Genomic Relationships and Biases in the Evaluation of Sow Litter Size

S. Forni[*], I. Aguilar[‡†], I. Misztal[†] and N. Deeb[*]

Introduction

Traditional genetic evaluations combine only phenotypic data and probabilities that genes are identical by descent using pedigree information. Genomic information can be added to evaluations and the observed proportion of chromosome segments shared by individuals can replace probabilities. This may increase the rate of genetic improvement each generation by raising the accuracy of estimated breeding values (EBVs) and reducing the generation interval (VanRaden (2008)). The integration of genotype information into the numerator relationship matrix (**A**) using a genomic relationship matrix (**G**) is an alternative to exploit the advantages of genomics when the entire population cannot be genotyped (Legarra et al. (2009); Christensen and Lund (2010)). Differences in the scale of pedigree-based and genomic relationship coefficients were reported (Aguilar et al. (2010)). How these differences can affect parameter estimates is still uncertain. The objectives of this research were to contrast pedigree and genotype-based relationship coefficients in a swine population, and compare the outcomes of a genetic evaluation when the standard relationships were modified by genomic coefficients.

Material and methods

Phenotypic records and model. Records of litter sizes from 338,346 sows, of which 1,919 were genotyped using the porcine 60k SNP chip, were evaluated. Genotypes of 70 sires were also available. The data set is privately-owned by PIC/Genus Plc. Analyses were carried out with the complete data set and with a subset of genotyped animals and their parents (n=5,090). A single-trait animal model was used to estimate variance components and breeding values. The fixed effects of parity order, age at farrowing (linear covariable), number of services, mating type (artificial insemination or natural service), contemporary group, sow line and sire line (mate) were included in the analysis. Contemporary groups were defined by: season, year and farrowing farm. EBVs' accuracies were estimated using: accuracy = 1 – sqrt(PEV/ σ_a^2), where sqrt= square root, PEV = prediction error variance and σ_a^2 = additive variance. PEV was obtained by inversion of the coefficient matrix of the mixed model equations.

Pedigree-based and genomic relationships. The numerator relationship matrix was built with pedigree information on 382,988 animals. A genomic relationship matrix was

---

[*] PIC/Genus Plc, Hendersonville, TN, United States
[‡] Instituto Nacional de Investigación Agropecuaria, Las Brujas, Uruguay
[†] Department of Animal and Dairy Science, University of Georgia, Athens, GA, United States

constructed following (VanRaden (2008)) and used allele frequencies in the base population equal to 0.5 (**G05**), equal to the observed average minor allele frequency (**GMF**) or observed allele frequencies (**GRF**). In addition, a modification in the denominator proposed by Gianola et al. (2009) to scale **G** to be analogous to **A** was also applied with realized frequencies (**GRF***). This algorithm assumes that gene frequencies in the base population are not independent. A normalized matrix (**GN**) was obtained by multiplying **GRF** by a constant to achieve an average diagonal of 1. The genomic matrix was weighted as **G** = 0.95**Gr** + 0.05**A** (**Gr** = matrix with raw genomic coefficients) and combined with pedigree-based coefficients in the inverse of matrix **H** as shown in Aguilar et al. (2010).

Results and discussion

Statistics of pedigree-based and genomic relationship coefficients are presented in Table 1. The values refer to the matrix containing relationships between genotyped animals only ($\mathbf{A}_{22}$ or **G**). Using **G05** and **GMF** both diagonal and off-diagonal elements were on average greater than the coefficients in **A**. The average minor allele frequency was equal to 0.264. Regardless of the denominator employed to scale **G,** the mean genomic coefficients obtained with realized frequencies were smaller than pedigree-based coefficients. Also, genomic coefficients presented greater variance.

Table 1: Statistics of relationship coefficients estimated using pedigree and genomic information

	N	Mean	Maximum	Minimum	Variance
Pedigree-based relationships (A)					
Diagonal	1989	1.000	1.075	1.000	0.00003
Off-diagonal	3954132	0.032	0.600	0.000	0.00172
Genomic relationships assuming allele frequency 0.5 (G05)					
Diagonal	1989	1.253	1.462	1.178	0.00083
Off-diagonal	3954132	0.595	1.231	0.387	0.00160
Genomic relationships using the average minor allele frequency (GMF)					
Diagonal	1989	1.697	1.894	1.632	0.00073
Off-diagonal	3954132	1.022	1.654	0.822	0.00155
Genomic relationships using realized allele frequencies (GRF)					
Diagonal	1989	0.936	1.228	0.837	0.00176
Off-diagonal	3954132	0.000	1.000	-0.198	0.00241
Normalized genomic relationships (GN)					
Diagonal	1989	1.002	1.314	0.895	0.00201
Off-diagonal	3954132	0.000	1.070	-0.212	0.00275
Genomic relationships using realized non-independent allele frequencies (GRF*)					
Diagonal	1989	0.505	0.663	0.436	0.00051
Off-diagonal	3954132	0.000	0.540	-0.105	0.00070

Pedigrees may include many generations but must end eventually. Traditional genetic evaluations assume that the founder individuals are the earliest generation recorded and they do not share genes from more remote ancestors. Relationship and inbreeding coefficients from later generations are estimated as deviances from the founders' relatedness. Genomic analysis may reveal that founding animals actually share genes identical by descent and shift the relationship coefficients up and down. Genomic and pedigree-based matrices should be compatible in scale to be integrated, but there is no reason for **G** to have a particular interpretation in terms of relationships (Legarra et al. (2009)). Estimates of variance components in the full data set (n = 338,346) are shown in Table2 and in the subset (n = 1,919) in Table 3.

Table 2: Variance components estimates for litter size using pedigree and genomic relationship coefficients and all phenotypes available

	Additive Variance (ste^1)	Residual Variance (ste^1)
A	1.26 (±0.03)	6.66 (±0.02)
G05	1.28 (±0.03)	6.65 (±0.03)
GMF	1.28 (±0.03)	6.65 (±0.03)
GRF	1.27 (±0.03)	6.65 (±0.03)
GN	1.27 (±0.03)	6.65 (±0.03)
GRF*	1.30 (±0.03)	6.64 (±0.03)

[1] standard errors; *following Gianola et al. (2009)

Choices of **G** with average diagonal coefficients different from 1 led to greater estimates of additive variance, mainly in the smaller data set. Relative inflation of variance estimates would not be a problem if the entire population was genotyped and all relationships were genomic-based, but this is not the case for most livestock populations. If analysis requires a unique additive variance for genotyped and non-genotyped animals, a possible solution to avoid this problem is re-scaling the genomic matrix to obtain average diagonal elements equal to 1. This was achieved by multiplying the genomic coefficients by a constant (equal to 1.07). Legarra et al. (2009) showed that a normalized **G** obtained using **GN** = **G**/trace(**G**) allows the same expectation of variance for breeding values of genotyped and non-genotyped animals. Standard errors of additive variance estimates in the subset were smaller when **GRF** and **G**N were used. This may indicate that more information is added by the genomic coefficients in the appropriate scale than standard probabilities.

Table 3: Variance components estimates for litter size using pedigree and genomic relationship coefficients and phenotypes of genotyped animals

	Additive Variance (ste^1)	Residual Variance (ste^1)
A	2.27 (±0.52)	5.30 (±0.44)
G05	3.43 (±0.56)	5.25 (±0.29)
GMF	3.43 (±0.56)	5.25 (±0.30)
GRF	2.41 (±0.39)	5.29 (±0.30)
GN	2.25 (±0.36)	5.30 (±0.30)
GRF*	4.46 (±0.73)	5.22 (±0.30)

[1] standard errors; *following Gianola et al. (2009)

Christensen and Lund (2010) also reported that parameter estimation was sensitive to the choice of the allele frequencies in a scenario with selection and where the base population was not genotyped. Aguilar et al. (2010) observed inflation of breeding values using genomic relationship coefficients and proposed to weight the difference between **G** and **A** to obtain optimal predictions. The primary influence of the weighting factor would be related to the proportion of the additive variance explained by the genomic information. It is not clear if the optimal dispersion parameters for the standard linear model using **A** should be the same after incorporation of **G**.

Estimates of accuracies obtained by inversion with different genomic matrices are presented in Table 4. Increase in accuracies for genotyped animals was observed despite the allele frequencies and scaling method employed. It is possible that differences in scale of relationship coefficients obtained using different sources of information led to biases in estimates of breeding values and accuracies, especially in the small data set. However, estimates of genomic breeding values were fairly similar. The correlation between EBVs and genomic breeding values were 0.79, 0.79, 0.78, 0.78, 0.79 for **G05**, **GMF, GRF, GN** and **GRF***, respectively. In a simulation study, VanRaden (2008) showed that accurate estimates of genomic inbreeding coefficients required very precise estimates of allele frequencies in the base population while genetic merit was less sensitive.

Table 4: Average breeding values' accuracies using pedigree and genomic relationship coefficients

Relationship Matrix	**Full pedigree (n=382,988)**	**Genotyped females (n=1,919)**	**Genotyped sires (n=70)**
A	0.21	0.22	0.62
G05	0.21	0.37	0.63
GMF	0.21	0.49	0.64
GRF	0.21	0.30	0.63
GN	0.21	0.28	0.63
GRF*	0.21	0.43	0.66

Conclusion

Genetic similarity can be defined in several ways using pedigree, genotype information or a combination of both. Genetic evaluations can benefit from the mixture of genomic and pedigree information, but such integration poses a challenge. When the genomic relationship coefficients are of a different scale than pedigree-based coefficients, parameter estimates may be biased, especially if data sets are small. A possible solution to remove the bias is to normalize the genomic relationship matrix.

References

Aguilar, I., Misztal I., Johnson D. L. *et al* (2010). *J. Dairy Sci.* 93: 743-752

Christensen, O. F., and Lund, M. S. (2010). *Genet. Sel. Evol.*, available online: http://www.gsejournal.org/content/42/1/2

Legarra, A., Aguilar, I., and Misztal I. (2009). *J. Dairy .Sci.*, 92:4656–4663.

Gianola, D., de los Campos, G., Hill, W. G. *et al* (2009). *Genetics* 183:347-363.

VanRaden, P. (2008). *J. Dairy Sci.,* 91:4414-4423.