# Effect of different genomic relationship matrices on accuracy and scale

Ching-Yi Chen, I. Misztal, I. Aguilar, Andres Legarra, W.M. Muir

**Effect of different genomic relationship matrices on accuracy and scale**

C. Y. Chen, I. Misztal, I. Aguilar, A. Legarra and W. M. Muir

The online version of this article, along with updated information and services, is located on the World Wide Web at:

http://www.journalofanimalscience.org/content/89/9/2673

# Effect of different genomic relationship matrices on accuracy and scale[1]

**C. Y. Chen,\*[2,3] I. Misztal,\* I. Aguilar,\*† A. Legarra,‡ and W. M. Muir§**

\*Department of Animal and Dairy Science, University of Georgia, Athens 30602-2771;
†Instituto Nacional de Investigación Agropecuaria, Las Brujas 90200, Uruguay;
‡INRA, UR631 Station d'Amélioration Génétique des Animaux (SAGA), BP 52627, 32326 Castanet-Tolosan,
France; and §Department of Animal Science, Purdue University, West Lafayette, IN 47907-1151

**ABSTRACT:** Phenotypic data on BW and breast meat area were available on up to 287,614 broilers. A total of 4,113 birds were genotyped for 57,636 SNP. Data were analyzed by a single-step genomic BLUP (ssGBLUP), which accounts for all phenotypic, pedigree, and genomic information. The genomic relationship matrix (**G**) in ssGBLUP was constructed using either equal (0.5; GEq) or current (GC) allele frequencies, and with all SNP or with SNP with minor allele frequencies (MAF) below multiple thresholds (0.1, 0.2, 0.3, and 0.4) ignored. Additionally, a pedigree-based relationship matrix for genotyped birds ($\mathbf{A}_{22}$) was available. The matrices and their inverses were compared with regard to average diagonal (AvgD) and off-diagonal (AvgOff) elements. In $\mathbf{A}_{22}$, AvgD was 1.004 and AvgOff was 0.014. In GEq, both averages decreased with the increasing thresholds for MAF, with AvgD decreasing from 1.373 to 1.020 and AvgOff decreasing from 0.722 to 0.025. In GC, AvgD was approximately 1.01 and AvgOff was 0 for all MAF. For inverses of the relationship matrices, all AvgOff were close to 0; AvgD was 2.375 in $\mathbf{A}_{22}$, varied from 11.563 to 12.943 for GEq, and increased from 8.675 to 12.859 for GC as the threshold for MAF increased. Predictive ability with all GEq and GC was similar except that at MAF = 0.4, they declined by 0.01 for BW and improved by 0.01 for breast meat area. Compared with BLUP, EBV in the ssGBLUP were, on average, increased by up to 1 additive SD greater with GEq and decreased by 2 additive SD less with GC. Genotyped animals were biased upward with GEq and downward with GC. The biases and differences in EBV could be controlled by adding a constant to GC; they were eliminated with a constant of 0.014, which corresponds to AvgOff in $\mathbf{A}_{22}$. Unbiased evaluation in the ssGBLUP may be obtained with GC scaled to be compatible with $\mathbf{A}_{22}$. The reduction of SNP with small MAF has a small effect on the real accuracy, but it may falsely increase the estimated accuracies by inversion.

**Key words:** chicken, genetic evaluation, genomic prediction

## INTRODUCTION

Genomic selection can be carried out either by estimating the effects of individual SNP or by using BLUP with a genomic relationship matrix, **G** (VanRaden et al., 2009b). For traits without large major genes, most methods provide similar accuracies (Hayes et al., 2009). Usually, not all animals have genotypes. A methodology exists to combine pedigree and genomic relationships (Legarra et al., 2009). A single-step approach that uses the combined relationships allows for a simple evaluation that takes all available phenotypic, pedigree, and genomic information and has no limitations on models for analyses (Aguilar et al., 2010; Chen et al., 2011).

One of the issues plaguing genomic predictions is bias. For instance, in dairy cattle the bulls evaluated genomically seem to have EBV biased upward (Van-

**Table 1.** Description of phenotypic records for genotyped birds and all birds

| Item[1] | No. of records | Mean | SD |
|---|---|---|---|
| Genotyped | | | |
| BW, 100 g | 4,113 | 25.79 | 3.20 |
| BM, cm² | 3,923 | 43.98 | 5.75 |
| All | | | |
| BW, 100 g | 287,614 | 26.25 | 4.91 |
| BM, cm² | 69,057 | 45.52 | 7.34 |

[1]BW = BW at 6 wk; BM = ultrasound area of breast meat.

Raden et al., 2009a). Bias from genomic evaluation has been reported in other studies (Aguilar et al., 2010; Forni et al., 2011). With bias, an accurate comparison of young and older animals is difficult.

The bias can partially be due to a genomic relationship matrix that is suboptimal. Such a matrix can be constructed by assuming different gene frequencies and can possibly be shifted and scaled, and SNP with small minor allele frequencies (**MAF**) may be ignored (Van-Raden, 2008). In Holsteins, Aguilar et al. (2010) found that the use of the equal allele frequencies resulted in the greatest realized accuracy and the smallest bias. In pigs, Forni et al. (2011) found that the best matrix used current allele frequencies and was scaled similarly to the numerator relationship matrix for genotyped animals.

The first goal of this study was to evaluate the effect of different options to construct **G** on statistics [means of the diagonal (**AvgD**) and off-diagonal (**AvgOff**)] of **G** and its inverse, and the accuracy of EBV and biases. The second goal was to investigate the origin of biases with some **G**. Broiler chickens from a pure line with 1.5% animals genotyped for 57,636 SNP were used to evaluate the different corrections on **G**.

## MATERIALS AND METHODS

Animal Care and Use Committee approval was not obtained for this study because the data were obtained from an existing database.

### Data

Phenotypic data of broiler chickens from a single pure line across 4 generations were provided by Cobb-Vantress Inc. (Siloam Springs, AR). Traits analyzed included BW at 6 wk (BW, 100 g) and ultrasound area of breast meat (**BM**, cm²). A total of 287,614 broilers were available for phenotypes. Body weight was recorded for all birds, whereas BM was recorded for only 69,057 broilers. There were 289,994 birds in the pedigree.

A total of 4,113 birds were genotyped for 57,636 SNP based on the SNP panel developed by Groenen et al. (2009). Descriptions of phenotypic records for genotyped birds and all birds are shown in Table 1.

The data set was split into training and validation data sets for the genetic prediction. The training data set consisted of 234,806 records in the first 3 generations, whereas the validation data set contained 806 genotyped birds in the last generation.

### Model

In our previous study (Chen et al., 2011), the single-trait model was implemented for a similar data set. In the current study, the multiple-trait model was

$$\mathbf{y_t} = \mathbf{X_t b_t} + \mathbf{Z_t u_t} + \mathbf{W_t mp_t} + \mathbf{e_t},$$

where **t** is 1 for BW and 2 for BM; **y** is the vector of observations for traits; **b** is the vector of fixed effects, including contemporary group (house-hatch) and sex; **u** is the vector of random additive genetic effects, combining polygenic and genomic breeding values; **mp** is the vector of random maternal permanent environmental effects; **X**, **Z**, and **W** are incidence matrices; and **e** is the vector of random residuals.

Effects of the multiple-trait model used in the present paper were the same as those used in the single-trait model by Chen et al. (2011), except that the maternal permanent environmental effect was also included for BM in the present paper. In Chen et al. (2011), the abnormality of accuracy obtained with a single-trait model may have been due to selection bias resulting from multiple-trait selection. For the next run of selection using the multiple-trait model, the anomaly was absent. Furthermore, the accuracy for a trait with missing records, BM, increased slightly with the multiple-trait model. Therefore, the multiple-trait model was used in the current study instead of the single-trait model.

For the regular BLUP analysis, the (co)variance matrix was assumed to be

$$\text{var}\begin{bmatrix} \mathbf{u} \\ \mathbf{mp} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \otimes \mathbf{D} & 0 & 0 \\ 0 & \mathbf{I} \otimes \mathbf{Q} & 0 \\ 0 & 0 & \mathbf{I} \otimes \mathbf{R} \end{bmatrix},$$

where **I** is an identity matrix, **A** is a numerator relationship matrix, and **D**, **Q**, and **R** are matrices of the additive, maternal permanent, and residual variances, respectively. In the single-step genomic BLUP (**ssG-BLUP**) proposed by Misztal et al. (2009), the **A** matrix was replaced by the **H** matrix with the following inverse (Aguilar et al., 2010):

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

where **H** is a modified relationship matrix incorporating genomic information, $\mathbf{A}_{22}^{-1}$ corresponds to the inverse of the numerator relationship matrix for genotyped birds, and **G** is a genomic relationship matrix.

Genomic relationship matrices were constructed as in VanRaden (2008) by using either equal (**GEq**) or current (**GC**) allele frequencies:

$$\mathbf{G} = \frac{\mathbf{QQ'}}{2\sum p_j(1-p_j)},$$

where $\mathbf{Q}$ is an incidence matrix for SNP effects with elements

$$\mathbf{q}_{ij} = \begin{cases} 0 - 2p_j & \text{if homozygous 11} \\ 1 - 2p_j & \text{if heterozygous 12 or 21,} \\ 2 - 2p_j & \text{if homozygous 22} \end{cases}$$

for bird i and SNP j with allele frequency $p_j$.

These matrices were constructed with the MAF ranging from none to 0.4. The purpose of the thresholds was to determine the effect of eliminating SNP with low MAF on GEq and GC, and to examine the effect of the smaller SNP size.

Estimates of variance components were obtained using the multiple-trait model with the complete data set, which included records across 4 generations. Genetic evaluations were performed by modified BLUP90IOD (Tsuruta et al., 2001; Misztal et al., 2002; Aguilar et al., 2010) using BLUP (no genomic information) and ssGBLUP (with genomic information). Predictive ability, $r(\hat{\mathbf{u}}, \mathbf{u} + \mathbf{mp} + \mathbf{e})$, was defined as the correlation between predicted breeding value and the observed phenotype, composed of the sum of the true breeding value, maternal permanent environmental effect, and residual, using the formula shown by Legarra et al. (2008). Accuracy, or correlation between predicted and true breeding values, was calculated as $r(\hat{\mathbf{u}}, \mathbf{u}) = r(\hat{\mathbf{u}}, \mathbf{u} + \mathbf{mp} + \mathbf{e}) / \mathbf{h}$, where $\mathbf{h}$ is the square root of heritability.

## RESULTS AND DISCUSSION

Estimates of the additive variance (heritability) were 1.3 (0.23) for BW and 5.0 (0.25) for BM. Estimates of the genetic correlations between BW and BM were 0.4. Table 2 shows the AvgD and AvgOff of $\mathbf{G}$ constructed with equal or current allele frequencies and with different cutoffs for MAF. Averages for $\mathbf{A}_{22}$ are presented for reference. With GEq, as MAF increased to 0.4, AvgD decreased from 1.37 to 1.02 and AvgOff decreased from 0.72 to 0.02. The differences from 1.0 and 0.0, respectively, decreased in GEq as the threshold was increased because the remaining SNP had average allele frequencies closer to 0.5. The differences between AvgD and AvgOff increased from 0.65 to 1.0. This difference, which was smaller than 1.0, was found by Forni et al. (2011) to be closely associated with upward biases in estimates of the additive variance. With GC, AvgD was close to 1.0 and AvgOff was 0, regardless of the threshold. This was expected because the contribution of SNP

**Table 2.** Means of the diagonal (AvgD) and off-diagonal (AvgOff) for the pedigree-based relationship matrix $\mathbf{A}_{22}$ and the genomic relationship matrix $\mathbf{G}$ for genotyped birds (n = 4,113)

| Item[1] | No. of SNP | AvgD | AvgOff | Diff[2] |
|---|---|---|---|---|
| $\mathbf{A}_{22}$ | — | 1.004 | 0.014 | 0.990 |
| **G** | | | | |
| GEq | | | | |
| All SNP | 57,636 | 1.373 | 0.722 | 0.651 |
| MAF0.1 | 41,749 | 1.197 | 0.368 | 0.829 |
| MAF0.2 | 32,885 | 1.121 | 0.217 | 0.904 |
| MAF0.3 | 22,619 | 1.062 | 0.101 | 0.961 |
| MAF0.4 | 11,253 | 1.020 | 0.025 | 0.995 |
| GC | | | | |
| All SNP | 57,636 | 1.017 | 0 | 1.017 |
| MAF0.1 | 41,749 | 1.015 | 0 | 1.015 |
| MAF0.2 | 32,885 | 1.013 | 0 | 1.013 |
| MAF0.3 | 22,619 | 1.012 | 0 | 1.012 |
| MAF0.4 | 11,253 | 1.008 | 0 | 1.008 |

[1]GEq = equal allele frequencies; MAF = minor allele frequencies; GC = current allele frequencies.
[2]Diff = AvgD − AvgOff.

to off-diagonals of $\mathbf{G}$ was proportional to the allelic frequencies, and they roughly canceled out. This was also the case with GEq for MAF >0.4, in which case the assumed frequency of 0.5 roughly corresponded to the true frequency. In addition, the differences between the averages were close to 1.0.

Table 3 shows the AvgD and AvgOff of the inverse of $\mathbf{G}$, with values for the inverse of $\mathbf{A}_{22}$ presented for reference. In $\mathbf{G}^{-1}$, all AvgOff were close to 0 and AvgD were between 8.6 and 12.9. For comparison, in $\mathbf{A}_{22}^{-1}$, AvgOff were close to 0 and AvgD were 2.37. With GEq, the greatest of AvgD was at MAF extremes. With GC, AvgD increased with the MAF cutoff.

The diagonal of the inverse of the numerator relationship matrix $\mathbf{A}_{22}$ has the following form:

$$a^{ii} = 4/(4 - n_p) + n_{pr1}/2 + n_{pr2}/3,$$

where $n_p$ is the number of known parents of bird i, $n_{pr1}$ is the number of progeny with the mate known, and $n_{pr2}$ is the number of progeny with the mate unknown. Assuming that differences between AvgD in the inverses of $\mathbf{A}_{22}$ and $\mathbf{G}$ are due to additional information in $\mathbf{G}$ relative to $\mathbf{A}_{22}$ as well as to additional factors, we can write

$$g^{ii} = a^{ii} + q_i/2 + e_i,$$

where $q_i$ is the extra information attributable to the genomic information in units of effective progeny, and $e_i$ is noise attributable to both incorrect scaling and the limited number of SNP.

When $\mathbf{G}$ is "adequately" scaled and derived from a large number of SNP, quantity $q_i$ can be estimated well because $e_i$ is likely close to 0. In separate experiments

**Table 3.** Means of the diagonal (AvgD) and off-diagonal (AvgOff) for the inverse of the pedigree-based relationship matrix $\mathbf{A}_{22}$ and the genomic relationship matrix $\mathbf{G}$ for genotyped birds (n = 4,113)

| Item[1] | No. of SNP | AvgD | AvgOff | Diff[2] |
|---|---|---|---|---|
| $\mathbf{A}_{22}^{-1}$ | — | 2.375 | −0.001 | 2.376 |
| $\mathbf{G}^{-1}$ | | | | |
| GEq | | | | |
| All SNP | 57,636 | 11.563 | −0.003 | 11.566 |
| MAF0.1 | 41,749 | 10.321 | −0.003 | 10.324 |
| MAF0.2 | 32,885 | 10.335 | −0.003 | 10.338 |
| MAF0.3 | 22,619 | 10.903 | −0.003 | 10.906 |
| MAF0.4 | 11,253 | 12.943 | −0.003 | 12.946 |
| GC | | | | |
| All SNP | 57,636 | 8.675 | −0.002 | 8.677 |
| MAF0.1 | 41,749 | 9.000 | −0.002 | 9.002 |
| MAF0.2 | 32,885 | 9.576 | −0.002 | 9.578 |
| MAF0.3 | 22,619 | 10.555 | −0.002 | 10.557 |
| MAF0.4 | 11,253 | 12.859 | −0.003 | 12.862 |

[1]GEq = equal allele frequencies; MAF = minor allele frequencies; GC = current allele frequencies.

[2]Diff = AvgD − AvgOff.

(results not shown), AvgD increased as the number of SNP was reduced. Thus, when the number of SNP used to create $\mathbf{G}$ decreased, the absolute value of $e_i$ increased. Assuming that GEq was scaled correctly, the increase in AvgD with the increasing threshold was likely due to extra noise in estimating $\mathbf{G}$ when the number of SNP was reduced. As the number of SNP was decreased, the numerical accuracy of $\mathbf{G}$ was decreased, creating "fake" relationships. Therefore, $e_i$ may have been reflecting information for false relationships when $\mathbf{G}$ was inaccurate (computed with too few SNP).

Yang et al. (2010) looked at underestimation of the additive variance by assuming that causal variants have small allele frequencies but that the SNP information is restricted to SNP with larger MAF. They used unrelated individuals and SNP information from 50 to 500 k and provided corrections for $\mathbf{G}$ based on MAF and the number of SNP. Corrections by Yang et al. (2010) did not improve the accuracy in the study by Forni et al. (2011).

Table 4 shows the predictive ability for the validation data with different $\mathbf{G}$. The predictive abilities were very similar across all $\mathbf{G}$. For equal allele frequencies, the predictive abilities reached the peak at the threshold of 0.1 to 0.3 for BW, and increased with the threshold for BM. Increasing the threshold decreased the amount of information in $\mathbf{G}$ but brought the average allele frequency close to 0.5. The somewhat greater accuracy for the cutoff of 0.4 in BM could be due to a major gene in the trait that has MAF close to 0.5 and then reduced noise in the estimation with a smaller number of SNP. Greater accuracies at larger MAF are in contrast with the assumptions of Yang et al. (2010). However, they looked for causative genes in a population of unrelated

individuals, whereas $\mathbf{G}$ here reflects improved pedigrees of closely related individuals.

With the current allele frequencies, the accuracies were the same for the cutoff <0.4 for BW, with a peak for BM at the threshold of 0.4. With GEq, the accuracies could be greater at low thresholds because the off-diagonals of $\mathbf{G}$ were not affected by incorrect computations of contributions from SNP with small allele frequencies.

In the dairy cattle population (Aguilar et al., 2010), the accuracy was greater with GEq than with GC. The reverse was true here. In dairy cattle, the parent average for young bulls was biased even without the genomic information, with the bias attributable to the preferential treatment and specific population structure of the dairy cattle population. It is possible that this bias was partially compensated by bias introduced by GEq. The population structure in chickens is much simpler than that in dairy cattle.

Moser et al. (2010) looked at the accuracy of GEBV for several dairy traits by using subsets of SNP information selected by different methods. The subset of 5,000 SNP allowed >90% accuracy to be achieved compared with the complete SNP information. Although the greatest accuracy for each trait was obtained with trait-specific SNP, very few of the highest ranked SNP for 1 trait were highest ranked for other traits. The accuracy with 5,000 SNP selected by MAF alone was greater for cows but was reduced for bulls. In this study, we experienced a marginal decline in accuracy below 22,000 SNP for BW, but not for BM.

Table 5 shows the average EBV for the validation population and for the complete population. If we assume that BLUP provides unbiased EBV, then, on average, all EBV with GEq are biased upward and all

**Table 4.** Predictive ability[1] with BLUP and single-step genomic BLUP (ssGBLUP) for BW and breast meat area (BM)

| Item[2] | BW | BM |
|---|---|---|
| BLUP | 0.19 | 0.20 |
| ssGBLUP | | |
| GEq | | |
| All SNP | 0.28 | 0.34 |
| MAF0.1 | 0.29 | 0.35 |
| MAF0.2 | 0.29 | 0.35 |
| MAF0.3 | 0.29 | 0.36 |
| MAF0.4 | 0.28 | 0.37 |
| GC | | |
| All SNP | 0.29 | 0.36 |
| MAF0.1 | 0.29 | 0.36 |
| MAF0.2 | 0.29 | 0.37 |
| MAF0.3 | 0.29 | 0.37 |
| MAF0.4 | 0.28 | 0.38 |

[1]Predictive ability is defined as correlations between the predicted breeding value and the sum of the true breeding value, maternal effect, and residual.

[2]GEq = equal allele frequencies; MAF = minor allele frequencies; GC = current allele frequencies.

**Table 5.** Means of EBV with BLUP and single-step genomic BLUP (ssGBLUP) for genotyped birds (n = 4,113) and all birds (n = 289,994) for BW and breast meat area (BM)

| Item[1] | Genotyped | | All | |
|---|---|---|---|---|
| | BW | BM | BW | BM |
| BLUP | 2.77 | 2.27 | 1.67 | 1.38 |
| ssGBLUP | | | | |
| GEq | | | | |
| All SNP | 3.94 | 3.21 | 2.51 | 2.15 |
| MAF0.1 | 3.89 | 3.21 | 2.54 | 2.16 |
| MAF0.2 | 3.82 | 3.17 | 2.51 | 2.13 |
| MAF0.3 | 3.61 | 3.02 | 2.38 | 2.03 |
| MAF0.4 | 3.00 | 2.53 | 1.93 | 1.67 |
| GC | | | | |
| All SNP | 0.45 | 0.32 | −0.05 | −0.09 |
| MAF0.1 | 0.46 | 0.32 | −0.05 | −0.10 |
| MAF0.2 | 0.46 | 0.32 | −0.05 | −0.10 |
| MAF0.3 | 0.46 | 0.32 | −0.05 | −0.09 |
| MAF0.4 | 0.46 | 0.32 | −0.04 | −0.08 |

[1]GEq = equal allele frequencies; MAF = minor allele frequencies; GC = current allele frequencies.

EBV with GC are biased downward. The biases are approximately 2 additive SD for BW and 1 SD for BM. Additionally, on average, the difference in EBV between genotyped birds and all birds appeared biased upwards with GEq and downwards with GC. Thus, the biases seemed to be dependent on the structure of $\mathbf{G}$. In particular, this could have been due to differences in AvgOff between $\mathbf{G}$ and $\mathbf{A}_{22}^{-1}$.

In the single-step method

$$\mathrm{var}(\mathbf{u}) = \mathbf{H},$$

where

$$\mathbf{H} = \begin{bmatrix} \mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G} \\ \mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G} \end{bmatrix}.$$

Assume that the genomic relationship matrix is computed with an offset as $\mathbf{G}^* = \mathbf{G} + \mathbf{11}'\alpha$, where $\alpha$ is a constant and $\mathbf{1}$ is a vector of ones (that is, adding $\alpha$ to all elements of $\mathbf{G}$). The new matrix $\mathbf{H}^*$ will be defined as in [1].

This is equivalent to fitting a model where $u_i^* = u_i + \delta_i\mu$, $\mathrm{var}(\mu) = \alpha$, where $\delta_i$ is 1 for genotyped birds ($\mathbf{u}_g$) and 0 for the remaining birds; thus,

$$\mathbf{G}^* = \mathrm{var}(\mathbf{u}_g^*) = \mathrm{var}(\mathbf{u}_g + \mathbf{1}\mu)$$

$$= \mathrm{var}(\mathbf{u}_g) + \mathbf{11}'\mathrm{var}(\mu) = \mathbf{G} + \mathbf{11}'\alpha.$$

In this case, $\mathbf{u}^*$ contains the mean of all genotyped birds. That mean is automatically applied to the EBV of young genotyped birds. Therefore, if the genotyped birds are selected, their difference from the mean is

added to all new genotyped birds, causing bias. Although the value $\alpha$ treated as variance causes regression of the estimate of $\mu$ toward zero, that regression is small for all but very small $\alpha$ because of the quantity of information that many genotyped birds contribute to a single variable $\mu$.

To test the hypothesis that a mean of $\mathbf{G}$ would affect bias in ssGBLUP, different $\alpha$ were applied to $\mathbf{G}$ with the current allele frequencies and all SNP. Only results for positive $\alpha$ are presented because $\mathbf{G}$ with $\alpha < 0$ was not positive definite (and does not make sense as the variance of $\mu$). Average EBV for genotyped birds and all birds, and for their differences, are presented in Table 6. As $\alpha$ was increased to 0.05, the averages for all birds changed by about 1 SD for both BW and BM. Thus, the averages, and thus biases (compared with BLUP), were strongly influenced by AvgOff. All averages were similar to those with BLUP at $\alpha = 0.014$. With this offset, AvgOff in $\mathbf{A}_{22}$ and $\mathbf{G} + \mathbf{11}'\alpha$ were identical. This suggests that the optimal $\mathbf{G}$ should have AvgD and AvgOff close to that of $\mathbf{A}_{22}$. Although similar AvgD − AvgOff in $\mathbf{G}$ and $\mathbf{A}_{22}$ ensured unbiased estimates of the additive variances, identical AvgOff seemed to remove biases for the EBV of genotyped birds. The average differences between the EBV of genotyped birds and all birds as a function of a difference between the AvgOff elements of $\mathbf{G}$ and $\mathbf{A}_{22}$ are shown in Figure 1 (using BW). The smallest bias was with the MAF cutoff of 0.4. In that case, both the AvgD and AvgOff of $\mathbf{G}$ were the closest to those in $\mathbf{A}_{22}$.

Adding $\alpha$ had a very small effect on the predictive abilities. The predictive ability for BW was 0.29 for any $\alpha$. The predictive ability for BM was 0.36 for $\alpha < 0.015$, and was decreased to 0.35 for larger $\alpha$. Thus, the offset in $\mathbf{G}$ had a negligible effect on the accuracy of the validation population in this study.

The bias attributable to the offset in $\mathbf{G}$ mainly affects the genotyped birds, although it also affects the ungenotyped birds through the $\mathbf{H}$ matrix. The size of that bias depends on the strength of selection, the superiority of the genotyped as opposed to the ungenotyped birds, and the amount of information in older generations. In populations under weak or no selection, and when genotyped birds are a random sample from the population, the genotyped birds may have an average merit close to that of ungenotyped birds. Thus, any bias incurred by the offset of $\mathbf{G}$ is likely to be small or none. Under strong selection, the offset will bias the EBV of young genotyped birds. Additionally, it may reduce the accuracy of EBV of the older birds and, subsequently, their progenies if the genomic information is a large fraction of the total information. In this study, the offset affected only the bias and not the accuracy of the newest generation because the phenotypic information for older birds was large (about 100 times more records than genotypes). The stronger bias for BM suggests that selection for this trait was stronger than for BW.

Vitezica et al. (2010) looked at the biases and accuracy of BLUP, ssGBLUP, and a 2-step method with

Chen et al.

**Table 6.** Means of EBV for genotyped birds (n = 4,113) and all birds (n = 289,994) with various differences in the average off-diagonal (AvgOff) between the genomic relationship matrix **G** and the pedigree-based relationship matrix $\mathbf{A}_{22}$ for BW and breast meat area (BM)

| Item[1] | AvgOff | | BW | | | BM | | |
|---|---|---|---|---|---|---|---|---|
| | $\mathbf{G}(\alpha)$[2] | $\mathbf{G} - \mathbf{A}_{22}$[3] | Genotyped | All | Diff[4] | Genotyped | All | Diff |
| BLUP | — | — | 2.77 | 1.67 | 1.10 | 2.27 | 1.38 | 0.89 |
| ssGBLUP (GC) | 0 | −0.014 | 0.45 | −0.05 | 0.50 | 0.32 | −0.09 | 0.41 |
| | 0.003 | −0.011 | 1.51 | 0.78 | 0.73 | 1.23 | 0.63 | 0.60 |
| | 0.007 | −0.007 | 2.28 | 1.37 | 0.91 | 1.86 | 1.12 | 0.74 |
| | 0.014 | 0 | 2.87 | 1.83 | 1.04 | 2.39 | 1.54 | 0.85 |
| | 0.050 | 0.036 | 3.47 | 2.28 | 1.19 | 2.93 | 1.96 | 0.97 |
| | 0.100 | 0.086 | 3.69 | 2.45 | 1.24 | 3.08 | 2.08 | 1.00 |
| | 0.200 | 0.186 | 3.81 | 2.54 | 1.27 | 3.17 | 2.15 | 1.02 |

[1]ssGBLUP = single-step genomic BLUP; GC = current allele frequencies with all SNP.

[2]Genomic matrix with $\alpha$ ranging from 0 to 0.2.

[3]AvgOff of $\mathbf{A}_{22} = 0.014$.

[4]Diff = means of EBV for genotyped birds − means of EBV for all birds.

simulated data under weak and strong selection. The data consisted of 10 generations of animals with 10% of animals genotyped with **G** constructed using base allele frequencies. The EBV by BLUP was always unbiased. The accuracy by ssGBLUP was greater by the 2-step method under weak selection, but it was less under strong selection. After adjusting the offset in **G**, the accuracy of ssGBLUP increased, especially under strong selection, and exceeded that of the 2-step method (Vitezica et al., 2011). They also provided a formal proof for the calculation of the offset and showed that it was the same as that developed in human studies by Powell et al. (2010). Therefore, the correct offset of **G** is critical for small populations, especially under strong selection when the genomic information is a large fraction of the population and where progeny sizes are small. The use of adjusted **G** was also evaluated for Holsteins for type by using a large number of phenotypes and genotypes (Tsuruta et al., 2011). The bias removed by adjusting the offset in **G** was small because of the large progeny sizes for bulls.

The population in this study consisted of a single line of a single breed. When 2 lines of chickens were evaluated simultaneously, even though using different **G**, the rankings within each line were almost identical but the means of the lines were different (Simeone et al., 2011). This was due to different allele frequencies and different selection pressures in each line. In a joint evaluation of multiple lines or multiple breeds, sections of **G** corresponding to each line and each cross may require separate adjustments (e.g., as in Harris and Johnson, 2010).

Removing SNP with MAF $\leq 0.3$ in a genomic evaluation of a commercial chicken population changed the realized accuracy of prediction only marginally. Those accuracies were slightly greater when assuming the current rather than the equal gene frequencies, especially when SNP with low MAF were retained. Removing alleles with low MAF increased the average elements of the inverse of **G** created with the current allele frequencies. This increase attributable to false relationships introduced by less accurate **G** caused inflated accuracies computed by inversion. In the genomic evaluation, differences in average EBV of birds with and without the genomic information varied with **G** and the threshold of MAF. These differences depended on the level of selection and the data structure, and they could be controlled by adding a constant to **G** constructed by assuming current allele frequencies. When AvgOff elements in $\mathbf{A}_{22}$ and **G** were identical, the differences were the same as in BLUP. Assuming that BLUP is unbiased, an unbiased genomic evaluation could be obtained by adding a constant to **G** derived by assuming current allele frequencies.
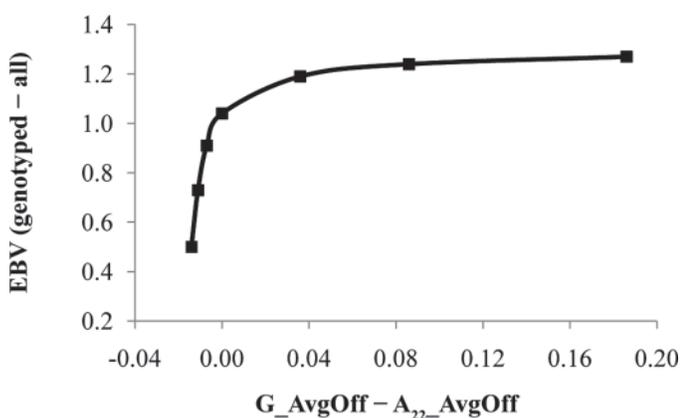


**Figure 1.** The average difference between the EBV of genotyped birds and all birds as a function of the difference between mean off-diagonal (AvgOff) elements of the genomic (**G**) and relationship ($\mathbf{A}_{22}$) matrices. The average difference between the EBV of genotyped birds and all birds in BLUP was 1.10.

## LITERATURE CITED

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752.

Chen, C. Y., I. Misztal, I. Aguilar, S. Tsuruta, T. H. E. Meuwissen, S. E. Aggrey, T. Wing, and W. M. Muir. 2011. 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.

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.

Groenen, M. A. M., P. Wahlberg, M. Foglio, H. H. Cheng, H. J. Megens, R. P. M. A. Crooijmans, F. Besnier, M. Lathrop, W. M. Muir, G. K. Wong, I. Gut, and L. Andersson. 2009. A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. Genome Res. 19:510–519.

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. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92:433–443.

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., C. Robert-Granié, E. Manfredi, and J. M. Elsen. 2008. Performance of genomic selection in mice. Genetics 180:611–618.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92:4648–4655.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. H. Lee. 2002. BLUPF90 and related programs (BGF90). Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France. Commun. No. 28-07.

Moser, G., M. Khatkar, B. Hayes, and H. Raadsma. 2010. Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. Genet. Sel. Evol. 42:37.

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.

Simeone, R., I. Misztal, I. Aguilar, and Z. Vitezica. 2011. Evaluation of a multi-line broiler chicken population using a single-step genomic evaluation procedure. J. Anim. Breed. Genet. (Accepted)

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.

Tsuruta, S., I. Misztal, and I. Stranden. 2001. Use of the preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications. J. Anim. Sci. 79:1166–1172.

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

VanRaden, P. M., M. E. Tooker, and J. B. Cole. 2009a. Can you believe those genomic evaluations for young bulls? J. Dairy Sci. 92(E-Suppl. 1):314. (Abstr.)

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

Vitezica, Z. G., I. Aguilar, and A. Legarra. 2010. One-step vs. multi-step methods for genomic prediction in presence of selection. Proc. 9th World Congr. Genet. Appl. Livest. Prod., Leipzig, Germany.

Vitezica, Z. G., I. Aguilar, I. Misztal, and A. Legarra. 2011. Selection bias in genomic predictions of single-step and multi-step methods. Genet. Res. (Accepted)

Yang, J., B. Benyamin, B. P. McEvoy, S. D. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard, and P. M. Visscher. 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42:565–569.

**References**

This article cites 17 articles, 4 of which you can access for free at:
http://www.journalofanimalscience.org/content/89/9/2673#BIBL