ORIGINAL ARTICLE

Evaluation of the utility of diagonal elements of the genomic relationship matrix as a diagnostic tool to detect mislabelled genotyped animals in a broiler chicken population

R. Simeone[1], I. Misztal[1], I. Aguilar[1,2] & A. Legarra[3]

1 Department of Animal and Dairy Science, University of Georgia, Athens, USA
2 Instituto Nacional de Inverstigación Agropecuaria, Las Brujas, Uruguay
3 L'Institut National de la Recherche Agronomique, Castanet-Tolosan, France

Summary

This study explored distributions of diagonal elements of genomic relationship matrix ($\mathbf{G}$), evaluated the utility of $\mathbf{G}$ as a diagnostic tool to detect mislabelled animals in a genomic dataset and evaluated the effect of mislabelled animals on the accuracy of genomic evaluation. Populations of 10 000 animals were simulated with 60 000 SNP varying in allele frequency at each locus between 0.02 and 0.98. Diagonal elements of $\mathbf{G}$ were distributed with a single peak (mean = 1.00 ± 0.03) and ranged from 0.84 through 1.36. Mixed populations were also simulated: 7 000 animals with frequencies of second alleles ranging from 0.02 through 0.98 were combined with 1750 or 7000 animals with frequencies of second alleles ranging from 0.0 through 1.0. The resulting distributions of diagonal elements of $\mathbf{G}$ were bimodal. Body weight at 6 weeks was provided by Cobb-Vantress for broiler chickens, of which 3285 were genotyped for 57 636 SNP. Analysis used a combined genomic and pedigree relationship matrix; $\mathbf{G}$ was scaled using current allele frequencies. The distribution of diagonal elements was multimodal and ranged from 0.54 to 3.23. Animals with diagonal elements >1.5 were identified as coming from another chicken line or as having low call rates. Removal of mislabelled animals increased accuracy by 0.01. For the studied type of population, diagonal elements of $\mathbf{G}$ may be a useful tool to help identify mislabelled animals or secondary populations.

Introduction

Genomic selection has been used successfully in dairy cattle and other species (Hayes *et al.* 2009a; VanRaden *et al.* 2009; Chen *et al.* 2010). Genomic proofs for young animals may be almost as accurate as progeny-tested animals (Meuwissen *et al.* 2001; Schaeffer 2006). Initially, genomic predictions were derived by estimating the effects of SNP markers; however, predictions by genomic BLUP with a geno-mic relationship matrix ($\mathbf{G}$) may have similar or increased accuracy (Hayes *et al.* 2009b).

$\mathbf{G}$ may be viewed as a matrix of realized relationships. When constructed from large SNP arrays, $\mathbf{G}$ is likely to reflect real relationships better than a pedigree-based relationship matrix ($\mathbf{A}$) because it contains information on unrecorded pedigrees and on the Mendelian sampling (Hayes *et al.* 2009b). $\mathbf{G}$ is scaled to resemble a pedigree relationship matrix; the coefficients of the matrix are 0.0 to 1.0 on the

off-diagonals, forcing diagonal elements to be 1.0 (VanRaden 2008). Scaling can be performed using current allele frequencies, and estimates of additive variance are similar to those obtained through traditional evaluation (Forni *et al.* 2010).

Allele frequencies have been used to identify population structure in admixed populations (Pritchard *et al.* 2000; Rosenberg *et al.* 2001). Harris & Johnson (2010) have indicated that multi-breed populations will have animals with distorted diagonal elements if breed is not taken into account. Wiggans *et al.* (2010) have shown that methods exist to detect mislabelled samples within and across populations using allele frequencies and off-diagonal elements of **G**. Another method of population identification may therefore be to use the diagonal elements of **G** to separate populations. Mislabelled animal can occur for a variety of reasons: incorrect reporting of a wrong animal's population or line, obtaining DNA from the wrong animal, switching DNA samples in the laboratory or other data processing errors. It is unknown how small secondary populations or mislabelled animals belonging to a line that has been under different selection pressures for many generations than the line of interest will be expressed in the diagonal elements of **G** or how they might impact predictions in genomic analysis. Moreover, the theoretical distribution of diagonal elements of **G** for single and multiple populations is unknown. The objectives of this study were to explore by simulation the distributions of the diagonal elements of **G** for single and multiple populations, to evaluate the utility of **G** as a diagnostic tool in the detection of different populations within a genomic dataset and to evaluate the effect of mislabelled genotyped animals on the accuracy of genomic evaluation.

Materials and methods

Theoretical distribution of diagonal elements of G

Assume that all genotyped biallelic loci are present with the same allele frequency, p = 0.5. If genotypes are coded as 0, 1 and 2, **M**, a matrix of marker alleles, $= \begin{cases} 0 \\ 1 \\ 2 \end{cases}$ and **P**, equal to two times the second allele frequency, $= \begin{cases} 1 \\ 1 \\ 1 \end{cases}$ (VanRaden 2008). Matrix **Z**, an incidence matrix for SNP effects, is then equal to **M**−**P**; for each possible genotype, $\mathbf{Z} = \begin{cases} -1 \\ 0 \\ 1 \end{cases}$ and

under Hardy–Weinberg equilibrium , genotype frequencies are $\begin{matrix} aa \\ aA \\ AA \end{matrix} = \begin{cases} 1/4 \\ 1/2 \\ 1/4 \end{cases}$. Then, $E(z^2) = \frac{1}{2}$ and $Var(z^2) = \frac{1}{4}$. It follows that $E(\sum_{i=1}^{n} z_i^2 = \frac{n}{2})$ and $Var(\sum_{i=1}^{n} z_i^2 = \frac{n}{4})$, where n is the number of SNP. Therefore, $\sum z_i^2 \approx N(\frac{n}{2}, \frac{n}{4})$; scaling z_i^2 for a diagonal of 1.0 gives $\sum \frac{z_i^2}{\frac{n}{2}} \approx N(1, \frac{1}{n})$. This suggests that the diagonal elements of **G** should be distributed normally with a sharp peak and variance decreasing with increased numbers of SNPs. Theoretical distributions of diagonal elements of **G** for 60 000 SNPs assuming equal allele frequencies at each locus are presented in Figure 1. With differing allele frequencies at each locus, the distribution widens but remains normally distributed. Multiple peaks would not occur.

Simulated data

Populations were simulated to find the distribution of the diagonal elements of **G**. Each population had either three or six generations, of which 10% were male and had a total of 10 000 animals. The three-generation population consisted of one unrelated base population of 3000 animals and two more generations of 3000 and 4000 animals. The six-generation population consisted of one unrelated base population of 1000 animals and five more generations of 1000, 1500, 1500, 2000, and 3000 animals. Mating was random; dams could have multiple progeny, and sires and dams could be used as parents in multiple generations. The genome was simulated with 60 000 SNP. The frequency of the second allele at each locus varied uniformly between 0.02 and 0.98, 0.1 and 0.9, 0.25 and 0.75, and 0.45 and 0.55.
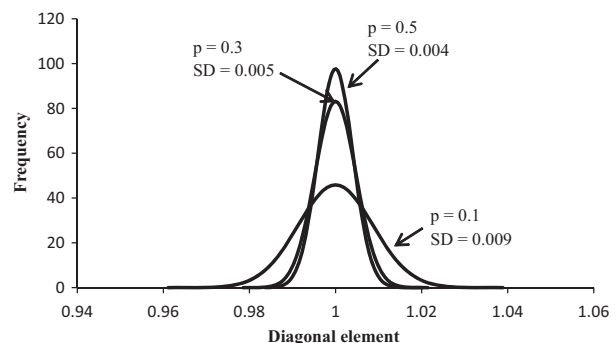
Figure 1 Theoretical distribution of diagonal elements of G for 60 000 SNP assuming equal allele frequencies (p) at each locus.

Progeny genotypes were created by selecting one of each of the parental chromosomes. Recombination rate between loci varied between 0.02 and 0.1.

Four mixed populations were simulated with either three or six generations. First, a 7000 animal pedigree was simulated with the frequency of the second allele ranging from 0.02 through 0.98. Second, a 1750 or 7000 animal pedigree was simulated with the frequency of the second allele ranging from 0.0 through 1.0. Third, the secondary populations were added to the original 7000 animal populations and resulted in two datasets with either 8750 animals (G3_1750, G6_1750) or 14 000 animals (G3_7000, G6_7000). For each population, the frequency of the second allele at each locus of the combined population was calculated.

Field data

Body weight at 6 weeks (100 g) was provided by Cobb-Vantress for broiler chickens in three generations. Details of the dataset and methods used in the analysis are provided in Chen *et al.* (2010). Data consisted of 3285 animals genotyped for 57 636 SNP. Monomorphic loci or loci with an allele frequency <0.02 were removed from that data, leaving 48 006 SNP for the analysis.

For the initial analysis, data were divided into a training set, consisting of all phenotyped and genotyped animals from the first two generations, n = 2485, and a validation set, consisting of phenotyped and genotyped animals from the third generation, n = 799. This was known as ALL data. Subsequent analysis involved removing observations of animals with high diagonal elements from the data and regrouping into training (n = 2397) and validation (n = 798). This was known as CLEANED data.

Models and analysis

Body weight was analysed using an animal model in a single-step procedure (Aguilar *et al.* 2010) as in Chen *et al.* (2010). The genomic relationship matrix was constructed as $\mathbf{G} = zz'/2\sum p_i(1 - p_i)$ using current allele frequencies from the population; this scales $\mathbf{G}$ so that the average of the diagonal elements is 1.0 (VanRaden 2008). The distributions of the diagonal elements of $\mathbf{G}$ were plotted for each dataset. A 99% range of diagonal elements for simulated data was computed by multiplying the SD of the diagonal elements of $\mathbf{G}$ by 2.567, from the normal distribution. To visualize differences in the genotypes

that might cause differences in the diagonal elements of $\mathbf{G}$, the distributions of the second allele frequencies were plotted.

Predictions of EBV were made for the validation data using the training data for ALL and CLEANED. Accuracy was defined as the correlation between the predicted breeding value and the true breeding value: $r(\hat{u}, u) = \frac{r(\hat{u}, u+e)}{h}$, where h is the square root of the heritability and u is obtained as an animal's phenotypic observation minus fixed effects (Legarra *et al.* 2008; Chen *et al.* 2010).

Results and discussion

The distribution of diagonal elements for six generations of animals with allele frequencies ranging from 0.02 through 0.98 is shown in Figure 2. The mean of this distribution is 1.00 (0.03), corresponding to the mean of the diagonal elements in the traditional relationship matrix, $\mathbf{A}$. Diagonal values range from a minimum of 0.88 to a maximum of 1.36; 99% of the diagonal elements should fall within 0.91 and 1.09. Statistics of the diagonal distributions of multiple populations were examined to ensure that the distribution was consistent regardless of population number, generation size or allele frequency (Table 1). Little variation in the distribution of the diagonal elements of $\mathbf{G}$ is seen within populations simulated over three or six generations. Diagonal elements of animals simulated over three generations ranged from 0.98 to 1.26, suggesting minimal inbreeding or change in allele frequencies, which
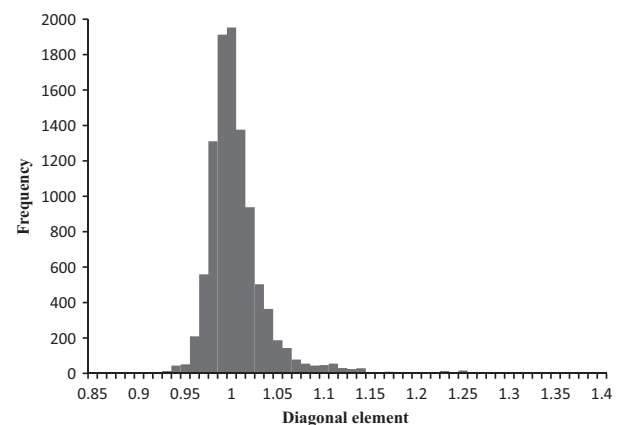
Figure 2 Distribution of diagonal elements for six generations of animals (n = 10 000). Base population (n = 1000) was assumed to be unrelated; subsequent populations were generated based on parents in previous generations (n = 1000, 1500, 1500, 2000, 3000). The genome was simulated with 60 000 SNP ranging in frequency between 0.02 and 0.98, with recombination frequency between 0.02 and 0.1.

Table 1 Statistics of the diagonal distributions of multiple populations[1]

No. generations	Allele frequencies	Mean (SD)	Min	Max	99% range
3[2]	0.02–0.98	1.00 (0.01)	0.98	1.25	0.98–1.02
	0.25–0.75	1.00 (0.01)	0.98	1.26	0.98–1.02
	All frequencies	1.00 (0.01)	0.98	1.26	0.98–1.02
6	0.02–0.98	1.00 (0.03)	0.88	1.36	0.91–1.09
	0.10–0.90	1.00 (0.04)	0.87	1.28	0.91–1.09
	0.25–0.75	1.00 (0.04)	0.86	1.32	0.89–1.11
	0.45–0.55	1.00 (0.03)	0.84	1.30	0.91–1.09
	All frequencies	1.00 (0.03)	0.84	1.36	0.91–1.09

[1]Populations were simulated over three or six generations with different allele frequencies. The distributions of the diagonal elements of G were then plotted over all generations.
[2]Animals with allele frequencies of 0.02–0.98, 0.10–0.90 and 0.45–0.55 had identical results.

would be expressed as very low or very high diagonal elements; diagonal elements of animals simulated over six generations ranged from 0.84 to 1.36, suggesting slightly more inbreeding and changes in allele frequencies.

While the majority of diagonal elements were very close to 1.0, each population had animals with diagonal elements outside of the 99% range (Table 1, Figure 2). With three generations, eleven animals had diagonal elements ranging from 1.12 through 1.25 for animals simulated with frequencies between 0.02 and 0.98, 0.10 and 0.90, and 0.45 and 0.55; nine animals had diagonal elements ranging from 1.12 to 1.26 for animals simulated with frequencies between 0.25 and 0.75. With six generations, greater numbers of animals had diagonal elements outside of the 99% range. For animals with frequencies between 0.02 and 0.98, one animal had a diagonal element of 0.88, and 234 animals had elements ranging from 1.10 to 1.36; animals with frequencies between 0.10 and 0.90 had five animals with elements ranging from 0.87 to 0.90, and 221 animals had elements ranging from 1.10 to 1.28; animals with frequencies between 0.25 and 0.75 had two animals with elements of 0.86 and 0.88, and 201 animals had elements ranging from 1.12 through 1.32; animals with frequencies between 0.45 and 0.55 had 17 animals with elements ranging from 0.84 to 0.90, and 207 animals had elements ranging from 1.10 to 1.30. Individuals with low diagonals occurred in the first generation because of animals whose genotypes made them appear to be less related to the rest of the population. Individuals with higher diagonals occurred because of the mating of relatives, including half-sib, sire-progeny and grand-

sire-grandprogeny in later generations. These events arose owing to the smaller number of males than females in each population, particularly as generation size grew past that of the previous generation.

Diagonal elements from multiple populations within the same dataset should show different distributions when **G** is scaled using current allele frequencies from the combined populations. Figure 3 shows the distributions of the diagonal elements of **G** for each of the four combined population datasets. The distribution of the diagonal elements is bimodal for each dataset. The elements of G3_1750 ranged from 0.93 to 0.96, with an average of 0.94 (0.004) for the original 7000 member population, and from 1.82 to 1.96, with an average of 1.84 (0.02) for the subset 1750 member population; the elements of G3_7000 ranged from 1.18 to 1.21, with an average of 1.19 (0.004) for the original 7000 member population, and from 1.15 to 1.25 for the subset 7000 member population, with an average of 1.17 (0.01). The elements of G6_1750 ranged from 0.89 to 1.27, with an average of 0.97 (0.05) for the original 7000 member population, and from 1.57 to 1.95, with an average of 1.71 (0.04) for the subset 1750 member population; the elements of G6_7000 ranged from 1.18 to 1.49, with an average of 1.24 (0.04) for the original 7000 member population, and from 1.03 to 1.36, with an average of 1.10 (0.03) for the subset 7000 member population.

The diagonal distributions show no overlap between the two populations in G3_1750 and G6_1750; however, the diagonal distributions show slight overlap between the two populations in G3_7000 and G6_7000. This is likely because **G** is scaled using allele frequencies estimated from the complete population. With an equal number of animals from both populations contributing to the allele frequency estimation, there is less difference in scaling between the two populations, and the differences are more difficult to detect than for populations with smaller subsets; moreover, with greater numbers of genotyped animals, it is more likely that some genotypes between populations will be similar. It may be easier to discern populations when allele frequencies between them are very different; similarities in allele frequencies or genotypes will allow similarly scaled diagonal elements. Animals of the same breed but of different lines may be difficult to separate because they may share more alleles in the same frequency than those in the simulation.

The genomic relationship matrix of field data was examined. Figure 4 shows the distribution of the
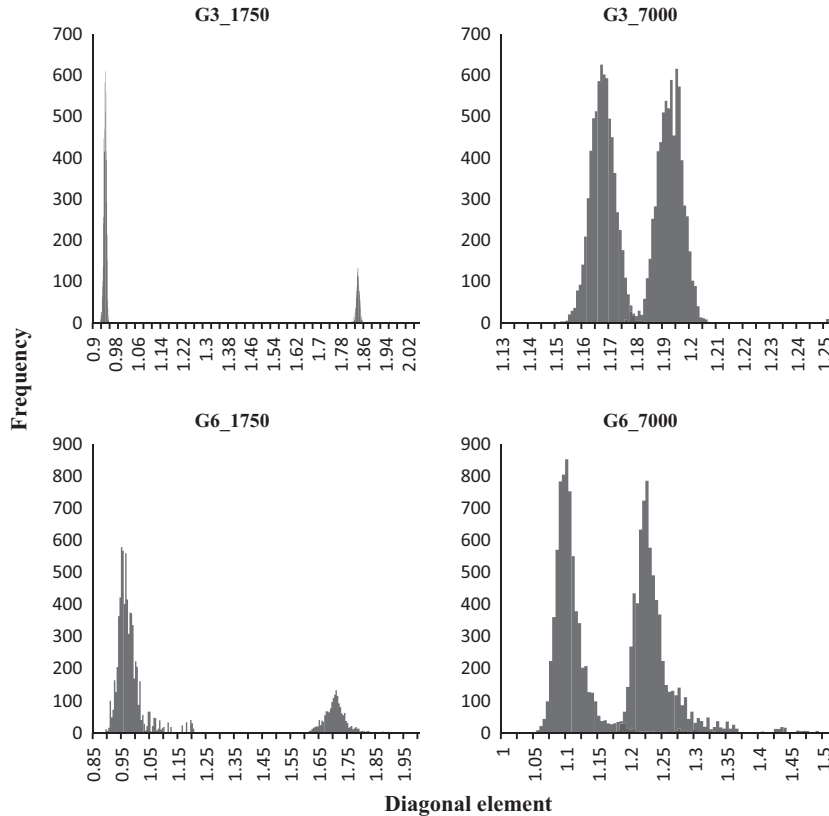
Figure 3 Distributions of the diagonal elements of G for each of the four combined-population datasets. Distributions are shown on different scales.

diagonal elements of **G**. Animals with diagonal elements >1.5 were considered to have abnormally large diagonal elements. Of the 3284 animals, 3195 had diagonal elements ranging from 0.54 to 1.19, 88 had diagonal elements ranging from 1.73 to 2.09 and 1 had a diagonal element of 3.23. In contrast, diagonal elements for **A** had a mean of 1.00 and ranged from 1.00 to 1.09. **G** and **A** were not similar in distribution or scale, suggesting that diagonal elements were sensitive to discrepancies in the data that were not identified using expected relationships in **A**. Consultation with the genotyping laboratory revealed a problem with the initial genotypes. The animal with a diagonal element of 3.23 was identified as having a low call rate. Three of the remaining 88 animals' genotypes were confirmed; the high diagonal elements may have been because of a problematic record during data collection. The final 85 animals were identified as being mislabelled and coming from a second line of broiler chickens under different selection pressures; all 85 animals were from the training dataset. **G** was recomputed for CLEANED. While the overall mean of the diagonal elements increased from 1.03 in ALL to 1.10 in CLEANED, the maximum diagonal element
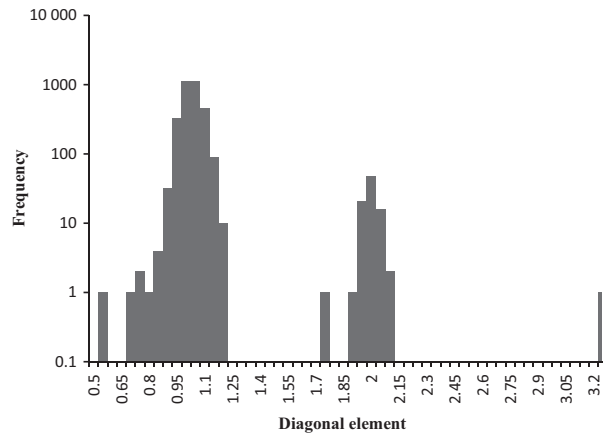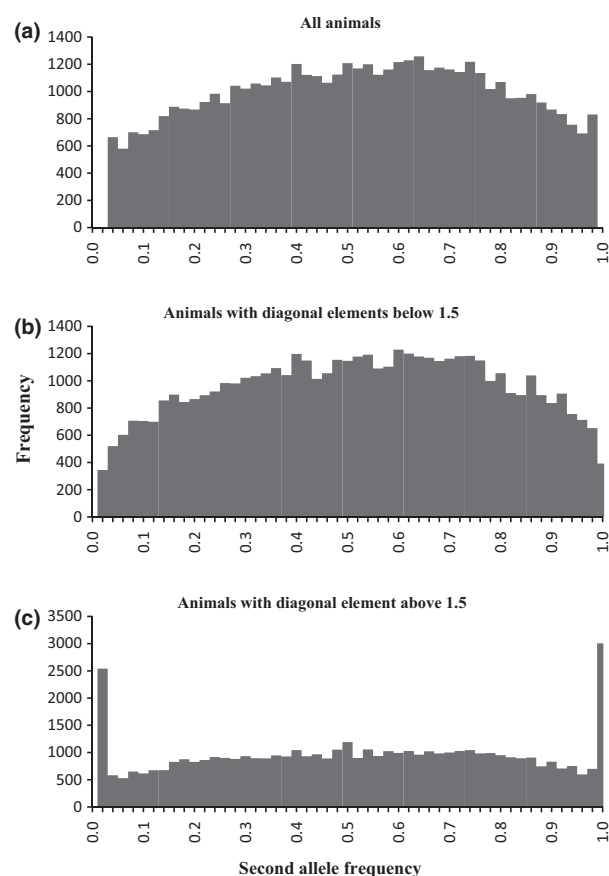

Figure 4 Distribution of the diagonal elements of G for field data. Results are shown in a logarithmic scale.

decreased from 3.23 to 1.20 and the variance decreased from 0.025 to 0.002.

The distributions of the frequency of the second allele at each locus for all animals, animals with diagonal elements of **G** < 1.5 and animals with

diagonal elements of **G** > 1.5 are shown in Figure 5. For all animals, the distribution of the second allele frequency had a mean of 0.51 (0.26) and ranged from 0.02 to 0.98 (Figure 5a). The distribution of the frequencies of the second alleles for animals with diagonal elements below 1.5 had a mean of 0.51 (0.26) and was similar to that of the distribution of second allele frequencies for all animals, although it did have a small number of loci with frequencies below 0.02 and above 0.98 (Figure 5b); the distribution of the frequencies of the second alleles for animals with diagonal elements above 1.5 had a mean of 0.52 (0.30) and had a large number of animals with second allele frequencies below 0.02 and above 0.98 (Figure 5c).

The large peaks below 0.02 and above 0.98 of the animals with high diagonal elements indicate that the frequencies of the second alleles are different than those of the animals with diagonal elements below 1.5. Including these animals in the evaluation

(a) All animals

(b) Animals with diagonal elements below 1.5

(c) Animals with diagonal element above 1.5

Second allele frequency

Figure 5 Distributions of the frequency of the second allele at each locus for all animals (a), animals with diagonal elements of G < 1.5 (b), and animals with diagonal elements of G > 1.5 (c).

affected the calculated second allele frequency of the population. Markers were included that should not have been, as seen by loci with no variation in the animals with diagonal elements below 1.5; differences in allele frequencies between the populations averaged across the entire population allowed some uninformative markers to be included in the evaluation.

Estimates from training data were used to make predictions on the validation data. A traditional evaluation using only phenotypic information was also completed for comparison. Accuracy using a traditional evaluation was 0.48. Accuracy using ALL was 0.64, while accuracy using CLEANED was 0.65. The small increase in accuracy was most likely because the genotypes of the misidentified animals were contributing noise that was averaged over many animals and only slightly affected the predictions.

Accurate genomic evaluations depend on error-free genomic data. The presence of a secondary breed or line mislabelled as belonging to another dataset can impact the results by altering the allele frequency of the population (Bonin *et al.* 2004; Hirschhorn & Daly 2005). SNP effects calculated from one breed or population of animals do not provide accurate GEBVs for animals of another breed or population (Goddard *et al.* 2006). Breeds of animals or populations that developed independently of each other will likely have different allele frequencies, and the phase of the SNP markers may be different between populations (Goddard *et al.* 2006). Differences in allele frequencies will affect the construction of **G**.

The presence of a small secondary population within a larger population did not greatly impact the genetic evaluation. Even with incorrect animals included in the analysis, the genomic evaluation outperformed the traditional evaluation, likely because **G** captured more relationships than would have been captured by **A** alone (Hayes *et al.* 2009b). Adding more phenotypes may also have increased the accuracy of the evaluation (Hayes *et al.* 2009a), but the slight improvement in accuracy observed using CLEANED indicates that the misidentified animals were detrimental and should have been removed. A greater number of misidentified animals may have further decreased accuracy or biased the results and re-ranked animals because of false relationships created by SNPs that were identical by state and not descent. It is important to note that the check using diagonal elements represents only one layer of quality control. If a dataset includes animals mislabelled to the wrong genetic line, it is possible that other errors exist, such as samples being mislabelled to the wrong animal, further reducing

the accuracy. Use of the diagonal elements should be used as a preliminary check of the data; more detailed checks can be completed using allele frequencies, pedigrees and off-diagonal elements of **G** (Wiggans *et al.* 2010).The simulated datasets indicated that diagonal elements of **G** should be distributed with a single peak. VanRaden *et al.* (2008) have indicated that using incorrect allele frequencies has an impact on the estimate of genomic inbreeding coefficients. Homozygosity for rare alleles results in increased inbreeding coefficients because inbreeding is a measure of homozygosity if it is identical by descent (VanRaden 2008). The choice of scaling factor of **G** is important to the evaluation, as an incorrect scaling factor may bias the results (Aguilar *et al.* 2010; Forni *et al.* 2010). Even a small subset of animals impacted the calculated second allele frequencies of the population enough to include uninformative markers; moreover, the same small subset was detectable when the diagonal elements of **G** were plotted as a second distribution >1.0 for field and simulated data. The differences in the diagonal elements arose because the scaling factor was calculated using allele frequencies of the population. This indicates **G** may be a useful diagnostic tool to help recognize mislabelled animals or secondary populations. This study showed that the method worked for a small number of animals deviating strongly from the allele frequencies of the whole population. Further studies may verify the sensitivity of the genomic relationship matrix in identifying multiple populations.

Conclusions

The diagonal elements of the genomic relationship matrix for animals from a simulated population had a narrow distribution centred on 1.00. As the number of generations simulated increased, the distribution of diagonal elements widened. This corresponded with the theory that diagonal elements should be centred on 1.0. Diagonal elements corresponding to animals from different populations may be higher than average. A distribution of diagonal elements of **G** with multiple peaks may be attributed to an admixed population, unreliable sample data or a low call rate. A second population of mislabelled animals may be identifiable in a larger population if the allele frequencies between the two populations are very different. Removal of mislabelled animals and genotypes will increase the accuracy of prediction, with the level of increase dependent on the number of removed genotypes and the total number of phenotypes observed.

References

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

Bonin A., Bellemain E., Eidesen P.B., Pompanon F., Brochmann C., Taberlet P. (2004) How to track and assess genotyping errors in population genetics studies. *Mol. Ecol.*, **13**, 3261–3273.

Chen C.Y., Misztal I., Aguilar I., Tsuruta S., Meuwissen T.H.E., Aggrey S.E., Muir W.M. (2010) Genome wide marker assisted selection in chicken: making the most of all data, pedigree, phenotypic, and genomic in a simple one step procedure. In: Proceeding of the 9th World Congress on Genetics Applied to Livestock Production. Leipzig, Germany, 1–6 August 2010.

Forni S., Aguilar I., Misztal I., Deeb N. (2010) Genomic relationships and biases in the evaluation of sow litter size. In: Proceeding of the 9th World Congress on Genetics Applied to Livestock Production. Leipzig, Germany, 1–6 August 2010.

Goddard M.E., Hayes B., McPartlan H., Chamberlain A.J. (2006) Can the same genetic markers be used in multiple breeds? In: Proceeding of the 8th World Congress on Genetics Applied to Livestock Production. Belo Horizonte, Brazil, 13–18 August 2006.

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

Hayes B.J., Bowman P.J., Chamberlain A.J., Goddard M.E. (2009a) Invited review: genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.*, **92**, 433–443.

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

Hirschhorn J.N., Daly M.J. (2005) Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.*, **6**, 95–108.

Legarra A., Robert-Granie C., Manfredi E., Elsen J.-M. (2008) Performance of genomic selection in mice. *Genetics*, **180**, 611–618.

Meuwissen T.H.E., Hayes B.J., Goddard M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, **157**, 1819–1829.

Pritchard J.K., Stephens M., Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Rosenberg N.A., Burke T., Elo K., Feldman M.W., Freidlin P.J., Groenen M.A.M., Hillel J., Maki-Tanila A., Tixier-Boichard M., Vignal A., Wimmers K., Weigend S. (2001) Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics*, **159**, 699–713.

Schaeffer L.R. (2006) Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet.*, **123**, 218–223.

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

VanRaden P.M., Tooker M.E., Gengler N. (2008) Effects of allele frequency estimation on genomic predictions and inbreeding coefficients. *J. Dairy Sci.*, **91** (E-Suppl. 1): 506 (Abstract 522).

VanRaden P.M., Van Tassell C.P., Wiggans G.R., Sonstegard T.S., Schnabel R.D., Taylor J.F., Schenkel F.S. (2009) Invited review: reliability of genomic predictions for north american holstein bulls. *J. Dairy Sci.*, **92**, 16–24.

Wiggans G.R., VanRaden P.M., Bacheller L.R., Tooker M.E., Hutchison J.L., Cooper T.A., Sonstegard T.S. (2010) Selection and management of DNA markers for use in genomic evaluation. *J. Dairy Sci.*, **93**, 2287–2292.