

Options for evaluating multiple breeds in a single-step GBLUP for US dairy population

A. Cesarani^{1*}, D. Lourenco¹, P.M. VanRaden², E.L. Nicolazzi³, A. Legarra⁴, S. Tsuruta¹,
and I. Misztal¹

¹Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602; ²Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD, USA; ³Council on Dairy Cattle Breeding, Bowie, MD 20716; ⁴INRA, UMR1388 GenPhySE, Castanet-Tolosan, France; *acesarani@uniss.it

Abstract

This work aimed to develop single-step multibreed genomic predictions for US dairy population. About 45M yield phenotypes recorded between 2000 and 2020 in purebred Ayrshire, Brown Swiss, Guernsey, Holstein, and Jersey animals were analyzed. Genotypes at 79,294 SNPs were available for 3.9M animals. A 3-traits model with milk, fat, and protein yields was applied to four different scenarios: each breed separately; all five breeds together; Aushire, Brown Swiss, and Guernsey together; Holstein and Jersey together. For each scenario, BLUP and single-step genomic BLUP were carried out and were validated using the correlation between adjusted phenotypes (for cows) or daughter yield deviations (for bulls) and (genomic) breeding values. As expected, all ssGBLUP scenarios led to higher values compared to BLUP. The ssGBLUP multibreed evaluation for US dairy population is feasible, but model fine-tuning is needed to achieve good validation results.

Introduction

In 2007, a multibreed genetic evaluation was implemented for US dairy cattle, and it was extended to a multistep genomic evaluation in 2009 (VanRaden et al., 2007; 2009). Currently, US dairy cattle are evaluated with a two-step method: a multibreed BLUP model followed by SNP effects estimation and direct genomic values computation within each breed. Single-step genomic BLUP (ssGBLUP) has been adopted in lieu of the multistep method for genetic evaluations in several farm animal species over the years. ssGBLUP has been applied with satisfactory results in almost all livestock species (e.g., Macedo et al., 2020; Pimentel et al., 2021). Genomic models have been tested also in a multibreed context (e.g., Winkelman et al., 2015; Khansefid et al., 2020). The aim of this work was to develop ssGBLUP multibreed genomic predictions for US dairy cattle. In particular, we used the algorithm for proven and young (APY) to compute the inverse of the genomic relationship matrix.

Materials & Methods

Data. Official data used in the multibreed genomic evaluations for US dairy cattle breeds were provided by the Council on Dairy Cattle Breeding (CDCB, Bowie, MD). Only purebred Ayrshire (AY), Brown Swiss (BS), Guernsey (GU), Holstein (HO), and Jersey (JE) animals were considered. Phenotypes for milk (MY), fat (FY), and protein (PY) yields recorded from 2000 to 2020 were considered. The 305-d yields (about 45M recorded on 19.4M cows) for the first 5 lactations included projected records for the final lactation and for lactations not yet completed by June 2020. About 3.9M animals were genotyped at various densities and imputed to 79,294 selected SNPs within each breed. Only purebred animals were included in the evaluation.

Analysis. Two evaluation methods were considered: i) traditional BLUP, with unknown parent groups (UPG); ii) ssGBLUP, with QP transformation of unknown parent groups applied to **A** and **A₂₂**, implemented as in Tsuruta et al. (2019) and Cesarani et al. (2021). UPG were defined based on breed, sex, and year of birth for a total of 8 UPG for each breed. The two methods were carried out for: i) each breed separately (SINGLE); ii) the 5 breeds together (5_BREEDS); iii) AY, BS, and GU (AY_BS_GU); iv) HO and JE together (HO_JE). The APY algorithm was applied with 15k core animals in SINGLE (for HO and JE), 30k for AY_BS_GU and HO_JE, and 45k core animals for 5_BREEDS. A 3-trait repeatability animal model was applied with herd-management, age-parity, inbreeding coefficient and heterosis covariates, and UPG as fixed effects and herd-sire combination, animal, and permanent environment as random effects. For the multibreed models, all effects were made breed-specific by adding the breed code to each level of the effects.

Validation. Each scenario was run two times: FULL, with all phenotypes; REDUCED, removing the phenotypes of cows born in the last 4 years (2015-2018). Validation of each model was assessed by the correlation between adjusted phenotypes in FULL and (genomic) estimated breeding values ((G)EBV) in REDUCED for cows (predictive ability), and by the correlation of daughter yield deviation (DYD) in FULL on the (G)EBV in REDUCED for bulls. Validation cows were genotyped females born in 2015-2018 with no records in REDUCED, whereas validation bulls had at least 10 (AY, BS, GU), or 50 (HO, JE) daughters in FULL and no daughters in REDUCED.

Results

Correlations between raw EBV from BLUP estimated in all the different models were 1.00 for all breeds and traits, whereas correlations between raw GEBV estimated in ssGBLUP SINGLE and ssGBLUP 5_BREEDS ranged from 0.95 (MY for AY and BS) to 1.00 (HO and JE). Moreover, we also computed the slope of the regression of (G)EBV in 5-BREEDS on (G)EBV in SINGLE for each breed: slopes for BLUP were always 1, confirming that in practice each breed was independently evaluated, as breeds were not connected by the pedigree. Acceptable slope values were also found for ssGBLUP as values ranged from 0.93 (MY and PY for BS) to 1.02 (MY for AY). Table 2 shows the validation results for both cows and bulls. As expected, ssGBLUP SINGLE gave better results compared to BLUP for all considered breeds and traits. Within ssGBLUP, higher values were observed for 5_BREEDS compared to SINGLE for all breeds, expect for BS bulls, for which a slight decrease was observed in the multibreed scenario. Results in AY_BS_GU and HO_JE were similar to SINGLE. However, results for AY, and GU are strongly affected by the very low number of bulls.

Table 2. Correlations between adjusted phenotypes for cows or daughter yield deviations for bulls and EBV from BLUP or GEBV from ssGBLUP in the five considered breeds.

Breed	Trait	N animals		BLUP		ssGBLUP							
		cows	bulls	cows	bulls	SINGLE		5_BREEDS		AY_BS_GU		HO_JE	
						cows	bulls	cows	bulls	cows	bulls	cows	bulls
AY	milk	181	17	0.45	0.45	0.47	0.52	0.55	0.85	0.47	0.58	–	–
	fat			0.51	0.45	0.54	0.56	0.52	0.86	0.53	0.58	–	–
	prot			0.52	0.55	0.54	0.66	0.55	0.93	0.54	0.69	–	–
BS	milk	2,423	107	0.23	0.39	0.39	0.63	0.40	0.61	0.39	0.63	–	–
	fat			0.27	0.37	0.42	0.55	0.41	0.55	0.42	0.58	–	–

	prot			0.27	0.46	0.44	0.67	0.43	0.65	0.44	0.67	–	–
	milk			0.24	0.42	0.35	0.54	0.44	0.64	0.35	0.51	–	–
GU	fat	750	28	0.27	0.45	0.33	0.56	0.39	0.63	0.34	0.53	–	–
	prot			0.24	0.42	0.34	0.50	0.39	0.53	0.34	0.48	–	–
	milk			0.31	0.53	0.55	0.88	0.55	0.88	–	–	0.55	0.88
HO	fat	577,340	3,278	0.34	0.56	0.55	0.88	0.55	0.87	–	–	0.55	0.88
	prot			0.33	0.55	0.52	0.86	0.52	0.86	–	–	0.52	0.86
	milk			0.32	0.66	0.5	0.84	0.51	0.87	–	–	0.52	0.87
JE	fat	90,666	471	0.29	0.60	0.46	0.80	0.47	0.82	–	–	0.47	0.82
	prot			0.35	0.67	0.51	0.83	0.52	0.86	–	–	0.52	0.86
Mean				0.33	0.50	0.46	0.69	0.48	0.76	0.42	0.58	0.52	0.86
SD				0.09	0.09	0.08	0.15	0.06	0.14	0.08	0.07	0.03	0.02

Table 3 reports the slope of DYD or adjusted phenotypes on (G)EBV for bulls and cows, respectively. Results were similar to those in Table 2, where ssGBLUP SINGLE led to better results, on average (i.e., values closer to 1) compared to BLUP. All regression coefficients estimated for HO and JE were within 15% of the unit, considered acceptable values (Tsuruta et al., 2011). In this case, results for AY and GU were also affected by the limited number of validation animals.

Table 3. Regression coefficients (b1) of bull daughter yield deviations (DYD) on EBV from BLUP and on GEBV from ssGBLUP for validation bulls and regression coefficients of cow adjusted phenotypes on EBV from BLUP and on GEBV from ssGBLUP for validation cows.

Breed	Trait	N animals		ssGBLUP									
				BLUP		SINGLE		5_BREEDS		AY_BS_GU		HO_JE	
		cows	bulls	cows	bulls	cows	bulls	cows	bulls	cows	bulls	cows	bulls
	milk			1.03	0.94	0.96	0.88	0.88	1.37	0.98	0.96	–	–
AY	fat	181	17	0.95	0.79	0.89	0.86	0.76	1.25	0.90	0.92	–	–
	prot			0.98	1.13	0.92	1.04	0.80	1.34	0.93	1.06	–	–
	milk			0.72	0.64	0.96	0.94	0.90	0.81	0.97	0.93	–	–
BS	fat	2,423	107	0.72	0.55	0.91	0.80	0.84	0.72	0.92	0.80	–	–
	prot			0.73	0.70	0.96	0.93	0.89	0.82	0.96	0.92	–	–
	milk			0.75	0.75	0.91	1.04	1.07	1.14	0.91	1.03	–	–
GU	fat	750	28	0.71	0.67	0.74	0.84	0.93	0.89	0.77	0.85	–	–
	prot			0.71	0.71	0.84	0.9	1.05	1.02	0.85	0.92	–	–
	milk			1.01	0.86	1.10	1.03	1.11	1.03	–	–	1.10	1.04
HO	fat	577,340	3,278	1.03	0.91	1.03	0.99	1.03	0.97	–	–	1.02	0.99
	prot			1.02	0.87	1.08	1.03	1.08	1.03	–	–	1.07	1.02
	milk			1.02	1.06	1.03	1.01	1.07	1.05	–	–	1.09	1.08
JE	fat	90,666	471	0.89	0.89	0.88	0.86	0.93	0.92	–	–	0.97	0.97
	prot			1.01	1.06	0.99	0.99	1.05	1.06	–	–	1.07	1.08
Mean				0.89	0.84	0.95	0.94	0.96	1.03	0.91	0.93	1.05	1.03

SD		0.14	0.17	0.09	0.08	0.11	0.19	0.07	0.08	0.05	0.05
----	--	------	------	------	------	------	------	------	------	------	------

Regarding computing time, about 7 hours were needed to solve the MME with ssGBLUP in HO SINGLE (3.4M genotypes); this time increased by a factor of 9 for the 5_BREEDS scenario.

Discussion

The use of genomic information is expected to increase the reliability of breeding value prediction in livestock. ssGBLUP gives higher validation accuracies in several livestock, but applications of this method in large-scale dairy evaluations have been limited. Recently, Cesarani et al. (2021) showed the superior performance of ssGBLUP compared to BLUP in US Holsteins. The present study gave insights into using this genomic model in a multibreed context. As expected, using genomic information within each breed improves validation statistics for both bulls and cows. Results for Holsteins were similar in all scenarios involving this breed because of the greatest number of genotypes and records. This study demonstrates that large-scale multibreed genomic evaluations are feasible, as we obtained results in the multibreed scenarios that were at least similar to the ones in single-breed evaluations. Fine-tuning of models and a better representation of each considered breed in the core animals are needed to avoid reducing the reliability of smaller breeds when evaluated together with numerically dominant breeds. This is the first step towards the multi breed evaluation, in which crossbreed and external information will be considered.

References

- Cesarani A., Masuda Y., Tsuruta S., Nicolazzi E.L., VanRaden P.M., et al. (2021) *J. Dairy Sci.* 104:5843-5853. <https://doi.org/10.3168/jds.2020-19789>
- Khansefid M., Goddard M.E., Haile-Mariam M., Konstantinov K.V., Schrooten C., et al. (2020) *Front. Genet.* 11:598580. <https://doi.org/10.3389/fgene.2020.598580>.
- Macedo F.L., Christensen O.F., Astruc J.M., Aguilar I., Masuda Y., et al. (2020) *Genet. Sel. Evol.* 52:1-10. <https://doi.org/10.1186/s12711-020-00567-1>
- Pimentel E.D.C.G., Edel C., Krogmeier D., Emmerling R., and Götz K.U. (2021) *Interbull Bulletin*, (56), 121-124.
- Tsuruta S., Lourenco D.A.L., Masuda Y., Misztal I., and Lawlor T.J. (2019) *J. Dairy Sci.* 102:9956-9970. <https://doi.org/10.3168/jds.2019-16789>
- Tsuruta S., Misztal I., Aguilar I., and Lawlor T.J. (2011) *J. Dairy Sci.* 94:4198-4204. <https://doi.org/10.3168/jds.2011-4256>.
- VanRaden P.M., Tooker M.E., Cole J.B., Wiggans G.R., and Megonigal Jr J.H. (2007) *J. Dairy Sci.* 90:2434-2441. <https://doi.org/10.3168/jds.2006-704>.
- VanRaden P.M., Van Tassell C.P., Wiggans G.R., Sonstegard T.S., Schnabel R.D., et al. (2009) *J. Dairy Sci.* 92:16-24. <https://doi.org/10.3168/jds.2008-1514>
- Winkelman A.M., Johnson D.L., and Harris B.L. (2015) *J. Dairy Sci.* 98:659-675. <https://doi.org/10.3168/jds.2014-8560>.