Walugembe, M., Diaz, I., Bhatia, V., Cross, A., Hidalgo, J., Lourenco, D., Rathje, T., Borg, R. (2026) Strategies for genomic prediction of a sow retention trait in a multi-breed pig population using single-step GBLUP, *Proceedings 13th World Congress on Genetics Applied to Livestock Production*: 24180 https://doi.org/10.31274/wcgalp.24180 ©2026 *Walugembe, et al.* CC BY-NC-ND

Strategies for genomic prediction of a sow retention trait in a multi-breed pig population using single-step GBLUP

M. Walugembe[1*], I.S. Diaz[1], V. Bhatia[1], A.J. Cross[1], J. Hidalgo[2], D. Lourenco[2], T.A. Rathje[1], and R.C. Borg[1]

[1]DNA Genetics, Columbus, NE, USA; [2]University of Georgia, Athens, GA, USA; [*]Corresponding author: MWalugembe@dnaswinegenetics.com

Introduction

Commercial sow longevity is important in pig production due to its influence on the number of produced and weaned piglets in the sow's reproductive life. Sows may be removed voluntarily to meet the herd's replacement rates; however, this is usually not the case, as they are involuntarily culled due to other reasons before reaching a desired parity. Sow longevity traits are often measured and selected in purebred nucleus farms; however, these sows are unable to fully express traits that are measured late in life. Hence, there is a need to measure longevity traits in commercial herds for more accurate genomic predictions (Poulsen *et al.,* 2020). Improving commercial sow longevity is often challenging through direct selection (Tart *et al.,* 2013) because it takes a long time for the traits to become available.

Commercial sows are crosses between breeds and may trace back to the nucleus pedigrees; thus, it is crucial to utilize the most appropriate genomic prediction methods. The Meta-founders (MF) approach in single-step genomic best linear unbiased prediction (ssGBLUP) has been recommended in multi-breed evaluations that combine genomic and pedigree infrmation because it ensures the compatibility between **G** and **A** (Legarra *et al*., 2015; Masuda *et al*., 2022). In multi-breed animals, the genomic regions are inherited from their respective parental breeds (Sevillano et al., 2019). Therefore, it is important to improve the compatibility of **G** and **A** in multi-breed populations as reported by Lourenco *et al*., (2016). Other approaches have used Unknown parent groups (UPGs) in pedigree-based animal models to account for genetic trend change not accounted for by known pedigree relationships (Westell *et al*., 1988). The incorporation of UPGs into ssGBLUP modifies H^{-1} (Misztal et al., 2013) and may decrease genetic trend bias and inflation of genomic estimated breeding values (GEBV) as noted by Tsuruta *et al*., (2014).

Our objective in this study was to evaluate the feasibility of multi-breed genomic predictions for a crossbred sow retention trait using ssGBLUP. Compatibility of **G** and **A** in the evaluations was ensured by using different MF definitions. The MF approach was compared with UPGs, multi-breed, and single-breed evaluations, and the four scenarios were compared based on estimates of bias and dispersion.

Materials & Methods

Data and Variance Components. Sow retention phenotypic data of 30,000 pedigreed F1 sows were obtained from five different commercial sow farms from 2019 to 2024. The F1 pedigrees trace back to Yorkshire and Landrace nucleus purebred lines. The F1 sows did not have genotypes; however, nearly 200,000 genotypes were utilized from the two pure nucleus breeds.

A pedigree-based animal model (presented below) was used to estimate variance components for the sow retention trait, which was a binary trait defined as the sow's ability to reach the fourth parity, and at least 55% of the sows made it to the fourth parity in this dataset.

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D_1}\boldsymbol{u} + \mathbf{D_2}\boldsymbol{i} + \mathbf{D_3}\boldsymbol{j} + \boldsymbol{e} \quad (1)$$

where $\boldsymbol{y}$, $\boldsymbol{\beta}$, $\boldsymbol{u}$, $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{e}$ are the phenotype, fixed, additive animal genetic, maternal additive genetic, non-additive random, and residual effects vectors, respectively. $\mathbf{X}$, $\mathbf{D_1}$, $\mathbf{D_2}$, and $\mathbf{D_3}$ are the incidence matrices relating the effects contained in $\boldsymbol{\beta}$, $\boldsymbol{u}$, $\boldsymbol{i}$, and $\boldsymbol{j}$, to the phenotypes in $\boldsymbol{y}$. For the estimation of variance components, we used a Bayesian approach via Gibbs sampling and a linear model; thus, heritabilities were estimated on the observed scale. Genetic relationships among animals were incorporated using different pedigrees, but without genomic information.

Estimation of genomic breeding values and scenario comparisons. Four scenarios were used to evaluate the feasibility of multi-breed ssGBLUP evaluations for the F1 sow longevity trait using the BLUPF90 software (Lourenco *et al*. 2022). The scenarios were defined as: ssGBLUP_MF, ssGBLUP_UPGs, ssGBLUP with two breeds, and ssGBLUP with each breed individually. The ssGBLUP_MF utilized a combined pedigree from both breeds with two Metafounders, ssGBLUP_UPGs had unknown parents groups applied, ssGBLUP with two breeds used a combined two breed pedigree but without assignment of MF or UPGs, and ssGBLUP with each breed individually only used a single breed pedigree with the other breed (sire or dam) assigned to missing in the first generation. Combined breed analyses utilized a combined G matrix using combined data. For the MF, a matrix (gamma.mat) was calculated to obtain the within and across population relationship as described by Legarra *et al*., (2015)

A linear regression (LR) validation, as proposed by Legarra and Reverter, (2018) was used to compare scenarios. Two datasets, "whole" and "partial" were used for the analyses. The whole (w) dataset included all the animals' phenotype information, whilst the partial (p) dataset omitted phenotypic information from 2019 to 2020 (i.e., from validation animals). The omitted phenotypic information accounted for about 13.4% of all the phenotypic records. The bias and dispersion linear estimators for the validation animals were calculated as shown below (Legarra and Reverter, 2018; Londoño-Gil *et al*., 2025).

$$Dispersion = \frac{cov(\hat{u}_p, \hat{u}_w)}{var\ (\hat{u}_p)} \quad (2)$$

$$Bias = \frac{\bar{\hat{u}}_p - \bar{\hat{u}}_w}{\sigma_u} \quad (3)$$

Where, $\hat{u}$ and $\bar{\hat{u}}$ are GEBV and average GEBV, respectively, for the whole and partial datasets. σ_u is the square root of genetic variance. A Pearson's correlation between whole and partial GEBV was computed for the validation animals as well.

Results

Estimates of variance components and heritabilities for the sow retention trait are shown in Table 1. The heritability estimates were low and generally similar among the different pedigrees used for the analyses. Standard errors were too small (0.00) and are not shown in the table.

The dispersion and bias of genomic predictions for the validation of animals across the four scenarios are shown in Table 2. No dispersion was observed for ssGBLUP_MF, ssGBLUP with two breeds approaches, whereas ssGBLUP_UPGs yielded dispersion close to the ideal value (i.e., 1.0). However, it was the worst for the ssGBLUP with individual Yorkshire and Landrace pedigrees. Optimum (0.0) and close to optimum biases were observed in the ssGBLUP_UPGs and in the ssGBLUP with two breeds scenarios, respectively. However, a slight negative bias was observed with the MF scenario. Pearson's correlations for validation GEBV were highest for

ssGBLUP_MF and ssGBLUP with two breeds, and it was zero for ssGBLUP with individual breed pedigrees.

Table 1. Variance component estimates and heritabilities for the sow retention trait using different breed pedigrees.

Variance	YY[1]	LL[2]	Combined YY & LL
σ_r^2	0.002	0.002	0.002
σ_m^2	0.010	0.012	0.008
σ_d^2	0.021	0.015	0.018
σ_e^2	0.214	0.218	0.221
h^2	0.080	0.060	0.070

σ_e^2 - residual error variance; σ_d^2 – direct additive (animal) genetic variance; σ_m^2 – maternal genetic variance (dam litter)*;* σ_r^2 – non-additive random variance (herd seasonality effect)*,* [1]Yorkshire, and [2]Landrace

Table 2. Dispersion and bias for the genomic predictions of sow retention obtained with different model strategies.

Scenarios	**Dispersion**	**Bias**	**Correlation**
ssGBLUP_MF	1.00	-0.16	0.67
ssGBLUP_upgs	0.81	0.00	0.56
ssGBLUP with two breeds	1.00	0.07	0.66
ssGBLUP_YY	0.00	0.00	0.00
ssGBLUP_LL	0.04	0.00	0.00

Discussion

The low heritability estimates (Table 1) for the sow retention trait in this study were consistent with those reported in other studies (Hong *et al.,* 2019). Low heritability implies that sow retention could be highly influenced by the environment or management of sows at the farms. Nucleus selections focus on maternal and performance traits that may have unfavorable genetic correlations with longevity (Lopez-Serrano *et al*., 2000). Limited trait data, together with the genotype x environment interactions, make genetic improvement of commercial sow retention a challenging task.

The zero Pearson's correlation among GEBV for the ssGBLUP with individual breed pedigrees for the validation animals suggests that this scenario is not appropriate, in agreement with the LR metrics as dispersion was noticeable. Because commercial sows are often an F1 cross from purebreds, a pedigree that utilizes all available breed base populations should be used in any evaluation method. There were differences in the LR estimates among the four genetic evaluation scenarios for the sow retention trait (Table 2). Among the ssGBLUP_MF, ssGBLUP_UPGs, and ssGBLUP with two breeds pedigree scenarios, the one with UPGs had dispersion of less than one (0.81), i.e., GEBV of validation animals were overestimated. Combining breeds with or without MF has been reported to have higher prediction accuracy (Londoño-Gil *et al.,* 2025). Moreover, combining genetically close populations into reference populations increased prediction accuracy with MF as reference population size increased (Wicki *et al*., 2023). Similarly, in our study near- or optimum dispersion metrics and Pearson's correlations were observed when breeds were combined. Overall, the MF yielded the best metrics compared to the UPGs scenario. Although the ssGBLUP_UPGs scenario had no bias (0) compared to other scenarios, and this agrees with

previous literature (Tsuruta *et al*., 2014; Masuda *et al*., 2022). This could be explained by the minimal discrepances between **G** and **A**, and there is limited missing in the pedigree.

Conclusions

Findings from this study indicate that it is important to combine both Yorkshire and Landrace breeds in this single F1 sow retention trait evaluation. An appropriate methodology should be chosen to avoid loss of precision in predicting breeding values. The MF or UPGs approaches may be beneficial; however, each has its own limitations.

References

Hong, J. K., Y. M. Kim, K. H. Cho, E. S. Cho, D. H. Lee, and T. J. Choi. 2019. J. Anim. Sci. 32:1077–1083. doi:10.5713/ajas.18.0789.

Legarra, A., O. F. Christensen, Z. G. Vitezica, I. Aguilar, and I. Misztal. 2015. Genetics. 200:455–468. doi:10.1534/genetics.115.177014.

Legarra, A., and A. Reverter. 2018. Genetics Selection Evolution. 50. doi:10.1186/s12711-018-0426-6.

Londoño-Gil, M., R. López-Correa, I. Aguilar, C. U. Magnabosco, J. Hidalgo, F. Bussiman, F. Baldi, and D. Lourenco. 2025. 142:43–56. doi:10.1111/jbg.12882.

Lopez-Serrano, ´ M, N. Reinsch, H. Looft, and E. Kalm. 2000. Genetic correlations of growth, backfat thickness and exterior with stayability in large white and landrace sows. Available from: www.elsevier.com/locate/livprodsci

Lourenco, D. A. L., S. Tsuruta, B. O. Fragomeni, C. Y. Chen, W. O. Herring, and I. Misztal. 2016. J. Anim. Sci. 94:909–919. doi:10.2527/jas.2015-9748.

Lourenco D., Tsuruta S., Aguilar I., Masuda Y., Bermann M. et al. (2022) Proc. 12th WCGALP, Rotterdam, The Netherlands.

Masuda, Y., P. M. VanRaden, S. Tsuruta, D. A. L. Lourenco, and I. Misztal. 2022a. J. Dairy Sci. 105:923–939. doi:10.3168/jds.2021-20293.

Misztal, I., Z. G. Vitezica, A. Legarra, I. Aguilar, and A. A. Swan. 2013. Unknown-parent groups in single-step genomic evaluation. Journal of Animal Breeding and Genetics. 130:252–258. doi:10.1111/jbg.12025.

Poulsen, B. G., B. Nielsen, T. Ostersen, and O. F. Christensen. 2020. J. Anim. Sci. 98:1–8. doi:10.1093/JAS/SKAA183.

Sevillano, C. A., H. Bovenhuis, and M. P. L. Calus. 2019. Front. Genet. 10. doi:10.3389/fgene.2019.00418.

Tart, J. K., R. K. Johnson, J. W. Bundy, N. N. Ferdinand, A. M. McKnite, J. R. Wood, P. S. Miller, M. F. Rothschild, M. L. Spangler, D. J. Garrick, S. D. Kachman, and D. C. Ciobanu. 2013. Anim. Genet. 44:387–397. doi:10.1111/age.12028.

Tsuruta, S., I. Misztal, D. A. L. Lourenco, and T. J. Lawlor. 2014. J. Dairy Sci. 97:5814–5821. doi:10.3168/jds.2013-7821.

Westell, R. A., R. L. Quaas, and L. D. Van Vleck. 1988. J. Dairy Sci. 71:1310–1318. doi:10.3168/jds.S0022-0302(88)79688-5.

Wicki, M., J. Raoul, and A. Legarra. 2023. J. Dairy Sci. 106:5570–5581. doi:10.3168/jds.2022-23114.