

### 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.

The 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 (each associated with one record), without availability of genotyping data, is hereafter detailed. Consider  $N_A$  unrelated genotyped breed A reference animals and  $N_{AB}$  unrelated genotyped crossbred AB reference animals. It is assumed that the breed A (B) specific effect  $\beta_{c_k}^{*(A)}$  ( $\beta_{c_k}^{*(B)}$ ) of each  $k^{\text{th}}$  independent locus explains an equal amount of the breed A (B) specific additive genetic variance for the crossbred performance trait  $\sigma_{c_A}^2$  ( $\sigma_{c_B}^2$ ), i.e.  $\sigma_{c_A}^2 = Me^{(A)}\sigma_{\beta_c^{*(A)}}^2$  ( $\sigma_{c_B}^2 = Me^{(B)}\sigma_{\beta_c^{*(B)}}^2$ ) with  $Me^{(A)}$  ( $Me^{(B)}$ ) being the effective number of chromosome segments underlying the crossbred performance trait for the breed A (B). It is also assumed that the effect  $\beta_{a_k}^{*(A)}$  of each  $k^{\text{th}}$  independent locus explains an equal amount of the additive genetic variance for the purebred performance trait  $\sigma_{a_A}^2$ , i.e.  $\sigma_{a_A}^2 = Me\sigma_{\beta_a^{*(A)}}^2$  with  $Me$  being the effective number of chromosome segments underlying the purebred performance trait for the breed A. The (co)variances among  $\beta_a^{*(A)}$ ,  $\beta_c^{*(A)}$ , and  $\beta_c^{*(B)}$ , are assumed to be:

$$Var \begin{bmatrix} \beta_a^{*(A)} \\ \beta_c^{*(A)} \\ \beta_c^{*(B)} \end{bmatrix} = \begin{bmatrix} \mathbf{I}\sigma_{\beta_a^{*(A)}}^2 & \mathbf{I}\sigma_{\beta_a^{*(A)}\beta_c^{*(A)}} & \mathbf{0} \\ \mathbf{I}\sigma_{\beta_a^{*(A)}\beta_c^{*(A)}} & \mathbf{I}\sigma_{\beta_c^{*(A)}}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_{\beta_c^{*(B)}}^2 \end{bmatrix}.$$

The entries of the matrix  $\mathbf{Z}_A^{*(A)}$ ,  $\mathbf{Z}_{AB}^{*(A)}$ , and  $\mathbf{Z}_{AB}^{*(B)}$  are defined as previously in Appendices A and B.

The genomic breeding value ( $c_{a_i}^{(A)}$ ) for the  $i^{\text{th}}$  purebred selection candidate can be predicted as

follows:

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

where  $\mathbf{z}_{a_i}^{(A)}$  is a row vector of the standardized genotypes for the  $Me^{(A)}$  independent loci of the  $i^{\text{th}}$  selection candidate of breed A; and the vector  $\widehat{\boldsymbol{\beta}}_c^{*(A)}$  is the vector of the predictions of  $\boldsymbol{\beta}_c^{*(A)}$ .

Following the mixed model theory [17, 19], the reliability of  $\widehat{c}_{a_i}^{(A)}$ ,  $r_{c_{a_i}}^2$ , is equal to:

$$\begin{aligned} r_{c_{a_i}}^2 &= \frac{\text{Var}\left(\widehat{c}_{a_i}^{(A)}\right)}{\text{Var}\left(c_{a_i}^{(A)}\right)} = \frac{\text{Var}\left(\mathbf{z}_{a_i}^{*(A)} \widehat{\boldsymbol{\beta}}_c^{*(A)}\right)}{\text{Var}\left(\mathbf{z}_{a_i}^{*(A)} \boldsymbol{\beta}_c^{*(A)}\right)} \\ &= \frac{\mathbf{z}_{a_i}^{*(A)} \text{Var}\left(\widehat{\boldsymbol{\beta}}_c^{*(A)}\right) \mathbf{z}_{a_i}^{*(A)'} }{\mathbf{z}_{a_i}^{*(A)} \text{Var}\left(\boldsymbol{\beta}_c^{*(A)}\right) \mathbf{z}_{a_i}^{*(A)'} } = \frac{\text{Var}\left(\widehat{\boldsymbol{\beta}}_{c_k}^{*(A)}\right)}{\text{Var}\left(\boldsymbol{\beta}_{c_k}^{*(A)}\right)} = r_{\boldsymbol{\beta}_c}^2, \end{aligned}$$

because it is assumed that the effect  $\boldsymbol{\beta}_{c_k}^{*(A)}$  of each  $k^{\text{th}}$  independent locus explains an equal amount of  $\sigma_{c_A}^2$ , and that the accuracy of the estimated effect,  $r_{\boldsymbol{\beta}_c}^2$ , is the same for each locus.

The reliability  $r_{\boldsymbol{\beta}_c}^2$  can be approximated as follows. The prediction of  $\boldsymbol{\beta}_{c_k}^{*(A)}$  for the  $k^{\text{th}}$  locus can be performed from the phenotypes,  $\mathbf{y}_A$  and  $\mathbf{y}_{AB}$ , corrected for all other fixed effects, for the breed A specific effects other than the  $k^{\text{th}}$  effect,  $\widehat{\boldsymbol{\beta}}_{a_{\neq k}}^{*(A)}$  and  $\widehat{\boldsymbol{\beta}}_{c_{\neq k}}^{*(A)}$ , as well as for the breed B specific effects,  $\widehat{\boldsymbol{\beta}}_c^{*(B)}$ ,  $\widehat{\mathbf{y}}_A^*$  and  $\widehat{\mathbf{y}}_{AB}^*$  respectively, using the model:

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

where the vectors  $\mathbf{z}_{A_k}^{*(A)}$  and  $\mathbf{z}_{AB_k}^{*(A)}$  contain the standardized breed A genotypes of breed A reference animals and breed A alleles of crossbred AB reference animals, respectively; and

$\boldsymbol{\varepsilon}_{A_k}$  and  $\boldsymbol{\varepsilon}_{AB_k}$  are the residual vectors.

For simplicity, it will be assumed that  $Var\left(\begin{bmatrix} \boldsymbol{\varepsilon}_{A_k} \\ \boldsymbol{\varepsilon}_{AB_k} \end{bmatrix}\right) = \begin{bmatrix} \mathbf{I}\sigma_{\varepsilon_A}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_{\varepsilon_{AB}}^2 \end{bmatrix}$ . Following the mixed

model theory [17,19], the prediction of  $\begin{bmatrix} \beta_{a_k}^{*(A)} \\ \beta_{c_k}^{*(A)} \end{bmatrix}$ ,  $\begin{bmatrix} \hat{\beta}_{a_k}^{*(A)} \\ \hat{\beta}_{c_k}^{*(A)} \end{bmatrix}$ , is equal to:

$$\begin{aligned} & \begin{bmatrix} \hat{\beta}_{a_k}^{*(A)} \\ \hat{\beta}_{c_k}^{*(A)} \end{bmatrix} = \\ & \left( \begin{bmatrix} \mathbf{z}_{A_k}^{*(A)'} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{AB_k}^{*(A)'} \end{bmatrix} \begin{bmatrix} \mathbf{I}\sigma_{\varepsilon_A}^{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_{\varepsilon_{AB}}^{-2} \end{bmatrix} \begin{bmatrix} \mathbf{z}_{A_k}^{*(A)} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{AB_k}^{*(A)} \end{bmatrix} + \right. \\ & \left. \begin{bmatrix} \sigma_{\beta_a}^2 & \sigma_{\beta_a} \sigma_{\beta_c} \\ \sigma_{\beta_a} \sigma_{\beta_c} & \sigma_{\beta_c}^2 \end{bmatrix}^{-1} \right)^{-1} \begin{bmatrix} \mathbf{z}_{A_k}^{*(A)'} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{AB_k}^{*(A)'} \end{bmatrix} \begin{bmatrix} \mathbf{I}\sigma_{\varepsilon_A}^{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_{\varepsilon_{AB}}^{-2} \end{bmatrix} \begin{bmatrix} \widehat{\mathbf{y}}_A^* \\ \widehat{\mathbf{y}}_{AB}^* \end{bmatrix}, \quad (C1) \end{aligned}$$

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

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

where  $PEV_{\hat{\beta}_{c_k}^{*(A)}}$  is the prediction error variance of  $\hat{\beta}_{c_k}^{*(A)}$  and is equal to the diagonal element of

inverse of the left-hand-side of the mixed model equations (C1) [17].

The inverse of left-hand-side of the mixed model equations (C1) can be simplified as:

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

$$\begin{aligned}
&= \left( \begin{bmatrix} \mathbf{z}_{A_k}^{*(A)'} \mathbf{z}_{A_k}^{*(A)} \sigma_{\varepsilon_A}^{-2} & 0 \\ 0 & \mathbf{z}_{AB_k}^{*(A)'} \mathbf{z}_{AB_k}^{*(A)} \sigma_{\varepsilon_{AB}}^{-2} \end{bmatrix} \right. \\
&\quad \left. + \frac{1}{\sigma_{\beta_a}^{*(A)} \sigma_{\beta_c}^{*(A)} - \sigma_{\beta_a \beta_c}^{*(A)}} \begin{bmatrix} \sigma_{\beta_c}^{*(A)} & -\sigma_{\beta_a \beta_c}^{*(A)} \\ -\sigma_{\beta_a \beta_c}^{*(A)} & \sigma_{\beta_a}^{*(A)} \end{bmatrix} \right)^{-1} \\
&= \left( \begin{bmatrix} N_A \sigma_{\varepsilon_A}^{-2} & 0 \\ 0 & \frac{N_{AB}}{2} \sigma_{\varepsilon_{AB}}^{-2} \end{bmatrix} + \frac{1}{\sigma_{\beta_a}^{*(A)} \sigma_{\beta_c}^{*(A)} - \sigma_{\beta_a \beta_c}^{*(A)}} \begin{bmatrix} \sigma_{\beta_c}^{*(A)} & -\sigma_{\beta_a \beta_c}^{*(A)} \\ -\sigma_{\beta_a \beta_c}^{*(A)} & \sigma_{\beta_a}^{*(A)} \end{bmatrix} \right)^{-1}
\end{aligned}$$

Defining the correlation between  $\beta_a^{*(A)}$  and  $\beta_c^{*(A)}$  as  $r_{\beta_a \beta_c}^{*(A)} = \frac{\sigma_{\beta_a \beta_c}^{*(A)}}{\sigma_{\beta_a}^{*(A)} \sigma_{\beta_c}^{*(A)}}$ , it follows that:

$$\frac{1}{\sigma_{\beta_a}^{*(A)} \sigma_{\beta_c}^{*(A)} - \sigma_{\beta_a \beta_c}^{*(A)}} = \frac{1}{\sigma_{\beta_a}^{*(A)} \sigma_{\beta_c}^{*(A)} (1 - r_{\beta_a \beta_c}^{*(A)})} = \frac{1}{\tau}, \text{ and}$$

$$\begin{aligned}
&\left( \begin{bmatrix} \mathbf{z}_{A_k}^{*(A)'} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{AB_k}^{*(A)'} \end{bmatrix} \begin{bmatrix} \mathbf{I} \sigma_{\varepsilon_A}^{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \sigma_{\varepsilon_{AB}}^{-2} \end{bmatrix} \begin{bmatrix} \mathbf{z}_{A_k}^{*(A)} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{AB_k}^{*(A)} \end{bmatrix} + \begin{bmatrix} \sigma_{\beta_c}^{*(A)} & \sigma_{\beta_a \beta_c}^{*(A)} \\ \sigma_{\beta_a \beta_c}^{*(A)} & \sigma_{\beta_a}^{*(A)} \end{bmatrix} \right)^{-1} \\
&= \tau \left( \begin{bmatrix} \tau N_A \sigma_{\varepsilon_A}^{-2} + \sigma_{\beta_c}^{*(A)} & -\sigma_{\beta_a \beta_c}^{*(A)} \\ -\sigma_{\beta_a \beta_c}^{*(A)} & \tau \frac{N_{AB}}{2} \sigma_{\varepsilon_{AB}}^{-2} + \sigma_{\beta_a}^{*(A)} \end{bmatrix} \right)^{-1} \\
&= \frac{\tau}{\left( \tau N_A \sigma_{\varepsilon_A}^{-2} + \sigma_{\beta_c}^{*(A)} \right) \left( \tau \frac{N_{AB}}{2} \sigma_{\varepsilon_{AB}}^{-2} + \sigma_{\beta_a}^{*(A)} \right) - \sigma_{\beta_a \beta_c}^{*(A)}} \begin{bmatrix} \tau \frac{N_{AB}}{2} \sigma_{\varepsilon_{AB}}^{-2} + \sigma_{\beta_a}^{*(A)} & \sigma_{\beta_a \beta_c}^{*(A)} \\ \sigma_{\beta_a \beta_c}^{*(A)} & \tau N_A \sigma_{\varepsilon_A}^{-2} + \sigma_{\beta_c}^{*(A)} \end{bmatrix}
\end{aligned}$$

The prediction error variance of  $\hat{\beta}_{c_k}^{*(A)}$ ,  $PEV_{\hat{\beta}_{c_k}^{*(A)}}$ , is then equal to

$$PEV_{\hat{\beta}_{c_k}^{*(A)}} = \frac{\tau \left( \tau N_A \sigma_{\varepsilon_A}^{-2} + \sigma_{\beta_c}^{*(A)} \right)}{\left( \tau N_A \sigma_{\varepsilon_A}^{-2} + \sigma_{\beta_c}^{*(A)} \right) \left( \tau \frac{N_{AB}}{2} \sigma_{\varepsilon_{AB}}^{-2} + \sigma_{\beta_a}^{*(A)} \right) - \sigma_{\beta_a \beta_c}^{*(A)}}, \text{ and the reliability of } \hat{\beta}_{c_k}^{*(A)} \text{ is equal to:}$$



$$= \frac{\frac{1}{2} \sigma_{\beta_c^{*(A)}}^2 \sigma_{\varepsilon_{AB}}^{-2} \left( \sigma_{\beta_a^{*(A)}}^2 \sigma_{\varepsilon_A}^{-2} + \frac{1}{N_A} \right) + \sigma_{\beta_a^{*(A)}}^2 \sigma_{\varepsilon_A}^{-2} r_{\beta_a^{*(A)} \beta_c^{*(A)}}^2 \left( \frac{1}{2} \sigma_{\beta_c^{*(A)}}^2 \sigma_{\varepsilon_{AB}}^{-2} + \frac{1}{N_{AB}} \right) - \sigma_{\beta_a^{*(A)}}^2 \sigma_{\beta_c^{*(A)}}^2 \sigma_{\varepsilon_A}^{-2} \sigma_{\varepsilon_{AB}}^{-2} r_{\beta_a^{*(A)} \beta_c^{*(A)}}^2}{\left( \sigma_{\beta_a^{*(A)}}^2 \sigma_{\varepsilon_A}^{-2} + \frac{1}{N_A} \right) \left( \frac{1}{2} \sigma_{\beta_c^{*(A)}}^2 \sigma_{\varepsilon_{AB}}^{-2} + \frac{1}{N_{AB}} \right) - \frac{1}{2} \sigma_{\beta_a^{*(A)}}^2 \sigma_{\beta_c^{*(A)}}^2 \sigma_{\varepsilon_A}^{-2} \sigma_{\varepsilon_{AB}}^{-2} r_{\beta_a^{*(A)} \beta_c^{*(A)}}^2}$$

It is worth noting that the denominator:

$$\left( \sigma_{\beta_a^{*(A)}}^2 \sigma_{\varepsilon_A}^{-2} + \frac{1}{N_A} \right) \left( \frac{1}{2} \sigma_{\beta_c^{*(A)}}^2 \sigma_{\varepsilon_{AB}}^{-2} + \frac{1}{N_{AB}} \right) - \frac{1}{2} \sigma_{\beta_a^{*(A)}}^2 \sigma_{\beta_c^{*(A)}}^2 \sigma_{\varepsilon_A}^{-2} \sigma_{\varepsilon_{AB}}^{-2} r_{\beta_a^{*(A)} \beta_c^{*(A)}}^2$$

is the determinant of the 2-by-2 matrix:

$$\begin{bmatrix} \sigma_{\beta_a^{*(A)}}^2 \sigma_{\varepsilon_A}^{-2} + \frac{1}{N_A} & \sqrt{\frac{1}{2} \sigma_{\beta_a^{*(A)}}^2 \sigma_{\beta_c^{*(A)}}^2 \sigma_{\varepsilon_A}^{-2} \sigma_{\varepsilon_{AB}}^{-2} r_{\beta_a^{*(A)} \beta_c^{*(A)}}^2} \\ \sqrt{\frac{1}{2} \sigma_{\beta_a^{*(A)}}^2 \sigma_{\beta_c^{*(A)}}^2 \sigma_{\varepsilon_A}^{-2} \sigma_{\varepsilon_{AB}}^{-2} r_{\beta_a^{*(A)} \beta_c^{*(A)}}^2} & \frac{1}{2} \sigma_{\beta_c^{*(A)}}^2 \sigma_{\varepsilon_{AB}}^{-2} + \frac{1}{N_{AB}} \end{bmatrix}$$

Therefore, after reordering the terms, it follows that:

$$r_{\beta_c^{*(A)}} =$$

$$\begin{bmatrix} \sqrt{\frac{\sigma_{\beta_a^{*(A)}}^2 r_{\beta_a^{*(A)} \beta_c^{*(A)}}^2}{\sigma_{\varepsilon_A}^2}} \\ \sqrt{\frac{\sigma_{\beta_c^{*(A)}}^2}{2\sigma_{\varepsilon_{AB}}^2}} \end{bmatrix}' \begin{bmatrix} \frac{\sigma_{\beta_a^{*(A)}}^2}{\sigma_{\varepsilon_A}^2} + \frac{1}{N_A} & \sqrt{\frac{\sigma_{\beta_a^{*(A)}}^2 \sigma_{\beta_c^{*(A)}}^2 r_{\beta_a^{*(A)} \beta_c^{*(A)}}^2}{2\sigma_{\varepsilon_A}^2 \sigma_{\varepsilon_{AB}}^2}} \\ \sqrt{\frac{\sigma_{\beta_a^{*(A)}}^2 \sigma_{\beta_c^{*(A)}}^2 r_{\beta_a^{*(A)} \beta_c^{*(A)}}^2}{2\sigma_{\varepsilon_A}^2 \sigma_{\varepsilon_{AB}}^2}} & \frac{\sigma_{\beta_c^{*(A)}}^2}{2\sigma_{\varepsilon_{AB}}^2} + \frac{1}{N_{AB}} \end{bmatrix}^{-1} \begin{bmatrix} \sqrt{\frac{\sigma_{\beta_a^{*(A)}}^2 r_{\beta_a^{*(A)} \beta_c^{*(A)}}^2}{\sigma_{\varepsilon_A}^2}} \\ \sqrt{\frac{\sigma_{\beta_c^{*(A)}}^2}{2\sigma_{\varepsilon_{AB}}^2}} \end{bmatrix}$$

By approximating the residual variances  $\sigma_{\varepsilon_A}^2$  and  $\sigma_{\varepsilon_{AB}}^2$  by the corresponding  $\sigma_{P_A}^2$  and  $\sigma_{P_{AB}}^2$ , and

because  $r_{c_{ai}}^2 = r_{\beta_c^{*(A)}}^2$ ,  $\sigma_{c_A}^2 = Me^{(A)} \sigma_{\beta_c^{*(A)}}^2$ ,  $\sigma_{a_A}^2 = Me \sigma_{\beta_a^{*(A)}}^2$ , and, following Wientjes et al.

$$[14], r_{\beta_a^{*(A)} \beta_c^{*(A)}} = \frac{\sigma_{\beta_a^{*(A)} \beta_c^{*(A)}}}{\sigma_{\beta_a^{*(A)}} \sigma_{\beta_c^{*(A)}}} = r_{PC}^{(A)} \text{ with } r_{PC}^{(A)} \text{ being the breed A specific genetic correlation}$$

between the purebred and crossbred performance traits, the predicted reliability of the genomic estimated breeding values for the  $i^{\text{th}}$  selection candidate of breed A is equal to, without availability of genotyping data:

$$r_{c_{ai}}^2 = r_{\beta_c^*(A)}^2 = \begin{bmatrix} r_{PC}^{(A)} \sqrt{\frac{h_{aA}^2}{Me}} & \sqrt{\frac{h_{cA}^2}{2Me^{(A)}}} \end{bmatrix} \begin{bmatrix} \frac{h_{aA}^2}{Me} + \frac{1}{N_A} & r_{PC}^{(A)} \sqrt{\frac{h_{cA}^2}{2Me^{(A)}} \frac{h_{aA}^2}{Me}} \\ r_{PC}^{(A)} \sqrt{\frac{h_{cA}^2}{2Me^{(A)}} \frac{h_{aA}^2}{Me}} & \frac{h_{cA}^2}{2Me^{(A)}} + \frac{1}{N_{AB}} \end{bmatrix}^{-1} \begin{bmatrix} r_{PC}^{(A)} \sqrt{\frac{h_{aA}^2}{Me}} \\ \sqrt{\frac{h_{cA}^2}{2Me^{(A)}}} \end{bmatrix},$$

where  $h_{aA}^2$  is the heritability of the purebred performance trait for the breed A and  $h_{cA}^2$  is the breed A specific heritability for the crossbred performance trait.