Proceedings of the World Congress on Genetics Applied to Livestock Production, 11.994

Incorporation of external information for international multi-breed beef cattle genetic evaluations using multi-trait single-step Bayesian regression model

M. Saatchi[1,2], L. Hyde[2], W. Shafer[2] D. Garrick[3,4] & B. Golden[4]

[1]Department of Animal Science, Iowa State University, Ames, IA 50011, USA
msaatchi@iastate.edu (Corresponding Author)
[2]American Simmental Association, Bozeman, MT 59715, USA
[3,4]Institute of Vet, Animal and Biomedical Science, Massey University, Palmerston, 4442, NZ
[4]Theta Solutions, LLC, Atascadero, CA 93422, USA

Summary

The single-step Bayesian regression (SS-BR) model benefits from a wider class of variable selection models, while this is a limitation with the single-step GBLUP model. We used an extended SS-BR model for multi-breed multi-trait analysis that includes extra polygenic effects on all genotyped and non-genotyped individuals as well as the external EPDs. This model was applied to the database of the International Genetic Solutions (IGS), an international collaboration between 12 beef breed associations with the largest multi-breed beef database in the world (over 16 million registered animals). The objective of this study was to investigate the impact of using two sources of external information: 1) external breed contrast adjustments reported by the US Meat Animal Research Center and 2) external EPDs derived from non-IGS partner breeds, on the predictions of multi-breed multi-trait SS-BR genomic-enhanced EPDs (GE-EPDs). In general, the correlations between SS-BR GE-EPDs with the conventional pedigree-based EPDs were higher for the SS-BR model without using any external information. However, the correlations were lower between GE-EPDs and the external EPDs in the model without using any external information. These results show that the markers can capture the breed differences in the multi-breed SS-BR and no external breed contrast adjustment is needed. However, the incorporation of the external EPDs are required to leverage the additional information that can be obtained from externally generated EPDs.

Keywords: single-step, multi-breed, multi-trait, beef cattle, external EPD, breed differences

Introduction:

Genomic profiles provide additional information about the genetic merit of a DNA tested animal and increase the accuracy of EPDs, which are called Genomic Enhanced EPD or GE-EPDs, of selection candidates. The single-step Bayesian regression model (SS-BR; Fernando et al., 2014) benefits from a wider class of variable selection models, while this is a limitation with the single-step GBLUP model. Saatchi et al. (2014) detected a few large-effect quantitative trait loci (QTL) that explain the majority of genetic variations for most growth and carcass traits in ten US beef cattle breeds. International Genetic Solutions (IGS), a progressive international collaboration between 12 breed associations across the world (www.internationalgeneticsolutions.com), decided to implement the multi-breed SS-BR model for their international genetic evaluation to take advantage of this knowledge.

The objective of this study was to investigate the impact of using two sources of external information on the predictions of GE-EPDs from multi-breed multi-trait SS-BR models. The two sources of external information were: 1) The breed contrast estimates reported by the US Meat Animal Research Center (USMARC), 2) The external EPDs for animals that have EPDs outside the IGS genetic evaluation system. The first is predicated on the hypothesis that the breed differences are more accurately estimated from data collected at research centers than the field collected data. The second, external EPDs, are incorporated to leverage the additional information that can be obtained from externally generated EPDs (Quaas & Zhang, 2006).

Materials and Methods

Animals and Data

A total of 16,784,173 animals from 12 collaborating beef breed associations with pedigree information were available in the IGS database at the time of study. The total number of individuals with own records on birth weight (BW), carcass weight (CW) ribeye area (REA) and ultrasonic ribeye area (UREA), were 9,673,144; 63,456; 56,559 and 438,913; respectively. The total number of genotyped animals were 78,854; which were genotyped with different SNP arrays and marker densities, mostly with the GGP-LD (GeneSeek; Lincoln, NE) and Illumina BovineSNP50 (Illumina; San Diego, CA). All genotypes were filtered and then imputed to 52,662 high quality markers according to the Saatchi, et al. (2014). The imputation was performed within each breed using FImpute software (Salgolzaei, et al. 2014). After imputation only 2250 highly influential markers (known as MSRP markers, Saatchi & Garrick, 2014) were used for analysis. Animals that were more than 3 generations removed from an animal with at least one phenotype were removed, leaving 11,598,426 animals for the analysis.

Statistical Model

The single-step Bayesian regression method (Fernando et al., 2014; Fernando et al., 2016) was extended for multi-trait analysis using BayesC with $\pi = 0$ (Cheng et al., 2017) and to include extra polygenic effects on all genotyped and non-genotyped individuals. The model equation was:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{M}\boldsymbol{\alpha} + \mathbf{U}\boldsymbol{\epsilon} + \mathbf{W}\boldsymbol{\delta} + \mathbf{e},$$

where $\mathbf{y}$ is the phenotypic vector for both genotyped and non-genotyped individuals for all 4 traits, $\mathbf{X}$ is the design matrix for contemporary group fix effects, $\boldsymbol{\beta}$ is the contemporary group fix effects for both genotyped and non-genotyped individuals for all 4 traits, $\mathbf{Z}$ is the design matrix, $\mathbf{M} = \begin{bmatrix} \widehat{\mathbf{M}}_{\mathrm{n}} \\ \mathbf{M}_{\mathrm{g}} \end{bmatrix}$, where $\mathbf{M}_{\mathrm{g}}$ is the matrix of marker covariate for genotyped animals and $\widehat{\mathbf{M}_{\mathbf{n}}} = \mathbf{A}_{ng}\mathbf{A}_{gg}^{-1}\mathbf{M}_{\mathrm{n}}$, representing imputed marker covariates for non-genotyped animals derived from genotyped relatives for all 4 traits, $\boldsymbol{\alpha}$ is the marker effect**,** $\boldsymbol{\epsilon}$ is the imputation residual, and $\mathbf{U}$ is the design matrix allocating records to breeding values of non-genotyped animals for all 4 traits. The $\mathbf{W}$ is the design matrix and $\boldsymbol{\delta}$ is the extra polygenic effects for all 4 traits.

The priors for $\boldsymbol{\beta}$ are flat priors. The priors for marker effects $\boldsymbol{\alpha}$ is $\alpha|\sigma_{g\alpha}^2 \sim N\left(0, \left(c * \left(\sum_{m=1}^{2250} 2p_m q_m\right)^{-1}\right)\sigma_g^2\right)$ where σ_g^2 is the additive genetic variance, the prior for $\boldsymbol{\epsilon}$ is $\boldsymbol{\epsilon}|\sigma_{g\epsilon}^2 \sim N(0, c * (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})\sigma_g^2)$, for $\boldsymbol{\delta}$ is $\delta|\sigma_{g\delta}^2 \sim N(0, cA\sigma_g^2)$, $c = 0.5$ and $(\sigma_g^2|\nu_g, S_g^2) \sim \nu_g S_g^2 \chi_{\nu_g}^2$. The prior for e_i is $e_i|\sigma_e^2 \sim_{\square id} N(0, \sigma_e^2)$ and $(\sigma_e^2|\nu_e, S_e^2) \sim \nu_e S_e^2 \chi_{\nu_e}^2$.

In the first analysis, the raw phenotypes were used without using any external information. While in the second analysis, an animal's phenotypes for BW, CW and REA traits were pre-adjusted based on the USMARC breed contrasts and the animal's breed composition (estimated from the pedigree). The external EPDs for animals with available EPDs from breed associations external to IGS were used in the second analysis using a modified method of Quaas and Zhang (2006), which does not require the inversion of the relationship matrix among the individuals with the external EPDs. This modification allowed us to include more animals with the external EPDs in the analysis (21,874 animals in total).

The analyses were performed using the BOLT software (Golden et al., 2016). The results were compared to the EPDs reported by the American Simmental Association (www.simmental.org/site/index.php/multi-breed-evaluation) from the conventional pedigree-based multi-breed national genetic evaluation in spring 2017, which include external EPDs.

Results and Discussions

The Pearson correlations between the GE-EPDs calculated from SS-BR models with and without external information (breed contrasts adjustment and external EPDs) and the conventional pedigree-based EPDs are shown in Table 1. The regression coefficient of SS-BR GE-EPDs on conventional EPDs are also shown in the Table 1. In general, the correlations are higher for the SS-BR model without breed adjustment and external EPDs in comparison to the SS-BR model without using external information (except for rib eye area, where the correlations were lower). Similar results were also observed for the genetic trends obtained from the SS-BR model without using any external information, which were more similar to the genetic trends reported from the IGS conventional genetic evaluation (results are not shown). However, the regression coefficients are much lower in the SS-BR model without using external information, reflecting a higher shrinkage in GE-EPDs in this model. These results show that in the multi-breed SS-BR model, the markers can capture the breed differences and no external breed contrast adjustment is needed.

Table 1 – The overall correlations and regression coefficients between SS-BR GE-EPDs and the conventional pedigree-based EPDs.

		With external information		Without external information	
Trait		r	b	r	b
Birth weight	All	0.95	1.10	0.97	0.97
	Genotyped	0.95	1.00	0.95	0.92
	Non-genotyped	0.95	1.10	0.97	0.97
Carcass weight	All	0.64	0.62	0.70	0.38
	Genotyped	0.48	0.56	0.48	0.34
	Non-genotyped	0.64	0.62	0.70	0.38
Ribeye area	All	0.62	0.92	0.56	0.40
	Genotyped	0.65	0.86	0.53	0.45
	Non-genotyped	0.62	0.92	0.56	0.40

As expected, the correlations were lower for genotyped animals versus the non-genotyped animals because deviations exist in GE-EPDs compare the expected pedigree-based EPDs by adding the extra DNA information and accounting for the Mendelian sampling in the genotyped animals.

Table 2 - The correlations and regression coefficients between SS-BR GE-EPDs and external EPDs by different levels of GE-EPDs accuracies.

		With external information			Without external information		
Trait	Accuracy	N	r	b	N	r	b
Birth weight	.40-.49	1889	0.84	0.94	1506	0.53	0.50
	.90-.99	38	1.00	0.97	38	0.92	0.79
Carcass weight	.30-.39	590	0.94	1.01	656	0.37	0.19
	.60-.69	24	0.95	0.90	31	0.39	0.32
Ribeye area	.30-.39	365	0.92	1.02	456	0.31	0.18
	.60-.69	41	0.98	0.96	15	0.51	0.27

The Pearson correlations and coefficient regressions between the SS-BR GE-EPDs and external EPDs are shown in Table 2. It shows that the correlations and regressions are much lower for the SS-BR model without using the external information. These results demonstrate that the external EPDs could be used in the SS-BR to improve the GE-EPD predictions using the independent information available on such animals.

References

Cheng, H., K. Kizilkaya, J. Zheng, D.J. Garrick & R. L. Fernando, 2017. Multple-trait Bayesian Regression Methods with Mixture Priors for Genomic Prediction. bioRxiv, doi: http://dx.doi.org/10.1101/102962.

Fernando, R.L., H. Cheng, B.L. Golden & D.J. Garrick, 2016. Computational strategies for alternative single-step Bayesian regression models with large numbers of genotyped and non-genotyped animals. Genet. Sel. Evol. 48:96.

Fernando, R.L., J.C.M. Dekkers & D.J. Garrick, 2014. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. Genet. Sel. Evol. 46:50.

Golden, B.L., R.L. Fernando & D.J. Garrick, 2016. Bolt and an alternative approach to genomic EPDs. Proceedings of the Beef Improvement Federation, Manhattan, KS, USA, 102–106.

Quaas, R. L. & Z. Zhang, 2006. Multiple-breed genetic evaluation in the us beef cattle context: methodology. In proceeding of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brasil.

Saatchi M. and D. J. Garrick, 2014. Developing a reduced SNP panel for low-cost genotyping in beef cattle. ISU Animal Industry Report, Leaflet-R2854.

Saatchi M., R.D. Schnabel, J.F. Taylor & D.J. Garrick, 2014. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. BMC Genomics, 15:442.

Sargolzaei, M., J.P. Chesnais & F.S. Schenkel, 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics, 15:478.