

RESSALVA

Atendendo solicitação do(a)
autor(a), o texto completo desta tese
será disponibilizado somente a partir
de 09/05/2026.

SÃO PAULO STATE UNIVERSITY
JABOTICABAL CAMPUS

**STRATEGIES FOR GENOMIC PREDICTIONS OF AN
INDICINE MULTIBREED POPULATION USING SINGLE-
STEP GBLUP**

Marisol Londono Gil

Animal Scientist

MSc in Science - Biotechnology

2025

**T
E
S
E
/
L
O
N
D
O
Ñ
O
G
I
L
M.
2
0
2
5**

SÃO PAULO STATE UNIVERSITY
JABOTICABAL CAMPUS

**STRATEGIES FOR GENOMIC PREDICTIONS OF AN
INDICINE MULTIBREED POPULATION USING SINGLE-
STEP GBLUP**

Student: Marisol Londono Gil

Advisor: Prof. Dr. Fernando Baldi

Coadvisor: Prof. Dra. Daniela Lourenco

Doctoral Thesis presented to the Animal Science Graduate Program of the Faculty of Agricultural and Veterinary Sciences – UNESP, Jaboticabal Campus, as part of the requirements for obtaining the title of Doctor in Animal Breeding and Genetics.

2025

CATALOG CARD

L847s Londoño Gil, Marisol
Strategies for genomic predictions of an indicine multibreed population using single-step GBLUP / Marisol Londoño Gil. -- Jaboticabal, 2025
136 p. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (UNESP), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal
Orientadora: Fernando Sebastian Baldi Rey
Coorientadora: Daniela Andressa Lino Lourenço

1. Accuracy. 2. Metafounders. 3. Reference population. 4. SNP effects. 5. Multi-breed evaluation. I. Título.

POTENTIAL IMPACT OF THIS RESEARCH

Brazil is home to one of the largest cattle populations in the world, with more than 238 million animals spread across the country. Among the most important zebu breeds are Nellore, Guzerat, Tabapua, and Brahman. These breeds are well adapted to tropical climates and perform well even under challenging pasture conditions. However, when it comes to applying genomic tools, that is, predicting an animal's potential based on its DNA, there is a significant imbalance. Nellore has a big amount of genomic and performance data available, while the other breeds lag far behind, limiting the accuracy and reliability of genetic evaluations for young animals that could become the next generation of sires.

This research addresses that gap by proposing a multi-breed genomic evaluation strategy. Instead of evaluating each breed separately, the study integrates data from all four breeds into a single, unified evaluation model. Using a method known as metafounders, the genetic differences between breeds are respected and accounted for, while still allowing the information to be shared across populations. This makes genetic predictions more accurate and stable, especially for underrepresented breeds like Guzerat and Tabapua.

Another major contribution of this work is the use of indirect predictions (IP), a method that estimates the genetic merit of young animals without performance or pedigree records. These predictions are calculated using information from older, well-evaluated animals and allow for early selection decisions. This is particularly valuable in commercial herds, where rapid and cost-effective decision-making is critical. IP enables producers to identify promising animals at an earlier age, reducing feeding and management costs, and accelerating the pace of genetic improvement. Moreover, the method simplifies the use of genomic selection in routine evaluations. Since IP relies on previously estimated SNP effects, there is no need to recalculate entire genomic evaluations each time a new animal is genotyped. This greatly reduces computational costs and makes genomic tools more accessible to a wider range of breeders.

The real-world impact of this research is already visible. The 2024 Sire Report published by ANCP (National Association of Breeders and Researchers) officially adopted the multi-breed evaluation strategy developed in this study. This has given breeders access to more reliable and science-based tools for selection, even for breeds that previously lacked robust genomic information. Furthermore, countries that collaborate with ANCP and work with Brahman cattle, for example, can now use indirect predictions for selection, even if they don't yet have performance data, as long as genomic data is available.

Looking ahead, this research has the potential to transform how genetic improvement is approached in tropical cattle systems worldwide. As more countries and organizations adopt multi-breed evaluations and indirect prediction strategies, we can expect a wider and fairer distribution of the benefits of genomics, particularly in low- and middle-income countries where data collection remains a challenge.

The methodology developed here also opens the door to international genetic evaluations between indicine breeds, promoting genetic exchange and collaboration across countries. This could increase the genetic diversity and resilience of cattle populations, supporting long-term sustainability in the face of climate change and evolving market demands.

Additionally, by enabling faster and earlier selection, this approach can significantly reduce the generation interval, making it possible to achieve genetic gains more quickly and efficiently. This is a game changer not just for elite breeding programs, but also for small and medium producers who can now access high-impact genetic tools without needing complex infrastructure.

Ultimately, this research provides a scalable and inclusive framework for implementing genomic selection in real-world settings. As technology becomes more accessible and costs continue to drop, the strategies proposed here can serve as a model for other livestock species and agricultural sectors aiming to increase productivity, sustainability, and food security through science-based innovation.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: STRATEGIES FOR GENOMIC PREDICTIONS OF AN INDICINE MULTIBREED POPULATION USING SINGLE-STEP GBLUP

AUTORA: MARISOL LONDOÑO GIL

ORIENTADOR: FERNANDO SEBASTIAN BALDI REY

COORIENTADORA: DANIELA ANDRESSA LINO LOURENÇO

Aprovada como parte das exigências para obtenção do Título de Doutora em Genética e Melhoramento Animal, pela Comissão Examinadora:

Prof. Dr. FERNANDO SEBASTIAN BALDI REY (Participação Virtual)
Departamento de Zootecnia / FCAV UNESP Jaboticabal

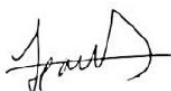
Documento assinado digitalmente
 **FERNANDO SEBASTIAN BALDI REY**
Data: 12/05/2025 06:22:37-0300
Verifique em <https://validar.iti.gov.br>



Prof. Dr. JORGE ANGEL HIDALGO MORENO (Participação Virtual)
Department of Animal and Dairy Science / University of Georgia (UGA) - Athens/GA/EUA

Prof. Dr. DANISIO PRADO MUNARI (Participação Virtual)
Departamento de Ciencias Exatas / FCAV UNESP Jaboticabal

Documento assinado digitalmente
 **DANISIO PRADO MUNARI**
Data: 09/05/2025 20:15:37-0300
Verifique em <https://validar.iti.gov.br>



Pesquisador Dr. IGNACIO AGUILAR (Participação Virtual)
Instituto Nacional de Investigación Agropecuaria (INIA) / Tacuarembó/Uruguay



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal



Pós-doutorando LÚCIO FLÁVIO MACÊDO MOTA (Participação Virtual)
Departamento de Zootecnia / FCAV UNESP Jaboticabal



Documento assinado digitalmente

LUCIO FLAVIO MACEDO MOTA

Data: 10/05/2025 14:03:01-0300

Verifique em <https://validar.itl.gov.br>

Jaboticabal, 09 de maio de 2025

AUTHOR'S CURRICULAR DATA

Marisol Londoño Gil was born on August 31, 1993, in Entreríos, Antioquia, Colombia, daughter of Francisco Rodrigo Londoño Londoño and Luz Marina de Jesús Gil Lopera. She earned her bachelor's degree in Animal Science from the Universidad Nacional de Colombia, Medellín campus, in May 2016. During her undergraduate studies, she gained practical experience working as an Animal Scientist at Ganadería Rusticana from January 2016 to January 2018. In February 2018, she began her master's degree in Science – Biotechnology at the same institution, under the supervision of Prof. Dr. Luis Gabriel González Herrera and Prof. Dr. Albeiro López Herrera. She completed her master's in June 2020, earning meritorious distinction for her dissertation.

In August 2020, she started her Ph.D. in Animal Breeding and Genetics at São Paulo State University “Júlio de Mesquita Filho” (UNESP), under the supervision of Prof. Dr. Fernando Sebastián Baldi Rey and Prof. Dr. Daniela Andressa Lino Lourenço. Her doctoral studies were initially funded by a CAPES scholarship, later replaced by a FAPESP scholarship (FAPESP process 2020/14275-5). As part of her Ph.D. research, she conducted an internship at the University of Georgia (UGA) in the United States in 2023, supported by a FAPESP BEPE scholarship (FAPESP process 2022/15168-3), under the guidance of Prof. Dr. Daniela Andressa Lino Lourenço. Throughout her doctoral studies, she actively participated in various research projects, conferences, courses, and symposiums, further strengthening her expertise in animal breeding and genetics.

“Remember to look up at the stars and not down at your feet. Try to make sense of what you see and wonder about what makes the universe exist. Be curious. And however difficult life may seem, there is always something you can do and succeed at.

It matters that you don't just give up”.

Stephen Hawking

With heartfelt gratitude, I dedicate this work to my unwavering pillars of strength, my beloved parents, Luz Marina Gil Lopera and Rodrigo Londoño Londoño. Your enduring support has illuminated even the darkest paths of my journey.

ACKNOWLEDGMENTS

To the graduate program in Animal Breeding and Genetics at the Faculty of Agricultural and Veterinary Sciences of the Sao Paulo State University (FCAV/UNESP) for the opportunity to complete this stage of my life.

To the São Paulo Research Foundation FAPESP for the scholarships during my doctoral degree (grant # 2020/14275-5) and my internship abroad (grant # 2022/15168-3) which allowed me great professional growth. And also, the grant that allowed to have more genotyped animals for the analysis (grant # 2021/04807-2).

To the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) which provided funds to have more data for the research (grant # 402313/2021-6).

To the National Association of Breeders and Researchers (ANCP), for providing the data needed for this research.

To the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior -Brazil (CAPES – PROEX) who partially funded my studies at the Sao Paulo State University.

To my advisor Prof. Dr. Fernando Baldi for the time, trust, patience and advice and his example of a professional who encouraged me to pursue my goals every day during my learning and PhD process.

To my co-advisor Prof. Dr. Daniela Lourenco for accepting me to do my internship in the Department of Animal and Dairy Science at University of Georgia, teaching and giving me so many opportunities that help me to achieve my doctoral degree.

To the University of Georgia for accepting me as Research Visitor for 15 months.

To the members of the thesis defense committee Danisio Munari, Jorge Hidalgo, Lúcio Mota, and Ignacio Aguilar, Fernando Baldi, thank you for your valuable contributions, for helping me improve my thesis, and for granting me the approval of my Ph.D.

To my friends and colleges at Sao Paulo State University, such as Ana Maria Fernanda Ramirez, Daniel Cardona, Joao Barbosa, Roney Teixeira, Gabriel Gubiani, and Larissa Bordin Temp, who were really important and helpful during my doctorate journey.

To my colleges at the University of Georgia, such as Zuleica Trujano, Arielly Garcia, Sergio Sanchez, Fernando Bussman and Jorge Hidalgo, whom I'm really grateful by their support during my internship and my doctoral projects.

To my family, especially my mom, Luz Marina Gil Lopera; my dad, Rodrigo Londoño Londoño; my brother, Andrés Londoño Gil; my sisters, Natalia Andrea Londoño Gil and María Londoño Gil; my nephew, Jerónimo Rodrigues Londoño; and my nieces, Juliana Rodrigues Londoño and Hannah Cabrera Londoño, whom I love with all my heart and who are the reason I'm here today.

SUMMARY

CHAPTER 1: GENERAL CONSIDERATIONS	1
1.1 INTRODUCTION	1
1.2 OBJECTIVES	4
1.2.1 General objective.....	4
1.2.2 Specific objectives	4
1.3 LITERATURE REVIEW	5
1.3.1 Beef production and the role of Indicine breeds in Brazil.....	5
1.3.2 History of genetic evaluations.....	8
1.3.3 Single-step genomic best linear unbiased predictor (ssGBLUP)	10
1.3.4 Genomic evaluation challenges in small and multi-breed populations.....	11
1.3.5 Strategies used in multi-breed genomic predictions	13
1.3.6. Accuracy of genomic predictions and IP:.....	18
CHAPTER 2: STRATEGIES FOR GENOMIC PREDICTIONS OF AN INDICINE MULTI-BREED POPULATION USING SINGLE-STEP GBLUP.....	30
2.1 INTRODUCTION	31
2.2 MATERIAL AND METHODS	33
2.2.1 Data.....	33
2.2.2 Variance components and heritabilities estimation.....	35
2.2.3 Scenarios for the estimation of genomic breeding values	35
2.2.4 Comparison of scenarios.....	39
2.3 RESULTS AND DISCUSSION	40
2.3.1 Relationships within and between Metafounders	41
2.3.2 Comparison of scenarios.....	43
2.3.3 Compatibility between G and A₂₂	48
2.4. CONCLUSIONS	52
2.5. REFERENCES	52
CHAPTER 3 - INDIRECT GENOMIC PREDICTIONS FOR INDICINE CATTLE BREEDS WITH SNP EFFECTS FROM A MULTI-BREED GENOMIC EVALUATION	57

3.1 INTRODUCTION	58
3.2 MATERIAL AND METHODS	60
3.2.1 Data	60
3.2.2 Analyses	62
3.2.3 Validation	65
3.3 RESULTS AND DISCUSSION	67
3.3.1 SNP Effects	67
3.3.2 Evaluation of Indirect Predictions	73
3.4 CONCLUSIONS	78
3.5 REFERENCES	78
 CHAPTER 4 - ACCURACY OF INDIRECT PREDICTIONS FOR INDICINE CATTLE BREEDS BASED ON A MULTI-BREED GENOMIC EVALUATION	 83
4.1 INTRODUCTION	85
4.2 MATERIAL AND METHODS	86
4.2.1 Data	86
4.2.2 Scenarios	87
4.2.3 Benchmark GEBV and accuracy	89
4.2.4 Indirect predictions and accuracy	90
4.2.5 Validation of IP accuracies	91
4.3 RESULTS AND DISCUSSION	92
4.4 CONCLUSION	100
4.5 REFERENCES	101
 CHAPTER 5 – FINAL CONSIDERATIONS	 105
5.1 KEY FINDINGS	105
5.2 IMPLICATIONS FOR BREEDING PROGRAMS	106
5.3 FUTURE RESEARCH DIRECTIONS	107
5.4 CONCLUSION	107
 APPENDIX	 109

STRATEGIES FOR GENOMIC PREDICTIONS OF AN INDICINE MULTIBREED POPULATION USING SINGLE-STEP GBLUP

ABSTRACT – Genomic selection has changed animal breeding by enabling the early identification of superior individuals based on their genetic merit. However, this approach faces considerable challenges when applied to indicine cattle, particularly populations such as Nellore, Brahman, Guzerat, and Tabapua. Brahman, Guzerat, and Tabapua present small reference populations, making genomic selection difficult and inaccurate. A multi-breed evaluation could solve this problem, but it must consider the heterogeneity in allele frequencies among breeds and the incompatibility between genomic (**G**) and pedigree (**A**) relationship matrices. This work investigated strategies, feasibility, and effectiveness for genomic prediction in indicine cattle by focusing on a multi-breed population that includes Nellore, Brahman, Guzerat, and Tabapua using the single-step genomic best linear unbiased predictor (ssGBLUP). Data for approximately 4.2 M pedigree records, 329 K phenotypic observations, and 63 K genotyped animals were used. The traits analyzed were adjusted weight at 210 days (W210), adjusted weight at 450 days (W450), and scrotal circumference at 365 days (SC365). Different scenarios were evaluated, comparing single-breed and multi-breed approaches. Multi-breed evaluations were conducted both with and without the inclusion of metafounders (MF) to account for genetic differences across breeds. At the same time, without MF, the inclusion of breed-specific allele frequencies was analyzed or not adjusted at all. Comparative analyses of the different scenarios were performed using the linear regression (LR) method to assess bias, dispersion, and accuracy. In addition to improving direct genomic evaluations, this work addressed the challenge of obtaining genomic predictions for young, genotyped animals lacking phenotypic data or pedigree information by computing indirect predictions (IP). These IP were derived as the sum of SNP effects weighted by gene content. The evaluation of the different scenarios was made using the LR method. Moreover, the accuracy of IP (ACC_{IP}) was estimated from the prediction error covariance (PEC) of SNP effects and validated through correlations with the accuracy of genomic estimated breeding values (ACC_{GEBV}). The results of this work showed that multi-breed evaluations significantly improved prediction accuracy compared to single-breed approaches,

particularly helping breeds with smaller genomic datasets such as Guzerat, and Tabapua. Incorporating MF reduced bias and mitigated under- or over-dispersion in the predictions, with a modest increase in IP accuracy. The single-breed evaluations frequently showed lower accuracy, especially for Guzerat and Tabapua which have limited genomic data. The study demonstrated that structured multi-breed genomic evaluations are a robust and effective strategy for improving genomic selection in Brazilian indicine cattle. While using genomic information across multiple breeds, this approach improved the accuracy of predictions for traits of economic importance and allowed rapid IP for young animals. Consequently, Brazilian indicine multi-breed evaluations can help selection decisions and potentially accelerate genetic progress, especially in breeds with constrained reference populations.

Keywords: Accuracy, allele frequencies, Brahman, genomic selection, Guzerat, metafounders, multi-breed evaluation, Nellore, reference population, SNP effects, Tabapua.

ESTRATÉGIAS PARA PREDIÇÃO GENÔMICA EM POPULAÇÕES MULTIRRACIAIS DE ZEBUÍNOS APLICANDO O MÉTODO DO PASSO ÚNICO GENÔMICO BLUP

RESUMO – A seleção genômica transformou a seleção de animais ao permitir a identificação precoce de indivíduos superiores com base em seu mérito genético. Contudo, essa abordagem enfrenta desafios consideráveis quando aplicada ao gado zebuino, especialmente em populações como Nelore, Brahman, Guzerá e Tabapuã. Brahman, Guzerá e Tabapuã apresentam populações de referência pequenas, o que torna a sua seleção genômica difícil e imprecisa. Uma avaliação multirracial poderia resolver esse problema, mas deve considerar a heterogeneidade nas frequências alélicas que podem existir entre as raças e a incompatibilidade entre as matrizes de relacionamento genômico (**G**) e de pedigree (**A**). Este trabalho pesquisou estratégias para as predições genômicas de gado zebuino, em uma população multirracial que incluiu o Nelore, Brahman, Guzerá e Tabapuã, utilizando o preditor linear não viesado genômico de uma única etapa (ssGBLUP). Foram utilizados dados de aproximadamente 4,2 milhões de registros de pedigree, 329 mil observações fenotípicas e 63 mil animais genotipados. As características analisadas foram o peso ajustado aos 210 dias (W210), o peso ajustado aos 450 dias (W450) e a circunferência escrotal aos 365 dias (SC365). Diferentes cenários foram avaliados, comparando abordagens de raça única e multirracial. As avaliações multirraciais foram conduzidas tanto com a inclusão de metafundadores (MF), para levar em conta as diferenças e semelhanças genéticas entre as raças, quanto sem eles. Sem MF, foi analisada o ajuste ou não de frequências alélicas específicas de cada raça. Análises comparativas dos diferentes cenários foram realizadas utilizando o método de regressão linear (LR) para avaliar viés, dispersão e acurácia. Além das avaliações genômicas diretas, este trabalho abordou o desafio de obter predições genômicas para animais jovens genotipados que carecem fenótipos ou pedigree, calculando predições indiretas (IP). Essas IP foram derivadas como a soma dos efeitos dos SNPs ponderados pelo conteúdo gênico. A avaliação dos diferentes cenários foi realizada utilizando o método LR. Ademais, a acurácia das IP (ACC_{IP}) foi estimada a partir da covariância do erro de predição (PEC) dos efeitos dos SNPs e validada através de correlações com a acurácia dos valores genômicos estimados (ACC_{GEBV}). Os resultados deste trabalho

demonstraram que as avaliações multirraciais melhoraram significativamente a precisão das predições em comparação com as abordagens de raça única, beneficiando particularmente as raças com conjuntos de dados genômicos menores, como Guzerá e Tabapuã. A inclusão de MF reduziu o viés e mitigou a sub ou superdispersão nas predições, com um modesto aumento na acurácia das IP. As avaliações de raça única frequentemente apresentaram menor precisão, especialmente para Guzerá e Tabapuã, que possuem dados genômicos limitados. O estudo demonstrou que as avaliações genômicas multirraciais estruturadas são uma estratégia robusta e eficaz para melhorar a seleção genômica no gado zebuíno brasileiro. Ao utilizar informações genômicas de múltiplas raças, essa abordagem melhorou a precisão das predições para características de importância econômica e permitiu a rápida obtenção de IP para animais jovens. Consequentemente, as avaliações multirraciais de gado zebuíno brasileiro podem auxiliar nas decisões de seleção e potencialmente acelerar o progresso genético, especialmente em raças com populações de referência pequenas.

Palavras-chave: Acurácia, avaliação multirracial, Brahman, efeitos dos SNPs, frequências alélicas, Guzerá, metafundadores, Nelore, população de referência, seleção genômica, Tabapuã.

LISTS OF ABBREVIATIONS

A: Pedigree-based relationship matrix

ACC_{GEBV} : Accuracy of GEBV

ACC_{IP} : Accuracy of IP

ANCP: National Association of Breeders and Researchers *Associação Nacional de Criadores e Pesquisadores*

BLUP: Best linear unbiased prediction

C: Coefficient matrix of the mixed model equations

EBV: Estimated breeding value

G: Genomic relationship matrix.

GBLUP: Genomic best linear unbiased prediction

GEBV: Genomic estimated breeding value

H: Combined genomic and pedigree-based relationship matrix

IBGE: Brazilian Institute of Geography and Statistics

IP: Indirect predictions

LD: Linkage disequilibrium

LHS: Left hand side

LR: Linear regression

MAF: Minor allele frequency

MAS: Marker assisted selection

MF: Metafounders

MME: Mixed model equations

N: Number of records

PCA: Principal component analysis

PEC: Prediction error covariance

PEV: Prediction error variance

QTL: Quantitative trait loci

SC365: Fcrotal circumference at 365 days

SD: standard deviation

SNP: Single nucleotide polymorphism

ssGBLUP: Single step genomic best linear unbiased prediction

SVD: Singular value decomposition

UPG: Unknown parent groups

USA: United States of America

USDA: United States Department of Agriculture

W210: Adjusted weight at 210 days

W450: Adjusted weight at 450 days

CHAPTER 1: GENERAL CONSIDERATIONS

1.1 INTRODUCTION

Indicine Beef cattle breeds (*Bos taurus indicus*) are socially and economically important worldwide, dominating meat production systems in tropical and subtropical regions. They serve as a crucial resource for food security in communities within developing countries (Thornton, 2010). The dominance of Zebu breeds is mainly due to their superior ability to tolerate high temperatures and humidity, which are common in these regions. Additionally, these cattle exhibit greater resistance to ticks and other environmental challenges than taurine breeds (*Bos taurus taurus*), typically found in temperate regions (Utsunomiya et al., 2019).

According to the United States Department of Agriculture (USDA, 2024), Brazil is the leading beef exporter and producer worldwide, with production systems predominantly relying on indicine breeds (*Bos taurus indicus*) adapted to tropical conditions. The Nellore breed, which accounts for around 80% of Brazil's approximately 200 million cattle, is the most prominent, followed by Brahman, Guzerat, and Tabapua (Cunha et al., 2023; de Souza et al., 2021). Most technological developments and applications in genomic evaluations have focused on Nellore. Despite these efforts, the other breeds are still in the early stages of genomic evaluations, limiting their potential gains from genomic selection, mainly due to the small size of their reference populations.

A key challenge in these breeds is the limited size of their reference populations, markedly when pedigree, phenotypic, and genomic data are scarce. This limitation makes it difficult to develop a sufficiently large reference population with high accuracy. One strategy to improve genomic evaluations is to combine these breeds into a metapopulation, which may increase the predictive power without additional data recording costs (Steyn et al., 2019). However, integrating multiple breeds may introduce compatibility issues between genomic (**G**) and pedigree (**A**) relationship matrices, as allele frequencies used for scaling **G** are typically estimated based on genotyped individuals across all populations.

To mitigate these compatibility issues in multi-breed evaluations, Legarra et al. (2015) proposed the use of metafounders (MF) in single-step genomic best linear unbiased prediction (ssGBLUP). MF are pseudo-individuals that act as founders of base populations, which can be considered simultaneously as the sire and dam of the base population. This base population becomes related and can be inbred (Legarra et al., 2015), ensuring compatibility between **G** and **A** matrices in genomic evaluations. Therefore, this methodology has been recommended when combining genomic and pedigree information for individuals, as in ssGBLUP, and when several base populations are combined (Legarra et al., 2015), as in the multi-breed evaluations.

Another strategy for improving the compatibility is to include breed-specific allele frequencies when constructing **G**. This is important because genomic regions inherited from different breeds may show unique allele frequencies (Lourenco et al., 2016). The genome of multi-breed populations combines genomic regions inherited from their parental breeds. Thus, the frequency and effects of single nucleotide polymorphism (SNP) alleles may vary significantly across different breeds, depending on their genetic diversity and history (Sevillano et al., 2019). By considering these differences, it may be possible to enhance the accuracy of genomic predictions in multi-breed populations.

While incorporating multi-breed evaluations, MF, and breed-specific allele frequencies can improve compatibility and the accuracy of genomic evaluations for small breeds, a significant challenge remains in accurately predicting young animals. For closely related breeds such as indicine cattle, across-breed genomic predictions are expected to be highly accurate (Steyn et al., 2019), meaning that merging them in a single evaluation should not compromise prediction power. However, young animals without phenotypes or progeny data are often excluded from official evaluations (Tsuruta et al., 2021), despite many commercial farms increasingly genotyping them for herd management. Including these animals in the evaluation could reduce the accuracy of genomic breeding values (GEBVs) and may introduce bias (Lourenco et al., 2018), highlighting the growing need for reliable indirect predictions (IP) in these individuals (Londoño-Gil et al., 2023; Tsuruta et al., 2021).

IP can help identify superior young candidates early, especially in regions with

limited genomic evaluations. They are calculated as the sum of SNP effects weighted by gene content and can be derived from GEBVs. In genomic BLUP (GBLUP) or ssGBLUP, SNP effects are not directly available but can be backsolved from GEBVs using formulas (Strandén & Garrick, 2009; VanRaden, 2008). Once SNP effects are computed, IP for young animals are obtained as described before. This approach is commonly used in dairy cattle genomic predictions to release fast interim predictions based on IP, though SNP effects are typically estimated through multi-step procedures.

The accuracy of predicted breeding values, and therefore IP, is also important for better selection decisions. Henderson (1984) showed that accuracies of EBVs can be calculated based on the prediction error variance (PEV) by inverting the coefficient matrix of the BLUP mixed model equations (MME). Similarly, Strandén & Christensen (2011) and Tier et al. (2018) showed how to calculate the accuracy of IP using the SNP prediction error covariance (PEC). Calculating PEC for SNP effects in ssGBLUP requires considering the number of genotyped animals (Gualdrón Duarte et al., 2014). In principle, the exact computation of SNP PEC requires obtaining the full PEC matrix for all genotyped animals from the inverse of the MME (Garcia et al., 2022). Thus, the accuracy of IP is also feasible to obtain in the multi-breed scenario.

In the context of multi-breed evaluations, particularly relevant for Brazil's livestock sector, improving the accuracy of genomic predictions is essential. Evaluating multiple breeds together may increase the precision of genetic predictions, providing more accurate estimates for selection decisions. Precise genomic evaluations improve genomic breeding values (GEBVs) and may provide more accurate indirect predictions (IP), which is important to identify young superior animals earlier, especially in regions with less established genomic infrastructure. The different strategies to do so, therefore, help to increase evaluation accuracies, making these methods important for future breeding decisions and genetic progress for these populations, especially the ones of small size.

4.4 CONCLUSION

This study demonstrated that incorporating MF in multi-breed genomic evaluations significantly improved the accuracy of IP in indicine cattle breeds with limited reference populations. This scenario provided the highest accuracy estimates, particularly benefiting smaller breeds like Guzerat and Tabapua by using the larger

reference population of Nellore. However, some bias remains in the accuracy estimation of IP, which requires further investigation.

Additionally, using SNP-derived PEC estimates to compute IP accuracy proved to be a reliable approach, reinforcing the importance of multi-breed evaluations in genomic selection. These results suggest that indirect predictions can be a valuable tool for genomic selection when appropriate methodologies, such as multi-breed ssGBLUP with MF, are applied. Integrating these strategies into breeding programs can enhance selection efficiency, reduce uncertainty, and accelerate genetic progress in the livestock industry.

4.5 REFERENCES

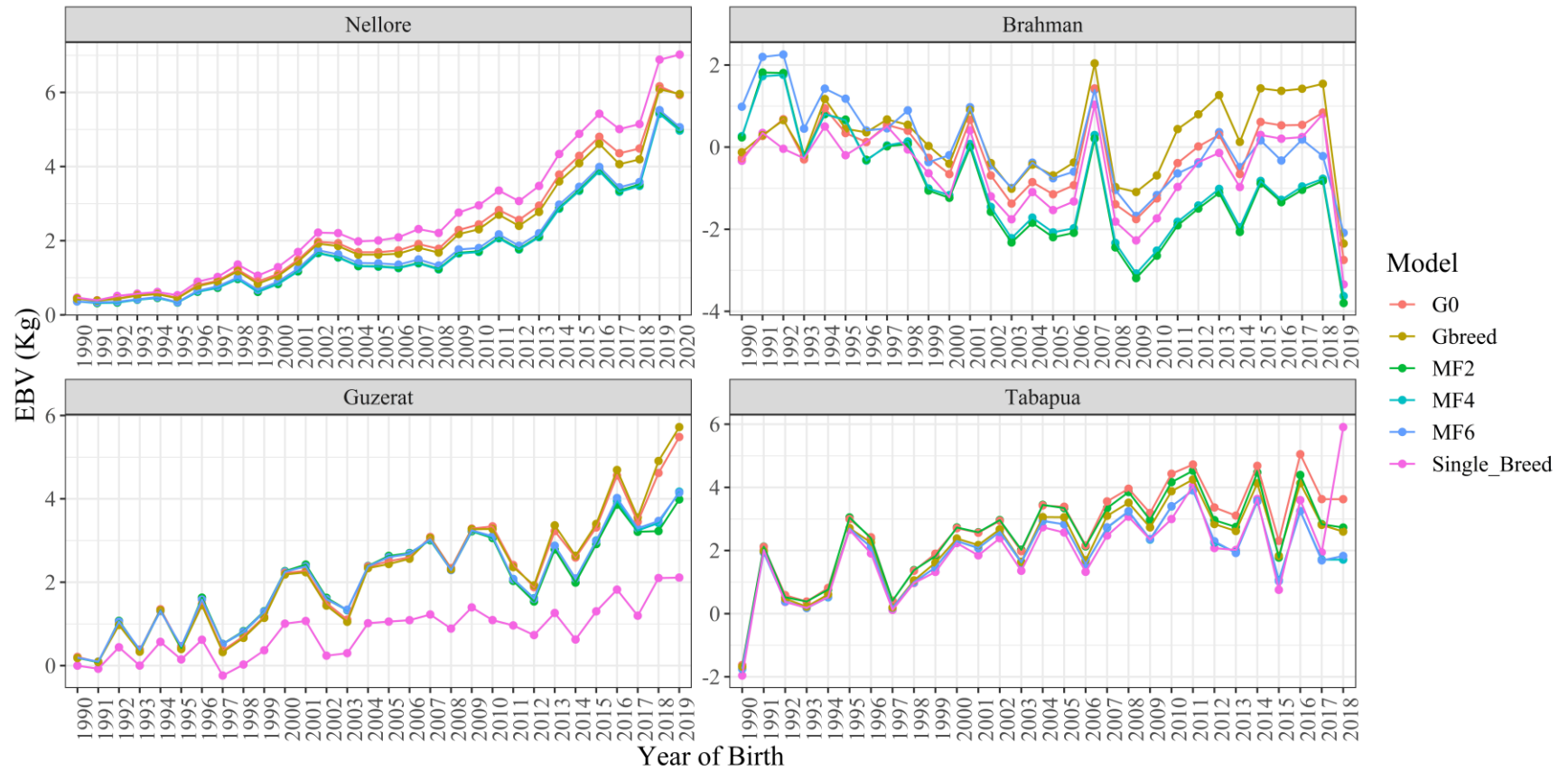
- Aguilar, I., Fernandez, E. N., Blasco, A., Ravagnolo, O., & Legarra, A. (2020). Effects of ignoring inbreeding in model-based accuracy for BLUP and SSGBLUP. **Journal of Animal Breeding and Genetics**, 137(4), 356–364. <https://doi.org/10.1111/jbg.12470>
- Aguilar, I., Legarra, A., Cardoso, F., Masuda, Y., Lourenco, D., & Misztal, I. (2019). Frequentist p-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle. **Genetics Selection Evolution**, 51(1). <https://doi.org/10.1186/s12711-019-0469-3>
- Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., & Lawlor, T. J. (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(2), 743–752. <https://doi.org/10.3168/jds.2009-2730>
- Anglhuber, C., Edel, C., Pimentel, E. C. G., Emmerling, R., Götz, K. U., & Thaller, G. (2024). Definition of metafounders based on population structure analysis. **Genetics Selection Evolution**, 56(1). <https://doi.org/10.1186/s12711-024-00913-7>
- Bijma, P. (2012). Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations. **Journal of Animal Breeding and Genetics**, 129(5), 345–358. <https://doi.org/10.1111/j.1439-0388.2012.00991.x>
- Brito, L. F., Clarke, S. M., McEwan, J. C., Miller, S. P., Pickering, N. K., Bain, W. E., Dodds, K. G., Sargolzaei, M., & Schenkel, F. S. (2017). Prediction of genomic breeding values for growth, carcass and meat quality traits in a multi-breed sheep population using a HD SNP chip. **BMC Genetics**, 18(1). <https://doi.org/10.1186/s12863-017-0476-8>
- Christensen, O. F., Madsen, P., Nielsen, B., Ostersen, T., & Su, G. (2012). Single-step methods for genomic evaluation in pigs. **Animal**, 6(10), 1565–1571. <https://doi.org/10.1017/S1751731112000742>

- Cole, J. B., Dürr, J. W., & Nicolazzi, E. L. (2021). Invited review: The future of selection decisions and breeding programs: What are we breeding for, and who decides? **Journal of Dairy Science**, 104(5), 5111–5124. <https://doi.org/10.3168/jds.2020-19777>
- Dekkers, J. C. M., Su, H., & Cheng, J. (2021). Predicting the accuracy of genomic predictions. **Genetics Selection Evolution**, 53(1). <https://doi.org/10.1186/s12711-021-00647-w>
- Garcia, A., Aguilar, I., Legarra, A., Tsuruta, S., Misztal, I., & Lourenco, D. (2022). Theoretical accuracy for indirect predictions based on SNP effects from single-step GBLUP. **Genetics Selection Evolution**, 54(1), 66. <https://doi.org/10.1186/s12711-022-00752-4>
- Gorjanc, G., Bijma, P., & Hickey, J. M. (2015). Reliability of pedigree-based and genomic evaluations in selected populations. **Genetics Selection Evolution**, 47(1). <https://doi.org/10.1186/s12711-015-0145-1>
- Gualdrón Duarte, J. L., Cantet, R. J., Bates, R. O., Ernst, C. W., Raney, N. E., & Steibel, J. P. (2014). Rapid screening for phenotype-genotype associations by linear transformations of genomic evaluations. **BMC Bioinformatics**, 15(1), 246. <https://doi.org/10.1186/1471-2105-15-246>
- Hayes, B. J., Bowman, P. J., Chamberlain, A. C., Verbyla, K., & Goddard, M. E. (2009). Accuracy of genomic breeding values in multi-breed dairy cattle populations. **Genetics Selection Evolution**, 41(1), 1–9. <https://doi.org/10.1186/1297-9686-41-51>
- Henderson, C. R. (1984). **Applications of Linear Models in Animal Breeding** (1st ed.). Guelph University.
- Jin, L., Xu, L., Jin, H., Zhao, S., Jia, Y., Li, J., & Hua, J. (2024). Accuracy of Genomic Predictions Cross Populations with Different Linkage Disequilibrium Patterns. **Genes**, 15(11). <https://doi.org/10.3390/genes15111419>
- Khansefid, M., Goddard, M. E., Haile-Mariam, M., Konstantinov, K. V., Schrooten, C., de Jong, G., Jewell, E. G., O'Connor, E., Pryce, J. E., Daetwyler, H. D., & MacLeod, I. M. (2020). Improving Genomic Prediction of Crossbred and Purebred Dairy Cattle. **Frontiers in Genetics**, 11. <https://doi.org/10.3389/fgene.2020.598580>
- Legarra, A., Aguilar, I., & Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. **Journal of Dairy Science**, 92(9), 4656–4663. <https://doi.org/10.3168/jds.2009-2061>
- Legarra, A., Lourenco, D., & Vitezica, Z. (2024). **Bases for Genomic Prediction** (pp. 1–167).
- Li, L., Gurman, P. M., Swan, A. A., & Tier, B. (2023). Approximating prediction error variances and accuracies of estimated breeding values from a SNP–BLUP model for genotyped individuals. **Animal Production Science**, 63(11), 1086–1094. <https://doi.org/10.1071/an23027>
- Londoño-Gil, M., López-Correa, R., Aguilar, I., Magnabosco, C. U., Hidalgo, J., Bussiman, F., Baldi, F., & Lourenco, D. (2024). Strategies for genomic

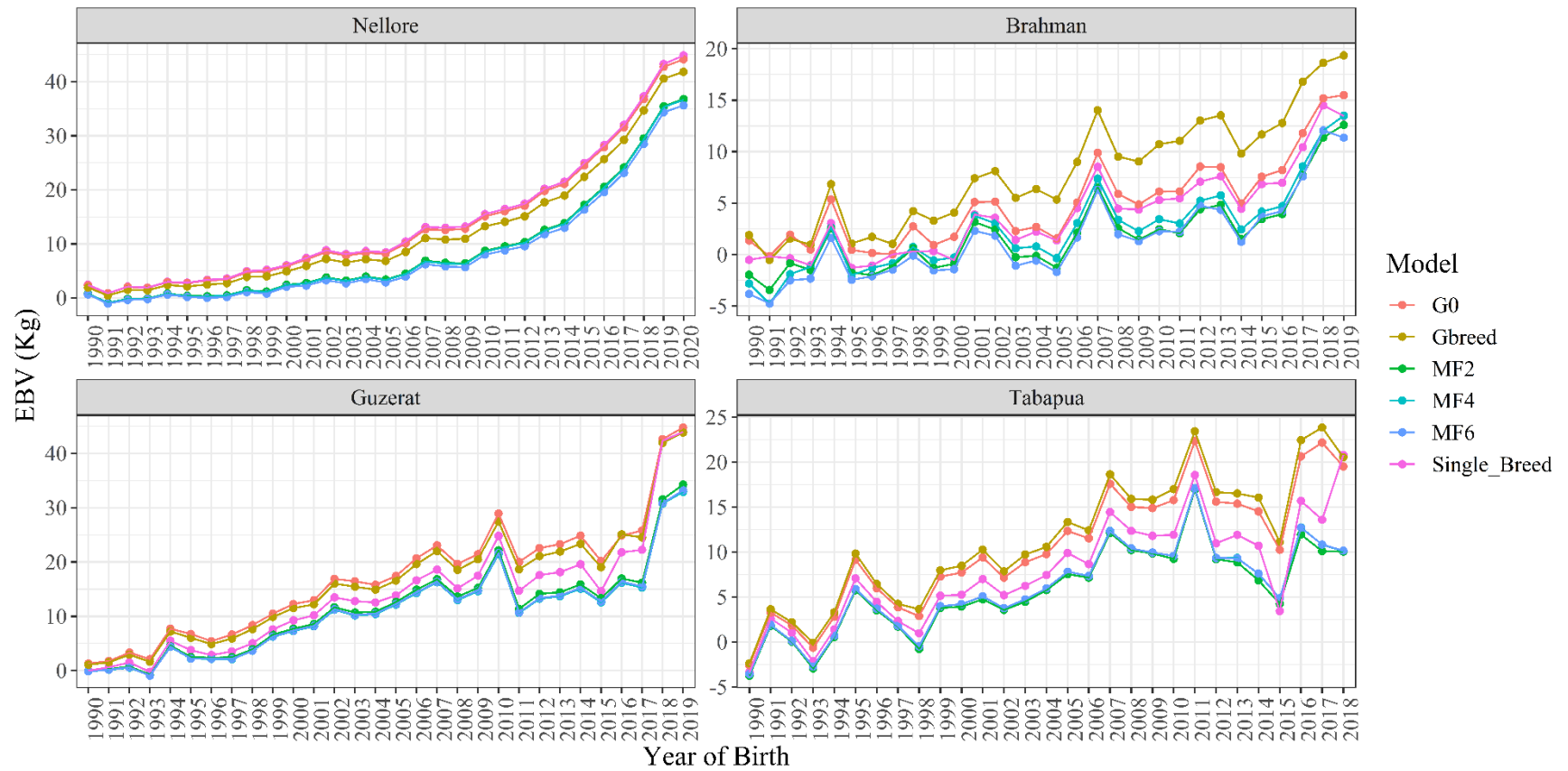
- predictions of an indicine multi-breed population using single-step GBLUP. **Journal of Animal Breeding and Genetics**. <https://doi.org/10.1111/jbg.12882>
- Lourenco, D. A. L., Legarra, A., Tsuruta, S., Moser, D., Miller, S., & Misztal, I. (2018). Tuning Indirect Predictions Based on SNP Effects from Single-Step GBLUP. **Interbull Bulletin**, 53, 48–53. <https://journal.interbull.org/index.php/ib/article/view/1448/1511>
- Lourenco, D., Legarra, A., Tsuruta, S., Moser, D., Miller, S., & Misztal, I. (2018). Tuning indirect predictions based on SNP effects from ssGBLUP.
- Lourenco, D., Tsuruta, S., Masuda, Y., Berman, M., Legarra, A., & Misztal, I. (2022, July 3). Recent updates in the BLUPF90 software suite. **World Congress on Genetics Applied to Livestock Production**: 3 - 8 July.
- Misztal, I., Tsuruta, S., Lourenco, D. a L., Aguilar, I., Legarra, A., & Vitezica, Z. (2015). **Manual for BLUPF90 family of programs**. 125.
- Misztal, I., Tsuruta, S., Lourenco, D., Aguilar, I., Legarra, A., & Vitezica, Z. (2014). **BLUPF90 family of programs**. In Athens: University of Georgia (p. 125). http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf
- Mohammad Rahimi, S., Rashidi, A., & Esfandyari, H. (2020). Accounting for differences in linkage disequilibrium in multi-breed genomic prediction. **Livestock Science**, 240. <https://doi.org/10.1016/j.livsci.2020.104165>
- Strandén, I., & Christensen, O. F. (2011). Allele coding in genomic evaluation. **Genetics Selection Evolution**, 43(1). <https://doi.org/10.1186/1297-9686-43-25>
- Strandén, I., & Garrick, D. J. (2009). Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. **Journal of Dairy Science**, 92(6), 2971–2975. <https://doi.org/10.3168/jds.2008-1929>
- Takeda, M., Inoue, K., Oyama, H., Uchiyama, K., Yoshinari, K., Sasago, N., Kojima, T., Kashima, M., Suzuki, H., Kamata, T., Kumagai, M., Takasugi, W., Aonuma, T., Soma, Y., Konno, S., Saito, T., Ishida, M., Muraki, E., Inoue, Y., ... Uemoto, Y. (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(1). <https://doi.org/10.1186/s12864-021-08121-z>
- Tier, B., Meyer, K., & Swan, A. A. (2018). On implied genetic effects, relationships and alternate allele coding. **Proceedings of the World Congress on Genetics Applied to Livestock Production**, 11, 127.
- van den Berg, I., Meuwissen, T. H. E., MacLeod, I. M., & Goddard, M. E. (2019). Predicting the effect of reference population on the accuracy of within, across, and multibreed genomic prediction. **Journal of Dairy Science**, 102(4), 3155–3174. <https://doi.org/10.3168/jds.2018-15231>
- Van Vleck, L. D. (1993). **Selection index and introduction to mixed model methods for genetic improvement of animals: The green book** (SF105.26.V). CRC Press, Inc.
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. **Journal of Dairy Science**, 91(11), 4414–4423. <https://doi.org/10.3168/jds.2007-0980>

- Wang, H., Misztal, I., Aguilar, I., Legarra, A., & Muir, W. M. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research**, 94(2), 73–83. <https://doi.org/10.1017/S0016672312000274>
- Xu, L., Wang, Z., Zhu, B., Liu, Y., Li, H., Bordbar, F., Chen, Y., Zhang, L., Gao, X., Gao, H., Zhang, S., Xu, L., & Li, J. (2019). Theoretical evaluation of multi-breed genomic prediction in Chinese indigenous cattle. **Animals**, 9(10). <https://doi.org/10.3390/ani9100789>

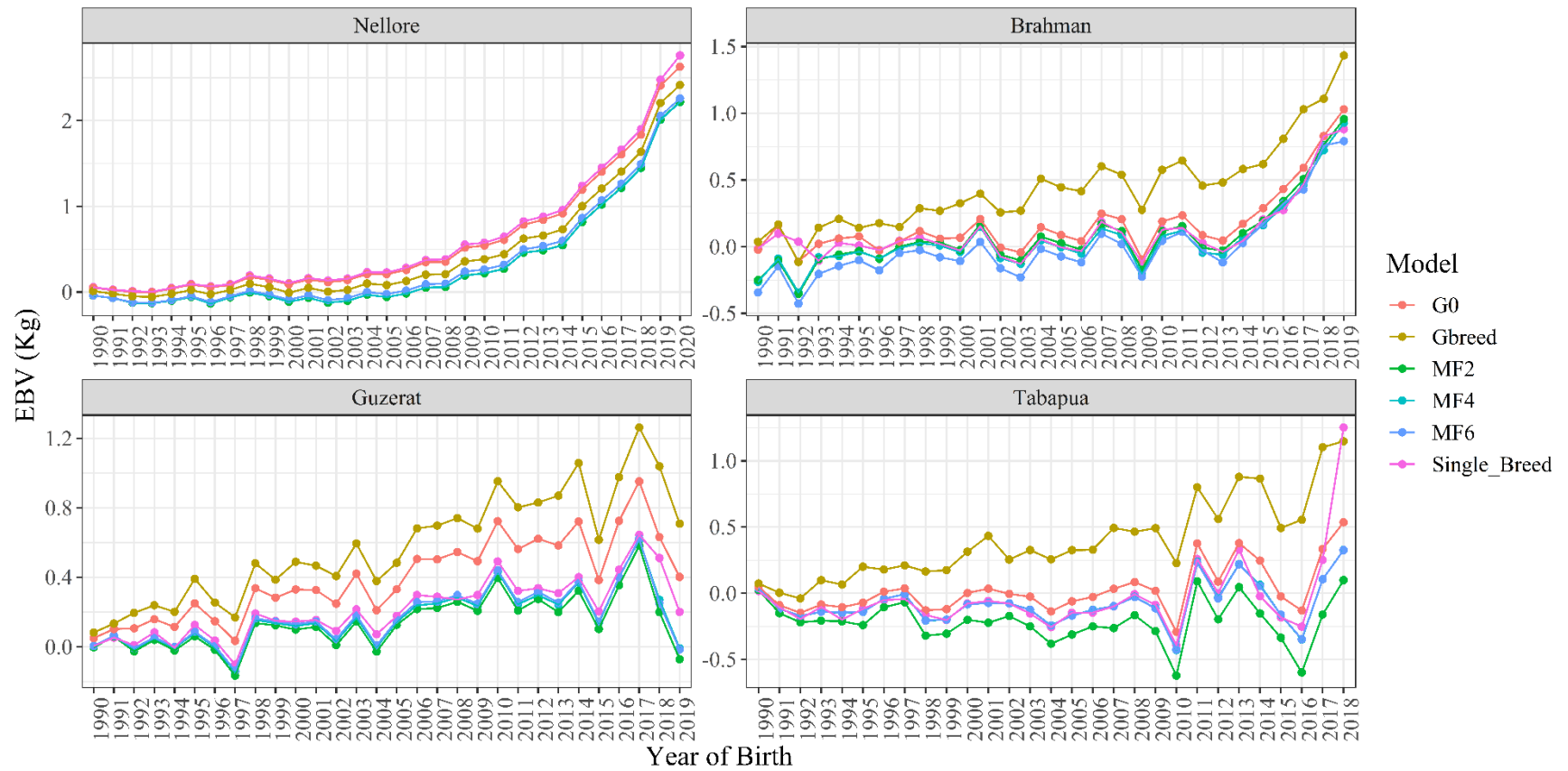
APPENDIX



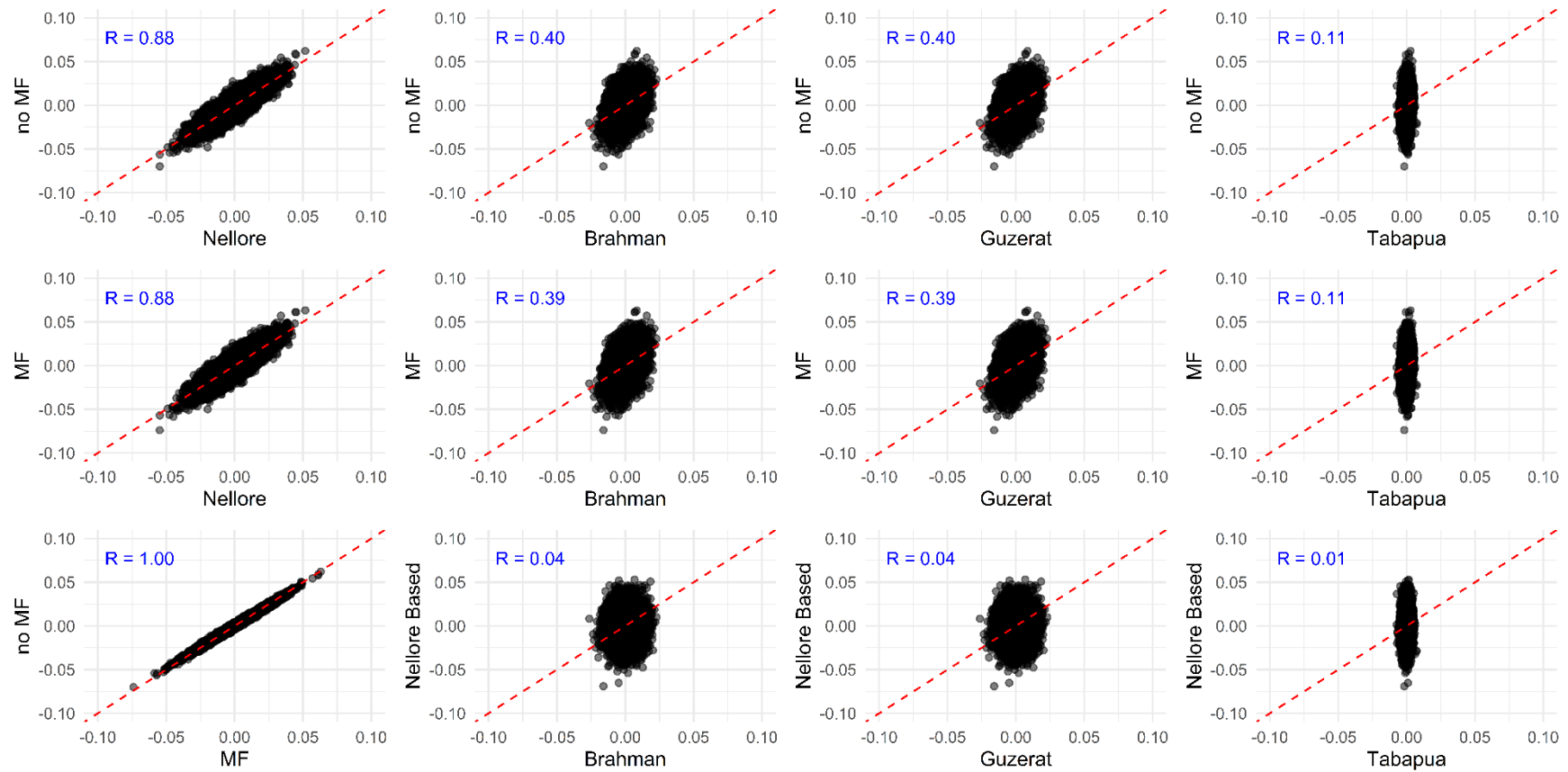
Supp. Figure 2.1. Maternal genetic trends for the trait adjusted weight at 210 (W210) days of age for Bulls of the Breeds Nellore, Brahman, Guzerat and Tabapua, in the six models evaluated.



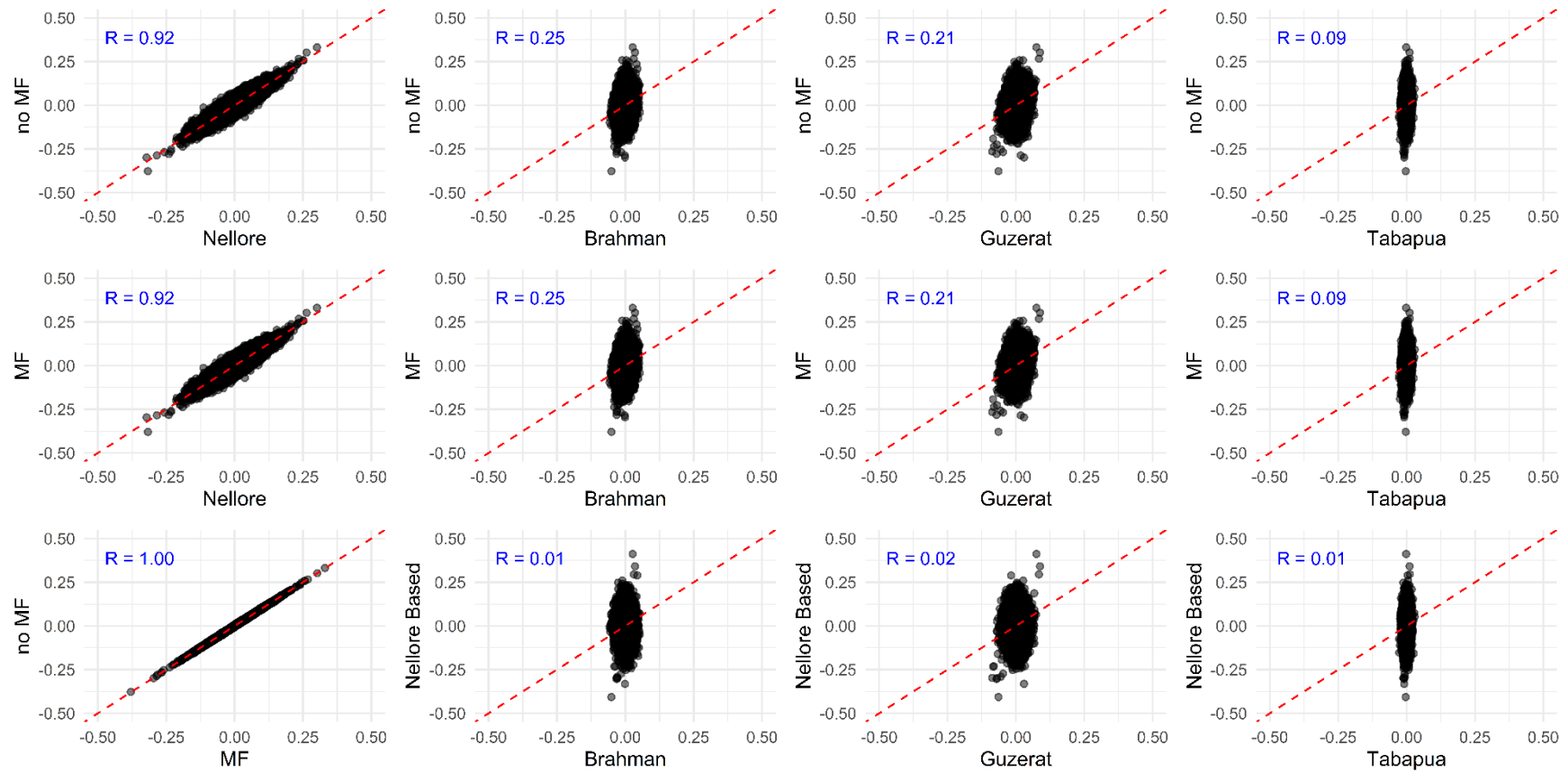
Supp. Figure 2.2. Genetic trends for the trait adjusted weight at 450 (W450) days of age for Bulls of the Breeds Nelore, Brahman, Guzerat and Tabapua, in the six models evaluated.



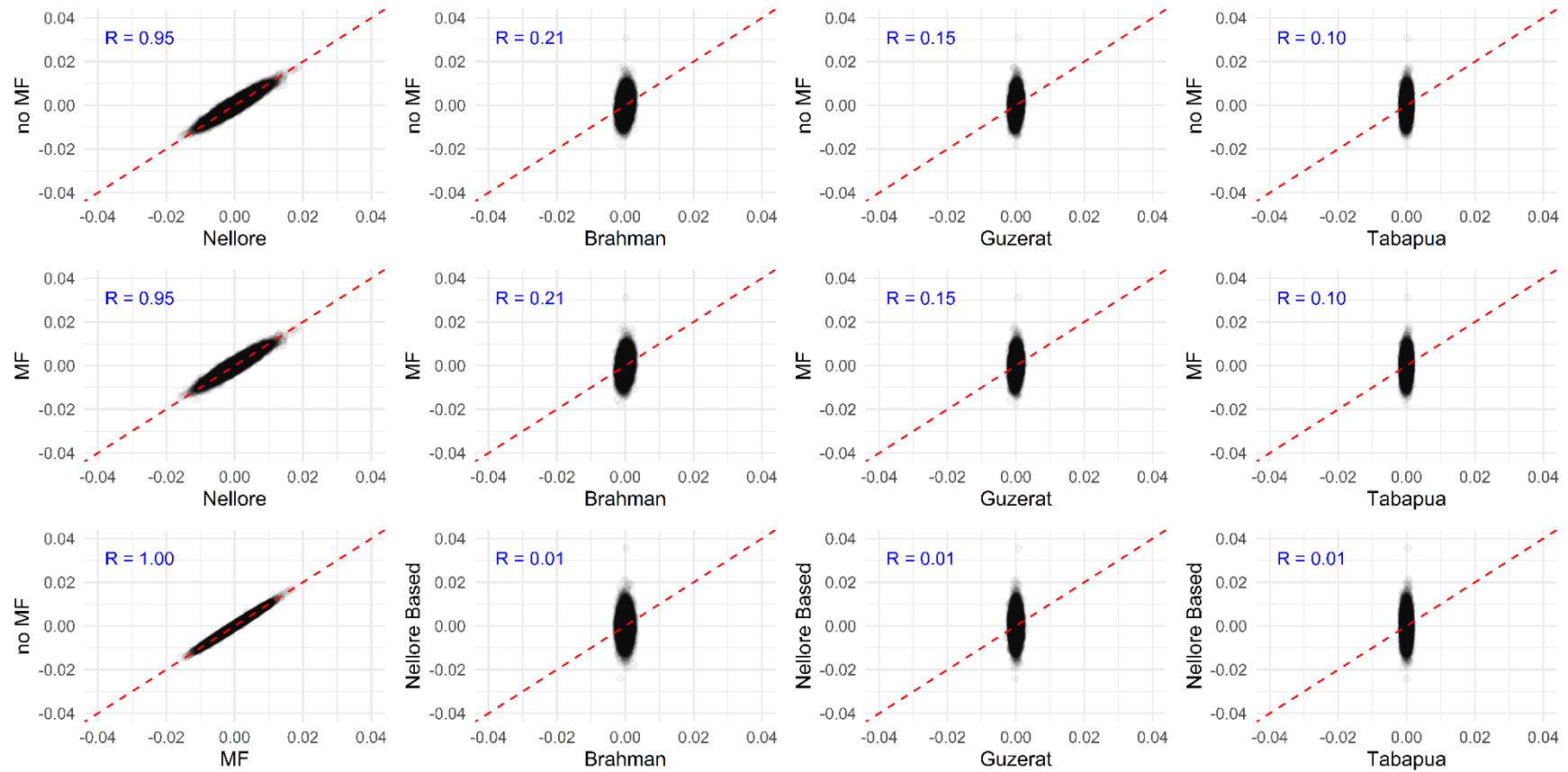
Supp. Figure 2.3. Genetic trends for the trait scrotal circumference at 365 (SC365) days of age for Bulls of the Breeds Nellore, Brahman, Guzerat and Tabapua, in the six models evaluated.



Supp. Figure 3.1. SNP effects correlations (R) between scenarios for the trait maternal adjusted weight at 210 days of age (W_{210m}). No MF: SNP effects based on multi-breed GEBV without metafounders and using the default $\mathbf{G}$, 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.



Supp. Figure 3.2. SNP effects correlations (R) between scenarios for the trait adjusted weight at 450 days of age (W_{450}). No MF: SNP effects based on multi-breed GEBV without metafounders and using the default $\mathbf{G}$, 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.



Supp. Figure 3.3. SNP effects correlations (R) between scenarios for the trait scrotal circumference at 365 days of age (SC365). No MF: SNP effects based on multi-breed GEBV without metafounders and using the default $\mathbf{G}$, 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 scenario