Use of genetic variance to determine weighting factors for genomic information used in single step genomic evaluation

Y. Zhang, A.A. Swan & R.G. Banks
Animal Genetics and Breeding Unit, University of New England, Armidale NSW, 2350 Australia
yzhang4@une.edu.au (Corresponding Author)

Summary

Single step genomic BLUP (SS-GBLUP) has recently been implemented in the Australian beef cattle and sheep industries. Selection of the weighting factor () to combine the genomic and pedigree relationship matrixes could affect the accuracies of EBV. Cross validation has been used as a method to determine the optimal . This study estimated genetic variances using models fitting the genomic relationship matrix (G, Vg), the pedigree relationship matrix (A, Va), or both (Vt), and used the ratio Vg/Vt as an alternative approach to identify the optimal . Results showed that the ratio of genetic variances (Vg/Vt) aligned reasonably well with the optimal values of for most traits. Compared with cross validation, this approach was computationally fast, and can be used as an alternative to identify the optimal values of in SS-GBLUP.

Keywords: Single step genomic BLUP, genetic variance weighting, accuracy, beef

Introduction

Application of genomic selection in livestock has the potential to increase genetic gain in comparison to pedigree BLUP evaluations. Genomic BLUP (GBLUP) works in the similar way to pedigree BLUP, but substitutes the pedigree based relationship matrix (A) with the genomic relationship matrix (G). The recently developed single step genomic BLUP (SS-GBLUP) by Legarra *et al.* (2009) and Christensen and Lund (2010) makes use of genotypes, all phenotypes and pedigree in a single analysis, aiming to streamline the application and enhance the accuracy of EBVs. The variance matrix (H) combines A and G (Aguilar *et al.*, 2010). An appropriate weighting of pedigree and genomic information when constructing G is required because genomic markers do not explain all of the additive genetic variation (e.g. Goddard *et al.,* 2011, Legarra *et al.,* 2014). A modified genomic relationship matrix is typically used, as , where is the fraction of the additive genetic variance explained by markers, ranging between 0 and 1, and is the genomic relationship matrix unweighted by pedigree information, and is the pedigree relationship matrix for genotyped animals.. Studies have demonstrated different methods to calculate : 1) use of ratio of variances due to genetic marker Vg to the total genetic variance (Vt, Aguilar *et al.,* 2010, Legarra *et al.,* 2014); 2) goodness of model fit (Vandenplas *et al.*, 2017, Christensen & Lund, 2010), 3) empirical assessment, in which λ is picked by highest accuracies or reliabilities of EBV, using cross-validation (Zhang *et al.,* 2017), and 4) Goddard *et al.* (2011) suggested that λ could be estimated by sampling G and regressing G on A. SS-GBLUP has been implemented in

Australian sheep and beef cattle evaluations with optimal λ identified using cross validations (McMillan & Swan, 2017, Zhang *et al.*, 2017). This study approximated λ values in beef cattle using genetic variances from pedigree based relationship matrixes or/and genomic marker based matrixes, and the outcomes were compared with the results from cross validation.

Materials and methods

Phenotypes, pedigree and genotypes for this study were retrieved from the BREEDPLAN database for Angus cattle, with phenotypes pre-adjusted for all fixed effects but contemporary group. Traits and number of genotypes are shown in Table 1, including growth, ultrasound scanning body composition, carcase, scrotal circumference and gestation length traits. The pedigree for each trait was traced back three generations from the animals with records. Genotype management and generation of GRM have been previously described by Connors *et al.* (2017). Relationship matrixes were built for genotyped animals from the pedigree (A) or from genomic markers (G). Genetic variances were estimated by fitting REML models in WOMBAT (Meyer, 2007) with A (Va), G (Vg) or both (Vt). Total genetic variance Vt was the sum of variance components from the model fitting both A and G. The ratio of Vg to Vt was then calculated for each trait. The model fitted contemporary group as the sole fixed effect, and the additive genetic breeding value as a random effect. For comparison, the optimal λ was calculated by assessing accuracies of EBV from 5-fold cross validations, as described previously (Zhang *et al.*, 2017).

Table 1. Summary of data for each trait, numbers of animals (N), mean (Mean), standard deviation (Std), and number of genotyped animals.

Trait	N	Mean	Std	Ng
Final weight: 600 day (FWD, kg)	671,867	536.4	111.6	8,950
Mature cow weight (MCW, kg)	98,658	650.8	92.8	1,757
Bull scan eye muscle area (BEA, cm^2)	362,016	83.3	11.6	7,440
Bull scan p8 fat depth (BP8, mm)	361,111	4.8	2.2	7,410
Bull scan rib fat (BRF, mm)	361,333	3.8	1.6	7,420
Heifer scan eye muscle area (HEA, cm2)	325,955	63.4	10.3	6,096
Heifer scan p8 fat depth (HP8, mm)	326,294	7.0	3.3	6,120
Heifer scan rib fat (HRF, mm)	325,884	5.4	2.4	6,126
Carcase weight (CWT, kg)	13,821	412.2	64.8	1,696
Carcase eye muscle area (CEA, cm2)	5,366	81.8	8.6	1,032
Carcase intramuscular fat (CIM, mm)	10,516	9.6	4.5	1,619
Carcase P8 depth (CP8, mm)	11,745	19.4	6.0	1,600
Gestation length (GLD, days)	371,388	280.3	4.8	7,408
Scrotal size (SS, cm)	295,968	36.0	2.9	6,663

Results and discussion

As shown in Table 2, most traits were moderately to highly heritable, with estimates ranging from 0.25 to 0.67. Heritability ranged from 0.32 to 0.61 for carcase traits, 0.32 to 0.61 for growth traits, 0.25 to 0.48 for bull scanning and 0.27 to 0.56 for heifer scanning traits, 0.45 for scrotal size, and 0.47 to 0.67 for gestation length. For most traits, genetic variance estimated using G only (Vg) accounted for 60% or more total genetic variance (Vt). Heritability from models fitting G only were in a similar range to those fitting A, but were slightly lower for most traits.

Table 2 Variance components from GRM, NRM and GRM+NRM models.

Trait	GRM			NRM			GRM+NRM		
	Vg	Vp	h^2	Va	Vp	h^2	Vt	Vp	Vg/Vt
FWD	548	1472	0.37	627	1491	0.42	688	1498	0.75
MCW	2103	3854	0.55	2155	3884	0.56	2303	3884	0.84
CWT	447	865	0.52	538	887	0.61	466	867	0.93
CEA	21.56	47.49	0.45	24.07	48.45	0.50	22.68	47.58	0.92
CIM	2.74	8.48	0.32	4.62	8.60	0.54	4.42	8.55	0.25
CP8	10.45	20.97	0.50	12.91	21.30	0.61	14.16	21.31	0.63
CRF	7.71	18.60	0.41	6.96	18.70	0.37	7.71	18.60	1.00
BEA	11.15	44.27	0.25	12.63	44.96	0.28	11.72	44.36	0.93
BP8	0.66	2.01	0.33	0.88	2.04	0.43	0.89	2.04	0.62
BRF	0.30	0.97	0.31	0.47	1.00	0.48	0.50	1.00	0.42
HEA	7.75	28.36	0.27	9.08	28.70	0.32	8.72	28.47	0.84
HP8	1.93	4.79	0.40	2.46	4.86	0.51	2.60	4.88	0.64
HRF	0.83	2.11	0.39	1.21	2.17	0.56	1.16	2.15	0.58
GLD	8.45	17.99	0.47	12.63	18.81	0.67	12.32	18.71	0.60
SS	2.30	5.09	0.45	2.28	5.18	0.44	2.41	5.11	0.95

Vg genetic variance from GRM model, Va from NRM model, total genetic variance (Vt) is calculated as the sum of Vg and Va from the GRM +NRM model. Vg/Vt is the proportion of Vg from GRM model to Vt.

In Table 3 the values of λ at the maximal accuracies ranged from 20 to 100, and 70 to 80 for most traits. As the maximum accuracy of EBV was approached, the response surface generally approached an asymptote, such that the range in λ encompassing $r_{max} \pm 0.01$ was large, for example, 40 to 90 for CEA or 70 to 100 for MCW. Therefore, accuracy was relatively insensitive over a large range in λ values. The maximal accuracies of EBV from SS-GBLUP were always higher than those from GBLUP.

The maximum SS-GBLUP cross-validation accuracies (r_{max}) of EBV ranged from 0.28 to 0.41 for carcase traits, 0.65 to 0.81 for growth traits, 0.60 to 0.81 for scanning; and 0.70 and 0.74 for SS and GLD. The ratio of genetic variances (Vg/Vt) were all within the range of λ values encompassing $r_{max} \pm 0.01$.

The ratio of Vg/Vt aligned reasonably well with the optimal ranges of λ values across traits. Averaged across all traits both Vg/Vt and λ at the maximum cross-validation accuracy were 0.73, suggesting a potential value of λ of around 0.7 for SS-GBLUP evaluations for this breed. This approach can be used as an alternative to determine the λ weighing factor for genomic information in SS-GBLUP.

Table 3 Proportion of genetic variances Vg from GRM to Vt from GRM and NRM, comparing with λ at the maximum accuracy r_{max} and accuracies of EBV from cross validation from SS-GBLUP and from GBLUP.

Trait	Vg/Vt		SS-GBLUP			GBLUP	
		λ	r_{min}	r_{max}	λ range	r_{min}	r_{max}
FWD	0.75	70	0.66	0.81	40-90	0.29	0.61
MCW	0.84	90	0.48	0.65	70-100	0.37	0.49
CWT	0.93	80	0.23	0.38	50-100	0.16	0.37
CEA	0.92	60	0.21	0.40	40-90	0.09	0.37
CIM	0.25	20	0.31	0.41	10-40	0.27	0.37
CP8	0.63	90	0.13	0.36	60-100	0.06	0.27
CRF	1.00	70	0.15	0.28	40-100	0.12	0.26
BEA	0.95	90	0.62	0.75	50-100	0.45	0.62
BP8	0.62	70	0.59	0.71	50-90	0.45	0.57
BRF	0.42	70	0.50	0.60	50-90	0.54	0.62
HEA	0.84	60	0.71	0.81	40-80	0.43	0.64
HP8	0.64	70	0.45	0.63	50-100	0.24	0.46
HRF	0.58	80	0.51	0.65	50-90	0.37	0.51
GLD	0.60	80	0.55	0.70	50-100	0.35	0.57
SS	0.95	100	0.53	0.74	60-100	0.40	0.65

r_{min} and r_{max} are minimal and maximal accurancies of EBV, λ range is the range in λ encompassing $r_{max} \pm 0.01$ derived from SS-GBLUP for all records, and r_{min} and r_{max} from GBLUP for genotyped animals.

List of References

Aguilar I., I. Misztal, D.L. Johnson, A. Legarra, S. Tsuruta & T.J. Lawlor, 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. of Dairy Sci.* **93**:743.

Christensen O.F. & M.S. Lund, 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42: 2.

Connors, N.K., J. Cook, C. Girard, B. Tier, K. Gore, D.J. Johnston and M.F. Fordosi, 2017. Development of the beef genomic pipeline for breedplan single step evaluation. *Proc. Assoc. Advmt. Anim. Breed. Genet.* **22**: 72.

Goddard M., B. Hayes & T. Meuwissen, 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. *J. of Anim. Breed. Genet.* **128**: 409.

Legarra A., I. Aguilar & I. Misztal, 2009. A relationship matrix including full pedigree and genomic information. *J. of Dairy Sci.* 92: 4656.

Legarra A., O.F.Christensen, I. Aguilar&I. Misztal, 2014. Single Step,a general approach forgenomic selection. *Livestock Science **166**:54-65.*

McMillan, A. & A. Swan, 2017. Weighting of genomic and pedigree relationships in single step evaluation of carcase traits in Australian sheep. *Proc. Assoc. Advmt. Anim. Breed. Genet.* **22** 130.

Meyer K., 2007. WOMBAT—A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *J. of Zhejiang Uni. Sci. B* 8: 815.

Vandenplas J., M. Spehar, K. Potocnik, N. Gengler & G. Gorjanc, 2017. National single step genomic method that integrates multi-national genomic information. *J. of Dairy Sci.* **100**:465-478.

Zhang, Y.D., A. Swan, D.J. Johnston & C.J. Girard, 2017. Weighting factors for genomic information used in single step genomic selection in Australian Beef. *Proc. Assoc. Advmt. Anim. Breed. Genet.* **22**: 70.