Sparse single-step genomic BLUP in crossbreeding schemes^{1,2}

Jérémie Vandenplas,³ Mario P. L. Calus, and Jan ten Napel

Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, 6700 AH Wageningen, The Netherlands

ABSTRACT: The algorithm for proven and young animals (APY) efficiently computes an approximated inverse of the genomic relationship matrix, by dividing genotyped animals in the so-called core and noncore animals. The APY leads to computationally feasible single-step genomic Best Linear Unbiased Prediction (ssGBLUP) with a large number of genotyped animals and was successfully applied to real single-breed or line datasets. This study aimed to assess the quality of genomic estimated breeding values (GEBV) when using the APY (GEBV_{APY}), in comparison to GEBV when using the directly inverted genomic relationship matrix ($\text{GEBV}_{\text{DIRECT}}$), for situations based on crossbreeding schemes, including F1 and

F2 crosses, such as the ones for pigs and chickens. Based on simulations of a 3-way crossbreeding program, we compared different approximated inverses of a genomic relationship matrix, by varying the size and the composition of the core group. We showed that GEBV_{APY} were accurate approximations of $\text{GEBV}_{\text{DIRECT}}$ for multivariate ssGBLUP involving different breeds and their crosses. GEBV_{APY} as accurate as $\text{GEBV}_{\text{DIRECT}}$ were obtained when the core groups included animals from different breed compositions and when the core groups had a size between the numbers of the largest eigenvalues explaining 98% and 99% of the variation in the raw genomic relationship matrix.

Key words: APY, genomic evaluation, single-step

© The Author(s) 2018. Published by Oxford University Press on behalf of the American Society of Animal Science. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

J. Anim. Sci. 2018.96:2060–2073

doi: 10.1093/jas/sky136

INTRODUCTION

Single-step genomic Best Linear Unbiased Prediction (ssGBLUP) is currently the method of choice to predict genomic breeding values in many species (Legarra et al., 2014). The main reason is that ssGBLUP enables simultaneous use of phenotypes from genotyped and nongenotyped animals by combining genomic and pedigree

relationship matrices. An inconvenience of ssGBLUP is that the inverse of a dense genomic relationship matrix ($\mathbf{G}$) is required, leading to a soft limit of approximately 100,000 genotyped animals for the currently available computers (Misztal et al., 2014).

Recently, Misztal et al. (2014) and Misztal (2016) proposed the so-called algorithm for proven and young animals (APY) to compute an approximated inverse of $\mathbf{G}$ ($\mathbf{G}_{\text{APY}}^{-1}$) for a large number of genotyped animals. The computation of $\mathbf{G}_{\text{APY}}^{-1}$ involves the inversion of a genomic relationship submatrix among a limited number of genotyped animals, called core animals, and the recursive computation of other coefficients for noncore animals. The APY was successfully applied on (large) real datasets with animals originating from a single breed or line (Fragomeni et al., 2015; Lourenco et al., 2015; Masuda et al., 2016; Ostensen et al.,

¹This study was financially supported by the Dutch Ministry of Economic Affairs (TKI-Toeslag; public-private partnership “Breed4Food” code BO-22.04-011-001-ASG-LR-3) and the Breed4Food partners, Cobb Europe, CRV, Hendrix Genetics, and Topigs Norsvin.

²We declare that there are no competing interests in the matter and results covered by this article.

³Corresponding author: jeremie.vandenplas@wur.nl

Received March 1, 2018.

Accepted April 16, 2018.

2016; Pocrnic et al., 2016b; Strandén et al., 2017). However, several livestock production systems, such as the ones for pigs and chickens, are based on well-structured crossbreeding schemes, generating production animals with a specific breed composition. In these cases, the ssGBLUP may include nongenotyped and genotyped animals from different breeds, as well as their crossbred progeny. Using the APY with such datasets is desirable for implementing ssGBLUP in crossbreeding schemes efficiently.

The aim of this study was to assess the quality of genomic estimated breeding values (GEBV) when using $\mathbf{G}_{\text{APY}}^{-1}$, in comparison to GEBV when using the direct inversion of $\mathbf{G}$ ($\mathbf{G}_{\text{direct}}^{-1}$), for situations based on well-structured crossbreeding schemes that include genotyped animals from a few different breeds and their F1 and F2 crosses. Influence of the selection strategy of the core animals and of the number of core animals were also investigated. All analyses were based on simulated data.

MATERIALS AND METHODS

Single-Step Genomic Best Linear Unbiased Prediction

The ssGBLUP method replaces the inverse of the pedigree relationship matrix for all animals ($\mathbf{A}^{-1}$) with the inverse of the combined pedigree–genomic relationship matrix ($\mathbf{H}^{-1}$), defined as follows (Aguilar et al., 2010; Christensen and Lund, 2010):

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

where $\mathbf{A}_{22}$ is the pedigree relationship matrix for the genotyped animals, $\mathbf{G} = (1 - w)\mathbf{G}_a + w\mathbf{A}_{22}$ with $\mathbf{G}_a$ being a genomic relationship matrix adjusted to be on the same scale as $\mathbf{A}_{22}$ and with w being the weight on the pedigree relationship matrix. Several approaches for computing $\mathbf{G}_a$ by adjusting a raw genomic relationship matrix $\mathbf{G}^*$ toward $\mathbf{A}_{22}$ were proposed in the literature (Powell et al., 2010; Vitezica et al., 2011; Christensen, 2012; Lourenco et al., 2016).

Highest computational costs for creating $\mathbf{H}^{-1}$ are the creation and the inversion of the dense matrices $\mathbf{G}$ and $\mathbf{A}_{22}$. Additional computational costs also appear during solving of the mixed model equations due to an increase of nonzero elements in $\mathbf{H}^{-1}$, increasing the number of operations per iteration, for example, of the preconditioned conjugate

gradient used to solve the mixed model equations (Ostersen et al., 2016).

Sparse Inversion of $\mathbf{G}$

The matrix $\mathbf{G}$ can be divided into 4 submatrices as follows:

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{cc} & \mathbf{G}_{cn} \\ \mathbf{G}_{cn}' & \mathbf{G}_{nn} \end{bmatrix}$$

where the subscript c refers to a group of genotyped animals called hereafter “core animals,” and the subscript n refers to a second group of genotyped animals called hereafter “noncore animals.”

Following Misztal (Misztal et al., 2014; Misztal, 2016), the inverse of $\mathbf{G}$, $\mathbf{G}^{-1}$, can be approximated using the APY as follows:

$$\mathbf{G}_{\text{APY}}^{-1} = \begin{bmatrix} \mathbf{G}_{cc}^{-1} + \mathbf{G}_{cc}^{-1}\mathbf{G}_{cn}\mathbf{M}^{-1}\mathbf{G}_{cn}'\mathbf{G}_{cc}^{-1} & -\mathbf{G}_{cc}^{-1}\mathbf{G}_{cn}\mathbf{M}^{-1} \\ -\mathbf{M}^{-1}\mathbf{G}_{cn}'\mathbf{G}_{cc}^{-1} & \mathbf{M}^{-1} \end{bmatrix}$$

where the matrix $\mathbf{M}$ is a diagonal matrix of size of the number of noncore animals and with a diagonal element for the i th noncore animal equal to $\mathbf{M}_{ii} = \text{diag}(\mathbf{G}_{nn_{ii}} - \mathbf{G}_{ci}'\mathbf{G}_{cc}^{-1}\mathbf{G}_{ci})$ with $\mathbf{G}_{ci}$ being the i th column of $\mathbf{G}_{cn}$. It is worth noting that the matrix $\mathbf{M}$ is an approximation of the Schur complement of $\mathbf{G}_{cc}$, that is, $\mathbf{S} = \mathbf{G}_{nn} - \mathbf{G}_{cn}'\mathbf{G}_{cc}^{-1}\mathbf{G}_{cn}$. Replacing $\mathbf{M}$ by $\mathbf{S}$ in the formula of $\mathbf{G}_{\text{APY}}^{-1}$ would lead to the computation of the inverse of $\mathbf{G}$, $\mathbf{G}^{-1}$.

The APY only requires the computation of the submatrices $\mathbf{G}_{cc}$, $\mathbf{G}_{cn}$ and of the diagonal elements of $\mathbf{G}_{nn}$, in addition to the inversion of the submatrix $\mathbf{G}_{cc}$. Thus, the computational costs of the APY are reduced in comparison to the setting up and the direct inversion of $\mathbf{G}$. Also, the memory costs of the APY are reduced because only submatrices, $\mathbf{G}_{cc}$ and $\mathbf{G}_{cn}$, must be stored and the matrix $\mathbf{G}_{\text{APY}}^{-1}$ is sparse thanks to the diagonal matrix $\mathbf{M}^{-1}$.

Simulated Data

Populations. The assessment of the quality of the genomic predictions from a sparse ssGBLUP in crossbreeding schemes was achieved by simulating a 3-way crossbreeding program with random selection (Fig. 1). Simulations of historic, purebred, and crossbred recent populations were performed using the QMSim software (Sargolzaei and Schenkel, 2009). For the historic population, 70 discrete random mating generations (i.e., generations 1 to 70) with a constant size of 18,840 individuals with equal number of individuals from each sex were simulated, followed by 10 generations (i.e., generations 71 to 80) in

Figure 1. Schematic representation of the simulation. The crossbreeding program started at generation 200 (generation numbers in bold). The number of males (M) and females (F) per generation and per breed (A, B, and C), or per cross [BC and A(BC)], is reported within brackets. Blue arrows denote the sires and dams of the next generation; red arrows denote the dams of the next generation; and green arrows denote the sires of the next generation.

which the effective population size was gradually reduced to 390 individuals. The next 20 generations (i.e., generations 81 to 100) were simulated to gradually expand the population size to 18,840. The last generation (i.e., generation 100) included 90 males and 18,750 females. Matings for all generations were based on the random union of gametes, which were randomly sampled from the pools of male and female gametes. To simulate the 3 breed populations (hereafter referred to as breeds A, B, and C), 3 random samples were drawn from the generation 100

of the historic population, each including 30 males and 6,250 females. Subsequently, within each breed, 100 generations (i.e., generations 101 to 200) of random mating were simulated before starting the 3-way crossbreeding program (Fig. 1). In each of the simulated 100 generations of random mating, each female had 1 male and 1 female offspring.

In the second step, a 3-way crossbreeding program was simulated (Fig. 1). Purebred (i.e., A, B, and C) animals that were used as founders of the pedigree (i.e., the first generation of the pedigree)

were from generations 200. For each breed, A, B, and C, the next 9 discrete generations (i.e., generations 201 to 209) of purebred animals were simulated by means of random selection and matings while maintaining a constant size of 30 males and 6,250 females. For mimicking a 3-way crossbreeding program, from the generation 205 until the generation 208, B and C purebred animals were randomly crossed to produce 4 generations (i.e., generations 206 to 209) of F1 animals, that is, 30 BC crossbred males and 6,250 BC crossbred females. These BC crossbred animals were then randomly mated to males from breed A to produce 4 generations (i.e., generations 206 to 209) of F2 animals, called A(BC) crossbred animals. For each generation, 6,280 A(BC) crossbred animals were simulated (Fig. 1). Purebred animals that were used as parents of crossbred animals could also be parents of purebred animals in the next generation. A total of 5 replicates were simulated using the QMSim software.

Genotypes. The genome was simulated using the QMSim software, simultaneously with the simulation of the historic, purebred, and crossbred recent populations. The genome consisted of 18 chromosomes designed to resemble the Sus Scrofa genome with a SNP density that was comparable to that of a 60k SNP chip. The SNP positions were randomized across the genome, and a recurrent mutation rate of 2.5×10^{-5} , as well as 1 mean crossover per 1 Morgan, was assumed. All SNPs that segregated in the last historical generation (i.e., generation 100) and with a minor allele frequency (MAF) higher than or equal to 0.05 were selected and used to simulate the genotypes of the purebred and crossbred animals. In addition to the SNPs, 4,500 QTL were simulated, and their positions were also randomized across the genome. Mutation rate and MAF of the QTL were the same as the ones for the simulated SNPs.

Phenotypes. For all purebred and crossbred animals, phenotypes for the breed composition to which they belonged were simulated under additive gene action using a custom Fortran program. This resulted in 5 traits: 1 trait for each of the

purebred performances A, B and C, and 1 trait for each of the crossbred performances BC and A(BC). Genetic correlations between traits were randomly sampled in the range of 0.2 to 0.8 from a uniform distribution. Simulated genetic correlations between purebred and crossbred traits were in the lowest range of reported values in the literature as reviewed by Wientjes and Calus (2017) (Table 1). Heritabilities (h_i^2) were randomly sampled in the range of 0.2 to 0.4 from a uniform distribution. Residual covariances were set to zero, as they would be in practice, because each animal has a phenotype for 1 of the 5 traits only. The same genetic correlations and heritabilities were used in all replicates and are reported in Table 1.

For each animal and for each of the 5 traits, a true breeding value (TBV) was simulated by summing a polygenic effect and the multiplication of the simulated allele substitution effects with the genotypes of the 4,500 QTL coded as 0, 1, and 2. This genotype multiplication allowed different genetic levels across breeds for the same trait because QTL allele frequencies differ across breeds. For each trait, the polygenic effect of each individual was equal to the sum of the average of polygenic effects of the parent and a Mendelian sampling term. The Mendelian sampling terms for the 5 traits were sampled from a multinormal distribution with means of 0 and variances equal to the Mendelian sampling variances (Mrode, 2005). Correlations between the simulated Mendelian sampling terms were assumed to be equal to the genetic correlations. The variance of the polygenic effect of each i th trait was assumed to be equal to 5% of the total additive genetic variance (σ_{Ai}^2).

The allele substitution effects of QTLs were sampled from a multinormal distribution with means of 0 and variances of 1. The correlations between allele substitution effects of the QTL underlying the 5 traits were equal to the genetic correlations. For each trait, the genetic variance explained by all QTLs was computed as the sum of the variances across all QTLs, assuming no correlation between the QTLs. The simulated additive genetic variance of each j th QTL was calculated

Table 1. Heritabilities (diagonal) and genetic correlations (off-diagonal) among the 5 simulated traits

Trait	Purebred A	Purebred B	Purebred C	Crossbred BC	Crossbred A(BC)
Purebred A	0.28				
Purebred B	0.46	0.39			
Purebred C	0.27	0.80	0.22		
Crossbred BC	0.33	0.58	0.30	0.36	
Crossbred A(BC)	0.55	0.31	0.26	0.69	0.23

as $\sigma_{gj}^2 = 2p_j(1-p_j)a_j^2$, where p_j is the allele frequency and a_j is the allele substitution effect of j th QTL. For each trait, the allele substitution effects were rescaled to obtain an additive genetic variance explained by the QTLs (σ_g^2) equal to 1. The part of the total additive genetic variance explained by the QTLs was assumed to be equal to 95% for each i th trait. Finally, the phenotypes for each trait for each animal were generated by summing the TBV and a residual error sampled from a normal distribution with a mean 0 and a variance equal to $\left(\frac{1}{h_i^2} - 1\right) \times \sigma_{Ai}^2$.

Datasets. For all the analyses, the pedigree included all the animals simulated for the creation of the 3-way crossbreeding program. The phenotype dataset included 126,000 records. Among all records, 100,000 records were associated with purebred (i.e., A, B, and C) animals randomly sampled among all purebred animals from generations 204 until 208. A total of 16,000 records were associated with A(BC) crossbred animals randomly sampled among all A(BC) crossbred animals from generations 206 until 209. Finally, 10,000 records were associated with BC crossbred dams. Average numbers of purebred and crossbred animals per generation with a phenotype are given in [Supplementary Table S1](#).

The genotype dataset included 89,000 genotypes. This included all 26,000 phenotyped BC and A(BC) crossbred animals. A total of 48,000 genotypes were from purebred (i.e., A, B, and C) animals randomly sampled among all purebred animals from the generations 205 until 208, regardless whether they had a phenotype or not. A total of 15,000 genotypes were from purebred (i.e., A, B, and C) animals randomly sampled among all purebred animals from generation 209. These 15,000 animals did not have phenotypes and are hereafter considered as selection candidates. Average numbers of purebred and crossbred animals per generation with a phenotype and a genotype are provided in [Supplementary Table S2](#).

Model and Scenarios Evaluated

Five-trait ssGBLUP was performed. The model for the i th trait [$i = A, B, C, BC, A(BC)$] was as follows:

$$\mathbf{y}_i = \mathbf{1}\mu_i + \mathbf{W}_i\mathbf{a}_i + \mathbf{e}_i$$

where for the i th trait, $\mathbf{y}_i$ is the vector of records, μ_i is the general mean, $\mathbf{a}_i$ is the vector of additive genetic effects, $\mathbf{e}_i$ is the vector of residuals, the

vector $\mathbf{1}$ is a vector of 1's relating the records to the general mean, and $\mathbf{W}_i$ is an incidence matrix relating the records to the animals. The variance components used for the simulations were used for the 5-trait ssGBLUP. The vector of additive genetic effects $\mathbf{a} = [\mathbf{a}'_A \ \mathbf{a}'_B \ \mathbf{a}'_C \ \mathbf{a}'_{BC} \ \mathbf{a}'_{A(BC)}]'$ followed a multivariate normal (MVN) distribution $MVN(\mathbf{0}, \mathbf{H}^{-1} \otimes \mathbf{\Gamma})$, where $\otimes$ is the Kronecker product, $\mathbf{\Gamma}$ is the additive genetic (co) variance matrix, and the vector of residuals $\mathbf{e} = [\mathbf{e}'_A \ \mathbf{e}'_B \ \mathbf{e}'_C \ \mathbf{e}'_{BC} \ \mathbf{e}'_{A(BC)}]'$ followed a MVN distribution $MVN(\mathbf{0}, \mathbf{I} \otimes \mathbf{R})$, where $\mathbf{R}$ is the residual (co)variance matrix.

Using all the 89,000 genotypes, the matrix $\mathbf{G}$ required for the computation of $\mathbf{H}^{-1}$ was computed without breed-specific adjustments, as suggested by [Lourenco et al. \(2016\)](#). This matrix was equal to $\mathbf{G} = 0.95\mathbf{G}_a + 0.05\mathbf{A}_{22}$ with the adjusted genomic relationship matrix $\mathbf{G}_a$ computed as follows:

$$\mathbf{G}_a = (1 - \bar{f}_p)\mathbf{G}^* + 2\bar{f}_p\mathbf{J}$$

where $\mathbf{G}^*$ is a raw genomic relationship matrix computed following the first method of [VanRaden \(2008\)](#) using current allele frequencies computed from all genotyped animals, $\mathbf{J}$ is a matrix of ones, and $\bar{f}_p$ is the average pedigree inbreeding coefficient across (core) genotyped animals. The matrix $\mathbf{H}^{-1}$ was constructed in 2 different ways. First, the complete $\mathbf{G}$ was directly inverted to obtain $\mathbf{G}_{\text{direct}}^{-1}$. Second, $\mathbf{G}_{\text{direct}}^{-1}$ was replaced by $\mathbf{G}_{\text{APY}}^{-1}$. Because the APY relies on the size and the composition of the set of core animals ([Misztal et al., 2014](#)), we investigated different numbers of core animals and different strategies to select the core animals. For all the strategies, the selection candidates were allowed to be considered as core animals. The number of core animals were 4,000, 6,000, 8,000, 10,000, and 13,000. For each size, 4 different strategies were applied to select the core animals. The core animals were randomly sampled 1) among all breed A genotyped animals (called “Breed A”), 2) among all purebred genotyped animals (called “Purebred”), or 3) among all purebred and crossbred genotyped animals (called “Purebred + Crossbred”). For the fourth strategy, a QR decomposition with pivoting of the transposed genotype matrix was applied to the animals. The QR decomposition with pivoting returns a permutation matrix such that the diagonal elements of the upper triangular matrix $\mathbf{R}$ are decreasing ([Golub and Van Loan, 1996](#)). The genotyped animals corresponding to the highest diagonal elements of the matrix $\mathbf{R}$ were chosen as core animals (called “QR”). The aim of this fourth strategy was to select core animals such that the

conditioning of the mixed model equations was improved, resulting in faster convergence, in comparison to the other 3 strategies (Fernando et al., 2016). All computations and analyses were run using our own custom programs for QR decomposition and statistical analyses: `calc_grm` (Calus and Vandenplas, 2016) for the computation of the different relationship matrices (i.e., $\mathbf{G}_{\text{direct}}^{-1}$, $\mathbf{G}_{\text{APY}}^{-1}$, and $\mathbf{A}_{22}^{-1}$) and MiXBLUP (ten Napel et al., 2016) for predicting the different GEBV. The matrices $(\mathbf{G}_{\text{direct}}^{-1} - \mathbf{A}_{22}^{-1})$ and $(\mathbf{G}_{\text{APY}}^{-1} - \mathbf{A}_{22}^{-1})$ were provided to MiXBLUP as external matrices.

Criteria

We evaluated the prediction of GEBV of genotyped selection candidates for the purebred A, B, and C performances and the crossbred A(BC) performances, for each set of core animals and each breed separately. Three criteria were computed from the GEBV of the selection candidates. First, the ratios between the accuracies of GEBV_{APY} from alternative core groups and the accuracies of $\text{GEBV}_{\text{DIRECT}}$ (i.e., from $\mathbf{G}_{\text{direct}}^{-1}$) were computed. Accuracies were computed as the Pearson correlation between GEBV and TBV. A ratio of accuracies smaller than 1 means that GEBV_{APY} is less accurate than $\text{GEBV}_{\text{DIRECT}}$. Second, regression coefficients of TBV on GEBV_{APY} and on $\text{GEBV}_{\text{DIRECT}}$ (hereafter called bias) were computed. Third, ratios between mean square errors (MSE) of GEBV_{APY} and MSE of $\text{GEBV}_{\text{DIRECT}}$ were computed. The MSE were computed as the mean of the squared differences between GEBV and TBV. All results were averaged across 5 replicates. Tukey's honest significant difference test (Tukey, 1949) was used to assess significance of differences between scenarios at a 5% significance level.

For situations with single breeds, the number of required core animals that gives accurate GEBV can be determined as the number of largest eigenvalues explaining 98% to 99% of the variation in $\mathbf{G}^*$ (Misztal, 2016; Pocrnic et al., 2016a,b). For investigating this relationship in situations involving multiple breeds and their F1 and F2 crosses, we computed the numbers of eigenvalues that explained 98% and 99% of the variation in $\mathbf{G}^*$ that included all the 89,000 genotyped purebred and crossbred animals. Computations were performed with `calc_grm` (Calus and Vandenplas, 2016). For each scenario, the number of eigenvalues was compared with the number of core animals needed such that the accuracies of GEBV_{APY} were equal to or higher than 99% of the accuracy

for $\text{GEBV}_{\text{DIRECT}}$ for both purebred and crossbred performance traits.

RESULTS

Characteristics of Simulated Data

The simulation yielded 3 breeds, A, B, and C, that were highly separated, as shown by the projections of genomic relationships into the 2 first eigenvectors for the first replicate (Fig. 2). The estimated global Wright's F_{st} statistics, that is a measure to quantify the level of genetic differentiation between the breeds, was equal to 0.35 on average across the 5 replicates. The global Wright's F_{st} statistics were estimated from the genotypes of all purebred animals of the generation 204 with the software Genepop (4.2) (Raymond and Rousset, 1995; Rousset, 2008). The mean absolute difference in allele frequencies between breeds was about 0.34 on average across the 5 replicates. All these observations suggest 3 genetically divergent populations. The average linkage disequilibrium, expressed as r^2 (Hill and Robertson, 1968), between adjacent SNP pairs with MAF > 0.05 and across chromosomes, was 0.25 for the 3 breeds on average across the 5 replicates. Genomic relationship matrices required for the singular value decomposition and genomic predictions were based on 52,518 SNPs on average across the 5 replicates.

Composition of the Core Groups

Four selection strategies were applied to compose the core groups: 1) the core animals were randomly selected among only breed A animals; 2) the core animals were randomly selected among purebred animals of breeds A, B, and C; 3) the core animals were randomly selected among purebred animals of breeds A, B, and C, and crossbred BC and A(BC) animals; and 4) the core animals were selected based on a QR decomposition of the genotype matrix. For the 4 selection strategies, Fig. 3 shows the proportions of core animals across the generations and across the breed compositions of a randomly chosen replicate for the scenario with 8,000 core animals. Similar results were obtained for the other replicates and sizes of core groups. Proportions of core animals were similar across the generations and across the breed compositions for the first 3 selection strategies. For the selection strategy based on QR decomposition, core animals were unequally spread across all generations

Figure 2. Projections of genomic relationships for purebred (A, B, and C) and crossbred [BC and A(BC)] genotyped animals into the 2 first eigenvectors for the first replicate.

Figure 3. Proportions of core animals per generation and breed composition of 1 replicate for the scenario using 8,000 core animals. Core animals were selected using 4 different strategies: 1) only from breed A animals (Breed A), 2) from purebred animals of breeds A, B, and C (PB), 3) from purebred animals of breeds A, B, and C, and crossbred BC and A(BC) animals (PB + CB), and 4) chosen based on a QR decomposition of the genotype matrix (QR). Darker colors represent higher proportions of core animals per generation and breed composition.

and breed compositions: the highest proportions of core animals selected within a generation and a breed composition were observed among the crossbred A(BC) animals and the first generation of genotyped purebred animals (Fig. 3).

Quality of GEBV With G_{direct}^{-1}

On average 5,000 genotyped selection candidates per breed were considered for computing accuracy, bias, and MSE (Table 2). For purebred performance, the accuracies were between 0.79 and 0.81. For crossbred performance, the accuracies were between 0.63 and 0.71. All sets of GEBV were almost unbiased (i.e., values for bias were close to 1) and had values of MSE close to 0 (Table 2).

Quality of GEBV With Only Breed A Core Animals

When the core groups included only breed A animals, GEBV_{APY} were predicted as accurately

as GEBV_{DIRECT} for the breed A selection candidates for both purebred and crossbred performance traits, as shown by the ratios between the accuracies of GEBV_{APY} and of GEBV_{DIRECT} (Fig. 4). In addition, GEBV_{APY} were unbiased, and MSE was close to 0 (Fig. 4; Tables 3 and 4; Supplementary Tables S3 to S6). However, GEBV_{APY} were less accurate and more biased than GEBV_{DIRECT} for the breed B and breed C selection candidates, as shown by low ratios of accuracies and high values for bias and ratios of MSE of GEBV_{APY} (Fig. 4; Tables 3 and 4; Supplementary Tables S3 to S6). Across core groups, GEBV_{APY} were from 18% to 40% less accurate than GEBV_{DIRECT}, and MSE of GEBV_{APY} were between 16% and 81% higher than the corresponding MSE of GEBV_{DIRECT}.

Quality of GEBV With Core Animals of Different Breed Compositions

Based on the 3 performance criteria, ratios of accuracies, bias, and ratios of MSE, the scenarios

Table 2. Accuracies, bias, and MSE¹ of GEBV² from the direct inversion of G (average for the 5 replicates; SD within brackets)

Selection candidates	Number	Purebred performance			Crossbred performance		
		Accuracy	Bias	MSE	Accuracy	Bias	MSE
Breed A	5,010 (24)	0.81 (0.02)	1.04 (0.05)	1.11 (0.69)	0.68 (0.04)	0.98 (0.08)	0.68 (0.51)
Breed B	4,975 (30)	0.85 (0.01)	1.06 (0.03)	1.16 (0.81)	0.63 (0.02)	0.95 (0.04)	0.90 (0.43)
Breed C	5,016 (45)	0.79 (0.04)	1.04 (0.03)	1.42 (0.74)	0.71 (0.04)	1.04 (0.07)	1.35 (1.18)

¹MSE = mean square errors.

²GEBV = genomic estimated breeding values.

Figure 4. Relative correlations of genomic estimated breeding values (GEBV) from different sizes of core groups with only breed A animals. Relative correlations for the purebred performance (PP) and crossbred performance (CP) traits are defined as the ratio between the accuracies of GEBV from alternative core groups and the corresponding accuracies of GEBV from G_{direct}^{-1} . Vertical columns depict the number of eigenvalues that explained 98% and 99% of the variation in G^* . Results are averages for the 5 replicates.

Table 3. Regression coefficients (average for the 5 replicates; SD within brackets) of TBV¹ on GEBV² from alternative core groups³ for genotyped selection candidates

Number of core animals	Purebred performance				Crossbred performance			
	Breed A	PB	PB + CB	QR	Breed A	PB	PB + CB	QR
Breed A selection candidates								
4,000	1.04 (0.06)	1.05 (0.06)	1.06 (0.06)	1.06 (0.06)	0.90 (0.08)	0.96 (0.08)	0.99 (0.09)	0.99 (0.08)
8,000	1.04 (0.05)	1.05 (0.06)	1.05 (0.06)	1.05 (0.06)	0.92 (0.08)	0.97 (0.08)	0.99 (0.08)	0.99 (0.08)
13,000	1.04 (0.05)	1.05 (0.05)	1.05 (0.05)	1.05 (0.05)	0.93 (0.08)	0.97 (0.08)	0.98 (0.08)	0.98 (0.08)
Breed B selection candidates								
4,000	1.49 (0.08)	1.06 (0.03)	1.06 (0.02)	1.07 (0.03)	1.62 (0.11)	0.91 (0.08)	0.93 (0.05)	0.94 (0.06)
8,000	1.46 (0.08)	1.06 (0.02)	1.06 (0.03)	1.06 (0.03)	1.58 (0.13)	0.93 (0.06)	0.95 (0.05)	0.95 (0.04)
13,000	1.43 (0.09)	1.06 (0.03)	1.06 (0.03)	1.06 (0.03)	1.54 (0.14)	0.93 (0.05)	0.95 (0.05)	0.95 (0.04)
Breed C selection candidates								
4,000	1.69 (0.15)	1.05 (0.04)	1.06 (0.04)	1.05 (0.03)	2.41 (0.30)	1.07 (0.08)	1.09 (0.07)	1.08 (0.07)
8,000	1.62 (0.12)	1.04 (0.04)	1.05 (0.04)	1.04 (0.03)	2.27 (0.19)	1.06 (0.07)	1.06 (0.07)	1.06 (0.07)
13,000	1.58 (0.11)	1.04 (0.03)	1.04 (0.04)	1.04 (0.03)	2.14 (0.14)	1.06 (0.07)	1.05 (0.07)	1.05 (0.07)

¹TBV = true breeding value.²GEBV = genomic estimated breeding values.³Core groups include 1) randomly selected breed A animals only (Breed A), 2) randomly selected purebred animals (PB), 3) randomly selected purebred and crossbred animals (PB + CB), and 4) animals selected based on a QR decomposition of the genotype matrix (QR).**Table 4.** Relative mean squares errors¹ (average for the 5 replicates; SD within brackets) of GEBV² from alternative core groups³ for genotyped selection candidates

Number of core animals	Purebred performance				Crossbred performance			
	Breed A	PB	PB + CB	QR	Breed A	PB	PB + CB	QR
Breed A selection candidates								
4,000	1.09 (0.33)	1.05 (0.08)	1.05 (0.04)	1.04 (0.06)	1.14 (0.29)	1.14 (0.20)	1.04 (0.04)	1.03 (0.04)
8,000	1.11 (0.34)	0.98 (0.05)	0.98 (0.06)	0.99 (0.07)	1.20 (0.28)	1.11 (0.22)	1.11 (0.23)	1.11 (0.22)
13,000	1.12 (0.33)	1.04 (0.19)	0.98 (0.06)	0.98 (0.06)	1.20 (0.28)	1.11 (0.22)	1.12 (0.22)	1.10 (0.23)
Breed B selection candidates								
4,000	1.75 (0.93)	1.12 (0.16)	1.05 (0.06)	1.08 (0.08)	1.34 (0.30)	0.99 (0.15)	1.04 (0.08)	1.02 (0.07)
8,000	1.81 (1.04)	1.09 (0.18)	1.09 (0.17)	1.09 (0.18)	1.27 (0.21)	0.99 (0.15)	0.95 (0.10)	0.95 (0.10)
13,000	1.77 (0.95)	1.15 (0.23)	1.08 (0.18)	1.08 (0.18)	1.24 (0.21)	1.00 (0.14)	0.93 (0.11)	0.94 (0.09)
Breed C selection candidates								
4,000	1.29 (0.43)	0.97 (0.08)	1.02 (0.05)	1.00 (0.05)	1.20 (0.44)	1.00 (0.16)	1.00 (0.08)	1.02 (0.07)
8,000	1.28 (0.37)	0.96 (0.08)	0.97 (0.09)	0.96 (0.09)	1.19 (0.38)	0.99 (0.16)	0.94 (0.09)	0.95 (0.09)
13,000	1.24 (0.34)	0.88 (0.16)	0.96 (0.08)	0.96 (0.09)	1.16 (0.35)	1.00 (0.16)	0.95 (0.09)	0.95 (0.08)

¹Results are expressed as the ratio between mean squares errors (MSE) of genomic estimated breeding values (GEBV) from alternative core groups and MSE of GEBV from the direct inversion of **G**.²GEBV = genomic estimated breeding values.³Core groups include 1) randomly selected breed A animals only (Breed A), 2) randomly selected purebred animals (PB), 3) randomly selected purebred and crossbred animals (PB + CB), and 4) animals selected based on a QR decomposition of the genotype matrix (QR).

with core animals of different breed compositions outperformed the scenarios with only breed A core animals for both purebred and crossbred performance traits. Use of core groups with core animals of different breed compositions allowed the prediction of $GEBV_{APY}$ that were unbiased, and (almost) as accurate as $GEBV_{DIRECT}$, for all selection candidates and performance traits. Indeed, the regression coefficients of TBV on $GEBV_{APY}$ were close to 1 (Table 3); the ratios of accuracies were higher than 0.97 for the purebred performance trait and higher

than 0.94 for the crossbred performance trait (Figs 5 and 6; Supplementary Table S3); and the MSE of $GEBV_{APY}$ were similar to MSE of $GEBV_{DIRECT}$ (Table 4; Supplementary Table S6). Ratios of accuracies close to, or higher than, 0.99 were then obtained for both traits when at least 8,000 core animals were used. The corresponding Pearson correlations between $GEBV_{APY}$ and $GEBV_{DIRECT}$, which is usually used as criteria in studies on real datasets (e.g., Ostensen et al., 2016; Strandén et al., 2017), were about 0.995 (Supplementary Material). It is

Figure 5. Relative correlations of genomic estimated breeding values (GEBV) from alternative core groups for the purebred performance traits. Core groups include randomly selected purebred and crossbred animals (Purebred + Crossbred core), randomly selected purebred animals (Purebred core), and animals selected based on a QR decomposition of the genotype matrix (QR core). Relative correlations are defined as the ratio between the accuracies of GEBV from alternative core groups and the corresponding accuracies of GEBV from $\mathbf{G}_{\text{direct}}^{-1}$. Vertical columns depict the number of eigenvalues that explained 98% and 99% of the variation in $\mathbf{G}^*$. Results are averages for the 5 replicates.

worth noting that the core size of 8,000 animals is between the numbers of eigenvalues that explained 98% and 99% of the variation in $\mathbf{G}^*$, which is about 6,498 and 9,213 eigenvalues on average across the 5 replicates, respectively (Figs 4 to 6).

Comparison of the 3 performance criteria for the purebred performance trait showed no difference among the 3 core selection strategies involving core animals of different breed compositions (Fig. 5; Table 3 and 4; Supplementary Tables S3 to S6). For the crossbred performance trait, the scenarios with purebred and crossbred core animals, either randomly chosen or chosen based on a QR decomposition, slightly outperformed the scenarios with only purebred core animals (Fig. 6). However, these outperformances were not always significant (Supplementary Material).

Quality of GEBV for Core and Noncore Selection Animals

Table 5 shows ratios of accuracies and MSE, and the regression coefficients for the scenario using 8,000 core animals randomly selected among purebred and crossbred animals. The regression

coefficients and ratios of MSE for GEBV_{APY} of core selection candidates and noncore selection candidates were similar. Ratios of accuracies for noncore selection candidates were slightly lower than the corresponding ratios for the core selection candidates, meaning that GEBV_{APY} of noncore selection candidates were slightly less accurate than those of core selection candidates, in comparison to $\text{GEBV}_{\text{DIRECT}}$. However, the differences between accuracies of GEBV_{APY} of core and noncore selection candidates were not significant following a Welch's t test (Welch, 1947) with a 5% significance level.

Convergence of ssGBLUP With Alternative Core Groups

Convergence of ssGBLUP with alternative core groups of 8,000 animals were compared against ssGBLUP using $\mathbf{G}_{\text{direct}}^{-1}$. The number of iterations of ssGBLUP using $\mathbf{G}_{\text{APY}}^{-1}$ was expressed as the ratio to the number of iterations of ssGBLUP using $\mathbf{G}_{\text{direct}}^{-1}$. Average values of this ratio across the 5 replicates (SD within brackets) were 0.85 (0.39) using breed A core animals, 1.05 (0.33) using purebred core animals, 0.95 (0.31) using

Figure 6. Relative correlations of GEBV from alternative core groups for the crossbred performance trait. Core groups include randomly selected purebred and crossbred animals (Purebred + Crossbred core), randomly selected purebred animals (Purebred core), and animals selected based on a QR decomposition of the genotype matrix (QR core). Relative correlations are defined as the ratio between the accuracies of GEBV from alternative core groups and the corresponding accuracies of GEBV from $\mathbf{G}_{\text{direct}}^{-1}$. Vertical columns depict the number of eigenvalues that explained 98% and 99% of the variation in $\mathbf{G}^*$. Results are averages for the 5 replicates.

Table 5. Quality of GEBV¹ using APY² for the core and noncore selection candidates³

Selection candidates	Number	Purebred performance			Crossbred performance		
		Accuracy ⁴	Regression coefficients	MSE ⁴	Accuracy ⁴	Regression coefficients	MSE ⁴
A core	453 (19)	0.999 (0.001)	1.02 (0.05)	0.960 (0.058)	0.997 (0.001)	0.98 (0.07)	1.087 (0.203)
A noncore	4,557 (32)	0.990 (0.003)	1.06 (0.06)	0.980 (0.066)	0.990 (0.006)	0.99 (0.08)	1.112 (0.232)
B core	456 (23)	0.998 (0.002)	1.02 (0.06)	1.076 (0.169)	0.995 (0.004)	0.88 (0.09)	0.946 (0.091)
B noncore	4,519 (37)	0.991 (0.001)	1.07 (0.03)	1.093 (0.170)	0.988 (0.004)	0.95 (0.04)	0.949 (0.096)
C core	322 (43)	0.998 (0.001)	1.07 (0.07)	0.961 (0.082)	0.999 (0.004)	1.03 (0.06)	0.939 (0.089)
C noncore	4,694 (27)	0.994 (0.001)	1.04 (0.04)	0.966 (0.087)	0.996 (0.003)	1.06 (0.07)	0.942 (0.093)

¹GEBV = genomic estimated breeding values.

²APY = algorithm for proven and young animals.

³Results (average for the 5 replicates; SD within brackets) are shown for the scenario using 8,000 core animals randomly selected among purebred and crossbred animals.

⁴Results for accuracies and mean square errors (MSE) are expressed as the ratio between accuracies (MSE) of GEBV using APY and accuracies (MSE) of GEBV using the direct inversion of $\mathbf{G}$.

purebred and crossbred animals, and 0.94 (0.30) using core animals selected based on a QR decomposition of the genotype matrix. In comparison to ssGBLUP with $\mathbf{G}_{\text{direct}}^{-1}$, use of the APY led to similar number of iterations to reach convergence. The selection strategy based on the QR decomposition led to similar convergence as the other selection strategies.

DISCUSSION

In this study, we showed that GEBV_{APY} were accurate approximations of $\text{GEBV}_{\text{DIRECT}}$ for multivariate ssGBLUP involving multiple breeds and their crosses. GEBV_{APY} as accurate as $\text{GEBV}_{\text{DIRECT}}$ were obtained when the core groups included animals from different breed compositions and when the core

groups had a size between the numbers of the largest eigenvalues explaining 98% and 99% of the variation in the raw (i.e., before blending with the pedigree relationship matrix) genomic relationship matrix ($\mathbf{G}^*$).

Composition of the Core Groups and Selection Strategies

The quality of the GEBV_{APY} for both purebred and crossbred performance traits was close to the $\text{GEBV}_{\text{DIRECT}}$ as long as all classes of purebred and crossbred animals were well represented in the core group. This was not the case if not all breeds were included in the core group. Such a situation where core animals are only from 1 breed could be obtained with a naive random selection strategy on a large genotype dataset that is dominated by 1 breed. Due to the properties of the simulated datasets, for example, similar numbers of genotyped animals per breed and per generation, a random selection of core animals across the full dataset led to similar proportions of core animals per breed composition and per generation. Based on a study involving single-breed ssGBLUP, [Ostersen et al. \(2016\)](#) advised that core groups should represent all generations. Including animals from each generation in the core group was also recommended by [Bradford et al. \(2017\)](#), especially when genotyped animals had incomplete pedigree, such as unknown parents. Incomplete pedigree could be common in crossbreeding schemes because pedigree data for crossbred animals in field conditions are difficult to collect ([Ibáñez-Escriche et al., 2009](#)). From our results with the selection strategy based on QR decomposition with pivoting, it seems that all generations, and all breed compositions, do not have to be similarly represented in core groups. Indeed, in comparison to a random selection, the selection strategy based on QR decomposition included higher proportions of crossbred A(BC) animals and of the first generation of genotyped purebred animals selected as core animals. One possible explanation is that genotypes of the crossbred A(BC) animals and of the first generation of genotyped purebred animals include a large proportion of the independent chromosome segments from all the genotyped purebred and crossbred animals. However, core groups including animals that were randomly selected and that represented similarly all generations and all breed composition gave results similar to the numerical strategy based on QR decomposition, which is computationally expensive. Therefore, a random selection of core animals by ensuring that core animals represent similarly all generations and all breed compositions is advisable for the implementation of

the APY in well-structured crossbreeding schemes as investigated in this study. More complex situations, such as multibreed (beef) cattle populations with a large variation in the observed breed compositions, would probably benefit from more advanced APY core selection approaches ([Mäntysaari et al., 2017](#)), such as the proposed numerical strategy based on QR decomposition.

Size of the Core Groups

For single-breed ssGBLUP, [Pocrnic et al. \(2016a,b\)](#) showed that the size of the core groups required to predict GEBV_{APY} at least as accurate as $\text{GEBV}_{\text{DIRECT}}$ was related to the dimensionality of the genomic information. In their studies, the most accurate GEBV_{APY} were obtained when the core size was at least equal to the number of largest eigenvalues that explained 98% of the variation in the raw genomic relationship matrix $\mathbf{G}^*$. In this study, GEBV_{APY} as accurate as $\text{GEBV}_{\text{DIRECT}}$ (i.e., with correlations between them ≥ 0.995) were obtained when the core sizes were between the numbers of largest eigenvalues that explained 98% and 99% of the variation in the raw genomic relationship matrix $\mathbf{G}^*$, provided that the composition of the core group represented the variation in all the breeds and crosses. Using a multibreed beef cattle population, [Mäntysaari et al. \(2017\)](#) also showed that a core size larger than the number of largest eigenvalues that explained 98% of the variation in $\mathbf{G}^*$ was needed to get correlations between GEBV_{APY} and $\text{GEBV}_{\text{DIRECT}}$ close to 1. Furthermore, [Mäntysaari et al. \(2017\)](#) observed that the correlation between GEBV_{APY} and $\text{GEBV}_{\text{DIRECT}}$ depended on the composition of the core groups, even with a core size close to the number of largest eigenvalues that explained 98% of the variation in $\mathbf{G}^*$. All these results suggest that the core size involving multiple breeds and crosses can be also approximated based on the dimensionality of the genomic information of all breeds and crosses together to ensure that the core size is optimal. It should be noted, however, that in crossbreeding situations relationships between the core size, the dimensionality of the genomic information, and some population parameters (e.g., number of independent segments, effective population size) are not as straightforward in as in single-breed situations ([Pocrnic et al., 2016a,b](#)).

CONCLUSIONS

We showed that the APY gives results equivalent to those obtained with the direct inversion of

the genomic relationship matrix when genotyped animals belong to a few different breeds and their F1 and F2 crosses, such as commonly observed in pig and poultry breeding programs. For such situations, we suggest that core animals could be randomly selected among all purebred and crossbred genotyped animals, while ensuring that they represent all generations and all breed compositions. It was also shown that selecting a number of core animals equal to the number of largest eigenvalues needed to explain 98% to 99% of the variation on the raw genomic relationship matrix is sufficient to achieve good quality of GEBV in crossbreeding schemes.

SUPPLEMENTARY DATA

Supplementary data are available at *Journal of Animal Science* online.

LITERATURE CITED

- Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752. doi:10.3168/jds.2009-2730
- Bradford, H. L., I. Pocrnić, B. O. Fragomeni, D. A. L. Lourenco, and I. Misztal. 2017. Selection of core animals in the algorithm for proven and young using a simulation model. *J. Anim. Breed. Genet.* 134:545–552. doi:10.1111/jbg.12276
- Calus, M. P. L., and J. Vandenplas. 2016. Calc_grm – A program to compute pedigree, genomic, and combined relationship matrices. ABGC, Wageningen UR Livest. Res., Wageningen, The Netherlands.
- Christensen, O. F. 2012. Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. *Genet. Sel. Evol.* 44:37. doi:10.1186/1297-9686-44-37
- Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42:2. doi:10.1186/1297-9686-42-2
- Fernando, R. L., H. Cheng, and D. J. Garrick. 2016. An efficient exact method to obtain GBLUP and single-step GBLUP when the genomic relationship matrix is singular. *Genet. Sel. Evol.* 48:80. doi:10.1186/s12711-016-0260-7
- Fragomeni, B. O., D. A. Lourenco, S. Tsuruta, Y. Masuda, I. Aguilar, A. Legarra, T. J. Lawlor, and I. Misztal. 2015. Hot topic: Use of genomic recursions in single-step genomic best linear unbiased predictor (BLUP) with a large number of genotypes. *J. Dairy Sci.* 98:4090–4094. doi:10.3168/jds.2014-9125
- Golub, G., and C. F. Van Loan. 1996. Matrix computations. 3rd ed. Johns Hopkins Univ. Press, Baltimore, MD.
- Hill, W. G., and A. Robertson. 1968. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38:226–231. doi:10.1007/BF01245622
- Ibáñez-Escriche, N., R. L. Fernando, A. Toosi, and J. C. Dekkers. 2009. Genomic selection of purebreds for crossbred performance. *Genet. Sel. Evol.* 41:12. doi:10.1186/1297-9686-41-12
- Legarra, A., O. F. Christensen, I. Aguilar, and I. Misztal. 2014. Single Step, a general approach for genomic selection. *Livest. Sci.* 166:54–65. doi:10.1016/j.livsci.2014.04.029
- Lourenco, D. A., S. Tsuruta, B. O. Fragomeni, C. Y. Chen, W. O. Herring, and I. Misztal. 2016. Crossbreed evaluations in single-step genomic best linear unbiased predictor using adjusted realized relationship matrices. *J. Anim. Sci.* 94:909–919. doi:10.2527/jas.2015-9748
- Lourenco, D. A., S. Tsuruta, B. O. Fragomeni, Y. Masuda, I. Aguilar, A. Legarra, J. K. Bertrand, T. S. Amen, L. Wang, D. W. Moser, et al. 2015. Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus. *J. Anim. Sci.* 93:2653–2662. doi:10.2527/jas.2014-8836
- Mäntysaari, E. A., R. D. Evans, and I. Strandén. 2017. Efficient single-step genomic evaluation for a multibreed beef cattle population having many genotyped animals. *J. Anim. Sci.* 95:4728–4737. doi:10.2527/jas2017.1912
- Masuda, Y., I. Misztal, S. Tsuruta, A. Legarra, I. Aguilar, D. A. L. Lourenco, B. O. Fragomeni, and T. J. Lawlor. 2016. Implementation of genomic recursions in single-step genomic best linear unbiased predictor for US Holsteins with a large number of genotyped animals. *J. Dairy Sci.* 99:1968–1974. doi:10.3168/jds.2015-10540
- Misztal, I. 2016. Inexpensive computation of the inverse of the genomic relationship matrix in populations with small effective population size. *Genetics* 202:401–409. doi:10.1534/genetics.115.182089
- Misztal, I., A. Legarra, and I. Aguilar. 2014. Using recursion to compute the inverse of the genomic relationship matrix. *J. Dairy Sci.* 97:3943–3952. doi:10.3168/jds.2013-7752
- Mrode, R. A. 2005. Linear models for the prediction of animal breeding values. 2nd ed. CABI Publ., Wallingford, UK.
- Ostersen, T., O. F. Christensen, P. Madsen, and M. Henryon. 2016. Sparse single-step method for genomic evaluation in pigs. *Genet. Sel. Evol.* 48:48. doi:10.1186/s12711-016-0227-8
- Pocrnić, I., D. A. Lourenco, Y. Masuda, A. Legarra, and I. Misztal. 2016a. The dimensionality of genomic information and its effect on genomic prediction. *Genetics* 203:573–581. doi:10.1534/genetics.116.187013
- Pocrnić, I., D. A. Lourenco, Y. Masuda, and I. Misztal. 2016b. Dimensionality of genomic information and performance of the algorithm for proven and young for different livestock species. *Genet. Sel. Evol.* 48:82. doi:10.1186/s12711-016-0261-6
- Powell, J. E., P. M. Visscher, and M. E. Goddard. 2010. Reconciling the analysis of IBD and IBS in complex trait studies. *Nat. Rev. Genet.* 11:800–805. doi:10.1038/nrg2865
- Raymond, M., and F. Rousset. 1995. GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *J. Hered.* 86:248–249. doi:10.1093/oxfordjournals.jhered.a111573
- Rousset, F. 2008. Genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* 8:103–106. doi:10.1111/j.1471-8286.2007.01931.x
- Sargolzaei, M., and F. S. Schenkel. 2009. Qmsim: A large-scale genome simulator for livestock. *Bioinformatics* 25:680–681. doi:10.1093/bioinformatics/btp045
- Strandén, I., K. Matilainen, G. P. Aamand, and E. A. Mäntysaari. 2017. Solving efficiently large single-step

- genomic best linear unbiased prediction models. *J. Anim. Breed. Genet.* 134:264–274. doi:10.1111/jbg.12257
- ten Napel, J., M. P. L. Calus, M. Lidauer, I. Stradén, E. A. Mäntysaari, H. A. Mulder, and R. F. Veerkamp. 2016. MiXBLUP, user-friendly software for large genetic evaluations systems. Version 2.0. Wageningen University and Research, Wageningen, The Netherlands.
- Tukey, J. W. 1949. Comparing individual means in the analysis of variance. *Biometrics* 5:99–114. doi:10.2307/3001913
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423. doi:10.3168/jds.2007-0980
- Vitezica, Z. G., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. *Genet. Res. (Camb)* 93:357–366. doi:10.1017/S001667231100022X
- Welch, B. L. 1947. The generalisation of student's problems when several different population variances are involved. *Biometrika* 34:28–35. doi:10.1093/biomet/34.1-2.28
- Wientjes, Y. C. J., and M. P. L. Calus. 2017. The purebred-cross-bred correlation in pigs: A review of theory, estimates, and implications. *J. Anim. Sci.* 95:3467–3478. doi:10.2527/jas.2017.1669