Proc. Assoc. Advmt. Anim. Breed. Genet: 24: 397-401

COMPARING GENOMIC WITH PEDIGREE RELATIONSHIP MATRICES AND PRELIMINARY GENOME WIDE ASSOCIATION IN SANTA GERTRUDIS BULLS

W.L.A. Tan[1], L.R. Porto-Neto[2], G. C. Gouveia[3], A. Reverter[2] and M.R.S. Fortes[1]

[1]The School of Chemistry and Molecular Biosciences, the University of Queensland, Brisbane, QLD 4072 Australia
[2]CSIRO Agriculture and Food, Brisbane, QLD 4067.
[3]Federal University of Minas Gerais, Belo Horizonte, MG, Brazil

SUMMARY

To estimate relatedness between individuals, one can choose between the **A** matrix from pedigree data, the **G** matrix built from genome-wide SNP markers, or even a combination of both. Comparing both matrices is of interest to explore the possibility of initiating a genomic breeding program. We use data on 929 Santa Gertrudis bulls with 617,348 SNP genotypes to compare the matrices. The **A** and **G** matrices were positively correlated (0.91; $P < 0.001$), and diagonal and off-diagonal coefficients were very similar. This result indicates that the pedigree is correct, although some genomic-estimated relationships did not agree with the pedigree-based matrix. Subsequently, a genome-wide association study was performed for scrotal circumference (SC) using **A** and **G** matrices. 100 SNPs were associated with SC with a corresponding FDR (<0.05) in GWAS using **G**, with the highest peak at BTA 5. Previously, the peak on BTA 5 has been associated with sheath score in Brahman and Tropical Composite (TC) cattle. BTA 5 has also been associated with SC in TC. For GWAS conducted using the **A**, 2883 SNPs were associated with SC with a corresponding FDR (<0.05), with the highest peak at BTA 5 and other peaks at BTA 22 and chromosome X. The peaks in chromosome X and BTA 22 was not observed in GWAS using **G**. This warrants further investigation into the differences in estimated SNP effects resulting from using different matrices in GWAS. Additionally, combining both the **A** and **G** in an **H** matrix may make more accurate predictions than using **G** alone. Further analysis is required to investigate the use of **H** and to verify the SNP associations identified in this study and across other breeds.

INTRODUCTION

Traditionally, most analyses in livestock relied on the calculation of the **A** matrix from pedigree data. Today, DNA markers are also used to estimate relatedness between individuals for various genetic analyses (Makgahlela *et al.* 2013). The relationship matrix estimated from genotypes is termed the **G** matrix. One advantage of estimating **G** is detecting alleles identical by state, traced to common ancestors that are unknown and therefore not captured in the pedigree (VanRaden 2008). The **G** matrix coefficients may be more precise than those in **A** and could correct pedigree relationships assigned wrongly in **A**. However, it is expensive and often impossible to genotype the entire population, and sometimes the **A** is still used. In genome-wide association studies (GWAS), relationship matrices account for population structures while testing for SNPs association. As using either **G** and **A** may affect the results of GWAS, it becomes crucial to determine which one might be a better fit for GWAS. In this study, we compare **A** and **G** matrices to inform future GWAS work.

MATERIALS AND METHODS

Animals and genotypes. Scrotal circumference (SC) was measured in 952 Santa Gertrudis bulls. Animals that were genotyped using 50K SNP panels were imputed to high density (770K) using a phased reference generated by Eagle2 (v2.4.1) and imputed using Minimac3 for autosomes and Minimac4 for the X chromosome (Loh *et al.* 2016; Das *et al.* 2016). SNPs with a call rate of less than 0.90 were removed before imputation. After quality control (minor allele frequency >0.05), 617,348 SNPs were available for analysis. Contemporary groups (CG) accounted for management, year of record, and age groups. CGs with less than five bulls were discarded, leaving a final dataset of 929 bulls for analysis.

G and A matrices. The **G** matrix was built using Method 1, as described by VanRaden (2008). Pedigree information on 4409 animals which were the genotyped animals and their ancestors, was used to compute the Wright coefficients in **A** described by Wright (1922) in SVS software (Release 8.9.0, Golden Helix, Inc.).

GWAS. GWAS was performed in SVS software (Release 8.9.0, Golden Helix, Inc.) using the Efficient Mixed-Model Association eXpedited method (EMMAX) (Kang *et al.* 2010). GWAS was performed for SC with **A** and **G** matrices separately. CG was fitted as a fixed effect, and age was fitted as a covariate. The first two principal components of the principal component analysis (PCA), calculated for the bulls included in this paper, were also fitted as a covariate. False Discovery Rate (FDR) was used to correct for multiple testing.

RESULTS AND DISCUSSION

Scatter and box plots comparing relationships estimated using both **A** and **G** matrices are shown in Figures 1 and 2. **A** and **G** matrix were highly positively correlated (0.91; < 0.001). The means of the diagonal and off-diagonal coefficients were very similar: they were 1 for **A** and 1.03 for **G** in the diagonal, and they were 0.02 for **A** and 0 for **G** in the off-diagonal. In some cases, according to **A,** unrelated animals were estimated to be similar to half-siblings in **G** (a coefficient of ~ 0.25). Whereas some animal pairs considered half-siblings in **A** (a coefficient of 0.25) were either full siblings or unrelated in **G**. In general terms, the genomic relationships (**G**) were within the expected range of the pedigree relationships. Thus, our analyses indicate that the pedigree for this population is accurate.

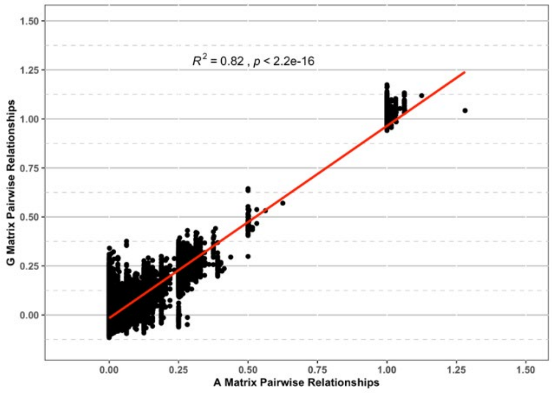

Figure 1. Scatter plot of relationships coefficients for all pairwise combinations of 929 Santa Gertrudis bulls estimated using SNP markers (G) and pedigree (A)

In the GWAS using the **A** matrix, a total of 2883 SNPs were detected with an FDR of less than 0.05 for SC, with the highest peak at BTA 5 and other peaks at BTA 22 and chromosome X. The peak SNP at BTA 5 for GWAS using the A matrix corresponded to an FDR of 0.0005 for BTA 5. In GWAS using the **G** matrix, a total of 100 SNPs were detected with an FDR of less than 0.05 for SC, with the highest peak at BTA 5. The peak SNP for GWAS using the G matrix corresponded to an FDR of 0.0011 for BTA 5. Other peaks were also observed in BTA 4 and BTA 9 for GWAS using the G matrix.

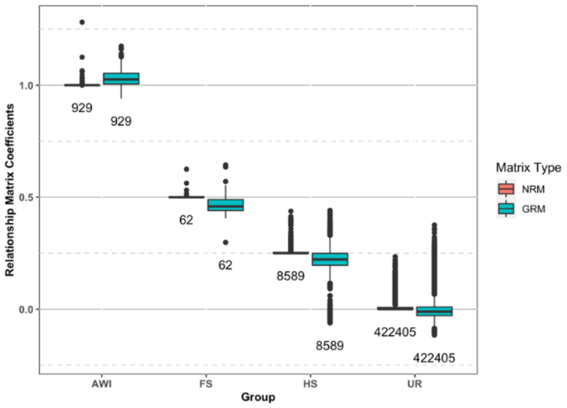

Figure 2. Boxplot comparing genomic relationship matrix (G) in blue against corresponding additive numeric matrix (A) coefficients in red of Santa Gertrudis bulls with the animal with itself (AWI), full-siblings (FS), half-siblings (HS), and unknown relationships (UR). The number of observations in each group has been annotated in black

The Manhattan plot observed when using another **G** matrix, calculated without X chromosome SNPs, was very similar to the **A** matrix plot, with peaks observed on X. Therefore, building **G** without SNPs in the X chromosome may lead to different GWAS results (Druet and Legarra 2020).

The largest concentration of SNPs associated with SC was located from 46.2 Mb to 50.2 Mb at BTA 5. The peak SNP for this region is found at 47.8 Mb, which accounted for 0.04% of the genetic variance in SC. Fortes *et al.* (2020) reported SNP associations with SC in BTA 5 in a study of Tropical Composite (TC) cattle which overlaps with the region reported for BTA 5 in our study. The significance of BTA 5 in cattle breeding and its association with other traits has been documented in previous studies. For example, Porto-Neto *et al*. (2014) reported significant SNP association at BTA 5 for sheath score in Brahmans and TC.

Our study found peak SNPs associated with SC in BTA 4 (40.1 Mb) and 9 (32.2 Mb) for GWAS conducted using either **A** or **G** matrix. The peak SNP in the BTA 4 accounted for 0.02% and 0.03% of the variance in SC for GWAS conducted using the **A** and G matrix, respectively. Whereas the peak SNP in BTA 9 accounted for 0.03% and 0.04% of the variance in SC for GWAS conducted using the **A** and **G** matrix, respectively. A crossbred beef cattle study also reported an association in BTA 9 for SC, and another study reported a region in BTA 4 associated with SC in Nelore Cattle (Sweett et al. 2020; Sbardella *et al.* 2021). However, these regions did not overlap with the regions reported in our study for BTA 4 and BTA 9.

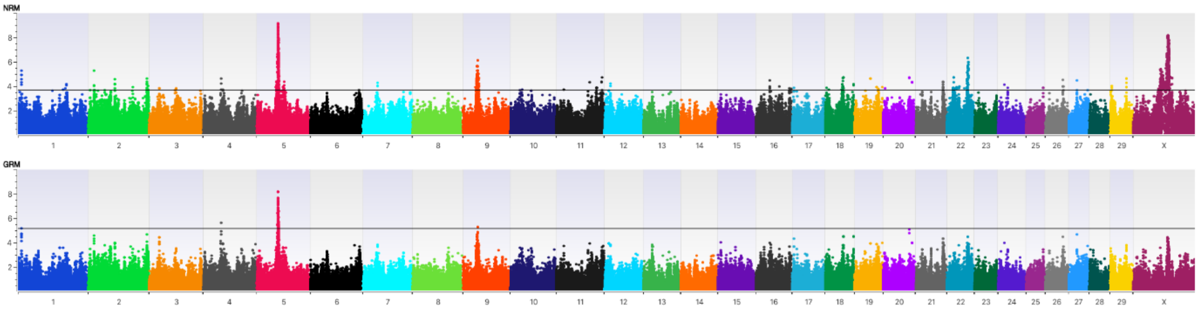

Figure 3. Manhattan plot of inverse log *P*-values for SC in GWAS using the A matrix (top) and the G (bottom). The inverse log *p*-values for each SNP are plotted on the y-axis for each chromosome. The chromosome number is labelled on the x-axis, and a variety of colours correspond to different chromosomes. The *P*-value line in black corresponds to an FDR (<0.05)

The heritability estimated using **G** and **A** were similar, although subtle differences exist between genetic parameters estimated using the different matrices. **G** accounts for a higher genetic variance and has a lower variance of heritability compared to **A**, as shown in Table 2. The differences between both estimates may have resulted from differences in relationship estimates between the **A** and **G** matrices. However, more research is required to understand the differences seen in the parameters estimated by both matrices. The heritability estimates are similar to those reported in TC (0.43) but are lower than those reported in Brahmans (0.75) and another composite breed (0.67) (Corbet *et al.* 2013; Roberts *et al.* 2010).

Table 2. Genetic parameters obtained using G and A matrices

Matrix Type	Heritability	Variance of heritability	V_g	V_e
A	0.429	0.012	2.879	3.828
G	0.466	0.005	3.125	3.578

CONCLUSIONS

This study shows that the matrix coefficients between animals using the **A** and **G** matrices are highly correlated but not identical. Utilising genomic information includes additional information, which results in capturing relationships otherwise undetected or missing in pedigree. GWAS reveals differences in SNP effect estimates resulting from the use of these different matrices. The **G** matrix may provide more accurate information and facilitate estimating genetic parameters and identifying significant SNPs. Future studies can explore combining both matrices in an **H** matrix, which might be better than the **G** matrix alone. The GWAS highlighted associations with SC in chromosomes 4, 5, 9 and X in a population of Santa Gertrudis bulls. Future studies can further investigate these associations and if they occur in other breeds.

ACKNOWLEDGEMENTS

The authors acknowledge that this study was funded as part of the MLA project L.GEN.1818 Bull fertility update: historical data, new cohort and advanced genomics. We thank MLA and all collaborators in this project for making the study possible.

REFERENCES

Corbet N.J., Burns B.M., Johnston D.J., Wolcott M.L., Corbet D.H., Venus B.K., Li Y., McGowan M.R., and Holroyd R.G. (2013) *Anim. Prod. Sci.* **53:** 101-113.

Druet T., and Legarra A. (2020) *Genet. Sel. Evol.* **52**: 50.

Das S., Forer L., Schönherr S., Sidore C., Locke A. E., Kwong A., Vrieze S. I., Chew E. Y., Levy S., McGue M., Schlessinger D., Stambolian D., Loh P.R., Iacono W. G., Swaroop A., Scott L.J., Cucca F., Kronenberg F., Boehnke M., Abecasis G.R., Fuchsberger C. (2016) *Nat. Genet.* **48**: 1284-1287

Fortes M.R.S., Porto-Neto L.R., Satake N., Nguyen L. T., Freitas A.C., Melo T.P., Scalez D.C.B., Hayes B.J., Raiden F.S.S., Reverter A., Boe-Hansen G.B. (2020) *Genet. Sel. Evol*. **52**: 46

Kang H.M., Sul J.H., Service S.K., Zaitlen N.A., Kong S.Y., Freimer N.B., Sabatti C., and Eskin E. (2010) *Nat. Genet.* **42**: 348-354.

Loh P.R., Danecek P., Palamara P.F., Fuchsberger C., Reshef Y.A., Finucane H.K., Schoenherr S., Forer Lukas., McCarthy S., Abecasis G.R., Durbin R., Price A.L. (2016) *Nat. Genet*. **48**: 1443–1448.

Porto-Neto L.R., Reverter A., Prayaga K.C., Chan E.K.F., Johnston D.J., Hawken R.J., Fordyce G., Garcia J.F., Sonstegard T.S., Bolormaa S., Goddard M.E., Burrow H.M., Henshall J.M., Lehnert S.A., Barendse W. (2014) *PLoS One* **9**: e113284.

Makgahlela M.L., Strandén I., Nielsen U.S., Sillanpää M.J., and Mäntysaari E.A. (2013) *J. Dairy Sci.* **96**: 5364-5375.

Roberts C., Geary T., MacNeil M., Waterman R., Roberts A., and Alexander L. (2010) *Proc. West. Sect. Ameri. Soc. Anim. Sci*. 61: 122-125.

Sbardella A.P., Watanabe R.N., Costa R.M., Bernardes P.A., Braga L.G., Rey F.S.B., Lôbo R.B., Munari D.P. (2021) *Animals*. **11**: 1386

Sweett H., Fonseca P.A.S., Suárez-Vega A., Livernois A., Miglior F., & Cánovas A. (2020) *Sci. Rep.* **10**: 20102-20102.

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

Wright S. (1922) *Am. Nat.* **56**: 330-338.