $z_j = 0, 1$ for paternal allele being 1 and 2, respectively, $z_{s_i^b}$ is defined similarly for the maternal allele, p_j^b denotes the allele frequency of breed b base population, and β_j^b is the breed b specific allele substitution effect which is assumed to be Gaussian distributed with mean 0. Breed-specific partial genetic effects are therefore as follows:

$$\begin{aligned}g_{A,i} &= \sum_j (z_{s_i^A} + z_{d_i^A} - 2p_j^A)\beta_j^A, \\ g_{AB,i}^A &= \sum_j (z_{s_i^A} - p_j^A)\beta_j^A, \\ g_{C(AB),i}^A &= \sum_{j:d_i^j \in A} (z_{d_i^j} - p_j^A)\beta_j^A,\end{aligned}$$

for breed A ,

$$\begin{aligned}g_{B,i} &= \sum_j (z_{s_i^B} + z_{d_i^B} - 2p_j^B)\beta_j^B, \\ g_{AB,i}^B &= \sum_j (z_{d_i^B} - p_j^B)\beta_j^B, \\ g_{C(AB),i}^B &= \sum_{j:d_i^j \in B} (z_{d_i^j} - p_j^B)\beta_j^B,\end{aligned}$$

for breed B , and

$$\begin{aligned}g_{C,i} &= \sum_j (z_{s_i^C} + z_{d_i^C} - 2p_j^C)\beta_j^C, \\ g_{C(AB),i}^C &= \sum_j (z_{s_i^C} - p_j^C)\beta_j^C,\end{aligned}$$

for breed C . The breed-segregation partial genetic effect is $g_{C(AB),i}^{AB} = 0.5 \sum_j r_{d_i^j} (\epsilon_j^A - \epsilon_j^B)$ where $r_{d_i^j} = 1$ when $d_i^j \in A$ and $r_{d_i^j} = -1$ when $d_i^j \in B$ and $\epsilon_j^A - \epsilon_j^B = p_j^A \beta_j^A - p_j^B \beta_j^B$.

An equivalent formulation of the model is to use marker-based partial relationship matrices instead of breed-specific allele substitution effects. Define breed b specific allele content matrices as matrix \mathbf{m}^b with entries $m_{ij}^b = z_j + z_{d_i^j}$ for purebred b animals, matrices \mathbf{z}^A and \mathbf{z}^B with entries z_j and $z_{d_i^j}$, respectively, for crossbred AB animals, matrix \mathbf{z}^C with entries z_j for crossbred $C(AB)$ animals, and finally matrices $\mathbf{z}_p^{A,i}$ and $\mathbf{z}_p^{B,i}$ with entries $z_{d_i^j}$ and $z_{d_i^j}$, respectively, for crossbred $C(AB)$ animals when the breed-specific allele is inherited and zero otherwise. From these breed-specific allele content matrices, marker-based breed-specific partial relationship matrices are constructed and these are as in the “Methods” section. Note that the scalings of these matrices are $s^b = \text{Var}(\beta_j^b)/\sigma_b^2$ for $b = A, B, C$. The breed-segregation partial relationship matrix is defined as $G_{i,i'}^{AB} = \sum_j r_{d_i^j} r_{d_{i'}^j} / (2n)$.

Independence between vectors, $\mathbf{g}^{(A)}$, $\mathbf{g}^{(B)}$, $\mathbf{g}^{(C)}$ and $\mathbf{g}_{C(AB)}^{AB}$ requires that allele substitution effects β_j^A , β_j^B and β_j^C are independent between breeds, and also that they are independent of the $\epsilon_j^A - \epsilon_j^B$. First, independence between β_j^A , β_j^B and β_j^C seems to somehow contradict the assumption of additive gene effect a_j being the same independent of breed of origin, but may be justified if there is no persistence of the phase between markers and QTL in different breeds. Second, $\epsilon_j^A - \epsilon_j^B = p_j^A \beta_j^A - p_j^B \beta_j^B$ being independent of β_j^A and β_j^B cannot strictly hold when β_j^A and β_j^B are also independent. However, as explained by de los Campos et al. [27] the fact that multiple markers are likely to track the same QTL questions the assumptions of additivity and independence of SNP allele substitution effects between loci, and therefore the

genomic model is a rough approximation to the reality of the additive model of causal effects.

Genomic model for crossbred $\mathcal{C}(AB)$ performance: common genetic approach

A genomic version of the model with common genetic effects across breeds can be formulated by replacing α_j^b by $(z_{s_j} - p_j^b)\beta_j$, $\alpha_{d_j}^b$ by $(z_{d_j} - p_j^b)\beta_j$, and ϵ_j^b by $(p_j^b - 1/2)\beta_j$ where $1/2$ is the common allele frequency and β_j is the allele substitution effect. Then,

$$\alpha_{s_j}^b + \epsilon_j^b = (z_{s_j} - 0.5)\beta_j, \quad \alpha_{d_j}^b + \epsilon_j^b = (z_{d_j} - 0.5)\beta_j.$$

The resulting model becomes:

$$\begin{aligned} g_i &= \mu + \sum_j (z_{s_j} - 0.5)\beta_j + (z_{d_j} - 0.5)\beta_j \\ &= \mu + \sum_j (m_{i,j} - 1)\beta_j. \end{aligned}$$

The marker-based relationship matrix is therefore constructed as usual across all genotyped animals:

$$\mathbf{G} = \frac{(\mathbf{m} - \mathbf{1}\mathbf{1}^T)(\mathbf{m} - \mathbf{1}\mathbf{1}^T)^T}{s},$$

where s is scaling parameter.

Received: 24 June 2015 Accepted: 4 December 2015

Published online: 22 December 2015

References

- Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001;157:1819–29.
- Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. *J Dairy Sci*. 2009;92:4656–63.
- 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;93:743–52.
- Christensen OF, Lund MS. Genomic prediction when some animals are not genotyped. *Genet Sel Evol*. 2010;42:2.
- 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, 1–6 August 2010; Leipzig; 2010. paper 0036. [urlhttp://www.kongressband.de/wcgalp2010/assets/pdf/0036](http://www.kongressband.de/wcgalp2010/assets/pdf/0036).
- 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.
- Wei M, van der Werf JHJ. Maximizing genetic response in crossbreds using both purebred and crossbred information. *Anim Prod*. 1994;59:401–13.
- 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;108:262–9.
- Baumung R, Sölkner J, Essl A. Correlation between purebred and crossbred performance under a two-locus model with additive by additive interaction. *J Anim Breed Genet*. 1997;114:89–98.
- Christensen OF, Madsen P, Nielsen B, Su G. Genomic evaluation of both purebred and crossbred performances. *Genet Sel Evol*. 2014;46:23.
- Stuber CW, Cockerham CC. Gene effects and variances in hybrid populations. *Genetics*. 1966;64:1279–86.
- Lo LL, Fernando RL, Grossman M. Covariance between relatives in multi-breed populations: additive model. *Theor Appl Genet*. 1993;87:423–30.
- García-Cortés LA, Toro MA. Multibreed analysis by splitting the breeding values. *Genet Sel Evol*. 2006;38:601–15.
- Legarra A, Christensen OF, Vitezica ZG, Aguilar I, Misztal I. Ancestral relationships using metafounders: finite ancestral populations and across population relationships. *Genetics*. 2015;200:455–68.
- 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.
- Colleau J-J. An indirect approach to the extensive calculation of relationship coefficients. *Genet Sel Evol*. 2002;34:409–21.
- Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*. 2014;15:478.
- Christensen OF, Madsen P, Nielsen B, Ostensen T, Su G. Single-step methods for genomic evaluation in pigs. *Animal*. 2012;6:1565–71.
- Christensen OF. Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. *Genet Sel Evol*. 2012;44:37.
- Brandt H, Täubert H. Parameter estimates for purebred and crossbred performances in pigs. *J Anim Breed Genet*. 1998;115:97–104.
- Kiszlinger HN, Farkas J, Köver G, Onika-Szvath S, Nagy I. Genetic parameters of growth traits from a joint evaluation of purebred and crossbred pigs. *Agric Cons Sci*. 2011;76:223–6.
- Wei M, van der Werf JHJ. Genetic correlation and heritabilities for purebred and crossbred performance in poultry egg production traits. *J Anim Sci*. 1995;73:2220–6.
- Zumbach B, Misztal I, Tsuruta S, Holl J, Heering 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.
- 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;79:3002–7.
- Bloemhof K, Kaue A, Knol EF, van Arendonk JAM, Misztal I. Heat stress effects on farrowing rate in sows: genetic parameter estimation using within-line and crossbred models. *J Anim Sci*. 2011;90:2009–119.
- de los Campos G, Sorensen D, Gianola D. Genomic heritability: what is it? *PLoS Genet*. 2015;11:1005048.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

