GENETIC CONTROL OF FERTILITY TRAITS ACROSS SPECIES: VARIANCE IN TROPICAL BEEF HEIFERS' AGE AT PUBERTY EXPLAINED BY GENES CONTROLING AGE AT MENARCHE IN WOMEN

R. Costilla, C. Warburton and B.J. Hayes

Centre for Animal Science, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, 4072 Australia

SUMMARY

Fertility traits are of paramount importance for humans and cattle. In cattle, they are one of the main profit drivers in the industry. Using data from genome-wide association studies (GWAS) from both species, we estimated the effect of genes associated with age at menarche in women (AaM) in the variance of age at puberty (AaP) in tropically adapted beef heifers. We found that variants within 100kb of AaM bovine orthologous genes explained 11.2% of the additive genetic variance of heifers AaP in the biggest cohort analysed. This represented about twice the variance explained by random gene-sets of similar size and number of SNPs ($P<0.2$). Our work suggests some potential of cross-species analyses to increase the cattle industry's productivity.

INTRODUCTION

Thanks to the recent advances in biomedical technology, the genetic basis of fertility in humans is better known now than ever. For instance, the biggest GWAS for female fertility to date with ~370,000 women, Day *et al.* (2017), reported hundreds of genomic loci associated with AaM in women, a female complex trait that is a milestone in pubertal development. An interesting question is, whether we can use the information coming out of the extremely powerful GWAS in humans to improve genomic predictions for related traits in cattle?

Given the evidence for genetic control of complex traits across mammalian species (Pryce *et al.* 2011; Bouwman *et al.* 2018), we hypothesised that genetic factors contributing to variation between individuals for age at puberty/age at menarche will be shared across humans and cattle. In humans, the heritability of AaM was estimated to be 0.32 (0.03) (Day *et al*. 2017). In cattle, AaP has been shown to be moderately to highly heritable in tropically adapted breeds (Johnston *et al.* 2009; Corbet *et al.* 2018) with heritabilities ranging from 0.22 (0.07) to 0.57 (0.12) for Santa Gertrudis and Brahman breeds respectively. Using bovine orthologous of genes associated with AaM, we estimated their contribution to the additive genetic variance of age at puberty (AaP) in heifers.

MATERIALS AND METHODS

Animals, genotypes and phenotypes. We used published data from several heifer populations: Beef Cooperative Research Centre for Beef Technology Brahman and Tropical Composite (CRC BB and CRC TC, respectively) and the Queensland Smart Futures population (Smart Futures). These herds contained heifers from several tropical beef breeds and were genotyped with the BovineSNP50 (CRC BB and CRC TC) and Geneseek GGP-LD array (Smart Futures). The Smart Futures heifers consisted of animals from three breeds: Brahman (979), Santa Gertrudis (1813) and Droughtmaster (914). Complete details for these animals and genotypes have been published elsewhere (Johnston *et al.* 2009; Corbet *et al.* 2018). In total, we used 3695, 960 and 868 animals from the Smart Futures, CRC TC and CRC BB herds. Genotypes were imputed twice up to 728,785 SNPs using Fimpute (Sargolzaei *et al.* 2014) and then to whole genome sequence using Minimac3 (Das *et al.* 2016). The phenotypes were age in days at first corpus luteum (AGECL) and corpus luteum score (CLscore)

at ~600 days for the CRC and Smart Futures cohorts, respectively. The AGECL is a count variable and CLscore is an ordinal variable ranging from 0 "infantile tract" to 5 "pregnancy > 10 weeks". These two heifers AaP phenotypes, CLscore and AGECL, exhibit a very high genetic correlation (-0.83(0.09), Engle *et al.* 2019).

Bovine orthologous AaM genes. Using coding variation (nonsynonymous SNPs), associated expression in neural tissues (eQTL) and chromatin interaction data (Hi-C), Day *et al.* (2017) implicated 233 protein-coding genes in the regulation of AaM in women. We mapped these genes to the UMD3.1 bovine genome using Biomart Ensemb 94 and filtered them out by conservation status (orthology confidence=1 and gene identity > 60%), rendering a total of 205 highly conserved orthologous AaM genes in the bovine genome. Then, we located variants (SNPs and INDELs) in or around + 100kb using imputed sequence data from the CRC BB, CRC TC and SMF cohorts.

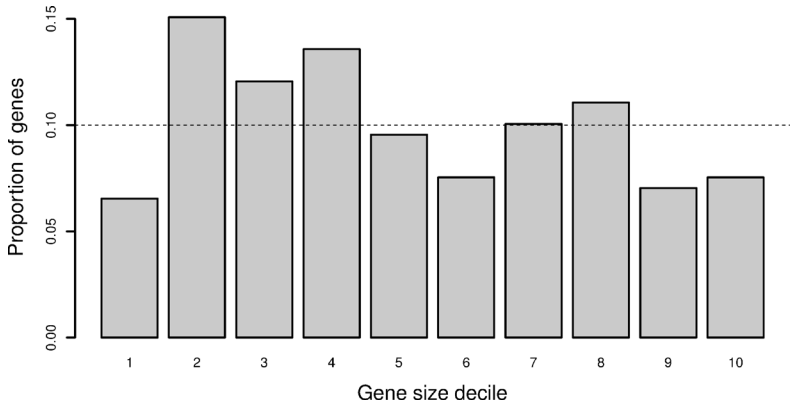

Figure 1. Gene size distribution (deciles) for bovine orthologous genes for age at menarche (AaM) in women

Statistical analysis. We estimated the variance of heifers' AaP explained by orthologous AaM genes using a model with two genomic relationship matrices (GRMs) constructed from the imputed to sequence genotypes described before. The first GRM is constructed from variants in or within 100kb of AaM genes and the second GRM from the remaining variants in the bovine genome. The model included additional continuous and categorical covariates as follows:

$$y = 1_n\mu + age + pc1 + pc2 + cgroup + g_1 + g_2 + \varepsilon$$

where y is a vector of phenotypes, μ the overall mean, 1_n is a vector of 1s, *age* is a vector with the heifers' age fitted as a continuous covariate, *pc1* and *pc2* the first and the second principal components (derived from the GRM), *cgroup* is a vector of contemporary groups that includes with herd, year, and season and is fitted as categorical covariate. g_1 and g_2 are vectors of random effects for the variants in or within 100kb of AaM genes and the remaining ones in the bovine genome with $g_1 \sim N(0, G_1\sigma^2_{g1})$ and $g_2 \sim N(0, G_2\sigma^2_{g2})$. ε is a vector of random residuals distributed $\varepsilon \sim N(0, \sigma^2_\varepsilon)$. G_1 and G_2 denote the corresponding GRM matrices constructed following the first method of VanRaden (2008) and σ^2_{g1}, σ^2_{g2}, σ^2_ε the corresponding genetic and error variances. We fitted the model separately for each cohort using GCTA (Yang *et al.* 2011).

In order to provide an appropriate comparison for the AaM genes, we also estimated the variance explained by 100 random gene-sets of similar length and SNP number, e.g. we performed a stratified random sampling by quantiles of gene size and number of SNPs, and ran a randomized permutation test for the percentage of AaP variance explained by AaM genes.

RESULTS AND DISCUSSION

Out of a total of 28.9 million imputed to sequence variants across all cohorts, there were 339,669 variants within +/- 100kb from 205 bovine orthologous AaM genes. The number of variants varied slightly within individual cohorts. Note also that in terms of gene physical size, AaM genes are over-represented in the lower deciles and thus tend to be smaller in size than the rest of protein-coding genes in the bovine genome (Figure 1).

Variants in AaM genes explained 2.5% phenotypic (11.2% genetic) variance of heifers AaP in the biggest cohort, Smart Futures (Table 1). This represented about twice the mean variance explained by variants in random gene-sets (1.2% phenotypic and 5.6% genetic) that had on average 379,325 variants. This result however did not reach significance in the randomized permutation test (P<0.2) (Figure 2). With regard to the CRC cohorts, variants in AaM genes explained negligible percentages when compared with variants in random gene-sets.

Table 1. SNP based heritability (h^2) partition for cohorts included in the meta-analysis

	Cohort					
	Smart Futures		CRC TC		CRC BB	
Component	h^2	se	h^2	se	h^2	se
AaM genes: V(G1)/Vp	0.025	0.019	0.005	0.059	0.015	0.058
Remaining: V(G2)/Vp	0.195	0.035	0.393	0.103	0.446	0.108
Overall: V(G1)+V(G2)/Vp	0.220	0.031	0.398	0.085	0.461	0.092
V(G1)/Vp for random gene-sets*	0.012	0.001	0.034	0.009	0.022	0.009

*Mean for 100 gene-sets (379,325 variants on average).

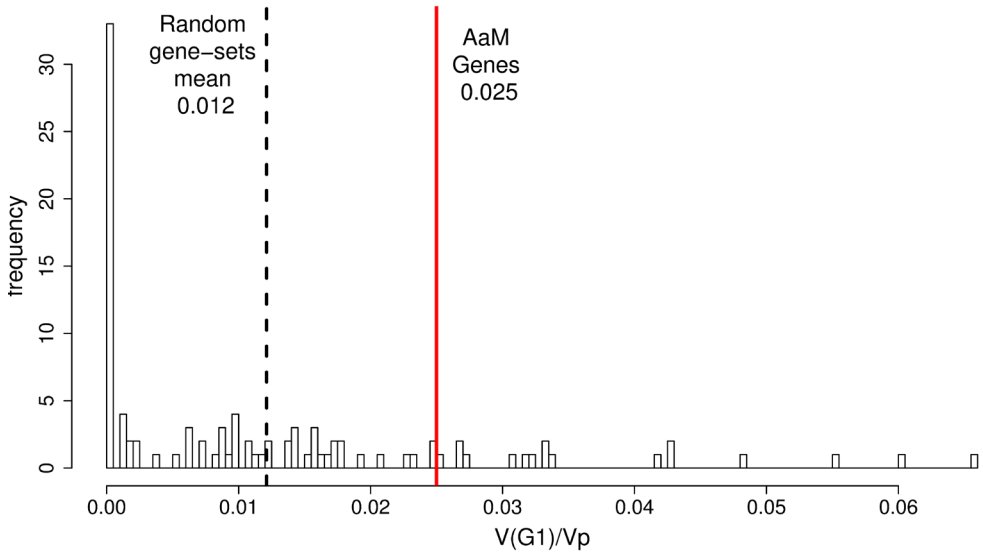

Figure 2. Randomised permutation test results for the Smart Futures cohort. Variance in heifers age at puberty (AaP) explained by age at menarche (AaM) genes (red line, 339,669 variants), and random gene-sets of similar size to AaM genes. Dotted lined displays the mean for 100 random gene-sets (379,325 variants on average)

Note that overall h^2 estimates by cohort: 0.220(0.031), 0.398(0.085), and 0.461(0.092) for Smart Futures, CRC TC and CRC BB, respectively, are consistent with previous estimates from published studies (Johnston *et al.* 2009; Corbet *et al.* 2018). In terms of individual genes, there were four genes in the AaM set (*ZNF654, LEPROT, CCDC40, CLUAP1)* that reached significance ($P<10^{-4}$) in the meta-analysis of AaP GWAS across the three cohorts. In humans, these genes are also associated with haemoglobin concentration (*ZNF654*), morbid obesity (*LEPROT),* blood protein levels (*CCDC40*), vital capacity and leukocyte count (*CLUAP1)* (Stelzer *et al.* 2016).

Taken together these results suggest that women's AaM genes are also associated with a similar phenotype in a different species, in this case fertility phenotypes in tropically adapted beef heifers. Importantly, however, is the issue of power for this complex trait as a large number of animals was required to pick up this signal, e.g. association was only presented in the biggest cohort with 3695 animals. An interesting extension would be to combined both CRC cohorts (Brahmans and composites) and performed the analyses presented here on this combined dataset.

CONCLUSIONS

Variants in AaM genes explained 2.5% phenotypic (11.2% genetic) variance of tropical beef heifers' AaP in the biggest cohort analysed here. This is about twice the variance explained by similar random gene-sets, although this result is not statistically significant ($P<0.2$), and the variance explained in the other cohorts was not different from zero. Some genes affecting AaM were also significant for AaP in heifers ($P<1 \times 10^{-4}$). Our work highlights the potential of cross-species analyses to increase the industry's productivity. Further research in terms of inclusion of variants in AaM genes in genomic prediction models is needed to achieve this potential.

ACKNOWLEDGEMENTS

We gratefully acknowledge the contribution of Drs. Nicholas J. Corbet, Jack M. Allen, Alan R. Laing, Geoffrey Fordyce, Russel Lyons, Michael R. McGowan, and Brian M. Burns who collected the data used in this study.

REFERENCES

Bouwman, A. C., Daetwyler H. D., Chamberlain A. J., Ponce C. H., Sargolzaei M., Schenkel F. S. *et al.* (2018) *Nat. Genet.* **50**: 362.

Corbet N. J., Allen J. M., Laing A. R., Fordyce G., McGowan M. R. and Burns B. M. (2018) *Anim. Prod. Sci*. **58**: 1735.

Das S., Forer L., Schonherr S., Sidore C., Locke A. E., *et al.* (2016) *Nat. Genet. **48***: 1284.

Day F. R., Thompson D. J., Helgason H., Chasman D.I., Finucane H., Sulem P.*et al.* (2017) *Nat. Genet.* **49**: 834.

Engle B. N., Corbet N. J., Allen J. M., Laing A. R., Fordyce G., McGowan M. R., Burns B. M., Lyons R. E. and Hayes B. J. (2019) *J. Anim. Sci.* **97**: 90.

Johnston D. J., Barwick S. A., Corbet N. J., Fordyce G., Holroyd R.G., Williams P.J. and Burrow H. M. (2009) *Anim. Prod. Sci*. **49**: 399.

Pryce, J. E., Hayes B. J., Bolormaa S. and Goddard M. E. (2011) *Genetics* **187**: 981.

Sargolzaei M., Chesnais J. P., and Schenkel F. S. (2014) *BMC Genomics*. **15**: 478.

Stelzer G., Rosen R., Plaschkes I., Zimmerman S., Twik M., Fishilevich S., *et al.* (2016) *Curr. Prot. in Bioinfo.* **54**: 1.30.1.

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

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