Proc. Assoc. Advmt. Anim. Breed. Genet. 25: 79 - 82

GENE NETWORK PREDICTION FOR BULL FERTILITY TRAITS

W.L.A Tan[1], N. Hudson[2], L.R. Porto-Neto[3], A. Reverter[3], J. Afonso[4], M.R.S. Fortes[1]

[1]The University of Queensland, School of Chemistry and Molecular Bioscience, St Lucia, QLD 4072 Australia
[2]The University of Queensland, School of Agriculture and Food Sciences, Gatton, QLD 4343 Australia
[3]CSIRO Agriculture & Food, St Lucia, QLD 4067 Australia
[4]Empresa Brasileira de Pesquisa Agropecuária, Pecuária Sudeste, São Carlos, São Paulo, 13560-970 Brazil

SUMMARY

Most beef breeding herds globally still use natural mating, and therefore, conception rates are influenced by bull fertility. Many indicator traits are captured in the Bull Breeding Soundness Evaluation (BBSE). This paper uses a set of BBSE phenotypes subjected to Genome-Wide Association Studies (GWAS) to predict a gene co-association network. Gene networks can be used to mine the genetic basis of complex traits, thereby deriving a better biological understanding of the underlying mechanisms and informing genomic predictions. Here we described how a dataset of BBSE traits in a multibreed population resulted in a network of 537 connected genes whose topology and prediction will serve as the starting point for future work.

INTRODUCTION

The standardised Bull Breeding Soundness Examination (BBSE) intends to evaluate bulls' traits relevant to fertility (Entwistle and Fordyce 2003). The quantitative traits of BBSE are heritable (0.17 to 0.57) (Corbet *et al.* 2013; Porto-Neto *et al.* 2023) and possibly suitable for improvement via genomic selection. Previously, we have performed a multibreed sequence level GWAS (~ 13 million SNPs), which includes data from 6,422 beef bulls. As a result, we identified 179440 variants associated with one or more of the seven BBSE traits tested (unpublished results). The traits were body weight, condition score, scrotal circumference, sheath score, and semen morphology. In an effort to take these results beyond simple associations with our phenotypes of interest and explore underlying biology, this study utilises an Association Weight Matrix (AWM) (Fortes *et al.* 2010) approach to identify co-associations between SNPs and build a gene network. SNP selection through the AWM could highlight genes that potentially explain a key fertility phenotype, giving us insight into the genetics of bull fertility.

MATERIALS AND METHODS

Animals and phenotypes. BBSE records from 6,422 bulls comprising six different breeds were included in this study. Two breeds were research herds from the Cooperative Research Centre for Beef Genetic Technologies (Beef CRC) consisting of 1,051 Brahman (BRH) and 1,819 Tropical Composite bulls (TRC). The remaining four breeds were obtained from industry, which consists of 1,288 Santa Gertrudis (SGT), 760 Droughtmasters (DMT), 844 Ultra blacks (UBK), and 660 Belmont Tropical Composite (BTC). Descriptive statistics of BBSE records obtained for these six populations are shown in Table 1. Phenotypes include body weight (Weight), body condition score (CS), scrotal circumference (SC), sheath score (Sheath), percent normal sperm (PNS), proximal droplets (PD) and mid-piece abnormalities (MP).

Genotypes. Most animals were genotyped at ~ 50K. A reference panel that utilised BeefCRC and industry animals that were at higher density (~700K) and sequence level (~25 million) were used to impute animals to higher density and, subsequently, to sequence level. The animals used in

the reference population was representative of the bulls used in this study (Porto-Neto *et al.* 2021). This was conducted using a phased reference generated by Eagle 2 (v2.4.1) and then imputed using Minimac3 for autosomes and Minimac 4 for Chromosome X. Imputation $r^2 > 0.8$, a call rate > 0.9 and a minor allele frequency > 0.01 were kept for further analysis. After quality control, a total of 13,398,171 SNPs, including 92,134 SNPs mapped onto the X chromosome. After running a Leave One Chromosome Out (LOCO) GWAS in GCTA (Yang *et al.* 2011), a total of 179,440 variants were significant ($P < 5 \times 10^{-8}$) for at least one trait. 19, 337 variants were significant for two or more traits.

Table 1. The number of records and descriptive statistics of the observed traits*

	N[A]	Mean[B]	SD[C]	Min[D]	Max[E]
Weight, kg	6014	391.59	98.65	124.00	810.00
CS, score	5917	2.96	0.37	2.00	4.00
SC, cm	6235	30.82	4.26	15.50	52.50
Sheath, score	6417	3.19	1.77	1.00	9.00
PNS, %	6055	61.76	27.53	0.00	100.00
PD, %	6052	13.50	19.96	0.00	96.00
MP, %	6052	11.39	11.04	0.00	83.00

[A] Number of records available for a trait. [B] Mean of a trait. [C] Standard deviation of a trait. [D] Minimum value of the trait. [E] Maximum value of the trait.

AWM-PCIT methodology. The AWM was constructed using the procedure described by (Fortes *et al.* 2010). This method applies a series of selection steps to choose relevant SNPs from the 179440 significant variants base on our previous GWAS study (Figure 1). Firstly, we only considered significant SNPs that mapped to genes expressed in the testis, which were previously reported by de Lima *et al.* (2021). PNS was chosen as the key phenotype for the AWM as sperm morphology is an important aspect of bull fertility that is heritable (0.24) and correlated with commonly used bull fertility indices (Attia *et al.* 2016; Butler *et al.* 2019; Porto-Neto *et al.* 2023). We selected SNPs that were associated with PNS ($P < 0.05$). If SNPs were not associated with PNS but with at least three other traits ($P < 0.05$), these SNPs were also kept. The final selection step for the AWM chose SNPs that map to coding regions or was within 2,500 bp of known genes. SNP-to-gene mapping was done using the Map2NCBI package (Hulsman Hanna and Riley 2014) in R. SNPs were grouped by gene to map one representative SNP per gene. This was achieved by selecting the SNP within each gene group associated with the highest number of phenotypes. Next, SNPs within each group were chosen using the most significant average *p*-value across traits. The result is a matrix with rows representing genes (I) and columns representing phenotypes (J). Each element (I, J) contains the association of the SNP to the phenotype. We applied the partial correlation and information theory (PCIT) algorithm described by Reverter and Chan (2008) to the AWM. This algorithm assigns zero for non-significant correlations and retains significant correlations to establish edges in the network (Reverter and Chan 2008). The PCIT algorithm allows for a less stringent threshold ($P < 0.05$) to be used, because SNPs are highlighted based on a number of features and not just it's association to the phenotype (Reverter and Chan 2008). The correlation values can be used as input for Cytoscape (Shannon *et al.* 2003) to establish gene interactions in the gene network analyses.

RESULTS AND DISCUSSION

The gene network constructed using the AWM is shown in Figure 2. The network contains 537 genes forming two distinct clusters with 279 genes on the left and 237 genes on the right. Among these genes, 21 are transcription factors (TF). This network can serve as a starting point for further downstream analysis that can serve two aims: biological discovery and genomic prediction.

For example, biological discovery with the STRING database (Szklarczyk *et al.* 2020) will perform functional enrichment analysis to derive biological information from the gene network. Recent efforts have shown that biological data and the discovery of causal variants can positively impact genomic prediction (Xiang *et al.* 2021). Botelho *et al.* (2021) proposed AWM weighted single step genomic best linear unbiased prediction (AWM-WssGBLUP) as a method to derive weights when building the genomic relationship matrix (G). However, this method did not significantly increase the predictive ability of genomic predictions in their dataset of boar taint compounds. Nonetheless, biological information can still be useful in genomic predictions. Tahir *et al.* (2022) showed that slight improvements in predictive accuracy could be attained using biologically informed SNPs in heifer fertility traits. The SNPs that underpin the network described here are leads for causal variants that could be used to improve predictions of bull fertility traits.

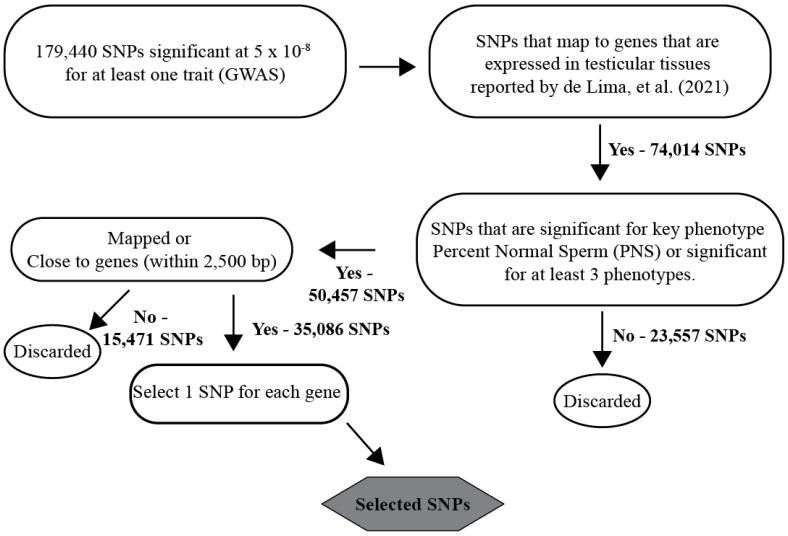

Figure 1. AWM SNP selection flow chart

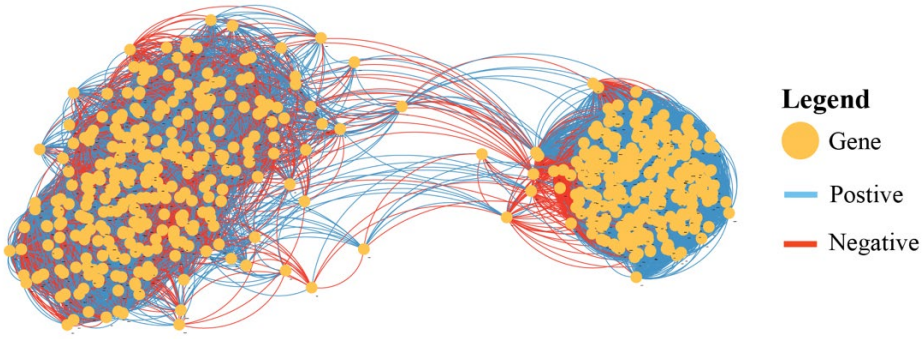

Figure 2. Gene network derived from the Association Weight Matrix (AWM)

CONCLUSION

The gene network created using the AWM highlights several genes and TFs associated with bull fertility traits. These genes and TFs, together with the significant SNP in our sequence-level GWAS, are promising leads to discover causal variants important for bull fertility. This network can be a starting point for further downstream analysis, giving insight into important molecular mechanisms for bull fertility traits.

ACKNOWLEDGMENTS

This work was supported by the Bull Fertility update project, an initiative from CSIRO, the University of Queensland, and Meat & Livestock Australia. The authors are thankful to the beef producers who took part in this collaborative effort and Meat & Livestock Australia that co-funded the project (L.GEN.1818).

REFERENCES

Attia S., Katila T. and Andersson M. (2016) *Reprod. Domest. Anim.* **51:** 54.

Botelho M.E., Lopes M.S., Mathur P.K., Knol E.F., Guimarães S.E.F., Marques D.B.D., Lopes P. S., Silva F.F. and Veroneze R. (2021) *J. Anim. Breed. Genet.* **138:** 442.

Butler M.L., Bormann J.M., Weaber R.L., Grieger D.M. and Rolf M.M. (2019) *Transl. Anim. Sci.* **4:** 423.

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.

de Lima A.O., Afonso J., Edson J., Marcellin E., Palfreyman R., Porto-Neto L.R., Reverter A. and Fortes M.R.S. (2021) *Front. Genet*. **12:** 610116.

Entwistle K. and Fordyce G. (2003) 'Evaluating and reporting bull fertility'. Australian Association of Cattle Veterinarians.

Fortes M.R.S., Reverter A., Zhang Y., Collis E., Nagaraj S.H., Jonsson N.N., Prayaga K.C., Barris W., Hawken R.J. (2010) *Proc. Natl. Acad. Sci.* **107:** 13642.

Hulsman Hanna L.L. and Riley D.G. (2014) *Livest. Sci*. **162:** 59.

Porto-Neto L.R., McWilliam S.M., Alexandre P.A., Reverter A., McGowan M., Fortes M.R.S., Hayes B. and Betram J. (2021) Meat and Livestock Australia, North Sydney.

Porto-Neto L.R., Alexandre P.A., Hudson N.J., Bertram J., McWilliam S.M., Tan A.W.L., Fortes M.R.S., McGowan M.R., Hayes B.J. and Reverter A. (2023) *PloS one* **18:** e0279398.

Reverter A. and Chan E.K.F. (2008) *Bioinformatics* **24:** 2491.

Shannon P., Markiel A., Ozier O., Baliga N.S., Wang J.T., Ramage D., Amin N., Schwikowski B. and Ideker T. (2003) *Genome Res.* **13:** 2498.

Szklarczyk D., Gable A.L., Nastou K.C., Lyon D., Kirsch R., Pyysalo S., Doncheva N.T., Legeay M., Fang T., Bork P., Jensen L.J. and von Mering C. (2020) *Nucleic Acids Res.* **49:** D605.

Tahir M.S., Porto-Neto L.R., Reverter-Gomez T., Olasege B.S., Sajid M.R., Wockner K.B., Tan A.W.L. and Fortes M.R.S. (2022) *J. Anim. Sci.* **100:** 340.

Xiang R., MacLeod I.M., Daetwyler H.D., de Jong G., O'Connor E., Schrooten C., Chamberlain A. J. and Goddard M.E. (2021) *Nat. Commun.* **12:** 860.

Yang J., Lee S.H., Goddard M.E. and Visscher P.M. (2011) *Am. J. Hum. Genet.* **88:** 76.