

Prediction of the reliability of genomic breeding values for crossbred performance

Jérémie Vandenplas, Jack J. Windig, Mario P. L. Calus

► **To cite this version:**

Jérémie Vandenplas, Jack J. Windig, Mario P. L. Calus. Prediction of the reliability of genomic breeding values for crossbred performance. *Genetics Selection Evolution*, BioMed Central, 2017, 49 (1), pp.43. 10.1186/s12711-017-0318-1 . hal-01522125

HAL Id: hal-01522125

<https://hal.archives-ouvertes.fr/hal-01522125>

Submitted on 12 May 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH ARTICLE

Open Access



Prediction of the reliability of genomic breeding values for crossbred performance

Jérémie Vandenplas^{*} , Jack J. Windig and Mario P. L. Calus

Abstract

Background: In crossbreeding programs, various genomic prediction models have been proposed for using phenotypic records of crossbred animals to increase the selection response for crossbred performance in purebred animals. A possible model is a model that assumes identical single nucleotide polymorphism (SNP) effects for the crossbred performance trait across breeds (ASGM). Another model is a genomic model that assumes breed-specific effects of SNP alleles (BSAM) for crossbred performance. The aim of this study was to derive and validate equations for predicting the reliability of estimated genomic breeding values for crossbred performance in both these models. Prediction equations were derived for situations when all (phenotyping and) genotyping data have already been collected, i.e. based on the genetic evaluation model, and for situations when all genotyping data are not yet available, i.e. when designing breeding programs.

Results: When all genotyping data are available, prediction equations are based on selection index theory. Without availability of all genotyping data, prediction equations are based on population parameters (e.g., heritability of the traits involved, genetic correlation between purebred and crossbred performance, effective number of chromosome segments). Validation of the equations for predicting the reliability of genomic breeding values without all genotyping data was performed based on simulated data of a two-way crossbreeding program, using either two closely-related breeds, or two unrelated breeds, to produce crossbred animals. The proposed equations can be used for an easy comparison of the reliability of genomic estimated breeding values across many scenarios, especially if all genotyping data are available. We show that BSAM outperforms ASGM for a specific breed, if the effective number of chromosome segments that originate from this breed and are shared by selection candidates of this breed and crossbred reference animals is less than half the effective number of all chromosome segments that are independently segregating in the same animals.

Conclusions: The derived equations can be used to predict the reliability of genomic estimated breeding values for crossbred performance using ASGM or BSAM in many scenarios, and are thus useful to optimize the design of breeding programs. Scenarios can vary in terms of the genetic correlation between purebred and crossbred performances, heritabilities, number of reference animals, or distance between breeds.

Background

Several livestock production systems are based on crossbreeding schemes (e.g., [1–3]), and take advantage of the increased performance of crossbred animals compared to purebred animals, along with breed complementarity. For such production systems based on crossbreeding, the breeding goal for the purebred populations is to optimize

the performance of crossbred descendants. However, the selection of purebred animals for crossbred performance has not been extensively implemented in livestock, partly due to the difficulty of routine collection of pedigree information on crossbred animals [4].

With the advent of genomic selection, various genomic prediction models have been proposed, which use phenotypic records of crossbred animals to increase the selection response for crossbred performance in purebred animals (e.g., [2, 4–6]). These approaches predict breeding values for crossbred performance of selection

*Correspondence: jeremie.vandenplas@wur.nl
Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands

candidates using the estimated allele substitution effects of many single nucleotide polymorphisms (SNPs). The SNP allele substitution effects are estimated from phenotypes of genotyped reference animals. In the context of crossbreeding, several breeds and their crosses are involved in genomic prediction, and purebred and crossbred performances are often considered to be different but correlated traits (e.g., [1, 3, 5, 7, 8]). Therefore, estimates of SNP allele substitution effects for purebred and crossbred performance traits may not be the same for purebred and crossbred populations, e.g., due to genotype by environment interactions. Assuming only additive gene action, one approach to accommodate this is to model differences between allele substitution SNP effects using a multivariate genomic model that assumes a correlation structure between the effects of SNPs across the purebred and crossbred populations, or equivalently, by assuming a genetic correlation structure across the trait measured in purebred and crossbred populations [9, 10]. These multivariate genomic models are referred to hereafter as across-breed SNP genotype models (ASGM), since the estimates of SNP allele substitution effects for the crossbred performance trait are also used to predict breeding values for crossbred performance of purebred selection candidates, regardless of their breed of origin [4, 6]. Thus, estimates of SNP effects for the crossbred performance trait using ASGM are not breed-specific. However, a number of factors may have an impact on the effect that can be measured for a SNP for the crossbred performance trait. First, the two parental alleles at a SNP in a crossbred animal may have different effects on the phenotype due to different levels of linkage disequilibrium (LD) with a quantitative trait locus (QTL) in the parental purebred populations. Second, different genetic backgrounds, such as dominance or epistatic interactions, can result in the effects of the same QTL to be different in purebred versus crossbred animals. And third, purebred and crossbred animals may be exposed to different environments, leading to genotype by environment interactions. Because of these reasons, estimated allele substitution effects at SNPs for the crossbred performance trait may be breed-specific. To accommodate all these differences, previously an approach was proposed [3–6] that estimates breed-specific allele substitution effects for the crossbred performance trait (BSAM), assuming that the breed origin of SNP alleles in crossbred animals is known. Results from simulations have shown that BSAM can result in greater accuracy of genomic estimated breeding values (EBV) of purebreds for crossbred performance than ASGM under some conditions [2, 4, 11].

In order to be able to evaluate many different breeding program designs that apply genomic prediction for

crossbreeding performance, it would be useful to be able to predict the reliability of genomic EBV using, for example, different genomic models or different breeding schemes. Prediction of reliability should preferably consider the genotype data of all reference animals and selection candidates when available, although it is also desirable to be able to predict the reliability when genotype data of, e.g., selection candidates is not available, i.e. when designing breeding programs. Various equations have been proposed in the literature to predict the reliability or the accuracy (i.e., the square root of reliability) of genomic EBV for (groups of) animals. The investigated genomic predictions rely on single-population genomic models [12, 13], and on ASGM [10, 14]. When genotypes are available for both reference animals and selection candidates, prediction equations are derived using selection index (SI) theory, while before availability of all genotyping data, they are derived using population parameters (e.g., heritability, number of reference animals) [10, 13, 15]. However, to our knowledge, equations for predicting the reliability of genomic EBV for crossbreeding performance for (groups of) animals have not yet been reported.

The primary aim of this study was to derive equations for predicting the reliability of genomic EBV for crossbred performance based on ASGM or BSAM. Prediction equations were derived for situations when all genotyping data are available for both reference animals and selection candidates (referred to as “with availability of genotyping data”), and for situations when the genotyping data are not available (referred to as “without availability of genotyping data”). The second aim was to compare the predictions of the reliability of genomic EBV without availability of genotyping data to the predictions obtained from the equations with availability of genotyping data, because the former are an approximation of the latter. Both reliabilities have the same expectation, since they both rely on prediction error variances (PEV) and assume absence of selection. Finally, the equation for predicting reliability without availability of genotyping data was used to investigate the expected ranges of reliabilities of genomic EBV using BSAM for a pig breeding program.

Methods

The first part of this section describes equations for predicting the reliability of genomic EBV for crossbred performance using ASGM or BSAM. For the derivations of these equations, we assumed a crossbreeding program with two breeds, A and B, with their F1 being crossbred AB animals. In order to simplify the derivation of the equations, we assumed that phenotypes are corrected for all fixed and random effects, other than additive genetic effects. Furthermore, reference animals are defined as

animals with genotypes and phenotypes, and selection candidates are defined as animals with genotypes but without their own phenotype. The assumption that all reference animals have genotypes is likely to be correct in the near future, as genotyping costs continue to decrease. The aim is to predict the reliability of genomic EBV for crossbred performance for selection candidates of breed A. For the reference population, three scenarios were investigated: (1) the reference population includes only breed A animals (PB–PB), i.e. purebred (PB) phenotypes are used to predict EBV for crossbred (CB) performance of PB selection candidates; (2) the reference population includes only crossbred AB animals (CB–PB), i.e. CB phenotypes are used to predict EBV for CB performance of PB selection candidates; and (iii) the reference population includes both crossbred AB and breed A animals (CB + PB–PB), i.e. CB and PB phenotypes are used to predict EBV for CB performance of PB selection candidates. These scenarios represent situations where crossbred animals are terminal animals in commercial herds of pigs and chickens. The second part of this section describes simulations of the three scenarios used to validate the prediction equations without availability of genotyping data. In the equations below, reference animals are indicated by uppercase letters, while selection candidates are indicated by lowercase letters.

Across-breed SNP genotype models

Equations for predicting the reliability of genomic EBV for crossbred performance using ASGM were developed for the three scenarios. As ASGM is assumed, breed A and crossbred AB animals can be considered as belonging to different populations, assuming the genetic correlation between the PB and CB performance traits (r_{PC}) to be the genetic correlation between these breed A and crossbred AB populations. Therefore, equations for predicting the reliability of genomic EBV for crossbred performance for the three scenarios using ASGM can be derived from previous studies by, for example, Daetwyler et al. [12] and Wientjes et al. [10], without availability of genotyping data, and by VanRaden [15] with availability of genotyping data.

PB–PB scenario

The PB–PB scenario considers breed A animals for both reference animals and selection candidates. Phenotypes are therefore associated with the purebred performance trait, while the trait of interest is the crossbred performance trait. Indeed, selection candidates must be selected to optimize crossbred performance of their crossbred descendants.

Wientjes et al. [10] developed equations for predicting the accuracy of across-population genomic EBV values

without and with availability of genotyping data. Assuming additive gene action, differences in allele substitution effects that underlie the population-specific trait of interest were modelled by the genetic correlation between traits, which implies a multivariate genomic model. Similarly, for the PB–PB scenario, differences in allele substitution effects that underlie the purebred and crossbred performance traits can be considered in terms of the genetic correlation between the purebred and crossbred performance traits (r_{PC}) [10, 16]. Therefore, following Wientjes et al. [10], with availability of genotyping data, the average predicted reliability of genomic EBV for crossbred performance across-breed A selection candidates using breed A reference animals can be computed as follows, based on SI theory:

$$\begin{aligned} r_{P_ASGM_with}^2 &= \frac{1}{N_a} \sum_i r_{P_ASGM_with_i}^2 \\ &= \frac{1}{N_a} \sum_i r_{PC}^2 \frac{\mathbf{G}_{a_i,A} \left(\mathbf{G}_{AA} + \mathbf{I} \frac{1-h_a^2}{h_a^2} \right)^{-1} \mathbf{G}_{A,a_i}}{\mathbf{G}_{a_i,a_i}}, \end{aligned} \quad (1)$$

where N_a is the number of breed A selection candidates; $h_a^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_{eA}^2}$ is the heritability of the purebred performance trait, with σ_a^2 being the genetic variance of the purebred performance trait, and σ_{eA}^2 the residual variance of purebred performance trait; $r_{PC} = \frac{\sigma_{a,c}}{\sqrt{\sigma_a^2 \sigma_c^2}}$ with $\sigma_{a,c}$ being the genetic covariance between the purebred and crossbred performance trait and σ_c^2 the genetic variance of the crossbred performance trait; matrix $\mathbf{G}_{A,A}$ is the $N_a \times N_a$ genomic relationship matrix for the N_a reference animals of breed A; vector $\mathbf{G}_{a_i,A}$ is the row corresponding to the i th selection candidate of breed A of the $N_a \times N_a$ genomic relationship matrix $\mathbf{G}_{A,A}$ between selection candidates of breed A and reference animals of breed A; \mathbf{G}_{a_i,a_i} is the diagonal element corresponding to the i th selection candidate of breed A of the $N_a \times N_a$ genomic relationship matrix $\mathbf{G}_{A,A}$ between selection candidates of breed A; and matrix \mathbf{I} is the identity matrix.

Matrices $\mathbf{G}_{A,A}$, $\mathbf{G}_{a,a}$ and $\mathbf{G}_{A,a}$ are parts of the genomic relationship matrix among all reference animals and selection candidates of breed A, i.e. $\mathbf{G} = \begin{bmatrix} \mathbf{G}_{A,A} & \mathbf{G}_{A,a} \\ \mathbf{G}_{A,a} & \mathbf{G}_{a,a} \end{bmatrix}$. Without loss of generality, and similar to Wientjes et al. [10], matrix \mathbf{G} is computed following the second method of VanRaden [15], i.e., $\mathbf{G} = \frac{\mathbf{Z}\mathbf{Z}'}{m}$ where m is the number of SNP genotypes, and matrix \mathbf{Z} contains the standardized genotypes as $Z_{lk} = \frac{M_{lk} - 2p_k}{\sqrt{2p_k(1-p_k)}}$, with M_{lk} being the SNP genotype (coded as 0 for one homozygous genotype, 1 for the heterozygous genotype, or 2 for the alternate

homozygous genotype) of the l th animal of breed A for the k th locus, and p_k is the allele frequency at the k th locus.

Without availability of genotyping data, the predicted reliability of genomic EBV for crossbred performance of breed A selection candidates and using breed A reference animals can be computed as [10]:

$$r_{P_ASGM_without}^2 = r_{PC}^2 \frac{N_A h_a^2}{N_A h_a^2 + Me_{a,A}}, \quad (2)$$

where $Me_{a,A}$ is the effective number of chromosome segments that are shared between selection candidates and reference animals of breed A. If the term r_{PC}^2 is ignored, or equal to 1, Eq. (2) has the same form as the equation proposed by Daetwyler et al. [12]. One of the assumptions in the derivation of this equation was that the error variance was approximately equal to the phenotypic variance, because only one locus was taken into account at a time and each locus explains only a small part of the additive genetic variance [10, 12, 14]. However, as explained by Daetwyler et al. [12] in the Appendix of their paper, this approximation results in slight underestimation of the predicted reliabilities, because the error variance decreases when multiple loci are used. In Additional file 1 of the current study, we proposed a derivation of Eq. (2) based on the mixed model theory and ignoring the term r_{PC}^2 . We assumed that a single population was used and that effects of all independent loci are estimated simultaneously. Our derivation leads to the same equation as proposed by Daetwyler et al. [12] in the Appendix of their paper, which corrects for the fact that the error variance decreases when multiple loci are used. This derivation using the mixed model theory can be extended for deriving prediction equations using ASGM, and it will be also the basis for deriving prediction equations using BSAM.

CB–PB scenario

The reference population for the CB–PB scenario includes genotyped crossbred AB animals that have phenotypes for the crossbred performance trait. The selection candidates are breed A animals that are related to the reference population and that must be selected to optimize crossbred performance of their crossbred AB descendants. Because the trait of interest is the crossbred performance trait and because allele substitution SNP effects are estimated from crossbred data, the average predicted reliability of genomic EBV for crossbred performance across breed A selection candidates using a crossbred AB reference population can be computed with availability of genotyping data as follows [10]:

$$\begin{aligned} r_{C_ASGM_with}^2 &= \frac{1}{N_a} \sum_i r_{C_ASGM_withi}^2 \\ &= \frac{1}{N_a} \sum_i \frac{\mathbf{G}_{a_i,AB} \left(\mathbf{G}_{AB,AB} + \mathbf{I} \frac{1-h_c^2}{h_c^2} \right)^{-1} \mathbf{G}_{AB,a_i}}{\mathbf{G}_{a_i,a_i}} \end{aligned} \quad (3)$$

where $h_c^2 = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_{ec}^2}$ is the heritability of the crossbred performance trait, with σ_{ec}^2 being the residual variance; matrix $\mathbf{G}_{AB,AB}$ is the $N_{AB} \times N_{AB}$ genomic relationship matrix between the N_{AB} crossbred AB reference animals; and vector $\mathbf{G}_{a_i,AB}$ is the row corresponding to the i th selection candidate of breed A of the $N_a \times N_{AB}$ genomic relationship matrix \mathbf{G}_{aAB} between breed A selection candidates and crossbred AB reference animals. Similarly to Wientjes et al. [10], the genomic relationship matrix between breed A selection candidates and crossbred AB reference animals, \mathbf{G} , is computed following the second method of VanRaden [15] but taking into account that the selection candidates and reference animals belong to two different populations. It then follows that $\mathbf{G} = \begin{bmatrix} \mathbf{G}_{AB,AB} & \mathbf{G}_{AB,a} \\ \mathbf{G}_{a,AB} & \mathbf{G}_{a,a} \end{bmatrix} = \frac{\mathbf{Z}\mathbf{Z}'}{m}$, where m is the number of SNPs and matrix \mathbf{Z} contains the standardized genotypes as $Z_{ljk} = \frac{M_{ljk} - 2p_{jk}}{\sqrt{2p_{jk}(1-p_{jk})}}$, with M_{ljk} being the SNP genotype (coded as previously) of the l th individual from the j th population (i.e., purebred or crossbred) for the k th locus, and p_{jk} is the allele frequency of the j th population at the k th locus.

Without availability of data, an equation that predicts the reliability of genomic EBV for crossbred performance of breed A selection candidates using N_{AB} crossbred AB reference animals can be simply written as follows:

$$r_{C_ASGM_without}^2 = \frac{N_{AB} h_c^2}{N_{AB} h_c^2 + Me_{a,AB}}, \quad (4)$$

where $Me_{a,AB}$ is the effective number of chromosome segments shared by breed A selection candidates and crossbred AB reference animals [10].

CB + PB–PB scenario

The reference population for the CB + PB–PB scenario includes animals of breed A with phenotypes for the purebred performance trait and crossbred AB animals with phenotypes for the crossbred performance trait. The selection candidates are animals of breed A that are related to the reference population. Since the crossbred performance trait is the trait of interest, the average

predicted reliability of genomic EBV across selection candidates for crossbred performance of breed A selection candidates using breed A and crossbred AB reference animals can be computed with availability of genotyping data as follows [10]:

$$\begin{aligned}
 r_{C+P_ASGM_with}^2 &= \frac{1}{N_a} \sum_i r_{C+P_ASGM_with_i}^2 \\
 &= \frac{1}{N_a} \sum_i \frac{1}{\mathbf{G}_{a_i,a_i}} \left[r_{PC} \mathbf{G}_{a_i,A} \quad \mathbf{G}_{a_i,AB} \right] \\
 &\quad \times \left[\begin{array}{cc} \mathbf{G}_{A,A} + \mathbf{I} \frac{1-h_a^2}{h_a^2} & \mathbf{G}_{A,AB} r_{PC} \\ r_{PC} \mathbf{G}_{AB,A} & \mathbf{G}_{AB,AB} + \mathbf{I} \frac{1-h_c^2}{h_c^2} \end{array} \right]^{-1} \left[\begin{array}{c} r_{PC} \mathbf{G}_{a_i,A} \\ \mathbf{G}_{a_i,AB} \end{array} \right].
 \end{aligned} \tag{5}$$

Without availability of genotyping data, the prediction equation for the reliability of genomic EBV for crossbred performance of breed A selection candidates using breed A and crossbred AB reference animals can be written as follows [14]:

$$\begin{aligned}
 r_{C+P_ASGM_without}^2 &= \left[r_{PC} \sqrt{\frac{h_a^2}{Me_{a,A}}} \quad \sqrt{\frac{h_c^2}{Me_{a,AB}}} \right] \\
 &\quad \times \left[\begin{array}{cc} \frac{h_a^2}{Me_{a,A}} + \frac{1}{N_a} & r_{PC} \sqrt{\frac{h_a^2 h_c^2}{Me_{a,A} Me_{a,AB}}} \\ r_{PC} \sqrt{\frac{h_a^2 h_c^2}{Me_{a,A} Me_{a,AB}}} & \frac{h_c^2}{Me_{a,AB}} + \frac{1}{N_{AB}} \end{array} \right]^{-1} \\
 &\quad \times \left[\begin{array}{c} r_{PC} \sqrt{\frac{h_a^2}{Me_{a,A}}} \\ \sqrt{\frac{h_c^2}{Me_{a,AB}}} \end{array} \right].
 \end{aligned} \tag{6}$$

Breed-specific allele substitution models

In crossbred populations, SNP effects may be breed-specific due to a number of factors [4], including different extents of LD between SNP and QTL between breeds, which can be accommodated by using BSAM, which fits breed-specific allele substitution effects [3, 4]. In this section, it is assumed that the breed origin of SNP alleles is known, as required by BSAM. Moreover, only the CB–PB and CB + PB–PB scenarios are considered, since the PB–PB scenario involves data on only one breed. To our knowledge, equations for predicting the reliability of genomic EBV using BSAM have not previously been developed.

CB–PB scenario

For the CB–PB scenario, assuming that each individual has one phenotypic record corrected for all effects other than the additive genetic effects, BSAM for the crossbred performance trait can be written as follows [3, 4]:

$$\mathbf{y}_{AB} = \mathbf{Z}_{AB}^{(A)} \boldsymbol{\beta}_c^{(A)} + \mathbf{Z}_{AB}^{(B)} \boldsymbol{\beta}_c^{(B)} + \mathbf{e}_{AB},$$

where \mathbf{y}_{AB} is the vector of corrected records of crossbred performance; $\mathbf{Z}_{AB}^{(A)}$ ($\mathbf{Z}_{AB}^{(B)}$) contains the standardized breed A (B) SNP alleles of each crossbred animal; $\boldsymbol{\beta}_c^{(A)}$ ($\boldsymbol{\beta}_c^{(B)}$)

is the vector of breed A (B)-specific allele substitution effects for all SNPs; and \mathbf{e}_{AB} is the residual vector. Entries of matrix $\mathbf{Z}_{AB}^{(A)}$ are defined as $Z_{ABlk}^{(A)} = \frac{M_{lk} - p_{Ak}}{\sqrt{2p_{Ak}(1-p_{Ak})}}$, where element M_{lk} is set to 0 or 1 when the k th locus of the l th individual has breed A allele 1 or 2, respectively; and p_{Ak} is the frequency at the k th locus for breed A. Matrix $\mathbf{Z}_{AB}^{(B)}$ is defined similarly. Expectations and variances of $\boldsymbol{\beta}_c^{(A)}$ and $\boldsymbol{\beta}_c^{(B)}$ are assumed to be $E \begin{bmatrix} \boldsymbol{\beta}_c^{(A)} \\ \boldsymbol{\beta}_c^{(B)} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$ and

$$\text{Var} \begin{bmatrix} \boldsymbol{\beta}_c^{(A)} \\ \boldsymbol{\beta}_c^{(B)} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \sigma_{\beta_c^{(A)}}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \sigma_{\beta_c^{(B)}}^2 \end{bmatrix} = \begin{bmatrix} \mathbf{I} \frac{\sigma_{c_A}^2}{m} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \frac{\sigma_{c_B}^2}{m} \end{bmatrix}$$

where $\sigma_{\beta_c^{(A)}}^2$ ($\sigma_{\beta_c^{(B)}}^2$) is the variance of the breed A (B)-specific allele substitution effect, and $\sigma_{c_A}^2$ ($\sigma_{c_B}^2$) is the additive genetic variance due to alleles from population A (B) in the crossbred population for the crossbred performance trait [3, 4].

Equivalently, BSAM for the crossbred performance trait can be written as [3]:

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

where $\mathbf{c}_{AB}^{(A)} = \mathbf{Z}_{AB}^{(A)} \boldsymbol{\beta}_c^{(A)}$ ($\mathbf{c}_{AB}^{(B)} = \mathbf{Z}_{AB}^{(B)} \boldsymbol{\beta}_c^{(B)}$) is the vector of breed A (B) of origin additive genetic effects for the crossbred performance trait. It then follows that expectations and variances of $\mathbf{c}_{AB}^{(A)}$ and $\mathbf{c}_{AB}^{(B)}$ are defined as

$$\begin{aligned}
 E \begin{bmatrix} \mathbf{c}_{AB}^{(A)} \\ \mathbf{c}_{AB}^{(B)} \end{bmatrix} &= \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \text{ and} \\
 \text{Var} \begin{bmatrix} \mathbf{c}_{AB}^{(A)} \\ \mathbf{c}_{AB}^{(B)} \end{bmatrix} &= \begin{bmatrix} \mathbf{Z}_{AB}^{(A)} \mathbf{Z}_{AB}^{(A)'} \frac{\sigma_{c_A}^2}{m} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{AB}^{(B)} \mathbf{Z}_{AB}^{(B)'} \frac{\sigma_{c_B}^2}{m} \end{bmatrix} \\
 &= \begin{bmatrix} \mathbf{G}_{AB,AB}^{(A)} \sigma_{c_A}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{AB,AB}^{(B)} \sigma_{c_B}^2 \end{bmatrix},
 \end{aligned}$$

where $\mathbf{G}_{AB,AB}^{(A)}$ ($\mathbf{G}_{AB,AB}^{(B)}$) is the breed A (B) partial genomic relationship among the N_{AB} crossbred AB animals [3]. These assumptions imply that $\mathbf{c}_{AB}^{(A)}$ and $\mathbf{c}_{AB}^{(B)}$ as well as $\boldsymbol{\beta}_c^{(A)}$ and $\boldsymbol{\beta}_c^{(B)}$, are independent of each other.

Based on SI theory, genomic EBV for crossbred performance of selection candidates from breed A ($\hat{c}_a^{(A)}$) can be predicted from records of crossbred AB reference animals as follows:

$$\hat{c}_a^{(A)} = \text{Cov}(\mathbf{c}_a^{(A)}, \mathbf{y}_{AB}) (\text{Var}(\mathbf{y}_{AB}))^{-1} \mathbf{y}_{AB}.$$

Using the model description, it can then be shown that the variance of \mathbf{y}_{AB} is equal to:

$$\text{Var}(\mathbf{y}_{AB}) = \mathbf{G}_{ABAB}^{(A)} \sigma_{c_A}^2 + \mathbf{G}_{ABAB}^{(B)} \sigma_{c_B}^2 + \mathbf{I} \sigma_{e_{AB}}^2,$$

and that the covariance between $\mathbf{c}_a^{(A)}$ and \mathbf{y}_{AB} is equal to:

$$\begin{aligned} \text{Cov}\left(\mathbf{c}_a^{(A)}, \mathbf{y}_{AB}\right) &= \text{Cov}\left(\mathbf{c}_a^{(A)}, \mathbf{c}_{AB}^{(A)} + \mathbf{c}_{AB}^{(B)} + \mathbf{e}_{AB}\right) \\ &= \text{Cov}\left(\mathbf{c}_a^{(A)}, \mathbf{c}_{AB}^{(A)}\right) + \text{Cov}\left(\mathbf{c}_a^{(A)}, \mathbf{c}_{AB}^{(B)}\right) \\ &\quad + \text{Cov}\left(\mathbf{c}_a^{(A)}, \mathbf{e}_{AB}\right) \\ &= \text{Cov}\left(\mathbf{c}_a^{(A)}, \mathbf{c}_{AB}^{(A)}\right) = \mathbf{G}_{a,AB}^{(A)} \sigma_{c_A}^2, \end{aligned}$$

with matrix $\mathbf{G}_{a,AB}^{(A)} = \frac{1}{m} \mathbf{Z}_a^{(A)} \mathbf{Z}_{AB}^{(A)'}$ being the breed A-specific partial genomic relationship matrix between the N_a selection candidates of breed A and the N_{AB} crossbred AB reference animals.

The reliability of $\hat{c}_{a_i}^{(A)}$ of the i th selection candidate of breed A is then equal to:

$$\begin{aligned} r_{C_BSAM_with_i}^2 &= \frac{\left(\text{Cov}\left(\hat{c}_{a_i}^{(A)}, c_{a_i}^{(A)}\right)\right)^2}{\text{Var}\left(\hat{c}_{a_i}^{(A)}\right) \text{Var}\left(c_{a_i}^{(A)}\right)} = \frac{\text{Var}\left(\hat{c}_{a_i}^{(A)}\right)}{\text{Var}\left(c_{a_i}^{(A)}\right)} \\ &= \frac{1}{\mathbf{G}_{a_i,a_i}^{(A)}} \mathbf{G}_{a_i,AB}^{(A)} \left(\mathbf{G}_{AB,AB}^{(A)} + \mathbf{G}_{AB,AB}^{(B)} \frac{\sigma_{c_B}^2}{\sigma_{c_A}^2} + \mathbf{I} \frac{\sigma_e^2}{\sigma_{c_A}^2} \right)^{-1} \mathbf{G}_{AB,a_i}^{(A)}. \end{aligned}$$

With availability of genotyping data, the average predicted reliability of genomic EBV across all breed A selection candidates is equal to:

$$\begin{aligned} r_{C_BSAM_with}^2 &= \frac{1}{N_a} \sum_i r_{C_BSAM_with_i}^2 = \frac{1}{N_a} \sum_i \frac{1}{\mathbf{G}_{a_i,a_i}^{(A)}} \mathbf{G}_{a_i,AB}^{(A)} \\ &\quad \times \left(\mathbf{G}_{AB,AB}^{(A)} + \mathbf{G}_{AB,AB}^{(B)} \frac{h_{c_B}^2}{h_{c_A}^2} + \mathbf{I} \frac{1 - \frac{1}{2}h_{c_A}^2 - \frac{1}{2}h_{c_B}^2}{h_{c_A}^2} \right)^{-1} \mathbf{G}_{AB,a_i}^{(A)}, \end{aligned} \tag{7}$$

where $h_{c_A}^2 = \frac{\sigma_{c_A}^2}{\frac{\sigma_{c_A}^2}{2} + \frac{\sigma_{c_B}^2}{2} + \sigma_{e_{AB}}^2}$ ($h_{c_B}^2 = \frac{\sigma_{c_B}^2}{\frac{\sigma_{c_A}^2}{2} + \frac{\sigma_{c_B}^2}{2} + \sigma_{e_{AB}}^2}$) is the breed A (B)-specific heritability of crossbred performance.

Since no equation has previously been proposed to predict the reliability of genomic EBV for BSAM without availability of genotyping data, here, we put forward a derivation based on mixed model theory [17], assuming that allele substitution effects for breeds A and B are estimated simultaneously. Equivalence between the mixed model and SI theories has previously been shown under certain conditions, including the use of the same estimates of the fixed effects [15, 17, 18]. Our derivation of the equation for predicting the reliability of genomic EBV for BSAM without availability of genotyping data [i.e., Eq. (8) below] is detailed in Additional file 2, and the result is briefly described in the following.

Consider N_{AB} unrelated genotyped crossbred AB reference animals. For simplicity, it is assumed that the breed A-specific effect $\beta_{c_k}^{*(A)}$ of each k th independent locus

explains an equal amount of the breed A-specific additive genetic variance $\sigma_{c_A}^2$, i.e., $\sigma_{c_A}^2 = Me_{a,AB}^{(A)} \sigma_{\beta_c^{*(A)}}^2$, with $Me_{a,AB}^{(A)}$ being the effective number of chromosome segments underlying the crossbred performance trait for breed A and segregating in both breed A selection candidates and crossbred AB reference animals. The same assumption is made for the breed B-specific effect $\beta_{c_k}^{*(B)}$. The genomic EBV ($c_{a_i}^{(A)}$) for the i th selection candidate of breed A can be predicted as follows:

$$\hat{c}_{a_i}^{(A)} = \mathbf{z}_{a_i}^{*(A)} \hat{\beta}_c^{*(A)},$$

where $\mathbf{z}_{a_i}^{*(A)}$ is a vector of the standardized genotypes for the $Me_{a,AB}^{(A)}$ independent loci of the i th selection candidate of breed A and $\hat{\beta}_c^{*(A)}$ is the vector of the predictions of $\beta_c^{*(A)}$. Following mixed model theory [17, 19], the reliability of $\hat{c}_{a_i}^{(A)}$ can be computed from the prediction error variance, $\text{Var}\left(\hat{c}_{a_i}^{(A)} - c_{a_i}^{(A)}\right)$, and is equal to:

$$\begin{aligned} r_{C_BSAM_without_i}^2 &= 1 - \frac{\text{Var}\left(\hat{c}_{a_i}^{(A)} - c_{a_i}^{(A)}\right)}{\text{Var}\left(c_{a_i}^{(A)}\right)} \\ &= \frac{\text{Var}\left(\hat{c}_{a_i}^{(A)}\right)}{\text{Var}\left(c_{a_i}^{(A)}\right)} = \frac{\text{Var}\left(\mathbf{z}_{a_i}^{*(A)} \hat{\beta}_c^{*(A)}\right)}{\text{Var}\left(\mathbf{z}_{a_i}^{*(A)} \beta_c^{*(A)}\right)} \\ &= \frac{\mathbf{z}_{a_i}^{*(A)} \text{Var}\left(\hat{\beta}_c^{*(A)}\right) \mathbf{z}_{a_i}^{*(A)'}}{\mathbf{z}_{a_i}^{*(A)} \text{Var}\left(\beta_c^{*(A)}\right) \mathbf{z}_{a_i}^{*(A)'}}. \end{aligned}$$

Assuming that the allele substitution effect $\beta_{c_k}^{*(A)}$ of each k th independent locus explains an equal amount of the breed A-specific additive genetic variance $\sigma_{c_A}^2$ and that the reliability of the estimated effect, $r_{\beta_c^{*(A)}}^2$ is the same for each locus, it follows that:

$$r_{C_BSAM_without_i}^2 = \frac{\text{Var}\left(\hat{\beta}_{c_k}^{*(A)}\right)}{\text{Var}\left(\beta_{c_k}^{*(A)}\right)} = r_{\beta_c^{*(A)}}^2.$$

Reliability $r_{\beta_c^{*(A)}}^2$ can be approximated as follows. Let $\widehat{\mathbf{y}}_{AB}^*$ be the vector of phenotypes corrected for all other fixed effects for the breed A-specific allele substitution effects other than the k th effect, $\hat{\beta}_{c_{\neq k}}^{*(A)}$, as well as for the breed B-specific allele substitution effects, $\hat{\beta}_c^{*(B)}$. The prediction of $\beta_{c_k}^{*(A)}$ for the k th locus can then be performed using the following model:

$$\widehat{\mathbf{y}}_{AB}^* = \mathbf{z}_{AB_k}^{*(A)} \beta_{c_k}^{*(A)} + \mathbf{e}_{AB_k},$$

where vector $\mathbf{z}_{AB_k}^{*(A)}$ contains the standardized breed A alleles of crossbred AB reference animals, and \mathbf{e}_{AB_k} is the residual vector.

The variance of $\widehat{\mathbf{y}}_{AB}^*$ is equal to:

$$\begin{aligned} \text{Var}\left(\widehat{\mathbf{y}}_{AB}^*\right) &= \text{Var}\left(\mathbf{z}_{AB_k}^{*(A)}\beta_{c_k}^{*(A)}\right) + \text{Var}\left(\mathbf{e}_{AB_k}\right) \\ &= \mathbf{z}_{AB_k}^{*(A)}\mathbf{z}_{AB_k}^{*(A)'}\sigma_{\beta_c}^{2*(A)} + \text{Var}\left(\mathbf{e}_{AB_k}\right), \end{aligned}$$

from which it follows that, after some algebra and assuming unrelated genotyped animals (see Additional file 2 for details):

$$\begin{aligned} \text{Var}\left(\mathbf{e}_{AB_k}\right) &= \text{Var}\left(\widehat{\mathbf{y}}_{AB}^*\right) - \mathbf{z}_{AB_k}^{*(A)}\mathbf{z}_{AB_k}^{*(A)'}\sigma_{\beta_c}^{2*(A)} \\ &= \mathbf{I}\left(\frac{\sigma_{c_A}^2}{2}\left(1 - r_{\beta_c}^{2*(A)}\right) + \frac{\sigma_{c_B}^2}{2}\left(1 - r_{\beta_c}^{2*(B)}\right) + \sigma_{e_{AB}}^2\right) \\ &\approx \mathbf{I}\left(\sigma_{P_{AB}}^2 - \frac{\sigma_{c_A}^2}{2}r_{\beta_c}^{2*(A)} - \frac{\sigma_{c_B}^2}{2}r_{\beta_c}^{2*(B)}\right) = \mathbf{I}\sigma_{\varepsilon_{AB}}^2, \end{aligned}$$

where $\sigma_{P_{AB}}^2$ ($\sigma_{\varepsilon_{AB}}^2$) is the phenotypic (residual) variance of the crossbred performance trait.

Therefore, following mixed model theory [17, 19], the prediction of $\hat{\beta}_{c_k}^{*(A)}$, $\hat{\beta}_{c_k}^{*(A)}$, is equal to:

$$\hat{\beta}_{c_k}^{*(A)} = \left(\mathbf{z}_{AB_k}^{*(A)'}\mathbf{z}_{AB_k}^{*(A)}\sigma_{\varepsilon_{AB}}^{-2} + \sigma_{\beta_c}^{-2*(A)}\right)^{-1}\sigma_{\varepsilon_{AB}}^{-2}\mathbf{z}_{AB_k}^{*(A)'}\widehat{\mathbf{y}}_{AB}^*$$

and the reliability of $\hat{\beta}_{c_k}^{*(A)}$ is equal to (see Additional file 2 for more details):

$$\begin{aligned} r_{\beta_c}^{2*(A)} &= \frac{\text{Var}\left(\beta_{c_k}^{*(A)}\right) - \text{Var}\left(\hat{\beta}_{c_k}^{*(A)} - \beta_{c_k}^{*(A)}\right)}{\text{Var}\left(\beta_{c_k}^{*(A)}\right)} \\ &= \frac{N_{AB}\sigma_{\varepsilon_{AB}}^{-2}\sigma_{\beta_c}^{2*(A)}}{N_{AB}\sigma_{\varepsilon_{AB}}^{-2}\sigma_{\beta_c}^{2*(A)} + 2}. \end{aligned}$$

It then follows that, without availability of genotyping data, the predicted reliability of the genomic EBV for breed A selection candidates is equal to (see Additional file 2 for more details):

$$\begin{aligned} r_{C_BSAM_without}^2 &= r_{\beta_c}^{2*(A)} \\ &= \frac{N_{AB}h_{c_A}^2}{N_{AB}h_{c_A}^2 + 2Me_{a,AB}^{(A)}\left(1 - \frac{1}{2}h_{c_A}^2r_{c_A}^{2*(A)} - \frac{1}{2}h_{c_B}^2r_{c_A}^{2*(B)}\right)}. \end{aligned}$$

By ignoring the term $\left(-\frac{1}{2}h_{c_A}^2r_{c_A}^{2*(A)} - \frac{1}{2}h_{c_B}^2r_{c_A}^{2*(B)}\right)$ for low $h_{c_A}^2$ and $h_{c_B}^2$, the prediction equation simplifies to:

$$r_{C_BSAM_without}^2 = \frac{N_{AB}h_{c_A}^2}{N_{AB}h_{c_A}^2 + 2Me_{a,AB}^{(A)}}. \quad (8)$$

CB + PB–PB scenario

For the CB + PB–PB scenario, the reference population includes breed A and crossbred AB reference animals, each with their own phenotypes. The BSAM for the crossbred performance trait, assuming that each individual has one record corrected for all effects other than additive genetic effects, can be written as follows [3, 4]:

$$\begin{aligned} \begin{bmatrix} \mathbf{y}_A \\ \mathbf{y}_{AB} \end{bmatrix} &= \begin{bmatrix} \mathbf{Z}_A^{(A)} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{AB}^{(A)} \end{bmatrix} \begin{bmatrix} \beta_a^{(A)} \\ \beta_c^{(A)} \end{bmatrix} \\ &+ \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{AB}^{(B)} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \beta_c^{(B)} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_A \\ \mathbf{e}_{AB} \end{bmatrix}, \end{aligned}$$

where \mathbf{y}_A is the vector of corrected records of purebred performance, $\mathbf{Z}_A^{(A)}$ contains the standardized SNP genotypes of breed A reference animals, and $\beta_a^{(A)}$ is the vector of breed A allele substitution effects for all SNPs for purebred performance.

Equivalently, the previous BSAM can be written as [3]:

$$\begin{bmatrix} \mathbf{y}_A \\ \mathbf{y}_{AB} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_A^{(A)} \\ \mathbf{c}_{AB}^{(A)} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{c}_{AB}^{(B)} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_A \\ \mathbf{e}_{AB} \end{bmatrix},$$

where $\mathbf{a}_A^{(A)} = \mathbf{Z}_A^{(A)}\beta_a^{(A)}$ is the vector of additive genetic effects for the purebred performance trait.

Expectations and variances and covariances of $\mathbf{a}_A^{(A)}$, $\mathbf{c}_{AB}^{(A)}$

and $\mathbf{c}_{AB}^{(B)}$ are assumed to be $E\begin{bmatrix} \mathbf{a}_A^{(A)} \\ \mathbf{c}_{AB}^{(A)} \\ \mathbf{c}_{AB}^{(B)} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$ and

$$\begin{aligned} \text{Var} \begin{bmatrix} \mathbf{a}_A^{(A)} \\ \mathbf{c}_{AB}^{(A)} \\ \mathbf{c}_{AB}^{(B)} \end{bmatrix} &= \begin{bmatrix} \mathbf{Z}_A^{(A)}\mathbf{Z}_A^{(A)'}\frac{\sigma_{a_A}^2}{m} & \mathbf{Z}_A^{(A)}\mathbf{Z}_{AB}^{(A)'}\frac{\sigma_{a_Ac_A}}{m} & \mathbf{0} \\ \mathbf{Z}_{AB}^{(A)}\mathbf{Z}_A^{(A)'}\frac{\sigma_{a_Ac_A}}{m} & \mathbf{Z}_{AB}^{(A)}\mathbf{Z}_{AB}^{(A)'}\frac{\sigma_{c_A}^2}{m} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}_{AB}^{(B)}\mathbf{Z}_{AB}^{(B)'}\frac{\sigma_{c_B}^2}{m} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{G}_{A,A}^{(A)}\sigma_{a_A}^2 & \mathbf{G}_{A,AB}^{(A)}\sigma_{a_Ac_A} & \mathbf{0} \\ \mathbf{G}_{AB,A}^{(A)}\sigma_{a_Ac_A} & \mathbf{G}_{AB,AB}^{(A)}\sigma_{c_A}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G}_{AB,AB}^{(B)}\sigma_{c_B}^2 \end{bmatrix}, \end{aligned}$$

where $\mathbf{G}_{A,A}^{(A)}$ is the breed A genomic relationship matrix between N_A reference animals of breed A and $\mathbf{G}_{A,AB}^{(A)}$ is the breed A-specific partial genomic relationship matrix between N_A selection candidates of breed A and N_{AB} crossbred AB reference animals [3].

Based on the SI theory, genomic EBV for the crossbred performance trait for breed A selection candidates ($\mathbf{c}_a^{(A)}$) can be predicted from records of breed A reference animals and of crossbred AB reference animals:

$$\hat{\mathbf{c}}_a^{(A)} = \text{Cov}\left(\mathbf{c}_a^{(A)}, \mathbf{y}\right)\left(\text{Var}\left(\mathbf{y}\right)\right)^{-1}\mathbf{y},$$

with $\mathbf{y} = \begin{bmatrix} \mathbf{y}_A \\ \mathbf{y}_{AB} \end{bmatrix}$.

Based on the model description, the variance of \mathbf{y} is equal to:

$$\begin{aligned} \text{Var}(\mathbf{y}) &= \text{Var}\left(\begin{bmatrix} \mathbf{y}_A \\ \mathbf{y}_{AB} \end{bmatrix}\right) \\ &= \begin{bmatrix} \mathbf{G}_{AA}^{(A)}\sigma_{c_A}^2 + \mathbf{I}\sigma_{e_A}^2 & \mathbf{G}_{A,AB}^{(A)}\sigma_{a_A c_A} \\ \mathbf{G}_{A,AB}^{(A)}\sigma_{a_A c_A} & \mathbf{G}_{AB,AB}^{(A)}\sigma_{c_A}^2 + \mathbf{G}_{AB,AB}^{(B)}\sigma_{c_B}^2 + \mathbf{I}\sigma_{e_{AB}}^2 \end{bmatrix}, \end{aligned}$$

since:

$$\begin{aligned} \text{Cov}(\mathbf{y}_A, \mathbf{y}_{AB}) &= \text{Cov}\left(\mathbf{a}_A^{(A)} + \mathbf{e}_A, \mathbf{c}_{AB}^{(A)} + \mathbf{c}_{AB}^{(B)} + \mathbf{e}_{AB}\right) \\ &= \text{Cov}\left(\mathbf{a}_A^{(A)}, \mathbf{c}_{AB}^{(A)} + \mathbf{c}_{AB}^{(B)}\right) \\ &= \text{Cov}\left(\mathbf{a}_A^{(A)}, \mathbf{c}_{AB}^{(A)}\right) = \mathbf{G}_{A,AB}^{(A)}\sigma_{a_A c_A}. \end{aligned}$$

Similarly, the covariance between $\mathbf{c}_a^{(A)}$ and \mathbf{y} is equal to:

$$\text{Cov}\left(\mathbf{c}_a^{(A)}, \mathbf{y}\right) = \begin{bmatrix} \mathbf{G}_{a_A, A}^{(A)}\sigma_{a_A c_A} & \mathbf{G}_{a_A, AB}^{(A)}\sigma_{c_A}^2 \end{bmatrix}.$$

The reliability of $\hat{c}_{a_i}^{(A)}$ of the i th selection candidate of breed A is then equal to:

$$\begin{aligned} r_{C+P_BSAM_with_i}^2 &= \frac{\left(\text{Cov}\left(\hat{c}_{a_i}^{(A)}, c_{a_i}^{(A)}\right)\right)^2}{\text{Var}\left(\hat{c}_{a_i}^{(A)}\right)\text{Var}\left(c_{a_i}^{(A)}\right)} = \frac{\text{Var}\left(\hat{c}_{a_i}^{(A)}\right)}{\text{Var}\left(c_{a_i}^{(A)}\right)} \\ &= \frac{1}{\text{Var}\left(c_{a_i}^{(A)}\right)} \begin{bmatrix} \mathbf{G}_{a_i, A}^{(A)}\sigma_{a_A c_A} & \mathbf{G}_{a_i, AB}^{(A)}\sigma_{c_A}^2 \end{bmatrix} \\ &\quad \times \begin{bmatrix} \mathbf{G}_{A, A}^{(A)}\sigma_{a_A}^2 + \mathbf{I}\sigma_{e_A}^2 & \mathbf{G}_{A, AB}^{(A)}\sigma_{a_A c_A} \\ \mathbf{G}_{A, AB}^{(A)}\sigma_{a_A c_A} & \mathbf{G}_{AB, AB}^{(A)}\sigma_{c_A}^2 + \mathbf{G}_{AB, AB}^{(B)}\sigma_{c_B}^2 + \mathbf{I}\sigma_{e_{AB}}^2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{G}_{A, a_i}^{(A)}\sigma_{a_A c_A} \\ \mathbf{G}_{AB, a_i}^{(A)}\sigma_{c_A}^2 \end{bmatrix} \\ &= \frac{1}{\mathbf{G}_{a_i, a_i}^{(A)}} \begin{bmatrix} r_{PC}^{(A)}\mathbf{G}_{a_i, A}^{(A)} & \mathbf{G}_{a_i, AB}^{(A)} \end{bmatrix} \\ &\quad \times \begin{bmatrix} \mathbf{G}_{A, A}^{(A)} + \mathbf{I}\frac{1-h_{a_A}^2}{h_{a_A}^2} & r_{PC}^{(A)}\mathbf{G}_{A, AB}^{(A)} \\ r_{PC}^{(A)}\mathbf{G}_{AB, A}^{(A)} & \mathbf{G}_{AB, AB}^{(A)} + \mathbf{G}_{AB, AB}^{(B)}\frac{h_{c_B}^2}{h_{c_A}^2} + \mathbf{I}\frac{1-\frac{1}{2}h_{c_A}^2 - \frac{1}{2}h_{c_B}^2}{h_{c_A}^2} \end{bmatrix}^{-1} \begin{bmatrix} r_{PC}^{(A)}\mathbf{G}_{A, a_i}^{(A)} \\ \mathbf{G}_{AB, a_i}^{(A)} \end{bmatrix}, \end{aligned}$$

where the breed A-specific genetic correlation between purebred and crossbred performance traits $r_{PC}^{(A)}$ is equal to $r_{PC}^{(A)} = \frac{\sigma_{a_A c_A}}{\sqrt{\sigma_{a_A}^2\sigma_{c_A}^2}}$.

With availability of genotyping data, the average predicted reliability of genomic EBV across all breed A selection candidates is therefore equal to:

$$\begin{aligned} r_{C+P_BSAM_with}^2 &= \frac{1}{N_a} \sum_i r_{C+P_BSAM_with_i}^2 = \frac{1}{\mathbf{G}_{a_i, a_i}^{(A)}} \begin{bmatrix} r_{PC}^{(A)}\mathbf{G}_{a_i, A}^{(A)} & \mathbf{G}_{a_i, AB}^{(A)} \end{bmatrix} \\ &\quad \times \frac{1}{N_a} \sum_i \begin{bmatrix} \mathbf{G}_{A, A}^{(A)} + \mathbf{I}\frac{1-h_{a_A}^2}{h_{a_A}^2} & r_{PC}^{(A)}\mathbf{G}_{A, AB}^{(A)} \\ r_{PC}^{(A)}\mathbf{G}_{AB, A}^{(A)} & \mathbf{G}_{AB, AB}^{(A)} + \mathbf{G}_{AB, AB}^{(B)}\frac{h_{c_B}^2}{h_{c_A}^2} + \mathbf{I}\frac{1-\frac{1}{2}h_{c_A}^2 - \frac{1}{2}h_{c_B}^2}{h_{c_A}^2} \end{bmatrix}^{-1} \\ &\quad \times \begin{bmatrix} r_{PC}^{(A)}\mathbf{G}_{A, a_i}^{(A)} \\ \mathbf{G}_{AB, a_i}^{(A)} \end{bmatrix} \end{aligned} \tag{9}$$

Without availability of genotyping data, the prediction equation for the reliability of genomic EBV based on BSAM, $r_{C+P_BSAM_without_i}^2$ can be derived similarly to the prediction equation for the CB–PB scenario, $r_{C_BSAM_with_i}^2$. The derivation is based on mixed model theory and assumes that independent allele substitution effects for breeds A and B for both purebred and crossbred performances were estimated simultaneously. The detailed derivation can be found in Additional file 3.

Consider N_A unrelated genotyped breed A reference animals and N_{AB} unrelated genotyped crossbred AB reference animals. Similar to the CB–PB scenario, the genomic EBV ($c_{a_i}^{(A)}$) for the i th selection candidate of breed A can be predicted as $\hat{c}_{a_i}^{(A)} = \mathbf{z}_{a_i}^{*(A)}\hat{\beta}_c^{*(A)}$ and its reliability is equal to:

$$r_{C+P_BSAM_without_i}^2 = \frac{\text{Var}\left(\hat{c}_{a_i}^{(A)}\right)}{\text{Var}\left(c_{a_i}^{(A)}\right)} = \frac{\text{Var}\left(\hat{\beta}_{c_k}^{*(A)}\right)}{\text{Var}\left(\beta_{c_k}^{*(A)}\right)} = r_{\beta_c^{*(A)}}^2.$$

Reliability $r_{\beta_c^{*(A)}}^2$ can be approximated as follows. The prediction of $\beta_{c_k}^{*(A)}$ for the k th independent locus can be performed using the phenotypes of both purebred and crossbred performances, $\begin{bmatrix} \mathbf{y}_A^* \\ \mathbf{y}_{AB}^* \end{bmatrix}$, corrected for all other

fixed effects, as well as for the breed B-specific allele substitution effects and correlated effects, using the model:

$$\begin{bmatrix} \widehat{\mathbf{y}}_A^* \\ \widehat{\mathbf{y}}_{AB}^* \end{bmatrix} = \begin{bmatrix} \mathbf{z}_{A_k}^{*(A)} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{AB_k}^{*(A)} \end{bmatrix} \begin{bmatrix} \widehat{\beta}_{a_k}^{*(A)} \\ \widehat{\beta}_{c_k}^{*(A)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{A_k} \\ \boldsymbol{\varepsilon}_{AB_k} \end{bmatrix},$$

where $\boldsymbol{\varepsilon}_{A_k}$ is the residual vector. For simplicity, we will assume that $Var\left(\begin{bmatrix} \boldsymbol{\varepsilon}_{A_k} \\ \boldsymbol{\varepsilon}_{AB_k} \end{bmatrix}\right) = \begin{bmatrix} \mathbf{I}\sigma_{\varepsilon_A}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_{\varepsilon_{AB}}^2 \end{bmatrix}$, with $\sigma_{\varepsilon_A}^2$

being the residual variance associated with \mathbf{y}_A^* . Then, following mixed model theory [17, 19], the prediction of

$\begin{bmatrix} \widehat{\beta}_{a_k}^{*(A)} \\ \widehat{\beta}_{c_k}^{*(A)} \end{bmatrix}$ is equal to:

$$\begin{bmatrix} \widehat{\beta}_{a_k}^{*(A)} \\ \widehat{\beta}_{c_k}^{*(A)} \end{bmatrix} = \left(\begin{bmatrix} \mathbf{z}_{A_k}^{*(A)'} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{AB_k}^{*(A)'} \end{bmatrix} \begin{bmatrix} \mathbf{I}\sigma_{\varepsilon_A}^{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_{\varepsilon_{AB}}^{-2} \end{bmatrix} \begin{bmatrix} \mathbf{z}_{A_k}^{*(A)} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{AB_k}^{*(A)} \end{bmatrix} + \begin{bmatrix} \sigma_{\beta_a}^{2*(A)} & \sigma_{\beta_a}^{*(A)}\sigma_{\beta_c}^{*(A)} \\ \sigma_{\beta_a}^{*(A)}\sigma_{\beta_c}^{*(A)} & \sigma_{\beta_c}^{2*(A)} \end{bmatrix}^{-1} \right)^{-1} \begin{bmatrix} \mathbf{z}_{A_k}^{*(A)'} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{AB_k}^{*(A)'} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{y}}_A^* \\ \widehat{\mathbf{y}}_{AB}^* \end{bmatrix},$$

and the reliability of $\widehat{\beta}_{c_k}^{*(A)}$ is equal to:

$$r_{\beta_c}^{2*(A)} = \frac{Var(\widehat{\beta}_{c_k}^{*(A)}) - Var(\widehat{\beta}_{c_k}^{*(A)} - \beta_{c_k}^{*(A)})}{Var(\widehat{\beta}_{c_k}^{*(A)})} = \frac{\sigma_{\beta_c}^{2*(A)} - PEV_{\widehat{\beta}_{c_k}^{*(A)}}}{\sigma_{\beta_c}^{2*(A)}},$$

where $PEV_{\widehat{\beta}_{c_k}^{*(A)}}$ is the prediction error variance of $\widehat{\beta}_{c_k}^{*(A)}$ and is equal to the diagonal element of the inverse of the left-hand side of the mixed model equations associated with the prediction of $\begin{bmatrix} \widehat{\beta}_{a_k}^{*(A)} \\ \widehat{\beta}_{c_k}^{*(A)} \end{bmatrix}$. After some algebra,

which is detailed in Additional file 3, it follows that the predicted reliability of genomic EBV for breed A selection candidates without data is equal to:

$$r_{C+P_BSAM_without}^2 = r_{\beta_c}^{2*(A)} = \left[r_{PC}^{(A)} \sqrt{\frac{h_{a,A}^2}{Me_{a,A}}} \sqrt{\frac{h_{c,A}^2}{2Me_{a,AB}^{(A)}}} \right] \times \begin{bmatrix} \frac{h_{a,A}^2}{Me_{a,A}} + \frac{1}{N_A} & r_{PC}^{(A)} \sqrt{\frac{h_{c,A}^2}{2Me_{a,AB}^{(A)}} \frac{h_{a,A}^2}{Me_{a,A}}} \\ r_{PC}^{(A)} \sqrt{\frac{h_{c,A}^2}{2Me_{a,AB}^{(A)}} \frac{h_{a,A}^2}{Me_{a,A}}} & \frac{h_{c,A}^2}{2Me_{a,AB}^{(A)}} + \frac{1}{N_{AB}} \end{bmatrix}^{-1} \times \begin{bmatrix} r_{PC}^{(A)} \sqrt{\frac{h_{a,A}^2}{Me_{a,A}}} \\ \sqrt{\frac{h_{c,A}^2}{2Me_{a,AB}^{(A)}}} \end{bmatrix} \quad (10)$$

Computation of the effective number of chromosome segments (Me)

The prediction equations without availability of genotyping data require the effective number of chromosome segments that are independently segregating in a population, including selection candidates (S) and reference animals (R) (i.e., that are shared between the two

populations), $Me_{S,R}$. The value of $Me_{S,R}$ can be computed as proposed by Wientjes et al. [14]:

$$Me_{S,R} = \frac{1}{S_{G_{S,R}-A_{S,R}}^2},$$

where $G_{S,R}$ is the genomic relationship matrix between selection candidates and reference animals, $A_{S,R}$ is the pedigree relationship matrix, and $S_{G_{S,R}-A_{S,R}}^2$ is the empirical variance of the differences between corresponding elements of $G_{S,R}$ and $A_{S,R}$. In our study, $G_{S,R}$ and $A_{S,R}$ were scaled to the same base population by rescaling the inbreeding level in $G_{S,R}$ to the inbreeding in $A_{S,R}$ as follows [20]:

$$G_{S,R}^* = (1 - \bar{F})G_{S,R} + 2\bar{F}J,$$

where \bar{F} is the average pedigree inbreeding level computed from the pedigree, and J is a matrix filled with 1s.

The proposed computation of Me requires genotypes for both selection candidates and reference animals, which may be inconsistent with its use in the computation of reliabilities without availability of genotyping data. However, it is reasonable to assume that genotypes are already available for a limited number of animals, for example at least 100, that have the right family structure that is representative of the evaluated scenario, such that an accurate approximation of Me can be computed [14].

The effective number of chromosome segments originating from a specific breed (b) and that are shared between purebred selection candidates (S) of this breed and crossbred reference animals (Rc), $Me_{S,Rc}^{(b)}$ is required for the prediction equations for BSAM. In this study, $Me_{a,AB}^{(A)}$ is required in Eqs. (8) and (10) and was assumed to be equal to $Me_{a,A}$, which is required in Eqs. (2) and (6). The equality $Me_{a,AB}^{(A)} = Me_{a,A}$ was assumed since the selection candidates were the same for Eqs. (2), (4), (6), (8), and (10) the number of reference animals R and Rc was large, and the parents of breed A and crossbred AB reference animals were sampled from the same finite pool.

Simulated data

Data were simulated to validate Eqs. (2), (4), (6), (8), and (10), which predict the reliability of genomic EBV for crossbred performance using ASGM or BSAM, without availability of genotyping data. Two extreme scenarios were considered, in which either two closely-related or two unrelated breeds were used to produce crossbred

animals. The reliabilities predicted by Eqs. (2), (4), (6), (8), and (10) were validated against the reliabilities computed with the corresponding prediction equations with availability of genotyping data, that is, Eqs. (1), (3), (5), (7), and (9). The reliabilities predicted by equations with availability of genotyping data are equivalent to those computed from PEV associated with selection candidates of a genomic best linear unbiased prediction including both reference animals and selection candidates, based on phenotypes corrected with the best linear unbiased estimates of the fixed effects, and assuming the absence of selection [15].

Populations

First, historical and breed populations were simulated using the QMSim software [21]; second, a two-way crossbreeding program with five generations of random selection was simulated using a customized Fortran program. For the historical population, 1000 discrete random mating generations with a constant size of 10,000 individuals were simulated, which was followed by 1000 generations in which the population size was gradually decreased to 2000 individuals. In these 2000 historical generations, half of the simulated animals were males and the other half were females. Offspring were produced by the random union of gametes from the male and female gametic pools, and the number of offspring was equal to the number of animals required in the next generation. To simulate the two breed populations, A and B, two random samples were drawn from the last generation of the historical population (i.e., generation 2000), each including 500 males and 500 females. Subsequently, within each of the breeds, 10 or 100 generations of random mating were simulated before the two-way crossbreeding scheme was begun. These two scenarios (i.e., a common origin either 10 or 100 generations ago) will be referred to as related and unrelated breeds, respectively. For the 10 and 100 generations of random mating, a litter of four offspring (two males and two females) per female was simulated. From these offspring, 500 males were selected at random for the next generation. The number of females selected randomly for the next generation was gradually increased from 500 to 800 during the first four generations of the simulation of the breed populations, in order to enlarge the size of the population (Fig. 1).

In a second step, a two-way crossbreeding program with five generations of random selection was simulated. The animals of breeds A and B that were used to start the crossbreeding program were sampled from generation 2010 for the related breeds and from generation 2100 for

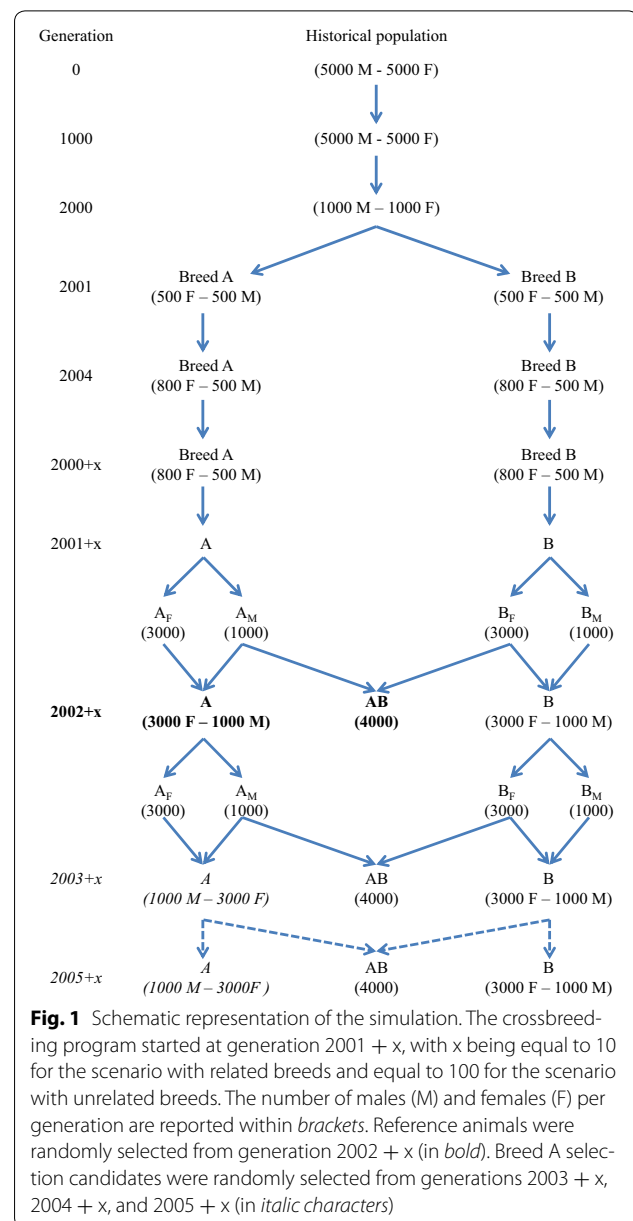


Fig. 1 Schematic representation of the simulation. The crossbreeding program started at generation 2001 + x, with x being equal to 10 for the scenario with related breeds and equal to 100 for the scenario with unrelated breeds. The number of males (M) and females (F) per generation are reported within brackets. Reference animals were randomly selected from generation 2002 + x (in bold). Breed A selection candidates were randomly selected from generations 2003 + x, 2004 + x, and 2005 + x (in italic characters)

the unrelated breeds. During the crossbreeding program, and for both breeds, animals of breeds A and B were randomly selected and mated to simulate the next generation of a constant size of 1000 males and 3000 females for each breed. From each of these five generations, animals of breeds A and B were randomly crossed to produce five generations of 4000 crossbred AB animals. Purebred animals used as parents of crossbred animals could also be parents of the next generation of purebred animals (Fig. 1).

Genotypes

The total length of the simulated genome was 10 Morgans (M) (10 chromosomes of 1 M and 4000 SNPs each). The positions of SNPs and of recombinations were randomized per chromosome and a recurrent mutation rate of 2.5×10^{-4} was assumed. All SNPs with a minor allele frequency (MAF) higher than or equal to 0.05 in the last historical generation (i.e., generation 2000) and were used to simulate the SNP genotypes of the purebred and crossbred animals. For subsequent analyses, 2000 SNPs were randomly selected from these SNPs for each chromosome. The breed origin of each allele for each crossbred animal was recorded. All scenarios (including the historical populations) were replicated 10 times.

Validation of prediction equations without availability of genotyping data

The validation required a set of known genotypes, as described previously, but no phenotype, since the reliabilities predicted without availability of genotyping data were validated against the reliabilities predicted with availability of genotyping data. However, estimates of heritabilities and genetic correlations between purebred and crossbred performance were required. Heritabilities of 0.20, 0.40, and 0.95 were used for both the purebred and crossbred performance traits. A high heritability, such as 0.95, and a single record per reference animal can be assumed when phenotypes of reference animals are derived from highly reliable EBV (e.g., deregressed EBV) [10]. Genetic correlations between purebred and crossbred performance traits were assumed to be equal to 0.30 or 0.70.

In the simulated data, two groups of reference animals and one group of selection candidates were defined for each scenario of related and unrelated breeds. For the scenarios with related and unrelated breeds, the two groups of reference animals were randomly selected from generations 2012 and 2102, respectively. For scenarios PB–PB and CB–PB, the two groups of reference animals included 2000 and 4000 animals that were randomly chosen from breed A and crossbred AB animals, respectively. For scenario CB + PB–PB, the first group included 4000 randomly chosen breed A animals and 2000 randomly chosen crossbred AB animals and the second group included 4000 breed A animals and 4000 crossbred AB animals. For the selection candidates for scenarios PB–PB, CB–PB and CB + PB–PB, 1000 breed A animals were randomly selected from each generation, starting from generation 2013 for the related breeds scenario and from generation 2103 for the unrelated breeds scenario, to create the groups of selection candidates. In the following, selection candidates from generations 2013 or 2103 are referred to as “G1” selection candidates. Similarly,

selection candidates from generations 2014 and 2104 and from generations 2015 and 2105 are referred to as “G2” and “G3” selection candidates, respectively.

For each ‘reference population-selection candidates’ combination and for each scenario, reliabilities of the genomic EBV for crossbred performance were computed using Eqs. (1), (3), (5), (7), and (9) for the scenarios in which all data was available, and using Eqs. (2), (4), (6), (8), and (10) for scenarios without availability of genotyping data. The required genomic relationship matrices and values of Me were computed using our in-house software `calc_grm` [22]. The predicted reliabilities were averaged across the 10 replicates.

Application of a prediction equation

The proposed equations can be used to investigate the reliability of genomic EBV for crossbred performance in crossbreeding schemes. As an illustration, Eq. (10), which predicts the reliability of genomic EBV using both purebred and crossbred animals as reference animals by BSAM, was used to predict the reliability of genomic EBV for a pig production system for which 10,000 breed A animals were previously genotyped and phenotyped. The aim was to investigate the effect of the addition of crossbred AB animals to the reference population on the reliability of genomic EBV for crossbred performance. A heritability of 0.20 was assumed for both purebred and crossbred performance traits and the genetic correlation between purebred and crossbred performance traits for breed A, ($r_{PC}^{(A)}$), ranged from 0.0 to 1.0. Both values of Me required by Eq. (10) (i.e. $Me_{a,AB}^{(A)}$ and $Me_{a,A}$) were assumed to be equal to 476.6, based on the equation $Me = 2N_eL / (\ln(4N_eL))$ [23], with N_e being the effective population size and L being the total length of the genome in M. For N_e and L , we assumed values of 80 and 27 respectively, based on the study of Landrace pigs by Uimari and Tapio [24] and the study by Lin et al. [25]. The use of equal values of Me for the purebred and crossbred populations was based on the assumption that breed A parents of purebred and crossbred animals were sampled from the same pool.

Results

This section first presents the results of the validation of the equations for predicting reliability without availability of genotyping data. As defined previously, the reliabilities without availability of genotyping data were validated against the reliabilities computed with availability of genotyping data. The second part of this section describes the increase in reliabilities from the addition of crossbred animals to a purebred reference population in a pig breeding program.

PB–PB scenario

For the PB–PB scenario, the results show that reliabilities predicted without availability of genotyping data were of the same order of magnitude as reliabilities computed with availability of genotyping data (Figs. 2, 3). For the scenario with related breeds and $r_{PC} = 0.3$ (Fig. 2), the predicted reliabilities with availability of genotyping data were around 0.01 for $h_a^2 = 0.2$, in the range [0.02; 0.03] for $h_a^2 = 0.4$, and in the range [0.04; 0.05] for $h_a^2 = 0.95$, across all three groups of G1, G2 or G3 selection candidates and with 2000 reference animals from breed A. When $r_{PC} = 0.7$ (Fig. 3), the corresponding

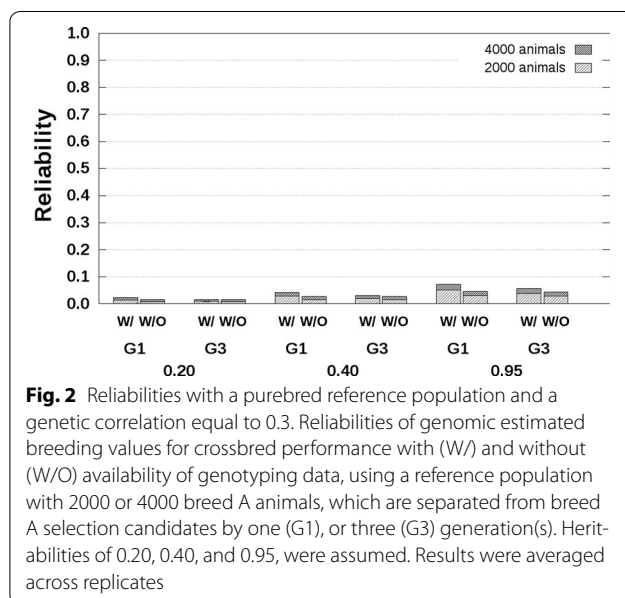


Fig. 2 Reliabilities with a purebred reference population and a genetic correlation equal to 0.3. Reliabilities of genomic estimated breeding values for crossbred performance with (W/) and without (W/O) availability of genotyping data, using a reference population with 2000 or 4000 breed A animals, which are separated from breed A selection candidates by one (G1), or three (G3) generation(s). Heritabilities of 0.20, 0.40, and 0.95, were assumed. Results were averaged across replicates

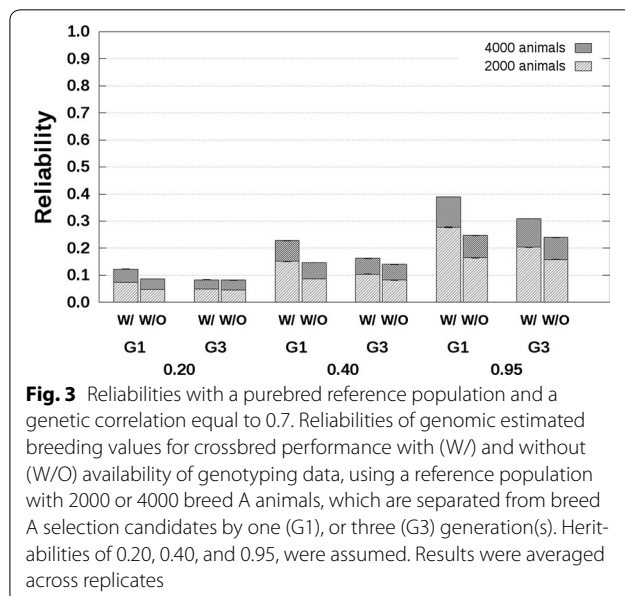


Fig. 3 Reliabilities with a purebred reference population and a genetic correlation equal to 0.7. Reliabilities of genomic estimated breeding values for crossbred performance with (W/) and without (W/O) availability of genotyping data, using a reference population with 2000 or 4000 breed A animals, which are separated from breed A selection candidates by one (G1), or three (G3) generation(s). Heritabilities of 0.20, 0.40, and 0.95, were assumed. Results were averaged across replicates

predicted reliabilities with availability of genotyping data were in the range [0.05; 0.07] for $h_a^2 = 0.2$, in the range [0.10; 0.15] for $h_a^2 = 0.4$, and in the range [0.20; 0.28] for $h_a^2 = 0.95$. For both scenarios with $r_{PC} = 0.3$ and $r_{PC} = 0.7$, the addition of 2000 breed A reference animals slightly increased the predicted reliabilities (Figs. 2, 3).

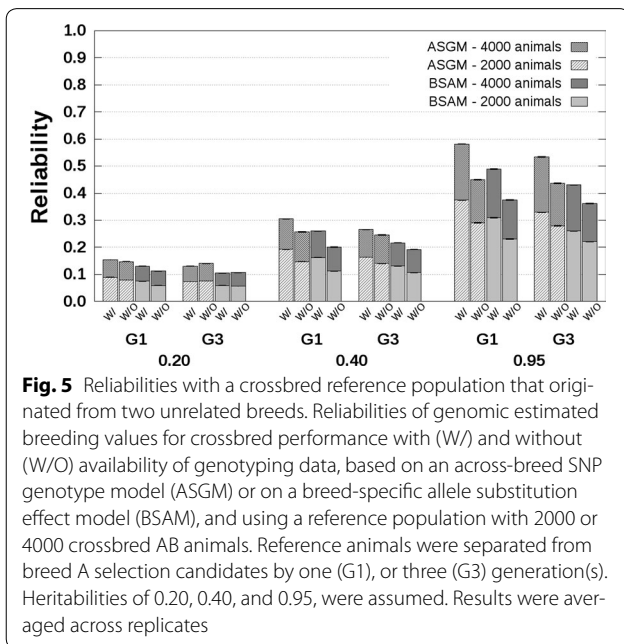
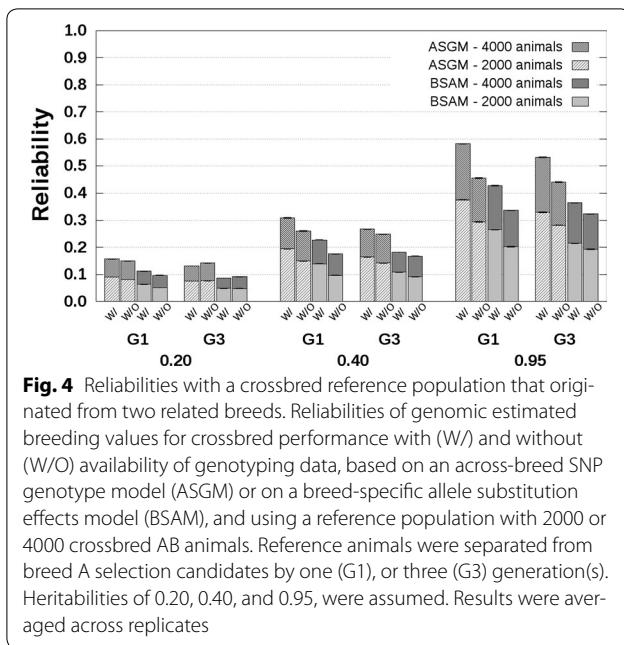
Reliabilities predicted without availability of genotyping data were always lower than those predicted with availability of genotyping data, which agrees with theory (see “PB–PB scenario” section in the “Methods” section). For the scenario with related breeds and $r_{PC} = 0.3$ (Fig. 2), the differences between reliabilities predicted without and with availability of genotyping data were around 0.00 for $h_a^2 = 0.2$, in the range [−0.02; 0.00] for $h_a^2 = 0.4$, and in the range [−0.02; −0.01] for $h_a^2 = 0.95$ across all three groups of G1, G2 or G3 selection candidates and with 2000 breed A reference animals. When $r_{PC} = 0.7$ (Fig. 3), the corresponding differences between reliabilities predicted without and with availability of genotyping data were in the range [−0.03; 0.00] for $h_a^2 = 0.2$, in the range [−0.06; −0.02] for $h_a^2 = 0.4$, and in the range [−0.11; −0.04] for $h_a^2 = 0.95$. The largest differences between reliabilities predicted without and with availability of genotyping data were always observed for the G1 selection candidates.

Similar results were obtained for the scenario with unrelated breeds (see Additional file 4: Tables S1, S2). Such similar results were expected since the distance between breeds is not taken into account by ASGM. The SD of the reliabilities across replicates were in the range [0.000; 0.001] (see Additional file 4: Tables S1, S2).

CB–PB scenario

Reliabilities with and without availability of genotyping data are presented in Fig. 4 for related breeds and in Fig. 5 for unrelated breeds. The CB–PB scenario included both ASGM and BSAM. For both models, the reliabilities predicted without availability of genotyping data underestimated the reliabilities predicted with availability of genotyping data. Underestimation was close to 0 when $h_c^2 = 0.20$, and increased up to 0.1 with increasing h_c^2 and number of crossbred AB reference animals. Similar to the PB–PB scenario, the underestimation of reliabilities predicted with availability of genotyping data by the reliabilities predicted without availability of genotyping data was largest for the G1 selection candidates.

For the G1 selection candidates, the reliabilities for ASGM with availability of genotyping data were around 0.09 with 2000 crossbred reference animals, independent of the relationship between the breeds, and around 0.16 with 4000 crossbred reference animals, using $h_c^2 = 0.20$ (Figs. 4, 5). Differences between the reliabilities predicted without and with availability of genotyping data were



around -0.01 for both 2000 and 4000 crossbred reference animals. The corresponding reliabilities using $h_c^2 = 0.95$ were around 0.37 and 0.58 with 2000 and 4000 crossbred reference animals, respectively. The corresponding differences between reliabilities predicted without and with availability of genotyping data were in the range $[-0.13; -0.08]$.

For G1 selection candidates with related breeds, the reliabilities for BSAM with availability of genotyping data

were around 0.06 and 0.11 with 2000 and 4000 crossbred reference animals, respectively, when using $h_c^2 = 0.20$ (Fig. 4). Differences between reliabilities predicted without and with availability of genotyping data were around -0.01 with both 2000 and 4000 crossbred reference animals. The corresponding reliabilities using $h_c^2 = 0.95$ were around 0.27 and 0.43 with 2000 and 4000 crossbred reference animals, respectively. Corresponding differences between reliabilities predicted without and with availability of genotyping data were in the range $[-0.09; -0.06]$. Similar differences were observed with unrelated breeds (Fig. 5). The SD of reliabilities across replicates were in the range $[0.000; 0.002]$ (see Additional file 4: Tables S3, S4).

A comparison of reliabilities with availability of genotyping data between ASGM and BSAM showed that ASGM consistently performed better than BSAM. However, reliabilities for BSAM increased with increasing distance between breeds, while reliabilities for ASGM were only slightly affected (Figs. 4, 5). The increase in reliabilities with increasing distance between breeds, which compensates for the larger number of effects fitted in BSAM compared to ASGM, is in agreement with previous studies, e.g., Ibanez-Escriche et al. [4].

CB + PB–PB scenario

The CB + PB–PB scenario included both breed A and crossbred AB animals in the reference population. The number of breed A reference animals was always 4000. The number of crossbred AB animals was equal to 2000 or 4000. The CB + PB–PB scenario also included both ASGM and BSAM.

For related breeds, reliabilities without and with availability of genotyping data are presented in Fig. 6 for $r_{PC} = 0.3$ and in Fig. 7 for $r_{PC} = 0.7$. Reliabilities predicted without and with availability of genotyping data were of the same order of magnitude, for both ASGM and BSAM. Differences between the two predicted reliabilities were in the range $[-0.09; 0.05]$. Similar to previous results, these differences increased with heritability. Reliabilities for BSAM with availability of genotyping data were about 0.03 to 0.04 lower than the corresponding reliabilities for ASGM when heritabilities were assumed to be 0.20. This difference between reliabilities for BSAM and for ASGM increased with increasing heritability and r_{PC} , and with decreasing distance between breeds. For example, reliabilities for BSAM with availability of genotyping data were between 0.07 and 0.12 points lower than the corresponding reliabilities for ASGM when heritabilities were equal to 0.95. These lower reliabilities for BSAM can be attributed to the additional breed-specific effects fitted in the model for a given number of records. Similar

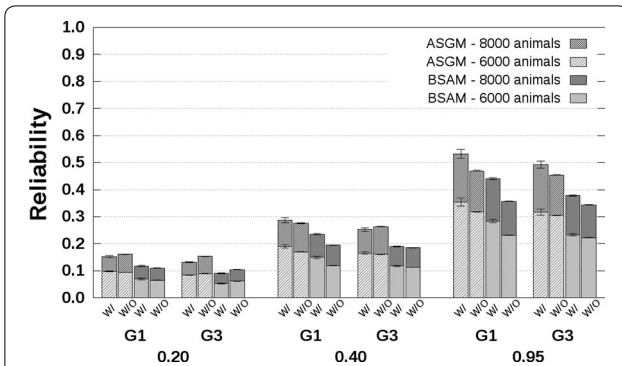


Fig. 6 Reliabilities with a mixed reference population assuming two related breeds and a genetic correlation of 0.3. Reliabilities of genomic estimated breeding values for crossbred performance with (W) and without (W/O) availability of genotyping data, based on an across-breed SNP genotype model (ASGM) or on a breed-specific allele substitution effects model (BSAM). The reference population included 4000 breed A animals and 2000 or 4000 crossbred AB animals. Reference animals were separated from breed A selection candidates by one (G1), or three (G3) generation(s). Heritabilities of 0.20, 0.40, and 0.95, were assumed. Results were averaged across replicates

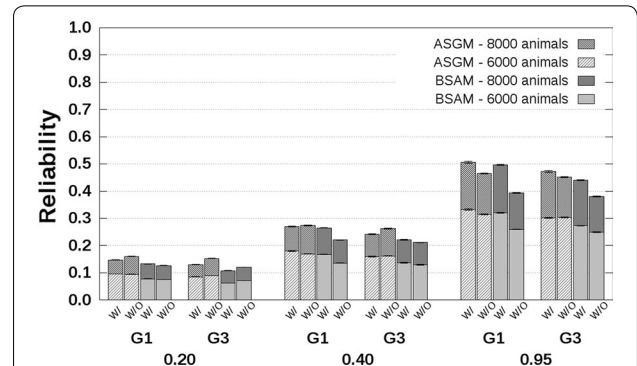


Fig. 8 Reliabilities with a mixed reference population assuming two unrelated breeds and a genetic correlation of 0.3. Reliabilities of genomic estimated breeding values for crossbred performance with (W) and without (W/O) availability of genotyping data, based on an across-breed SNP genotype model (ASGM) or on a breed-specific allele substitution effects model (BSAM). The reference population included 4000 breed A animals and 2000 or 4000 crossbred AB animals. Reference animals were separated from breed A selection candidates by one (G1), or three (G3) generation(s). Heritabilities of 0.20, 0.40, and 0.95, were assumed. Results were averaged across replicates

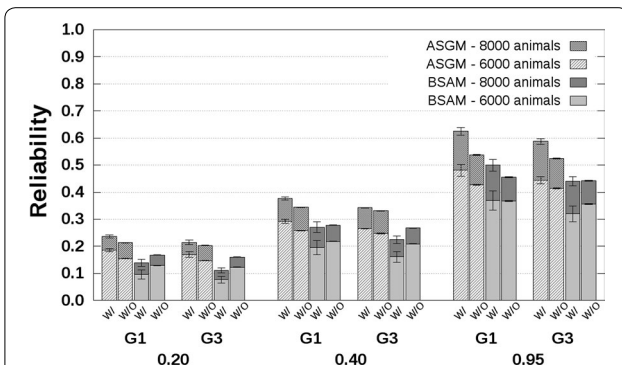


Fig. 7 Reliabilities with a mixed reference population assuming two related breeds and a genetic correlation of 0.7. Reliabilities of genomic estimated breeding values for crossbred performance with (W) and without (W/O) availability of genotyping data, based on an across-breed SNP genotype model (ASGM) or on a breed-specific allele substitution effects model (BSAM). The reference population included 4000 breed A animals and 2000 or 4000 crossbred AB animals. Reference animals were separated from breed A selection candidates by one (G1), or three (G3) generation(s). Heritabilities of 0.20, 0.40, and 0.95, were assumed. Results were averaged across replicates

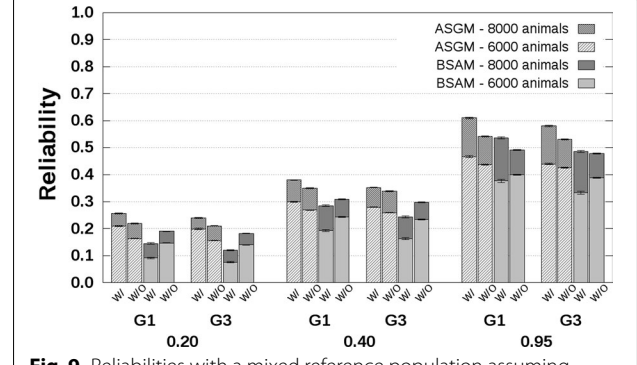


Fig. 9 Reliabilities with a mixed reference population assuming two unrelated breeds and a genetic correlation of 0.7. Reliabilities of genomic estimated breeding values for crossbred performance with (W) and without (W/O) availability of genotyping data, based on an across-breed SNP genotype model (ASGM) or on a breed-specific allele substitution effects model (BSAM). The reference population included 4000 breed A animals and 2000 or 4000 crossbred AB animals. Reference animals were separated from breed A selection candidates by one (G1), or three (G3) generation(s). Heritabilities of 0.20, 0.40, and 0.95, were assumed. Results were averaged across replicates

trends were observed for reliabilities without availability of genotyping data.

With unrelated breeds, the reliabilities without and with availability of genotyping data, averaged across replicates, are presented in Fig. 8 for $r_{PC} = 0.3$ and in Fig. 9 for $r_{PC} = 0.7$. Differences between reliabilities predicted without and with availability of genotyping data were in

the range $[-0.04; 0.07]$ for all scenarios with heritabilities equal to 0.20 and to 0.40, and in the range $[-0.10; 0.05]$ for all scenarios with heritabilities equal to 0.95. Reliabilities predicted with availability of genotyping data for BSAM with unrelated breeds were higher by about 0.01 to 0.06 than the reliabilities predicted with availability of genotyping data for BSAM with related breeds. Similar trends were observed for reliabilities predicted

without availability of genotyping data, showing a reliable prediction of reliability computed with availability of genotyping data. The SD of reliabilities across replicates were in the range [0.000; 0.006] (see Additional file 4: Tables S5, S6, S7, S8).

Reliabilities in a pig-breeding program

Predicted reliabilities for BSAM when up to 10,000 crossbred AB animals were added to a reference population of 10,000 breed A animals are in Fig. 10, showing that predicted reliabilities increased when crossbred AB animals were added to the reference population. The increase in reliabilities decreased with increasing $r_{PC}^{(A)}$. The reliabilities obtained for ASGM with $r_{PC}^{(A)} = 0.92$ based on 10,000 breed A reference animals (and no crossbred AB reference animals) (0.68) was the same as that for BSAM based on only 10,000 crossbred AB reference animals (i.e., with $r_{PC}^{(A)} = 0.0$). Therefore, for $r_{PC}^{(A)} < 0.92$, BSAM with only crossbred reference animals can be at least as accurate as ASGM with a larger number of purebred reference animals.

Discussion

In this study, the term “reliability” refers to the precision of genomic EBV obtained by relating their PEV to the additive genetic variance of the base population, i.e., assuming absence of selection. Equations for predicting the reliability of genomic EBV for crossbred performance are proposed for reference populations that include purebred animals, crossbred animals, or both. Reliabilities were predicted for two models: ASGM and BSAM. For

the BSAM, we used the true breed-of-origin of all alleles for the crossbred animals, which would have to be estimated in practice, which may negatively impact the reliability obtained. However, we expect this to have only a very minor effect, since we showed in previous studies that it is possible to accurately derive breed-of-origin of alleles in three-breed crossbred pigs [26, 27].

Reliabilities of genomic EBV can be predicted when genotype data are already available, i.e., with availability of genotyping data, or without availability of genotyping data. For scenarios without availability of genotyping data, it is assumed that the required genetic parameters are computed using pedigree instead of genomic data, or that estimates are available from the literature. The results of this study showed that the reliabilities of genomic EBV for crossbred performance predicted without availability of genotyping data were of the same order of magnitude as those predicted with availability of genotyping data. Therefore, while prediction of reliability should preferably take the genotype data of selection candidates into account when available, both methods can predict the reliability of genomic EBV for crossbred performance for different reference populations, heritabilities, and r_{PC} . The derived equations can therefore be useful to optimize the design of breeding programs.

Reliabilities predicted without and with availability of genotyping data

The aim of this study was to predict the precision of genomic EBV based on PEV in the absence of selection. Thus, the derivation of our prediction equations without and with availability of genotyping data was based on the SI and mixed model theories and assumed that phenotypes were corrected for all fixed and random effects other than the considered genetic additive effects. The equivalence between SI and mixed model theories under certain conditions, such as the use of the same estimates for the fixed effects, has previously been shown by several studies (e.g., [15, 17, 28, 29]). Therefore, reliabilities predicted with availability of genotyping data would be expected to be close to reliabilities computed from PEV obtained from genomic best linear unbiased prediction, in the absence of selection. Equations for predicting the reliability of genomic EBV without availability of genotyping data were validated against the equations for predicting reliability with availability of genotyping data, and not against the reliability of selection, i.e., the squared correlation between estimated and true genomic breeding values, which is often obtained by cross-validation. Indeed, the reliability of genomic EBV is not equivalent to the reliability of selection for populations that are under selection, although they are equivalent for populations without selection [30–33]. Reliability of selection

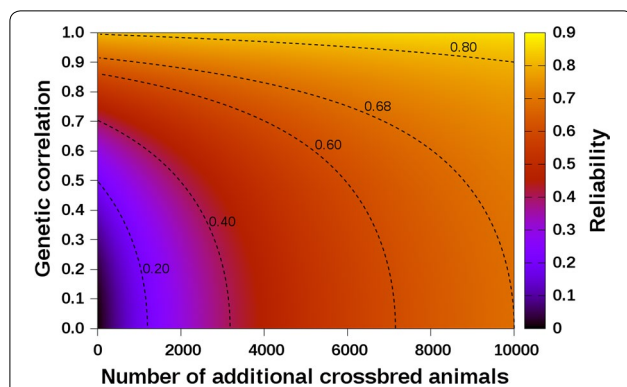


Fig. 10 Reliabilities with additional crossbred reference animals and different genetic correlations. Reliabilities predicted without availability of genotyping data for genomic estimated breeding values of crossbred performance using a breed-specific allele substitution effects model. The reference population included 10,000 purebred animals and a number of crossbred animals that varied from 0 to 10,000. A heritability of 0.20 was assumed for both purebred and crossbred performance traits, and the genetic correlation between purebred and crossbred performance traits varied from 0.0 to 1.0. All the required values of M_e were assumed to be equal to 476.6

can be predicted from the reliability of genomic EBV by considering the intensity of selection using, e.g., the equations proposed by Dekkers [30] and Bijma [31].

Our prediction equations with availability of genotyping data can be extended to situations when phenotypes are already available for selection candidates and even to situations when some reference animals are not genotyped. First, in our study, selection candidates were defined as animals with genotypes but without phenotypes. In the context of poultry and pig breeding, this reflects for instance carcass, disease, or fertility traits. However, for growth-related traits, phenotypes are typically available for selection candidates at the time of selection. For these situations, prediction equations with availability of genotyping data can be used simply by including the selection candidates with genotypes and phenotypes as both reference animals and selection candidates, such that they are used both for computing the genomic relationship matrix between reference animals and selection candidates and the genomic relationship matrix between reference animals (e.g., for computing both $G_{ai,A}$ and $G_{A,A}$ for the PB–PB scenario using ASGM). Second, situations in which some reference animals are not genotyped can also be modelled by our prediction equations with availability of genotyping data through the use of a combined pedigree-genomic relationship matrix H [34–36] instead of the genomic relationship matrix G used in this study. Third, our prediction equations with availability of genotyping data can be extended to situations in which reference animals have repeated phenotypes, or pseudo-phenotypes, such as deregressed proofs and associated weights. For instance, the predicted reliability for the PB–PB scenario using ASGM can be computed as:

$$r_{P_ASGM_with}^2 = \frac{1}{N_a} \sum_i r_{PC}^2 \frac{\mathbf{G}_{ai,A} \mathbf{W}' (\mathbf{W} \mathbf{G}_{A,A} \mathbf{W}' + \mathbf{R} \frac{1-h_a^2}{h_a^2})^{-1} \mathbf{W} \mathbf{G}_{A,ai}}{\mathbf{G}_{ai,ai}}$$

where \mathbf{W} is the incidence matrix relating (pseudo-)phenotypes to reference animals and \mathbf{R} is a diagonal matrix with elements equal to 1 for real phenotypes, or equal to the inverse of weights associated with pseudo-phenotypes [17, 28]. Unlike the prediction equations with availability of genotyping data, the extension of prediction equations without availability of genotyping data to more complex scenarios is not as straightforward.

We also assumed that all additive genetic variance was captured by the SNPs in the derivation of the prediction equations. When only a portion of the additive genetic variance is captured by the SNPs, the prediction equations need to take this into account, as proposed by Goddard et al. [13] and Wientjes et al. [14]. This proportion could be empirically estimated when the reference

population includes only one population by comparing predicted and realized (cross-validation) reliabilities [14].

For most scenarios, predicted reliabilities without availability of genotyping data underestimated the reliabilities predicted with availability of genotyping data (Figs. 2, 3, 4, 5, 6, 7, 8, 9). While this is in agreement with the theory, only a part of the underestimation is due to the fact that the decrease of the error variance when multiple loci are used was ignored (see “Methods” section; [12, 13]). This underestimation is greater when heritability and reliability increase to a value of 1 [13], as observed in our results. Most of the underestimation is, however, primarily due to an overestimation of Me , especially for the PB–PB and CB–PB scenarios with only one generation separating reference animals and selection candidates, for which the largest underestimations were observed. For instance, the fractional underestimation of $r_{P_ASGM_with}^2$ that can be attributed to not considering the reduction in error variance $(1 - h_a^2 r_{P_ASGM_without}^2)$ (see Additional file 1) is approximately equal to 0.03 (i.e., 3% error) for the PB–PB scenario with one generation separating 4000 reference animals and selection candidates, $h_a^2 = 0.95$, $r_{PC} = 1.0$, and $Me_{a,A} = 3730$ (obtained from one random replicate). This does, however, explain only part of the fractional underestimation of about 0.57 that is observed in Fig. 3. Thus, the underestimation appears to be mainly due to the overestimation of Me , particularly when only one generation separates the reference animals and selection candidates. Indeed, while estimates of Me increased with decreasing predicted reliabilities with availability of genotyping data, the results show that the reliabilities predicted without availability of genotyping data decreased at a lower rate than reliabilities predicted with availability of genotyping data when the relationships between the reference population and the selection candidates decreased. Further work to improve estimation of Me is needed, especially for scenarios in which reference animals and selection candidates are highly related.

While predicted reliabilities without availability of genotyping data were underestimated for most scenarios, overestimations were observed for some scenarios with reference populations that included both purebred and crossbred animals (Figs. 6, 7, 8, 9). These overestimations may be the result of estimation of Me and assumptions taken for the derivation of the equations without availability of genotyping data (e.g., a diagonal residual (co)variance matrix for corrected phenotypes associated with BSAM and using purebred and crossbred reference animals).

Potential use of the prediction equations

The equations derived in this study can be used to compare the effects of modifying the values of various factors (e.g., r_{PC} , numbers of reference animals, or relationships

between the reference population and the selection candidates) on the reliability of genomic EBV for crossbred performance and for the optimization of the design of breeding programs. However, the effects of some factors should be compared carefully. For example, the results show that the prediction equations without availability of genotyping data should be used with care for the comparison of the effects of different relationships between the reference population and the selection candidates. The prediction equations without availability of genotyping data should also be used with care for the comparison of the reliabilities of the ASGM and BSAM models, especially when the reference population includes both purebred and crossbred animals (e.g., for the PB + CB–PB scenario with unrelated breeds and $r_{PC} = 0.7$; Fig. 9). Nevertheless, the prediction equations without availability of genotyping data can still provide some insight into the reliability of both models in different scenarios. For instance, the results (Figs. 4, 5, 6, 7, 8, 9) showed that reliabilities for BSAM tended to increase with increasing distance between breeds, while the reliabilities for ASGM were only slightly affected. The increase in reliabilities with increasing distance between breeds, which compensates for fitting more effects in BSAM in comparison to ASGM, is in agreement with previous studies, e.g., Ibanez-Escriche et al. [4]. For instance, assume that a reference population of a fixed number of crossbred AB animals is available, and that heritabilities of crossbred performance traits estimated for ASGM and BSAM for the breed A are equal. Therefore, from Eq. (4), $r_{C_ASGM_without}^2 = \frac{N_{AB}h_c^2}{N_{AB}h_c^2 + Me_{a,AB}}$, and Eq. (8), $r_{C_BSAM_without}^2 = \frac{N_{AB}h_c^2}{N_{AB}h_c^2 + 2Me_{a,AB}^{(A)}}$, it follows that the reliability of genomic EBV based on BSAM would be higher than the reliability based on ASGM if $Me_{a,AB} > 2Me_{a,AB}^{(A)}$. This will be the case if the LD patterns between breeds A and B are sufficiently different, which is more likely in the case when the breeds have diverged for many generations [37]. This is in agreement with our results (Figs. 4, 5) and previous studies based on simulated data (e.g., [4, 11]) which show that reliabilities for BSAM increase with increasing distance between breeds. The additional effects fitted in BSAM are taken into account in Eq. (8) by the factor of 2, which was also considered by van Grevenhof and van der Werf [38], who evaluated the benefit of including crossbred animals in the reference population of a crossbreeding program using genomic selection.

Computation of *Me*

The evaluation of different scenarios based on the prediction equations without availability of genotyping data requires accurate estimates of all parameters, and

especially of *Me* (e.g., [10, 14, 39, 40]). Parameters such as heritabilities and correlations, if estimated inaccurately, would similarly bias reliabilities predicted without and with availability of genotyping data, since these parameters are used in both equations. However, *Me*, the effective number of segments that are shared and segregating in both selection candidates and reference animals, is only used when predicting reliability without availability of genotyping data, and has a large impact. In our study, the estimates of *Me* were computed from the differences between genomic and pedigree relationships between reference animals and selection candidates, as proposed by Wientjes et al. [14]. However, our results showed that these estimates of *Me* did not adequately consider the close relationships that can exist between reference animals and selection candidates. As already proposed by Daetwyler et al. [39] and Brard and Ricard [40], another approach would be to reverse the prediction equations without availability of genotyping data for computing *Me*. Required reliabilities and other parameters should be obtained from a reference population and different generations of selection candidates in which genomic prediction is already applied. However, estimates of *Me* obtained by the reversion of prediction equations would be underestimated, since this would include a correction for the fact that the error variance decreases when multiple loci are used, which is trait-dependent.

This study has introduced the concept of the effective number of chromosome segments originating from a specific breed (*b*), and shared by selection candidates (*S*) from this breed and crossbred reference animals (*Rc*), $Me_{S,Rc}^{(b)}$. This $Me_{S,Rc}^{(b)}$ is different from $Me_{S,Rc}$ as defined previously, since the latter does not take the breed origin of the chromosome segments of the crossbred animals into consideration. Indeed, each purebred population has its own value of *Me*, while the genome of crossbred animals combines segments from the different populations they originated from. Thus, the value of $Me_{S,Rc}$ includes both the effective number of chromosome segments segregating in breed *b*, and the effective number of chromosome segments segregating in the other breed(s) of origin for the crossbred animals, while $Me_{S,Rc}^{(b)}$ only involves the effective number of chromosome segments segregating in breed *b*. For this study, it was assumed that $Me_{S,Rc}^{(b)}$ (i.e., $Me_{a,AB}^{(A)}$) was equal to $Me_{S,R}$ (i.e., $Me_{a,A}$) for which the breed *b* selection candidates and reference animals (*R*) share the same parents as the crossbred *Rc* animals. This assumption was valid based on the results obtained. In practice, such an assumption would not be possible, since the purebred and crossbred reference animals may not share the same parents, or reference animals may belong to different generations. Further research on accurate estimation of *Me* is therefore required.

Conclusions

Several equations for predicting the reliability of genomic EBV for crossbred performance based on ASGM or on BSAM were derived for three different scenarios. These three scenarios involved a reference population that included only purebred animals, only crossbred animals, or both. The prediction equations were derived for application either without or with availability of genotyping data. Results showed that the reliabilities predicted without availability of genotyping data were of the same order of magnitude as the predictions of reliabilities predicted with availability of genotyping data. Thus, the proposed equations applied either without or with availability of genotyping data can be used to evaluate the effects of several parameters on the reliability of genomic EBV for crossbred performance (e.g., the genetic correlation between purebred and crossbred performances, heritabilities of the traits, number of reference animals, distance between breeds), and for the optimization of the design of breeding programs. Moreover, we showed that model BSAM can outperform model ASGM for a breed, if the effective number of chromosome segments originating from this breed and shared by selection candidates of this breed and crossbred reference animals is less than half the effective number of all chromosome segments that are independently segregating in these same animals, provided all other parameters remain equal. It is necessary to improve estimation of the effective number of chromosome segments to predict the reliability of genomic EBV without availability of genotyping data more accurately.

Additional files

Additional file 1. Derivation of the equation for predicting the reliability of genomic estimated breeding values without availability of data. Based on the mixed model theory, a derivation of the equation for predicting the reliability of genomic estimated breeding values is detailed in this document, assuming that effects of all independent loci are estimated simultaneously, and assuming a single population.

Additional file 2. Derivation of the prediction equation without availability of data, using a breed-specific allele substitution effects model and only crossbred reference animals. A derivation of the equation that predicts the reliability of genomic estimated breeding values for crossbred performance using a breed-specific allele substitution effects model and only crossbred animals without availability of genotyping data is detailed in this document.

Additional file 3. Derivation of the prediction equation without availability of data, using a breed-specific allele substitution effects model, and purebred and crossbred reference animals. A derivation of the equation for predicting the reliability of genomic estimated breeding values for crossbred performance using a breed-specific allele substitution effects model and a reference population including both purebred and crossbred animals, without availability of genotyping data, is detailed in this document.

Additional file 4: Table S1. Reliabilities for crossbred performance with a purebred reference population and a genetic correlation equal to 0.3.

Table S2. Reliabilities for crossbred performance with a purebred reference population and a genetic correlation equal to 0.7. **Table S3.** Reliabilities for crossbred performance with a crossbred reference population originated from two related breeds. **Table S4.** Reliabilities for crossbred performance with a crossbred reference population originated from two unrelated breeds. **Table S5.** Reliabilities for crossbred performance with a mixed reference population assuming two related breeds and a genetic correlation of 0.3. **Table S6.** Reliabilities for crossbred performance with a mixed reference population assuming two related breeds and a genetic correlation of 0.7. **Table S7.** Reliabilities for crossbred performance with a mixed reference population assuming two unrelated breeds and a genetic correlation of 0.3. **Table S8.** Reliabilities for crossbred performance with a mixed reference population assuming two unrelated breeds and a genetic correlation of 0.7.

Authors' contributions

JV derived the equations, performed the analyses, and drafted the manuscript. JJW wrote the simulation program. All authors discussed the design of the simulations. All authors provided valuable insights throughout the analysis and writing process. All authors read and approved the final manuscript.

Acknowledgements

Financial support from the Dutch Ministry of Economic Affairs, Agriculture, and Innovation (Public-private partnership "Breed4Food" Code BO-22.04-011-001-ASG-LR-3) is acknowledged. Discussions with Yvonne Wientjes and Piter Bijma, and useful comments of the two anonymous reviewers are acknowledged.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Received: 5 July 2016 Accepted: 27 April 2017

Published online: 12 May 2017

References

- Wei M, van der Werf JHJ. Maximizing genetic response in cross-breeds using both purebred and crossbred information. *Anim Sci*. 1994;59:401–13.
- Toosi A, Fernando RL, Dekkers JCM. Genomic selection in admixed and crossbred populations. *J Anim Sci*. 2010;88:32–46.
- Christensen OF, Madsen P, Nielsen B, Su G. Genomic evaluation of both purebred and crossbred performances. *Genet Sel Evol*. 2014;46:23.
- Ibáñez-Escriche N, Fernando RL, Toosi A, Dekkers JC. Genomic selection of purebreds for crossbred performance. *Genet Sel Evol*. 2009;41:12.
- Dekkers JCM. Marker-assisted selection for commercial crossbred performance. *J Anim Sci*. 2007;85:2104–14.
- Zeng J, Toosi A, Fernando RL, Dekkers JC, Garrick DJ. Genomic selection of purebred animals for crossbred performance in the presence of dominant gene action. *Genet Sel Evol*. 2013;45:11.
- Lourenco DAL, Tsuruta S, Fragomeni BO, Chen CY, Herring WO, Misztal I. Crossbred evaluations in single-step genomic best linear unbiased predictor using adjusted realized relationship matrices. *J Anim Sci*. 2016;94:909–19.
- Hidalgo AM, Bastiaansen JWM, Lopes MS, Harlizius B, Groenen MAM, de Koning D-J. Accuracy of predicted genomic breeding values in purebred and crossbred pigs. *G3 (Bethesda)*. 2015;5:1575–83.
- Karoui S, Carabaño MJ, Díaz C, Legarra A. Joint genomic evaluation of French dairy cattle breeds using multiple-trait models. *Genet Sel Evol*. 2012;44:39.

10. Wientjes YC, Veerkamp RF, Bijma P, Bovenhuis H, Schrooten C, Calus MP. Empirical and deterministic accuracies of across-population genomic prediction. *Genet Sel Evol*. 2015;47:5.
11. Esfandyari H, Sørensen AC, Bijma P. A crossbred reference population can improve the response to genomic selection for crossbred performance. *Genet Sel Evol*. 2015;47:76.
12. Daetwyler HD, Villanueva B, Woolliams JA. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One*. 2008;3:e3395.
13. Goddard ME, Hayes BJ, Meuwissen THE. Using the genomic relationship matrix to predict the accuracy of genomic selection. *J Anim Breed Genet*. 2011;128:409–21.
14. Wientjes YCJ, Bijma P, Veerkamp RF, Calus MPL. An equation to predict the accuracy of genomic values by combining data from multiple traits, populations, or environments. *Genetics*. 2016;202:799–823.
15. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci*. 2008;91:4414–23.
16. Vitezica ZG, Varona L, Elsen JM, Misztal I, Herring W, Legarra A. Genomic BLUP including additive and dominant variation in purebreds and F1 crossbreds, with an application in pigs. *Genet Sel Evol*. 2016;48:6.
17. Henderson CR. Applications of linear models in animal breeding. 2nd ed. Guelph: University of Guelph; 1984.
18. de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*. 2012;193:327–45.
19. Henderson CR. Best linear unbiased estimation and prediction under a selection model. *Biometrics*. 1975;31:423–47.
20. Powell JE, Visscher PM, Goddard ME. Reconciling the analysis of IBD and IBS in complex trait studies. *Nat Rev Genet*. 2010;11:800–5.
21. Sargolzaei M, Schenkel FS. QMSim: a large-scale genome simulator for livestock. *Bioinformatics*. 2009;25:680–1.
22. Calus MPL, Vandenplas J. Calc_grm—a program to compute pedigree, genomic, and combined relationship matrices. Wageningen: ABGC, Wageningen UR Livestock Research; 2016.
23. Goddard ME. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*. 2009;136:245–57.
24. Uimari P, Tapio M. Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds. *J Anim Sci*. 2011;89:609–14.
25. Lin Z, Hayes BJ, Daetwyler HD. Genomic selection in crops, trees and forages: a review. *Crop Pasture Sci*. 2014;65:1177–91.
26. Sevillano CA, Vandenplas J, Bastiaansen JWM, Calus MPL. Empirical determination of breed-of-origin of alleles in three-breed cross pigs. *Genet Sel Evol*. 2016;48:55.
27. Vandenplas J, Calus MPL, Sevillano CA, Windig JJ, Bastiaansen JWM. Assigning breed origin to alleles in crossbred animals. *Genet Sel Evol*. 2016;48:61.
28. Mrode RA. Linear models for the prediction of animal breeding values. 2nd ed. Wallingford: CABI Publishing; 2005.
29. Strandén I, Garrick DJ. Technical note: derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *J Dairy Sci*. 2009;92:2971–5.
30. Dekkers JCM. Asymptotic response to selection on best linear unbiased predictors of breeding values. *Anim Sci*. 1992;54:351–60.
31. Bijma P. Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations. *J Anim Breed Genet*. 2012;129:345–58.
32. Van Grevenhof EM, Van Arendonk JA, Bijma P. Response to genomic selection: the Bulmer effect and the potential of genomic selection when the number of phenotypic records is limiting. *Genet Sel Evol*. 2012;44:26.
33. Gorjanc G, Bijma P, Hickey JM. Reliability of pedigree-based and genomic evaluations in selected populations. *Genet Sel Evol*. 2015;47:65.
34. 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 evaluation of Holstein final score. *J Dairy Sci*. 2010;93:743–52.
35. Christensen OF, Lund MS. Genomic prediction when some animals are not genotyped. *Genet Sel Evol*. 2010;42:2.
36. Legarra A, Christensen OF, Aguilar I, Misztal I. Single step, a general approach for genomic selection. *Livest Sci*. 2014;166:54–65.
37. de Roos APW, Hayes BJ, Spelman RJ, Goddard ME. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics*. 2008;179:1503–12.
38. van Grevenhof IE, van der Werf JH. Design of reference populations for genomic selection in crossbreeding programs. *Genet Sel Evol*. 2015;47:14.
39. Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA. The impact of genetic architecture on genome-wide evaluation methods. *Genetics*. 2010;185:1021–31.
40. Brard S, Ricard A. Is the use of formulae a reliable way to predict the accuracy of genomic selection? *J Anim Breed Genet*. 2015;132:207–17.

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

