Gordo, D., Sanglard, L., Bernal Rubio, Y., Passafaro, T., Gonzalez-Peña, D., Ampofo, I., Fragomeni, B., Andersen, K., Di Croce, F. (2026) Evaluating the use of metafounders in an industry multibreed genetic evaluation in beef cattle, *Proceedings 13th World Congress on Genetics Applied to Livestock Production* 23989: https://doi.org/10.31274/wcgalp.23989 ©2026 Gordo, et al. CC BY-NC-ND

Evaluating the use of metafounders in an industry multibreed genetic evaluation in beef cattle

D.G.M. Gordo[1]*, L.P. Sanglard[1], Y.L. Bernal Rubio[1], T. Passafaro[1], D. Gonzalez-Peña[1], I. Ampofo[2], B. Fragomeni[2], K.J. Andersen[1], and F. Di Croce[1]

*[1]Zoetis, Kalamazoo, USA; [2]University of Connecticut, Storrs, USA; *Corresponding author: danielgustavo.mansangordo@zoetis.com*

Introduction

Single-step genomic BLUP (ssGBLUP) is widely used for genetic evaluation in livestock due to its ability to combine genotyped and non-genotyped animals in one analysis, yielding more accurate predictions than traditional BLUP (Aguilar *et al*., 2010). However, the hybrid relationship matrix (H) used in ssGBLUP relies on assumptions that may not hold, especially when pedigree data are missing for genotyped animals, causing incompatibilities between genomic (G) and pedigree (A) relationship matrices and resulting in biased genomic estimated breeding values (Garcia-Baccino *et al*., 2017). These incompatibilities arise partly because A assumes founders as the base population, while G is centered on current allele frequencies, an issue intensified in multibreed populations with heterogeneous base populations (Garcia-Baccino *et al*., 2017). Further, pedigree information often traces back to multiple unrelated base populations, while genomic data may reveal relationships among these founders, intensifying the imbalance and bias. Unknown parent groups have partly addressed this by modeling genetic merit differences across missing parent classes, but still assume unrelated base populations. To better model relatedness in base populations, Legarra *et al*. (2015) introduced the use of metafounders (MF), pseudo-individuals representing the base populations with relationships among them, improving G and A compatibility. Metafounders generalize unknown parent groups by allowing relationships within and between base populations, potentially reducing bias without loss of accuracy (Kluska *et al*., 2021). While previous research has highlighted the potential benefits of metafounders, their implementation in practical, multibreed industry-based populations is limited. Therefore, we assess the effect of incorporating metafounders into base population modeling on the predictive performance and potential bias of Expected Progeny Differences (EPD) for Birth Weight (BW), Fertility (FERT), and Marbling (MARB) in an industry multibreed beef cattle population. Specifically, we evaluate whether metafounders can improve the ranking ability of EPD and correct potential upward inflation of EPD arising from incomplete pedigree information.

Materials & Methods

Genotype data from 286,294 animals and 42,915 SNPs (after quality control) were used, together with pedigree records spanning up to five generations. Within the genotyped population, 174,640 animals had both parents missing, 12,208 were missing only the sire, and 76,664 were missing only the dam. Genomic breed composition estimates — expressing each animal's proportional ancestry across breeds — were available for all 286,294 genotyped animals. These estimates spanned 14 breed components: Angus (AN), Red Angus (AR), Hereford (HH), Gelbvieh (GV), South Devon (DS), Charolais (CH), Simmental (SM), Limousin (LM), Brown Swiss (BS), Jersey (JE), Holstein (HO), Brahman (BR), Nellore (NE), and Wagyu (WG). Three scenarios were evaluated: (I) no metafounders (No-MF) included in the pedigree; (II) MF11, with 11

metafounders defined by individual breeds or breed groups (dairy, indicus, or Others); and (III) MF5, with metafounders grouped by breed origin (British, Continental, Dairy, Indicus, and Others). Metafounders for missing parents were assigned based on the genomic breed composition of the progeny: if one parent was unknown and the progeny had more than 80% of a particular breed, the major breed metafounder group was assigned to the unknown parent. If both parents were unknown and the progeny's breed composition was split between 20–80% for two breeds, the metafounder for one breed was assigned to the sire and the other to the dam. If only one parent was unknown and the breed composition was split between 20–80% for two breeds, the complementary breed metafounder was assigned to the unknown parent. If the breed composition was undetermined—split across all breeds with less than 20% for any breed—the 'Others' metafounder's group was assigned to the unknown parent. Validation assessed predictive performance for three traits: birth weight (BW, n = 1,294,693), fertility (FERT, n = 1,372,651), and carcass marbling (MARB, n = 45,866). Genomic EBV were calculated using a single-step GBLUP animal model with the H matrix and the APY algorithm, employing a 30K random core to accommodate the large genotype dataset. Analyses were performed using the BLUPF90 family of programs (Misztal *et al*., 2002). The validation population included animals born after 2020 (FERT, n = 47,510), 2022 (MARB, n = 6,134), and 2023 (BW, n = 22,216). Adjusted phenotypes were obtained for all animals incorporating all available information (phenotypes, pedigree, and genotypes), while predictions excluded phenotypes for validation animals. Estimated breeding values were divided by two to represent EPD, ranked, and divided into quartiles. The mean EPD, adjusted phenotype (y*), and observed phenotype (y) per quartile were computed to evaluate: (1) ranking ability, via consistency of phenotypic trends across EPD quartiles; and (2) EPD bias, via comparison of average EPD levels across scenarios — as incomplete pedigree is known to cause bias in ssGBLUP-derived GEBV (Masuda *et al*., 2022; Garcia-Baccino *et al*., 2017).

Table 1. Number of records, fixed and random effects, and phenotypic means of the selected traits in the genetic evaluation.

Trait[1]	N. records	Fixed effects[2]	Random effects[3]	Mean (Std)
BW	1,294,693	CG, ET, sex, age of dam (linear and quadratic)	σ^2_a, σ^2_m, σ^2_c, σ^2_e	35.1 (5.2)
MARB	45,866	RH, age at harvest, CG, sex	σ^2_a, σ^2_e	5.8 (1.3)
FERT	1,372,651 (215,236)[4]	CG, AFC, RH, YS, age, record	σ^2_a, σ^2_{pe}, σ^2_e	0.3 (0.5)

[1] BW, birth weight (kg); MARB, marbling (BIF score); FERT, fertility (no. calves). Correlated traits in multi-trait models: BW with post-weaning gain (PWG) and birth-weaning gain (WG); MARB with ultrasound intramuscular fat (UIMF).[2] CG, contemporary group; ET, embryo transfer status; RH, retained heterosis; AFC, age at first calving; YS, year/season. [3] σ^2a, additive genetic; σ^2m, maternal genetic; σ^2c, maternal permanent environment; σ^2pe, permanent environment; σ^2e, residual variance. [4] Number of records (number of animals).

Results & Discussion

A key challenge in applying MF in a multibreed industry genetic evaluation context is identifying the breeds of unknown parents, as distinguishing specific breed contributions from the dam versus the sire is difficult, particularly in animals with a two-breed background (Garcia-Baccino *et al*., 2017; Legarra *et al*., 2015). Despite this challenge, MF relationships were successfully estimated and varied from 0.40 to 0.49 for MF11 and from 0.40 to 0.47 for MF5, with the lowest values observed between Indicus and dairy breeds. These findings align with Junqueira *et al*. (2020), who reported relationships ranging from 0.34 (Zebu–Hereford) to 0.46 (Hereford–Bradford), and are higher than those reported by Kluska *et al*. (2021) for composite beef cattle populations (0.15–

0.38), reflecting the expected relatedness among breeds of similar vs. distant origins (e.g., Indicus vs. Taurine). The correlations between EPD obtained from Scenario I, without MF, and those from Scenarios II and III, including MF, ranged from 0.93 (BW) to 0.96 (MARB). The high correlation between scenarios indicates good agreement in animal ranking across all three approaches. However, the absolute EPD values differed notably between scenarios: in Scenario I (No-MF), EPD means were systematically higher than in Scenario II (MF11); for example, the top BW quartile averaged 5.2 kg vs. 1.0 kg (Table 2).

Table 2: Mean expected progeny differences (EPD, mean adjusted phenotypes (y*), and observed phenotype (y) by EPD ranked quartiles for birth weight.

Scenarios	Quantiles	All animals			Missing both parents			Missing one parent		
		EPD	y*	y	EPD	y*	y	EPD	y*	y
I	Top 25%	5.2	8.6	88.8	5.2	8.5	88.7	5.3	8.5	88.7
	Mid 25 – 50%	0.2	1.5	79.9	0.2	1.4	79.8	0.3	1.4	79.8
	Mid 50 – 75%	-2.8	-2.9	74.7	-2.9	-2.9	74.6	-2.8	-2.9	74.6
	Bottom 75%	-7.6	-9.4	67.1	-7.7	-9.4	67.1	-7.6	-9.4	67.1
II	Top 25%	1.0	3.8	88.5	1.1	3.8	88.4	1.1	3.8	88.4
	Mid 25 – 50%	-4.3	-3.5	80.0	-4.3	-3.6	79.9	-4.2	-3.5	79.9
	Mid 50 – 75%	-7.4	-8.1	74.8	-7.5	-8.2	74.8	-7.4	-8.2	74.8
	Bottom 75%	-12.2	-15.0	67.2	-12.3	-15.1	67.1	-12.3	-15.1	67.1

This shift is consistent with an upward inflation of EPD when pedigree is incomplete: when parentage is unknown, the base population is implicitly anchored at zero genetic merit, inflating EPD relative to a properly modeled baseline. The inclusion of metafounders, by explicitly modeling relationships among base population founders, provides a more appropriate reference point and reduces this inflation. Despite this difference in EPD scale, ranking ability — measured by the differentiation between top and bottom quartiles — was largely preserved for BW and FERT across scenarios.

Table 3: Mean expected progeny differences (EPD, mean adjusted phenotypes (y*), and observed phenotype (y) by EPD ranked quartiles for fertility.

Scenarios	Quantiles	All animals			Missing both parents			Missing one parent		
		EPD	y*	y	EPD	y*	y	EPD	y*	y
I	Top 25%	0.35	0.40	1.80	0.37	0.43	1.92	0.33	0.40	1.90
	Mid 25 – 50%	0.25	0.28	1.57	0.26	0.27	1.73	0.24	0.27	1.81
	Mid 50 – 75%	0.19	0.19	1.39	0.20	0.19	1.52	0.18	0.18	1.72
	Bottom 75%	0.11	0.11	1.25	0.13	0.12	1.67	0.10	0.11	1.74
II	Top 25%	0.37	0.37	1.61	0.31	0.29	1.64	0.31	0.38	1.86
	Mid 25 – 50%	0.27	0.29	1.53	0.21	0.18	1.57	0.21	0.24	1.81
	Mid 50 – 75%	0.20	0.23	1.48	0.13	0.24	1.83	0.14	0.19	1.80
	Bottom 75%	0.12	0.16	1.40	0.04	0.10	1.80	0.06	0.09	1.70

For MARB (Table 4), however, Scenario I showed greater differentiation (2.8 vs. 1.9 score points), suggesting that EPD scale correction may come at a modest cost to discriminative power for some traits. Similarly, Kluska *et al*. (2021) reported that including metafounders in the Montana

composite beef cattle evaluation did not improve the accuracy of genomic predictions. Regarding the number of metafounders, results for Scenario II (MF11) were similar or slightly better than those for Scenario III (MF5); therefore, only results for Scenario II are presented here. This is consistent with Junqueira *et al*. (2020), who also observed that finer breed group resolution (11 vs. 5 metafounders) often yielded better or comparable accuracy. Unlike our study, Junqueira *et al*. (2020) reported accuracy improvements of 20% and 2% for Hereford and Braford, respectively, which were more pronounced for traits with moderate to high heritability.

Table 4: Mean expected progeny differences (EPD, mean adjusted phenotypes (y*), and observed phenotype (y) by EPD ranked quartiles for marbling.

Scenarios	Quantiles	All animals			Missing both parents			Missing one parent		
		EPD	y*	y	EPD	y*	y	EPD	y*	y
I	Top 25%	0.9	0.9	6.4	0.7	0.7	6.3	1.0	1.0	7.0
	Mid 25 – 50%	0.1	0.2	5.4	0.0	0.0	5.2	0.3	0.3	5.9
	Mid 50 – 75%	-0.3	-0.2	4.7	-0.4	-0.4	4.7	-0.2	-0.2	5.1
	Bottom 75%	-0.9	-0.8	4.2	-1.0	-0.9	4.1	-0.9	-0.8	4.2
II	Top 25%	0.7	0.5	6.0	0.3	0.2	5.7	0.7	0.6	6.5
	Mid 25 – 50%	-0.1	-0.1	5.3	-0.1	-0.3	5.2	0.2	0.0	5.9
	Mid 50 – 75%	-0.3	-0.4	4.9	-0.4	-0.5	4.9	-0.2	-0.4	5.2
	Bottom 75%	-0.9	-1.0	4.4	-0.9	-1.0	4.4	-0.8	-0.9	4.6

Conclusion

This study assessed the effect of metafounders on EPD predictive performance and bias for economically relevant traits in a multibreed industry genetic evaluation. Metafounders corrected the upward EPD inflation observed in the No-MF scenario while largely preserving ranking ability across traits. Overall, while the metafounders approach offers a biologically meaningful framework, its practical benefits appear limited in this large-scale commercial context, possibly due to the difficulty of accurately assigning breed of origin to unknown parents in a largely crossbred population, which likely reduces the precision of metafounder assignment and their ability to improve predictions beyond what genomic information already provides. Further refinement in breed-of-origin assignment may be necessary to fully realize its potential.

References

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, *et al*. (2010) J. Dairy Sci. 93(2):743–752. doi.org/10.3168/jds.2009-2730

Garcia-Baccino CA, Legarra A, Christensen OF, Misztal I, Pocrnic I, *et al*. (2017) Genet. Sel. Evol. 49(1):34. doi.org/10.1186/s12711-017-0309-2

Junqueira VS, Lopes PS, Lourenco D, Silva FFE, Cardoso FF. (2020) Front. Genet. 11:556399. doi.org/10.3389/fgene.2020.556399

Kluska S, Masuda Y, Ferraz JBS, Tsuruta S, Eler JP, *et al*. (2021) Front. Genet. 12:678587. doi.org/10.3389/fgene.2021.678587

Legarra A, Christensen OF, Vitezica ZG, Aguilar I, Misztal I. (2015) Genetics 200(2):455–468. doi.org/10.1534/genetics.115.177014

Masuda Y, VanRaden PM, Tsuruta S, Lourenco DAL, Misztal I. (2022) J. Dairy Sci. 105(2):923–939. doi.org/10.3168/jds.2021-20293

Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T. (2002) Proc. of 7th WCGALP, Montpellier, France.