Proc. Assoc. Advmt. Anim. Breed. Genet. 25: 338 - 341

ASSESSING THE INFLUENCE OF BAYESR AND GBLUP ON SNP EFFECTS USED IN THE CORRELATION SCAN METHODOLOGY

B.S. Olasege[1,2], I. van den Berg[3], E.J. Breen[3], M. Haile-Mariam[3,6], P.N. Ho[3], L.R. Porto-Neto[2], B.J. Hayes[4], M. Goddard[3,5], J.E. Pryce[3,6] and M.R.S. Fortes[1,4]

[1]The University of Queensland, School of Chemistry and Molecular Biosciences, QLD 4072, Australia
[2]CSIRO Agriculture and Food, St Lucia, QLD 4067, Australia
[3]Agriculture Victoria, AgriBio, Centre for AgriBioscience, VIC. 3083, Australia
[4]The University of Queensland, Queensland Alliance for Agriculture and Food Innovation, St Lucia, QLD 4072, Australia
[5]Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, VIC. 3052, Australia
[6]School of Applied Systems Biology, La Trobe University, VIC. 3083, Australia

SUMMARY

The Correlation Scan (CS) identifies local genomic regions that disproportionately contribute to the genetic correlation between traits using SNP effects generated from GBLUP. BayesR, has been shown to precisely localise SNP effects, and the BayesR SNP effects size are often less shrunk than GBLUP. Therefore, we aimed to compare the SNP effects generated from GBLUP and BayesR models on the resulting localised genomic regions using the CS method. Single-trait and bivariate models were used to analyse fertility data from Brahman cows (age at detection of first corpus luteum; 996 animals) and bulls (insulin-like growth factor measured from blood; 1022 animals) genotyped with the Illumina BovineHD (770K) SNP chip. We observed that the local correlation (**r**) estimates were larger with GBLUP than BayesR. There were considerable differences in the **r** estimates on chromosome 5, 14, and X. Further analysis into the distribution of the SNP effects of a QTL region on chromosome 14 highlights the effect that each method had on CS results. GBLUP spread the effect across neighbouring SNPs, while BayesR localised the effect to a small number of SNPs, reducing the **r** estimates. The differences between GBLUP and BayesR were reduced with BayesR bivariate model. As BayesR bivariate model tended to select common SNPs as having non-zero effects on both traits compared to BayesR single-trait, the patterns of the **r** estimates were larger in the bivariate model. Other metrics from the BayesR bivariate model identified similar regions as the GBLUP in CS results. Our results showed that BayesR SNP effects can be used in our CS, but the bivariate model is recommended.

INTRODUCTION

Estimated genetic correlations between traits are useful parameters for developing and optimising animal breeding programs (Petrini *et al.* 2016). However, little is known about the local genomic regions that disproportionately contribute to these overall genetic correlations. With the widespread use of genomic data, the knowledge of local regions affecting trait correlations could allow breeders to make a more targeted genomic selection. The Correlation Scan (CS) identifies local genomic regions that contribute to estimates of the genetic correlations between traits (Olasege *et al.* 2022). The CS framework was developed using SNP effects generated from GBLUP, but it is possible to extend it for Bayesian approaches. BayesR has been shown to precisely localise SNP effects and the effect sizes are less shrunk than GBLUP (Kemper *et al.* 2015). Therefore, we used BayesR (single and bivariate models) to generate the SNP effects for the CS and compared the observed results with those obtained from GBLUP.

MATERIALS AND METHODS

The two traits used for this study were age at detection of first corpus luteum in cow (AGECL, n=980) and blood concentration of insulin-like growth hormone measured in bulls (IGF1b, n=964) from a Brahman population. A detailed description of the traits is provided by Olasege et al. (2021). The estimated genome-wide genetic correlation between these traits was -0.65 (Olasege *et al.* 2021).

SNP effects for the CS were calculated using single-trait and bivariate GBLUP (Olasege *et al.* 2022) and BayesR (Breen *et al.* 2022) models, with BovineHD 770K SNP chip. The posterior inclusion probability (PIP) and Q2 probability (the probability that the SNPs are associated with either of the traits) were also obtained from bivariate BayesR. Then local correlations (**r**) were estimated using the SNP effects using each model. The method to estimate **r** (correlation of 500 SNP effects in sliding windows between the two traits) has been previously detailed by Olasege *et al.* (2022).

RESULTS AND DISCUSSION

Single- and bivariate **r** estimates for the BayesR model are presented in Figure 1. The GBLUP single-trait result has been published (Olasege *et al.* 2022). The bivariate result for the GBLUP model looks identical to the single trait (result not shown). GBLUP yielded larger **r** estimates than BayesR. While both models identified similar windows, there were considerable differences in the **r** estimates on chromosome 5, 14, and X. For example, a QTL region including ***PLAG1*** (Fortes *et al.* 2012; Hawken *et al.* 2012) on chromosome 14 was not identified by the BayesR single-trait model. However, with BayesR bivariate model, this region was signalled.

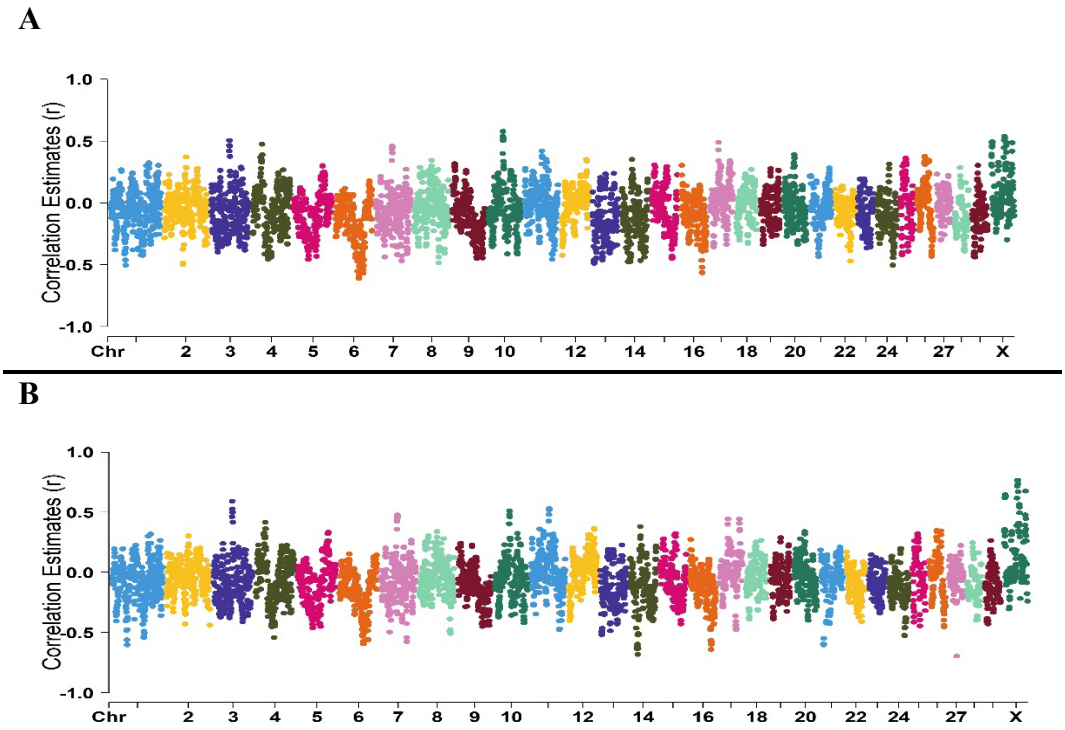

Figure 1. Genome-wide plots of the local correlation (r) estimates for age at first corpus luteum and blood concentration of insulin growth hormone for BayesR model using SNP effects from single-trait (A) and bivariate model (B)

By investigating the 100 SNP effects surrounding the ***PLAG1*** region between GBLUP (single-trait) and BayesR (single- and bivariate model), we found that GBLUP (Figure 2A; **r** = 0.96) spread the effect across neighbouring SNPs, while BayesR SNP effects were localised to a small number of SNPs. BayesR bivariate (Figure 2C; **r** = 0.76) identified similar SNPs for each trait as having non-zero effects whereas BayesR single trait (Figure 2B; **r** =0.23) often picked different sets of SNPs. Leveraging on the PIP and Q2 probability from BayesR bivariate model, the regions identified as the most significant from GBLUP CS were also signalled using Q2*PIP, showing that these two metrics could complement the CS method (Figure 3).

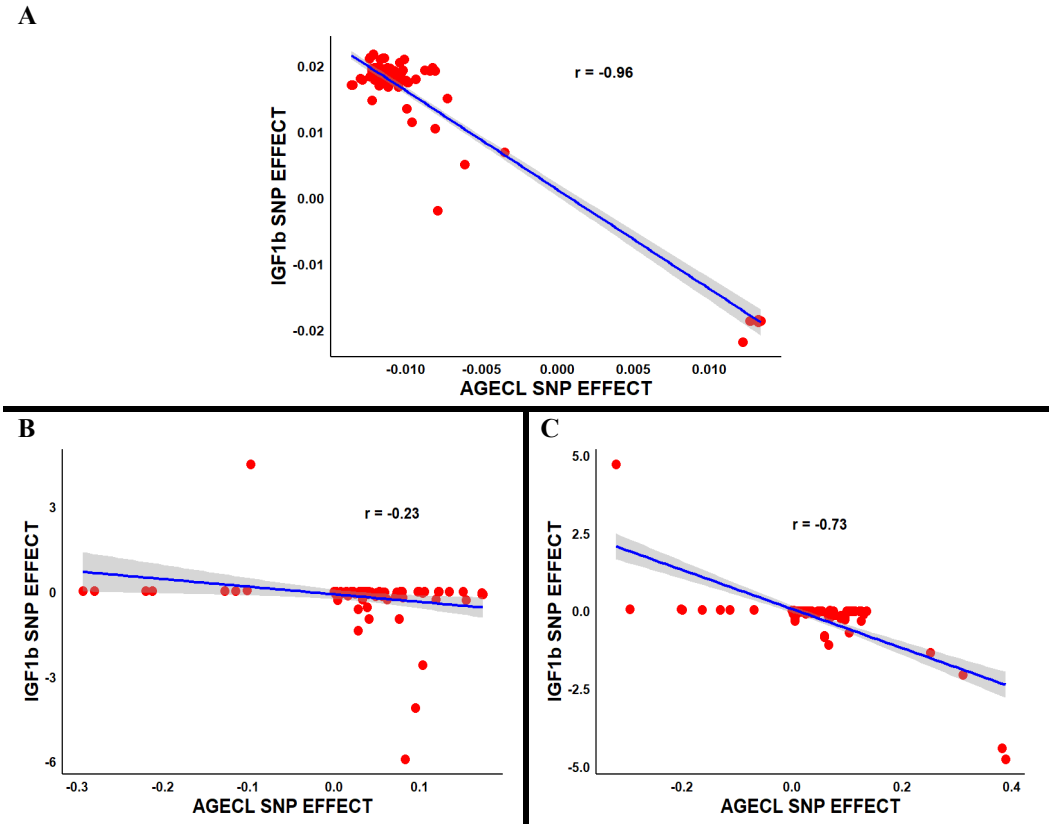

Figure 2. The regression of the distribution of the 100 SNP effects within the boundary of the *PLAG1* gene between age at first corpus luteum and (AGECL) and blood concentration of insulin growth hormone (IGF1b) using GBLUP single-trait (A), BayesR single-trait (B), and BayesR bivariate Model (C)

CONCLUSIONS

The differences in model assumptions led to differences in local correlations estimated using GBLUP and BayesR. GBLUP spreads the effect across neighbouring SNPs, whereas BayesR localised the effect to a small number of SNPs. With bivariate BayesR, SNP effects tend to be allocated to common SNPs across the traits, while BayesR single trait may select different SNPs for each trait, resulting in reduced **r** estimates. Our results showed that BayesR SNP effects can be used for the CS, but the bivariate model is recommended. Q2 and PIP from BayesR bivariate model could complement the CS method for insights into important QTLs.

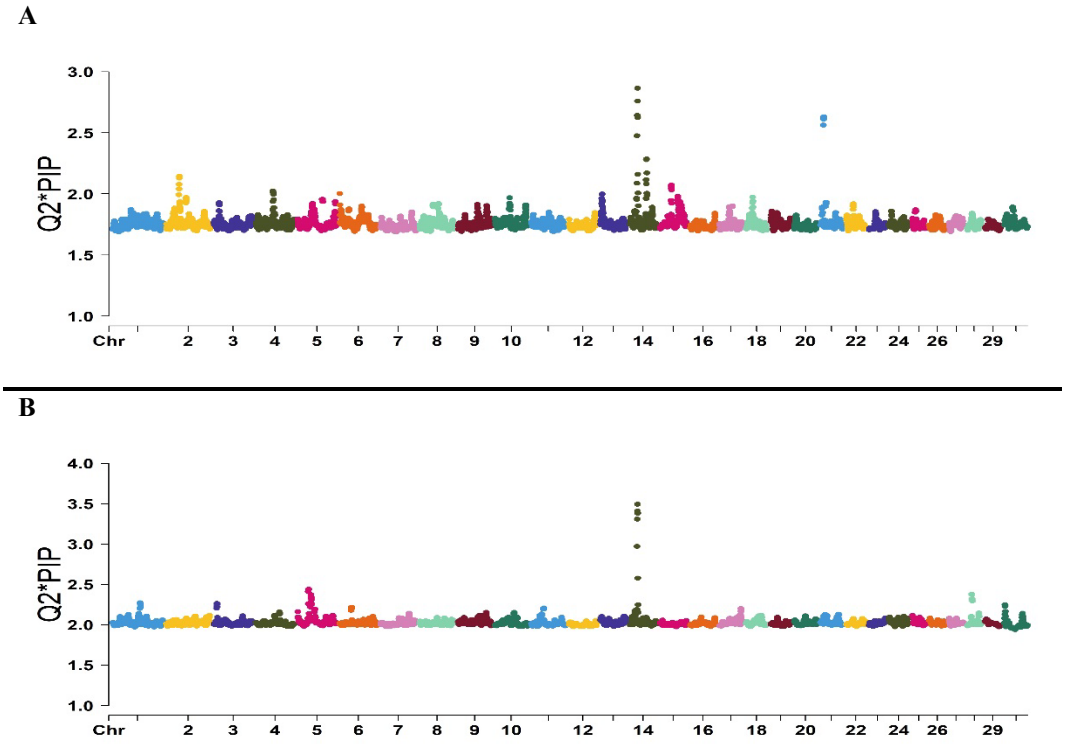

Figure 3. The posterior inclusion probability (PIP) weighted by the Q2 probability for age at detection of first corpus luteum (A) and blood concentration of insulin growth hormone (B) from BayesR bivariate model

ACKNOWLEDGEMENTS

This research was conducted using the legacy database of the Cooperative Centre for Beef Genetics Technologies and their core partners, including Meat and Livestock Australia. The authors acknowledge all the Cooperative Centre for Beef Genetics Technologies participants for their efforts in conducting the field trials and recording the phenotypes.

REFERENCES

Breen E.J., MacLeod I.M., Ho P.N., Haile-Mariam M., Pryce J.E., Thomas C.D., Daetwyler H.D., and Goddard, M.E. (2022) *Comm. Bio*. **5:** 661.

Fortes M.R., Reverter A., Hawken R.J., Bolormaa S. and Lehnert S.A. (2012) *Bio. Reprod*. **87**: 58.

Hawken R.J., Zhang Y.D., Fortes M.R.S., Collis E., Barris W.C., Corbet N.J., Williams P.J., Fordyce G., Holroyd R.G., Walkley J.R.W. and Barendse W. (2012) *J. Anim. Sci*. **90:** 1398.

Kemper K.E., Reich C.M., Bowman P.J., Vander Jagt C.J., Chamberlain A.J., Mason B.A., Hayes B.J. and Goddard M.E. (2015) *Genet. Sel. Evol.* **47:** 1.

Olasege B.S., Porto-Neto L.R., Tahir M.S., Gouveia G.C., Cánovas A., Hayes B.J. and Fortes M.R. (2022) *BMC Genom*. **23:** 1.

Olasege B.S., Tahir M.S., Gouveia G.C., Kour J., Porto-Neto L.R., Hayes B.J. and Fortes M.R. (2021) *Anim. Prod. Sci.* **16:** 1863.

Petrini J., Iung L., Rodriguez M.A., Salvian M., Pértille F., Rovadoscki G.A., Cassoli L.D., Coutinho L.L., Machado P.F., Wiggans G.R. and Mourão, G.B. (2016) *J. Anim. Breed. Genet.* **133:** 384.