Implication of the order of blending and tuning when computing the genomic relationship matrix in single-step GBLUP

Taylor Mcwhorter, Matias Bermann, Andre Garcia, Andrés Legarra, Ignacio Aguilar, Ignacy Misztal, Daniela Lourenco

Received: 4 May 2022 | Accepted: 12 July 2022

DOI: 10.1111/jbg.12734

Journal of Animal Breeding and Genetics WILEY

ORIGINAL ARTICLE

Implication of the order of blending and tuning when computing the genomic relationship matrix in single-step GBLUP

Taylor M. McWhorter[1] | **Matias Bermann**[1] | **Andre L. S. Garcia**[1] | **Andrés Legarra**[2] | **Ignacio Aguilar**[3] | **Ignacy Misztal**[1] | **Daniela Lourenco**[1]

[1]Department of Animal and Dairy Science, University of Georgia, Athens, Georgia, USA

[2]UMR GenPhySE, Institut National de la Recherche Agronomique, Castanet-Tolosan, France

[3]Department of Animal Breeding, Instituto Nacional de Investigacion Agropecuaria, Montevideo, Uruguay

Correspondence
Taylor M. McWhorter, Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602, USA.
Email: taylor.mcwhorter@uga.edu

Funding information
National Institute of Food and Agriculture

Abstract

Single-step genomic BLUP (ssGBLUP) relies on the combination of the genomic (**G**) and pedigree relationship matrices for all (**A**) and genotyped ($\mathbf{A}_{22}$) animals. The procedure ensures **G** and $\mathbf{A}_{22}$ are compatible so that both matrices refer to the same genetic base ('tuning'). Then **G** is combined with a proportion of $\mathbf{A}_{22}$ ('blending') to avoid singularity problems and to account for the polygenic component not accounted for by markers. This computational procedure has been implemented in the reverse order (blending before tuning) following the sequential research developments. However, blending before tuning may result in less optimal tuning because the blended matrix already contains a proportion of $\mathbf{A}_{22}$. In this study, the impact of 'tuning before blending' was compared with 'blending before tuning' on genomic estimated breeding values (GEBV), single nucleotide polymorphism (SNP) effects and indirect predictions (IP) from ssGBLUP using American Angus Association and Holstein Association USA, Inc. data. Two slightly different tuning methods were used; one that adjusts the mean diagonals and off-diagonals of **G** to be similar to those in $\mathbf{A}_{22}$ and another one that adjusts based on the average difference between all elements of **G** and $\mathbf{A}_{22}$. Over 6 million Angus growth records and 5.9 million Holstein udder depth records were available. Genomic information was available on 51,478 Angus and 105,116 Holstein animals. Average realized relationship estimates among groups of animals were similar across scenarios. Scatterplots show that GEBV, SNP effects and IP did not noticeably change for all animals in the evaluation regardless of the order of computations and when using blending parameter of 0.05. Formulas were derived to determine the blending parameter that maximizes changes in the genomic relationship matrix and GEBV when changing the order of blending and tuning. Algebraically, the change is maximized when the blending parameter is equal

to 0.5. Overall, tuning **G** before blending, regardless of blending parameter used, had a negligible impact on genomic predictions and SNP effects in this study.

KEYWORDS

genetic base, indirect predictions, residual polygenic effect, scaling of genomic matrices, single-step genomic best linear unbiased prediction

1 | INTRODUCTION

Single-step genomic best linear unbiased prediction (ssGBLUP) allows for the combination of phenotypes, pedigree and single nucleotide polymorphism (SNP) information to obtain the genomic estimated breeding values (GEBV) for all the animals in an evaluation (Misztal et al., 2009). In this method, the inverse of the pedigree relationship matrix, $\mathbf{A}^{-1}$, is replaced by the inverse of the realized relationship matrix, $\mathbf{H}^{-1}$(Aguilar et al., 2010). The $\mathbf{H}^{-1}$ is composed of $\mathbf{A}^{-1}$, the inverse of the genomic relationship matrix (**G**) and the inverse of the pedigree relationship matrix among genotyped animals $(\mathbf{A}_{22})$.

When constructing $\mathbf{H}^{-1}$ in ssGBLUP, two main challenges arise: (i) the singularity of **G** (VanRaden, 2008), and (ii) the compatibility between **G** and $\mathbf{A}_{22}$ (VanRaden, 2008; Vitezica et al., 2011). The singularity of **G** is solved in a procedure called *blending*, in which a weighted sum of **G** and a positive-definite matrix is used. A common choice of blending is described in VanRaden (2008), where the blended genomic relationship matrix $\mathbf{G}_b = (1-\beta)\mathbf{G} + \beta\mathbf{A}_{22}$, with $0 < \beta < 1$ known as the blending parameter or proportion. This blending method is equivalent to fitting a residual polygenic effect in a SNP-BLUP model (Liu et al., 2016). Another option is to blend **G** with an identity matrix instead of $\mathbf{A}_{22}$. Although the second option is computationally more convenient, the first option is still preferred (Legarra et al., 2022). The compatibility between **G** and $\mathbf{A}_{22}$ is contingent on the genetic bases of the two matrices as well as the frequencies of the selected SNP and centring of the genotypes in the construction of **G** (VanRaden, 2008; Vitezica et al., 2011). The differences in the genetic base between **G** and $\mathbf{A}_{22}$ arise because **G** is not constructed with the allele frequencies of the base population (VanRaden, 2008) and the pedigree used to construct $\mathbf{A}_{22}$ might be incomplete (Misztal et al., 2013). More often, **G** is constructed with allele frequencies based on the current, observed population ('centred' coding) or a constant 0.5, because estimating the allele frequencies in the base population is not straightforward (Gengler et al., 2007). Utilizing centred coding assumes that the expectation of breeding values for genotyped animals is 0 which poses another obstacle. When the population has undergone selection, more recent animals should have higher genetic values than the base. To account for this fact, several methods were proposed to establish compatibility between the relationship matrices of genotyped animals, either to scale **G** to $\mathbf{A}_{22}$ (Christensen et al., 2012; Vitezica et al., 2011) or to scale $\mathbf{A}_{22}$ to **G** (Christensen, 2012; Legarra et al., 2015). All these methods implicitly estimate the difference between the genetic bases in **G** and in **A** (Vitezica et al., 2011). The methods belonging to the first group consist of multiplying and adding constants obtained from $\mathbf{A}_{22}$ to the elements of **G**. The methods of the second group consist of modifying the entire $\mathbf{A}^{-1}$ with parameters calculated from genomic information (Garcia-Baccino et al., 2017). Although the second group of methods provides a sturdy framework based on genetic theory, there is a need to estimate extra parameters in the model (Garcia-Baccino et al., 2017). This is a difficult task for large-scale genetic evaluations. Hereafter, the first group of methods will be referred to as the *tuning* of **G**.

In the early stages of ssGBLUP implementation, blending was applied before tuning. Blending before tuning was implemented as the default order of operations in the BLUPF90 family of programs (Misztal et al., 2014) until version 1.306 (2022) of the genomic library and has been observed in literature (Masuda et al., 2016; Pocrnic, 2017). However, blending **G** before tuning may add bias to GEBV because the original **G** is not in the same scale as $\mathbf{A}_{22}$. Thus, the blended **G** already contains a portion of $\mathbf{A}_{22}$ and refers to a different genetic base. Further tuning this resulting **G** with a residual polygenic effect included creates a theoretical inconsistency which results in a suboptimal correction for the difference in the genetic base between the two matrices. The inconsistency is expected to be small when the blending parameter β has a small value, for example, 0.05 as in VanRaden (2008). However, other evaluations (Andersen et al., 2021; Interbull Centre, 2021; Alkhoder & Liu, 2021) use higher numbers of 0.20 or more for blending. These theoretical inconsistencies may impact GEBV and indirect predictions (IP) as IP are a linear function of GEBV and $\mathbf{G}^{-1}$. It is therefore more appropriate to first, make the original (unblended, untuned) **G** resemble $\mathbf{A}_{22}$ and later blend it with a polygenic fraction of $\mathbf{A}_{22}$. Thus, the objectives of this study were to

investigate the impact of changing the order of blending and tuning on the genomic relationship matrix and genomic predictions. Changes in GEBV and IP were evaluated using American Angus Association and Holstein Association USA, Inc. data. Mathematical expressions were derived to determine the blending parameter that maximizes changes in the genomic relationship matrix and genomic predictions when changing the order of blending and tuning.

2 | MATERIALS AND METHODS

The data were obtained from existing databases; therefore, approval from the Animal Care and Use Committee was not obtained for this study.

2.1 | Description of data

Phenotypic data of Angus animals were provided by American Angus Association (AAA, Saint Joseph, MO) and included 6,189,661 birth weight (BW) records, 6,890,625 weaning weight (WW) records and 3,387,252 post-weaning gain (PWG) records. There were 8,236,425 animals in the pedigree born between 1955 and 2014. A total of 51,478 genotyped animals were available after 55 individuals were removed due to an animal call rate lower than 0.90. Angus cattle were genotyped or imputed to 54,609 SNP markers based on a combination of several different SNP chips. SNP with call rate lower than 0.9, minor allele frequencies (MAF) less than 0.05 and monomorphic SNP were removed in quality control (Wiggans et al., 2009). Genotypes and phenotypes of a set of 19,056 young, genotyped animals born in 2013–2014 were extracted from the existing data for computing IP. These genotypes were not used for the estimation of marker effects. Pedigrees of these animals were kept for simplicity, but it is known that these pedigrees carried negligible information (Lourenco et al., 2015).

Holstein cattle phenotypic data were provided by the Holstein Association USA, Inc. and included 5,932,709 udder depth (UD) records. The pedigree contained 8,338,450 animals born between 1983 and 2014. A total of 105,116 Holstein animals were genotyped and then imputed for 60,671 SNP based on a combination of several different SNP chips (Wiggans et al., 2016). Quality control was carried out by the data providers following Wiggans et al. (2009). Again, a separated dataset of young, genotyped Holsteins was created for IP as described by the Interbull validation method (Mäntysaari et al., 2010). The group of animals for IP consisted of 1,711 genotyped bulls with no progeny records in 2010, and at least 50 daughters with records in 2014. Thus, progeny records of these individuals did not contribute to their IP.

Both data providers conducted quality control on phenotypic data prior to sharing. Descriptive statistics of phenotypes are shown in Table 1 for both datasets.

2.2 | Models

The multiple-trait model used to evaluate all three AAA growth traits was as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{u} + \mathbf{Z}_2\mathbf{mat} + \mathbf{Z}_3\mathbf{mpe} + \mathbf{e} \quad (1)$$

where $\mathbf{y} = \left(\mathbf{y}'_{\text{bw}}\ \mathbf{y}'_{\text{ww}}\ \mathbf{y}'_{\text{pwg}}\right)'$ was the vector of phenotypes, $\mathbf{b}$ was the vector of fixed effects of contemporary groups, $\mathbf{u} = (\mathbf{u}'_{\text{bw}}\ \mathbf{u}'_{\text{ww}}\ \mathbf{u}'_{\text{pwg}})'$ was the vector of additive direct genetic effects, $\mathbf{mat} = (\mathbf{mat}'_{\text{bw}}\ \mathbf{mat}'_{\text{ww}}\ \mathbf{0}')'$ was the vector of maternal genetic effects, $\mathbf{mpe} = (\mathbf{0}'\ \mathbf{mpe}'_{\text{ww}}\ \mathbf{0}')'$ was the vector of maternal permanent environmental effects, $\mathbf{e}$ was the vector of error terms and $\mathbf{X}$, $\mathbf{Z}_1$, $\mathbf{Z}_2$ and $\mathbf{Z}_3$ were incidence matrices. All random effects were assumed to be multivariate normally distributed with null expectation and the following covariance structure:

$$\text{Var}\begin{pmatrix}\mathbf{u}\\ \mathbf{mat}\\ \mathbf{mpe}\\ \mathbf{e}\end{pmatrix} = \text{Var}\begin{pmatrix}\mathbf{u}_{\text{bw}}\\ \mathbf{u}_{\text{ww}}\\ \mathbf{u}_{\text{pwg}}\\ \mathbf{mat}_{\text{bw}}\\ \mathbf{mat}_{\text{ww}}\\ \mathbf{mpe}_{\text{ww}}\\ \mathbf{e}\end{pmatrix} = \begin{pmatrix}
\mathbf{H}\sigma^2_{\text{u}_{\text{bw}}} & \mathbf{H}\sigma_{\text{u}_{\text{bw}},\text{u}_{\text{ww}}} & \mathbf{H}\sigma_{\text{u}_{\text{bw}},\text{u}_{\text{pwg}}} & \mathbf{H}\sigma_{\text{u}_{bw},\text{mat}_{\text{bw}}} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
 & \mathbf{H}\sigma^2_{\text{u}_{\text{ww}}} & \mathbf{H}\sigma_{\text{u}_{ww},\text{u}_{pwg}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
 & & \mathbf{H}\sigma^2_{\text{u}_{\text{pgw}}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
 & & & \mathbf{H}\sigma^2_{\text{mat}_{\text{bw}}} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
 & & & & \mathbf{H}\sigma^2_{\text{mat}_{\text{ww}}} & \mathbf{0} & \mathbf{0}\\
 & & \text{symmetric} & & & \mathbf{I}\sigma^2_{\text{mpe}_{\text{ww}}} & \mathbf{0}\\
 & & & & & & \mathbf{R}\otimes\mathbf{I}
\end{pmatrix} \quad (2)$$

where $\mathbf{H}$ is the ssGBLUP covariance matrix defined in Legarra et al. (2009), σ_{i_j} denotes the covariance components of the ith effect for the jth combination of traits and $\mathbf{R}$ is the error covariance matrix among traits.

TABLE 1 Mean and standard deviations (SD) of phenotypes for genotyped and all Angus and Holstein cattle

Item	No. of animals with records	No. of records	Mean	*SD*
Angus				
Genotyped	51,478			
BW		50,388	35.85	4.16
WW		51,425	301.91	43.16
PWG		35,995	194.30	72.53
All	6,948,617			
BW		6,189,661	36.46	4.43
WW		6,890,625	263.12	44.62
PWG		3,387,252	162.26	67.12
Holstein				
Genotyped	105,116			
UD		37,292	32.97	8.29
All	10,067,745			
UD		10,067,731	28.85	8.57

Abbreviations: BW, birth weight (kg); PWG, post weaning gain (kg/day); UD, udder depth (score based on distance from udder floor to point of the hock); WW, weaning weight (kg).

The model used to evaluate UD for Holstein cattle was as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{W}_1\mathbf{u} + \mathbf{W}_2\mathbf{hs} + \mathbf{W}_3\mathbf{pe} + \mathbf{e} \tag{3}$$

where **y** was the vector of udder depth observations, **b** was the vector of fixed effects of herd-year-season, age × parity and stage of lactation × parity; **u** was the vector of additive genetic effects, **hs** was the vector of herd × sire interaction effects, **pe** was the vector of permanent environmental effects, **e** was the vector of errors and **X**, $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3$ were incidence matrices for vectors **b**, **u**, **hs** and **pe** respectively. As before, all random effects were assumed to be multivariate normally distributed with null expectation and the following covariance structure:

$$\text{Var}\begin{pmatrix}\mathbf{u}\\ \mathbf{hs}\\ \mathbf{pe}\\ \mathbf{e}\end{pmatrix} = \begin{pmatrix}\mathbf{H}\sigma^2_{\text{u}} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ & \mathbf{I}\sigma^2_{\text{hs}} & \mathbf{0} & \mathbf{0}\\ & & \mathbf{I}\sigma^2_{\text{pe}} & \mathbf{0}\\ & \text{symmetric} & & \mathbf{I}\sigma^2_{\text{e}}\end{pmatrix} \tag{4}$$

2.3 | Construction of genomic relationship matrix (G) and genomic predictions

In all analyses, **G** was created using the first method of VanRaden (2008):

$$\mathbf{G} = \frac{(\mathbf{M} - 2\mathbf{P})(\mathbf{M} - 2\mathbf{P})'}{2\sum_{i=1}^{m} p_i(1 - p_i)} \tag{5}$$

where **M** contains genotypes coded as {0, 1, 2} and **P** is a matrix whose columns contain observed allele frequencies across the entire dataset of the second allele at a locus p_i.

Here we describe how the blending and tuning procedures work. A blended matrix is calculated as

$$\mathbf{G}_{\text{b}} = (1 - \beta)\mathbf{G} + \beta\mathbf{A}_{22} \tag{6}$$

with $0 < \beta < 1$. A tuned matrix is obtained by scaling **G** to $\mathbf{A}_{22}$ as

$$\mathbf{G}_{\text{x}} = \mathbf{1}\mathbf{1}'\alpha_x + t_x\mathbf{G} \tag{7}$$

with constants α_x and t_x calculated from the following tuning methods. The subscript x was either *c* or *v* for Christensen et al. (2012) or Vitezica et al. (2011) tuning methods respectively. Christensen et al. (2012) adjust the means of diagonal elements ($\overline{diag\mathbf{G}}$) and of all elements ($\overline{\mathbf{G}}$) of **G** to be similar to those in $\mathbf{A}_{22}$ with constants α_c and t_c solving the two equations:

$$\overline{diag\mathbf{G}} \times t_c + \alpha_c = \overline{diag\mathbf{A}_{22}} \tag{8}$$

$$\overline{\mathbf{G}} \times t_c + \alpha_c = \overline{\mathbf{A}_{22}} \tag{9}$$

which results in $t_c = \frac{(\overline{diag\mathbf{A}_{22}} - \overline{\mathbf{A}_{22}})}{(\overline{diag\mathbf{G}} - \overline{\mathbf{G}})}$ or, equivalently after some algebra $\frac{(\overline{diag\mathbf{A}_{22}} - \overline{offdiag\mathbf{A}_{22}})}{(\overline{diag\mathbf{G}} - \overline{offdiag\mathbf{G}})}$ and $\alpha_c = \overline{\mathbf{A}_{22}} - \overline{\mathbf{G}} \times t_c$.

Vitezica et al. (2011) adjusts based on Wright's fixation index (F_{st}) with constants α_v and t_v:

$$\alpha_v = \overline{\mathbf{A}_{22}} - \overline{\mathbf{G}} \tag{10}$$

$$t_v = \left(1 - \frac{\alpha_v}{2}\right) \tag{11}$$

In this case, α_v is Wright's F_{st} which represents the difference between genomic and pedigree bases. In the case of Hardy–Weinberg equilibrium, both methods give nearly the same values of α_v and t_v.

It must be noted that although the algebra is quite general, these equations were conceived to be used with observed frequencies, in which case α_v and α_c are positive by construction. Otherwise, for instance if allele frequencies used to build **G** are 0.5, α_v or α_c may be negative and this may result in not positive definite $\mathbf{G}_x$.

For actual building of the final **G** matrix, there are two options. The first one compares relationship matrices and achieves compatibility, then modifies the resulting matrix. In this case, tuning is applied before blending and the final matrix can be understood as

$$\mathbf{G}_{xb} = (1 - \beta)\mathbf{G}_x + \beta\mathbf{A}_{22} \tag{12}$$

Alternatively, if blending is applied before tuning, the final matrix is

$$\mathbf{G}_{bx} = \mathbf{1}\mathbf{1}'\alpha_x + t_x\mathbf{G}_b \tag{13}$$

For each dataset, four sets of GEBV were calculated: $\mathbf{G}_{bc}$, $\mathbf{G}_{cb}$, $\mathbf{G}_{bv}$ and $\mathbf{G}_{vb}$, where the subscript indicates the order ('b' from blending comes earlier or later) and the form of tuning ('c' or 'v'—either Christensen et al. (2012) or Vitezica et al. (2011)). Expanded formulas for each genomic relationship matrix can be found in Appendix A.

In the initial analysis, a blending parameter of $\beta = 0.05$ (VanRaden, 2008) was used to calculate all four sets of GEBV. Then, we derived equations which determined the blending parameter β that maximizes changes in the genomic relationship matrix and genomic predictions when changing the order of blending and tuning using Christensen et al. (2012) and Vitezica et al. (2011) methods (Appendices B and C). Analyses were repeated using the blending parameter that maximizes changes, $\beta = 0.5$. After calculating each set of GEBV using the BLUP90IOD2 program (Misztal et al., 2014), a separate analysis was conducted to compute IP. GEBV were recalculated for each set but excluded genotypes and all phenotypes of the young, genotyped animals. SNP effects and IP were computed from the recalculated GEBV using POSTGSF90 (Misztal et al., 2014) and PREDF90 (Misztal et al., 2014) respectively. The calculation of SNP effects ($\hat{\mathbf{a}}$) given the calculated GEBV ($\hat{\boldsymbol{u}}$) also considered blending and tuning (Lourenco et al., 2020):

$$\hat{\boldsymbol{a}}_{xx} \mid \hat{\boldsymbol{u}}_{xx} = t_x(1 - \beta)(\mathbf{M} - 2\mathbf{P})' \frac{1}{2\sum p_i(1 - p_i)} \mathbf{G}_{xx}^{-1}\hat{\mathbf{u}}_{xx} \tag{14}$$

where the subscript xx varies from 'bc', 'cb', 'bv', or 'vb' according to the blending and tuning scenario. The 'marker' estimated breeding values on a genomic base ($\widehat{\mathbf{IP}}^*$) were calculated as

$$\widehat{\mathbf{IP}}^*_{xx} \mid \hat{\mathbf{a}}_{xx} = (\mathbf{M} - 2\mathbf{P})'\hat{\mathbf{a}}_{xx} \tag{15}$$

To transform $\widehat{\mathbf{IP}}^*$ to be 'marker' estimated breeding values on a pedigree scale and comparable to GEBV, a constant ($\hat{\mu}$) is back solved as the difference between pedigree and genomic bases and added (Legarra et al., 2022):

$$\hat{\mu}_{xx} \mid \hat{\mathbf{u}}_{xx} = E\left[\hat{\mu}_{xx} | \mathbf{u}_{xx} = \hat{\mathbf{u}}_{xx}\right] = \alpha_x(1 - \beta)\mathbf{1}'\mathbf{G}_{xx}^{-1}\hat{\boldsymbol{u}}_{xx} \tag{16}$$

The final 'marker' estimated breeding value on pedigree scale ($\widehat{\mathbf{IP}}$) is obtained as

$$\widehat{\mathbf{IP}}_{xx} = \hat{\mu}_{xx} + \widehat{\mathbf{IP}}^*_{xx} \tag{17}$$

GEBV, IP* and IP were compared among the four scenarios for both datasets.

2.4 | Average genomic relationships among genotyped relatives

The average estimate of realized relationships of genotyped animals for self-relationships and between unilineal (e.g., parent–offspring, half-sibs, grandparent–grandoffspring) and bilineal (e.g., full-sibs) relatives (Forneris et al., 2016) were compared for each method. The estimated realized relationships for genotyped animals are the proportions of genome-shared identical-by-state and were obtained from the diagonal (e.g., self) or off-diagonal (e.g., unilineal and bilineal relatives) elements of each **G** matrix. We obtained estimates of these values in each genomic relationship matrix after tuning and blending operations. These estimates were compared to the expectation of actual relationships.

3 | RESULTS AND DISCUSSION

3.1 | Changes in G between methods

In formulas found in Appendix A, (A1) to (A4), all the components are fixed for a given **G** and $\mathbf{A}_{22}$ except for β. Therefore, the four matrices can be understood as a function of the proportion of the residual polygenic effect in the model. Consequently, the differences between the matrices in formulas (A1) to (A4), which are the differences in GEBV when changing the order between blending and tuning, are dependent on β. When approaching the boundaries of β, that is, zero and one, the resulting matrix

is unique regardless of which procedure is performed first. Logically, when $\beta \to 1$, the residual polygenic effect is assumed to explain all the genetic variation and $\mathbf{G} \to \mathbf{A}_{22}$ for both tuning methods. As shown in Appendix B, the maximum change in GEBV when changing the order between blending and tuning using the method of Christensen et al. (2012) is attained when:

$$\beta = \frac{\sqrt{f(\mathbf{G})f(\mathbf{A}_{22})} - f(\mathbf{G})}{f(\mathbf{A}_{22}) - f(\mathbf{G})} \tag{18}$$

For the Angus dataset, this value was equal to 0.5001, whereas for the Holstein data was equal to 0.5016. The maximum absolute difference in the covariance matrix when changing the order between blending and tuning for the tuning method of Christensen et al. (2012) is bounded by the following equation:

$$\beta(1-\beta)\left|\frac{f(\mathbf{G}) - f(\mathbf{A}_{22})}{(1-\beta)f(\mathbf{G}) + \beta f(\mathbf{A}_{22})}\right|\left(\max_{i,j}\left|g_{ij}\right|\frac{f(\mathbf{A}_{22})}{f(\mathbf{G})} + \max_{i,j} a_{22_{ij}} + n^{-2}\,\mathbf{1}'\mathbf{A}_{22}\mathbf{1}\right) \tag{19}$$

For the Angus data, this value was equal to 1.29×10^{-4} for $\beta = 0.05$ and 6.82×10^{-4} for $\beta = 0.5001$, whereas for the Holstein data was 1.59×10^{-3} for $\beta = 0.05$ and 9.06×10^{-3} for $\beta = 0.5016$. Appendix C shows the maximum change in GEBV when changing the order between blending and tuning and using the tuning method of Vitezica et al. (2011) also occurs when β is equal to 0.5. The maximum absolute difference in the covariance matrix is bounded by the following equation:

$$\left(\beta - \beta^2\right)|h|\left(1 + F_{\max}\right) \tag{20}$$

where $F_{\max}$ denotes the maximum average of the genomic and pedigree inbreeding coefficients among the genotyped animals and h is half the average difference among the two matrices:

$$h = \frac{\mathbf{1}'(\mathbf{A}_{22} - \mathbf{G})\mathbf{1}}{2n^2} \tag{21}$$

where n is the number of genotyped animals. For the Angus data, this value was equal to 1.26×10^{-3} for $\beta = 0.05$

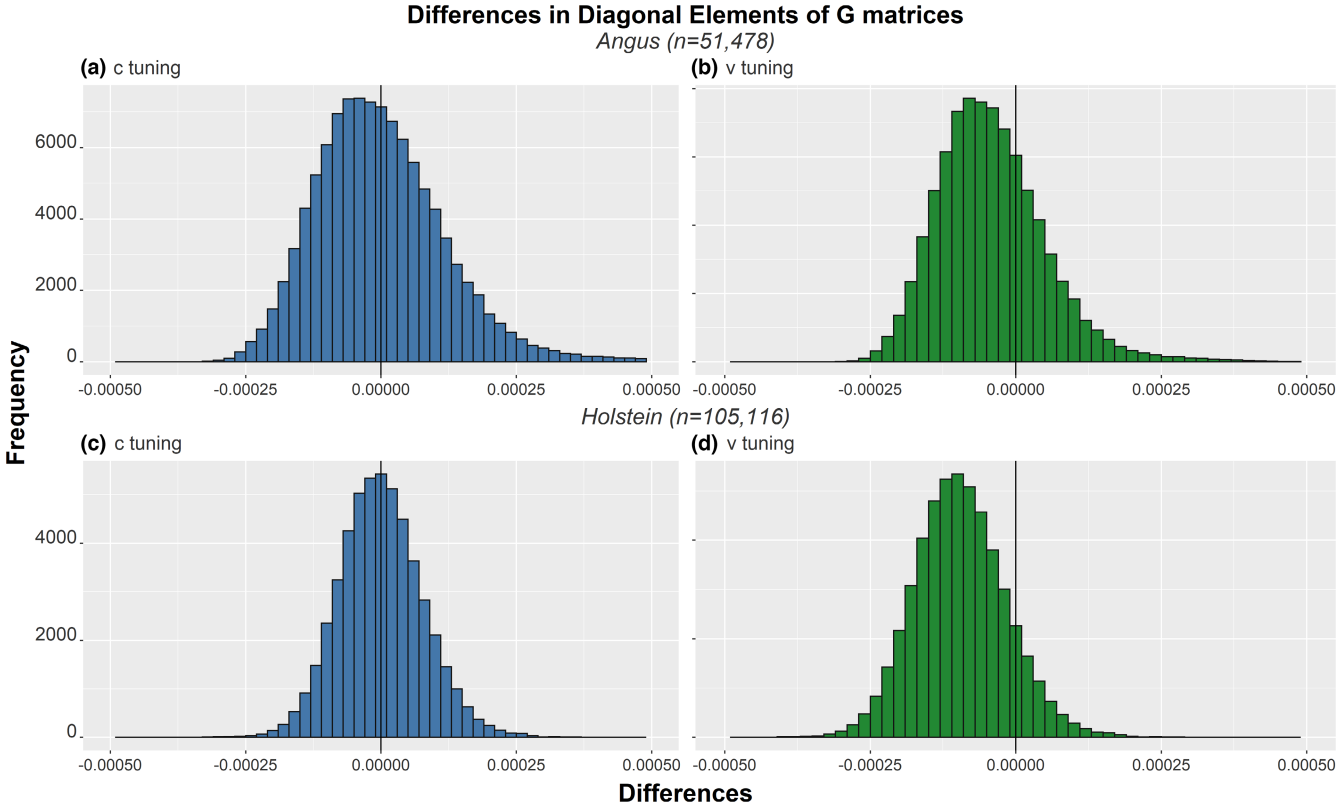

FIGURE 1 Differences in diagonal elements of genomic relationship matrices, **G**, with $\beta = 0.05$ in blending first and tuning first scenarios using two different tuning methods, c tuning (Christensen et al., 2012) and v tuning (Vitezica et al., 2011), for genotyped Angus (a and b) and Holstein (c and d)

TABLE 2 Expected and estimated realized (genomic) relationships[a] when changing the order between blending and tuning for two tuning methods using blending parameters, $\beta = 0.05$

Relationship	*N*	Expected relationship	Genomic relationships			
			G_{bc}	G_{cb}	G_{bv}	G_{vb}
			Mean (*SD*)	Mean (*SD*)	Mean (*SD*)	Mean (*SD*)
Angus						
Self	51,478	1.00	1.05 (0.04)	1.05 (0.04)	1.05 (0.04)	1.05 (0.04)
Parent–offspring	35,684	0.50	0.55 (0.05)	0.55 (0.05)	0.55 (0.05)	0.55 (0.05)
Full-sibs	7,672	0.50	0.55 (0.06)	0.55 (0.06)	0.55 (0.06)	0.55 (0.06)
Half-sibs	9,781,177	0.25	0.31 (0.04)	0.31 (0.04)	0.31 (0.04)	0.31 (0.04)
Grandparent–grandoffspring	3,788	0.25	0.33 (0.05)	0.33 (0.05)	0.33 (0.05)	0.33 (0.05)
Holstein						
Self	105,116	1.00	1.03 (0.04)	1.03 (0.04)	1.04 (0.04)	1.04 (0.04)
Parent–offspring	150,427	0.50	0.53 (0.04)	0.53 (0.04)	0.54 (0.04)	0.54 (0.04)
Full-sibs	57,314	0.50	0.53 (0.05)	0.53 (0.05)	0.54 (0.06)	0.54 (0.06)
Half-sibs	46,657,929	0.25	0.29 (0.04)	0.29 (0.04)	0.29 (0.04)	0.29 (0.04)
Grandparent–grandoffspring	134,832	0.25	0.29 (0.05)	0.29 (0.05)	0.29 (0.05)	0.29 (0.05)

[a]For the construction of the genomic relationship matrix $\mathbf{G}_{xx}$, the subscript c refers to Christensen et al. (2012) tuning, v refers to Vitezica et al. (2011) tuning, and b refers to blending $\mathbf{G}$ and $\mathbf{A}_{22}$. The order of the subscript refers to the order operations were performed.

TABLE 3 Statistics[a] for changes in GEBV for Angus and Holstein data sets when changing the order between blending and tuning for two tuning methods using blending parameters, $\beta = \{0.05, 0.5\}$

Data	Trait	Tuning method[b]	Absolute mean difference		Absolute maximum difference	
			$\beta = 0.05$	$\beta = 0.5$	$\beta = 0.05$	$\beta = 0.5$
			Genotyped animals			
Holstein	UD	c	0.006	0.001	0.01	0.02
		v	0.004	0.000	0.01	0.02
Angus	BW	c	0.000	0.001	0.01	0.02
		v	0.000	0.000	0.01	0.02
	WW	c	0.000	0.002	0.01	0.03
		v	0.000	0.001	0.01	0.03
	PWG	c	0.000	0.001	0.01	0.03
		v	0.000	0.001	0.01	0.03
			Non-genotyped animals			
Holstein	UD	c	0.006	0.001	0.01	0.02
		v	0.005	0.000	0.01	0.01
Angus	BW	c	0.000	0.001	0.00	0.01
		v	0.000	0.000	0.00	0.01
	WW	c	0.000	0.001	0.01	0.02
		v	0.000	0.000	0.01	0.02
	PWG	c	0.000	0.001	0.01	0.03
		v	0.000	0.000	0.01	0.01

[a]All statistics are expressed in terms of the genetic standard deviation of the trait.

[b]c refers to the tuning method of Christensen et al. (2012), whereas v refers to the tuning method of Vitezica et al. (2011).

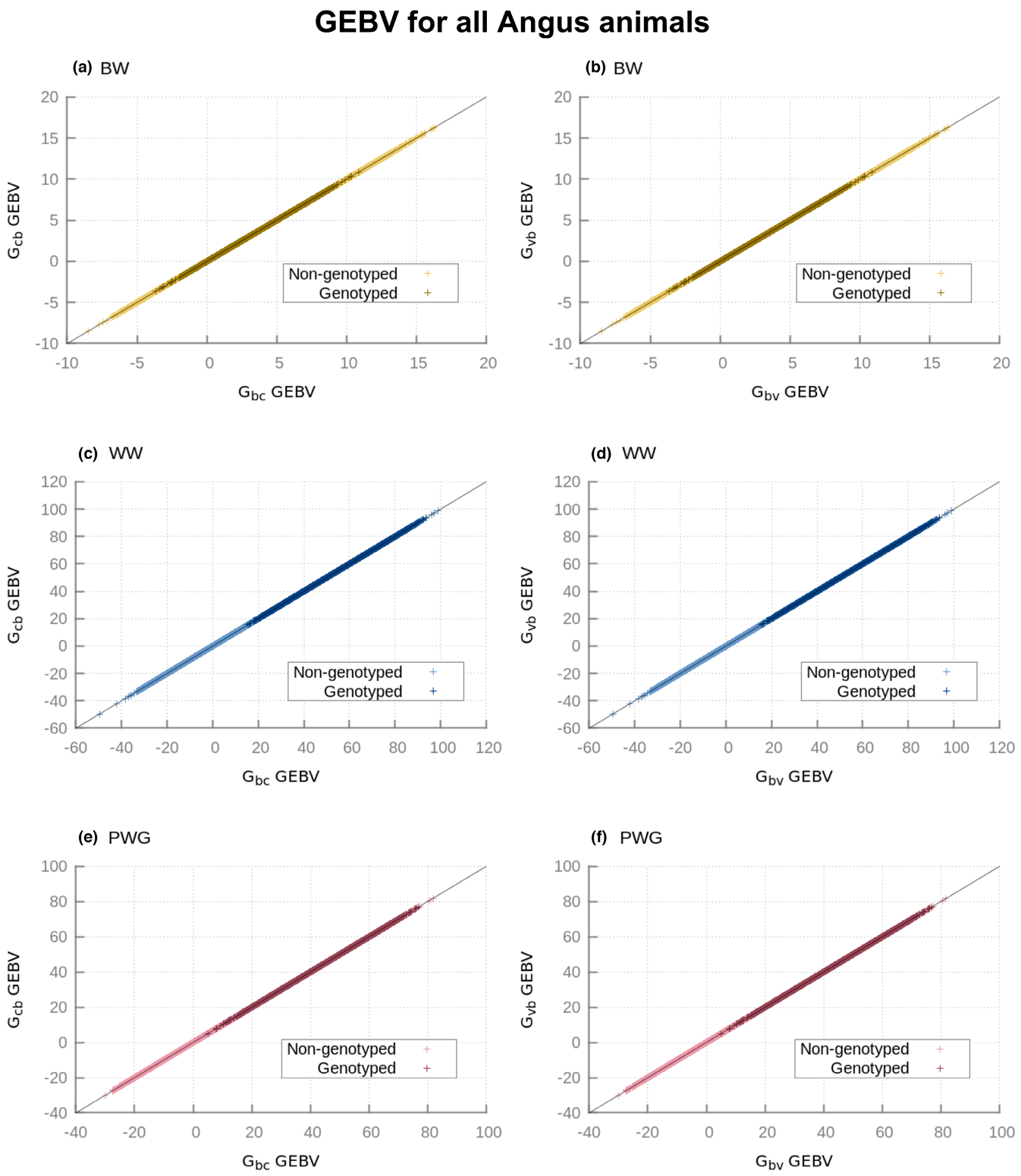

FIGURE 2 Scatter plots of GEBV for 51,478 genotyped and 8,184,542 non-genotyped Angus animals' birth weight (BW), weaning weight (WW) and post weaning gain (PWG) using $\beta = 0.05$ and blending before tuning ($\mathbf{G}_{bx}$) against predictions using tuning before blending ($\mathbf{G}_{xb}$) for two tuning methods; subscript x varies between c for Christensen et al. (2012) or v for Vitezica et al. (2011) tuning

and 6.65×10^{-3} for $\beta = 0.5$, whereas for the Holstein data was 1.12×10^{-3} for $\beta = 0.05$ and 5.94×10^{-3} for $\beta = 0.5$.

Across scenarios, the differences between GEBV or IP arise due to the different covariance matrices employed for the prediction. Therefore, if the matrices do not vary across the different combinations of blending and tuning, then the GEBV or IP will not vary. The average of off-diagonal elements of **G** in all four scenarios in the

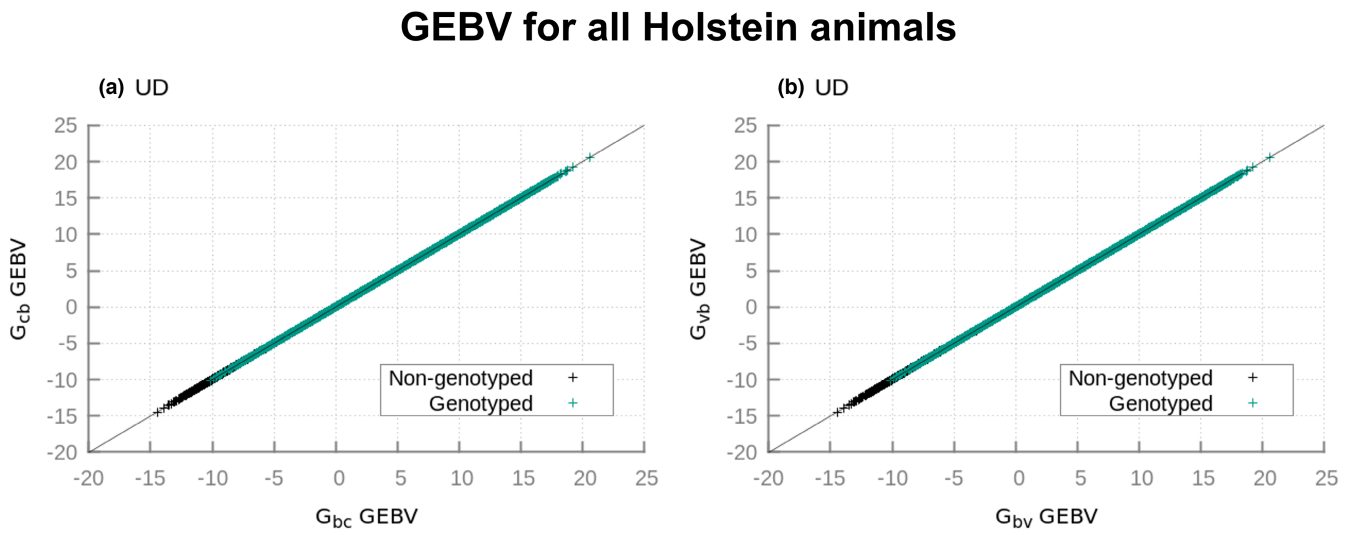

FIGURE 3 Scatter plots of GEBV for 105,116 genotyped and 8,233,333 non-genotyped Holstein animals' udder depth (UD) using $\beta = 0.05$ and blending before tuning ($\mathbf{G}_{bx}$) against predictions using tuning before blending ($\mathbf{G}_{xb}$) for two tuning methods; subscript x varies between c for Christensen et al. (2012) or v for Vitezica et al. (2011) tuning

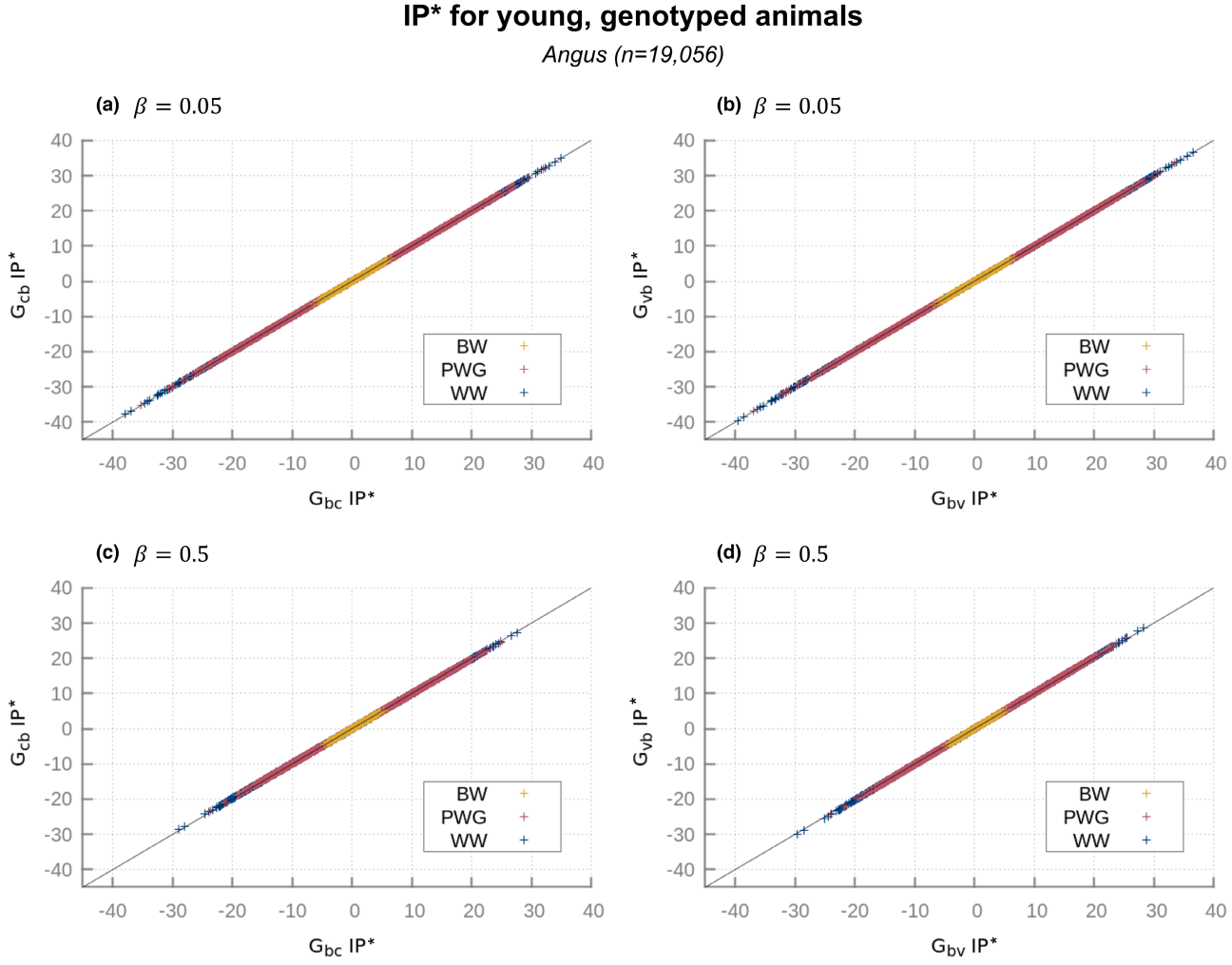

FIGURE 4 Scatter plots of indirect predictions on the genomic base (IP*) for young, genotyped Angus animals' birth weight (BW), weaning weight (WW) and post weaning gain (PWG) calculated using different blending parameters (β) and blending before tuning ($\mathbf{G}_{bx}$) against predictions using tuning before blending ($\mathbf{G}_{xb}$) for two tuning methods; subscript x varies between c for Christensen et al. (2012) or v for Vitezica et al. (2011) tuning

Angus and Holstein datasets were 0.091 and 0.073 respectively. The distribution of the difference of diagonal elements when blending first versus tuning first is displayed in Figure 1. It can be observed that the differences are only in the ten-thousandths place. Consequently, it is expected that the average variation of the predictions across scenarios will be small. This is logical because the goal of the employed tuning methods is to equalize the averages of **G** and $\mathbf{A}_{22}$ (Christensen et al., 2012; Vitezica et al., 2011).

3.2 | Genomic relationships

Table 2 shows average estimated realized (genomic) relationships for genotyped animals with respective standard deviation (*SD*) for animals grouped based on pedigree information. Genomic relationships were greater than the expected values, with a *SD* that varied from 0.04 to 0.06. In the Angus dataset, the values were similar across scenarios for self-relationships, parent–offspring, full-sibs, half-sibs and grandparent–grandoffspring. Minuscule differences were observed in the ten-thousandths place (results not shown), with slightly greater values for the tuning method proposed by Vitezica et al. (2011). In the Holstein dataset, average genomic relationships were slightly higher (0.01) when using the previously mentioned tuning method.

3.3 | Genomic predictions

3.3.1 | GEBV

Table 3 shows the statistics for changes in GEBV when changing the order between blending and tuning, for both tuning methods and $\beta = \{0.05, 0.5\}$. As expected,

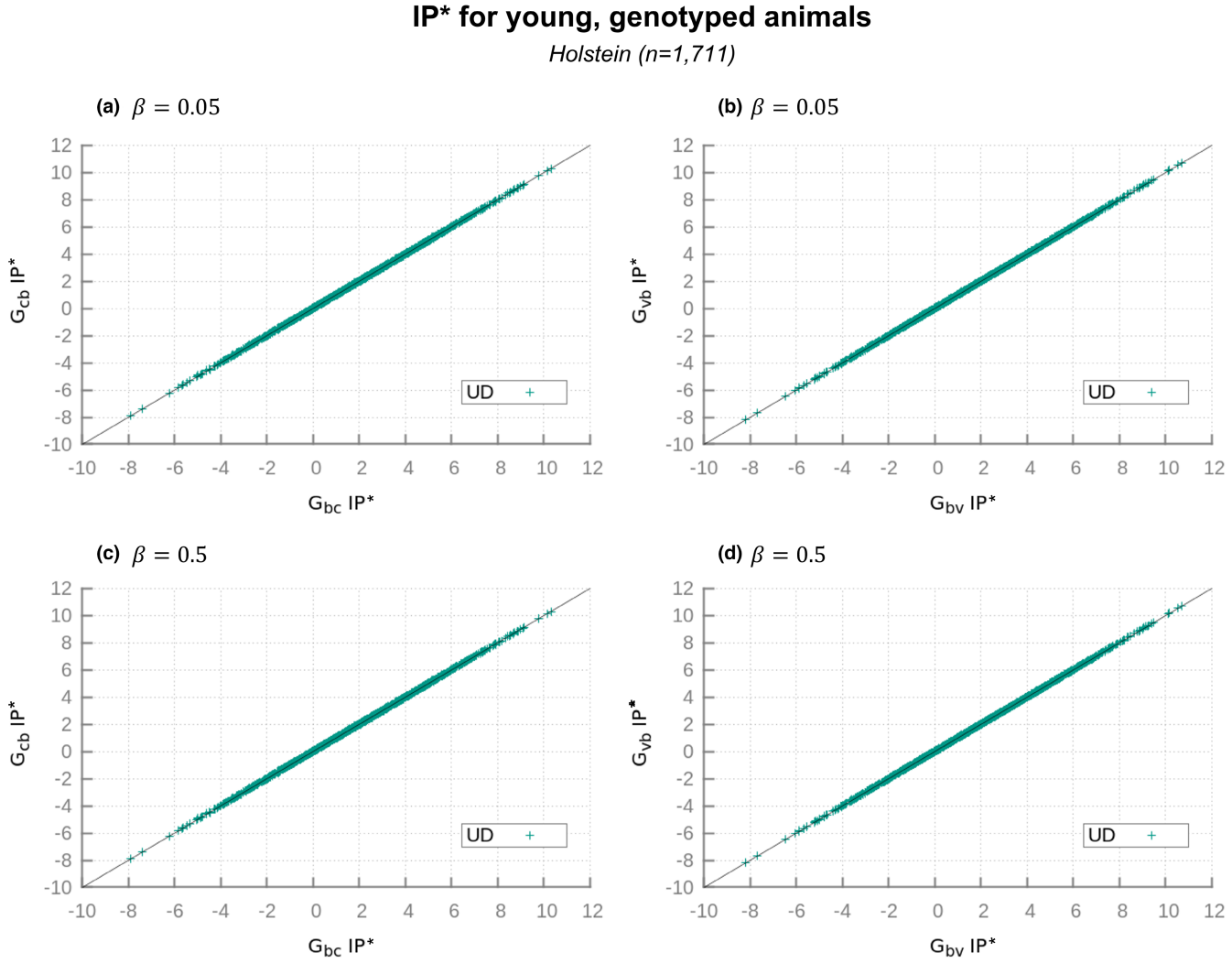

FIGURE 5 Scatter plots of indirect predictions on the genomic base (IP*) for young, genotyped Holstein bulls' udder depth (UD) calculated using different blending parameters (β) and blending before tuning ($\mathbf{G}_{bx}$) against predictions using tuning before blending ($\mathbf{G}_{xb}$) for two tuning methods; subscript x varies between c for Christensen et al. (2012) or v for Vitezica et al. (2011) tuning

changes in GEBV were negligible because the maximum change in the genomic relationship matrix is small. However, changes were greater for β near 0.5 as shown in Appendices B and C. Scatter plots of GEBV when using blending first versus tuning first for the two tuning methods and $\beta = 0.05$ are shown in Figure 2 for Angus and Figure 3 for Holstein. Scatterplots for GEBV using $\beta = 0.5$ are not shown but appear visually identical to Figures 2 and 3 despite the differences shown in Table 3. As expected, in light of Appendices B and C and Table 3, when changing the order between blending and tuning, GEBV were almost identical. For all the animals in the evaluation, the GEBV did not differ considerably when changing the order of blending and tuning for both datasets. The same results were observed for the IP*of young, genotyped animals.

3.3.2 | IP* and SNP effects

Graphs for IP*, 'marker' estimated breeding values on genomic base for the young, genotyped animals are in Figure 4 for Angus and Figure 5 for Holstein. The correlation for young, genotyped IP* when using $\beta = 0.05$ and $\beta = 0.05$ differed across scenarios in the ten thousandth places (not shown), however, when rounded to the thousandths place, the correlations were the same: 0.93, 0.96, 0.95 and 0.94 for UD, BW, WW and PWG respectively. The number of genotyped individuals did not affect these changes. The Holstein data contained over twice the number of genotyped individuals as the Angus dataset, yet no considerable differences were observed for either. Results are consistent even when doubling the number of genotyped individuals; this is expected to hold for any number of genotyped animals.

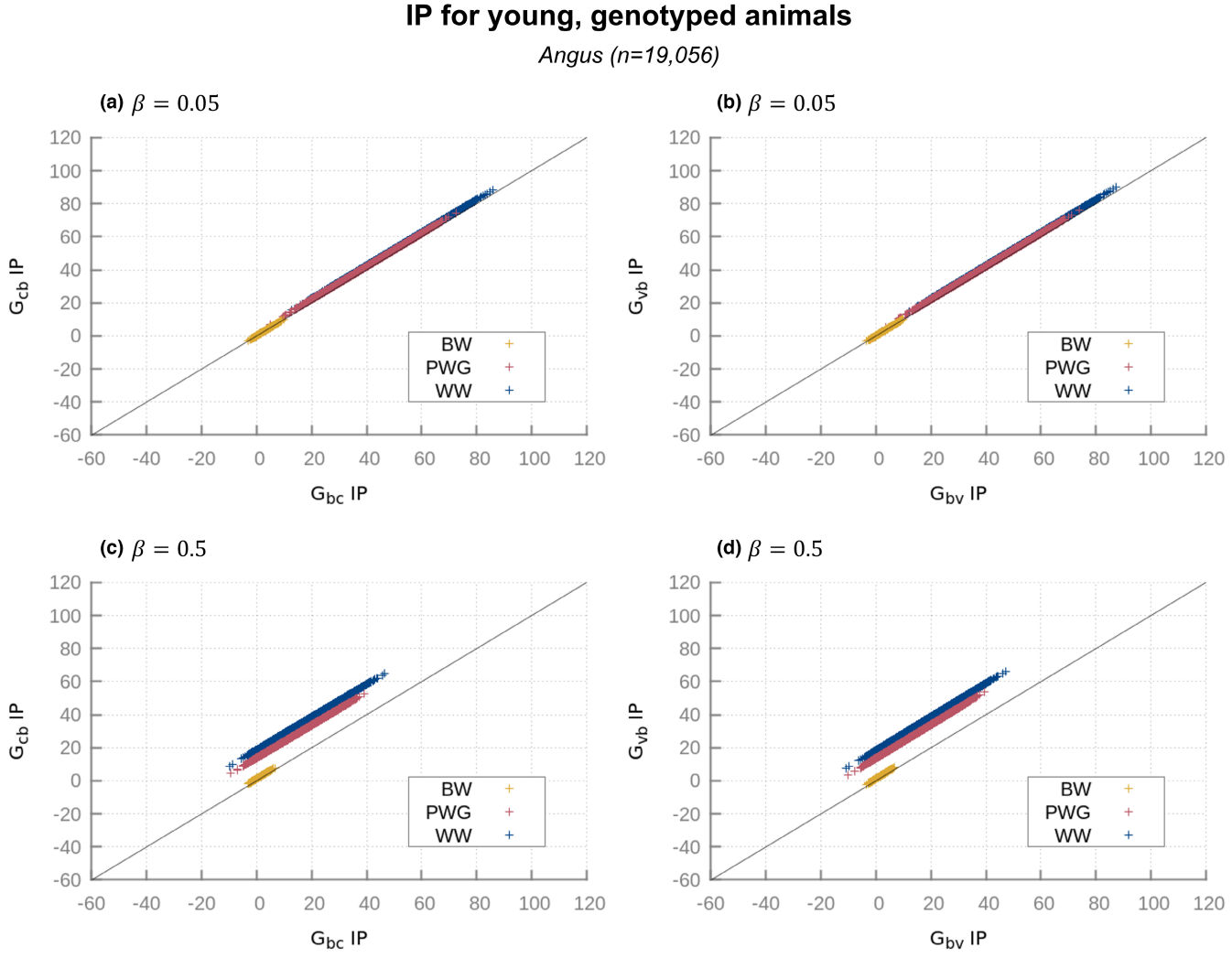

FIGURE 6 Scatter plots of indirect predictions on pedigree scale (IP) young, genotyped Angus animals' birth weight (BW), weaning weight (WW), and post weaning gain (PWG) calculated using different blending parameters (β) and blending before tuning ($\mathbf{G}_{bx}$) against predictions using tuning before blending ($\mathbf{G}_{xb}$) for two tuning methods; subscript x varies between c for Christensen et al. (2012) or v for Vitezica et al. (2011) tuning

IP* are calculated from SNP effects that are backsolved from GEBV; therefore, IP* are a linear function of $\mathbf{G}^{-1}$ (Stranden & Garrick, 2009). Given the small changes in **G** when tuning versus blending first, SNP effects were not expected to differ among scenarios. SNP effects from tuning first and blending first methods and for both tuning methods all had correlations >0.99. For both Holstein and Angus data, SNP effects for UD, BW, WW and PWG had slightly greater variation when using Vitezica et al. (2011) tuning; however, the increase in standard deviation was small (i.e., 0.0003, 0.0006, 0.0030 and 0.0025 respectively) (Figure 5).

3.3.3 | IP

Graphs for IP, 'marker' estimated breeding values on the pedigree scale for the young, genotyped animals are in Figures 6 and 7 for Angus and Holstein respectively. Predictions are shifted by a constant when the order of blending and tuning is changed, and this change is exacerbated when $\beta = 0.5$ is used. This shift is due to the constant $\hat{\mu}$ added to IP*. The $\hat{\mu}$ is a function of the difference between genomic and pedigree bases and a proportion $(1 - \beta)$ of the inverse of the final **G** (after blending and tuning). When **G** is blended before tuning, $\hat{\mu}$ is biased because the blended **G** already contains a portion of $\mathbf{A}_{22}$ and refers to a different genetic base. Tuning $\mathbf{G}_b$ with a residual polygenic effect results in a suboptimal correction when back solving for the difference in the genetic base between the two matrices. The inconsistency was small when using a blending parameter $\beta = 0.05$, but more pronounced when using $\beta = 0.50$.

3.4 | Blending parameter β

When ssGBLUP is applied, **G** is positive definite if observed allele frequencies are used (resulting in $\alpha_v > 0$ or $\alpha_c > 0$) and the blending parameter β is higher than

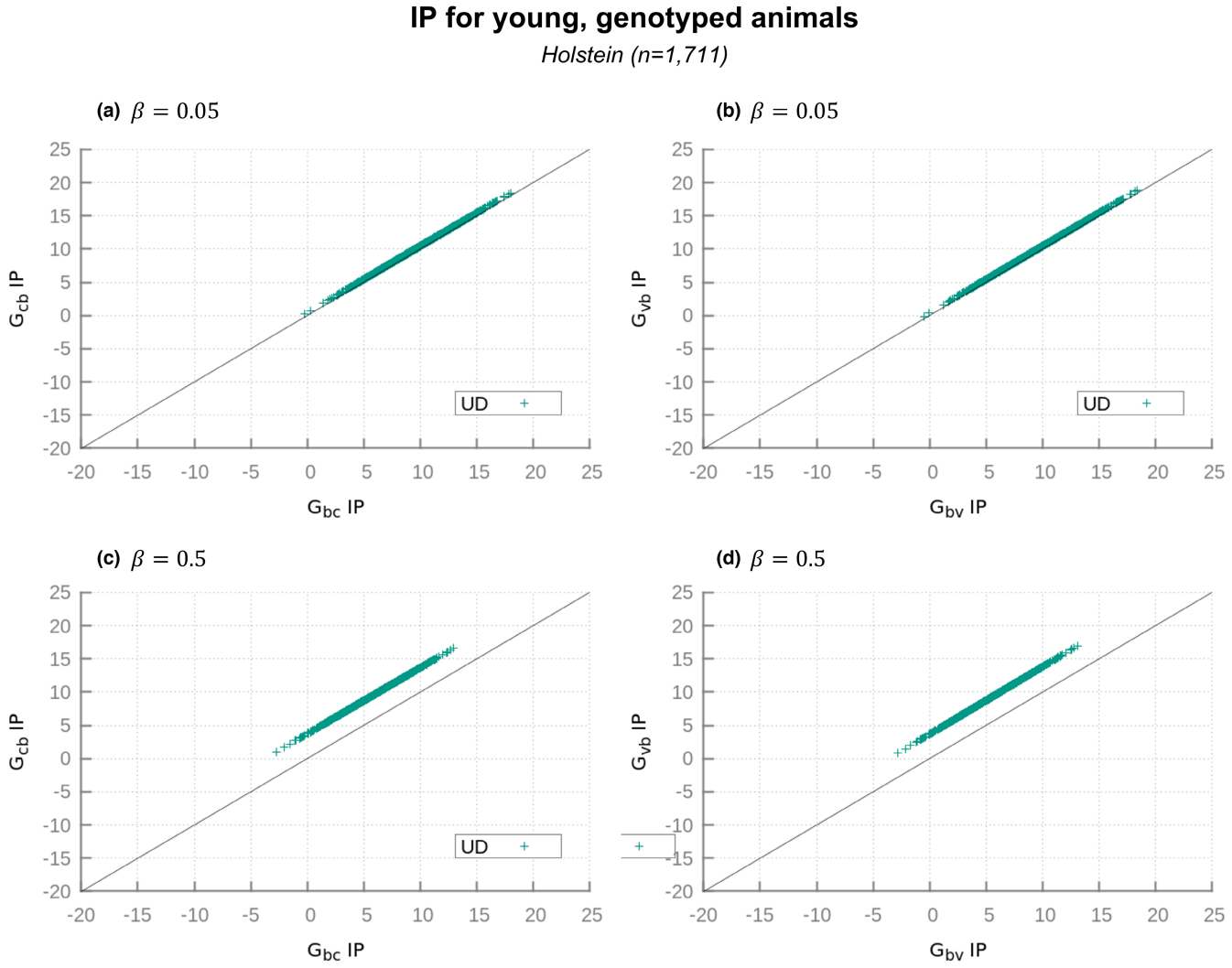

FIGURE 7 Scatter plots of indirect predictions on pedigree scale (IP) young, genotyped Holstein bulls' udder depth (UD) calculated using different blending parameters (β) and blending before tuning ($\mathbf{G}_{bx}$) against predictions using tuning before blending ($\mathbf{G}_{xb}$) for two tuning methods; subscript x varies between c for Christensen et al. (2012) or v for Vitezica et al. (2011) tuning

0, for example, equal to 0.05. The choice for β can be attained by maximum likelihood (Christensen & Lund, 2010) or empirically to minimize bias by cross-validation (Liu et al., 2016; Mäntysaari et al., 2010; McMillan & Swan, 2017). VanRaden (2008) showed that $\beta = SD^2 / \left(SD^2 + 0.125/m\right)$, where m is the number of markers and SD is the standard error of predicting the true fraction of shared DNA; this is the SD in Table 2. Utilizing this formula in the present study and in VanRaden (2008), β would be close to 0.001; however, VanRaden (2008) observed that $\beta = 0.05$ resulted in slightly more reliable predictions and suggested this value could be used in cattle populations. In this study, 5 and 50% of $\mathbf{A_{22}}$ were added to $\mathbf{G}$ in different scenarios and evaluated. An alternative to adding $\mathbf{A_{22}}$ is to add a proportion of the identity matrix to prevent singularity, in other words, to add a small constant to the diagonal values of $\mathbf{G}$. Himmelbauer et al. (2021) compared Mendelian sampling of bull families when a constant of 0.001 and 0.01 was added to the diagonal elements of $\mathbf{G}$. GEBV from both of these scenarios were correlated >0.99, however, when 0.01 was added to $\mathbf{G}$, some traits resulted in bias in Mendelian samplings and more outliers were observed from bull lines with many genotyped progeny. To reduce bias for bulls with many genotyped progeny, the constant added to diagonal elements of $\mathbf{G}$ should be small and used as a means to make $\mathbf{G}$ non-singular.When tuning is needed in ssGBLUP, changing the order of blending and tuning did not create considerable changes in genomic relationships, GEBV and IP*. However, when IP* are adjusted to a pedigree scale for comparison with GEBV, blending first resulted in biased estimates.

4 | CONCLUSIONS

When constructing the realized relationship matrix, scaling $\mathbf{G}$ to $\mathbf{A}_{22}$ prior to adding a weighted $\mathbf{G}$ and positive-definite $\mathbf{A}_{22}$ is theoretically more correct and robust. For the datasets used in this study, negligible differences were observed in GEBV for all animals and IP on genomic base for young, genotyped animals when changing the order between blending and tuning given the small proportion of residual polygenic effect assumed for the models. However, blending prior to tuning for IP adjusted to the pedigree scale results in biased estimates. Mathematical expressions were derived, which indicate that the blending parameter that maximizes these differences when changing the order of blending and tuning is 0.5. Although the changes were minimal or non-existent in our study, applying tuning before blending avoids theoretical inconsistencies when computing the difference in genetic base.

ACKNOWLEDGEMENTS

The authors thank American Angus Association (Saint Joseph, MO), Holstein Association USA, Inc. (Brattleboro, VT) and the Council on Dairy Cattle Breeding (CDCB; Bowie, MD) for providing access to data for the purpose of this analysis. This study was partially funded by Agriculture and Food Research Initiative Competitive Grant no. 2020-67015-31030 from the U.S. Department of Agriculture, Australian Government's National Institute of Food and Agriculture (Washington, DC).

CONFLICT OF INTEREST

A.L.S.G discloses that he is now employed by the American Angus Association that provides genetic evaluations of Angus cattle. This does not represent a conflict of interest in this study. The remaining authors declare no real or perceived conflicts of interest.

AUTHOR CONTRIBUTIONS

DL conceived and planned the research. TMM performed the analyses and summarized the results. DL and ALSG helped perform analyses. Theory refined by all authors. MB derived appendices. TMM drafted and completed the manuscript. All authors read, contributed to and approved the final manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from American Angus Association and Holstein Association USA. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the author(s) with the permission of American Angus Association and Holstein Association USA.

ORCID

Taylor M. McWhorter https://orcid.org/0000-0002-6158-5669
Matias Bermann https://orcid.org/0000-0002-5374-0710
Andre L. S. Garcia https://orcid.org/0000-0001-9778-7978
Andrés Legarra https://orcid.org/0000-0001-8893-7620
Ignacio Aguilar https://orcid.org/0000-0002-1038-4752
Ignacy Misztal https://orcid.org/0000-0002-0382-1897
Daniela Lourenco https://orcid.org/0000-0003-3140-1002

REFERENCES

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., & Lawlor, T. J. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science*, *93*, 743–752. https://doi.org/10.3168/jds.2009-2730

Alkhoder, H., & Liu, Z. (2021). Application of a single-step SNP BLUP model to conformation traits of German Holsteins. *Interbull Bulletin*, *56*, 30–40.

Andersen, T., Nielsen, U. S., Aamand, G. P., Mäntysaari, E., Pösö, J., & Rius-Vilarrasa, E. (2021). Using single-step genetic evaluation for type traits in the Nordic countries. *Interbull Bulletin*, *56*, 90–93.

Christensen, O. F. (2012). Compatibility of pedigree-based and marker-based relationship matrices for a single-step genomic evaluation. *Genetics Selection Evolution*, *44*, 37. https://doi.org/10.1186/1297-9686-44-37

Christensen, O. F., & Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. *Genetics, Selection, Evolution*, *42*, 2. https://doi.org/10.1186/1297-9686-42-2

Christensen, O. F., Madsen, P., Nielsen, B., Ostersen, T., & Su, G. (2012). Single-step methods for genomic evaluation in pigs. *Animal*, *6*(10), 1565–1571. https://doi.org/10.1017/S1751731112000742

Forneris, N. S., Steibel, J. P., Legarra, A., Vitezica, Z. G., Bates, R. O., Ernst, C. W., Basso, A. L., & Cantet, R. J. C. (2016). A comparison of methods to estimate genomic relationships using pedigree markers in livestock populations. *Journal of Animal Breeding and Genetics*, *133*, 452–462. https://doi.org/10.1111/jbg.12217

Garcia-Baccino, C. A., Legarra, A., Christensen, O. F., Misztal, I., Pocrnic, I., Vitezica, Z. G., & Cantet, R. J. C. (2017). Metafounders are related to F_{st} fixation indices and reduce bias in single-step genomic evaluations. *Genetics, Selection, Evolution*, *49*, 34. https://doi.org/10.1186/s12711-017-0309-2

Gengler, N., Mayeres, P., & Szydlowski, M. (2007). A simple method to approximate gene content in large pedigree populations: Application to the myostatin gene in dual-purpose Belgian blue cattle. *Animal*, *1*, 21–28. https://doi.org/10.1017/S1751731107392628

Gohberg, I., Goldberg, S., & Krupnik, N. (2000). Trace class and Hilbert-Schmidt operators in Hilbert space. In *Traces and determinants of linear operators. Operator theory advances and applications* (Vol. *116*, p. 63). Birkhäuser. https://doi.org/10.1007/978-3-0348-8401-3

Golub, G. H., & Van Loan, C. F. (2013). Unsymmetric eigenvalue problems. In *Matrix computations* (4th ed., p. 357). John Hopkins University Press.

Harville, D. A. (2008). Eigenvalues and eigenvectors. In *Matrix algebra from a statistician's perspective* (p. 539). Springer. https://doi.org/10.1007/0-387-22677-X_21

Himmelbauer, J., Schwarzenbacher, H., & Fuerst, C. (2021). Implementation of single-step evaluations for fitness traits in the German and Austrian Fleckvieh and Brown swiss populations. *Interbull Bulletin*, *56*, 82–89.

Horn, R. A., & Johnson, C. R. (2013). Norms for vectors and matrices. In *Matrix analysis* (2nd ed., p. 321). Cambridge University Press.

Interbull Centre. (2021). *The national genomic evaluation forms provided by countries*. https://interbull.org/ib/nationalgenoforms

Legarra, A., Aguilar, I., & Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, *92*, 4656–4663. https://doi.org/10.3168/jds.2009-2061

Legarra, A., Christensen, O. F., Vitezica, Z. G., Aguilar, I., & Misztal, I. (2015). Ancestral relationships using metafounders: Finite ancestral populations and across population relationships. *Genetics*, *200*, 455–468. https://doi.org/10.1534/genetics.115.177014

Legarra, A., Lourenco, D. A. L., & Vitezica, Z. G. (2022). *Bases for genomic predictions*. http://genoweb.toulouse.inra.fr/~alegarra/GSIP.pdf

Liu, Z., Goddard, M. E., Hayes, B. J., Reinhardt, F., & Reents, R. (2016). Technical note: Equivalent genomic models with a residual polygenic effect. *Journal of Dairy Science*, *99*, 2016–2025. https://doi.org/10.3168/jds.2015-10394

Lourenco, D., Legarra, A., Tsuruta, S., Masuda, Y., Aguilar, I., & Misztal, I. (2020). Single-step genomic evaluations from theory to practice: Using SNP chips and sequence data in BLUPF90. *Genes*, *11*, 790. https://doi.org/10.3390/genes11070790

Lourenco, D. A. L., Tsuruta, S., Fragomeni, B. O., Masuda, Y., Aguilar, I., Legarra, A., Bertrand, J. K., Amen, T. S., Wang, L., Moser, D. W., & Misztal, I. (2015). Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus. *Journal of Animal Science*, *93*(6), 2653–2662. https://doi.org/10.2527/jas.2014-8836

Mäntysaari, E. A., Zengting, L., & VanRaden, P. (2010). Interbull validation test for genomic evaluations. *Interbull Bulletin*, *41*, 17–22.

Masuda, Y., Misztal, I., Tsuruta, S., Legarra, A., Aguilar, I., Lourenco, D. A. L., Fragomeni, B. O., & Lawlor, T. J. (2016). Implementation of genomic recursions in single-step genomic best linear unbiased predictor for US Holsteins with a large number of genotyped animals. *Journal of Dairy Science*, *99*, 1968–1974. https://doi.org/10.3168/jds.2015-10540

McMillan, A. J., & Swan, A. A. (2017). Weighting of genomic and pedigree relationships in single step evaluation of carcass traits in Australian sheep. *Proceedings of the 22nd Association for the Advancement of Animal Breeding and Genetics, Townsville, Australia*, *22*, 557–560.

Misztal, I., Legarra, A., & Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science*, *92*, 4648–4655. https://doi.org/10.3168/jds.2009-2064

Misztal, I. S. Tsuruta, D. A. L. Lourenco, Y. Masuda, I. Aguilar, A. Le, and Z. Vitezica. 2014. *Manual for BLUPF90 family of programs*. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

Misztal, I., Vitezica, Z. G., Legarra, A., Aguilar, I., & Swan, A. A. (2013). Unknown-parent groups in single-step genomic evaluation. *Journal of Animal Breeding and Genetics*, *130*, 252–258. https://doi.org/10.1111/jbg.12025

Pocrnic, I. (2017). *Genomic selection and impact of limited dimensionality of genomic information* [Doctoral dissertation]. University of Georgia. https://getd.libs.uga.edu/pdfs/pocrnic_ivan_201712_phd.pdf

Strandén, I., & Garrick, D. J. (2009). Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *Journal of Dairy Science*, *92*, 2971–2975. https://doi.org/10.3168/jds.2008-1929

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*, 4414–4423. https://doi.org/10.3168/jds.2007-0980

Vitezica, Z. G., Aguilar, I., Misztal, I., & Legarra, A. (2011). Bias in genomic prediction for populations under selection.

Genetical Research, *93*, 357–366. https://doi.org/10.1017/S001667231100022X

Wiggans, G. R., Cooper, T. A., VanRaden, P. M., Van Tassell, C. P., Bickhart, D. M., & Sonstegard, T. S. (2016). Increasing the number of single nucleotide polymorphisms used in genomic evaluation of dairy cattle. *Journal of Dairy Science*, *99*, 4504–4511. https://doi.org/10.3168/jds.2015-10456

Wiggans, G. R., Sonstegard, T. S., Vanraden, P. M., Matukumalli, L. K., Schnabel, R. D., & Taylor, J. F. (2009). Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. *Journal of Dairy Science*, *92*, 3431–3436. https://doi.org/10.3168/jds.2008-1758

How to cite this article: McWhorter, T. M., Bermann, M., Garcia, A. L. S., Legarra, A., Aguilar, I., Misztal, I., & Lourenco, D. (2022). Implication of the order of blending and tuning when computing the genomic relationship matrix in single-step GBLUP. *Journal of Animal Breeding and Genetics*, *00*, 1–19. https://doi.org/10.1111/jbg.12734

APPENDIX A

CONSTRUCTION OF GENOMIC RELATIONSHIP MATRICES USING A DIFFERENT ORDER OF TUNING AND BLENDING AND TWO TUNING METHODS

The first set of GEBV ($\mathbf{G}_{bc}$) was obtained by blending **G** first and then tuning it following the method of Christensen et al. (2012):

Writing the value of $t_c = \frac{(\overline{diag\mathbf{A}_{22}} - \overline{\mathbf{A}_{22}})}{(\overline{diag\mathbf{G}} - \overline{\mathbf{G}})}$ as $t_c = \frac{f(\mathbf{A}_{22})}{f(\mathbf{G})}$ where $f: \mathbb{R}^{nxn} \rightarrow \mathbb{R}$ such that $f(\mathbf{M}) = \text{tr}(\mathbf{M}) - n^{-1}\mathbf{1}'\mathbf{M}\mathbf{1}$ where n is the number of genotyped individuals, the resulting $\mathbf{G}_{bc}$ matrix used for the calculation of the first set of GEBV was as follows:

$$\mathbf{G}_{bc} = \left[(1-\beta)\mathbf{G} + \beta\mathbf{A}_{22}\right]\frac{f(\mathbf{A}_{22})}{(1-\beta)f(\mathbf{G}) + \beta f(\mathbf{A}_{22})} + k\frac{1-\beta}{(1-\beta)f(\mathbf{G}) + \beta f(\mathbf{A}_{22})}\mathbf{1}\mathbf{1}' \tag{A1}$$

where $k = \frac{\text{tr}(\mathbf{G})\mathbf{1}'\mathbf{A}_{22}\mathbf{1} - \text{tr}(\mathbf{A}_{22})\mathbf{1}'\mathbf{G}\mathbf{1}}{n^2}$.

The second set of GEBV ($\mathbf{G}_{cb}$) was calculated by tuning **G** with the method of Christensen et al. (2012) first and then blending. In this case, the obtained **G** matrix was of the form:

$$\mathbf{G}_{cb} = (1-\beta)\left[\mathbf{G}\frac{f(\mathbf{A}_{22})}{f(\mathbf{G})} + \frac{k}{f(\mathbf{G})}\mathbf{1}\mathbf{1}'\right] + \beta\mathbf{A}_{22} \tag{A2}$$

The other two sets, $\mathbf{G}_{bv}$ and $\mathbf{G}_{vb}$, were analogous to the first two but used the tuning method of Vitezica et al. (2011) instead of the method of Christensen et al. (2012), giving the matrices $\mathbf{G}_{bv}$ and $\mathbf{G}_{vb}$ respectively:

$$\mathbf{G}_{bv} = \left[1 - (1-\beta)h\right]\left[(1-\beta)\mathbf{G} + \beta\mathbf{A}_{22}\right] + (1-\beta)2h\mathbf{1}\mathbf{1}' \tag{A3}$$

and

$$\mathbf{G}_{vb} = (1-\beta)\left[(1-h)\mathbf{G} + 2h\mathbf{1}\mathbf{1}'\right] + \beta\mathbf{A}_{22} \tag{A4}$$

where $h = \frac{\mathbf{1}'(\mathbf{A}_{22} - \mathbf{G})\mathbf{1}}{2n^2}$, half the average difference among the two matrices.

APPENDIX B

DERIVATION FOR MAXIMIZING DIFFERENCES BETWEEN SCENARIOS USING CHRISTENSEN ET AL. (2012) TUNING

Shown are the derivations for the proportion of residual polygenic effect (β) that maximizes the differences between genomic relationship matrices for the tuning method of Christensen et al. (2012), and an upper bound for the maximum change in the genomic relationship matrix when changing the order of blending and tuning.

β that maximizes the metric between $\mathbf{G}_{bc}$ and $\mathbf{G}_{cb}$

Keeping the same notation as in (A1) and (A2), that difference is evaluated in terms of the metric defined as follows:

$$d_1(\mathbf{G}_{bc}, \mathbf{G}_{cb}) := \max_{i,j}\left|\mathbf{G}_{bc_{ij}} - \mathbf{G}_{cb_{ij}}\right| = \max_{i,j}\left|(1-\beta)g_{ij}\left(\frac{f(\mathbf{A}_{22})}{(1-\beta)f(\mathbf{G}) + \beta f(\mathbf{A}_{22})} - \frac{f(\mathbf{A}_{22})}{f(\mathbf{G})}\right) + \beta\, a_{22_{ij}}\left(\frac{f(\mathbf{A}_{22})}{(1-\beta)f(\mathbf{G}) + \beta f(\mathbf{A}_{22})} - 1\right) + (1-\beta)k\left(\frac{1}{(1-\beta)f(\mathbf{G}) + \beta f(\mathbf{A}_{22})} - \frac{1}{f(\mathbf{G})}\right)\right| \tag{B1}$$

It is important to remark that $\max_{i,j}$ refers to a particular choice of g_{ij} and $a_{22_{ij}}$. g_{ij} refers to the genomic relationship between animal i and animal j and $g(\beta)$ is the beta that maximizes the metric between $\mathbf{G}_{bc}$ and $\mathbf{G}_{cb}$. Because g_{ij} and $a_{22_{ij}}$ are not functions of β, it is valid to proceed with the calculations while ignoring the maximum at the front of the expression. Therefore, to find the value of β that maximizes $d_1\left(\mathbf{G}_{bc}-\mathbf{G}_{cb}\right)$, it is necessary to:

$$\operatorname{argmax}\left\{d_1\left(\mathbf{G}_{bc}-\mathbf{G}_{cb}\right)=g(\beta) \text{ subject to } 0<\beta<1\right\} \tag{B2}$$

Let $g_{bc}(\beta,x,y)=\beta-x^2$ and $g_{cb}(\beta,x,y)=1-\beta-y^2$, then (B2) is equivalent to:

$$\operatorname{argmax}\left\{g(\beta) \text{ subject to } g_{bc}(\beta,x,y)=0 \text{ and } g_{cb}(\beta,x,y)=0\right\} \tag{B3}$$

To solve (B3) the following system must be solved:

$$\begin{cases} \nabla g=\theta_1\nabla g_{bc}+\theta_2\nabla g_{cb} \\ \beta-x^2=0 \\ 1-\beta-y^2=0 \end{cases} \tag{B4}$$

where ∇ is the gradient, and θ_1 and θ_2 are Lagrange Multipliers. Taking the derivative of (B1) with respect to β:

$$\frac{\partial d_1(\mathbf{G}_{bc},\mathbf{G}_{cb})}{\partial\beta}=\pm\left(g_{ij}\left[\frac{f(\mathbf{A}_{22})\left[f(\mathbf{G})-f(\mathbf{A}_{22})\right]}{\left[(1-\beta)f(\mathbf{G})+\beta f(\mathbf{A}_{22})\right]^2}-\frac{f(\mathbf{A}_{22})f(\mathbf{G})}{\left[(1-\beta)f(\mathbf{G})+\beta f(\mathbf{A}_{22})\right]^2}+\frac{f(\mathbf{A}_{22})}{f(\mathbf{G})}\right]+a_{22_{ij}}\left[\frac{f(\mathbf{A}_{22})f(\mathbf{G})}{\left[(1-\beta)f(\mathbf{G})+\beta f(\mathbf{A}_{22})\right]^2}-1\right]+k\left[\frac{1}{f(\mathbf{G})}-\frac{f(\mathbf{A}_{22})}{\left[(1-\beta)f(\mathbf{G})+\beta f(\mathbf{A}_{22})\right]^2}\right]\right) \tag{B5}$$

The other two coordinates of ∇g are equal to zero. Also, $\nabla g_{bc}=\langle 1,-2x,0\rangle$ and $\nabla g_{cb}=\langle -1,0,-2y\rangle$. Because the inequality in (B2) is strict, $\theta_1=\theta_2=0$. Therefore, the system (B4) reduces to:

$$\begin{cases} \nabla g=0 \\ \beta-x^2=0 \\ 1-\beta-y^2=0 \end{cases} \tag{B6}$$

The first equation in (B6) is solved by setting the derivative in (B5) to zero and rearranging the terms:

$$a_{22_{ij}}f\left(\mathbf{A}_{22}\right)f(\mathbf{G})-g_{ij}f\left(\mathbf{A}_{22}\right)^2-f\left(\mathbf{A}_{22}\right)k=\left(a_{22_{ij}}-g_{ij}\frac{f\left(\mathbf{A}_{22}\right)}{f(\mathbf{G})}-\frac{k}{f(\mathbf{G})}\right)\left[(1-\beta)f(\mathbf{G})+\beta f\left(\mathbf{A}_{22}\right)\right]^2 \tag{B7}$$

The equality in (1.7) will be reached when:

$$\begin{aligned} \left[(1-\beta)f(\mathbf{G})+\beta f\left(\mathbf{A}_{22}\right)\right]^2&=f(\mathbf{G})f\left(\mathbf{A}_{22}\right) \\ \beta&=\frac{\sqrt{f(\mathbf{G})f\left(\mathbf{A}_{22}\right)}-f(\mathbf{G})}{f\left(\mathbf{A}_{22}\right)-f(\mathbf{G})} \end{aligned} \tag{B8}$$

The solutions for the auxiliary variables x and y are as follows:

$$\begin{cases} x=\sqrt{\dfrac{\sqrt{f(\mathbf{G})f\left(\mathbf{A}_{22}\right)}-f(\mathbf{G})}{f\left(\mathbf{A}_{22}\right)-f(\mathbf{G})}} \\ y=\dfrac{\sqrt{f(\mathbf{G})f\left(\mathbf{A}_{22}\right)}-f\left(\mathbf{A}_{22}\right)}{f\left(\mathbf{A}_{22}\right)-f(\mathbf{G})} \end{cases} \tag{B9}$$

For a positive semi-definite matrix $\mathbf{M}$, $f(\mathbf{M}) \geq 0$. The argument is as follows. Let λ_{max} be the largest eigenvalue of $\mathbf{M}$; then $f(\mathbf{M}) \geq 0$ because of the finiteness of $\text{tr}(\mathbf{M})$ (Gohberg et al., 2000, p. 63) and that (i) $n^{-1}\mathbf{1}'\mathbf{M}\mathbf{1} \leq \lambda_{max}$ (Harville, 2008, p. 539), (ii) $\lambda_{max} < \infty$ (Golub & Van Loan, 2013; pp. 357) and (iii) $\text{tr}(\mathbf{M}) \geq \lambda_{max}$. The inequality is strict when $\mathbf{M}$ has at least two non-zero eigenvalues.

Because $f(\mathbf{A}_{22}), f(\mathbf{G}) > 0$, it worth noting that (B8) guarantees $\beta \in (0,1)$ and that when $f(\mathbf{A}_{22}) \approx f(\mathbf{G})$, $\beta \to 0.5$. Finally, to check whether (B8) represents a local maximum of $d_1(\mathbf{G}_{bc},\mathbf{G}_{cb})$, the second derivative of $d_1(\mathbf{G}_{bc},\mathbf{G}_{cb})$ with respect to β is as follows:

$$\frac{\partial d_1(\mathbf{G}_{bc},\mathbf{G}_{cb})}{\partial \beta^2} = \pm\left(g_{ij}\left[\frac{2f(\mathbf{A}_{22})^2\left[f(\mathbf{A}_{22})-f(\mathbf{G})\right]}{\left[(1-\beta)f(\mathbf{G})+\beta f(\mathbf{A}_{22})\right]^3}\right] + a_{22_{ij}}\left[\frac{2f(\mathbf{A}_{22})\left[f(\mathbf{G})-f(\mathbf{A}_{22})\right]}{\left[(1-\beta)f(\mathbf{G})+\beta f(\mathbf{A}_{22})\right]^3}\right] + k\left[\frac{2f(\mathbf{A}_{22})\left[f(\mathbf{A}_{22})-f(\mathbf{G})\right]}{\left[(1-\beta)f(\mathbf{G})+\beta f(\mathbf{A}_{22})\right]^3}\right]\right) \quad \text{(B10)}$$

Evaluating it at (B8) gives:

$$\frac{\partial d_1(\mathbf{G}_{bc},\mathbf{G}_{cb})}{\partial \beta^2}\left(\beta = \frac{\sqrt{f(\mathbf{G})f(\mathbf{A}_{22})}-f(\mathbf{G})}{f(\mathbf{A}_{22})-f(\mathbf{G})}\right) = \pm\frac{2f(\mathbf{A}_{22})\left[f(\mathbf{A}_{22})-f(\mathbf{G})\right]}{\left[f(\mathbf{G})f(\mathbf{A}_{22})\right]^{3/2}}\left(f(\mathbf{A}_{22})g_{ij} - a_{22_{ij}} + k\right) \quad \text{(B11)}$$

The last term is positive because it is dominated by the positive constants $f(\mathbf{A}_{22})g_{ij}$ and k. Then, assuming $g_{ij} > a_{22_{ij}}$ and given that $k > 0$, it can be checked that if $f(\mathbf{A}_{22}) > f(\mathbf{G})$ the plus-minus sign is negative, and that if $f(\mathbf{A}_{22}) < f(\mathbf{G})$, the plus-minus sign is positive. Therefore, by the second derivative criterion, (B8) maximizes (B1).

Upper bound for $d_1(\mathbf{G}_{bc},\mathbf{G}_{cb},)$

After applying triangle inequality and re-arranging terms, (B1) results in:

$$d_1(\mathbf{G}_{bc},\mathbf{G}_{cb}) \leq \beta(1-\beta)\max_{i,j}\left|g_{ij}\right|f(\mathbf{A}_{22})\left|\frac{f(\mathbf{G})-f(\mathbf{A}_{22})}{f(\mathbf{G})\left[(1-\beta)f(\mathbf{G})+\beta f(\mathbf{A}_{22})\right]}\right| + \beta(1-\beta)\max_{i,j}a_{22_{ij}}\left|\frac{f(\mathbf{A}_{22})-f(\mathbf{G})}{(1-\beta)f(\mathbf{G})+\beta f(\mathbf{A}_{22})}\right| + \beta(1-\beta)k\left|\frac{f(\mathbf{G})-f(\mathbf{A}_{22})}{f(\mathbf{G})\left[(1-\beta)f(\mathbf{G})+\beta f(\mathbf{A}_{22})\right]}\right| \quad \text{(B12)}$$

Given that $k \leq n^{-2}\,\mathbf{1}'\mathbf{A}_{22}\mathbf{1}f(\mathbf{G})$, (B12) is bounded above by the following:

$$d_1(\mathbf{G}_{bc},\mathbf{G}_{cb}) \leq \beta(1-\beta)\left|\frac{f(\mathbf{G})-f(\mathbf{A}_{22})}{(1-\beta)f(\mathbf{G})+\beta f(\mathbf{A}_{22})}\right|\left(\max_{i,j}\left|g_{ij}\right|\frac{f(\mathbf{A}_{22})}{f(\mathbf{G})} + \max_{i,j}a_{22_{ij}} + n^{-2}\,\mathbf{1}'\mathbf{A}_{22}\mathbf{1}\right) \quad \text{(B13)}$$

APPENDIX C

DERIVATION FOR MAXIMIZING DIFFERENCES BETWEEN SCENARIOS USING VITEZICA ET AL. (2011) TUNING

Shown are the derivations for the proportion of residual polygenic effect (β) that maximizes the differences between genomic relationship matrices for the tuning method of Vitezica et al. (2011), and an upper bound for the maximum change in the genomic relationship matrix when changing the order of blending and tuning. In this Appendix, the difference between the matrices is evaluated in terms of the metric induced by the Frobenius norm (Horn & Johnson, 2013, p. 321).

β that maximizes the metric between $\mathbf{G}_{bv}$ and $\mathbf{G}_{vb}$

The metric between $\mathbf{G}_{bv}$ and $\mathbf{G}_{vb}$ is defined as follows:

$$d_2(\mathbf{G}_{bv}-\mathbf{G}_{vb}) := \|\mathbf{G}_{bv}-\mathbf{G}_{vb}\|_F^2 = \text{tr}\left(\left[(1-\beta)\beta h(\mathbf{G}+\mathbf{A}_{22})\right]'\left[(1-\beta)\beta h(\mathbf{G}+\mathbf{A}_{22})\right]\right) = \beta^4h^2\|\mathbf{G}+\mathbf{A}_{22}\|_F^2 - 2\beta^3h^2\|\mathbf{G}+\mathbf{A}_{22}\|_F^2 + \beta^2h^2\|\mathbf{G}+\mathbf{A}_{22}\|_F^2 = \beta^4c - 2\beta^3c + \beta^2c = g(\beta) \quad \text{(C1)}$$

where $c = h^2\left\|\mathbf{G}+\mathbf{A}_{22}\right\|_F^2 \geq 0$. Thus, $d_2\left(\mathbf{G}_{bv}-\mathbf{G}_{vb}\right)$ is a polynomial of β. To find the value of β that maximizes $d_2\left(\mathbf{G}_{bv}-\mathbf{G}_{vb}\right)$, it is necessary to:

$$\operatorname{argmax}\left\{d_2\left(\mathbf{G}_{bv}-\mathbf{G}_{vb}\right)=g(\beta) \text{ subject to } 0<\beta<1\right\} \tag{C2}$$

Let $g_{bv}(\beta,x,y)=\beta-x^2$ and $g_{vb}(\beta,x,y)=1-\beta-y^2$, then (C2) is equivalent to:

$$\operatorname{argmax}\left\{g(\beta) \text{ subject to } g_{bv}(\beta,x,y)=0 \text{ and } g_{vb}(\beta,x,y)=0\right\} \tag{C3}$$

To solve (C3) the following system must be solved:

$$\begin{cases} \nabla g=\theta_1\nabla g_{bv}+\theta_2\nabla g_{vb} \\ \beta-x^2=0 \\ 1-\beta-y^2=0 \end{cases} \tag{C4}$$

where ∇ is the gradient, and θ_1 and θ_2 are Lagrange Multipliers. The gradients of the three functions over β, x and y are as follows:

$$\begin{gathered} \nabla g=4\beta^3c-6\beta^2c+2\beta c,0,0 \\ \nabla g_{bv}=1,-2x,0 \\ \nabla g_{vb}=-1,0,-2y \end{gathered} \tag{C5}$$

Therefore, the following system needs to be solved:

$$\begin{cases} 2\beta^3-3\beta^2+\beta-\dfrac{\theta_1-\theta_2}{2c}=0 \\ \beta-x^2=0 \\ 1-\beta-y^2=0 \\ -2\theta_1x=0 \\ -2\theta_2y=0 \end{cases} \tag{C6}$$

Because the inequality in (C2) is strict, $\theta_1=\theta_2=0$. Therefore, $x=y=\sqrt{0.5}$, and $\beta=0.5$. Finally, the criterion of the second derivative of $g(\beta)$ is used to assess if a local maximum is located at $\beta=0.5$. Hence:

$$\begin{gathered} g(\beta)=\beta^4c-2\beta^3c+\beta^2c \\ g'(\beta)=4\beta^3c-6\beta^2c+2\beta c \\ g''(\beta)=12\beta^2c-12\beta c+2c \end{gathered} \tag{C7}$$

Then, $g''(0.5)=-c$, and because c is positive, $\beta=0.5$ maximizes $d_2\left(\mathbf{G}_{bv}-\mathbf{G}_{vb}\right)$.

Value for $\max_{i,j}\left|\mathbf{G}_{bv_{ij}}-\mathbf{G}_{vb_{ij}}\right|$

Given:

$$\max_{i,j}\left|\mathbf{G}_{bv_{ij}}-\mathbf{G}_{vb_{ij}}\right|=\max_{i,j}\left|(1-\beta)\beta h\left(g_{ij}+a_{22_{ij}}\right)\right|=\left(\beta-\beta^2\right)|h|\max_{i,j}g_{ij}+a_{22_{ij}}=\left(\beta-\beta^2\right)\left|Avg\left(\mathbf{G}-\mathbf{A}_{22}\right)\right|\left(1+F_{max}\right) \tag{C8}$$

where F_{max} is the maximum average of the genomic and pedigree inbreeding coefficients among genotyped animals.