

UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CÂMPUS DE JABOTICABAL

**GENOMIC RANDOM REGRESSION MODELS FOR MILK
PRODUCTION ON THE TEST-DAY OF GIROLANDO CATTLE**

Ana Carolina Almeida Rollo de Paz

Animal Scientist

2025

UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CÂMPUS DE JABOTICABAL

**GENOMIC RANDOM REGRESSION MODELS FOR MILK
PRODUCTION ON THE TEST-DAY OF GIROLANDO CATTLE**

Ana Carolina Almeida Rollo de Paz

Advisor: Prof. Dr. Danísio Prado Munari

Co-advisors: Profa. Dra. Lenira El Faro Zadra

Prof. Dr. Marcos Vinícius Gualberto Barbosa da Silva

Thesis presented to the Faculdade de
Ciências Agrárias e Veterinárias –
Unesp, Campus of Jaboticabal, as part
of the requirements for obtaining the
title of Doctor of Animal Science.

2025

P348g Paz, Ana Carolina Almeida Rollo de
Genomic random regression models for milk production on
the test-day of Girolando cattle / Ana Carolina Almeida Rollo de
Paz. -- Jaboticabal, 2025
91 p.

Tese (doutorado) - Universidade Estadual Paulista (UNESP),
Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal
Orientador: Danísio Prado Munari
Coorientadora: Lenira El Faro Zadra

1. Animal genetics. 2. Dairy cattle breeding. 3. Genomics. 4.
Dairy cattle milk yield. I. Título.

REGISTRO DE IMPACTO

Este trabalho aprimora métodos de estimação de parâmetros genéticos em bovinos leiteiros por meio de modelos de regressão aleatória genômica, com potencial impacto na seleção genética, aumento da produtividade e sustentabilidade da cadeia leiteira.

IMPACT STATEMENT

This study enhances methods for estimating genetic parameters in dairy cattle using genomic random regression models, with potential impact on genetic selection, increased productivity, and sustainability in the dairy industry.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: GENOMIC RANDOM REGRESSION MODELS FOR MILK PRODUCTION ON THE TEST-DAY OF GIROLANDO CATTLE

AUTORA: ANA CAROLINA ALMEIDA ROLLO DE PAZ

ORIENTADOR: DANISIO PRADO MUNARI

COORDENADOR: LENIRA EL FARO ZADRA

COORDENADOR: MARCOS VINÍCIUS GUALBERTO BARBOSA DA SILVA

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

Documento assinado digitalmente
 **DANISIO PRADO MUNARI**
Data: 18/02/2025 11:41:53-0300
Verifique em <https://validar.itl.gov.br>

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

Documento assinado digitalmente
 **AMANDA MARCHI MAIORANO**
Data: 17/02/2025 08:47:44-0300
Verifique em <https://validar.itl.gov.br>

Profa. Dra. AMANDA MARCHI MAIORANO (Participação Virtual)
Departamento de Zootecnia / Universidade Federal de Uberlândia (UFU) - Uberlândia/MG

Documento assinado digitalmente
 **SABRINA LUZIA CAETANO**
Data: 17/02/2025 12:59:24-0300
Verifique em <https://validar.itl.gov.br>

Profa. Dra. SABRINA LUZIA CAETANO (Participação Virtual)
Departamento de Ciências Exatas / FCAV UNESP Jaboticabal

Documento assinado digitalmente
 **MARCOS ELI BUZANSKAS**
Data: 17/02/2025 22:51:48-0300
Verifique em <https://validar.itl.gov.br>

Prof. Dr. MARCOS ELI BUZANSKAS (Participação Virtual)
Departamento de Melhoramento e Nutrição Animal / FMVZ UNESP Botucatu



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal



Prof. Dr. SALVADOR BOCCALETTI RAMOS (Participação Virtual)
Departamento de Ciências Exatas / FCAV UNESP Jaboticabal



Documento assinado digitalmente

SALVADOR BOCCALETTI RAMOS

Data: 18/02/2025 08:05:42 -0300

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

Jaboticabal, 14 de fevereiro de 2025

AUTHOR CURRICULUM INFORMATION

Ana Carolina Almeida Rollo de Paz - born on January 16, 1995, in Taquaritinga, SP, she is the daughter of Maria Elisa Almeida Rollo de Paz and Fabio Paro de Paz. She earned her degree in Animal Science from São Paulo State University (UNESP) – Faculty of Engineering, Ilha Solteira, in 2017, where she was a FAPESP scholarship recipient in both Technical Training and Undergraduate Research projects. In 2020, she completed her Master's degree in Sustainable Animal Production at the Institute of Animal Science, with the thesis entitled "Residual Feed Intake, Pre- and Postnatal Nutrition, and Growth of Nelore Calves." Since 2021, she has been pursuing a Ph.D. in Animal Science at UNESP – Faculty of Agricultural and Veterinary Sciences in Jaboticabal. During her doctoral studies, she was awarded a CAPES scholarship in Brazil and completed a 12-month research internship at the University of Guelph, Canada, through the CAPES-PrInt program. She has also worked as a professor in Biostatistics, Statistics, Applied Statistics, Introduction to Computing, and Applied Computing. She is currently employed as a Data Analyst at Campanelli in Bebedouro, SP.

EPIGRAPH

"Entregue o seu caminho ao Senhor; confie nele e ele agirá." – Salmo 37:5

"Porque o Senhor é bom, e eterna a sua misericórdia; e a sua verdade
estende-se de geração em geração." – Salmo 100:5

DEDICATION

Dedico este trabalho aos meus pais, Maria Elisa e Fabio, que sob muito sol, fizeram-me chegar até aqui na sombra.

ACKNOWLEDGEMENTS

Em primeiro lugar, agradeço a Deus, fonte de força e sabedoria, por guiar meus passos e iluminar meu caminho ao longo dessa jornada. Sem Sua graça e proteção, nada disso seria possível.

Aos meus pais, Maria Elisa e Fabio, por todo o amor, apoio e incentivo incondicionais ao longo da minha vida. Vocês são meu alicerce e minha inspiração constante. À minha irmã, Mariana, e aos meus sobrinhos, Leonardo e Rafael, por estarem sempre ao meu lado, compartilhando alegrias e oferecendo carinho nos momentos mais difíceis.

Ao meu marido, Júnior, meu mais profundo agradecimento pelo amor, paciência, compreensão e por ser meu porto seguro em todos os momentos desta caminhada. Sua presença foi fundamental para que eu pudesse superar os desafios e chegar até aqui. Obrigada por acreditar em mim, por me motivar a seguir em frente e por compartilhar comigo cada conquista ao longo desse percurso.

Aos meus amigos do EAGMA, Ana Carolina, Edimar, Rafael e Larissa, e às minhas queridas amigas Lorena e Catiúcia, meu sincero agradecimento pela amizade, pelo apoio incondicional e pelas conversas que trouxeram leveza aos dias difíceis. Minha gratidão também à Ninette, que foi essencial durante o ano que passei no Canadá. Ela foi uma presença fundamental, oferecendo apoio em todos os momentos e tornando aquela experiência ainda mais especial. Um agradecimento especial à Julia, pelo suporte emocional, pelas palavras de incentivo que me ajudaram a seguir em frente e por estar ao meu lado nos momentos mais desafiadores. Você foi essencial em todo este processo.

Aos meus coorientadores, Dra. Lenira e Dr. Marcos Vinícius, agradeço pelas valiosas orientações, pela paciência e pelos ensinamentos que enriqueceram este trabalho. Ao Dr. Flavio, agradeço por me acolher na University of Guelph, oferecendo uma experiência enriquecedora e oportunidades que contribuíram significativamente para o meu crescimento pessoal e profissional.

Ao meu orientador, Dr. Danísio, meu mais profundo agradecimento pela dedicação, pelo apoio incansável durante esses quatro anos e pelas diversas oportunidades que me foram proporcionadas. Sua orientação foi muito mais que acadêmica; foi também um suporte humano que fez toda a diferença na minha formação e no sucesso desta tese.

Agradeço à CAPES pelo financiamento da minha bolsa de estudos, tanto no Brasil quanto no Canadá, permitindo-me dedicar integralmente à pesquisa e ampliar minhas experiências acadêmicas.

A todos que, de alguma forma, fizeram parte desta jornada, minha mais sincera gratidão. Cada palavra de incentivo, gesto de apoio e contribuição direta ou indireta tornou possível a realização deste sonho. Muito obrigada!

CONTENTS

CHAPTER 1 – GENERAL CONSIDERATIONS¹	1
1. INTRODUCTION	1
2. LITERATURE REVIEW	3
2.1. RANDOM REGRESSION MODELS	3
2.2. COVARIANCE FUNCTIONS	5
2.3. GENOMIC INFORMATION	9
2.4. GENOMIC SELECTION	11
2.4.1. GENOMIC RANDOM REGRESSION MODELS	12
2.5. FINAL CONSIDERATIONS	13
3. REFERENCES	15
CHAPTER 2 – GENETIC PARAMETER ESTIMATES FOR TEST-DAY MILK YIELD IN GIROLANDO CATTLE USING RANDOM REGRESSION MODELS	22
1. INTRODUCTION	24
2. MATERIALS AND METHODS	25
2.1. DATA	25
2.2. ANALYSIS MODELS	26
2.3. MODEL COMPARISON	28
2.4. ESTIMATION OF VARIANCE COMPONENTS AND BREEDING VALUES	29
3. RESULTS AND DISCUSSION	30
4. CONCLUSIONS	39
5. REFERENCES	40
CHAPTER 3 – POPULATION STRUCTURE AND GENOMIC PARAMETER ESTIMATES FOR TEST-DAY MILK YIELD IN GIROLANDO CATTLE USING RANDOM REGRESSION MODELS	45
1. INTRODUCTION	47
2.1. Phenotypic data	48
2.2. Genotypic data	49
2.3. Population structure and genetic diversity analysis	50
2.4. Analysis model	50
2.5. Estimation of (co)variance components	53
3. RESULTS AND DISCUSSION	54

5. REFERENCES.....	66
Appendix.....	72
CHAPTER 4 – FINAL CONSIDERATIONS	73

GENOMIC RANDOM REGRESSION MODELS FOR TEST-DAY MILK PRODUCTION IN GIROLANDO CATTLE

ABSTRACT – This study presents an integrated approach for evaluating milk yield and the genetic structure of the Girolando cattle population using random regression models (RRM) based on test-day milk yield (TDMY) records. These models allow for the adjustment of additive genetic and permanent environmental effects over the lactation period, providing more accurate estimates of genetic values and parameters compared to traditional methodologies. A total of 480,364 TDMY records from 76,688 first-lactation cows across 913 herds, with calvings occurring between 1991 and 2021, were analyzed. The study was conducted in two stages: the first employed the single-step BLUP (ssBLUP) model using phenotypic and pedigree data; the second applied the single-step genomic BLUP (ssGBLUP), incorporating genotypic information from 25,936 animals. The animals represented nine genetic groups formed by different proportions of Holstein (H) and Gir (G) ancestry. The average daily milk yield was 22.3 kg, peaking at 23.8 kg between 91 and 120 days in milk. The most suitable model was a fourth-order Legendre orthogonal polynomial with heterogeneous residual variances modeled across five classes. The average heritability and repeatability estimates were 0.56 and 0.84, respectively. The genomic analysis revealed high genetic variability, with genotypic frequencies indicating increased heterozygosity in crossbred groups. The mean inbreeding coefficient was close to zero, suggesting that inbreeding is effectively managed through mating strategies. The genetic differentiation index (F_{st}) ranged from 0.003 to 0.21, with the highest divergence observed between the pure Holstein and Gir groups, and the lowest among crossbred groups. Genetic correlations between adjacent lactation stages ranged from 0.64 to 0.99, indicating strong persistence of permanent environmental effects over time. In conclusion, the Girolando population exhibits high genetic diversity, with a positive impact of heterosis especially in crossbred animals. Incorporating genomic information through the ssGBLUP model improved the accuracy of genetic evaluations compared to traditional approaches based solely on pedigree. The combined use of RRM and genomic data is a robust and promising tool for enhancing genetic selection strategies in dairy cattle, particularly under tropical production conditions.

Keywords: animal breeding, dairy cattle, genetic diversity, genomic selection, longitudinal traits, population structure, test-day milk yield.

MODELOS DE REGRESSÃO ALEATÓRIA GENÔMICA PARA PRODUÇÃO DE LEITE NA DATA DO CONTROLE DE BOVINOS GIROLANDO

RESUMO – Este estudo apresenta uma abordagem integrada para a avaliação da produção de leite e da estrutura genética da população Girolando, utilizando modelos de regressão aleatória (MRA) aplicados a dados de produção de leite no dia do controle (PLDC). Esses modelos permitem o ajuste dos efeitos genéticos aditivos aleatórios e ambientais permanentes ao longo da lactação, proporcionando estimativas mais precisas de valores genéticos e parâmetros genéticos em comparação com metodologias tradicionais. Foram analisados 480.364 registros de PLDC de 76.688 vacas em primeira lactação, oriundas de 913 rebanhos, com partos ocorridos entre 1991 e 2021. O estudo foi conduzido em duas etapas: na primeira, utilizou-se o modelo ssBLUP com dados fenotípicos e de pedigree; na segunda, foi aplicado o modelo ssGBLUP, incorporando dados genotípicos de 25.936 animais. Os animais avaliados pertenciam a nove grupos genéticos resultantes de diferentes proporções entre as raças Holandesa (H) e Gir (G). A produção média diária de leite foi de 22,3 kg, com pico de 23,8 kg entre 91 e 120 dias de lactação. O modelo mais adequado foi o polinômio ortogonal de Legendre de quarta ordem, com variâncias residuais heterogêneas em cinco classes. As estimativas médias de herdabilidade e repetibilidade foram 0,56 e 0,84, respectivamente. Com a inclusão de dados genômicos, observaram-se frequências genotípicas indicando alta heterozigosidade nos grupos mestiços, e o coeficiente médio de endogamia foi próximo de zero, sugerindo controle efetivo dos acasalamentos. O índice de diferenciação genética (F_{st}) variou de 0,003 a 0,21, com maior divergência entre os grupos Holandês e Gir, e menor entre grupos mestiços. As correlações genéticas entre períodos adjacentes de lactação variaram de 0,64 a 0,99, evidenciando persistência dos efeitos ambientais permanentes ao longo do tempo. Conclui-se que a população Girolando apresenta elevada diversidade genética, com impacto positivo da heterose nos grupos mestiços. A inclusão de informações genotípicas na avaliação genética, por meio do modelo ssGBLUP, aumentou a precisão das estimativas em relação às abordagens baseadas exclusivamente em pedigree. O uso combinado de MRA e dados genômicos

constitui uma ferramenta robusta e promissora para a seleção genética mais eficiente em sistemas de produção leiteira tropical.

Palavras-chave: avaliação genética, diversidade genética e estrutura populacional, gado leiteiro, Girolando, modelos de regressão aleatória, produção de leite no dia do controle.

CHAPTER 1 – GENERAL CONSIDERATIONS¹

1. INTRODUCTION

The Girolando, developed in Brazil from a crossbreed between the Holstein and Gir breeds in the 1940s, reflects the search for an efficient solution for milk production in tropical and subtropical regions (Freitas et al., 2002). This strategic crossbreeding sought to combine the Gir breed's adaptability with the Holstein breed's high productivity, resulting in animals with productivity, fertility, and vigor. With racial compositions that vary in the proportion of Holstein (0.25 to 0.875) in the cross with Gir, the Girolando stands out as a bovine population adapted to the tropical environment (Freitas et al., 2002; Silva et al., 2023).

Daily milk yields, often assessed through test-day milk yield (TDMY), are considered longitudinal measures in dairy cattle. As Kirkpatrick & Heckman (1989) indicated, these measures can be analyzed using random regression models (RRM), also known as infinite-dimensional models, which can provide a more comprehensive and detailed understanding of production performance throughout lactation.

From the RRM, it is possible to estimate the breeding values for the lactation curves of each animal using the random regression coefficients to estimate the breeding values for the components of the curve, such as the peak and persistence of lactation. Lactation persistency measures the cow's ability to maintain its production after peak lactation, i.e., the higher the persistency of a cow's lactation, the slower the milk production decreases after reaching peak lactation (Ptak & Schaeffer, 1993; Jamrozik & Schaeffer, 1997). The importance of RRM for Girolando or even Zebu dairy breeds is increased by the need to include the shape of the lactation curve in genetic evaluations, in addition to 305-day cumulative milk yield.

RRM is adjusted according to the days in milk and uses ordinary polynomials or linear functions that model the trajectory of the lactation curve about the population average (fixed regressions) and for each animal (random

¹: This chapter was adapted from the article entitled “Utilizing random regression models for genetic evaluation of milk production in dairy cattle: a comprehensive approach”, submitted to the Journal of Applied Animal Research.

regression) according to Jamrozik & Schaeffer (1997). In this way, it is possible to estimate a genetic lactation curve for each animal using the (co)variances over time, and the breeding values are obtained from the genetic regression coefficients. Legendre's orthogonal polynomials have been the most widely used in RRM, as they are orthogonal and have the advantage of reducing correlations between the estimated coefficients and modeling the shape of the lactation curve (Pool et al., 2000; Schaeffer, 2004).

RRM has been used in various animal production systems. In dairy farming, RRM is fitted to milk production data to estimate variance components and genetic parameters (El Faro et al., 2008; Cobuci et al., 2011; Pereira et al., 2013; Buaban et al., 2020; Mbuthia et al., 2021; Khanal et al., 2022; Ashrafian et al., 2023) and to use this information in genetic evaluations for milk yield and the components of the lactation curve (Cobuci et al., 2007; Koivula et al., 2015; Kang et al., 2018; Buaban et al., 2020; González-Herrera et al., 2022; Ashrafian et al., 2023).

The use of RRM, according to Meyer (2004), increases accuracy by 6 to 8% for estimates of the direct effects of animals with only one or two TDMY. Studies indicate that the inclusion of genomic information in genetic evaluations can increase the accuracy of the genomic estimated breeding values (GEBV) by up to 70% and from 50 to 100% in the rate of genetic gain per year of production traits (VanRaden et al., 2009; García-Ruiz et al., 2016) and with the consequent increase in the accuracy of GEBV in the selection of young animals, which can reduce the generation interval (Meuwissen et al., 2001; Pryce & Daetwyler, 2012).

This review provides a comprehensive overview of RRM in the context of animal breeding. Topics approached include (i) a historical introduction of the development of the first RRM to the most recent advances; (ii) a discussion of how covariance functions are used to model the trajectory of repeated measurements, highlighting their importance in the analysis of longitudinal data; and (iii) explanation of how genomic information is integrated into genetic selection, using RRM to improve the accuracy and effectiveness of selection for desirable traits.

2. LITERATURE REVIEW

2.1. RANDOM REGRESSION MODELS

According to Meyer (2005), RRM is an alternative for better data use and adjustment when modeling longitudinal data. It uses all the records of the animal and its relatives, increasing the selection accuracy (Laird & Ware, 1982), which is compared to a multivariate analysis; each observation is considered a single trait (Meyer & Hill, 1997). The first RRM was proposed to solve the over-parameterization of multitrait models (Henderson, 1982) and was developed based on repeatability models, which estimate the genetic effect at different times in the animal's life without increasing the complexity of the model, which is one of the differentials of these models (Lidauer et al., 2003).

RRM was first used by Schaeffer & Dekkers (1994) as an alternative for describing the lactation curve of animals and evaluating them for milk production. It replaced the traditional models that used the 305-day cumulative milk yield or the multitrait TDMY.

Ptak & Schaeffer (1993) introduced an alternative model consolidating multiple TDMY into a single record per lactation. This model did not account for the specific effects of each TDMY, but it did incorporate projection factors to estimate the total 305-day milk yield for incomplete lactations.

With RRM, it is possible to predict two sets of curves associated with milk production, in which one set called fixed regression is adjusted for cows belonging to the same calving age subclass and contemporary group or even a single lactation curve for the population. The other set, called random regression for each individual, is calculated as the deviation from the fixed curve for the subclass to which the animal belongs (Jamrozik & Schaeffer, 1997; Schaeffer, 2004). The equation can be described as single-trait RRM for milk yield data:

$$y_{ijk} = \sum_{q=1}^Q b_{qi} z_{qi} + \sum_{r=1}^R a_{rj} z_{rj} + \sum_{s=1}^S p_{sj} z_{sj} + e_{ijk} \quad [1]$$

where y_{ijk} is the k-th record of animal j in the i-th group; b_{qi} is the q-th fixed regression coefficient for the i-th group; a_{rj} is the r-th random regression

coefficient for the additive genetic effect of animal j ; p_{sj} is the s -th random regression coefficient for the permanent environment effect of animal j ; e_{ijk} is the residual effect; z_{qi} , z_{rj} , and z_{sj} are the covariates of order Q , R , and S , used to describe the fixed curve and the additive genetic and permanent environment random curves of the animal.

The RRM uses the Mixed Model Equations (MME) (Henderson, 1963; 1973), and these can be described as follows:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}W \\ Z'R^{-1}X & Z'R^{-1}Z + A^{-1} \otimes G_0^{-1} & Z'R^{-1}W \\ W'R^{-1}X & W'R^{-1}Z & W'R^{-1}W + I \otimes P_0^{-1} \end{bmatrix} * \begin{bmatrix} b \\ a \\ p \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ W'R^{-1}y \end{bmatrix} \quad [2]$$

where y is the vector of observations; X , Z , and W are the incidence matrices for the fixed effects (contemporary group, age of cow at calving as a covariate, and regression coefficients of days in milk), (b) , the additive genetic random regression coefficients (a) and the random regression coefficients of the permanent environment of animal (p) , respectively; A is the genetic relationship matrix based on pedigree information; I is the identity matrix; G_0 and P_0 are the (co)variance matrices of the animal's additive genetic and permanent environment random regression coefficients, respectively; R is the residual variance matrix.

The RRM assumptions can be described as follows:

$$E[y] = Xb$$

and

$$Var \begin{bmatrix} a \\ p \\ e \end{bmatrix} = \begin{bmatrix} A \otimes G_0 & 0 & 0 \\ 0 & I \otimes P_0 & 0 \\ 0 & 0 & R \end{bmatrix} \quad [3]$$

where e is the vector of residuals; the other terms were defined in [equation 2], G_0 and P_0 contain the additive genetic and permanent environment (co)variances of the animal between the random regression coefficients. The matrices of the animal's additive genetic and permanent environment (co)variances of the animal for TDMY can be obtained:

$$\Sigma = TG_0T'$$

and

$$\phi = TP_0T' \quad [4]$$

where Σ is the matrix of genetic (co)variances between TDMY; ϕ is the matrix of permanent environmental (co)variance of the animal; T is the matrix of independent covariates for all points (days in milk).

With the solutions to the regression coefficients for each animal, it is possible to predict the estimated breeding values (EBV) for all points in the trajectory. The vector of the animal's EBV can be calculated using the equation:

$$EBV_j = T\hat{a}_j \quad [5]$$

where $\hat{a}_j$ is the vector of EBV for the coefficients of the covariance function for the j -th animal, and T is the days in milk.

Several authors have demonstrated the efficiency of using RRM in dairy cattle, where RRM improves the solutions of MME, the computational demand is lower, the estimation of genetic parameters is more accurate and with less bias, and the TDMY are better adjusted compared to other multitrait test-day models (Lidauer et al., 2003; Schaeffer, 2004; Cobuci et al., 2011; Oliveira et al., 2019).

2.2. COVARIANCE FUNCTIONS

The (co)variances between traits, which are measured at different ages, as a function of these ages are called the covariance function (CF), which is the finite-dimensional equivalent of a (co)variance matrix in a multitrait model (Kirkpatrick & Heckman, 1989). According to Meyer (1998), the CF is a continuous function that provides the (co)variances of traits measured at different points in a trajectory and explains the covariance between measurements taken at certain ages as a function of those ages.

According to Kirkpatrick & Heckman (1989), there are three main advantages of CFs compared to other conventional models: (i) CFs produce a description for each point along a continuous scale of measurements (time),

which facilitates interpolation to find covariances between unmeasured ages; (ii) Each CF has a set of eigenvalues and eigenfunctions that are comparable to eigenvectors and provide information on the direction in which the average curve is more likely to be altered by selection due to its greater genetic variance; (iii) CF models make the estimation and use of relevant parameter data more efficient. In addition, negative sample correlations can be minimized, which can avoid overparameterization, reduce residual variance, adjust lower order polynomials, and considering the exact date the data was collected, regardless of whether its collection dates are inaccurate (Meyer & Hill, 1997; Kirkpatrick et al.; 1994).

A CF, which results in the covariance between ages m and l , can be represented as:

$$f(a_m^*, a_l^*) = \sum_{i=0}^{k-l} \sum_{j=0}^{k-l} \phi_i(a_m^*) \phi_j(a_l^*) K_{ij} \quad [6]$$

where ϕ_i is the i -th polynomial; k is the order of the polynomial being adjusted; a^* are the ages adjusted for the range in which the polynomial is defined; and K_{ij} are the elements of the matrix K , which is the matrix of coefficients of the covariance function.

In matrix notation:

$$\Sigma = \Phi K \Phi' e K = \Phi^{-1} \Sigma (\Phi^{-1})' \quad [7]$$

where $k = t$ (number of ages); Σ is the observed covariance matrix; Φ is a $t * k$ matrix of orthogonal polynomials (Meyer & Hill, 1997).

Kirkpatrick et al. (1990; 1994) modeled CF using orthogonal polynomials, called Legendre Orthogonal Polynomials (ϕ_k), which presented the following formula:

$$\phi_k(a_i^*) = \frac{1}{2} \sqrt{\frac{2k+1}{2}} \sum_{m=0}^{[k/2]} (-1)^m \binom{k}{m} \binom{2k-2m}{k} (a_i^*)^{k-2m} \quad [8]$$

where a_i^* are the ages adjusted for the range in which the polynomial is defined $[-1 \text{ to } +1]$. The ages are standardized as follows:

$$a_i^* = \frac{2(a_i - a_{\min})}{a_{\max} - a_{\min}} - 1 \quad [9]$$

where a_i is the age to be adjusted, and $a_{\max}$ and $a_{\min}$ are the maximum and minimum ages in the data.

RRM is a specific case of CFs (Meyer & Hill, 1997; Meyer, 1998), and the Restricted Maximum Likelihood (REML) methodology can be used to estimate the coefficients of RRM covariance functions directly. This can be done by estimating variance components using a reparameterization of the REML methodology in finite-dimensional models (Meyer & Hill, 1991). The coefficient matrix will be positive definite, which is one of the advantages of using REML.

The method created by Meyer & Hill (1997) can be summarized below:

$$y = Xb + Za + e \quad [10]$$

where y is the vector of observations; b is the vector of fixed effects; a is the vector of additive random effects of the animal; e is the vector of residual random effects; X and Z are incidence matrices for the fixed and random effects, respectively; and $\text{var}(y) = ZGZ' + R$, $\text{var}(a) = G$, with $G = A \otimes G_0$. If the matrices are the same for all traits, $\text{var}(e) = R = I \otimes R_0$, with A the numerator of the relationship matrix between the animals; G_0 the matrix of additive genetic variances and covariances between the t traits; R_0 the matrix of residual variances and covariances between the t traits and $\otimes$ the Kronecker product. Thus, the logarithm of the REML function is:

$$\ln \mathcal{L} = -1/2 [\text{const} + N \ln|R_0| + N_a \ln|G_0| + t \ln|A| + \ln|C| + y'Py] \quad [11]$$

where N is the number of animals in the analysis, C is the coefficient matrix in the MME, and $y'Py$ is the sum of squares of the residuals. Meyer & Hill (1997) rewrote the above equation as a function of the (co)variance matrices between the CF

coefficients, K_a and K_e , for the additive and residual genetic CF, respectively, in place of G_0 and R_0 . Thus:

$$\ln \mathcal{L} = -1/2 [\text{const} + N \ln|K_e| + N_a \ln|K_a| + (N + N_a) \ln|\Phi\Phi'| + t \ln|A| + \ln|C| + y'Py] \quad [12]$$

where Φ is the $t \times k$ matrix (days in milk $\times$ coefficient matrix in the MME) with the orthogonal polynomials.

It is possible to separate the environment effects into permanent and temporary, with $e = r + \varepsilon$, and $\Sigma_e = \Sigma_r + \Sigma_\varepsilon$, where r is the vector of permanent environment effects of the animal; ε the vector of temporary environment effects; Σ_r the matrix of permanent environment (co)variances of the animal; and $\Sigma_\varepsilon = \text{Diag}\{\sigma_{\varepsilon_i}^2\}$ the matrix of temporary environment variances. Thus, the log of the REML function can be rewritten as:

$$\ln \mathcal{L} = -1/2 [\text{const} + N \ln|\Phi K_r \Phi' + \text{Diag}\{\sigma_{\varepsilon_i}^2\}| + N_a \ln|K_a| + N_a \ln|\Phi\Phi'| + t \ln|A| + \ln|C| + y'Py] \quad [13]$$

where K_r is the matrix of coefficients of the permanent environment covariance function of the animal.

Polynomial regressions generally describe growth or lactation curves. Regressions on time (age) are generally considered fixed effects to account for general trends within fixed-effects classes. Most studies have used orthogonal Legendre polynomials to model longitudinal data, as there is no need to make assumptions about the shape of the curve or the trajectory of the trait over time. However, higher-order polynomials are generally required to model the effects of the permanent environment of animals (Meyer, 2000).

In studies using RRM that considered the residual variances as homogeneous, the genetic variances of the animal were overestimated throughout lactation (Jamrozik & Schaeffer, 1997; Jamrozik et al., 1997). The estimation of variance components may be affected by the fact that residual variances throughout lactation are considered homogeneous.

Using residual variances as heterogeneous can improve the partitioning of phenotypic variances. However, on the other hand, the number of parameters

that must be estimated to maximize the likelihood function increases (Meyer & Kirkpatrick, 2005). According to El Faro et al. (2008), using this methodology for genetic evaluations with large data can be difficult due to the number of parameters in the model, the covariance structure, and the incidence matrices being denser than conventional models.

Several authors have shown that Legendre's orthogonal polynomials fit RRM test-day models well. However, there is no consensus on the best choice for adjusting the order of the polynomials, as this depends on the structure of the data and the population structure (Bignardi et al., 2009; Cobuci et al., 2011; Pereira et al., 2013).

Many early studies spent a great deal of time determining the orders of the polynomials to adjust the effects of the additive genetic and the permanent environment of the animal in the case of milk production. The literature also shows that the best fit for the model is one in which heterogeneous residual variances are assumed, grouped into periods, for the different stages of lactation with no need to use different residual variances for each days in milk (Bignardi et al., 2009; Cobuci et al., 2011; Pereira et al., 2013; Flores & Van der Werf, 2015; Scaletz et al., 2018; Jattawa et al., 2021).

More parsimonious models that do not alter the quality of fit for the random effects considered are needed once it is observed that the use of highly parameterized models tends to generate erratic estimates between the beginning and the end of lactation, in addition to many oscillations in their magnitudes and signs.

2.3. GENOMIC INFORMATION

Genomic technologies have become essential for genetic evaluation in milk production (Schultz & Weigel, 2018). Genomic Selection (GS) and Genome-Wide Association Study (GWAS) are the most widely used techniques. With these two techniques, it is possible to predict GEBV and identify the molecular markers related to the trait of interest.

GS and GWAS are complementary analyses requiring genomic information from SNP (Single Nucleotide Polymorphism) molecular markers. An SNP is defined as a substitution for a single nucleotide (Adenine, Cysteine,

Thymine, and Guanine) for another at a given location in the DNA, and the least frequent allele must be present in at least 1% of the population (Brookes, 1999).

The effects of SNP are used to predict GEBV, making it possible to identify and select the best animals. Genomic analysis can identify genetic markers associated with the traits of interest, allowing for selecting animals with greater genetic potential associated with productive and reproductive performance (Gutierrez-Reinoso et al., 2021). It is also possible to quantify the variance components of the trait being analyzed.

According to Hayes et al. (2013), GWAS studies of dairy cattle have some specificities that do not occur in other species. The intensive use of the same commercial bulls within the same herd contributes to an elevated level of linkage disequilibrium (LD), an increase in inbreeding, and a reduction in the effective size of the population, which leads to a robust statistical analysis of the association between SNP and phenotype (Hayes et al., 2013).

With SNP panels, it is possible to understand better traits related to fertility and longevity, which are difficult to measure and have low heritability, as well as complex traits that are expensive to measure, such as feed intake, methane emissions, and fatty acid content. Identifying these regions of interest in the bovine genome optimizes constructing specific and customized SNP panels for these traits (Iung et al., 2018; Steri et al., 2018).

However, SNP studies in dairy cattle face challenges because they have small reference populations, and most validations are created by randomly sampling databases with all genotyped animals. In addition, genomic data from sires and cows in the reference populations is often pooled, which can affect the accuracy of genomic predictions. In such cases, genomic predictions are less accurate than reference populations with more genotypic information about the cows (Mrode et al., 2019; Bolormaa et al., 2022).

The inclusion of genomic information from cows in reference populations led to a fivefold increase in the size of the reference population and an increase of up to 12% in the accuracy of genomic predictions compared to reference populations that only used sire genotypes (Calus et al., 2013; Mrode et al., 2019; Bolormaa et al., 2022). Boison et al. (2017) evaluated including younger bulls' genotypes in the cows' genomic evaluation. The authors found that the accuracy of GEBV was higher when the genotypes of younger animals were included in

the genomic analysis. The application of GS in dairy cattle has shown promising results, and these techniques are likely to be increasingly used to improve the genomic selection of these animals.

2.4. GENOMIC SELECTION

Studies on genomic selection began with research by Nejati-Javaremi, Smith and Gibson (1997), Lande & Thompson (2000), and Meuwissen et al. (2001). GS is a technique that uses markers to predict GEBV, and it was first implemented in genomic evaluation in dairy cattle (Schaeffer, 2006). GS methods have developed rapidly and make it possible to estimate the GEBV of bulls for milk production at the beginning, which can replace progeny testing in populations with a large amount of phenotype and genealogy information (Schaeffer, 2006; Hayes et al., 2009; Venot et al., 2016; Wiggans et al., 2017).

The first methods implemented for genomic evaluation involved several steps (multi-step) up to the prediction of the GEBV. The EBV for each milk production record was initially estimated using RRM and then adjusted to be used as pseudo-phenotypes, which were then used to calculate the GEBV through the traditional best linear unbiased predictor (BLUP) model. Subsequently, these pseudo-phenotypes were integrated into a genomic prediction model to estimate the effects of the markers on the reference population (VanRaden, 2008; Hayes et al., 2009).

Some animals may have phenotypic records but not be genotyped, so Misztal et al. (2009) and Aguilar et al. (2010) proposed a methodology for integrating phenotype, genotype, and pedigree records of animals called single-step genomic best linear unbiased prediction (ssGBLUP). In ssGBLUP, the relationship matrix based on pedigree (A) and the relationship matrix based on genomic information (G) are combined into a single H matrix (Legarra et al., 2009):

$$H = A + \begin{bmatrix} A_{12}A_{22}^{-1} & 0 \\ 0 & I \end{bmatrix} [G \quad -A_{22}] [I \quad I] \begin{bmatrix} A_{22}A_{21}^{-1} & 0 \\ 0 & I \end{bmatrix} \quad [14]$$

where I is an identity matrix; subscripts 1 (2) are animals without (with) genomic information; A is the pedigree matrix; A^{-1} is its inverse; and G is the genomic relationship matrix.

The inverse of H (H^{-1}) is described as:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \tau(\alpha G + \beta A_{22})^{-1} - \omega A_{22}^{-1} \end{bmatrix} \quad [15]$$

where A_{22} and A_{22}^{-1} are parts of the matrix A and A^{-1} , respectively, which contain the information of the animals with genomic information; α and β are parameters used for the inversion of the matrix G ; τ and ω are parameters used for the reduction of genetic variance and for the pedigree, respectively, to allow the matrices G^{-1} and A_{22}^{-1} to be compatible.

The compatibility between the G^{-1} , A^{-1} and A_{22}^{-1} matrices is related to the completeness of the pedigree (Mizstal et al., 2017). Christensen et al. (2012) showed that adjusting the G^{-1} matrix about the expected values in the A_{22}^{-1} matrix reduces scale differences, which improves the accuracy of GEBV estimates and reduces the bias of evaluations using the ssGBLUP methodology.

Aguilar et al. (2010) concluded that, despite the complexity of obtaining the H matrix and its inverse H^{-1} used in ssGBLUP, the method is viable and can provide a unified structure compared to multi-step procedures, with single-step simplifying and eliminating several parameters, which allows genetic evaluations to be more accurate. In addition, ssGBLUP can explain selection bias when selection is based only on genotypes (VanRaden, 2012).

2.4.1. GENOMIC RANDOM REGRESSION MODELS

The ssGBLUP methodology is based on the infinitesimal model, which assumes equal variance for all SNP effects (Wang et al., 2012). To predict GEBV using ssGBLUP based on RRM, a fixed regression (population mean) and genomic random regression coefficients are adjusted. The same RRM used for genetic evaluation can be used for genomic predictions based on ssGBLUP, using [equation 1]. The MME for genomic evaluation using ssGBLUP is similar to

the MME in traditional genetic evaluation. The only difference is that matrix A (pedigree) is replaced by matrix H (genomic) in [equation 2].

Mizstal et al. (2014) demonstrated that one of the main advantages of the ssGBLUP methodology is the ability to use the same statistical model when genomic information is added. In addition, ssGBLUP simplifies the analysis with RRM since it uses a single step in the estimation of GEBV, and it is possible to consider the same statistical model used in other genetic evaluations. Also, including genotypes as genomic information in evaluating longitudinal traits using RRM may be a desirable alternative for obtaining GEBV estimates and more accurate selection decisions.

Mehrban et al. (2019), comparing ssGBLUP methodologies with traditional BLUP, obtained more accurate GEBV estimates with ssGBLUP due to including genomic information in the analysis. These authors concluded that ssGBLUP can be considered a promising methodology for predicting GEBV. Similarly, other authors have concluded that using the ssGBLUP method in RRM provides high accuracy and less bias when estimating GEBV for longitudinal data when compared to other methods (Lourenco et al., 2014; Guarini et al., 2018; Piccoli et al., 2018; Koivula et al., 2021; Oliveira et al., 2019).

Thus, it is possible to estimate EBV more accurately by incorporating genomic information into genetic evaluation using ssGBLUP. The ssGBLUP methodology simplifies the analysis by using a single step to estimate the GEBV and allows the same statistical model to be used as in traditional genetic evaluations.

2.5. FINAL CONSIDERATIONS

This revision aimed to comprehensively analyze the current knowledge and application of RRM in animal breeding. Studies using longitudinal traits have shown that RRM is a viable alternative for estimating genetic parameters, allowing animals to be selected based on the complete pattern of the production curve using genomic information. Thus, including genomic information in evaluating longitudinal traits through RRM represents a promising alternative for more precise selection decisions.

Despite the advances in RRM applications, more research is still needed to combine RRM and genomic information successfully. Future research should provide a better understanding of the temporal variation of biological processes and the physiological implications underlying longitudinal traits, which will allow for improved methodologies and a more effective application of RRM in the context of animal breeding.

One of the major challenges in this field is the application of RRM in crossbred populations. The genetic heterogeneity and variability of breed composition in crossbred animals introduce additional complexity in modeling genetic parameters over time. Furthermore, the lack of structured and consistent pedigree or genomic data in crossbred populations can limit the effectiveness of genomic prediction models. These factors highlight the need for tailored approaches and robust statistical methods to ensure that RRM can be effectively applied in diverse and crossbred animal populations.

3. REFERENCES

- Aguilar, I, Misztal, I, Johnson, DL, Legarra, A, Tsuruta, S and Lawlor, TJ (2010) A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science** 93, 743-752.
- Ashrafiyan, A, Kashan, NEJ, Abbasi, MA, Sadeghi, AA and Rokouei, M (2023) Study of persistency of lactation and survival of Iranian Holstein dairy cattle using random regression model. **Canadian Journal of Animal Sciences** 103, 405-415.
- Bignardi, AB, El Faro, L, Cardoso, VL, Machado, PF and Albuquerque, LG (2009) Random regression models to estimate test-day milk yield genetic parameters Holstein cows in Southeastern Brazil. **Livestock Science** 123, 1-7.
- Boison, SA, Utsunomiya, ATH, Santos, DJA, Neves, HHR, Carneiro, R, Mészáros, G, Utsunomiya, YT, do Carmo, AS, Verneque, RS, Machado, MA, Panetto, JCC, Garcia, JF, Solkner, J and Silva, MVGB (2017) Accuracy of genomic predictions in Gyr (*Bos indicus*) dairy cattle. **Journal of Dairy Science** 100, 5479-5490.
- Bolormaa, S, MacLeod, IM, Khansefid, M, Marett, LC, Wales, WJ, Miglior, F, Baes, CF, Schenkel, FS, Connor, EE, Manzanilla-Pech, CIV, Stothard, P, Herman, E, Nieuwhof, GJ, Goddard, ME and Pryce, JE (2022) Sharing of either phenotypes or genetic variants can increase the accuracy of genomic prediction of feed efficiency. **Genetics Selection Evolution** 54, 1-17.
- Brookes, AJ (1999) The essence of SNPs. **Gene** 234, 177-186.
- Buaban, S, Puangdee, S, Duangjinda, M and Boonkum, W (2020) Estimation of genetic parameters and trends for production traits of dairy cattle in Thailand using a multiple-trait multiple-lactation test day model. **Asian-Australasian Journal of Animal Science** 33, 1387-1399.
- Calus, MPL, de Haas, Y and Veerkamp, RF (2013) Combining cow and bull reference populations to increase accuracy of genomic prediction and genome-wide association. **Journal of Dairy Science** 96, 6703-6715.
- Christensen, OF, Madsen, P, Nielsen, B, Ostersen, T and Su, G (2012) Single-step methods for genomic evaluation in pigs. **Animal** 6, 1565-1571.

- Cobuci, JA, Costa, CN, Braccini Neto, J and Freitas, AF (2011) Genetic parameters for milk production by using random regression models with different alternatives of fixed regression modeling. **Revista Brasileira de Zootecnia** 40, 557-567.
- Cobuci, JA, Euclides, RF, Costa, CN, Torres, RA, Lopes, OS and Pereira, CS (2007) Genetic evaluation for persistency of lactation in Holstein cows using a random regression model. **Genetics and Molecular Biology** 30, 349-355.
- El Faro, L, Cardoso, VL and Albuquerque, LG (2008) Variance component estimates applying random regression models for test-day milk yield in Caracu heifers (*Bos taurus* Artiodactyla, Bovidae). **Genetics and Molecular Biology** 31, 665-673.
- Flores, EB and Van der Werf, J (2015) Random regression test day models to estimate genetic parameters for milk yield and milk components in Philippine dairy buffaloes. **Journal of Animal Breeding and Genetics** 132, 289-300.
- García-Ruiz, A, Cole, JB, VanRaden, PM, Wiggans, GR, Ruiz-López, FJ and Tassell, PV (2016) Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. **Proceedings of the National Academy of Sciences** 113, E3995-E4004.
- González-Herrera, LG, Pereira, RJ, El Faro, L and Albuquerque, LG (2022) Genetic evaluation of lactation persistency in the Gyr breed by using a two-trait random regression model. **Animal Production Science** 62, 216-224.
- Guarini, AR, Lourenco, DAL, Brito, LF, Sargolzaei, M, Baes, CF, Migliot, F, Misztal, I and Schenkel, FS (2018) Comparison of genomic predictions for lowly heritable traits using multi-step and single-step genomic best linear unbiased predictor in Holstein cattle. **Journal of Dairy Science** 101, 8076-8086.
- Gutierrez-Reinoso, MA, Aponte, PM and Garcia-Herreros, M (2021) Genomic Analysis, Progress and Future Perspectives in Dairy Cattle Selection: A Review. **Animals** 11, 1-21.
- Hayes, BJ (2013) Overview of statistical methods for genome-wide association studies (GWAS). **Methods in Molecular Biology** 1019, 146-169.

- Hayes, BJ, Bowman, PJ, Chamberlain, AJ and Goddard, ME (2009) Invited review: Genomic selection in dairy cattle: Progress and challenges. **Journal of Dairy Science** 92, 433-443.
- Henderson, CR (1963) Selection Index and Expected Genetic Advance. **Statistical Genetics and Plant Breeding** 982, 1-23.
- Henderson, CR (1973) Sire evaluation and genetic trends. **Journal of Animal Science** 1973, 10-41.
- Henderson, CR (1982) Analysis of Covariance in the Mixed Model: Higher-Level, Nonhomogeneous, and Random Regression. **International Biometric Society** 38, 623-640.
- Iung, LHS, Petrini, J, Ramírez-Díaz, J, Salvian, M, Ravadoscki, GA, Pilonett, F, Dauria, BD, Machado, PF, Coutinho, LL, Wiggans, GR and Mourão, GB (2018) Genome-wide association study for milk production traits in a Brazilian Holstein population. **Journal of Dairy Science** 102, 5305-5314.
- Jamrozik, J, Kistemaker, GJ, Dekkers, JCM and Schaeffer, LR (1997) Comparison of Possible Covariates for Use in a Random Regression Model for Analyses of Test Day Yields. **Journal of Dairy Science** 80, 2550-2556.
- Jamrozik, J and Schaeffer, LR (1997) Estimates of genetic parameters for a test day model with random regressions for production of first lactation Holsteins. **Journal of Dairy Science** 80, 762-770.
- Jattawa, D, Elzo, MA, Koonawootrittriron, S and Suwanasopee, T (2021) Genomic-polygenic evaluations using random regression models with Legendre polynomials and linear splines for milk yield and fat percentage in the Thai multibreed dairy cattle population. **Livestock Science** 251.
- Kang, H, Ning, C, Zhou, L, Zhang, S, Yan, Q and Liu, JF (2018) Short communication: Single-step genomic evaluation of milk production traits using multiple-trait random regression model in Chinese Holsteins. **Journal of Dairy Science** 101, 1-7.
- Khanal, P, Gaddis, KLP, Vandehaar, MJ, Weigel, KA, White, HM, F Peñagaricano, F, Koltes, JE, Santos, JEP, Baldwin, RL, Burchard, JF, Dürr, JW and Tempelman, RJ (2022) Multiple-trait random regression modeling of feed efficiency in US Holsteins. **Journal of Dairy Science** 105, 5954-5971.

- Kirkpatrick, M and Heckman, NA (1989) Quantitative genetic model for growth, shape and other infinite-dimensional characters. **Journal of Mathematical Biology** 27, 429-450.
- Kirkpatrick, M, Hill, WG and Thompson, R (1994) Estimating the covariance structure of traits during growth and ageing, illustrated with lactation in dairy cattle. **Genetics Research** 64, 57-69.
- Kirkpatrick, M, Lofsvold, D and Bulmer, M (1990) Analysis of the inheritance, selection and evolution of growth trajectories. **Genetics** 124, 979-993.
- Koivula, M, Strandén, I, Aamand, GP and Mantysaari, EA (2021) Practical implementation of genetic groups in single-step genomic evaluations with Woodbury matrix identity-based genomic relationship inverse. **Journal of Dairy Science** 104, 10049-10058.
- Koivula, M, Strandén, I, Poso, J and Mantysaari, EA (2015) Single-step genomic evaluation using multitrait random regression model and test-day data. **Journal of Dairy Science** 98, 2775-2784.
- Laird, NM and Ware, JH (1982) Random-Effects Models for Longitudinal Data. **International Biometrical Society** 38, 963-974.
- Lande, R and Thompson, R (2000) Efficiency of Marker-Assisted Selection in the Improvement of Quantitative Traits. **Genetics Society of America** 124, 743-756.
- Legarra, A, Aguilar, I and Misztal, I (2009) A relationship matrix including full pedigree and genomic information. **Journal of Dairy Science** 92, 4656-4663.
- Lidauer, M, Mantysaari, EA and Strandén, I (2003) Comparison of test-day models for genetic evaluation of production traits in dairy cattle. **Livestock Production Science** 79, 73-86.
- Lourenco, DAL, Misztal, I, Tsuruda, S, Aguilar, I, Ezra, E, Ron, M, Shirak, A and Weller, JI (2014) Methods for genomic evaluation of a relatively small genotyped dairy population and effect of genotyped cow information in multiparity analyses. **Journal of Dairy Science** 97, 1742-1752.
- Mbuthia, JM, Mayer, M and Reinsch, N (2021) Modeling heat stress effects on dairy cattle milk production in a tropical environment using test-day records and random regression models. **The International Journal of Animal Biosciences** 15, 1-10.

- Mehrban, H, Lee, DH, Naserkheil, M, Moradi, MH and Ibanez-Escriche, N (2019) Comparison of conventional BLUP and single-step genomic BLUP evaluations for yearling weight and carcass traits in Hanwoo beef cattle using single trait and multi-trait models. **PLoS ONE** 14, 1-13.
- Meuwissen, THE, Hayes, BJ and Goddard, ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. **Genetics Society of America** 157, 1819-1829.
- Meyer, K (1998) "DXMRR" – A program to estimate covariance functions for longitudinal data by restricted maximum likelihood. **Animal Genetics and Breeding Unit University of New England, Armidale, Australia.**
- Meyer, K (1998) Estimating covariance functions for longitudinal data using a random regression model. **Genetics Selection Evolution** 30, 221-240.
- Meyer, K (2000) Random regressions to model phenotypic variation in monthly weights of Australian beef cows. **Livestock Production Science** 65, 19-38.
- Meyer, K (2004) Scope for a random regression model in genetic evaluation of beef cattle for growth. **Livestock Production Science** 86, 69-83.
- Meyer, K (2005) Random regression analyses using B-splines to model growth of Australian Angus cattle. **Genetics Selection Evolution** 37, 473.
- Meyer, K and Hill, WG (1991) Mixed model analysis of a selection experiment for food intake in mice. **Genetics Research** 57, 71-81.
- Meyer, K and Hill, WG (1997) Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by restricted maximum likelihood. **Livestock Production Science** 47, 185-200.
- Meyer, K and Kirkpatrick, M (2005) Restricted maximum likelihood estimation of genetic principal components and smoothed covariance matrices. **Genetics Selection Evolution** 37, 1-30.
- Misztal, I, Bradford, HL, Lourenco, DAL, Tsuruta, S, Masuda, Y, Legarra, A and Lawlor, TJ (2017) Studies on inflation of GEBV in single-step GBLUP for type. **Interbull Bulletin** 51, 38–42.
- Misztal, I, Legarra, A and Aguilar, I (2009) Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. **Journal of Dairy Science** 92, 4656-4663.

- Misztal, I, Legarra, A and Aguilar, I (2014) Using recursion to compute the inverse of the genomic relationship matrix. **Journal of Dairy Science** 97, 3943-3952.
- Mrode, R, Ojango, JMK, Okeyo, AM and Mwacharo, JM (2019) Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: Current status and future prospects. **Frontiers in Genetics** 9, 1-11.
- Nejati-Javaremi, A, Smith, C, Gibson, JP (1997) Effect of total allelic relationship on accuracy of evaluation and response to selection. **Journal of Animal Science** 75, 1738-1745.
- Oliveira, HR, Lourenco, DAL, Masuda, Y, Misztal, I, Tsuruta, S, Jamrozik, J, Brito, LF, Silva, FF and Schenkel, FS (2019) Application of single-step genomic evaluation using multiple-trait random regression test-day models in dairy cattle. **Journal of Dairy Science** 102, 1-13.
- Pereira, RJ, Bignardi, AB, El Faro, L, Verneque, RS, Vercesi Filho, AE and Albuquerque, LG (2013) Random regression models using Legendre polynomials or linear splines for test-day milk yield of dairy Gyr (*Bos indicus*) cattle. **Journal of Dairy Science** 96, 565-574.
- Piccoli, ML, Brito, LF, Braccini, J, Brito, FV, Cardoso, FF, Cobuci, JA, Sargolzaei, M and Schenkel, FS (2018) A comprehensive comparison between single- and two-step GBLUP methods in a simulated beef cattle population. **Canadian Journal of Animal Science** 98, 565-571.
- Pool, MH, Janss, LL and Meuwissen, TH (2000) Genetic parameters of Legendre polynomials for first parity lactation curves. **Journal of Dairy Science** 83, 2640-2649.
- Pryce, JE and Daetwyler, HD (2012) Designing dairy cattle breeding under genomic selection: a review of international research. **Animal Production Science** 52, 107-114.
- Ptak, E and Schaeffer, LR (1993) Use of test day yield for genetic evaluation of dairy sires and cows. **Livestock Production Science** 34, 23-34.
- Scaletz, DCB, Fragomeni, BO, Santos, DCC, Passafaro, TL, Alencar, MM and Toral, FLB (2018) Random regression models with B-splines to estimate genetic parameters for body weight of young bulls in performance tests. **Revista Brasileira de Zootecnia** 47, 1-9.

- Schaeffer, LR (2004) Application of random regression models in animal breeding. **Livestock Production Science** 86, 35-45.
- Schaeffer, LR (2006) Strategy for applying genome-wide selection in dairy cattle. **Journal of Animal Breeding and Genetics** 123, 218-223.
- Schaeffer, LR and Dekkers, JC (1994) Random regressions in animal models for test-day production in dairy cattle. **World Congress on Genetics Applied to Livestock Production** 5, 443–446.
- Schultz, NE and Weigel, KA (2018) An improved genomic prediction model in populations featuring shared environments and familial relatedness. **Proceedings of the World Congress of Genetics Applied to Livestock Production**, Auckland, New Zealand.
- Steri, R, Moioli, B, Catillo, G, Galli, A and Buttazzoni, L (2018) Genome-wide association study for longevity in the Holstein cattle population. **Animal** 3, 1-8.
- VanRaden, PM (2008) Efficient methods to compute genomic predictions. **Journal of Dairy Science** 91, 4414-4423.
- VanRaden, PM (2012) Avoiding bias from genomic pre-selection in converting daughter information across countries. **Interbull Meeting** 45, 1–5.
- VanRaden, PM, Tassell, CPV, Wiggans, GR, Sonstegard, TS, Schnabel, RD, Taylor, JF and Schenkel, FS (2009) Invited review: Reliability of genomic predictions for North American Holstein bulls. **Journal of Dairy Science** 92, 16-24.
- Venot, E, Boichard, D, Ducrocq, V, Croiseau, P, Fritz, S, Baur, A, Saintilan, R, Tribout, T, Sanchez, MP, Boulesteix, P, Astruc, JM, Legarra, A, Robert-Granié, C, Carillier, C, Palhière, I, Tortereau, F, Rupp, R, Larroque, H, Loywyck, V and Mattalia, S (2016) French genomic experience: genomics for all ruminant species. **Interbull Meeting** 48, 1-6.
- Wang, H, Misztal, I, Aguilar, I, Legarra, A and Muir, WM (2012) Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetic Research Cambridge** 94, 73-83.
- Wiggans, GR, Cole, JB, Hubbard, SM and Sonstegard, TS (2017) Genomic selection in dairy cattle: The USDA experience. **Annual Review of Animal Biosciences** 5, 309-327.

CHAPTER 2 – GENETIC PARAMETER ESTIMATES FOR TEST-DAY MILK YIELD IN GIROLANDO CATTLE USING RANDOM REGRESSION MODELS

Abstract - The Girolando breed, which results from the crossbreeding between the Gir and Holstein breeds, plays a crucial role in Brazilian dairy farming, being responsible for about 80% of the country's milk production. The aim of this study was to evaluate the milk production on the control day (TDMY) in Girolando cows. A total of 480,364 TDMY records from 76,688 cows in their first lactation, originating from 913 herds with calvings occurring between 1991 and 2021, were analyzed. The animals belonged to nine genetic groups, composed of different combinations of the Holstein (H) and Gir (G) groups: $1/2H + 1/2G$, $1/4H + 3/4G$, $1/8H + 7/8G$, $3/4H + 1/4G$, $3/8H + 5/8G$, $5/8H + 3/8G$, and $7/8H + 1/8G$. The analysis was performed using random regression models (RRM), considering fixed effects of heterozygosity and age at calving (linear and quadratic) as covariates, along with random effects of additive genetic and permanent environmental effects of the cow. Nine distinct models were evaluated and compared based on the Deviance Information Criterion (DIC), and the estimation of variance components and breeding values (EBVs) was performed using the Gibbs sampling method, utilizing the GibbsF90+ software. The results showed an average production of 22.3 kg/day, with a peak of 23.8 kg/day between days 91 and 120 of lactation. The most suitable model was the fourth-order Legendre orthogonal polynomial, with heterogeneous residual variances in five classes. The average heritability was 0.56, reaching 0.66 in TDMY 5, and the repeatability was 0.84. The high genetic correlations between adjacent periods (0.64 to 0.99) indicate a strong genetic association between these periods, especially those that are closer together. It is concluded that the TDMY model provides a more detailed genetic evaluation of the Girolando breed, surpassing traditional models based on the cumulative 305-day production, highlighting the effectiveness of using RRM in the genetic improvement of dairy cattle in this breed.

Keywords: Genetic parameters, Girolando, model comparison, test-day milk yield.

Resumo - A raça Girolando, resultado do cruzamento entre as raças Gir e Holandesa, desempenha um papel crucial na pecuária leiteira brasileira, sendo responsável por cerca de 80% da produção de leite do país. O objetivo deste estudo foi avaliar a produção de leite no dia do controle (PLDC) em vacas Girolando. Foram analisados um total de 480.364 registros de PLDC de 76.688 vacas em primeira lactação, originárias de 913 rebanhos, com partos ocorridos entre 1991 e 2021. Os animais pertenciam a nove grupos genéticos, compostos por diferentes combinações das raças Holandesa (H) e Gir (G): $1/2H + 1/2G$, $1/4H + 3/4G$, $1/8H + 7/8G$, $3/4H + 1/4G$, $3/8H + 5/8G$, $5/8H + 3/8G$ e $7/8H + 1/8G$. A análise foi realizada utilizando modelos de regressão aleatória (MRA), considerando os efeitos fixos da heterozigosidade e da idade ao parto (linear e quadrático) como covariáveis, além dos efeitos aleatórios genético aditivo e ambiental permanente da vaca. Nove modelos distintos foram avaliados e comparados com base no Critério de Informação da Deviance (DIC), e a estimação dos componentes de variância e dos valores genéticos (VGs) foi realizada por meio do método de amostragem de Gibbs, utilizando o software GibbsF90+. Os resultados mostraram uma produção média de 22,3 kg/dia, com pico de 23,8 kg/dia entre os dias 91 e 120 de lactação. O modelo mais adequado foi o polinômio ortogonal de Legendre de quarta ordem, com variâncias residuais heterogêneas em cinco classes. A herdabilidade média foi de 0,56, atingindo 0,66 no PLDC 5, e a repetibilidade foi de 0,84. As altas correlações genéticas entre períodos adjacentes (0,64 a 0,99) indicam uma forte associação genética entre esses períodos, especialmente aqueles mais próximos. Conclui-se que o modelo de PLDC proporciona uma avaliação genética mais detalhada da raça Girolando, superando os modelos tradicionais baseados na produção acumulada de 305 dias, destacando a eficácia do uso de MRA no melhoramento genético de bovinos leiteiros dessa raça.

Palavras-chave: Comparação de modelos, Girolando, parâmetros genéticos, produção de leite no dia do controle

1. INTRODUCTION

In tropical regions, using crossbred animals is a viable alternative for milk production systems (Marchioretto et al., 2023). Girolando cattle, a cross between Gir and Holstein breeds, are widely used on dairy farms due to their high adaptability to the climate, milk production, and reproductive efficiency in tropical systems (Canaza-Cayo et al., 2014). Currently, most Brazilian dairy herds consist of Girolando animals with Holstein proportions ranging between 1/4 to 7/8; while the 5/8 group accounts for 80% of the country's milk production (Canaza-Cayo et al., 2016).

Genetic improvement programs for Zebu dairy cattle use cumulative milk production up to 305 days of lactation as the selection criterion, estimated from test-day milk yield (TDMY) collected throughout lactation. This approach is particularly relevant for Girolando cattle, where genetic evaluations are conducted using an animal model based on cumulative milk production during the first 305 days of lactation (Canaza-Cayo et al., 2016; Daltro et al., 2021; Barbero et al., 2022).

The TDMY, an alternative to cumulative milk yield over 305 days, has been used as a selection criterion in European-origin cattle. Predicting milk yield based on cumulative production over 305 days assumes that milk yield is a representative average of each cow's lactation period, which may not be true (Yamazaki et al., 2018).

Some benefits of adopting TDMY as a selection criterion include a better definition of contemporary groups and removing environmental effects, leading to more accurate genetic evaluations for cows and bulls (Schaeffer et al., 2000). Among the genetic evaluation methodologies that use TDMY, random regression models (RRMs) are generally more parsimonious and effectively adjust for the cows' permanent environmental effects throughout lactation. Covariances are modeled using repeated records, resulting in fewer parameters than multi-trait models (Pereira et al., 2019; Cole et al., 2023).

One benefit of applying RRM to TDMY is that it partitions phenotypic variance based on the lactation curve, which can improve the accuracy of breeding values for milk yield in dairy cattle (Schaeffer et al., 2000). High and positive genetic correlations between TDMY suggest that selecting for increased

milk yield at a specific lactation stage can positively impact other lactation periods, resulting in greater genetic gains (Pereira et al., 2019). This study aimed to estimate variance components and genetic parameters for TDMY using RRM and predict breeding values for the trait in Girolando dairy cattle and assist in the breed's selection process.

2. MATERIALS AND METHODS

2.1. DATA

Embrapa Dairy Cattle and the Brazilian Association of Girolando Breeders provided the phenotypic data from dairy TDMY of Girolando cows. To estimate variance components, 480,364 TDMY records were used from 76,688 first lactations (only to the first calving) of Girolando cows, with calving ages between 22 and 48 months and calving between 1991 and 2021, distributed across 913 herds. Milk production per TDMY varied between 4 and 40 kg, and the TDMY were recorded between the 5th and the 305th day of lactation for cows with more than three measured TDMY. The proportions of cows according to the number of milk TDMY were: 8.5%, 11.8%, 12.0%, 11.9%, 11.6%, 10.8%, 10.1%, 9.1%, 7.6%, and 6.6% for 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 TDMY, respectively.

Animals formed contemporary groups (CG) from the same herd, year of calving, test-day date (day-month-year), and genetic group. These effects were previously tested and retained in the CG definition based on their significance and contribution to model fit. CGs with fewer than three animals were eliminated. The genetic groups (Table 1) were identified as follows: Holstein (H); Gir (G); $1/2H + 1/2G$; $1/4H + 3/4G$; $1/8H + 7/8G$; $3/4H + 1/4G$; $3/8H + 5/8G$; $5/8H + 3/8G$; and $7/8H + 1/8G$. After identifying the pedigree relationships for all animals over three generations, the pedigree file comprised 252,389 animals, including 8,312 bulls and 67,549 cows.

Table 1 - Number of observations by genetic group

Genetic Groups	Number of Animals
H	569
G	158
$1/2H + 1/2 G$	33,585
$1/4H + 3/4G$	1,055

1/8H + 7/8G	131
3/4H + 1/4G	25,754
3/8H + 5/8G	1,149
5/8H + 3/8G	9,462
7/8H + 1/8G	4,825

Heterosis (HET) was calculated as follows: $HET = ((HP * GM) + (HM * GP))$, where HP is the Holstein breed composition from paternal contribution; GM is the Gir breed composition from maternal contribution; HM is the Holstein breed composition from maternal contribution; GP is the Gir breed composition from paternal contribution. HET values ranged from 0.005 to 1.

2.2. ANALYSIS MODELS

The milk TDMY were analyzed using RRM. Heterozygosity and the linear and quadratic effects of the cow's age at calving were considered fixed effects, and the average lactation curve within each genetic group was treated as a fixed curve. The additive genetic and permanent environmental effects of animal were modeled using Legendre orthogonal polynomials. The CG effect was considered random. The Legendre orthogonal polynomials were tested in orders two, three, or four, and residual variances were modeled as homogeneous or heterogeneous with five or ten classes.

The general random regression model used in the analyses is represented by:

$$y_{ijklm} = GC_i + \sum_{n=1}^2 b_n x_k^n + \sum_{o=1}^c \beta_{jo} \varphi_o(t) + \sum_{o=1}^c \alpha_{ko} \varphi_o(t) + \sum_{o=1}^c \rho_{ko} \varphi_o(t) + e_{ijklm}$$

in which y_{ijklm} is the m-th observation on the lactation day t , of animal l , with heterozygosity k , belonging to genetic group j , and CG i ; HET_k is the effect of the k-th heterozygosity; b_n is the regression coefficient for the linear ($n=1$) and quadratic ($n=2$) effects of the age x_k of the cow at calving on TDMY (in months); β_{jo} is the set of regression coefficients for the mean trajectory of genetic group j ; $\varphi_o(t)$ is the covariate of the regression function according to the lactation day t ; α_{ko} , ρ_{ko} are the sets of random regression coefficients for the mean trajectory of the random genetic additive effect and the permanent environmental effect of the cow, for each cow l ; and e_{ijklm} is the error associated with the m-th observation

on lactation day t , of animal l , with heterozygosity k , belonging to genetic group j , and CG i . The number of regression coefficients c varies according to the functions used in the genetic-additive and permanent environmental random regressions of the cow.

Nine models were tested to evaluate the orders of the Legendre orthogonal polynomials and the classes of residual variances (homogeneous and heterogeneous). In the case of heterogeneous residual variances, these were organized into classes according to the lactation days. When five classes were used, the lactation days were divided as follows: 5-60, 61-120, 121-180, 181-240, and 241-305 lactation days, with each class containing two milking TDMY. The definition of 10 classes was made so that each one included a milking TDMY: 5-30; 31-60; 61-90; 91-120; 121-150; 151-180; 181-210; 211-240; 241-270; and 271-305 lactation days.

When evaluating the different structures of residual variances, models were used in which the random regressions, both for the genetic additive random effect and for the permanent environmental random effect of the cow, were described using Legendre orthogonal polynomials of the same degree as the random effects.

The model assumptions are:

$$E[y] = X\beta$$

$$var \begin{bmatrix} \delta \\ a \\ p \\ e \end{bmatrix} = \begin{bmatrix} I \otimes GC_0 & 0 & 0 & 0 \\ 0 & A \otimes G_0 & 0 & 0 \\ 0 & 0 & I \otimes P_0 & 0 \\ 0 & 0 & 0 & R \otimes R_0 \end{bmatrix}$$

where $E[y]$ is the expectation of y ; y is the vector of TDMY; X is the incidence matrix for the fixed effects of the vector β ; δ , a , p , and e are the vectors of the random coefficients for the CG effects, additive genetics, permanent environmental effects of the cow, and residuals; CG_0 , G_0 , and P_0 are the covariance matrices for the random regression coefficients for the CG effects, genetic additive, and permanent environmental effects of the cow, respectively; R_0 is the covariance matrix for the residuals; I is an identity matrix with dimensions equal to the number of levels of the effects; A is the pedigree relationship matrix; R is the diagonal matrix of residual variances; $\otimes$ is the Kronecker product.

In matrix notation, the model is represented by:

$$y = Xb + Za + Wp + Hc + e$$

where y is the vector of observations; b is the vector of solutions for the fixed effects; a is the vector of random coefficients for the genetic additive effects; p is the vector of random coefficients for the permanent environmental effects of the cow; c is the vector of random coefficients for the CG; e is the vector of residuals; and X, Z, W, H and are the corresponding incidence matrices.

The assumptions for the prior distributions for the unknown parameters were considered according to:

$y|b, a, p, \sigma_{e_1}^2, \dots, \sigma_{e_x}^2 \sim NMV(Xb + Za + Wp, R), R = \text{diag}\{\sigma_{e_x}^2\}$, with x going from 1 to n , where n is the number of residual variance classes.

$a|\Lambda_a \sim NMV(0, G)$, with $G = \Lambda_a \otimes A$, where A is the pedigree relationship matrix, and Λ_a is the matrix with the covariances between the genetic-additive random regression coefficients; $p|\Lambda_p \sim NMV(0, P)$, with $P = \Lambda_p \otimes I$, where I is the identity matrix and Λ_p is the matrix with the covariances between the random regression coefficients for the permanent environmental effects of the cow; $\Lambda_a|v_a, S_a^2 \sim W^{-1}(v_a, v_a S_a^2)$, $\Lambda_p|v_p, S_p^2 \sim W^{-1}(v_p, v_p S_p^2)$ and $R|v_e, S_e^2 \sim W^{-1}(v_e, v_e S_e^2)$, in which v_a, S_a^2 ; v_p, S_p^2 ; v_e, S_e^2 represent the degrees of confidence and the prior values for the covariances of the cow's genetic-additive and permanent environmental random regression coefficients and residual variances, respectively; $b \propto$ is a constant. NMV and W^{-1} refer to the multivariate normal and inverse Wishart distributions.

2.3. MODEL COMPARISON

To compare the tested models, the DIC was used:

$$DIC = \bar{D}(\theta) + p_D = 2\bar{D}(\theta) - D(\bar{\theta})$$

where $\bar{D}(\theta) = E_{\theta|y}[D(\theta)]$, i.e., the posterior expectation of the Bayesian deviance, and $D(\theta) = -2\log p(y|\theta)$, which represents the model fit quality. The effective number of parameters represents the penalty for increasing model complexity: $p_D = \bar{D}(\theta) - D(\bar{\theta})$, where θ is the parameter vector of the model and $D(\bar{\theta})$ is the Bayesian deviance evaluated at the posterior means of the

parameters. Models with lower DIC values were preferred (Spiegelhalter et al., 2002).

2.4. ESTIMATION OF VARIANCE COMPONENTS AND BREEDING VALUES

The computational analyses were performed using a Bayesian method via Gibbs sampling, employing the GibbsF90+ program (Misztal et al., 2014). Inferences about the parameters were made based on their corresponding posterior marginal distributions. For each evaluated model, a chain of 1,000,000 samples was generated; conservatively, the first 200,000 samples were discarded and considered burn-in. Convergence was monitored by graphical inspection of the samples versus iterations, as well as by the criteria proposed by Heidelberger & Welch (1983) and Geweke (1992), calculated using the BOA package in R (Smith, 2007). A sample discard interval of 50 iterations was considered, leaving 16,000 samples from which the inferences were made. Genetic parameters for lactation days were estimated in 30-day intervals for each sample.

Posterior means were used as point estimates of the (co)variance components. The cow's additive genetic, permanent environmental, phenotypic, and residual (co)variance matrices for all lactation days (from 5 to 305 days) were obtained following the approach applied by Druet et al. (2003). Heritabilities were defined as the ratio between additive genetic and phenotypic variance for each lactation day. Repeatabilities were defined as the ratio between the sum of additive genetic variance and permanent environmental variance of the cow and phenotypic variance for each lactation day. Correlations between lactation days i and j were calculated as the covariance $cov(i, j)$ ratio to the square root of the product of the variances of lactation days i and j .

The solutions for the genetic effects on each lactation day were used to calculate the estimated breeding values (EBVs):

$$EBV_i^j = \varphi_j \hat{a}_i$$

where $\hat{a}_i$ is the vector of solutions estimated by the orthogonal regression coefficients of animal i , and φ_j is the vector of orthogonal coefficients of the j -th TDMY. The EBVs were used to rank the animals, and the top 30 males and top 30 females were selected for Spearman's rank correlation analysis. The analysis

was performed separately for males and females and across four distinct genetic groups, chosen based on data availability.

3. RESULTS AND DISCUSSION

The overall mean and standard deviation for the milk production TDMY were 22.3 kg and 7.33 kg, respectively. The lowest mean milk yields were observed for TDMY 1 (20.3 ± 6.39 kg), TDMY 9 (21.6 ± 6.95 kg), and TDMY 10 (20.6 ± 6.87 kg). The peak milk production was achieved during TDMY 4, between 91 and 120 days in lactation, with a mean of 23.8 ± 7.62 kg. After TDMY 5 (23.6 ± 7.64 kg), milk production decreased in subsequent TDMY (Table 2). All analyses showed convergence, indicating model stability and reliability of the estimated parameters.

Table 2 - Number of observations, observed means, standard deviations, and coefficients of variation for milk production in TDMY (1 to 10).

TDMY	Days in Milk	Number of Observations	Mean $\pm$ Standard Deviation (kg)	CV (%)
1	5 to 30	40,763	20.3 ± 6.39	31.5
2	31 to 60	56,667	22.4 ± 7.01	31.2
3	61 to 90	57,639	23.4 ± 7.43	31.7
4	91 to 120	57,203	23.8 ± 7.62	32.1
5	121 to 150	55,654	23.6 ± 7.64	32.4
6	151 to 180	51,793	23.2 ± 7.52	32.4
7	181 to 210	48,479	22.7 ± 7.35	32.4
8	211 to 240	43,690	22.2 ± 7.16	32.3
9	241 to 270	36,608	21.6 ± 6.95	32.1
10	271 to 305	31,868	20.6 ± 6.87	33.4

A fixed regression for each of these genetic groups was considered in the analysis model (Figure 1) due to the discrepancies in milk production levels among the various genetic groups. Cows in the genetic group 1/2H + 1/2G had the highest mean milk production (23.2 ± 7.36 kg), while cows in the G genetic group had the lowest mean milk production (16.7 ± 4.47 kg).

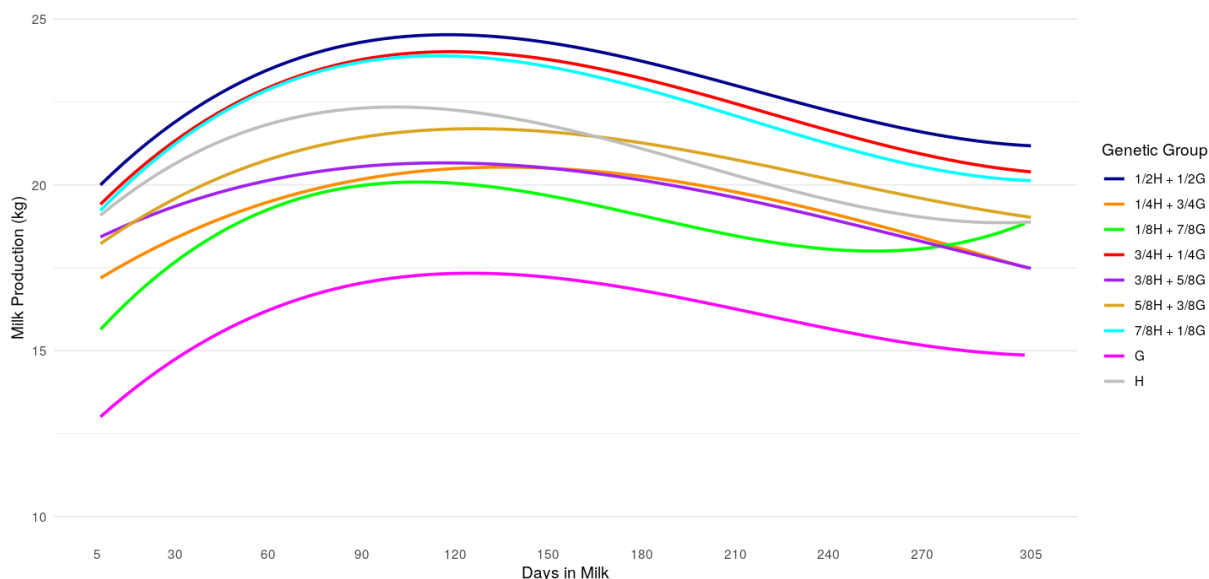


Figure 1. Observed means of milk production for different genetic groups of Holstein (H) and Gir (G) during lactation (5 to 305 days).

Animals with a higher proportion of Holstein (H, $1/2H + 1/2G$, $3/4H + 1/4G$, $5/8H + 3/8G$, and $7/8H + 1/8G$) produced more milk than animals with a higher proportion of Gir (G, $1/4H + 3/4G$, $1/8H + 7/8G$, and $3/8H + 5/8G$). As a result, genetic groups with a higher proportion of Holstein produced, on average, 17.2 kg more than groups with a higher proportion of Gir.

Based on the Deviance Information Criterion (DIC) for the various models tested (Table 3), models with fourth-order Legendre orthogonal polynomials provided the best fit for the data in this study. As shown in the literature (Pereira et al., 2013; Canaza-Cayo et al., 2015; Li et al., 2020; González-Herrera et al., 2021), models with the highest tested order of Legendre orthogonal polynomials exhibited the lowest DIC values. Among the fourth-order Legendre orthogonal polynomial models, heterogeneous residual variances with five classes had the lowest DIC value. Thus, five classes were sufficient to adjust this effect for this dataset. This demonstrated that residual variances were not uniform throughout lactation. Similar results were reported by Pereira et al. (2013) and Júnior et al. (2023).

Table 3. Comparison between models and Deviance Information Criterion (DIC) values for the analysis of TDMY

Order of Legendre Orthogonal Polynomial	Number of Residual Variance Classes	DIC
2	1	1,991,768.23
2	5	1,992,128.33
2	10	1,993,018.44
3	1	1,968,568.30
3	5	1,969,347.96
3	10	1,966,177.33
4	1	1,957,926.08
4	5	1,957,884.95
4	10	1,958,061.36

Residual variances with one class are homogeneous, while those with more than one class are heterogeneous.

Li et al. (2020) state that selecting the best analysis model requires balancing the best fit with model complexity. Models with higher orders of Legendre orthogonal polynomials have more parameters to estimate, which can result in greater computational demand. Estimates of phenotypic and additive genetic variances slightly increased at the beginning of lactation. Estimates of the cow's permanent environmental variances peaked at the end of lactation, and the magnitude of the residual variance varies between 7.7 and 8.3 throughout the lactation period (Figure 2).

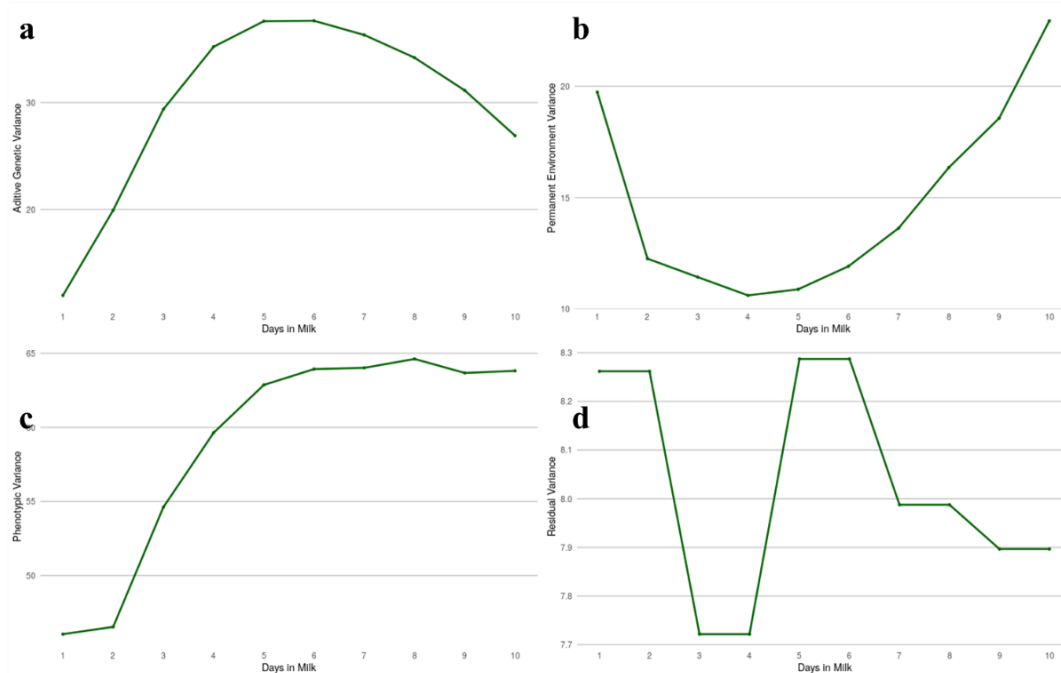


Figure 2. Mean estimates of additive genetic **(a)**, permanent environmental of the animal **(b)**, phenotypic **(c)**, and residual variances **(d)** over lactation.

In the present study, the permanent environmental effect of the cow variances were higher than the additive genetic variances during the early days of lactation. After this period, the additive genetic variances became more pronounced. Previous research with dairy cattle (Costa et al., 2008; El Faro et al., 2008; Canaza-Cayo et al., 2015; and Barbero et al., 2022) showed that estimates of genetic and phenotypic variances were higher at the beginning of lactation, compared to the permanent environmental variances of the cow, which were higher at the end of lactation. This may indicate that genetic and environmental factors affect the milk production curve differently (Marumo et al., 2022). RRM's have effectively estimated variances and identified those that best fit the milk production curve, as reported by Costa et al. (2008), El Faro et al. (2008), and Canaza-Cayo et al. (2015).

Lower heritability estimates at the beginning of lactation were caused by a smaller proportion of phenotypic variance attributable to additive genetic variance, possibly due to the smaller number of milk production observations during this period (Meyer, 1998). The average heritability estimate throughout the entire lactation period was 0.56 ± 0.11 , reaching its maximum value of 0.66 at TDMY 5 (121 to 150 days of lactation) and its minimum value (0.30) at TDMY 1 (Figure 3). Repeatability estimates varied similarly to heritability throughout lactation. The average repeatability estimate for the entire lactation period was 0.84 ± 0.03 , reaching its maximum value of 0.86 at TDMY 6, 7, 8, 9, and 10 (151 to 305 days of lactation) and its minimum value (0.79) at TDMY 1.

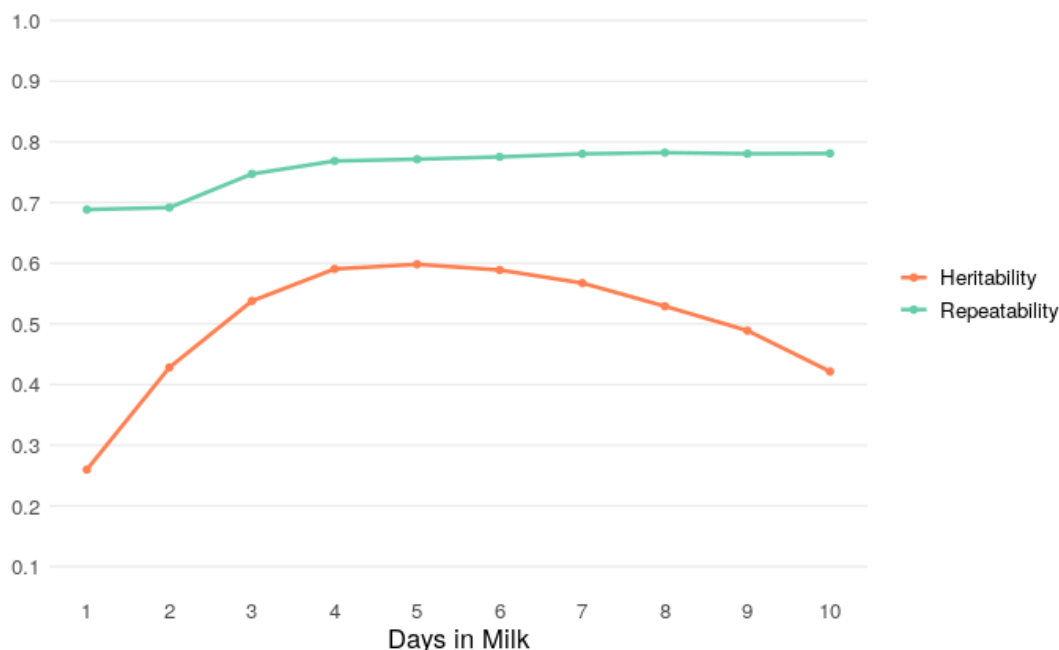


Figure 3. Mean estimates of heritability and repeatability during the lactation period.

In the literature, considerably variable estimates of heritability and repeatability for cumulative milk production up to 305 days have been obtained for Girolando animals. Canaza-Cayo et al. (2015) and Santos et al. (2021) reported estimates of 0.23 and 0.26, respectively, for milk production. On the other hand, Daltro et al. (2020) used an analysis model for P305 and obtained a lower estimate (0.14 ± 0.01). It is important to note that heritability estimates for milk production in Girolando cattle showed wide variation, with estimates ranging from 0.18 to 0.29 (Canaza-Cayo et al., 2015; Canaza-Cayo et al., 2016). Júnior et al. (2023) evaluated Girolando, Gir, and Holstein breeds and estimated heritabilities of 0.09, 0.17, and 0.12 and repeatabilities of 0.31, 0.35, and 0.28, respectively, for P305. Barbero et al. (2022) reported heritability and repeatability estimates of 0.29 and 0.94, respectively, for P305.

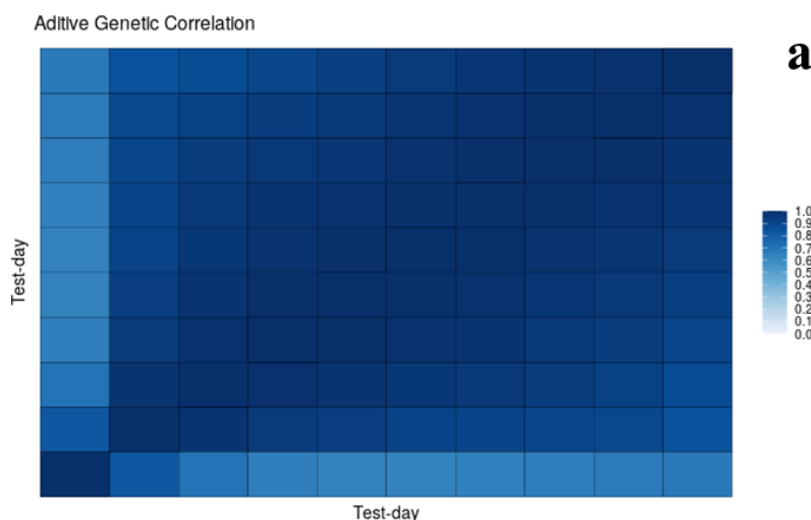
Costa et al. (2019) reported average repeatability estimates of 0.60 for TDMY in Holstein cattle. Buitenhuis and Poulsen (2023), evaluating TDMY of Danish Holstein, Jersey, and Red breeds, reported repeatability values of 0.48, 0.42, and 0.49, respectively, and heritability estimates of 0.16, 0.14, and 0.22, respectively. This study's heritability and repeatability estimates differed from those previously described in the literature for Girolando cattle. This may be explained by the larger number of observations used in the analyses and by the fact that the test-day model used can obtain more accurate estimates for the

entire lactation curve period compared to models that consider cumulative milk production up to 305 days (Ptak and Schaeffer, 1993; Lidauer et al., 2003; Canaza-Cayo et al., 2016).

The high heritability estimates suggest that the variation in milk production is largely determined by additive genetic factors, indicating that the selection process in the Girolando breed has been effective in improving milk production. This implies that most of the observed differences in milk production between individuals are attributed to additive genetic variation, highlighting the potential of genetic selection to drive improvements in milk yield across generations.

Another fact to consider is that heritability and repeatability estimates can be affected using genetic groups in the CG or as a fixed effect in the model (Canaza-Cayo et al., 2016). Scarso et al. (2017) and Visentin et al. (2017) concluded that about 40% of the variability in heritability and repeatability can be attributed to environmental effects. Repeatability is an important phenotypic parameter for longitudinal data because it predicts the animal's potential production based on previous productions (Cilek and Sahin, 2009).

The estimates of phenotypic correlations between milk production at different TDMY ranged from 0.30 to 0.85, with the highest magnitudes occurring between adjacent TDMY (Figure 4). Similar trends were observed for genetic correlations (0.64 to 0.99) and for cow's permanent environmental correlations (0.10 to 0.95). These last results indicate that persistent environmental effects have a lasting impact on milk production over time (Leclerc et al., 2008).



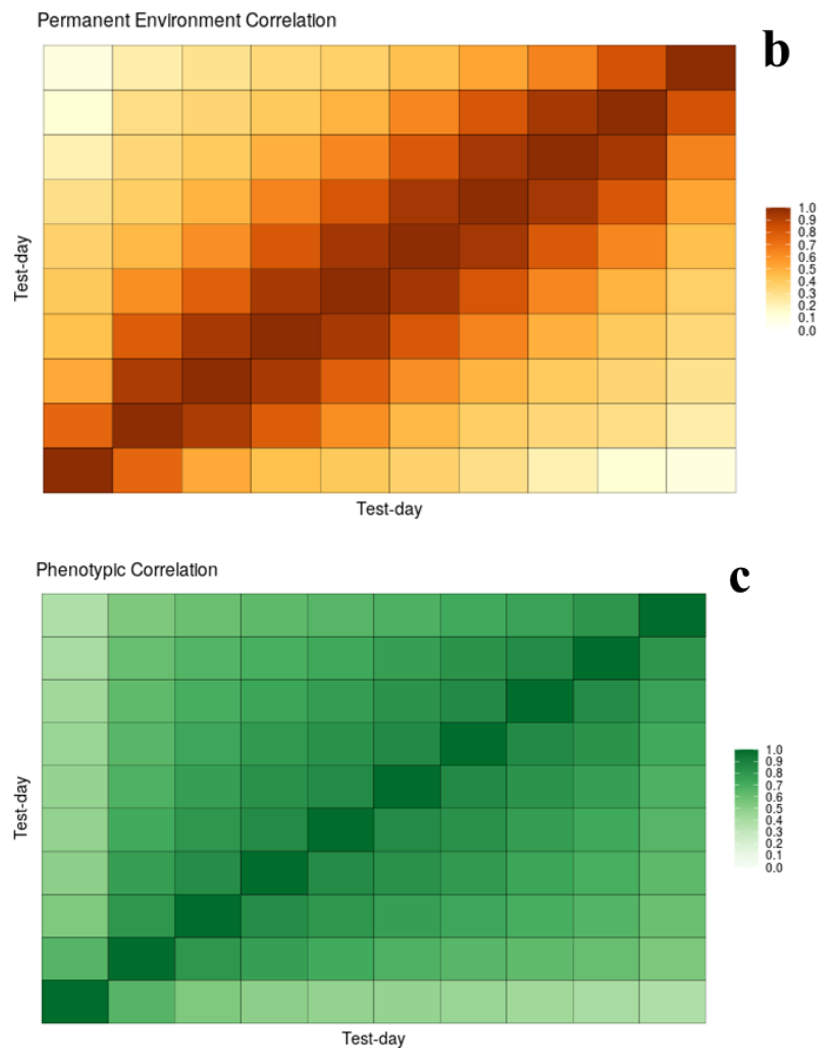


Figure 4. Mean estimates of additive genetic **(a)**, permanent environmental of the animal **(b)**, and phenotypic correlations **(c)** for milk production across TDMY.

Uribe and Lembeye (2020) estimated genetic correlations between milk production TDMY ranging from 0.37 to 1.00, with the highest estimates between adjacent TDMY. Bignardi et al. (2011) also estimated high genetic correlations between adjacent TDMY, which decreased as the distance between periods increased. Other authors have reported similar genetic correlations between the same periods (Cobuci et al., 2011; Gonzalez-Herrera et al., 2015; Singh et al., 2016; Santos et al., 2021).

Bignardi et al. (2011) and Uribe and Lembeye (2020) estimated negative genetic correlations between early and late TDMY, attributing these results to the difficulty in modeling milk yields on the initial TDMY caused by postpartum stress and negative energy balance during this period. A negative energy balance occurs when the energy required to maintain milk production exceeds the energy

from the cow's diet. Consequently, the cow resorts to its body fat reserves as an energy source (Goff and Horst, 1996; Coffey et al., 2002).

The Spearman correlations (Table 4) between different lactation periods ranged from 0.33 to 1.00 for females and from 0.82 to 1.00 for males in the genetic groups $1/2H + 1/2G$, $3/4H + 1/4G$, $5/8H + 3/8G$, and $7/8H + 1/8G$. The high coefficients for males indicate stability in the EBVs rankings for milk production across lactation phases. This suggests that performance in partial periods (start, middle, and end) is a good predictor of total production. For males, correlations were consistently high, with values close to 1.00 across all groups, reflecting almost absolute stability in production rankings throughout the analyzed periods.

Table 4. Spearman correlation coefficients between estimated breeding values for lactation periods in males (above the diagonal) and females (below the diagonal).

1/2H + 1/2G				
	5 – 105 DIM	106 – 205 DIM	206 – 305 DIM	5 – 305 DIM
5 – 105 DIM		0.91*	0.82*	0.94*
106 – 205 DIM	0.33 ^{NS}		0.94*	0.99*
206 – 305 DIM	0.41 ^{NS}	0.88*		0.94*
5 – 305 DIM	0.63*	0.89*	0.93*	
3/4H + 1/4G				
	5 – 105 DIM	106 – 205 DIM	206 – 305 DIM	5 – 305 DIM
5 – 105 DIM		0.91*	0.86*	0.93*
106 – 205 DIM	0.63*		0.98*	0.99*
206 – 305 DIM	0.36 ^{NS}	0.81*		0.98*
5 – 305 DIM	0.56*	0.96*	0.90*	
5/8H + 3/8G				
	5 – 105 DIM	106 – 205 DIM	206 – 305 DIM	5 – 305 DIM
5 – 105 DIM		0.94*	0.93*	0.95*
106 – 205 DIM	0.74*		0.99*	1.00*
206 – 305 DIM	0.58*	0.79*		0.99*
5 – 305 DIM	0.82*	0.94*	0.82*	
7/8H + 1/8G				
	5 – 105 DIM	106 – 205 DIM	206 – 305 DIM	5 – 305 DIM
5 – 105 DIM		0.98*	0.98*	0.98*
106 – 205 DIM	0.35 ^{NS}		1.00*	1.00*
206 – 305 DIM	0.44 ^{NS}	0.70*		1.00*
5 – 305 DIM	0.87*	0.60*	0.79*	

*: p-value <0.05; NS: not significant; DIM: days in milk.

On the other hand, females showed greater variability in correlations, especially in early lactation phases (Table 4). The correlations between total production and partial periods indicated that although early performance may be less consistent, correlations tend to strengthen in later periods. This suggests a strong correlation between the middle and end phases and total production, indicating that selecting animals based on performance during these phases may be more effective in identifying those with higher production potential (Cole et al., 2023; Boerner et al., 2023).

Using a test-day model in the genetic evaluation of Girolando cattle has resulted in more accurate estimates throughout the lactation curve than traditional cumulative models up to 305 days (Canaza-Cayo et al., 2015; Pereira et al., 2019). This finding is supported by the fact that the test-day model can estimate parameters and breeding values with higher accuracy, indicating greater

potential for evaluating milk production (Daltro et al., 2019). Furthermore, it was observed that using a test-day model results in higher accuracy in predicting breeding value estimates, leading to more efficient genetic evaluations (Pereira et al., 2019).

4. CONCLUSIONS

The results of this study indicate that RRM with fourth-order Legendre polynomials were effective in estimating variance components, parameters, and breeding values for milk production in Girolando cattle. The application of this model allowed greater precision in estimates throughout the lactation curve, highlighting it as a more suitable approach for genetic evaluation of milk production, especially when considering variation in the specific periods of early, middle, and late lactation.

The analyses revealed differences between breeding values for females and males across different genetic groups, reinforcing the importance of considering the distinct phases of the lactation curve in animal selection. These results highlight the need for appropriate genetic evaluation models, suggesting that RRM with a TDMY approach offers a more precise understanding of milk production. This is especially important as each genetic group shows a distinct lactation curve pattern, which must be considered to improve selection accuracy in the Girolando breed.

5. REFERENCES

- Barbero MMD, Fort NM, Schultz EB, de Melo ALP, Moura AM (2022) Estimation of genetic parameters in dairy production in Girolando cattle. **Brazilian Animal Science** 23:e-72300P.
- Bignardi AB, El Faro L, Santana Jr. ML, Rosa GJM, Cardoso VL, Machado PF, Albuquerque LG (2011) Bayesian analysis of random regression models using B-splines to model test-day milk yield of Holstein cattle in Brazil. **Livestock Science** 150:401-406.
- Boerner V, Nguyen TTT, Nieuwhof GJ (2023) Integration of Interbull's multiple across-country evaluation approach breeding values into the multiple-trait single-step random regression test-day genetic evaluation for the yield traits of Australian Red breeds. **J. Dairy Sci.** 106:1159-1167.
- Buitenhuis AJ, Poulsen NA (2023) Estimation of heritability for milk urea and genetic correlations with milk production traits in 3 Danish dairy breeds. **J. Dairy Sci.** 106:5562-5569.
- Canaza-Cayo AW, Cobuci JA, Lopes PS, Torres RA, Martins MF, Daltro DS, Silva MVGB (2016) Genetic trend estimates for milk yield production and fertility traits of the Girolando cattle in Brazil. **Livestock Science** 190:113-122.
- Canaza-Cayo AW, Lopes PS, Silva MVGB, Cobuci JA, Torres RA, Martins MF, Arbex WA (2014) Estrutura populacional da raça Girolando. **Ciência Rural** 44:2072-2077. <https://www.scielo.br/j/cr/a/PNshC3mykJQkxVmDyWGCWHk/abstract/?lang=pt>
- Canaza-Cayo AW, Lopes PS, Silva MVGB, Torres RA, Martins MF, Arbex WA, Cobuci JA (2015) Genetic Parameters for Milk Yield and Lactation Persistency Using Random Regression Models in Girolando Cattle. **Asian Australas. J. Anim. Sci.** 28:1407-1418.
- Cilek S, Sahin E (2009) Estimation of some genetic parameters (heritability and repeatability) for milk yield in the Anatolian population of Holstein cows. **Archiva Zootechnica** 12:57-64.

- Cobuci JA, Costa CN, Neto JB, de Freitas AF (2011) Genetic parameters for milk productions by using random regression models with different alternatives of fixed regression modeling. **R. Bras. Zootec.** 40:557-567.
- Coffey MP, Simm G, Brotherstone S (2002) Energy Balance Profiles for the First Three Lactations of Dairy Cows Estimated Using Random Regression. **J. Dairy Sci.** 85:2669-2678.
- Cole JB, Mikanjuola BO, Rochus CM, van Staaveren N, Baes C (2023) The effects of breeding and selection on lactation in dairy cattle. **Anim. Front.** 13:62-70.
- Costa A, Lopez-Villalobos N, Visentin G, De Marchi M, Cassandro M, Penasa M (2019) Heritability and repeatability of milk lactose and its relationships with traditional milk traits, somatic cell score and freezing point in Holstein cows. **Animal** 13:909-916.
- Costa CN, de Melo CMR, Packer IU, de Freitas AF, Teixeira NM, Cobuci JA (2008) Genetic parameters for test day milk yield of first lactation Holstein cows estimated by random regression using Legendre polynomials. **R. Bras. Zootec.** 37:602-608.
- Daltro DS, Negri R, Cobuci JA (2021) Heterosis effects on 305-day milk yield in a Girolando dairy cattle population in different lactation orders. **Livestock Science** 245:104428.
- Daltro DS, Silva MVGB, da Gama LT, Machado JD, Kern EL, Campos GS, Panetto JCC, Cobuci JA (2020) Estimates of genetic and crossbreeding parameters for 305-day milk yield of Girolando cows. **Italian Journal of Animal Science** 19:86-94.
- Daltro DS, Silva MVGB, Gama LT, Machado JD, Kern EL, Campos GS, Panetto JCC, Cobuci JA (2019) Estimates of genetic and crossbreeding parameters for 305-day milk yield of Girolando cows. **Italian Journal of Animal Science** 19:86-94.
- Druet T, Jaffrézic F, Boichard D, Ducrocq V (2003) Modeling Lactation Curves and Estimation of Genetic Parameters for First Lactation Test-Day Records of French Holstein Cows. **J. Dairy Sci.** 86:2480-2490.
- El Faro L, Cardoso VL, Albuquerque LG (2008) Variance component estimates applying random regression models for test-day milk yield in Caracu heifers

- (*Bos taurus* Artiodactyla, Bovidae). **Genetics and Molecular Biology** 31:665-673.
- Geweke, J. (1992) Evaluating the accuracy of sampling-based approaches to calculating posterior moments. **Bayesian Statistics** 4:169–193.
- Goff JP, Horst RL (1996) Physiological Changes at Parturition and Their Relationship to Metabolic Disorders. **J. Dairy Sci.** 80:1260-1268.
- Gonzalez-Herrera LG, El Faro L, Bignardi AB, Pereira RJ, Machado CHC, Albuquerque LG (2015) Random regression analysis of test-day milk yield in the first and second lactations of Brazilian Gyr cows. **Genet. Mol. Res.** 14:16497-16507.
- Heidelberger P, Welch PD (1983) Simulation run length control in the presence of an initial transient. **Opns Res.** 31:1109-1144.
- Júnior RNCC, Fernandes LS, Panetto JCC, Silva MVGB, Araújo CV, Silva AGM, Marques JRF, Silva WC, Araújo SI, Castro SRS, Silva LKX, Castro SV, Júnior JBL (2023) Heterogeneity of variance and genetic parameters for milk production in cattle, using Bayesian inference. **PLoS ONE** 18: e0288257.
- Li J, Gao H, Madsen P, Li R, Liu W, Bao P, Xue G, Gao Y, Di X, Su G (2020) Impact of the Order of Legendre Polynomials in Random Regression Model on Genetic Evaluation for Milk Yield in Dairy Cattle Population. **Front. Genet.** 11:586155.
- Lidauer M, Mantysaari EA, Strandén I (2003) Comparison of test-day models for genetic evaluation of production traits in dairy cattle. **Livestock Production Science** 79:73-86.
- Marchioretto PV, Rabel RAC, Allen CA, Ole-Neselle MMB, Wheeler MB (2023) Development of genetically improved tropical-adapted dairy cattle. **Anim. Front.** 13:7-15.
- Marumo JL, Lusseau D, Speakman JR, Mackie M, Hambly C (2022) Influence of environmental factors and parity on milk yield dynamics in barn-housed dairy cattle. **J. Dairy Sci.** 105:1225-1241.
- Meyer K (1998) “DXMRR” – A program to estimate covariance functions for longitudinal data by restricted maximum likelihood. **Animal Genetics and Breeding Unit** University of New England, Armidale, Australia.

- Misztal I, Legarra A, Aguilar I (2014) Using recursion to compute the inverse of the genomic relationship matrix. **Journal of Dairy Science** 97:3943-3952.
- Pereira RJ, Ayres DR, Júnior MLS, El Faro L, Vercesi Filho AE, Albuquerque LG (2019) Test-day or 305-day milk yield for genetic evaluation of Gir cattle. **Pesquisa Agropecuária Brasileira** 54:e00325.
- Pereira RJ, Bignardi AB, El Faro L, Verneque RS, Vercesi Filho AE, Albuquerque LG (2013) Random regression models using Legendre polynomials or linear splines for test-day milk yield of dairy Gyr (*Bos indicus*) cattle. **Journal of Dairy Science** 96:565-574.
- Ptak E, Schaeffer LR (1993) Use of test day yield for genetic evaluation of dairy sires and cows. **Livestock Production Science** 34:23-34.
- Santos EPB, Feltes GL, Negri R, Cobuci JA, Silva MVGB (2021) Estimation of genetic parameters for test-day milk yield in Girolando cows using a random regression model. **Arq. Bras. Med. Vet. Zootec.** 73:18-24.
- Scarso S, McParland S, Visentin G, Berry DP, McDermott A, De Marchi M (2017) Genetic and nongenetic factors associated with milk color in dairy cows. **J. Dairy Sci.** 100:7345-7361.
- Schaeffer LR, Jamrozik J, Kistemaker GJ, Van Doormaal BJ (2000) Experience with a Test-Day Model. **J. Dairy Sci.** 83:1135-1144.
- Singh H, Pannu U, Narula HK, Chopra A, Naharwara V, Bhakar SK (2016) Estimates of (co)variance components and genetic parameters of growth traits in Marwari sheep. **Journal of Applied Animal Research** 44:27-35.
- Smith BJ (2007) boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference. **J. Stat. Soft.** 21:1-37.
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit. **J. R. Statist. Soc. B** 64:583-639.
- Uribe H, Lembeye F (2020) Estimation of genetic parameters for milk yield using a random regression test-day model in first parity dairy cows under pasture-based systems of Los Ríos region in Chile. **Austral. J. Vet. Sci.** 52:103-107.
- Visentin G, McParland S, De Marchi M, McDermott A, Fenelon MA, Penasa M, Berry DP (2017) Processing characteristics of dairy cow milk are moderately heritable. **J. Dairy Sci.** 100:6343-6355.

Yamazaki T, Takeda H, Hagiya K, Yamaguchi S, Sasaki O (2018) Prediction of random-regression coefficient for daily milk yield after 305 days in Milk by using the regression-coefficient estimates from the first 305 days. **Asian-Australian J. Anim. Sci.** 31:1542-1549.

CHAPTER 3 – POPULATION STRUCTURE AND GENOMIC PARAMETER ESTIMATES FOR TEST-DAY MILK YIELD IN GIROLANDO CATTLE USING RANDOM REGRESSION MODELS

Abstract - Most Brazilian dairy herds are composed of Girolando animals, which are the result of crossbreeding between the Gir and Holstein breeds, with proportions of Holstein varying between 1/4 and 7/8. The 5/8 Holstein group represents 80% of the country's milk production. The aim of this study was to estimate variance components and analyze the genetic structure of the Girolando population based on genotypic and phenotypic data. A total of 480,364 test-day milk yield (TDMY) records from 76,688 cows in their first lactation, with calvings occurring between 1991 and 2021, were analyzed. The phenotypic data included nine genetic groups, formed by different combinations of the Holstein (H) and Gir (G) breeds: 1/2H + 1/2G, 1/4H + 3/4G, 1/8H + 7/8G, 3/4H + 1/4G, 3/8H + 5/8G, 5/8H + 3/8G, and 7/8H + 1/8G. The genotypic data involved 25,936 genotyped animals, which were later imputed, resulting in a final set of 37,911 SNPs from 5,339 animals. The population structure was analyzed using Principal Component Analysis (PCA) and the genetic differentiation index (F_{st}) in the Plink 1.9 software. The TDMY was modeled using an animal random regression model, considering fixed effects of heterozygosity and age at calving (linear and quadratic) as covariates, along with random additive genetic effects and permanent environmental effects of the cow, using fourth-order Legendre orthogonal polynomials. The ssGBLUP methodology was applied for the analysis. The genetic results indicated high variability in the population, with genotype frequencies of 0.07 (AA), 0.31 (AB), and 0.62 (BB), suggesting greater heterozygosity in the crossbred groups. The average observed and expected heterozygosity levels were 0.37 and 0.36, respectively, being higher in the 1/2H + 1/2G group, highlighting the benefits of hybrid vigor. The F_{st} index ranged from 0.003 to 0.21, indicating greater genetic differentiation between the Holstein and Gir breeds, while the crossbred groups showed closer genetic proximity. Milk production varied according to genetic composition, being highest in the 1/2H + 1/2G group (23.2 kg/day) and lowest in the Gir group (16.7 kg/day). The average heritability was 0.48, reaching 0.59 between days 91 and 120 of lactation, and repeatability was 0.84, reflecting high stability in production. High genetic correlations between adjacent records indicated a strong linear genetic association between TDMY. The estimates of heritability and repeatability, along with the increased production in 1/2H + 1/2G cows, suggest that, despite the influence of environmental factors, selection remains an efficient strategy for improving milk production.

Keywords: Genetic structure and diversity, genetic-genomic parameters, genomic selection, Gir x Holstein.

Resumo - A maioria dos rebanhos leiteiros brasileiros é composta por animais da raça Girolando, originada do cruzamento entre as raças Gir e Holandesa, com proporções de Holandês variando entre 1/4 e 7/8. O grupo 5/8 representa 80% da produção de leite do país. O objetivo deste estudo foi estimar os componentes de variância e analisar a estrutura genética da população Girolando a partir de dados genotípicos e fenotípicos. Foram analisados 480.364 registros de PLDC de 76.688 vacas em primeira lactação, com partos ocorridos entre 1991 e 2021. Os dados fenotípicos incluíram nove grupos genéticos, formados por diferentes combinações entre as raças Holandês (H) e Gir (G): 1/2H + 1/2G, 1/4H + 3/4G, 1/8H + 7/8G, 3/4H + 1/4G, 3/8H + 5/8G, 5/8H + 3/8G, 7/8H + 1/8G. Os dados genotípicos envolviam 25.936 animais genotipados, que foram posteriormente imputados, resultando em um conjunto final de 37.911 SNPs de 5.339 animais. A estrutura populacional foi analisada por Análise de Componentes Principais (PCA) e índice de diferenciação genética (F_{st}) no software Plink 1.9. A PLDC foi modelada por um modelo animal de regressão aleatória, considerando efeitos fixos de heterozigosidade e idade ao parto (linear e quadrático) como covariáveis, além de efeitos aleatórios genéticos aditivos e de ambiente permanente da vaca, com o uso de polinômios ortogonais de Legendre de quarta ordem. A metodologia ssGBLUP foi aplicada para a análise. Os resultados genéticos indicaram alta variabilidade na população, com frequências genotípicas de 0,07 (AA), 0,31 (AB) e 0,62 (BB), sugerindo maior heterozigosidade nos grupos mestiços. Os níveis médios de heterozigosidade observada e esperada foram 0,37 e 0,36, respectivamente, sendo mais elevados no grupo 1/2H + 1/2G, destacando os benefícios da heterose. O índice F_{st} variou de 0,003 a 0,21, indicando maior diferenciação genética entre as raças Holandês e Gir, enquanto os grupos mestiços apresentaram maior proximidade genética. A produção de leite variou conforme a composição genética, sendo maior no grupo 1/2H + 1/2G (23,2 kg/dia) e menor no grupo Gir (16,7 kg/dia). A herdabilidade média foi de 0,48, atingindo 0,59 entre 91 e 120 dias de lactação, e a repetibilidade foi de 0,84, refletindo alta estabilidade na produção. As correlações genéticas e fenotípicas elevadas entre controles adjacentes indicam a consistência da expressão genética na produção de leite. As estimativas de herdabilidade e repetibilidade, associadas ao aumento da produção nas vacas 1/2H + 1/2G, sugerem que, apesar da influência de fatores ambientais, a seleção genética continua sendo uma estratégia eficiente para melhorar a produção de leite.

Palavras-chave: Estrutura e diversidade genética, Gir x Holandês, parâmetros genéticos-genômicos, seleção genômica.

1. INTRODUCTION

The Girolando breed, resulting from the crossbreeding of the Gir and Holstein breeds and developed in the 1980s, stands out as a viable alternative for milk production systems in tropical regions (Panetto et al., 2022; Marchioretto et al., 2023). This breed combines the rusticity, heat resistance, and adaptability to adverse climatic conditions of the Gir breed with the high milk productivity of the Holstein breed (Canaza-Cayo et al., 2014a). This combination of characteristics makes Girolando one of the main dairy breeds in Brazil, responsible for over 80% of the milk produced in the country (Silva et al., 2018).

Genetic diversity, which encompasses the variation of traits within a population, is crucial for the potential response to future selections (Notter, 1999; Booy et al., 2008). However, the wide genetic variability in the Girolando breed, due to its various genetic groups, requires attention to ensure uniformity and predictability in milk production.

This represents a major challenge for genetic evaluation, as the use of composite breeds is still under extensive study in this context. These aspects are key for the efficient management and sustainability of dairy production systems (Daltro et al., 2019).

The incorporation of genomic information into phenotypic and pedigree records in the selection of dairy cattle has promoted advances in genetic improvement of the species, allowing faster and more accurate estimates of the genetic potential of animals (VanRaden, 2008) and consequently accelerating genetic improvement (Weigel, 2017). In this context, the use of molecular markers, such as single nucleotide polymorphisms (SNPs), has proven to be a valuable tool for analyzing the genetic structure of populations and estimating genetic parameters, as well as breeding values for traits of interest (Dekkers, 2012; Weigel, 2017; Gutierrez-Reinoso et al., 2021).

Dairy cattle breeding programs in tropical regions face specific challenges related to environmental constraints and limited available resources (Zumbach & Peters, 2002). Genotype-environment interactions are crucial in this context, as high-productivity breeds adapted to temperate climates tend to perform poorly in tropical conditions (Burrow, 2012). To overcome these limitations, combining advanced genomic selection techniques with conventional breeding methods

emerges as a promising strategy and can allow early identification of an animal's genetic potential before the manifestation of its phenotypic traits (Das et al., 2021).

Understanding genetic parameters plays a key role in comprehending economically important traits' genetic basis and defining more effective and sustainable selection strategies (Roka et al., 2024). Among the modern methodologies applied to genetic improvement, genomic random regression models (RRM) stand out, enabling more accurate estimates of the genetic values of animals for longitudinal data.

The Single-step Genomic Best Linear Unbiased Prediction (ssGBLUP) method has become widely used as a tool by efficiently integrating genomic, phenotypic, and pedigree information into a single analysis (Misztal et al., 2013a; Bermann et al., 2022). This methodology incorporates genomic data directly into the pedigree matrix and enhances the accuracy of genetic predictions, optimizing selection for traits of interest (Misztal et al., 2013a).

The objective of this study is to analyze the structure and genetic diversity of the Girolando breed and estimate variance components, parameters, and genomic breeding values for test-day milk yield so that it can aid in the selection procedures for improving the Girolando breed.

2. MATERIAL AND METHODS

The genotype and phenotype data for the milk production of Girolando cows were provided by Embrapa Gado de Leite and the Brazilian Association of Girolando Breeders.

2.1. Phenotypic data

For the estimation of variance components, 480,364 test-day milk yield (TDMY) were used, derived from 76,688 first lactations of Girolando cows, with ages at calving ranging from 22 to 48 months, calving between 1991 and 2021, and distributed across 913 herds. The milk production per control ranged from 4 to 40 kg, and the controls were recorded between the fifth and the 305th day of

lactation, for cows with at least three recorded controls. The proportions of cows according to the number of milk controls were: 8.5%, 11.8%, 12.0%, 11.9%, 11.6%, 10.8%, 10.1%, 9.1%, 7.6%, and 6.6% for 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 controls, respectively.

The contemporary group (CG) was formed by animals from the same herd, calving year, day-month-year of milk control, and genetic group. CG with fewer than three animals were excluded. The genetic groups (Table 1) were identified as follows: Holstein (H); Gir (G); 1/2H + 1/2G; 1/4H + 3/4G; 1/8H + 7/8G; 3/4H + 1/4G; 3/8H + 5/8G; 5/8H + 3/8G; 7/8H + 1/8G. After identifying the pedigree for all animals across three generations, the pedigree file consisted of 252,389 animals, including 8,312 bulls and 67,549 cows.

Heterozygosity (HET) was calculated as follows: $HET = ((HP * GM) + (HM * GP))$, where HP is the Holstein breed composition from the paternal contribution; GM is the Gir breed composition from the maternal contribution; HM is the Holstein breed composition from the maternal contribution; and GP is the Gir breed composition from the paternal contribution. The HET values ranged from 0.005 to 1. The HET effect was included to account for and correct the effects of the different genetic groups within the population (Appendix 1).

2.2. Genotypic data

The animals were genotyped with SNP panels of varying densities, including the Illumina BovineSNP50 BeadChip v2 (50 K), Illumina BovineHD BeadChip (HD), Zoetis Custom SNP chip ZL2 (20 K), Zoetis Custom SNP chip ZL5 (30 K), Zoetis Custom SNP chip ZBU (54 K), Zoetis Custom SNP chip ZM2 (70 K), and Zoetis Custom SNP chip ZBN (55 K).

After defining the Illumina BovineHD BeadChip (HD) panel as the standard, SNPs from other panels were extracted and/or imputed to the BovineHD chip using the Illumina 50 K panel as the marker panel. The FImpute 2.2 software (Sargolzaei et al., 2014) achieving an accuracy of 97%. The genotype file contained 45,843 SNPs from 25,936 animals, of which 5,342 had both phenotype and genotype information (Table 1).

Tabela 1 – Number of phenotype and phenotypes and genotypes observations by genetic group.

Genetic Groups	Animals with phenotypes	Animals with phenotypes and genotypes
H	569	10
G	158	13
1/2H + 1/2 G	33,585	2,409
1/4H + 3/4G	1,055	188
1/8H + 7/8G	131	3
3/4H + 1/4G	25,754	1,546
3/8H + 5/8G	1,149	82
5/8H + 3/8G	9,462	885
7/8H + 1/8G	4,825	206

Mendelian conflicts and the pedigree were checked, and any discrepant markers were defined as unknown and absent. After the quality control process, the data file contained 37,911 SNPs and 5,339 animals.

2.3. Population structure and genetic diversity analysis

Principal component analysis (PCA) was performed using the genotype matrix to characterize the population structure of the Girolando genetic groups. For the PCA, SNPs (single nucleotide polymorphisms) with minor allele frequency (MAF) lower than 5% and call rate for SNPs and samples lower than 90% (—geno) were removed. The genetic differentiation index (F_{st}) was calculated between all pairs of genetic groups. Additionally, inbreeding coefficients, as well as observed and expected heterozygosity levels, were estimated to assess the genetic diversity within and among the groups. Population structure and genetic diversity analyses were conducted using the Plink 1.9 software (Purcell et al., 2007).

2.4. Analysis model

The analysis of milk production on the control day was performed using a random regression animal model. The heterozygosity class and the linear and quadratic effects of the cow's age at calving were considered as fixed effects. The average lactation curves within each genetic group were modeled using Legendre orthogonal polynomials to account for differences in milk production levels between the various genetic groups.

The random effects of additive genetic and permanent environment of the cow effects were modeled using fourth-order Legendre orthogonal polynomials. Residual variances were considered heterogeneous, divided into five classes based on days in lactation: 5-60; 61-120; 121-180; 181-240; and 241-305 days in lactation. The effect of CG (contemporary group) was considered random.

The general random regression model used in the analyses was represented by:

$$y_{ijklm} = CG_i + \sum_{n=1}^2 b_n x_k^n + \sum_{o=1}^4 \beta_{jo} \varphi_o(t) + \sum_{o=1}^4 \alpha_{ko} \varphi_o(t) + \sum_{o=1}^4 \rho_{ko} \varphi_o(t) + e_{ijklm}$$

where y_{ijklm} is the m-th observation on the lactation day t of animal l , with heterozygosity k , belonging to the genetic group j and the contemporary group CG_i ; CG_i is the effect of the i-th contemporary group; b_n is the regression coefficient for the linear ($n=1$) and quadratic ($n=2$) effects of the cow's age at calving x_k on TDMY (in months); β_{jo} is the set of fixed regression coefficients for the average trajectory of genetic group j ; $\varphi_o(t)$ is the covariate of the regression function according to the day of lactation (t); α_{ko} , ρ_{ko} are the sets of random regression coefficients for the trajectory of the additive genetic and permanent environmental effects of the cow for each animal l ; and e_{ijklm} is the error associated with the m-th observation on lactation day t of animal l , with heterozygosity k , belonging to the genetic group j , and the contemporary group CG_i .

The assumptions of the model were:

$$E[y] = X\beta$$

and

$$var \begin{bmatrix} \delta \\ a \\ p \\ e \end{bmatrix} = \begin{bmatrix} I \otimes GC_0 & 0 & 0 & 0 \\ 0 & A \otimes G_0 & 0 & 0 \\ 0 & 0 & I \otimes P_0 & 0 \\ 0 & 0 & 0 & R \otimes R_0 \end{bmatrix}$$

where $E[y]$ is the expectation of y ; y is the vector of TDMY; X is the incidence matrix for the fixed effects of the vector β ; δ , a , p , and e are the vectors of the random coefficients for the CG effects, additive genetics, permanent environmental effects of the cow, and residuals; CG_0 , G_0 , and P_0 are the covariance matrices for the random regression coefficients for the CG effects, genetic additive, and permanent environmental effects of the cow, respectively; R_0 is the covariance matrix for the residuals; I is an identity matrix with dimensions equal to the number of levels of the effects; A is the pedigree relationship matrix; R is the diagonal matrix of residual variances; $\otimes$ is the Kronecker product.

In the ssGBLUP (single-step Genomic BLUP) methodology, the pedigree matrix A is replaced by matrix H , which combines pedigree information (A) and genomic data (G) into a single matrix (H) (Legarra et al., 2009):

$$H = A + \begin{bmatrix} A_{12}A_{22}^{-1} & 0 \\ 0 & 1 \end{bmatrix} [G - A_{22}][I \quad I] \begin{bmatrix} A_{22}A_{21}^{-1} & 0 \\ 0 & 1 \end{bmatrix}$$

where subscript 1 represents the non-genotyped animals and subscript 2 represents the genotyped animals. The matrix G is a genomic relationship matrix constructed using the method of VanRaden (2008). The inverse of matrix H is:

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

where A^{-1} is the inverse of the pedigree relationship matrix, A_{22}^{-1} is the inverse of the pedigree relationship matrix for the genotyped animals, and G^{-1} is the inverse of the genomic relationship matrix described by VanRaden (2008).

The computational analyses were performed using a Bayesian method, utilizing the GibbsF90+ program (Misztal et al., 2014). In matrix notation, the model is represented by:

$$y = Xb + Za + Wp + Hc + e$$

where y is the vector of observations; b is the vector of solutions for the fixed effects; a is the vector of random coefficients for the genetic additive effects; p is the vector of random coefficients for the permanent environmental effects of the cow; c is the vector of random coefficients for the CG; e is the vector of residuals; and X, Z, W, H and are the corresponding incidence matrices.

The assumptions for the prior distributions for the unknown parameters were considered according to:

- $y|b, a, p, \sigma_{e_1}^2, \dots, \sigma_{e_x}^2 \sim NMV(Xb + Za + Wp, R), R = diag\{\sigma_{e_x}^2\}$, with x going from 1 to n , where n is the number of residual variance classes.
- $a|\Lambda_a \sim NMV(0, G)$, com $G = \Lambda_a \otimes A$, where A is the pedigree relationship matrix, and Λ_a is the matrix with the covariances between the genetic-additive random regression coefficients;
- $p|\Lambda_p \sim NMV(0, P)$, com $P = \Lambda_p \otimes I$, where I is the identity matrix and Λ_p is the matrix with the covariances between the random regression coefficients for the permanent environmental effects of the cow;
- $\Lambda_a|v_a, S_a^2 \sim W^{-1}(v_a, v_a S_a^2)$, $\Lambda_p|v_p, S_p^2 \sim W^{-1}(v_p, v_p S_p^2)$ and
- $R|v_e, S_e^2 \sim W^{-1}(v_e, v_e S_e^2)$, em que v_a, S_a^2 ; v_p, S_p^2 ; v_e, S_e^2 represent the degrees of confidence and the prior values for the covariances of the cow's genetic-additive and permanent environmental random regression coefficients and residual variances, respectively;
- $b \propto$ is a constant.
- NMV and W^{-1} refer to the multivariate normal and inverse Wishart distributions.

Inferences about the parameters were made from their corresponding posterior marginal distributions. For each model evaluated, a chain of 200,000 samples was generated, and conservatively, the first 40,000 samples were discarded and considered as burn-in. Convergence was monitored by graphical inspection of the samples versus iterations and also by the criteria proposed by Heidelberger & Welch (1983) and Geweke (1992), calculated using the BOA package in R (Smith, 2007). A sample discard interval of 50 iterations was considered, leaving 16,000 samples on which the inferences were based. The genetic parameters for milk production on the control day of lactation were estimated at 30-day intervals for each sample.

2.5. Estimation of (co)variance components

The posterior means were used as a point estimate for the (co)variance components. The additive genetic, permanent environmental, phenotypic, and residual (co)variance matrices between milk production on all lactation days (from day 5 to day 305) were obtained following the approach applied by Druet et al.

(2003). Heritabilities were defined as the ratio of additive genetic variance to phenotypic variance for each lactation day. Repeatabilities were defined as the ratio of the sum of genetic and permanent environmental variances to the phenotypic variance for each lactation day. The correlations between milk production on lactation days i and j were calculated as the ratio of the covariance $cov(i, j)$ to the square root of the product of the variances for lactation days i and j .

3. RESULTS AND DISCUSSION

The average inbreeding coefficient among the different genetic groups was -0.04 (Figure 1). This value can indicate high genetic diversity in the analyzed population, suggesting that individuals have a low degree of relatedness, resulting in greater genetic distance between animals from different groups. Given the magnitude of the estimate, the average inbreeding coefficient can be considered null.

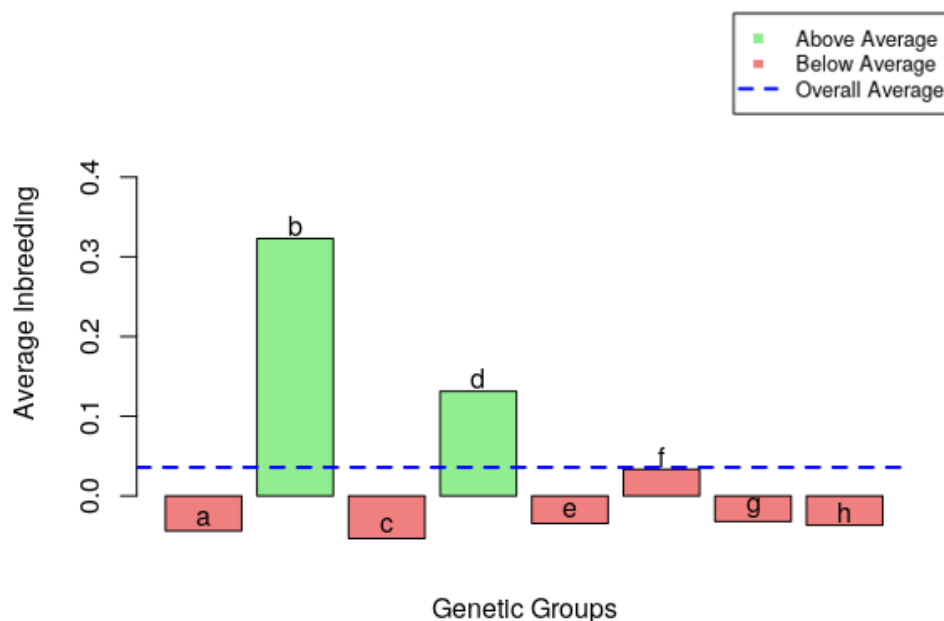


Figure 1. Average inbreeding of the different genetic groups: a: H; b: G; c: $1/2H + 1/2G$; d: $1/4H + 3/4G$; e: $3/4H + 1/4G$; f: $3/8H + 5/8G$; g: $5/8H + 3/8G$; and h: $7/8H + 1/8G$.

The estimates of inbreeding ranged from -0.05 to 0.32 for the $1/2H + 1/2G$ and G groups, respectively. These values fall outside the range reported in the

literature for the Girolando breed, in which pedigree-based inbreeding coefficients vary from 0.11 to 0.74 (Canaza-Cayo et al., 2014b; Rosa et al., 2023). For the Gir breed, the average inbreeding coefficient based on pedigree is 0.04 (Pereira et al., 2016), while for the Holstein breed, this value is 0.17 (Silva et al., 2019). The negative inbreeding estimate in the genetic groups with a higher degree of crossbreeding suggests a greater occurrence of heterozygosity resulting from the crossbreeding of genetically distinct breeds (Nishio et al., 2023). On the other hand, the higher inbreeding value indicates elevated homozygosity. These results are important for monitoring the effects of crossbreeding between breeds and avoiding potential negative impacts of inbreeding, such as inbreeding depression (Vanraden, 1992; Snelder & Larenstein, 2019).

The average observed and expected heterozygosity levels were 0.37 and 0.36, respectively. The observed heterozygosity ranged from 0.24 in the G group to 0.38 in the 1/2H + 1/2G group, while the expected average heterozygosity remained constant at 0.36 across all genetic groups (Figure 2).

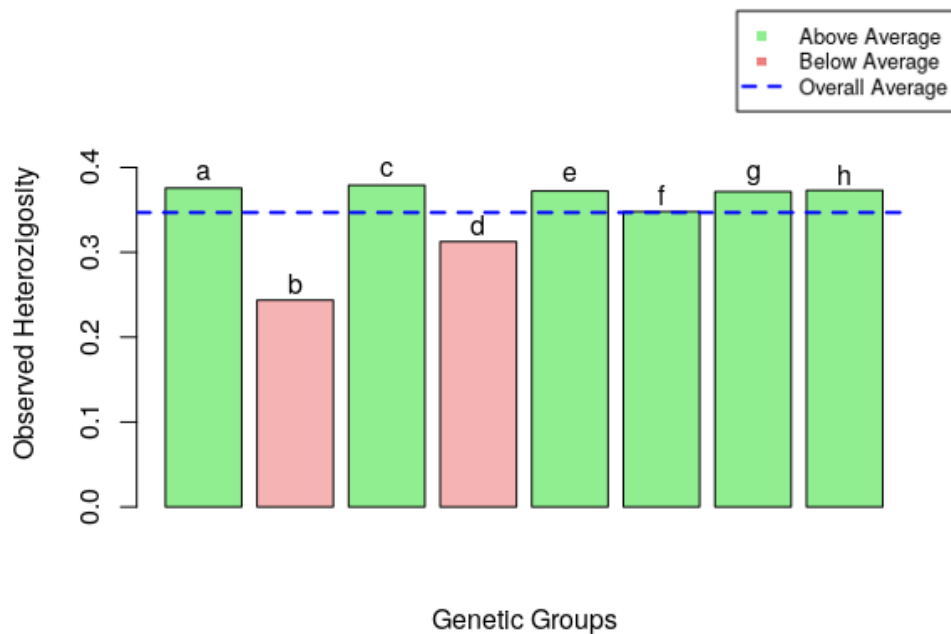


Figure 2. Average heterozygosity of the different genetic groups: a: H; b: G; c: 1/2H + 1/2G; d: 1/4H + 3/4G; e: 3/4H + 1/4G; f: 3/8H + 5/8G; g: 5/8H + 3/8G; h: 7/8H + 1/8G.

Observed heterozygosity directly reflects the proportion of heterozygous individuals in the population, serving as an important indicator of genetic variability (Harper & Hawksworth, 1994). The highest observed heterozygosity in the 1/2H + 1/2G group highlights the benefits of crossbreeding distinct breeds, which can result in heterosis. This effect is crucial in genetic improvement, particularly in production systems aiming for greater efficiency and adaptability of animals. Conversely, the lower heterozygosity in the G genetic group reinforces the importance of strategies to increase genetic variability within purebred populations (Kelleher et al., 2017).

The F_{st} indices (Table 2) indicated the degree of genetic differentiation among the genetic groups analyzed (Weir & Hill, 2002). Low values suggest more homogeneous populations, whereas high values indicate genetic isolation or significant differences between populations. The F_{st} values ranged from 0.003 (between the 3/4H + 1/4G and 5/8H + 3/8G groups) to 0.17 (between the H and G groups). These results demonstrate a high genetic differentiation between the Holstein and Gir breeds, while the lower values observed among groups with a higher frequency of crossbreeding between individuals of different breeds indicate greater genetic diversity and, consequently, greater genetic proximity.

Table 2. Average F_{st} index of the different genetic groups of Holstein (H) and Gir (G).

	G	1/2H+1/2G	1/4H+3/4G	3/4H+1/4G	3/8H+5/8G	5/8H+3/8G	7/8H+1/8G
H	0,17	0,06	0,11	0,02	0,07	0,03	0,01
G		0,05	0,02	0,09	0,04	0,07	0,11
1/2H+1/2G			0,02	0,01	0,01	0,01	0,03
G							
1/4H+3/4G				0,05	0,01	0,03	0,07
G							
3/4H+1/4G					0,03	0,003	0,004
G							
3/8H+5/8G						0,02	0,04
G							
5/8H+3/8G							0,01
G							

The PCA allowed for a clear identification of the eight evaluated genetic groups (Figure 3). The principal axes of the PCA, 1 and 2, explained 19.02% and 10.57% of the genetic variation, respectively, and were used to identify clustering or dispersion patterns among individuals.

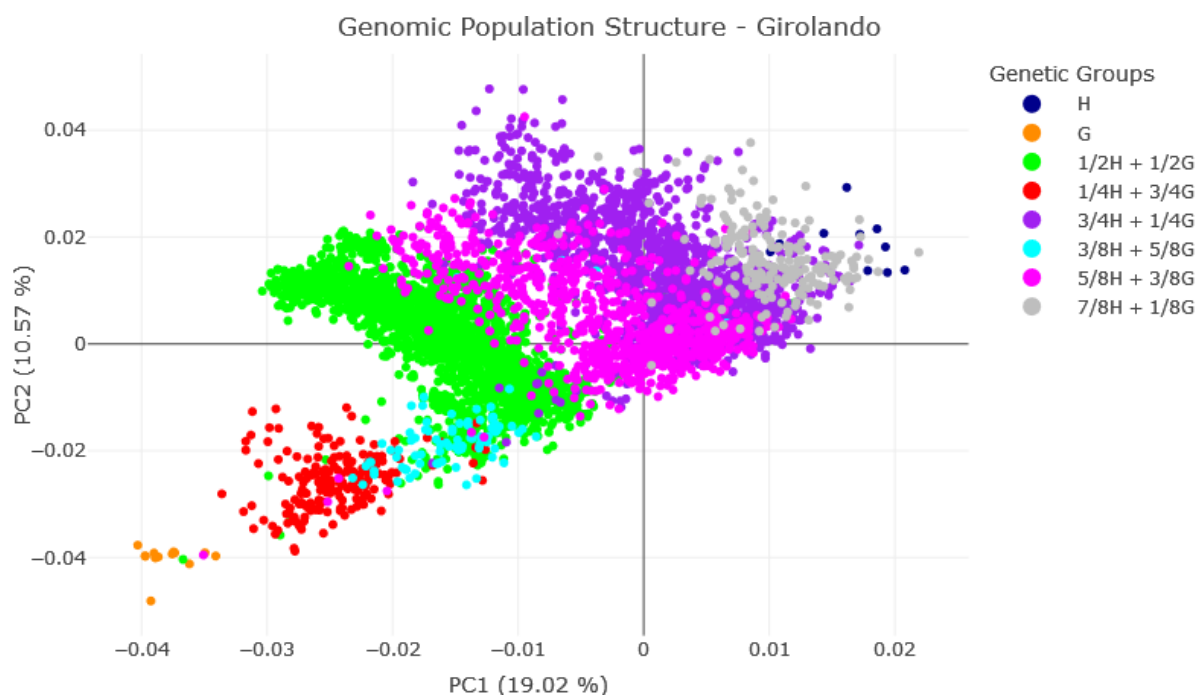


Figure 3. PCA of the different genetic groups of Holstein (H) and Gir (G).

From Figure 3, it was possible to visualize the clustering of individuals according to their genetic groups. Individuals with a higher proportion of Gir genetics occupied the lower-left region of the graph, while those with a higher proportion of Holstein genetics were located in the upper-right region. Intermediate groups, such as 3/4H + 1/4G or 5/8H + 3/8G, occupied central positions, indicating that they share genetic characteristics from both G and H breeds. The PCA highlighted the genetic differentiation between the groups, reflecting differences in genotypes and the impact of crossbreeding between the breeds. Furthermore, a progressive genotype difference among the genetic groups was revealed, with genetically closer groups exhibiting greater similarity (McVean, 2009).

The observed overall mean and standard deviation for TDMY were 22.3 kg and 7.33 kg, respectively. The lowest average milk yields were observed for TDMY 1 (20.3 ± 6.39 kg), 9 (21.6 ± 6.95 kg), and 10 (20.6 ± 6.87 kg). The peak

milk production was reached during TDMY 4, corresponding to 91 to 120 days in milk, with an average of 23.8 ± 7.62 kg. Shortly after, in TDMY 5 (23.6 ± 7.64 kg), milk production started to decline in the subsequent TDMY (Table 3).

Table 3 - Number of observations, observed means, standard deviations, and coefficients of variation for milk production on TDMY (1 to 10).

TDMY	Days in Milk	Number of Observations	Mean $\pm$ Standard Deviation (kg)	CV (%)
1	5 to 30	40,763	20.3 ± 6.39	31.5
2	31 to 60	56,667	22.4 ± 7.01	31.2
3	61 to 90	57,639	23.4 ± 7.43	31.7
4	91 to 120	57,203	23.8 ± 7.62	32.1
5	121 to 150	55,654	23.6 ± 7.64	32.4
6	151 to 180	51,793	23.2 ± 7.52	32.4
7	181 to 210	48,479	22.7 ± 7.35	32.4
8	211 to 240	43,690	22.2 ± 7.16	32.3
9	241 to 270	36,608	21.6 ± 6.95	32.1
10	271 to 305	31,868	20.6 ± 6.87	33.4

Cows from the 1/2H + 1/2G genetic group had the highest average milk production (23.2 ± 7.36 kg), while cows from the G genetic group had the lowest average milk production (16.7 ± 4.47 kg) (Figure 4).

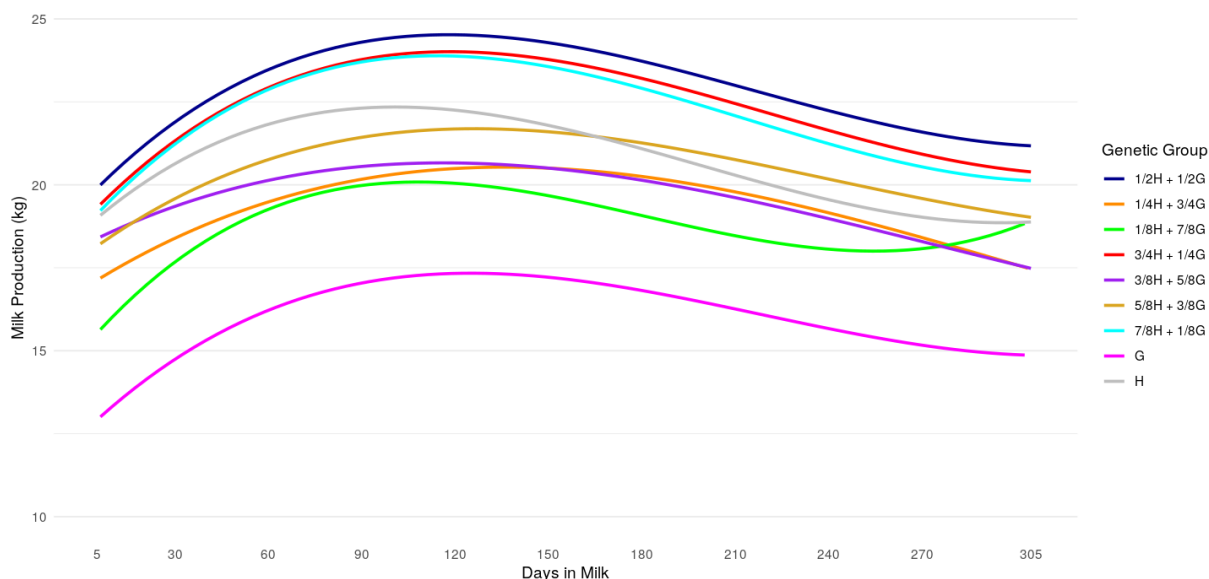


Figure 4. Observed means of milk production on the TDMY for different genetic groups of Holstein (H) and Gir (G) during lactation (5 to 305 days).

Regarding the variability in milk production among different Girolando genetic groups, the impact of genetic composition on milk productivity stands out. As adopted in this study, the use of fixed regression for each genetic group in the analysis model is recommended to adjust the phenotypic variation due to differences between genetic groups, which can contribute to a more accurate analysis of milk production traits. Koenig et al. (2005) emphasized that using fixed regression is an effective approach to isolate the additive genetic effect from environmental variations and can provide more precise estimates of additive genetic effects on milk production. In this context, several authors suggest that considering differences between genetic groups in statistical models, such as using specific fixed regression for each group, can lead to more reliable genetic evaluations and more effective selection strategies in genetic improvement programs (Costa et al., 2004; Facó et al., 2007; Canaza-Cayo et al., 2015; Barbero et al., 2022).

In the present study, cows from the 1/2H + 1/2G group had the highest average milk production (Figure 4), while those from the G genetic group showed the lowest, confirming the influence of heterosis and genetic composition on milk production. These results align with the observations of Misztal et al. (2013), who

highlighted the potential of crossbreeding to maximize heterosis and improve production efficiency.

For the 1/2H + 1/2G group, the balance between the superior dairy traits of the Holstein breed and the adaptability of the Gir breed seems to have provided this advantage. Meanwhile, the G genetic group showed lower milk production due to the predominance of Gir breed characteristics, which are more oriented toward resilience and adaptability in challenging environments, corroborating the findings of Madalena (2008). Compared to the results of Jacopini et al. (2016), who observed the highest production in the second lactation in the 7/8H + 1/8G group, followed by the 3/4H + 1/4G and 1/2H + 1/2G groups, the present study showed differences in milk productivity patterns among the genetic groups analyzed.

The average curves of the 1/2H + 1/2G, 3/4H + 1/4G, and 7/8H + 1/8G groups are similar, indicating comparable milk production performance in these genetic groups. This suggests that across different levels of Holstein genetic contribution, the gains in production are greater than in groups with a higher influence of the Gir breed. On the other hand, pure Holstein cattle do not perform well in the Brazilian tropical environment, reinforcing that the breed's adaptation to hotter and more challenging climates is limited.

Understanding the genetic interactions between the two breeds that influence milk production is essential for defining selection strategies in genetic improvement programs aimed at optimizing the productive efficiency of Girolando cattle. This knowledge can also guide mating decisions that maximize heterosis benefits, reduce the negative impacts of inbreeding, and contribute decisively to the long-term sustainability of dairy production, particularly in tropical environments (Brito et al., 2021).

The variances indicate that the contributions of additive genetic effects, permanent cow environment, and residual effects to the total variation in milk production fluctuated throughout lactation, reflecting the variations observed in the respective estimates (Figure 5).

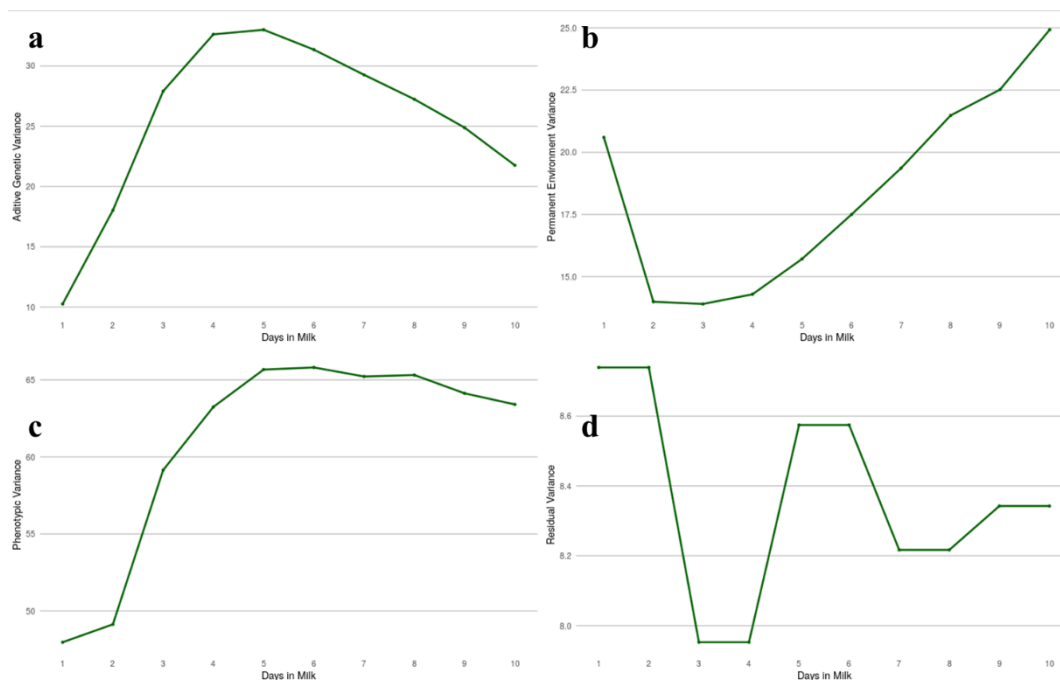


Figure 5. Mean estimates of additive genetic **(a)**, permanent environmental of the animal **(b)**, phenotypic **(c)**, and residual variances **(d)** over lactation.

The uniformity of residual variances throughout lactation, observed in this study, indicated that these factors remained stable over the entire period. However, additive genetic and permanent cow environment variances showed fluctuations, suggesting that the genetic and environmental factors influencing milk production did not remain constant throughout lactation. The variation within a restricted range in additive genetic variance reflects genetic differences among individuals, supporting the findings of Lynch and Walsh (1998), who argued that additive genetic variability is a key component for selection and improvement in breeding programs.

The small variability in residual variance suggests that the random regression model (RRM) employed, including the fixed effects in the model, was effective in capturing most of the variations in milk production, with only a small portion unexplained by additive genetic or environmental factors. This indicates that the model fits the data well (Hill, 2010). Such a result is important to justify the use of RRM in evaluating milk production during lactation in Girolando herds, which can contribute to a more robust and productive dairy industry, as Berry and Cromie (2007) highlighted when discussing the importance of selection strategies.

According to Costa et al. (2004), heterogeneity in genetic variances was identified between Holstein and crossbred cows, with the Gir breed showing higher heritability estimates for milk production. Studies have explored the genetic aspects of milk production and lactation uniformity in dairy cattle. Herrera et al. (2013) used random regression models to estimate genetic parameters for milk production in Holstein cattle, observing that heritability varied throughout lactation.

According to El-Awady (2013), moderate heritability estimates were obtained for milk production traits and high genetic correlations between milk production and other related traits. Braga et al. (2020) and Barbero et al. (2022) suggested that selection for milk production can lead to efficient genetic progress in Girolando herds. However, environmental factors still account for a significant portion of phenotypic variation, indicating that management practices remain crucial for improving productivity.

In this study, the average heritability estimate throughout the entire lactation period was 0.48 ± 0.11 , with a maximum value of 0.59 at TDMY 4 (91 to 120 days in lactation) and a minimum value of 0.26 at TDMY 1 (Figure 6). Repeatability estimates remained nearly constant throughout lactation. The average repeatability estimate was 0.84 ± 0.02 , with a maximum value of 0.86 at TDMY 4 and a minimum value of 0.76 at TDMY 1.

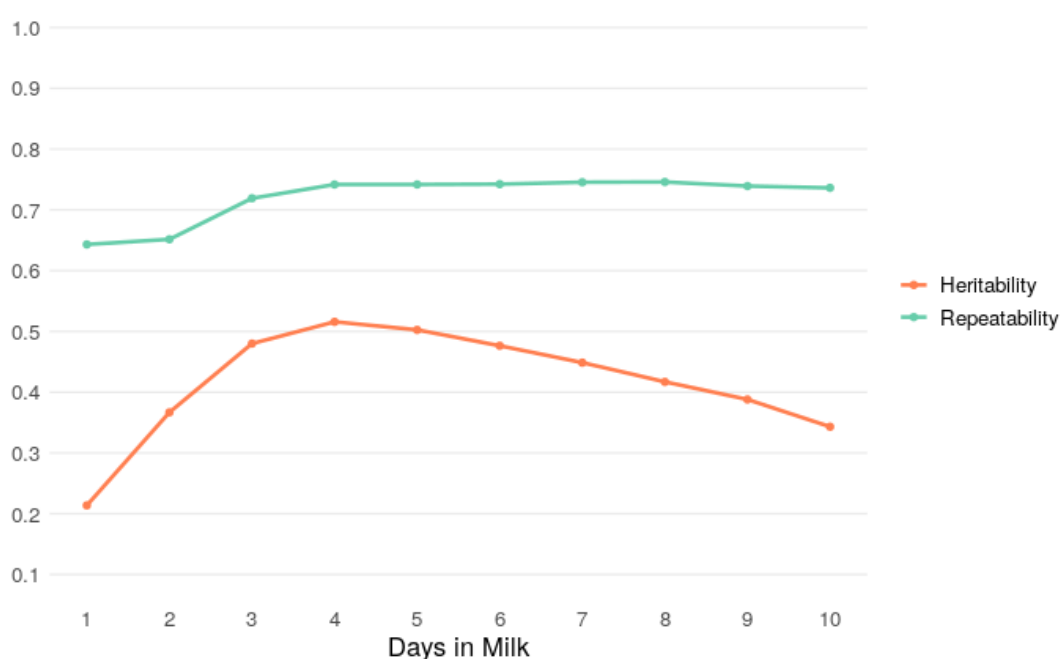


Figure 6. Average heritability and repeatability estimates during the lactation period.

Heritability and repeatability estimates are essential for understanding how additive genetic variability may change throughout lactation, especially in studies related to milk production in cattle. The average heritability of 0.48 estimated in this study indicated that approximately 48% of the phenotypic variance in milk production can be attributed to additive genetic effects. These estimates are higher than those reported in the literature (Ismael et al., 2021; Barbero et al., 2022), which may be due to differences in analysis methodology, the data structure from various crossbreeding combinations between Gir and Holstein breeds, and the number of observations used.

The variation in heritability estimates observed throughout lactation, with a peak at TDMY 4 and a lower value at the beginning of lactation, suggests that additive genetic expression may be more influential at specific stages of lactation. This could be due to the cows' physical condition variations, influenced by a negative energy balance at the start of lactation, or how environmental factors interact with genetic expression during different lactation periods. Heritability estimates for milk production have ranged from 0.18 to 0.31 (Canaza-Cayo et al., 2015; Barbero et al., 2022), indicating potential for genetic improvement in milk production.

Repeatability estimates of up to 0.84 reflect the correlation between repeated milk production measurements in the same animal over time (Misztal et al., 2013b). The observed variation in repeatability estimates throughout different lactation phases emphasizes the need to consider each phase separately in evaluating TDMY, as this metric indicates the consistency of phenotypic traits over the analyzed period.

The variation in repeatability throughout lactation highlights the importance of considering different lactation phases when assessing TDMY. Lactation can be divided into three phases: early, mid, and late. The early phase corresponds to the first months after calving when milk production peaks. The mid-phase follows the peak, characterized by more stable production. The late phase is marked by a gradual decline in production until the end of lactation (Oftedal & Iverson, 1995). Studies have shown that heritability and repeatability estimates for milk production vary across these phases. Heritability estimates tend to be higher at the beginning and middle of lactation while decreasing toward the end (Searle, 1961; Costa et al., 2005; Dionello et al., 2006). On the other hand, repeatability estimates tend to increase as more observations are incorporated, improving the precision of estimates (Searle, 1961; Van Vleck & Henderson, 1961).

Phenotypic correlation estimates between milk production at different TDMY ranged from 0.30 to 0.84, with the highest magnitudes between adjacent TDMY (Figure 7). Genetic correlations showed a similar variation to phenotypic correlations, with values

ranging from 0.53 to 0.99. Permanent cow environment correlations ranged from 0.18 to 0.97, with the highest magnitudes estimated between adjacent TDMY.

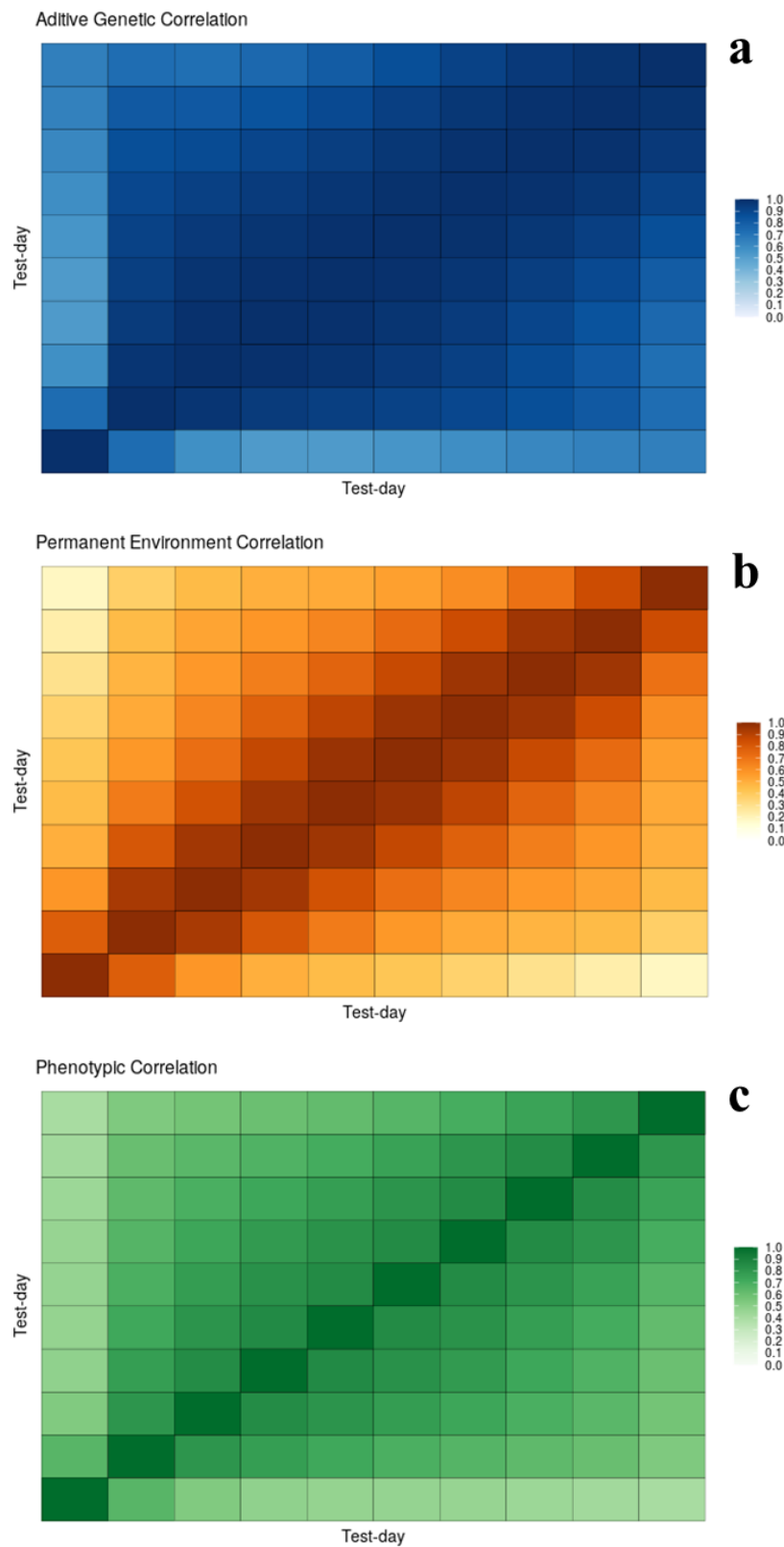


Figure 7. Average estimates of genetic (a), permanent cow environment (b), and phenotypic (c) correlations, respectively, between milk production on TDMY (1 to 10).

The high phenotypic correlations observed between TDMY, especially between consecutive periods, indicate a strong linear and proportional association between measurements over time. Genetic correlations between milk yields at different TDMY can be very high, ranging from 0.90 to 1.00, as reported by Haile-Mariam et al. (2015). Understanding these genetic and phenotypic associations contributes to developing long-term selection strategies in animal breeding programs (Barbosa et al., 2023) since genetic gain and correlated responses will be influenced by heritability (h^2) estimates and genetic correlations.

Genetic correlations and permanent cow environment correlations reinforce the influence of genetic factors and stable environmental conditions on Girolando milk production. The variation in genetic and permanent environmental correlations suggests that while a genetic component influences production, the environmental conditions under which the animals are maintained also play a significant role.

The highest magnitudes of these correlations between adjacent TDMY indicate that additive genetic effects and permanent cow environmental effects directly impact successive lactation phases. In studies that did not include genomic information, correlations showed greater variation, particularly during the early and late lactation periods. This difference can largely be attributed to the fewer records available in these periods, leading to less precise estimates (El Faro & Albuquerque, 2003; Melo et al., 2005; González Herrera et al., 2008). Including genomic data, these estimates exhibit less variation, suggesting greater stability in measuring additive genetic and permanent cow environmental effects throughout lactation (Veerkamp et al., 2011). This result highlights the importance of genomics in improving the accuracy of estimates, which can directly influence biological interpretation and management decision-making.

4. CONCLUSIONS

The analyses of the genetic structure and diversity of the Girolando population highlight high genetic variability among the different genetic groups. The continued use of Holstein and Gir bulls has ensured the preservation of genetic variability in the Girolando breed, with different contributions in the pedigree.

Milk production varied between the analyzed genetic groups, particularly the increase in production in the 1/2H + 1/2G cows, which resulted from the crossbreeding between the breeds. The estimates of heritability and repeatability indicated that, despite the influence of environmental factors, selection remains an efficient strategy for improving milk production.

5. REFERENCES

- Barbero MMD, Fort NM, Schultz EB, de Melo ALP, Moura AM (2022) Estimation of genetic parameters in dairy production in Girolando cattle. **Brazilian Animal Science** 23:e-72300P.
- Barbosa BL, Silva APS, Castro DP, Castro GC, Silva TS, Vieira RB, Sena LS, Torres TS, Pereira EHU, Silva LRG, Borges LS, Oliveira MB, Sarmento JLR (2023) Overview of the use of genomic data in animal breeding. **Ciência Rural** 53:e20220350.
- Bermann, M., Cesarani, A., Misztal, I., Lourenco, D. (2022) Past, present, and future developments in single-step genomic models. **Italian Journal of Animal**
- Berry DP, Cromie AR (2007) Associations between age at first calving and subsequent performance in Irish spring-calving Holstein-Friesian dairy cows. **Livestock Science** 106: 182-193, 2007.
- Booy, G., Hendriks, R.J.J., Smulders, M.J.M., Van Groenendael, J.M., Vosman, B. (2008) Genetic Diversity and the Survival of Populations. **Plant Biology**, 2:379-395.
- Braga AP, Carneiro Júnior JM, Pinheiro AK, Silva MS (2020) Genetic parameters of Girolando crossbred cows in dairy herds in the state of Acre, Brazil. **Arquivos de Ciências Veterinárias e Zoologia da UNIPAR** 21:e2311.
- Brito LF, Bedere N, Douhard F, Oliveira HR, Arnal M, Peñagaricano F, Schinckel AP, Baes CF, Miglior F (2021) Review: Genetic selection of high-yielding dairy cattle toward sustainable farming systems in a rapidly changing world. **Animal** 15:100292.
- Burrow HM (2012) Importance of adaptation and genotype x environment interactions in tropical beef breeding systems. **Animal** 6:729-740.
- Canaza-Cayo AW, Lopes PS, Silva MVGB, Cobuci JA, Torres RA, Martins MF, Arbex WA (2014b) Estrutura populacional da raça Girolando. **Ciência Rural** 44:2072-2077.
- <https://www.scielo.br/j/cr/a/PNshC3mykJQkxVmDyWGCWHk/abstract/?lang=pt>
- Canaza-Cayo AW, Lopes PS, Silva MVGB, Torres RA, Martins MF, Arbex WA, Cobuci JA (2015) Genetic Parameters for Milk Yield and Lactation

- Persistency Using Random Regression Models in Girolando Cattle. **Asian Australas. J. Anim. Sci.** 28:1407-1418.
- Canaza-Cayo, A.W., Lopes, P.S., Cobuci, J.A., Martins, M.F., Silva, M.V.G.B. (2014a) Genetic parameters of milk production and reproduction traits of Girolando cattle in Brazil. **Italian Journal of Animal Science** 17:22-30.
- Costa CN, Madalena FE, Teodoro RL, Lopes PS, Verneque RS, Freitas AF (2005) Genetic parameters for test-day milk yield of first lactation Holstein cows estimated by random regression using Legendre polynomials. **Revista Brasileira de Zootecnia** 34:388-398.
- Costa CN, Martinez ML, Verneque RS, Teodoro RL, Ledic IL (2004) Heterogeneidade de (Co)variância para as Produções de Leite e de Gordura entre Vacas Puras e Mestiças da Raça Gir. **R. Bras. Zootec.** 33:555-563.
- Daltro, D.S., Silva, M.V.G.B., Gama, T., Machado, J.D., Kernd, E.L., Campos, G.S., Panetto, J.C.C., Cobuci, J.A. (2019) Estimates of genetic and crossbreeding parameters for 305-day milk yield of Girolando cows. **Italian Journal of Animal Science** 19:86-94.
- Das DN, Karuthadurai T, Gnanasekaran S (2021) Chapter 10 – Genomic selection: a molecular tool for genetic improvement in livestock. **Advances in Animal Genomics** 141-163.
- Dekkers, J.C.M. (2012) Application of Genomics Tools to Animal Breeding. **Current Genomics** 13:207-212.
- Dionello NJL, Bergamaschi MACM, Fischer V, Ribas NP, Stumpf Júnior W (2006) Parâmetros genéticos para produção de leite no dia do controle e produção de leite acumulada até 305 dias em vacas da raça Holandesa no Rio Grande do Sul. **Revista Brasileira de Zootecnia** 35:1591-1601.
- Druet T, Jaffrézic F, Boichard D, Ducrocq V (2003) Modeling Lactation Curves and Estimation of Genetic Parameters for First Lactation Test-Day Records of French Holstein Cows. **J. Dairy Sci.** 86:2480-2490.
- El Faro L, Albuquerque LG (2003) Estimação de Parâmetros Genéticos para Produção de Leite no Dia do Controle e Produção Acumulada até 305 dias, para as Primeiras Lactações de Vacas da Raça Caracu. **Revista Brasileira de Zootecnia** 32:284-294.

- El-Awady HG (2013) Genetic aspects of lactation curve traits and persistency indices in Friesian cows. **Archiva Zootechnica** 16:15-29.
- Facó O, Martins Filho R., Lôbo RNB, Oliveira SMP, Martins GA (2007) Heterogeneidade de (co)variância para a produção de leite nos grupos genéticos formadores da raça Girolando. **Rev. Ciên. Agron.** 38:304-309.
- Geweke, J. (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. **Bayesian Statistics** 4:169–193.
- González Herrera LG, El Faro L, Albuquerque LG, Tonhati H, Machado CHC (2008) Estimativas de parâmetros genéticos para produção de leite e persistência da lactação em vacas Gir, aplicando modelos de regressão aleatória. **Revista Brasileira de Zootecnia** 37:1584-1594.
- Gutierrez-Reinoso, M.A., Aponte, P.M., Garcia-Herreros, M. (2021) Genomic Analysis, Progress and Future Perspectives in Dairy Cattle Selection: A Review. **Animals**, 11:599.
- Haile-Mariam M, Bishop SC, Jackson F, von Samson-Himmelstjerna G, Greer AW, Kenyon F, Sargison ND, Höglund J (2015) Comparison of genetic and non-genetic factors influencing nematode resistance in sheep in different countries. **Animal** 9:680-686.
- Harper JL, Hawksworth DL (1994) Biodiversity: measurement and estimation. **Phil. Trans. R. Soc. Lond. B** 345:5-12.
- Heidelberger P, Welch PD (1983) Simulation run length control in the presence of an initial transient. **Opns Res.** 31:1109-1144.
- Herrera AC, Múnera OD, Cerón-Muñoz MF (2013) Variance components and genetic parameters for milk production of Holstein cattle in Antioquia (Colombia) using random regression models. **Revista Colombiana de Ciencias Pecuarias** 26:90-97.
- Hill WG (2010) Understanding and using quantitative genetic variation. **Philosophical Transactions of the Royal Society B: Biological Sciences** 365:73-85.
- Ismael H, Janković D, Stanojević D, Bogdanović V, Trivunović S, Djedović R (2021) Estimation of heritability and genetic correlations between milk yield and linear type traits in primiparous Holstein-Friesian cows. **Revista Brasileira de Zootecnia** 50:e20200121.

- Jacopini LA, Barbosa SBP, Lourenço DAL, Silva MVGB (2016) Desempenho produtivo de vacas Girolando estimado pelo modelo de Wood ajustado por metodologia Bayesiana. **Archives of Veterinary Science** 21:43-54.
- Kelleher MM, Berry DP, Kearney JF, McParland S, Buckley F, Purfield DC (2017) Inference on population structure of purebred dairy and beef cattle using high-density genotype data. **Animal** 11:15-23.
- Koenig S, Simianer H, Willam A, Schaeffer LR, Kropf B (2005) Genetic evaluation of dairy cattle accounting for genotype by environment interactions. **Journal of Dairy Science** 88:883-893.
- Legarra, A, Aguilar, I and Misztal, I (2009) A relationship matrix including full pedigree and genomic information. **Journal of Dairy Science** 92, 4656-4663.
- Lynch M, Walsh B (1998) Genetics and Analysis of Quantitative Traits. **Sunderland: Sinauer Associates** 980p.
- Madalena FE (2008) Considerations on the impact of crossbreeding on the various animal genetic resources of the tropics. **Animal Genetic Resources Information** 42:49-63.
- Marchioretto PV, Rabel RAC, Allen CA, Ole-Neselle MMB, Wheeler MB (2023) Development of genetically improved tropical-adapted dairy cattle. **Anim. Front.** 13:7-15.
- McVean G (2009) A Genealogical Interpretation of Principal Components Analysis. **PLoS Genetics** 5:e1000686.
- Melo CMR, Packer IU, Costa CN, Machado PF (2005) Parâmetros Genéticos para as Produções de Leite no Dia do Controle e da Primeira Lactação de Vacas da Raça Holandesa. **Revista Brasileira de Zootecnia** 34:796-806.
- Misztal I, Legarra A, Aguilar I (2014) Using recursion to compute the inverse of the genomic relationship matrix. **Journal of Dairy Science** 97:3943-3952.
- Misztal I, Tsuruda S, Aguilar I, Legarra A, VanRaden PM, Lawlor TJ (2013b) Methods to approximate reliabilities in single-step genomic evaluation. **Journal of Dairy Science** 96:647-654.
- Misztal, I., Aggrey, S.E., Muir, W.M. (2013a) Experiences with a single-step genome evaluation. **Poultry Science**, 92:2530-2534.
- Nishio M, Inoue K, Ogawa S, Ichinoseki K, Arakawa A, Fukuzawa Y, Okamura T, Kobayashi E, Taniguchi M, Ishii K (2023) Comparing pedigree and genomic

- inbreeding coefficients, and inbreeding depression of reproductive traits in Japanese Black cattle. **BMC Genomics** 24:376.
- Notter, D.R. (1999) The importance of genetic diversity in livestock populations of the future. **Journal of Animal Science**, 77:61-69.
- Oftedal OT, Iverson SJ (1995) Comparative Analysis of Nonhuman Milks: A Phylogenetic Variation in the Gross Composition of Milks. **Food Science and Thecnology** 749-789.
- Panetto, J.C.C., Silva, M.V.G.B., Machado, M.A., Verneque, R.S., Fernandes, A.R., Machado, C.H.C., Martins, M.F., Faza, D.R.L.R., Borges, C.A.V., Ventura, H.T., Pereira, M.A., Oliveira, J.C., Glatzl Junior, L.A., Garcia, A.O., Leandro, F.D. (2022) Providing genetics for the dairy industry in the tropics – the Brazilian Dairy Gir Breeding Program. **World Congress on Genetics Applied to Livestock Production**.
- Pereira, R.J., Santana Jr, M.L, Ayres, D.R., Bignardi, A.B., Menezes, G.R.O., Silva, L.O.C., Machado, C.H.C., Josahkian, L.A., Albuquerque, L.G. (2016) Inbreeding depression in Zebu cattle traits. **Animal Breeding and Genetics** 1-11.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, Bakker PIW, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. **The American Journal of Human Genetics** 81:559-575.
- Roka, P., Shrestha, S., Adhikari, S. P., Neupane, A., Shreepaili, B., & Bista, M. K. (2024) A review on genetic parameters estimation, trait association, and multivariate analysis for crop improvement. **Archives of Agriculture and Environmental Science**, 9:618 – 625.
- Rosa MS, Andrade FM, Machado JD, Cobuci JA (2023) Population structure of Girolando cattle using pedigree analysis. **SSRN**.
- Sargolzaei M, Chesnais JP, Schenkel FS (2014) A new approach for efficient genotype imputation using information from relatives. **BMC Genomics** 15:478.
- Searle SR (1961) Part Lactations. II. Genetic and phenotypic studies of monthly milk fat yield. **Journal of Dairy Science** 44:282-295.

- Silva MHMA, Malhado CHM, Kern EL, Daltro DS, Cobuci JA, Carneiro PLS (2019) Inbreeding depression in Holstein cattle in Brazil. **Revista Brasileira de Zootecnia** 48:e20170212.
- Silva, M.V.G.B., Martins, M.F., Cembranelli, M.A.R., Panetto, J.C.C., Gonçalves, G.S., Machado, M.A., Faza, D.R.L.R., Alves, B.R.C. (2018) **Girolando breeding program – Sire Summary – Progeny Test Results**, 1^a Edition.
- Smith BJ (2007) boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference. **J. Stat. Soft.** 21:1-37.
- Snelder J., Larenstein V.H. (2019) Inbreeding versus Crossbreeding: the Potential Bias of Breeding Values in Dutch Dairy Cattle. **Student Undergraduate Research E-Journal!** 5, 1–4
- Van Vleck LD, Henderson CR (1961) Estimates of genetic parameters of some functions of part lactation milk records. **Journal of Dairy Science**
- VanRaden PM (1992) Accounting for Inbreeding and Crossbreeding in Genetic Evaluation of Large Populations. **Journal of Dairy Science** 75:3136-3144.
- VanRaden, PM (2008) Efficient methods to compute genomic predictions. **Journal of Dairy Science** 91 4414-4423.
- Veerkamp RF, Mulder HA, Thompson R, Calus MPL (2011) Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. **Journal of Dairy Science** 94:4189-4197.
- Weigel, K.A. (2017) Genomic selection of dairy cattle: A review of methods, strategies, and impact. **Journal of Animal Breeding and Genomics** 1:1-15.
- Weir BS, Hill WG (2002) Estimating F-Statistics. **Annu. Rev. Genet.** 36:721-750.
- Zumbach B, Peters KJ (2002) Sustainable breeding goals and selection criteria for Holstein Friesian cattle in the tropics. **Interbull Bulletin** 29:147-152.

Appendix

Appendix 1. Distribution of heterozygosity and number of animals by genetic group

Genetic groups	Heterozygosity	Number of animals
H	0,005	427
G	0,005	81
1/2H + 1/2G	0,005	16
1/2H + 1/2G	0,500	46
1/2H + 1/2G	1,000	25508
1/4H + 3/4G	0,500	556
1/8H + 7/8G	0,875	99
3/4H + 1/4G	0,005	3
3/4H + 1/4G	0,250	1726
3/4H + 1/4G	0,500	17247
3/8H + 5/8G	0,750	572
5/8H + 3/8G	0,005	12
5/8H + 3/8G	0,375	2083
5/8H + 3/8G	0,500	594
5/8H + 3/8G	0,750	4045
7/8H + 1/8G	0,005	8
7/8H + 1/8G	0,250	3320
7/8H + 1/8G	0,500	3

CHAPTER 4 – FINAL CONSIDERATIONS

The productive and genetic evaluation of Girolando cattle conducted in this study shows important contrasts between milk yield performance and the genetic structure of the different genetic compositions analyzed. While the milk production results revealed significant variation across lactation stages and among genetic groups, the indicators of genetic diversity — such as inbreeding coefficient, heterozygosity, and population structure — provided a broader understanding of genetic potential and the associated risks of inbreeding.

From a productive perspective, animals with a higher proportion of Holstein genetics, particularly the 1/2H + 1/2G group, showed superior milk yield averages, reaching 23.2 kg/day. In contrast, animals from the G group presented the lowest mean yield (16.7 kg/day), reflecting the breed's adaptive profile rather than high dairy productivity. The lactation curve exhibited a production peak between 91 and 120 days in milk (TDMY 4), followed by a gradual decline in subsequent periods. The use of fixed regressions by genetic group proved to be an appropriate modeling strategy, allowing for more accurate estimation of additive genetic effects.

The use of Random Regression Models (RRMs) was fundamental to the analytical approach adopted in this study. RRM are particularly suited for the longitudinal nature of test-day milk yield data, as they allow the modeling of genetic and environmental effects as continuous functions over time. This methodology enabled the estimation of time-specific genetic parameters and better captured the dynamic trajectory of milk production throughout lactation, surpassing the limitations of traditional cumulative models.

The joint interpretation of both data sets suggests that groups with greater genetic diversity — characterized by lower inbreeding levels and higher heterozygosity — also tended to perform better in milk production, reinforcing the benefits of heterosis in dairy systems. This outcome is consistent with the literature, which emphasizes the strategic use of crossbreeding to enhance productive performance and adaptability, especially under tropical environments.

Additionally, variance components throughout lactation indicated that additive genetic and permanent environmental effects varied according to the lactation stage, whereas residual variance remained relatively stable. The heritability (0.48 ± 0.11) and repeatability (0.84 ± 0.02) estimates were higher than those commonly reported for the breed, likely due to the application of the test-day model, the large volume of data, and the incorporation of genomic information.

In this context, the findings underscore the importance of integrating productive and genetic data in the design of breeding strategies. The combination of high milk yield and increased genetic diversity observed in crossbred groups, particularly the 1/2H + 1/2G, stands out as a promising alternative for sustainable selection programs. The incorporation of RRM, especially when combined with genomic information, proved to be a robust approach that enhances the accuracy of genetic evaluations and supports more efficient and precise decision-making in dairy cattle breeding under tropical production conditions.