Proc. Assoc. Advmt. Anim. Breed. Genet: 24: 345-348

ESTIMATION OF OPTIMUM POLYGENIC AND GENOMIC WEIGHTS IN SINGLE-STEP GENETIC EVALUATION OF CARCASS TRAITS IN AUSTRALIAN ANGUS BEEF CATTLE

A.M. Samaraweera[1], J.A. Torres-Vázquez[1], M G. Jeyaruban[1], D.J. Johnston[1] and V. Boerner[1]

[1]Animal Genetics Breeding Unit*, University of New England, Armidale, NSW, 2350 Australia

SUMMARY

Optimum polygenic and genomic weights enhance the accuracy of breeding value estimates in single-step genomic evaluations. This study estimated the contribution from marker information to total additive genetic variation referred as λ using an extended single-step model in a multi-trait variance component estimation based procedure using data for six Australian Angus carcase traits. The λ for these traits ranged from 0.54 (for carcass intramuscular fat) to 0.79 (for carcass eye muscle area). Heritabilities were similar between the pedigree only and the extended single-step multi-trait model when using the total genetic variance, and ranged from 0.37 (for carcass rib fat) to 0.53 (for carcass weight), suggesting that the single-step model did not explain more genetic variance than pedigree based models. Results suggest that the scalar λ in the current single-step routine evaluation could be replaced by an extended single-step model allowing for different proportions of the additive genetic co-variance explained by markers for all elements of the genetic co-variance matrix.

INTRODUCTION

Increasing availability of genomic information requires ongoing modification to incorporate genotypes efficiently in routine genetic evaluation of Australian beef cattle. Single-step genomic evaluation developed by Legarra *et al*. (2009) and Christensen and Lund (2010) combines both pedigree and genotypes in a unified analysis. This method integrates numerator relationship matrix ($\boldsymbol{A}$) and genomic relationship matrix ($\boldsymbol{G}$) into a single $\boldsymbol{H}$ matrix, depicting co-variance between both genotyped and non-genotyped animals in the analysis. An improved $\boldsymbol{G}$ matrix ($\boldsymbol{G_w}$) that can be obtained as $\lambda\boldsymbol{G} + (1-\lambda)\boldsymbol{C}$ was suggested by Christensen and Lund (2012), where $\boldsymbol{C}$ is often the numerator relationship matrix among the genotyped animals, and λ is a non-zero weight with $0 < \lambda < 1$. λ is usually referred to as the proportion of additive genetic variance explained by the marker effects. For current BREEDPLAN single-step multi-trait breeding value estimation λ is set to a scalar value of 0.5, implying that for all genetic co-variances in the model, the same proportion is explained by markers.

Previous studies aimed at estimating λ by a cross-validation grid-search procedure to maximise the accuracy of predicted breeding values expressed as $\left(cov(\hat{u}, y)/\sigma_{\hat{u}}\sigma_y\right) * \sqrt{1/h^2}$, with the cross-validations performed on single trait data sets using a genetic variance $H\sigma$ (McMillan *et al*., 2017; Zhang *et al*. 2017). However, the problem with the cross-validation approach is that contradicting values for λ in two single-trait analysis are difficult to accommodate when both traits are included in a multi-trait evaluation. Further, a multi-trait cross-validation grid-search would have to evaluate a high dimensional grid, which makes the approach computationally infeasible.

It can be shown that a model using $\boldsymbol{G_w} = \lambda\boldsymbol{G} + (1-\lambda)\boldsymbol{C}$ is simply the condensed form of the extended single-step model containing two genetic factors, one using $\boldsymbol{A}\otimes\boldsymbol{\Sigma_A}$ and the other using

* A joint venture of NSW Department of Primary Industries and the University of New England

$\boldsymbol{H}\otimes\boldsymbol{\Sigma_H}$ with $\boldsymbol{G_w} = \boldsymbol{G} + \boldsymbol{I}0.001$ where $\boldsymbol{\Sigma_A}$ and $\boldsymbol{\Sigma_H}$ are co-variance matrices and $\boldsymbol{I}$ is an identity matrix. The total genetic variance $\boldsymbol{\Sigma_G} = \boldsymbol{\Sigma_H} + \boldsymbol{\Sigma_A}$ with a scalar λ only being obtainable if $\boldsymbol{\Sigma_H} \oslash \boldsymbol{\Sigma_G} \equiv \boldsymbol{ii}'k$ where k is a scalar and $\boldsymbol{i}$ is an identity vector. Therefore, the partitioning of the genetic variance implicit in λ can be obtained by variance component estimation using the general model with two genetic factors, where the results may not support a scalar λ and in turn may require the use of the general model in genetic evaluation. However, the estimation of variance components for such a model via restricted maximum likelihood (REML) is challenging due to the mixed model coefficient matrix containing large non-zero blocks, and REML algorithms using the phenotypic co-variance matrix are severely limited with regard to the number of observations that can be accommodated.

This study investigated methods to optimally partition the genetic variances in Australian Angus carcass data. To overcome REML limitations, Bayesian methods were used.

MATERIALS AND METHODS

A total of 59,616 pre-corrected records (Graser *et al*. 2005) for Australian Angus carcass traits were analysed consisting of carcass weight (CWT), carcass rib fat (CRF), carcass P8 fat (CP8), carcass eye muscle area (CEA), carcass retail beef yield percentage (CMY), and carcass intramuscular fat (CIM). Numbers of phenotypes, genotypes, and pedigree information available for each trait are given in Table 1. The pedigree consisted of 2.6 million animals, 110,000 of which were genotyped with 56,009 markers per genotype.

Table 1. Number of phenotypic records, number of genotyped animals, and descriptive statistics for carcass traits, weight (CWT (kg)), rib fat (CRF (mm)), P8 fat (CP8 (mm)), eye muscle area (CEA (cm^2)), retail beef yield (CMY (%)), and intramuscular fat (CIM (%))

Trait	Records	Genotyped	Mean	Stddev	Minimum	Maximum
CWT	16875	3340	422.9	60.2	186.8	636.0
CRF	5319	1059	15.5	5.1	1.6	36.8
CP8	14793	3054	19.7	5.6	1.6	42.7
CEA	7392	839	83.9	9.1	41.8	120.7
CMY	2140	505	69.0	4.6	55.8	77.9
CIM	13097	2630	8.8	3.4	1.7	30.9

Models. A multi-trait linear mixed model (model 1) was fitted as follows:

$$\begin{pmatrix} y_1 \\ \cdot \\ y_6 \end{pmatrix} = \begin{pmatrix} X_1 & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & X_6 \end{pmatrix} \begin{pmatrix} b_1 \\ \cdot \\ b_6 \end{pmatrix} + \begin{pmatrix} Z_1 & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & Z_6 \end{pmatrix} \begin{pmatrix} u_1 \\ \cdot \\ u_6 \end{pmatrix} + \begin{pmatrix} e_1 \\ \cdot \\ e_6 \end{pmatrix}$$

where $[y_1,.,y_6]$ is a vector of phenotypic observations for traits 1 to 6, matrices $[X_1,.,X_6]$ and $[Z_1,.,Z_6]$ link fixed effects of contemporary group and random additive genetic effects, respectively to their respective observations, and $[e_1,.,e_6]$ is a vector of residuals. $[u_1,.,u_6]\sim N([0,.,0], \boldsymbol{A}\otimes\boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is the co-variance matrix between genetic factors and **A** is the pedigree derived co-variance matrix between animals.

The single-step multi-trait linear model (model 2) was fitted as follows:

$$\begin{pmatrix} y_1 \\ \cdot \\ y_6 \end{pmatrix} = \begin{pmatrix} X_1 & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & X_6 \end{pmatrix} \begin{pmatrix} b_1 \\ \cdot \\ b_6 \end{pmatrix} + \begin{pmatrix} Z_1 & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & Z_6 \end{pmatrix} \begin{pmatrix} u_1 \\ \cdot \\ u_6 \end{pmatrix} + \begin{pmatrix} Z_1 & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & Z_6 \end{pmatrix} \begin{pmatrix} g_1 \\ \cdot \\ g_6 \end{pmatrix} + \begin{pmatrix} e_1 \\ \cdot \\ e_6 \end{pmatrix}$$

where $[u_1,.,u_6]\sim N([0,.,0], \boldsymbol{A}\otimes\boldsymbol{\Sigma_A})$ is a vector of polygenic effects and $[g_1,.,g_6]\sim N([0,.,0], \boldsymbol{H}\otimes\boldsymbol{\Sigma_H})$ is a vector of genomic effects. Matrix $\boldsymbol{H}$ contains a genomic relationship matrix $\boldsymbol{G} = \boldsymbol{MM}' + \boldsymbol{C}$, where $\boldsymbol{M}$ is centred and scaled marker genotypes matrix and $\boldsymbol{C}$ is a diagonal matrix of small values (e.g. 0.0001) ensuring invertability of $\boldsymbol{G}$. The total additive genetic variance ($\boldsymbol{\Sigma_G}$) is equal to $\boldsymbol{\Sigma_A} + \boldsymbol{\Sigma_H}$, and when there are no genotyped animals model 2 essentially becomes

model 1. Therefore, an underlying assumption about λ is that $\boldsymbol{u}$ and $\boldsymbol{g}$ are vectors of orthogonal random effects.

Variance components based on model 1 and model 2 were estimated using Gibbs sampling. The analysis were conducted with model 1 using $\boldsymbol{A}$ and with model 2 using $\boldsymbol{A}$ and $\boldsymbol{H}$, where in both cases the blocks of $\boldsymbol{A}$ and $\boldsymbol{H}$ related to the union of all phenotyped individuals were extracted from $\boldsymbol{A}$ and $\boldsymbol{H}$ built using all animals in the pedigree and all available genotypes. For the pedigree model prior variances were calculated from the phenotypic variances. For the extended single-step model prior variances were those obtained from the pedigree model, with a prior variance partitioning equal to $\boldsymbol{\Sigma_H} = 0.1\boldsymbol{\Sigma_G}$ and $\boldsymbol{\Sigma_A} = 0.9\boldsymbol{\Sigma_G}$. However, for both models the prior weight was zero. Variance components and genomic weights were obtained by discarding the first 30,000 samples as burn-in and averaging the sum of every 100th sample from a total of 200,000 samples.

RESULTS AND DISCUSSION

The heritabilities for six carcass traits for model 1 which used $\boldsymbol{A}$ as the between animals co-variance matrix, and for model 2 where the variances were partitioned between the genomic and polygenic factor are presented in Table 2. The total heritabilities for six carcass traits in model 2 ranged from 0.37 for CRF to 0.53 for CWT , and they were almost identical to those derived from model 1 (Table 2).

The proportion of additive genetic variation explained by markers (λ) is greater for almost all carcass traits than the λ assumed in the current BREEDPLAN evaluations of 0.5, and ranged from 0.54 for CIM to 0.79 for CEA (Table 2). Therefore, future genetic evaluations should allow higher and different λ in BREEDPLAN routine genetic evaluation of carcass traits. These results suggest that the BREEDPLAN genetic evaluation model would have to allow for two genetic factors where the implications for model dimensionality, solver convergence rate, and breeding value accuracy must be investigated.

Table 2. Pedigree based heritability (h^2) when using model 1 and matrix A, and polygenic (h_A^2), genomic (h_H^2) and total heritability (h_G^2), genomic weights (λ) and phenotypic variances (σ_p^2) when using model 2 for 6 Australian Angus carcase traits

Parameter	CWT[1]	CRF[2]	CP8[3]	CEA[4]	CMY[5]	CIM[6]
h^2	0.51 (0.03)[a]	0.38 (0.05)	0.45 (0.03)	0.47 (0.04)	0.51 (0.07)	0.46 (0.03)
h_A^2	0.17 (0.03)	0.12 (0.04)	0.13 (0.03)	0.10 (0.04)	0.23 (0.07)	0.22 (0.04)
h_H^2	0.35 (0.02)	0.25 (0.03)	0.34 (0.03)	0.37 (0.03)	0.29 (0.07)	0.25 (0.02)
h_G^2	0.52 (0.03)	0.37 (0.04)	0.47 (0.03)	0.47 (0.04)	0.52 (0.05)	0.47 (0.03)
λ[†]	0.67 (0.04)	0.68 (0.08)	0.73 (0.06)	0.79 (0.08)	0.56 (0.13)	0.54 (0.05)
σ_p^2	844.40 (13)	16.80 (0.40)	21.89 (0.36)	46.51 (0.92)	2.72 (0.10)	5.50 (0.09)

[1]weight, [2]rib fat, [3]P8 fat, [4]eye muscle area, [5]retail beef yield, [6]intramuscular fat; [a]standard deviation from 1700 samples in parenthesis; [†]$\lambda = diag(\boldsymbol{\Sigma_H})/diag(\boldsymbol{\Sigma_G})$

Directions (and values) of between trait total genetic correlations from model 2 were similar to those from model 1 (Table 3). However, comparison of trait correlations between polygenic and genomic factor in model 2 shows that for many traits this correlation is in the in opposite direction (Table 4). One notable example is CEA where positive genetic correlations were observed for polygenic factor whereas those correlations were negative in genomic factor (Table 4). Global correlations between CEA and fat traits (CRF and CP8) were negative regardless of whether model 1 or model 2 was used. However, for model 2 genomic correlations remained negative whereas polygenic correlations turned positive. The opposite pattern was observed for correlations between CEA and CMY, where the global correlation remained positive, but was larger, and the polygenic

correlation turned negative. It needs to be confirmed whether these findings have a biological foundation or were caused by insufficient variance partitioning due to the low number of genotyped and phenotyped animals for CEA and CMY.

Table 3. Genetic correlation (lower triangle) when using model 1 and total genetic correlation (upper triangle) when using model 2 for 6 Australian Angus carcass traits

Trait[1]	CWT	CRF	CP8	CEA	CMY	CIM
CWT		-0.03 (0.02)[a]	-0.12 (0.02)	0.01 (0.02)	0.10 (0.02)	0.04 (0.02)
CRF	-0.02 (0.02)		0.55 (0.01)	-0.14 (0.02)	-0.43 (0.02)	0.01 (0.02)
CP8	-0.12 (0.02)	0.55 (0.01)		-0.19 (0.02)	-0.25 (0.02)	-0.02 (0.02)
CEA	0.08 (0.02)	-0.19 (0.02)	-0.23 (0.02)		0.39 (0.01)	0.05 (0.02)
CMY	0.11 (0.02)	-0.54 (0.03)	-0.29 (0.02)	0.43 (0.01)		-0.08 (0.02)
CIM	0.07 (0.02)	0.02 (0.02)	-0.07 (0.02)	0.03 (0.02)	-0.02 (0.02)	

[1]CWT, weight; CRF, rib fat; CP8, P8 fat; CEA, eye muscle area; CMY, retail beef yield; CIM, intramuscular fat; [a]standard deviation from 1700 samples in parenthesis

Table 4. Polygenic factor correlation (upper triangle) and genomic factor correlation (lower triangle) matrix when using model 2 for 6 Australian Angus carcass traits

Trait[1]	CWT	CRF	CP8	CEA	CMY	CIM
CWT		0.31(0.02)[a]	-0.13 (0.03)	0.22 (0.02)	-0.09 (0.03)	0.33 (0.02)
CRF	-0.20 (0.03)		0.49 (0.01)	0.27 (0.02)	-0.50 (0.04)	0.04 (0.02)
CP8	-0.12 (0.03)	0.57 (0.01)		0.06 (0.02)	-0.22 (0.03)	-0.15 (0.03)
CEA	-0.07 (0.03)	-0.30 (0.03)	-0.27 (0.03)		-0.21 (0.03)	0.10 (0.02)
CMY	0.22 (0.02)	-0.40 (0.04)	-0.27 (0.03)	0.69 (0.01)		-0.06 (0.03)
CIM	-0.15 (0.03)	0.00 (0.03)	0.05 (0.02)	0.03 (0.02)	-0.10 (0.03)	

[1]CWT, weight; CRF, rib fat; CP8, P8 fat; CEA, eye muscle area; CMY, retail beef yield; CIM, intramuscular fat; [a]standard deviation from 1700 samples in parenthesis

CONCLUSIONS

The proportion of additive genetic variation explained by markers (λ) ranged from 0.54 to 0.79 for the six carcass traits in Australian Angus beef cattle. This finding is significant because the current BREEDPLAN single-step evaluation uses a single λ for all traits, 0.5. The results of this study do not support the use of the same λ for all traits. However, accounting for these findings requires a change in the BREEDPLAN model which must be preceded by further investigations into computational feasibility and breeding value accuracy.

ACKNOWLEDGEMENTS

The authors would like to thank the Meat and Livestock Australia (MLA) for their financial support through L.GEN.1704, and Angus Australia and New South Wales Department of Primary Industries (NSW DPI) for providing data used in this study.

REFERENCES

Christensen O.F., and Lund M.S. (2010) *Genet. Sel. Evol.* **42**: 2.
Christensen O.F., Madsen P., Nielsen B., Ostersen T., and Su G. (2012) *Animal*. **6**: 1565.
Graser H.-U., Tier B., Johnston D.J., and Barwick S.A. (2005) *Aust. J. Exp. Agric*. **45**: 913.
Legarra A., Aguilar I., and Misztal I. (2009) *J. Dairy Sci.* **92**: 4656.
McMillan A., and Swan A. (2017) *Proc. Assoc. Advmt. Anim. Breed. Genet*. **22**:557.
Zhang Y., Swan A., Johnston D., and Girard C. (2017) *Proc. Assoc. Advmt. Anim. Breed. Genet*. **22**:309.