Journal of Animal Breeding and Genetics

WILEY

Journal of Animal Breeding and Genetics

ORIGINAL ARTICLE OPEN ACCESS

Indirect Genomic Predictions for Indicine Cattle Breeds With SNP Effects From a Multi-Breed Genomic Evaluation

Marisol Londoño-Gil[1,2] | Jorge Hidalgo[2] | Andres Legarra[2,3] | Claudio U. Magnabosco[4] | Fernando Baldi[1] | Daniela Lourenco[2]

[1]Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista Júlio de Mesquita Filho, Jaboticabal, São Paulo, Brazil | [2]Department of Animal and Dairy Science, University of Georgia, Athens, Georgia, USA | [3]Council on Dairy Cattle Breeding, Bowie, Maryland, USA | [4]Embrapa Cerrados, Brasilia, Brazil

Correspondence: Marisol Londoño-Gil (londono.gil@unesp.br)

Received: 18 February 2025 | **Revised:** 3 July 2025 | **Accepted:** 30 July 2025

Funding: This study was funded by São Paulo Research Foundation (FAPESP, grants #2020/14275-5, #2022/15168-3 and #2021/04807-2) and by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grant #402313/2021-6).

Keywords: genomics | metafounders | reference population | selection | young genotyped animals

ABSTRACT

Indirect predictions (IP) are used for young genotyped animals that lack phenotypes (of their own or from progeny) or are from commercial herds. The former can be left behind because they do not contribute to the official genomic evaluations. The latter are often excluded from the evaluations because they are not registered and may not have pedigree information. Including such animals could result in inflated and biased genomic breeding values (GEBV). In Brazil, pedigree, phenotype and genotype information is scarce for important breeds like Brahman, Guzerat, and Tabapua, while the Nellore breed has a substantial amount of information. IP for young animals of these breeds based on a larger reference population could enhance genomic selection accuracy. Our objective in this study was to compute IP for young genotyped Nellore, Brahman, Guzerat, and Tabapua animals using single- and multi-breed analyses, with or without metafounders (MF) to model genetic differences across breeds. Records from the four breeding programs of the National Association of Breeders and Researchers (ANCP—Ribeirão Preto, SP, Brazil) were used. Data included pedigree (4.2 M), phenotypes (329 K), and genotypes (63.5 K) across all breeds. The number of genotyped animals within each breed was 58,574 for Nellore, 3102 for Brahman, 1389 for Guzerat, and 427 for Tabapua. The analysed traits were adjusted weight at 210 (W210) and 450 (W450) days of age and the scrotal circumference at 365 (SC365) days of age. IP were derived as the sum of the SNP effects weighted by the gene content using different reference populations: multi-breed with or without MF, Nellore, or within-breed. Scenarios were compared using the linear regression (LR) method for bias, dispersion, and accuracy. Adding MF decreased bias and under- or overdispersion and slightly increased the accuracy of IP. Combining breeds increased the accuracy of IP, mainly benefiting breeds with a small number of genotypes. These findings suggest that when young genotyped animals are not included in an official multi-breed evaluation in zebuine cattle from Brazil, robust IP can be obtained with proper modelling, regardless of the breed. This helps obtain fast genomic predictions for young animals without overwhelming the evaluation system with too many animals.

1 | Introduction

Indirect predictions (IP) can help offload the computational burden of large-scale evaluations. This is because many genotyped individuals are young, meaning they have no phenotypes or progeny phenotypes. These animals must not be included in the official evaluations (Tsuruta et al. 2021). Including them increases computational cost without adding informative data to the model and may introduce bias and inflation in the GEBV because their lack of phenotypic contribution causes their predictions to regress toward the mean, altering the overall genetic evaluation (Tsuruta et al. 2021).

Additionally, commercial farms are genotyping animals for herd replacement and management, increasing the need for IP as those animals will never take part in an official genomic evaluation (Londoño-Gil et al. 2023; Tsuruta et al. 2021). Through IP, superior young selection candidates can be identified earlier, especially in places where genomic evaluations are not frequent. Furthermore, if these animals have incomplete pedigrees, excluding them from official evaluations may prevent a decline in accuracy and an increase in inflation and bias in genomic breeding values (GEBV) (Lourenco et al. 2018). IP are obtained as the sum of the SNP effects weighted by the gene content. These SNP effects can be calculated based on the GEBV from the official genomic evaluations. Hence, these IP provide quick genomic predictions for newly genotyped young animals without affecting routine evaluations or compromising the accuracy and stability of GEBV (Tsuruta et al. 2021).

To obtain reliable predictions, it is essential to calculate the SNP effects using a large group of phenotyped and genotyped animals encompassing most of the chromosome segments segregating in a population (Lourenco et al. 2015; Tsuruta et al. 2021). In Brazil, the number of phenotyped and genotyped animals is limited for breeds such as Brahman, Guzerat, and Tabapua, generating challenges in enhancing the accuracy of genomic selection for these breeds. The difficulty lies in constructing a large reference population within each single cattle breed, as Hayes et al. (2023) indicated. In contrast, the Nellore breed already benefits from a well-established and large reference population, which allows for more reliable GEBV and SNP effect estimation in single-breed evaluations. Moreover, the independent chromosome segments within breeds comprise high linkage disequilibrium (LD) blocks with low recombination rates, indicating a gradual formation of new segments, as evidenced by Hidalgo et al. (2023). Therefore, acquiring IP for these breeds through a broader reference population (Nellore), as in the multi-breed approach, could improve the accuracy of SNP effect estimation and genomic predictions.

Most multi-breed evaluations use a single genetic effect for all the genotyped animals and a single genomic relationship matrix (**G**), resulting in the same SNP effects for all breeds (Misztal et al. 2022). However, there are more appropriate approaches than this since the multi-breed model should include effects that account for breed differences while allowing for the compatibility of genomic and pedigree relationships (Misztal et al. 2022). Otherwise, the genetic predictions may be compromised (Steyn et al. 2019). Efficiently, the distinct genetic bases across breeds can be modelled with metafounders (MF) (Bermann et al. 2023; Legarra et al. 2015).

MF serve as virtual ancestors, simultaneously encapsulating the genetic contributions from paternal and maternal lines or breeds. This establishes relationships and facilitates considering inbreeding within and across base populations (Legarra et al. 2015). Consequently, they play a crucial role in addressing genetic differences between breeds and refining the accuracy of genomic selection. By integrating MF into genomic prediction, breeders can more precisely forecast individuals' genetic potential, regardless of their diverse backgrounds. Moreover, MF are instrumental in delineating the origins of cattle populations or their chromosome segments, offering insights into the biologically significant connections among distinct breeds shaped by evolutionary forces and geographical or genetic differences (Callister et al. 2022). Additionally, MF may enable a more precise representation of genetic diversity and relationships within a multi-breed population, indispensable for breeding programmes aiming to enhance specific traits across different breeds.

Therefore, this work aimed to compute IP for young genotyped Nellore, Brahman, Guzerat, and Tabapua animals using single-breed reference populations or a multi-breed reference population in the presence or absence of MF to model the genetic differences across breeds. A secondary objective was to determine the minimum number of genotyped animals in the official Nellore evaluation to estimate accurate SNP effects and IP.

2 | Material and Methods

This study did not require Animal Care and Use Committee approval, as the dataset was acquired from an existing database.

2.1 | Data

The National Association of Breeders and Researchers (*Associação Nacional de Criadores e Pesquisadores*—ANCP, Ribeirão Preto—SP, Brazil) provided the data, which comprised pedigree (4,207,516), phenotypic (328,748), and genomic (63,492) information. This association runs the breeding programs of Nellore (Nellore Brazil program), Brahman (Brahman Brazil program), Guzerat (Brahman Brazil program), and Tabapua (Tabapua Brazil program). The dataset contained phenotypes for adjusted weight at 210 (W210) and 450 (W450) days of age and for scrotal circumference adjusted at 365 (SC365) days of age. Animals were genotyped with the ZBN Zoetis chip with 74,653 SNP (Table 1). The criteria for the quality control of SNP were based on minor allele frequency (MAF), call rate, and Mendelian conflicts. Thus, markers with MAF <0.05, call rate <0.90, and monomorphic were excluded. Quality control was performed using the preGSF90 software from the BLUPF90 family of programs (Misztal et al. 2015).

The model used is presented below:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z}_1\mathbf{u} + \mathbf{Z}_2\mathbf{m} + \mathbf{Z}_3\mathbf{c} + \mathbf{e}$$

TABLE 1 | Descriptive data for adjusted weight at 210 (W210) and 450 (W450) days of age, and scrotal circumference at 365 (SC365) days of age for the Nellore Brahman, Guzerat, and Tabapua breeds.

Trait		Nellore	Brahman	Guzerat	Tabapua	Total
W210 (kg)	N° phenotyped	204,019	88,099	26,632	9998	328,748
	Mean	204.81	182.93	188.73	191.19	197.23
	SD	31.43	31.38	36.52	32.25	33.38
	N° genotyped	58,574	3102	1389	427	63,492
W450 (kg)	N° phenotyped	150,384	40,923	21,797	6699	219,803
	Mean	319.29	271.24	290.95	289.64	306.63
	SD	58.21	56.62	65.33	60.29	61.85
	N° genotyped	46,162	3102	1389	427	51,080
SC365 (cm)	N° phenotyped	100,513	16,536	7164	2168	126,381
	Mean	22.43	20.17	21.03	21.69	22.04
	SD	2.68	2.59	2.49	2.39	2.77
	N° genotyped	26,703	3102	1389	427	31,621

Abbreviations: N = Number of records, N° phenotyped and Genotyped = Number of phenotyped and genotyped individuals, SD = standard deviation, Total = Descriptive statistics for the four breeds combined.

where **y**, **β**, **u**, **m**, **c**, and **e** are the vectors of phenotypes, fixed effects (contemporary group and age of dam classes), direct additive genetic, maternal additive genetic, maternal permanent environmental, and residual effects, respectively, and **X**, $\mathbf{Z_1}$, $\mathbf{Z_2}$, and $\mathbf{Z_3}$ are the incidence matrices for the effects contained in **β**, **u**, **m**, and **c**. Maternal genetic and permanent maternal environmental effects were considered only for W210. The correlation between direct and maternal genetic effects was considered as zero.

2.2 | Analyses

Four different genomic analysis scenarios were performed using ssGBLUP (Aguilar et al. 2010) to predict the GEBV and to compute SNP effects and, therefore, IP for young genotyped Nellore, Brahman, Guzerat, and Tabapua animals. These young animals were born after 2021 (W210) or 2020 (W450 and SC3650). The SNP effects needed for IP were estimated based on the following scenarios:

1. Multi-breed ssGBLUP with no metafounders (no MF).
2. Multi-breed ssGBLUP with four metafounders, one for each breed (MF).
3. Single-breed Nellore ssGBLUP (Nellore-based).
4. Single-breed ssGBLUP (Single Breed).

Without MF (no MF, Nellore-based, and single-breed scenarios), the covariance structure among animals in ssGBLUP (Legarra et al. 2009) was defined as in Aguilar et al. (2010) and constructed using the BLUP90IOD3 program from the BLUPF90 suite of programs (Lourenco et al. 2022; Misztal et al. 2014):

$$\mathbf{H^{-1}} = \mathbf{A^{-1}} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G^{-1}} - \mathbf{A_{22}^{-1}} \end{bmatrix}$$

where **G** was constructed as in VanRaden (2008), that is, $\mathbf{G} = \mathbf{ZZ}'$, where:

$$\mathbf{Z} = \frac{(\mathbf{M} - \mathbf{P})}{\sqrt{k}}$$

with **M** being the matrix of n SNP genotypes for each animal, with elements indicating the gene content (i.e., copies of the reference allele) and set to 0, 1, and 2, for the homozygote, heterozygote, and alternative homozygote, respectively; p_j corresponds to the reference allele frequency at locus *j*. The **P** matrix contains elements equal to $2p_j$ to centre elements of **Z** by the mean of gene content, that is, twice the frequency of the reference allele at locus *j*. Additionally, *k* is $2\sum_{j=1}^{n} p_j(1-p_j)$. Allele frequencies were calculated based on the current genotyped population.

For the MF scenario, the MF were related by a matrix of additive relationships **A**(**Γ**) given by a positive definite matrix that was invertible, **Γ**, as follows:

$$\mathbf{\Gamma} = \begin{bmatrix} 0.71 & 0.66 & 0.68 & 0.67 \\ 0.66 & 0.69 & 0.68 & 0.66 \\ 0.68 & 0.68 & 0.76 & 0.69 \\ 0.67 & 0.66 & 0.69 & 0.73 \end{bmatrix}$$

where the relationships within MF (Diagonal of **Γ**) and between MF (off-diagonals of **Γ** belonged to Nellore, Brahman, Guzerat, and Tabapua) (Londoño-Gil et al. 2025). This matrix was computed using the GAMMAF90 program from the BLUPF90 suite of programs (Lourenco et al. 2022; Misztal et al. 2014). Then, the inverse relationship matrix with pedigree, genomics, and MF information $\mathbf{H}^{\Gamma-1}$ was calculated using the BLUP90IOD3 program from the BLUPF90 suite of programs (Lourenco et al. 2022; Misztal et al. 2014), as follows:

$$\mathbf{H}^{\Gamma-1} = \mathbf{A}^{\Gamma-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{05}^{\Gamma-1} - \mathbf{A}_{22}^{\Gamma-1} \end{bmatrix}$$

where $\mathbf{H}^{\Gamma-1}$ is the inverse matrix that combines the genomic and pedigree matrix, considering MF; $\mathbf{A}^{\Gamma-1}$ is an inverse matrix of additive relationships with MF; $\mathbf{G}_{05}^{\Gamma-1}$ is the inverse of the genomic relationship matrix of genotyped animals, with allele frequencies $p_j = 0.5$, and $\mathbf{A}_{22}^{\Gamma-1}$ is the matrix of additive relationships between genotyped animals, considering MF. $\mathbf{G}_{05} = \mathbf{ZDZ}' = \frac{2}{m}\mathbf{ZZ}'$, with $\mathbf{Z}$ coded as before, and m number of markers.

In the scenarios without MF, $\mathbf{G}$ was tuned to have the same means of diagonals and off-diagonals as $\mathbf{A}_{22}$, therefore:

$$\mathbf{G}^{\text{tunned}} = b\mathbf{G} + a$$

where a and b are the tuning parameters, calculated as in Christensen et al. (2012) by solving the system of equations $\text{Avg}(\text{diag}(\mathbf{G}^{\text{blended}}))b + a = \text{Avg}(\text{diag}(\mathbf{A}_{22}))$, $\text{Avg}(\text{offdiag}(\mathbf{G}^{\text{blended}}))b + a = \text{Avg}(\text{offdiag}(\mathbf{A}_{22}))$, where a adds the extra average relationship and at the same time models the difference in μ from the pedigree base to the genomic base; and the b considers the reduction in variance due to drift.

For all the scenarios and to avoid singularity problems, $\mathbf{G}$ $(\mathbf{G}^{\text{tunned}})$ was blended to be invertible with a small percentage ($\alpha = 0.05$) of the pedigree relationship matrix among genotyped animals, which corresponds to adding a residual polygenic effect as follows:

- MF scenario: $\mathbf{G}_{05}^{\text{blended}} = (1-\alpha)\mathbf{G}_{05} + \alpha\mathbf{A}_{\Gamma 22} = (1-\alpha)\left(\frac{2}{m}\mathbf{ZZ}'\right) + \alpha\mathbf{A}_{\Gamma 22}$
- Other scenarios (without metafounders): $\mathbf{G}^{\text{blended}} = (1-\alpha)\mathbf{G}^{\text{tunned}} + \alpha\mathbf{A}_{22}$

After obtaining the GEBV in each of the four scenarios, SNP effects were computed using POSTGSF90 (Lourenco et al. 2022; Misztal et al. 2014) as proposed by Wang et al. (2012) for the approaches without MF:

$$\hat{\mathbf{a}} \mid \hat{\mathbf{u}} = (1-\alpha)b\frac{1}{k}\mathbf{Z}'\mathbf{G}^{\text{blended}-1}\hat{\mathbf{u}}$$

and as proposed by Legarra et al. (2024) in the scenario with MF:

$$\hat{\mathbf{a}} \mid \hat{\mathbf{u}} = (1-\alpha)\mathbf{Z}'\frac{2}{m}\mathbf{G}_{05}^{\text{blended}}\hat{\mathbf{u}}$$

where $\hat{\mathbf{a}}$ is a vector of SNP effects, and $\hat{\mathbf{u}}$ is a vector of GEBV for the old genotyped animals in the evaluation system. For the MF scenario, there was no tuning in $\mathbf{G}$, since the changes to ensure the compatibility between $\mathbf{A}_{22}$ and $\mathbf{G}$ were made in $\mathbf{A}_{22}$ and $\mathbf{A}$ but not in $\mathbf{G}$.

Once the SNP effects were calculated, the IP was obtained via PREDF90 (Lourenco et al. 2022; Misztal et al. 2014) as in Lourenco et al. (2018):

$$\hat{\mathbf{u}}_{\text{IP}} = \hat{\mu}_g + \mathbf{Z}_v\hat{\mathbf{a}}$$

where $\hat{\mathbf{u}}_{\text{IP}}$ is a vector of estimated IP for the validation animals, $\mathbf{Z}_v$ is the matrix of SNP content for the validation animals, centred by the same allelic frequencies as the original $\mathbf{Z}$, $\hat{\mathbf{a}}$ is a vector of SNP effects, and $\hat{\mu}_g$ is the weighted average of GEBV for genotyped animals (without including the validation animals), computed as:

$$\hat{\mu}_g = \mathbf{1}'\mathbf{G}^{-1}\hat{\mathbf{u}}$$

where $\mathbf{1}$ is a vector of 1's. Adding $\hat{\mu}_g$ is necessary to guarantee that the IP is on the same scale as the GEBV from the official evaluations, considering the tuning of $\mathbf{G}$ to align with $\mathbf{A}$, a general adjustment to account for differences between genomic and pedigree bases (Lourenco et al. 2018). For the MF method $\hat{\mu}_g$ is always equal to zero, since in the MF scenario, there is no tuning (Legarra et al. 2015).

2.3 | Validation

All scenarios were compared using the linear regression method (LR) proposed by Legarra and Reverter (2018). The "whole" dataset included all available sources of information for the reference and validation animals (GEBVs, $\hat{u}_w$), whereas the "partial" dataset omitted all the sources of information for the validation animals, i.e., young genotyped animals born from 2021 onwards for W210 or 2020 for W450 and SC365, for which we obtained the $\hat{\mathbf{u}}_{\text{IP}}$ (IP). Table 2 presents the number of records in the whole and partial datasets. The LR estimators (Legarra and Reverter 2018) of accuracy, bias, and dispersion for the validation animals were computed as follows: $\text{accuracy} = \widehat{acc} = \sqrt{\frac{\text{cov}(\hat{u}_w, \hat{u}_{ip})}{(1-\overline{F})\sigma_u^2}}$; $\text{bias} = \hat{\Delta}_{w,ip} = \frac{\overline{\hat{u}}_{ip} - \overline{\hat{u}}_w}{\sigma_u}$; $\text{dispersion} = \hat{b}_{w,ip} = \frac{\text{cov}(\hat{u}_{ip}, \hat{u}_w)}{\text{var}(\hat{u}_{ip})}$; where $\hat{u}$ $\widehat{(u)}$ refers to the predicted breeding values (mean predicted breeding values), and the subscript indicates the whole (w, i.e., GEBV) or partial (i.e., IP) datasets. $\overline{F}$ is the mean inbreeding coefficient of the validation animals.

We computed Pearson correlations (using R; R Core Team 2023) among SNP effects from the different scenarios to study the agreement across models. In addition, to see the minimum number of genotyped animals necessary to obtain reliable IP for the Nellore breed, we estimated Pearson correlations among SNP effects and IP derived from GEBV of a variable number of genotyped animals. The latter analysis was performed after running the PREGSF90 software (Lourenco et al. 2022; Misztal et al. 2014) to conduct singular value decomposition (SVD) of $\mathbf{Z}$ to estimate the number of the largest eigenvalues explaining 90%, 95%, 98%, and 99% of the variation in $\mathbf{G}$. The number of the largest eigenvalues explaining an x% of the variation in $\mathbf{G}$ was used as the number of Nellore genotyped animals included in predicting GEBV. The GEBV were used to derive the SNP effects and IP, which were compared against the SNP effects and IP obtained using all the available Nellore genotypes, i.e., 50 k (58,574), representing our benchmark.

3 | Results and Discussion

3.1 | SNP Effects

Figure 1 and Figures S1–S3 depict the correlations among SNP effects through the different scenarios. High positive correlations (R close to 1) among multi-breed scenarios (MF/no MF)

TABLE 2 | Number of records in the whole and partial dataset and validation animals for whom the indirect predictions (IP) were estimated to validate using the Linear Regression method to assess statistics for the different scenarios for adjusted weight at 210 (W210), and 450 (W450) days of age and scrotal circumference at 365 (SC365) days of age.

Trait		Nellore	Brahman	Guzerat	Tabapua	4 Breeds
W210 (kg)	Whole	204,019	88,099	26,632	9998	328,748
	Partial	187,268	84,475	26,131	9895	307,769
	IP	12,936	330	151	61	[a]
W450 (kg)	Whole	150,384	40,923	21,797	6699	219,803
	Partial	126,987	37,039	21,168	6558	191,752
	IP	14,558	521	279	102	[a]
SC365 (cm)	Whole	100,513	16,536	7164	2168	126,381
	Partial	83,472	14,606	6794	2089	106,961
	IP	8615	271	152	61	[a]

Abbreviation: 4 Breeds, All four breeds in the multi-breed evaluation.
[a]The validation animals depended on the breed evaluated but are the sum of the four breeds.

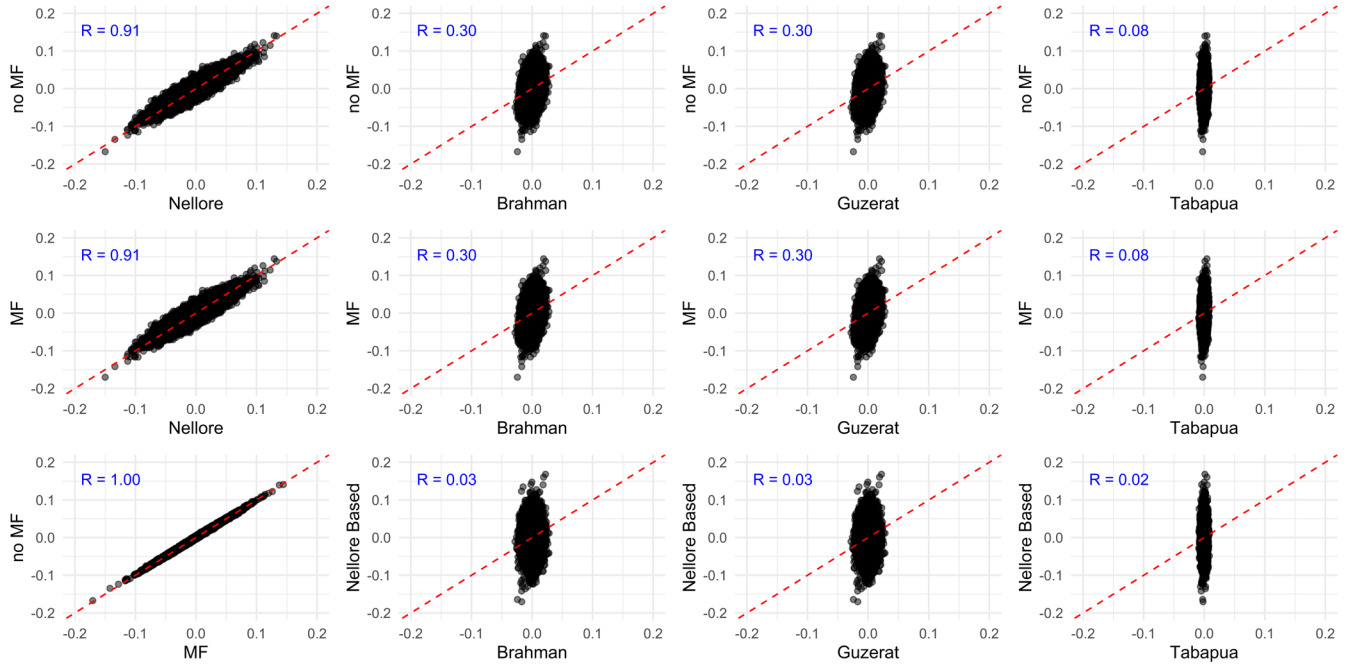

FIGURE 1 | SNP effects correlations (R) between scenarios for the trait adjusted weight at 210 days of age (W210). Each correlation represents corr (X, Y), where X and Y are vectors of SNP effects estimated under two different scenarios. No MF: SNP effects based on multi-breed GEBV without metafounders and using the default **G**, i.e., with current allele frequencies. MF: SNP effects based on GEBV obtained when using four metafounders, one for each breed. Nellore-based: Based on the Nellore GEBV for the small breeds Brahman, Guzerat, and Tabapua. Single Breed (breed name is shown): Based on each breed GEBV, when evaluated as a single breed. MF and no MF were both multi-breed scenarios. [Colour figure can be viewed at wileyonlinelibrary.com]

and between multi-breed and the single-breed Nellore suggest they encapsulate similar SNP effects. Given its substantial representation in genotyped samples, this reinforces the reliance of multi-breed scenarios on Nellore data. However, low positive correlations (≤ 0.30) between MF/no MF and single-breed Brahman, Guzerat, and Tabapua indicate the existence of some shared genetic effects with discernible breed-specific differences, potentially due to insufficient genotyped animals and limited phenotypic information within the smaller breeds to estimate SNP effects accurately.

Another possible explanation is that the assumption of shared SNP effects across breeds does not hold; i.e., the quantitative trait loci (QTL) affecting the traits may differ across breeds. Using Nellore as a base to predict the SNP effects resulted in near-non-existent correlations, ranging from 0.02 to 0.03 (Figure 1, 3rd row).

When multiple breed populations are combined into a reference population, the SNP effects estimates tend to be dominated by the breeds contributing more data, primarily capturing effects

of SNPs in LD with QTL across all breeds or within the largest population (Karaman et al. 2021). The SNP effects are better estimated for the dominant breed due to the larger reference population, which has more information and dominates the allele frequency and SNP effects. In this study, we used the MF approach to efficiently model the breed origin, adjusting SNP effects and pedigree relationships. With MF, the models can account for each breed's inherent differences and contributions (Christensen et al. 2012; Legarra et al. 2015). Therefore, accounting for genetic relationships among Nellore, Brahman, Guzerat, and Tabapua when incorporating MF in the multi-breed evaluation bridged these gaps, which were prominent when SNP effects were derived only from the Nellore reference population, resulting in very low correlations.

Despite these advantages, it is important to acknowledge some limitations of the multi-breed and MF approaches. Since SNP effects are estimated from back-solved GEBVs, they represent a weighted average across all genotyped animals in the reference population, largely influenced by the largest breed, which in our case is Nellore. As a result, the SNP effects derived from multi-breed evaluations, even with MF, may not fully reflect breed-specific effects. For example, in our study, the correlation of SNP effects between the MF or no-MF scenario and the single-breed Nellore scenario was 0.91, indicating some divergence. Nevertheless, this did not compromise the overall performance of the IP. Therefore, the benefits of MF should be interpreted within the context of both the prediction objective and the breed composition of the reference population.

Moreover, the effectiveness of multi-breed genomic evaluations is highly dependent on the genetic relatedness among the breeds involved. Multi-breed prediction uses shared LD patterns, which arise from common ancestry and historical recombination events (Brito et al. 2011). In our study, the four indicine breeds share a *Bos indicus* origin, but have different selection histories and breeding programmes (Londoño-Gil et al. 2025). While some level of genetic relatedness exists, breed-specific divergence may have resulted in differences in LD structure and allele frequencies across the genome. These factors likely contribute to the observed variation in SNP effect correlations between breeds.

Steyn et al. (2019) showed the potential negative effect on prediction accuracy if SNP effects differ among breeds, especially if the scaling between $\mathbf{A}_{22}$ and $\mathbf{G}$ is incorrect and the number of SNPs is insufficient to capture QTL effects, leading to wrongly estimated SNP effects. Adjusting $\mathbf{G}$ elements by scaling diagonal and off-diagonal values to match the averages observed in $\mathbf{A}_{22}$ might mitigate the bias due to those differences (Bermann et al. 2021; Vitezica et al. 2011). A constant, $\hat{\mu}_g$, is used to address variations in the mean breeding values when using pedigree and genomic information, representing the mean breeding value for genotyped animals in the absence of selection pressure, regardless of allele frequencies (Bermann et al. 2021).

Nevertheless, when employing MF, pedigree relationships derived from related base populations alter $\mathbf{A}_{22}$ and $\mathbf{A}$, with the overall means absorbing the covariances across individuals (i.e., covariance in $\boldsymbol{\Gamma}$). Since animals are likely related in these populations, and because the populations are under selection, the mean of the breeding values may deviate from zero. However, this deviation is adjusted for by the general mean of the model, thereby scaling the breeding values across the entire pedigree and ensuring compatibility between $\mathbf{G}$ and $\mathbf{A}_{22}$ (Legarra et al. 2015), which could improve the SNP effects estimation and the prediction accuracy.

Furthermore, we evaluated the impact of the number of genotyped animals used to estimate breeding values on SNP effects and IP accuracy. Based on the SVD of $\mathbf{Z}$ for the Nellore breed, the number of the largest eigenvalues explaining 90%, 95%, 98%, and 99% of the variation in $\mathbf{G}$ was 5,443, 8,770, 13,641, and 17,289, respectively.

Figures 2 and 3 show the correlations among SNP effects and IP. Figure 2 shows that in this Nellore population, more than 17 k (number of the largest eigenvalues explaining 99% of the variation in $\mathbf{G}$) genotyped animals are needed to estimate SNP effects as accurately as using all (50 k) available genotypes. The correlations were ≥ 0.90 using over 35 k genotyped animals and peaked at 1.0 with 45 k animals. Nevertheless, Figure 3 illustrates that the impact on the IP is less significant, where the correlations were higher. For example, the correlation between 17 k and 50 k genotyped animals was 0.71 for SNP effects and 0.90 for IP. This is because IP gather the combined effect of all SNP effects, and many possible combinations of SNP effects can yield the same IP.

These findings highlight the importance of having enough genotyped animals to estimate SNP effects accurately. In our study, we kept the available phenotypic information constant in all the scenarios, and increasing the number of genotypes showed the value of realised genomic relationships, which were needed to harness the phenotypic information in the most optimal way, connecting more animals through their shared alleles or genomic similarity. Therefore, future studies should focus on optimising the number of genotyped animals and the phenotypic records to balance accuracy and efficiency in estimating SNP effects.

This observation is consistent with previous studies demonstrating the importance of reference population size for accurate genomic predictions, particularly for multi-breed populations (Daetwyler et al. 2010; Hayes et al. 2009). For instance, Takeda et al. (2021) explored the size of the reference population for the expected accuracy of genomic prediction in Japanese Black cattle. They demonstrated that a reference population comprising more than 5000 animals with phenotypes and genotypes was necessary for accurate genomic predictions in traits with varying heritability.

Hidalgo et al. (2023) showed the value of a large (154 k genotypes and 820 k phenotypes for body weight) reference population in broilers; these authors demonstrated that when ample phenotypic and genomic data is used in the prediction of breeding values, these are predicted with high accuracy; therefore, the SNP effect derivation can be done from a subset of GEBV with a size equivalent to the number of largest eigenvalues explaining 99% of the variation in the genomic relationship matrix.

In our case, the combined genotype count of small breeds amounts to less than 5000 animals, which is insufficient for accurately estimating SNP effects or representing independent chromosome segments segregating in the populations. The SVD showed that 6445, 10,089, 15,095, and 18,726 animals are needed

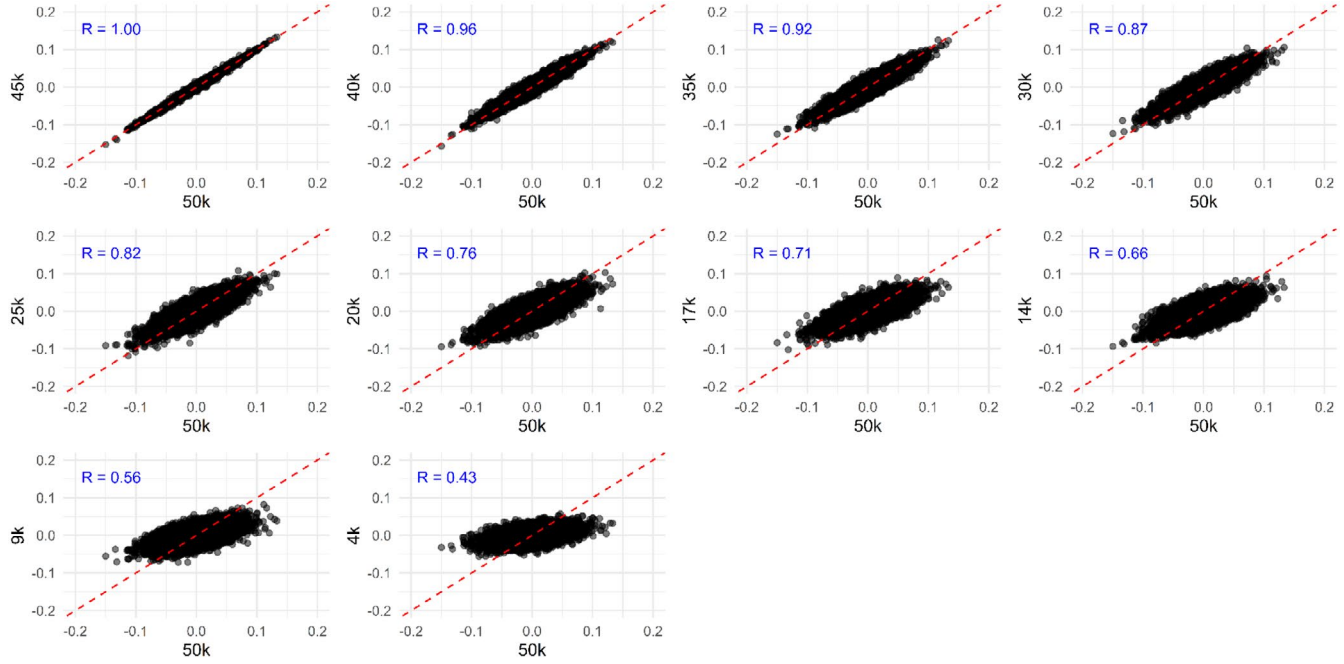

FIGURE 2 | Pearson's correlations (R) among adjusted weight at 210 days of age SNP effects derived from different numbers of genotyped animals in the Nellore breed. Each correlation represents corr (X, Y), where X and Y are vectors of SNP effects estimated using two different sample sizes of genotyped animals. [Colour figure can be viewed at wileyonlinelibrary.com]

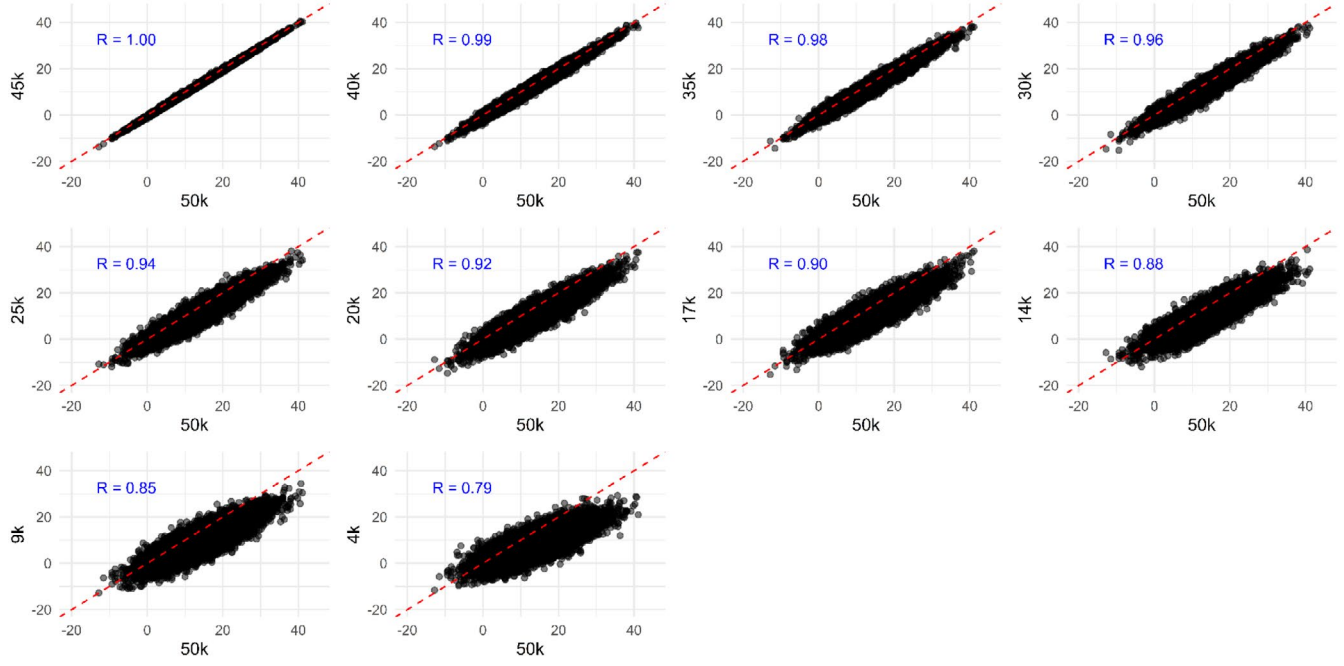

FIGURE 3 | Pearson's correlations (R) among adjusted weight at 210 days of age indirect predictions derived from different numbers of genotyped animals in the Nellore breed. Each correlation represents corr (X, Y), where X and Y are vectors of IP values estimated using two different sample sizes of genotyped animals. [Colour figure can be viewed at wileyonlinelibrary.com]

to explain 90%, 95%, 98%, and 99% of the genetic variation in the genomic relationship matrix of our multi-breed population. This emphasises the importance of using a large and genetically related reference population, such as Nellore in our case, to account for breed-specific effects that cannot be readily estimated when evaluated as a single breed.

3.2 | Evaluation of Indirect Predictions

Validation statistics for the IP are shown in Figure 4. Multi-breed scenarios consistently achieved the highest accuracy across most traits and breeds (W210, W450, SC365), with the MF scenario achieving the highest accuracy, closely followed

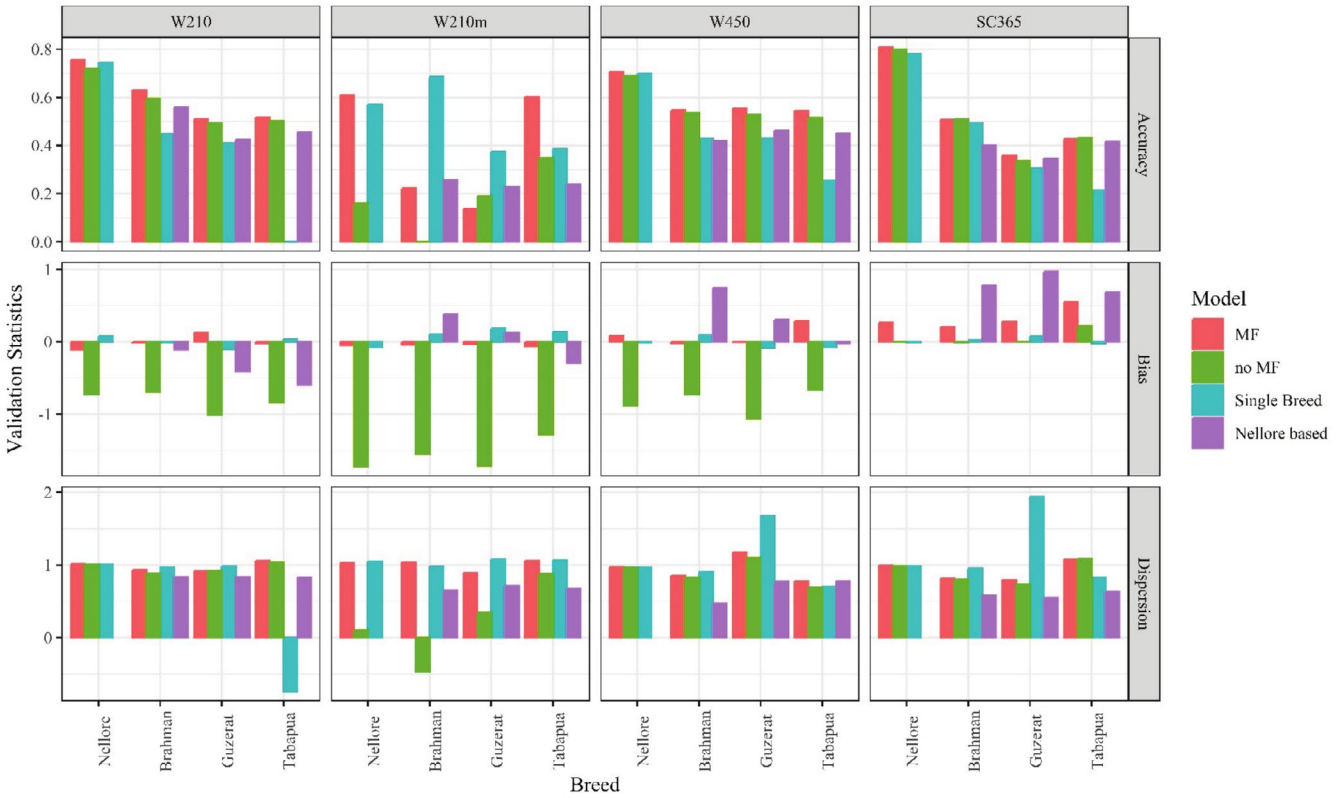

FIGURE 4 | Accuracy, bias, and dispersion for indirect predictions of the validation animals in each scenario and trait. W210, W210m, W450: Adjusted weight at 210 direct and maternal, and 450 days of age, respectively. SC365: Scrotal circumference at 365 days of age. MF: SNP effects and IP based on GEBV obtained when using four metafounders, one for each breed. no MF: Based on **G** default, with the current allele frequencies. Nellore-based: Based on the Nellore GEBV for the small breeds Brahman, Guzerat, and Tabapua. Single Breed: Based on each breed GEBV when evaluated as a single breed. MF and no MF were both multi-breed scenarios. [Colour figure can be viewed at wileyonlinelibrary.com]

by the no-MF scenario. This supports the hypothesis that breed-specific adjustments are beneficial. In fact, Londoño-Gil et al. (2025) found that multi-breed evaluations benefit from incorporating MF or unknown parent groups compared to ignoring breed differences. For smaller breeds, accuracy was notably higher when using multi-breed populations. This emphasises the importance of large populations to estimate SNP effects (and derive IP), while ensuring that the breed of interest is adequately represented. However, the comparison across scenarios reveals important nuances. While MF improves accuracy and reduces bias, the no-MF scenario still performs well in some cases, particularly when large datasets and strong pedigree connections are available.

The Nellore-based and single-breed scenarios generally exhibited lower accuracies (especially for Guzerat and Tabapua breeds). While applying Nellore-based GEBVs to small breeds demonstrated some utility, it did not accomplish the same level of accuracy as breed-specific evaluations in many cases, highlighting the limitations of across-breed genomic predictions. For instance, Nellore-based accuracy was higher than single-breed accuracy for W210 across all breeds; however, the opposite was observed for W210m. This discrepancy may indicate challenges in accurately estimating maternal effects, which require sufficient generations of data and multiple parities per cow. Furthermore, maternal effects skip one generation, which complicates their estimation, particularly when validations are limited or unevenly distributed by sex. Future studies should assess the distribution of validation animals (e.g., males vs. females) to ensure robust validation statistics for W210m.

The superiority of the MF scenario suggests that the remaining scenarios may not fully capture the comprehensive information encapsulated by the multi-breed approach, particularly for traits like W210 and W450, where the accuracy gap between MF and other scenarios was particularly pronounced. Accurate and unbiased prediction of breeding values for selection relies on phenotypic data connected through pedigree and genomic relationships of animals. The genetic similarity among animals, measured as the proportion of shared alleles, represents a main source of information in predicting breeding values because it connects phenotypic information across animals.

Contrary to traditional pedigrees that assume all the individuals in the base population or those not sharing known common ancestors as unrelated, genotypes have the potential to reveal existing relationships within and across the base populations (or breeds) and individuals unrelated in the pedigree. This concept is exploited in the metafounder theory (Legarra et al. 2015), which considers each ancestral population (or MF) as a finite-sized pool of gametes, with relationships within and across them. In addition to creating links between the populations, the MF approach decomposes the additive genetic relationships into breed-specific terms for each breed and a segregation term for the pair of breeds. It also accounts for different genetic base populations due to selection (changes in the mean of breeding

values and additive variance over time in a population) or crossbreeding (differences coming from different breeds), naturally overcoming possible compatibility issues between $\mathbf{A}_{22}$ and $\mathbf{G}$ matrices.

When MF are used, the overall mean breeding value ($\widehat{\mu_g}$) is absorbed into the model's general mean because MF accounts for ancestral relationships and adjusts the genetic base populations. This ensures compatibility between $\mathbf{G}$ and $\mathbf{A}_{22}$, aligning the scales of genomic and pedigree relationships. In scenarios without MF, $\widehat{\mu_g}$ is explicitly required to adjust for differences between genomic and pedigree-based breeding values, particularly when breed-specific SNP effects differ or when allele frequencies are inconsistent across populations (Steyn et al. 2019; Bermann et al. 2021).

These results indicate that using a multi-breed scenario can significantly improve the accuracy of IP. These findings are consistent with previous studies (Brito et al. 2011; Clasen et al. 2023; Daetwyler et al. 2010; Karaman et al. 2021) that reported improved accuracy and better calibration of genomic predictions by accounting for breed differences and using multi-breed reference populations. Furthermore, Bermann et al. (2023), Junqueira et al. (2020), and Legarra et al. (2015) demonstrated the utility of MF in accounting for breed-specific ancestral relationships and improving genomic predictions.

Regarding bias, the no-MF and Nellore-based approaches exhibited some bias, deviating strongly from zero for W210, W210m, and W450 (Figure 4), suggesting that suboptimal selection decisions will result when using these scenarios. The MF and single-breed scenarios across traits and breeds presented lower bias values, which were $\leq$0.13, 0.18, 0.28, and 0.55 for W210, W210m, W450, and SC365, respectively (Figure 4).

Dispersion, measured as the deviation of b1 from 1, reflects the spread of predictions around their expected values. Values closer to 1 indicate a desirable spread, which was predominantly observed across traits and breeds in the MF and no-MF scenarios (Figure 4). In contrast, the Nellore-based and single-breed scenarios tended to deviate from this ideal value, suggesting either over- or under-dispersion (Figure 4). These dispersion patterns may also reflect violations of the assumption of constant correlation across the genome, which underlies traditional GBLUP models (Teng et al. 2025). If SNP effects or LD structures differ substantially in certain genomic regions between breeds, assuming homogeneous correlations could lead to suboptimal predictions (Teng et al. 2025). Future models may benefit from region-specific or trait-specific modelling of correlations, especially in multi-breed settings where shared QTL architecture is uncertain. This emphasises the importance of within-breed performance in multi-breed scenarios in yielding more consistent predictions around the desirable values. Therefore, incorporating multi-breed evaluations and MF could help to mitigate IP biases while adjusting the breed differences (Legarra et al. 2014; Luan et al. 2012).

Some studies, like Garcia et al. (2022) have shown the importance of having enough genotypes in the genomic evaluation when computing the GEBV used to backsolve SNP effects. These authors reported a significant drop in the correlation between GEBV and IP, from 0.88 to 0.82, when reducing the reference population from 10 k to 2 k animals. This finding highlights the critical role of a large and diverse population in ensuring the accuracy of IP, particularly for small breeds. In our study, the use of single-breed populations (less than 5 k animals in small breeds when combined) resulted in lower accuracies, increased bias, and greater overdispersion.

These observations align with those of Himmelbauer et al. (2024), who simulated a crossbred dairy population with ~15% missingness in the pedigrees of grandparents and great-grandparents. While they found no differences in the accuracy of predictions with or without MF, they did report improvements in dispersion (0.93 vs. 1.00) and bias (0.20 vs. 0.0 SD units), demonstrating the value of leveraging multi-breed genomic evaluations to address challenges related to incomplete or limited pedigree data.

The results of this work demonstrated that accurate IP based on a multi-breed indicine population can be obtained and are valuable for predicting the performance of young genotyped animals. This has significant implications, particularly in countries such as Brazil, where the rapid adoption of advanced tools and technologies in livestock breeding programmes has gained momentum. For smaller breeding programmes, such as those for Guzerat and Tabapua, precise IP enables faster and more accurate genetic merit estimation, even in the absence of extensive phenotypic data. By ensuring adequate representation of the breed of interest within multi-breed reference populations, breeders can achieve higher accuracies and more reliable predictions, ultimately improving genetic progress and operational efficiency.

4 | Conclusions

Incorporating multiple breeds into a large joint indicine reference population improved accuracy, bias, and dispersion of genomic IP, especially for breeds with relatively few genotyped and phenotyped individuals. Combining breeds allowed the use of larger datasets, capturing broader genetic diversity and enabling more precise estimation of marker effects, thereby increasing the accuracy of IP. While including MF in the multi-breed evaluation showed some benefits in mitigating biases and improving dispersion, accurate IP can still be achieved without MF by leveraging well-structured multi-breed datasets, with higher correlations and relationships as in our case. However, this outcome may not be generalised to all multi-breed scenarios. When young genotyped animals are excluded from the official multi-breed evaluation, robust IP can still be obtained with proper modelling and a sufficiently large and diverse reference population. However, relying on a single breed with limited data or making predictions based solely on another breed results in lower accuracy of IP, emphasising the importance of multi-breed approaches for smaller breeding programmes. These findings highlight the value of multi-breed genomic evaluations in achieving reliable genetic predictions and advancing genetic progress across breeds.

Author Contributions

M.L.-G., J.H., C.U.M., A.L., F.B., and D.L. contributed to the study's conception and design. M.L.-G. and F.B. prepared the material, data

collection, and analysis. M.L.-G. wrote the first draft of the manuscript, and the other authors contributed to later versions. All authors read and approved the final manuscript.

Ethics Statement

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data supporting the findings of this study are available from the National Association of Breeders and Researchers (ANCP). The data sets generated and analysed during the current study are available through the corresponding author upon reasonable request with the permission of the ANCP.

References

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. "Hot Topic: A Unified Approach to Utilize Phenotypic, Full Pedigree, and Genomic Information for Genetic Evaluation of Holstein Final Score." *Journal of Dairy Science* 93, no. 2: 743–752. https://doi.org/10.3168/jds.2009-2730.

Bermann, M., I. Aguilar, D. Lourenco, I. Misztal, and A. Legarra. 2023. "Reliabilities of Estimated Breeding Values in Models With Metafounders." *Genetics Selection Evolution* 55, no. 1: 6. https://doi.org/10.1186/s12711-023-00778-2.

Bermann, M., D. Lourenco, and I. Misztal. 2021. "Technical Note: Automatic Scaling in Single-Step Genomic BLUP." *Journal of Dairy Science* 104, no. 2: 2027–2031. https://doi.org/10.3168/jds.2020-18969.

Brito, F. V., J. B. Neto, M. Sargolzaei, J. A. Cobuci, and F. S. Schenkel. 2011. "Accuracy of Genomic Selection in Simulated Populations Mimicking the Extent of Linkage Disequilibrium in Beef Cattle." *BMC Genetics* 12: 80. https://doi.org/10.1186/1471-2156-12-80.

Callister, A. N., M. Bermann, S. Elms, B. P. Bradshaw, D. Lourenco, and J. T. Brawner. 2022. "Accounting for Population Structure in Genomic Predictions of *Eucalyptus globulus*." *G3: Genes, Genomes, Genetics* 12, no. 9: jkac180. https://doi.org/10.1093/g3journal/jkac180.

Christensen, O. F., P. Madsen, B. Nielsen, T. Ostersen, and G. Su. 2012. "Single-Step Methods for Genomic Evaluation in Pigs." *Animal* 6, no. 10: 1565–1571. https://doi.org/10.1017/S1751731112000742.

Clasen, J. B., W. F. Fikse, G. Su, and E. Karaman. 2023. "Multibreed Genomic Prediction Using Summary Statistics and a Breed-Origin-Of-Alleles Approach." *Heredity* 131, no. 1: 33–42. https://doi.org/10.1038/s41437-023-00619-4.

Daetwyler, H. D., R. Pong-Wong, B. Villanueva, and J. A. Woolliams. 2010. "The Impact of Genetic Architecture on Genome-Wide Evaluation Methods." *Genetics* 185, no. 3: 1021–1031. https://doi.org/10.1534/genetics.110.116855.

Garcia, A., I. Aguilar, A. Legarra, S. Tsuruta, I. Misztal, and D. Lourenco. 2022. "Theoretical Accuracy for Indirect Predictions Based on SNP Effects From Single-Step GBLUP." *Genetics Selection Evolution* 54, no. 1: 66. https://doi.org/10.1186/s12711-022-00752-4.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. "Invited Review: Genomic Selection in Dairy Cattle: Progress and Challenges." *Journal of Dairy Science* 92, no. 2: 433–443. https://doi.org/10.3168/jds.2008-1646.

Hayes, B. J., J. Copley, E. Dodd, E. M. Ross, S. Speight, and G. Fordyce. 2023. "Multi-Breed Genomic Evaluation for Tropical Beef Cattle When no Pedigree Information Is Available." *Genetics Selection Evolution* 55, no. 1: 71. https://doi.org/10.1186/s12711-023-00847-6.

Hidalgo, J., D. Lourenco, S. Tsuruta, M. Bermann, V. Breen, and I. Misztal. 2023. "Derivation of Indirect Predictions Using Genomic Recursions Across Generations in a Broiler Population." *Journal of Animal Science* 101: 355. https://doi.org/10.1093/jas/skad355.

Himmelbauer, J., H. Schwarzenbacher, C. Fuerst, and B. Fuerst-Waltl. 2024. "Exploring Unknown Parent Groups and Metafounders in Single-Step Genomic BLUP: Insights From a Simulated Cattle Population." *Journal of Dairy Science* 107: 8170–8192. https://doi.org/10.3168/jds.2024-24891.

Junqueira, V. S., P. S. Lopes, D. Lourenco, F. F. e. Silva, and F. F. Cardoso. 2020. "Applying the Metafounders Approach for Genomic Evaluation in a Multibreed Beef Cattle Population." *Frontiers in Genetics* 11: 1–10. https://doi.org/10.3389/fgene.2020.556399.

Karaman, E., G. Su, I. Croue, and M. S. Lund. 2021. "Genomic Prediction Using a Reference Population of Multiple Pure Breeds and Admixed Individuals." *Genetics Selection Evolution* 53, no. 1: 46. https://doi.org/10.1186/s12711-021-00637-y.

Legarra, A., I. Aguilar, and I. Misztal. 2009. "A Relationship Matrix Including Full Pedigree and Genomic Information." *Journal of Dairy Science* 92, no. 9: 4656–4663. https://doi.org/10.3168/jds.2009-2061.

Legarra, A., O. F. Christensen, I. Aguilar, and I. Misztal. 2014. "Single Step, a General Approach for Genomic Selection." *Livestock Science* 166: 54–65. https://doi.org/10.1016/j.livsci.2014.04.029.

Legarra, A., O. F. Christensen, Z. G. Vitezica, I. Aguilar, and I. Misztal. 2015. "Ancestral Relationships Using Metafounders: Finite Ancestral Populations and Across Population Relationships." *Genetics* 200, no. 2: 455–468. https://doi.org/10.1534/genetics.115.177014.

Legarra, A., D. Lourenco, and Z. Vitezica. 2024. "Bases for Genomic Prediction." 1–167.

Legarra, A., and A. Reverter. 2018. "Semi-Parametric Estimates of Population Accuracy and Bias of Predictions of Breeding Values and Future Phenotypes Using the LR Method." *Genetics Selection Evolution* 50, no. 1: 53. https://doi.org/10.1186/s12711-018-0426-6.

Londoño-Gil, M., D. Cardona-Cifuentes, R. Espigolan, et al. 2023. "Genomic Evaluation of Commercial Herds With Different Pedigree Structures Using the Single-Step Genomic BLUP in Nelore Cattle." *Tropical Animal Health and Production* 55, no. 2: 95. https://doi.org/10.1007/s11250-023-03508-4.

Londoño-Gil, M., R. López-Correa, I. Aguilar, et al. 2025. "Strategies for Genomic Predictions of an Indicine Multi-Breed Population Using Single-Step GBLUP." *Journal of Animal Breeding and Genetics* 142: 43–56. https://doi.org/10.1111/jbg.12882.

Lourenco, D., S. Tsuruta, Y. Masuda, M. Berman, A. Legarra, and I. Misztal. 2022. "Recent Updates in the BLUPF90 Software Suite." World Congress on Genetics Applied to Livestock Production. 3–8.

Lourenco, D. A. L., A. Legarra, S. Tsuruta, D. Moser, S. Miller, and I. Misztal. 2018. "Tuning Indirect Predictions Based on SNP Effects From Single-Step GBLUP." *Interbull Bulletin* 53: 48–53. https://journal.interbull.org/index.php/ib/article/view/1448/1511.

Lourenco, D. A. L., S. Tsuruta, B. O. Fragomeni, et al. 2015. "Genetic Evaluation Using Single-Step Genomic Best Linear Unbiased Predictor in American Angus." *Journal of Animal Science* 93, no. 6: 2653–2662. https://doi.org/10.2527/jas.2014-8836.

Luan, T., J. A. Woolliams, J. Degård, et al. 2012. "The Importance of Identity-By-State Information for the Accuracy of Genomic Selection." *Genetics Selection Evolution* 44, no. 1: 28. https://doi.org/10.1186/1297-9686-44-28.

Misztal, I., Y. Steyn, and D. A. L. Lourenco. 2022. "Genomic Evaluation With Multibreed and Crossbred Data." *JDS Communications* 3, no. 2: 156–159. https://doi.org/10.3168/jdsc.2021-0177.

Misztal, I., S. Tsuruta, D. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2014. *BLUPF90 Family of Programs*, 125. University of Georgia. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf.

Misztal, I., S. Tsuruta, D. a. L. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2015. "Manual for BLUPF90 Family of Programs." 125.

R Core Team. 2023. "R: A Language and Environment for Statistical Computing (4.3.0)." R Foundation for Statistical Computing. https://www.r-project.org.

Steyn, Y., D. A. L. Lourenco, and I. Misztal. 2019. "Genomic Predictions in Purebreds With a Multibreed Genomic Relationship Matrix1." *Journal of Animal Science* 97, no. 11: 4418–4427. https://doi.org/10.1093/jas/skz296.

Takeda, M., K. Inoue, H. Oyama, et al. 2021. "Exploring the Size of Reference Population for Expected Accuracy of Genomic Prediction Using Simulated and Real Data in Japanese Black Cattle." *BMC Genomics* 22, no. 1: 799. https://doi.org/10.1186/s12864-021-08121-z.

Teng, J., T. Zhai, X. Zhang, et al. 2025. "Improving Multi-Trait Genomic Prediction by Incorporating Local Genetic Correlations." *Communications Biology* 8, no. 1: 1–14. https://doi.org/10.1038/s42003-025-07721-9.

Tsuruta, S., D. A. L. Lourenco, Y. Masuda, T. J. Lawlor, and I. Misztal. 2021. "Reducing Computational Cost of Large-Scale Genomic Evaluation by Using Indirect Genomic Prediction." *JDS Communications* 2, no. 6: 356–360. https://doi.org/10.3168/jdsc.2021-0097.

VanRaden, P. M. 2008. "Efficient Methods to Compute Genomic Predictions." *Journal of Dairy Science* 91, no. 11: 4414–4423. https://doi.org/10.3168/jds.2007-0980.

Vitezica, Z. G., I. Aguilar, I. Misztal, and A. Legarra. 2011. "Bias in Genomic Predictions for Populations Under Selection." *Genetics Research* 93, no. 5: 357–366. https://doi.org/10.1017/S001667231100022X.

Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2012. "Genome-Wide Association Mapping Including Phenotypes From Relatives Without Genotypes." *Genetics Research* 94, no. 2: 73–83. https://doi.org/10.1017/S0016672312000274.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** jbg70008-sup-0001-DataS1.docx.