J. Dairy Sci. 96:647–654
http://dx.doi.org/10.3168/jds.2012-5656

Methods to approximate reliabilities in single-step genomic evaluation

I. Misztal,*[1] S. Tsuruta,* I. Aguilar,† A. Legarra,‡ P. M. VanRaden,§ and T. J. Lawlor#
*Department of Animal and Dairy Science, University of Georgia, Athens 30602-2771
†Instituto Nacional de Investigación Agropecuaria, Las Brujas 90200, Uruguay
‡INRA, UR631-SAGA, BP 52627, 31326 Castanet-Tolosan Cedex, France
§Animal Improvement Programs Laboratory, Agricultural Research Service, US Department of Agriculture, Beltsville, MD 20705-2350
#Holstein Association USA Inc., Brattleboro, VT 05302-0808

ABSTRACT

Reliability of predictions from single-step genomic BLUP (ssGBLUP) can be calculated by matrix inversion, but that is not feasible for large data sets. Two methods of approximating reliability were developed based on the decomposition of a function of reliability into contributions from records, pedigrees, and genotypes. Those contributions can be expressed in record or daughter equivalents. The first approximation method involved inversion of a matrix that contains inverses of the genomic relationship matrix and the pedigree relationship matrix for genotyped animals. The second approximation method involved only the diagonal elements of those inverses. The 2 approximation methods were tested with a simulated data set. The correlations between ssGBLUP and approximated contributions from genomic information were 0.92 for the first approximation method and 0.56 for the second approximation method; contributions were inflated by 62 and 258%, respectively. The respective correlations for reliabilities were 0.98 and 0.72. After empirical correction for inflation, those correlations increased to 0.99 and 0.89. Approximations of reliabilities of predictions by ssGBLUP are accurate and computationally feasible for populations with up to 100,000 genotyped animals. A critical part of the approximations is quality control of information from single nucleotide polymorphisms and proper scaling of the genomic relationship matrix.

Key words: genomic prediction, reliability, single-step evaluation, best linear unbiased predictor

INTRODUCTION

A single-step genomic BLUP (**ssGBLUP**) is a modification of BLUP to use genomic information. In ssGBLUP, the pedigree-based numerator relationship matrix (**A**) and a relationship matrix based on genomic information (**G**) are combined into a single matrix **H** (Legarra et al., 2009). The inverse of **H** has a simple form and can substitute for the inverse of **A** in existing software (Aguilar et al., 2010). Compared with multistep methods (VanRaden, 2008), ssGBLUP is simpler and applicable to complicated models. The ssGBLUP has been successfully used for chickens (Chen et al., 2011b), pigs (Forni et al., 2011), and dairy cattle (Aguilar et al., 2010, 2011b; Tsuruta et al., 2011). The computing limit of ssGBLUP is currently up to about 100,000 genotypes of progeny-tested animals (Aguilar et al., 2011a), with no limit on the number of animals or traits. However, recent developments (Ducrocq and Legarra, 2011; Legarra et al., 2011) may allow ssGBLUP to be used with an unlimited number of genotypes.

In a genetic evaluation, computing reliability of EBV is of interest. When the system of equations is small, reliability can be computed by direct matrix inversion of the BLUP mixed-model equations (Henderson, 1984). When the system of equations is large, inversion is impossible and reliability needs to be approximated. Several approximation methods for animal models exist for nongenomic evaluations. The approximation method of Misztal and Wiggans (1988), which is easy to compute, involves the effective number of records and a sum of contributions to an animal from its parents and progeny. That approximation is iterative, although a noniterative modification exists (VanRaden and Wiggans, 1991). The approximation method of Misztal and Wiggans (1988) was extended to repeatability models (Wiggans et al., 1988; Misztal et al., 1993), multiple-trait models that include maternal effect (Strabel et al., 2001), and random regression models (Sánchez et al., 2008). The advantage of approximation is simplicity and computing ease.

An approximation of reliability when genomic information is available needs to fulfill a few obvious conditions. First, more genotypes should result in equal or higher reliability. Second, a young genotyped animal should create no additional information for other animals. Third, the extra information contributed to the reference population should be small or none for a

Received April 24, 2012.
Accepted September 18, 2012.

[1] Corresponding author: ignacy@uga.edu

young animal with ancestors that are not genotyped. However, a young animal should contribute information to its nongenotyped parents. For example, genotypes can be imputed for nongenotyped parents that have several genotyped progeny. Similarly, the single-step equations adjust parent EBV through linear rather than nonlinear imputation methods. Fourth, no extra reliability should be gained for an animal from different lines or breeds. The purpose of this study was to extend the approximation algorithm of Misztal and Wiggans (1988) to ssGBLUP.

MATERIALS AND METHODS

Data

Data were simulated using QMSim software (Sargolzaei and Schenkel, 2009) for an additive trait with heritability of 0.5, 2 chromosomes, and 60 QTL. Performance was simulated for 15,800 individuals in 5 generations, and 1,500 individuals of the last 3 generations were genotyped. Each animal in the simulation had a single phenotypic record. Details of the simulation were reported by Wang et al. (2012).

Derivation of Approximation Methods

Reliability of animal i (rel$_i$) can be approximated as $1 - [\alpha/(\alpha + d_i)]$, where α is the ratio of error variance to animal genetic variance and d_i is the amount of information for animal i in units of effective number of records (Misztal and Wiggans, 1988). The information can be calculated by inversion of the left-hand side (**LHS**) of the mixed-model equations as $\mathrm{LHS}_{uu}^{ii} = 1/\left(\alpha + d_i\right)$, where uu denotes the block of the LHS for the animal effect for the animal effect and ii denotes the diagonal element corresponding to animal i. Then, d_i can be partitioned as $d_i^r + d_i^p + d_i^g$, where d_i^r is the contribution from records (phenotypes), d_i^p is the contribution from pedigrees, and d_i^g is the contribution from genomic information. With pedigree information, contributions to an animal are from progeny and parents only. With genomic information, contributions are from all animals with genomic information.

For simplicity, assume a single-trait mixed model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e},$$

where $\mathbf{y}$ is a vector of observations, $\mathbf{b}$ is a fixed effect, $\mathbf{u}$ is the random additive animal effect, $\mathbf{X}$ and $\mathbf{Z}$ are incidence matrices relating $\mathbf{b}$ and $\mathbf{u}$ to $\mathbf{y}$, and $\mathbf{e}$ is the random residual effect. When relationships are known, LHS is

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\boldsymbol{\alpha} \end{bmatrix},$$

where $\mathbf{A}$ is the numerator relationship matrix, and the diagonal elements of the inverse of the LHS for animal i can be presented as

$$\mathrm{LHS}_{uu}^{ii} = 1/\left(\alpha + d_i^r + d_i^p\right). \qquad [1]$$

If $\mathbf{D}^r = \left\{d_i^r\right\}$ and $\mathbf{D}^p = \left\{d_i^p\right\}$ are known, Equation 1 can be simplified to $\mathrm{LHS}_{uu}^{ii} = \left[\left(\mathbf{D}_i^r + \mathbf{D}_i^p + \mathbf{I}\boldsymbol{\alpha}\right)^{-1}\right]_{ii}$, where $\mathbf{I}$ is an identity matrix, or approximated as

$$\mathrm{LHS}_{uu}^{ii} \approx \left[\left(\mathbf{D}_i^r + \mathbf{A}^{-1}\boldsymbol{\alpha}\right)^{-1}\right]_{ii}.$$

Misztal and Wiggans (1988) estimated the contributions from relationships separately for each relationship in an iterative formula:

$$\begin{bmatrix} 1.5\alpha + d_s - d_{s_i}^r & 0.5\alpha & -\alpha \\ 0.5\alpha & 1.5\alpha + d_d - d_{d_i}^r & -\alpha \\ -\alpha & -\alpha & 2\alpha + d_i - d_{i_s}^r - d_{i_d}^r \end{bmatrix}^{-1} = \begin{bmatrix} 1/\left(\alpha + d_s\right) & \ldots & \ldots \\ \ldots & 1/\left(\alpha + d_d\right) & \ldots \\ \ldots & \ldots & 1/\left(\alpha + d_i\right) \end{bmatrix},$$

where d_i, d_s, and d_d are total amounts of information from animal i and its sire (s) and dam (d), respectively; $d_{s_i}^r$ and $d_{d_i}^r$ are contributions to sire and dam information from records of animal i, respectively; and $d_{i_s}^r$ and $d_{i_d}^r$ are contributions to information for animal i from records of its sire and dam, respectively. Nonmatrix formulas for the same contributions, but expressed in daughter equivalents, were derived by VanRaden and Wiggans (1991).

When genomic information is available, the LHS of ssGBLUP is

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{H}^{-1}\boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\boldsymbol{\alpha} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}\boldsymbol{\alpha} \end{bmatrix},$$

where $\mathbf{A}_{22}$ is a pedigree-based numerator relationship matrix for genotyped animals (Aguilar et al., 2010). The diagonal elements of the inverse of the LHS for animal i now include an additional element because of the genomic information $\mathrm{LHS}_{uu}^{ii} = 1\big/\left(\alpha + d_i^r + d_i^p + d_i^g\right)$. If $\mathbf{D}^r$ and $\mathbf{D}^p$ are known, then the LHS can be approximated as

$$\mathrm{LHS}_{uu}^{ii} \approx \left\{\left[\mathbf{D}_i^r + \mathbf{D}_i^p + \left(\mathbf{I} + \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}\right)\boldsymbol{\alpha}\right]^{-1}\right\}_{ii}. \qquad [2]$$

In Equation 2, $\mathbf{G}$ accounts for genomic information and $\mathbf{A}_{22}$ accounts for an adjustment to prevent double counting of the relationship information contained in $\mathbf{G}$ and $\mathbf{A}$.

Based on Equation 2, an algorithm (**Approx1**) can be created to approximate reliabilities with genomic information:

1. Approximate reliabilities with an algorithm that ignores genomic information.
2. Convert those reliabilities to effective number of records for genotyped animals only: $d_i = \alpha[1/(1 - \mathrm{rel}_i) - 1]$.
3. Calculate the inverse: $\mathbf{Q}^{-1} = \left[\mathbf{D} + \left(\mathbf{I} + \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}\right)\boldsymbol{\alpha}\right]^{-1}$.
4. Calculate genomic reliabilities: $\mathrm{rel}_i = 1 - \alpha q^{ii}$.
5. Optionally adjust reliabilities of nongenotyped animals if those are functions of reliabilities of genotyped animals. This can be done by using complete contributions (including genomics) for d_i, d_s, and d_d.

An alternative algorithm (**Approx2**) can be used if the off-diagonals of $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ are ignored. Then $\mathbf{Q}^{-1}$ can be simplified to $\left\{\mathbf{D} + \left[\mathbf{I} + \mathrm{diag}\left(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}\right)\right]\boldsymbol{\alpha}\right\}^{-1}$. The algorithm Approx2 is based on observations that diagonal information in $\mathbf{G}^{-1}$ contains the information in $\mathbf{A}^{-1}$ plus genomic information (Chen et al., 2011b).

Genomic Contributions Based on Pairs of Animals

Contributions due to genomic information for one animal are from all the other genotyped animals, and the formulas above allow calculating the sum of all such contributions. However, knowledge of individual contributions can be useful in understanding the nature of the genomic information and also aid in selection of candidates for genotyping.

Let the genetic effects be split into 2 uncorrelated effects: $\mathbf{u} = \mathbf{u}^* + \mathbf{d}$, where $\mathrm{var}(\mathbf{u}^*) = \mathbf{A}_{22}$ and $\mathrm{var}(\mathbf{d}) = \mathbf{G} - \mathbf{A}_{22}$. Therefore, an extremely rough approximation can be obtained by using regular BLUP to estimate $\mathbf{u}^*$ and then estimating $\mathbf{d}$ separately with a model such as $\mathbf{y} - \mathbf{X}\hat{\mathbf{b}} - \mathbf{Z}\hat{\mathbf{u}}^* = \mathbf{Z}\mathbf{d} + \varepsilon$, where ε is a random residual. The approximate reliability of $\mathbf{u}$ that results from the sum of contributions for reliabilities of $\mathbf{u}^*$ and $\mathbf{d}$ is incorrect because $\mathbf{u}^*$ and $\mathbf{d}$ are correlated a posteriori even if they are uncorrelated a priori. Therefore, if $\mathbf{D}^r$ and $\mathbf{D}^p$ are known, an even rougher approximation of reliabilities is

$$\mathrm{LHS}_{uu}^{ii} \approx \left\{\left[\mathbf{D}_i^r + \mathbf{D}_i^p + \left(\mathbf{Z}'\mathbf{Z}\right)^{-1} + \left(\mathbf{G} - \mathbf{A}_{22}\right)^{-1}\boldsymbol{\alpha}\right]^{-1}\right\}_{ii}. \qquad [3]$$

An insight into sources of genomic information can be gained by examining $\mathbf{G}$ for animals i and j after adjustment for $\mathbf{A}_{22}$ (i.e., $\mathbf{G} - \mathbf{A}_{22}$ in Equation 3). Assuming equal genomic and pedigree-based inbreeding [i.e., $\mathrm{diag}(\mathbf{G}) = \mathrm{diag}(\mathbf{A}_{22})$], such a matrix is

$$\mathbf{S}_{ij} = \begin{bmatrix} 1 & g_{ij} - a_{22_{ij}} \\ \text{symmetric} & 1 \end{bmatrix},$$

with inverse

$$\mathbf{S}_{ij}^{-1} = \begin{bmatrix} 1/v & -\Delta_{ij}/v \\ \text{symmetric} & 1/v \end{bmatrix},$$

where S_{ij} is the relationship matrix for animals i and j, $\Delta_{ij} = g_{ij} - a_{22_{ij}}$ and $v = 1 - \Delta_{ij}^2$. For the mixed-model equations that involve animals i and j only, the LHS is

$$\begin{bmatrix} d_i & 0 \\ 0 & d_j \end{bmatrix} + \alpha\mathbf{S}_{ij}^{-1} = \begin{bmatrix} d_i + \left(\alpha/v\right) & -\alpha\Delta/v \\ \text{symmetric} & d_j + \left(\alpha/v\right) \end{bmatrix},$$

where d_i and d_j are total information for animals i and j except for this particular relationship. Then, the inverse of the LHS is

$$\left(1\Big/\left\{\left[d_i + \left(\alpha/v\right)\right]\left[d_j + \left(\alpha/v\right)\right] - \left(\alpha\Delta/v\right)^2\right\}\right)\begin{bmatrix} d_j + \left(\alpha/v\right) & \alpha\Delta/v \\ \text{symmetric} & d_i + \left(\alpha/v\right) \end{bmatrix},$$

and the reciprocal for each element of the LHS inverse is

$$\begin{aligned} d_i + \alpha + d_{ij}^g &= \left\{\left[d_i + \left(\alpha/v\right)\right]\left[d_j + \left(\alpha/v\right)\right] - \left(\alpha\Delta/v\right)^2\right\}\Big/\left[d_j + \left(\alpha/v\right)\right] \\ &= d_i + \left(\alpha/v\right) - \left\{\left(\alpha\Delta/v\right)^2\Big/\left[d_j + \left(\alpha/v\right)\right]\right\} \\ &= d_i + \alpha + \left[\alpha\left(1 - v\right)/v\right] - \left\{\left(\alpha\Delta/v\right)^2\Big/\left[d_j + \left(\alpha/v\right)\right]\right\}. \end{aligned}$$

The contribution from genomic information from animal j to i is

$$\begin{aligned} d_{ij}^g &= \left[\alpha\left(1 - v\right)/v\right] - \left\{\left(\alpha\Delta/v\right)^2\Big/\left[d_j + \left(\alpha/v\right)\right]\right\} \\ &= \alpha\left[\Delta^2\Big/\left(1 - \Delta^2\right)\right]\left[1 - \left(\alpha\Big/\left\{\left[\alpha/\left(1 - \Delta^2\right)\right] + d_j\right\}\right)\right]. \end{aligned} \qquad [4]$$

Assume that for a properly scaled **G**, the differences between **G** and $\mathbf{A}_{22}$ are small; such differences had a standard deviation of < 0.05 in dairy cattle (VanRaden, 2008). Then, because $1 - \Delta_{ij}^2 \approx 1$, Equation 4 can be simplified to

$$d_{ij}^g \approx \alpha\Delta^2\left[d_j\Big/\left(d_j + \alpha\right)\right] = \alpha\left(g_{ij} - a_{22_{ij}}\right)^2 \mathrm{rel}_j.$$

Summing contributions from all the genotyped animals, the total contribution to animal i from genomic information is

$$d_i^g = \sum_{j, j \neq i} d_{ij}^g \approx \alpha \sum_{j, j \neq i}\left[\left(g_{ij} - a_{22_{ij}}\right)^2 \mathrm{rel}_j\right]. \qquad [5]$$

Equation 5 was found to be inaccurate because of double counting and, therefore, was not used for comparisons. The total value of the reference population may be proportional to squared relationship differences times reliability, but an individual's genomic reliability also depends on its average relationship to the reference population (Liu et al., 2010; Wiggans and VanRaden, 2010). Thus, the overall $\sum\left(g_{ij} - a_{22_{ij}}\right)$ and an individual animal's $\sum g_{ij}$ or $\sum g_{ij}^2$ (without subtracting $a_{22_{ij}}$) may be useful. Two previous genomic reliability approximations did not require inversion. Using $\sum g_{ij}^2$ was found to give better results than using $\sum g_{ij}$, but weighting by rel_j did not help in the study of Liu et al. (2010). For official estimates of US reliability, $\sum\left[g_{ij}\left(\mathrm{rel}_i\right)\right]$ is used (Wiggans and VanRaden, 2010).

Factors that influence reliability from genomic information can be illustrated conceptually with Equation 5. First, genomic information is a function of squared differences between genomic and pedigree relationships. Therefore, relationships with such differences that are very small contribute little. For example, an animal with a difference of 0.02 contributes 9 times less than an animal with a difference of 0.06. Second, contributions are scaled by the square of reliability. Thus, an animal with a reliability of 0.99 (e.g., an old progeny-tested bull) contributes 3 times more than an animal with a reliability of 0.33. Third, a genotyping or pedigree error would inflate the contribution. For example, for a conflicting parent-progeny relationship, the difference would be close to 0.5, whereas the correct relationship would average 0.04. Subsequently, one error in pedigree could negate the contributions of >100 correct relationships. Finally, Equation 5 is sensitive to scaling of **G**. If **G** is constructed using incorrect gene frequencies, the relationship between unrelated animals will not be zero and can be as high as 0.6 if 0.5 gene frequencies are used (Forni et al., 2011). Indeed, Strandén and Christensen (2011) indicated that for regular BLUP that incorporates genomic information, EBV are identical despite assumed allelic frequencies, whereas computed reliabilities are not. A general explanation for the scaling sensitivity is that assuming different allelic frequencies implies different genetic base populations. Thus, scaling **G** to be compatible with $\mathbf{A}_{22}$ [e.g., as in Chen et al. (2011b) or Vitezica et al. (2011)] is crucial so that the genomic base is the same as that for pedigree relationships. For $\mathbf{G}^{-1}$, scaling seems less critical as its statistics are much less affected by gene frequencies.

Analyses

Total information per animal was calculated by inversion using an animal model with pedigree relationships only and using ssGBLUP. Contributions from genomics were calculated as differences in information from the 2 analyses. Approximations used nongenomic information from the pedigree-only analysis. Matrix **G** was constructed using current allele frequencies and subsequently rescaled so that means of diagonal and off-diagonal elements were identical to those of $\mathbf{A}_{22}$ (Chen et al., 2011a; Vitezica et al., 2011). Initially, ssGBLUP, Approx1, and Approx2 reliabilities were calculated from the sum of all contributions. For approximations only, reliabilities also were calculated with genomic contributions regressed to have a mean equal to that for ssGBLUP genomic contributions.

RESULTS AND DISCUSSION

Table 1 shows statistics for ssGBLUP and approximated genomic contributions as well as the correlations

Table 1. Statistics for genomic contributions, reliabilities, and reliabilities after rescaling for genomic contributions from 3 methods to estimate reliability

Estimate	Method[1]	Mean (±SE)	Range	Correlation with ssGBLUP estimate
Genomic contribution	ssGBLUP	2.4 ± 0.4	1.7–4.7	—
	Approx1	3.9 ± 0.6	2.9–8.3	0.92
	Approx2	8.6 ± 4.2	4.5–62	0.56
Reliability (%)	ssGBLUP	81 ± 2	77–90	—
	Approx1	85 ± 2	83–93	0.98
	Approx2	91 ± 2	86–98	0.72
Reliability after rescaling for genomic contribution (%)	Approx1	81 ± 2	78–92	0.99
	Approx2	81 ± 4	75–96	0.89

[1] Methods included single-step genomic BLUP (ssGBLUP), approximation using inversion of a matrix that contains inverses of the genomic relationship matrix and the pedigree relationship matrix for genotyped animals (Approx1), and approximation using only the diagonal elements of those inverses (Approx2).

between ssGBLUP and approximated contributions. Correlation with ssGBLUP genomic contributions was 0.92 for Approx1 and 0.56 for Approx2 contributions. Mean genomic contributions were inflated by 62% for Approx1 and by 258% for Approx2. Inflation resulted from ignoring off-diagonal elements in $\mathbf{Z}'\mathbf{Z}$ and $\mathbf{A}^{-1}$. Additional inflation would have resulted from ignoring the off-diagonals of $\mathbf{X}'\mathbf{X}$; however, the only fixed effect in this study was the mean.

Table 1 also shows statistics for ssGBLUP and approximated reliabilities as well as the correlations between ssGBLUP and approximated reliabilities. Correlation with ssGBLUP reliabilities was 0.98 for Approx1 and 0.72 for Approx2 reliabilities. Reliabilities from both approximation methods were inflated. After rescaling for genomic contributions (Table 1), reliabilities were no longer inflated, and correlation with ssGBLUP reliability increased to 0.99 for Approx1 and 0.89 for Approx2 reliabilities. In practice, the coefficient of regression is unknown and has to be derived (e.g., experimentally) to match realized reliabilities.

The Approx1 algorithm is computationally feasible when $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ are calculated as part of ssGBLUP. The Approx2 algorithm, which is a simplification of Approx1, generally offers little benefit over Approx1 except when diagonal elements of $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ can be computed at a lower cost. In general, the cost of either approximation is one inversion of a matrix the dimension of $\mathbf{G}^{-1}$ because the remaining costs are small. The time to invert $\mathbf{G}$ for 30,000 animals using an 8 core processor in 2010 was about 1 h (Aguilar et al., 2011a). Extrapolating, such time would be 1.5 d for 100,000 animals, although this time will be smaller with newer computers and more cores. Extra research is needed to determine if the approximations can be expressed in terms of ssGBLUP, which does not require inversion of $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ (Legarra and Ducrocq, 2012).

For Approx1 and Approx2, reliability calculated by inversion is assumed to reflect realized reliability. This was confirmed by Hayes et al. (2009) in a simulation study. However, predicted reliabilities were inflated compared with realized reliabilities in a study by VanRaden et al. (2009). Several explanations exist for the inflation. First, inflation could result from several approximations and assumptions inherent in multiple-step procedures. Second, genetic relationships fade over generations under selection (Muir, 2007) and, thus, contributions from older generations may be inflated. Third, effects of major genes (if they exist) may not be fully accounted for by the method. Fourth, the analysis model may be deficient (e.g., from ignoring selection, censoring, or preferential treatment). As an example, the genetic parameters for several chicken traits in 2 lines were different between complete data sets or genotyped subsets (Chen et al., 2011b), and origins of those differences were difficult to explain. Fifth, published reliabilities estimate the correlation between predicted and true breeding values in a hypothetical unselected population with random mating, whereas actual populations usually contain only selected candidates (i.e., young bulls selected based on parental information). Therefore, observed correlations tend to be reduced by selection. Differences among predicted and realized reliabilities were not obvious before the era of genomic selection, as interest in realized reliabilities was limited. Probably the best way to address the issue of inflated predicted reliabilities is by research on causes of inflation, both with and without genomic information.

The Approx1 and Approx2 algorithms are based on differences between $\mathbf{G}$ and $\mathbf{A}_{22}$. Chen et al. (2011a) found that number of SNP and assumed allele frequencies affected statistics of $\mathbf{G}$ and $\mathbf{G}^{-1}$. They recommended that $\mathbf{G}$ be constructed with current allele frequencies and then rescaled to match statistics of $\mathbf{A}_{22}$. They also found that decreasing the number of SNP

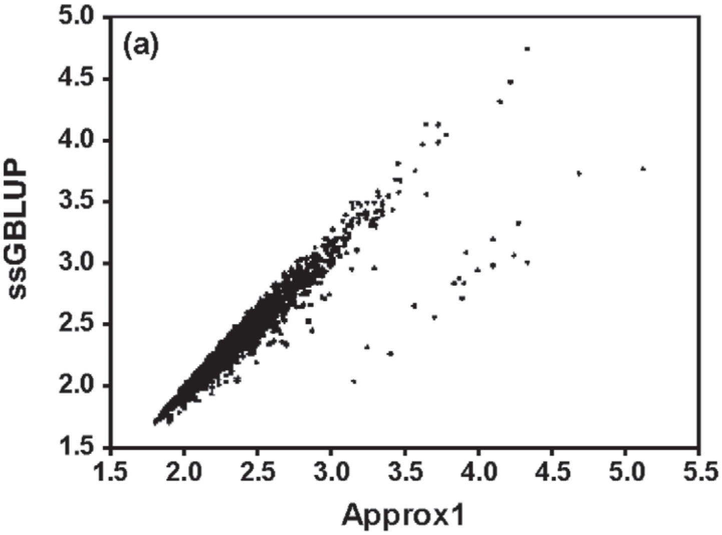

Figure 1. Genomic contributions from single-step genomic BLUP (ssGBLUP) compared with scaled genomic contributions from (a) approximation using inversion of a matrix that contains inverses of the genomic relationship matrix and the pedigree relationship matrix for genotyped animals (Approx1) or (b) approximation using only the diagonal elements of those inverses (Approx2); genomic contributions from Approx1 and Approx2 were regressed to have a mean equal to that for ssGBLUP genomic contributions.

when constructing **G** inflated **G** (although inflation was small when the number of SNP was >20,000). In populations with multiple lines with different allele frequencies (e.g., Simeone et al., 2012), **G** needs to be rescaled for different lines to avoid less accurate approximations of accuracy (e.g., Harris and Johnson, 2010). Wang and Misztal (2011) found that the standard deviation of a difference between elements of **G** and $\mathbf{A}_{22}$ was <0.04 for properly scaled **G**. A similar value was found by Hill and Weir (2011). Larger differences of up to 1.0 are the result of genotyping and pedigree mistakes, incomplete pedigree, and mixing of lines. Also, for identical twins and clones, $g_{ij} = 1.0$, but $a_{22_{ij}} = 0.5$ because the pedigree treats them as full sibs. Such differences can lead to poor approximations of reliability for selected animals.

Figure 1 shows ssGBLUP and approximated genomic contributions after scaling. Most of the Approx1 con-

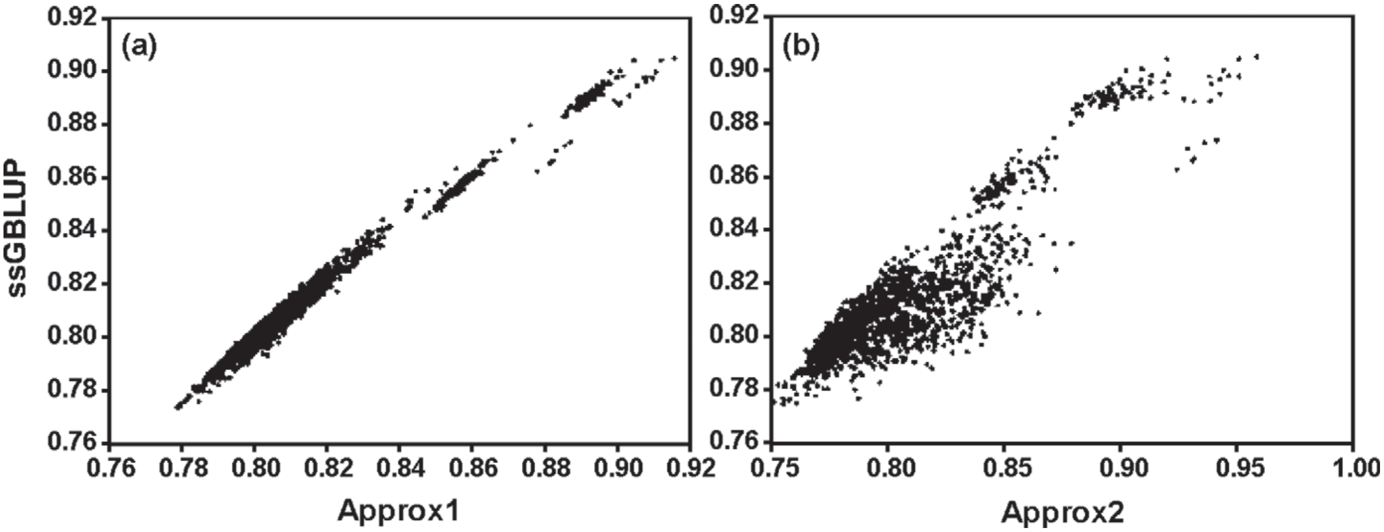

Figure 2. Genomic reliabilities from single-step genomic BLUP (ssGBLUP) compared with genomic reliabilities after scaling from (a) approximation using inversion of a matrix that contains inverses of the genomic relationship matrix and the pedigree relationship matrix for genotyped animals (Approx1) or (b) approximation using only the diagonal elements of those inverses (Approx2); genomic contributions from Approx1 and Approx2 were regressed to have a mean equal to that for ssGBLUP genomic contributions.

tributions were similar to ssGBLUP contributions, but some were inflated. For Approx2, the fit for most animals was not as good, and inflation for selected animals was larger. Reasons for inflation for some animals will be studied subsequently.

Figure 2 shows ssGBLUP and approximated reliabilities after scaling. The fit for Approx1 was very good, whereas that for Approx2 was not as good. The fit for reliabilities was better than for genomic contributions because of an upper bound of 1 and the stabilizing effect of contributions from records and pedigrees.

Although Approx1 showed a very good fit for the simulated data set, the fit in general is likely to depend on the population structure. More testing will determine the quality of Approx1 with different data sets. Such testing may also provide guidelines for optimal scaling especially when ssGBLUP reliabilities are too expensive to compute.

CONCLUSIONS

Two algorithms to approximate reliabilities from ssGBLUP were developed. The algorithm that used inversion of a matrix that contained the inverse of the genomic relationship matrix as well as the inverse of the pedigree relationship matrix for genotyped animals was relatively accurate and inexpensive for <100,000 genotypes. It required some heuristics to regress inflated genomic contributions. Reliability calculations have become more important because breeders could easily understand numbers of daughters or numbers of records in the past, but such measures no longer apply directly to genomic predictions.

ACKNOWLEDGMENTS

This project was supported by Agriculture and Food Research Initiative Competitive Grants no. 2009-65205-05665 and 2010-65205-20366 from the US Department of Agriculture National Institute of Food and Agriculture as well as by Holstein Association USA (Brattleboro, VT), and PIC (Hendersonville, TN). Discussions with G. R. Wiggans of the Animal Improvement Programs Laboratory (Beltsville, MD) were greatly appreciated. Manuscript review by S. M. Hubbard of the Animal Improvement Programs Laboratory is gratefully acknowledged.

REFERENCES

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.

Aguilar, I., I. Misztal, A. Legarra, and S. Tsuruta. 2011a. Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. J. Anim. Breed. Genet. 128:422–428.

Aguilar, I., I. Misztal, S. Tsuruta, G. R. Wiggans, and T. J. Lawlor. 2011b. Multiple trait genomic evaluation of conception rate in Holsteins. J. Dairy Sci. 94:2621–2624.

Chen, C. Y., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2011a. Effect of different genomic relationship matrices on accuracy and scale. J. Anim. Sci. 89:2673–2679.

Chen, C. Y., I. Misztal, I. Aguilar, S. Tsuruta, T. H. E. Meuwissen, S. E. Aggrey, T. Wing, and W. M. Muir. 2011b. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens. J. Anim. Sci. 89:23–28.

Ducrocq, V., and A. Legarra. 2011. An iterative implementation of the single step approach for genomic evaluation which preserves existing genetic evaluation models and software. Interbull Bull. 44:138–142.

Forni, S., I. Aguilar, and I. Misztal. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genet. Sel. Evol. 43:1.

Harris, B. L., and D. L. Johnson. 2010. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. J. Dairy Sci. 93:1243–1252.

Hayes, B. J., P. M. Visscher, and M. E. Goddard. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. Genet. Res. (Camb.) 91:47–60.

Henderson, C. R. 1984. Applications of Linear Models in Animal Breeding. University of Guelph, Guelph, ON, Canada.

Hill, W. G., and B. S. Weir. 2011. Variation in actual relationship as a consequence of Mendelian sampling and linkage. Genet. Res. (Camb.) 93:47–64.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663.

Legarra, A., and V. Ducrocq. 2012. Computational strategies for national integration of phenotypic, genomic, and pedigree data in a single-step best linear unbiased prediction. J. Dairy Sci. 95:4629–4645.

Legarra, A., I. Misztal, and I. Aguilar. 2011. The single step: Genomic evaluation for all. Book of Abstr. 62nd Annu. Mtg. Euro. Fed. Anim. Sci. No. 17:1. Wageningen Academic Publishers, Wageningen, the Netherlands.

Liu, Z., F. Seefried, F. Reinhardt, and R. Reents. 2010. Approximating reliabilities of estimated direct genomic values. Interbull Bull. 41:29–32.

Misztal, I., T. J. Lawlor, and T. H. Short. 1993. Implementation of single- and multiple-trait animal models for genetic evaluation of Holstein type traits. J. Dairy Sci. 76:1421–1432.

Misztal, I., and G. R. Wiggans. 1988. Approximation of prediction error variance in large-scale animal models. J. Dairy Sci. 71(Suppl. 2):27–32.

Muir, W. M. 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J. Anim. Breed. Genet. 124:342–355.

Sánchez, J. P., I. Misztal, and J. K. Bertrand. 2008. Evaluation of methods for computing approximate accuracies in maternal random regression models for growth trait in beef. J. Anim. Sci. 86:1057–1066.

Sargolzaei, M., and F. S. Schenkel. 2009. QMSim: A large-scale genome simulator for livestock. Bioinformatics 25:680–681.

Simeone, R., I. Misztal, I. Aguilar, and Z. G. Vitezica. 2012. Evaluation of a multi-line broiler chicken population using a single-step genomic evaluation procedure. J. Anim. Breed. Genet. 129:3–10.

Strabel, T., I. Misztal, and J. K. Bertrand. 2001. Approximation of reliabilities for multiple-trait models with maternal effects. J. Anim. Sci. 79:833–839.

Strandén, I., and O. F. Christensen. 2011. Allele coding in genomic evaluation. Genet. Sel. Evol. 43:25.

Tsuruta, S., I. Misztal, I. Aguilar, and T. J. Lawlor. 2011. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. J. Dairy Sci. 94:4198–4204.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423.

VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16–24.

VanRaden, P. M., and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. 74:2737–2746.

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.

Wang, H., and I. Misztal. 2011. Comparisons of numerator and genomic and relationship matrices. J. Dairy. Sci. 94(E-Suppl. 1):163. (Abstr.)

Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. Genet. Res. (Camb.) 94:73–83.

Wiggans, G. R., I. Misztal, and L. D. Van Vleck. 1988. Animal model evaluation of Ayrshire milk yield with all lactations, herd-sire interaction, and groups based on unknown parents. J. Dairy Sci. 71:1319–1329.

Wiggans, G. R., and P. M. VanRaden. 2010. Improved reliability approximation for genomic evaluations in the United States. J. Dairy Sci. 93(E-Suppl. 1):533. (Abstr.)