GMR

The Original by FUNPEC-RP

Genomic prediction of lactation curves of Girolando cattle based on nonlinear mixed models

F.R.F. Teixeira[1], M. Nascimento[2], P.R. Cecon[2], C.D. Cruz[3],
F.F. e Silva[4], A.C.C. Nascimento[2]; C.F. Azevedo[2], D.B.D. Marques[4],
M.V.G.B. da Silva[5], A.P.S. Carneiro[2] and D.M. Paixão[6]

[1] Universidade Federal do Piauí, Coordenação de Estatística, Campus Universitário inistro Petrônio Portella, Ininga, Teresina, PI, Brasil
[2] Universidade Federal de Viçosa, Departamento de Estatística, Viçosa, MG, Brasil
[3] Universidade Federal de Viçosa, Departamento de Biologia Geral, Viçosa, MG, Brasil
[4] Universidade Federal de Viçosa, Departamento de Zootencia, Viçosa, MG, Brasil
[5] Embrapa Gado de Leite, –Juíz de Fora, MG, Brasil
[6] Pacege/SEALQ/USP, Núcleo Acadêmico, Piracicaba, SP, Brasil

Corresponding author: F.R.F. Teixeira
E-mail: filiperfteixeira@ufpi.edu.br

ABSTRACT. Knowledge of lactation curves in dairy cattle is essential for understanding the animal production in milk production systems. Genomic prediction of lactation curves represents the genetic pattern of milk production of the animals in the herd. In this context, we made genomic predictions of lactation curves through genome-wide selection (GWS) to characterize the genetic pattern of lactation traits in Girolando cattle based on parameters estimated by nonlinear mixed effects (NLME) models. Data of 1,822 milk control records from 226 Girolando animals genotyped for 37,673 single nucleotide polymorphisms were analyzed. Nine NLME models were compared to identify the equation with the best fit. The lactation traits estimated by the best model were submitted to GWS analysis,

using the Bayesian LASSO method. Then, based on the genomic estimated breeding values (GEBVs) obtained, genomic predictions of lactation curves were constructed, and the genetic parameters were calculated. Wood's equation showed the best fit among the evaluated models. Heritabilities ranged from 0.09 to 0.29 for the seven lactation variables (initial production, rates of increase and decline, lactation peak, time to peak yield, persistence and total production). The correlations among GEBVs ranged from -0.85 to 0.98. The concordances between the best animals selected according to the selected traits were greater when the correlations between GEBVs for these traits were also high. Consequently, the methodology allowed us to identify the best nonlinear model and to construct the genetic lactation curves of a Girolando cattle population, as well as to assess the differences between animals and the association between lactation variables.

Key words: Heritability; Bayesian LASSO; Genome-wide selection

INTRODUCTION

About 80% of the milk produced in Brazil comes from cows that have Holstein and Gyr genes in their genetic composition (Daltro et al., 2019; Silva et al., 2015). The Girolando breed, predominant in Brasil, was developed beginning in the 1940s by crossing Gyr and Holstein (Canaza-Cayo et al., 2015).

Studies addressing the construction of lactation curves for dairy cattle usually aim to identify the lactation behavior of the herds (Bangar and Verma, 2017) and/or to evaluate the genetic parameters for the lactation traits under study (Macciotta et al., 2015). These approaches are important for characterizing the milk productivity of the herd and for assisting the choice of the equation that best describes its behavior.

Lactation curves can be characterized by nonlinear mixed equations (NLME) models. Based on the large amount of this class of models available in literature, their comparison becomes extremely important. In this sense, the models fit can be evaluated by goodness-of-fit measures, such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), residual standard deviation and Root Mean Square Error (RMSE), and the equation that best describes the productive behavior of the herd can be identified and used to estimate the lactation curve parameters.

As the differences between the measures used to construct the dairy cattle lactation curves usually occur due to genetic influences, methodologies that aim to find the functional relationship between phenotypes and DNA information, based on molecular markers, have been used (Macciotta et al., 2015; Cardona et al., 2015). Among them, the Genome Wide Selection - GWS (Meuwissen et al., 2001) is highlighted, which aims to predict the animal genetic merit by building statistical models using information directly from the genome through genetic markers. Thus, through the application of statistical methods applied to GWS (e.g., the Bayesian LASSO) considering the estimates of the lactation curve parameters as dependent variables, the prediction of Genomic Estimated Breeding Values

(GEBVs) and the estimation of genetic parameters for the lactation traits, as well as the construction of genomic prediction of lactation curves, are allowed.

In this context, we aimed to: i) compare nine different NLME models based on goodness-of-fit criteria to identify the equation that best fits the lactation data of a Girolando dairy cattle population; ii) use the model with the best goodness-of-fit to characterize the milk productivity of the herd; and iii) build the genomic prediction of lactation curves of the herd through GWS and estimate the heritabilities of the lactation variables and the correlations among the GEBVs of the animals for the lactation traits.

MATERIAL AND METHODS

Data

The data on milk production and molecular SNP markers were provided by EMBRAPA (National Center for Research on Dairy Cattle), located in Juiz de Fora, Minas Gerais, Brazil (21º46'55.8''S, 43º22'10.7''W). Initially, the available information corresponded to 94,263 milk control measures from 11,459 Girolando animals. Aiming to work only with individuals that had available molecular information - SNPs (single nucleotide polymorphisms) markers, data from 226 genotyped animals, composed of 1,822 records of milk control of Girolando cattle, were used. The animals were genotyped with an Illumina BovineSNP50 BeadChip, containing initially information of 54,609 SNPs markers. The filters used in quality control were: *mind* - 0.05 (maximum per-person missing); *geno* - 0.1 (maximum per-SNP missing); *maf* - 0.5 (minor allele frequency); and *me* - 0.05 (mendel error rate thresholds - per SNP, per family). After the quality control filters, 37,673 markers remained to apply the methodologies used in this study.

Nonlinear mixed models for lactation curves

According to Lindstrom and Bates (1990), the general nonlinear mixed model can be represented as follows:

$$y_{ij} = f(\phi_i; t_{ij}) + \varepsilon_{ij}, \qquad (Eq. 1)$$

wherein y_{ij} is the i-th observation of the response variable, represented by milk yield (kg day) at time j (in days); $f(\phi_i; t_{ij})$ is a pre-specified nonlinear equation (Table 1) that relates time to milk yield; ϕ_i is the vector of parameters, which varies according to the equation; and ε_{ij} is the error associated to the observation y_{ij}.

In the NLME approach, each parameter of the equations presented in Table 1 is decomposed, with fixed and random effects assigned to each of them. Considering, for example, Wood's equation (Wood, 1967), the decomposition can be performed as follows:

$$y_{ij} = \left(\mu_{\beta_1} + PL_i + \xi_{ai}^2\right) t_{ij}^{\left(\mu_{\beta_2} + PL_i + \xi_{bi}^2\right)} e^{-\left(\mu_{\beta_3} + PL_i + \xi_{ci}^2\right) t_{ij}} + \varepsilon_{ij}. \qquad (Eq. 2)$$

It is assumed that $\beta_{1i} = \mu_{\beta_1} + PL_i + \xi_{ai}^2$, $\beta_{2i} = \mu_{\beta_2} + PL_i + \xi_{bi}^2$ and $\beta_{3i} = \mu_{\beta_3} + PL_i + \xi_{ci}^2$, in which $\mu_{\beta_1}, \mu_{\beta_2}$ and μ_{β_3} are the averages of fixed effects for each parameter; PL_i is the effect of the production level of the i-th individual; ξ_{ai}^2, ξ_{bi}^2 and ξ_{ci}^2 are the

individual random effects for each parameter of the model; and ε_{ij} is the residual effect associated with observation y_{ij}. For the establishment of the production levels (PL_i's), the animals were grouped by a cluster analysis, using the Gower algorithm (Gower, 1971) and considering the variables: milking number (3 groups), age (4 groups), genetic grouping (3 groups) and contemporary groups (139 groups), resulting in a total of four different production levels.

Table 1. Nonlinear equations proposed to estimate lactation curves

Author (s)	Equation
Brody *et al.* (1923)	$y_t = \beta_1 e^{-\beta_3 t}$
Sikka (1950)	$y_t = \beta_1 e^{(\beta_2 t - \beta_3 t^2)}$
Nelder (1966)	$y_t = \dfrac{t}{\beta_1 + \beta_2 t + \beta_3 t^2}$
Wood (1967)	$y_t = \beta_1 t^{\beta_2} e^{-\beta_3 t}$
Cobby and Le Du (1978)	$y_t = \beta_1 - \beta_2 t - \beta_1 e^{-\beta_3 t}$
Dhanoa (1981)	$y_t = \beta_1 t^{\beta_2 \beta_3} e^{-\beta_3 t}$
Papajcsik and Bodero (1988)	$y_t = \beta_1 t e^{-\beta_3 t}$
Rook (1993)	$y_t = \beta_1 \left(\dfrac{1}{1 + \dfrac{\beta_2}{\beta_3 + t}} \right) e^{-\beta_4 t}$
Cappio-Borlino *et al.* (1995)	$y_t = \beta_1 t^{\beta_2 \exp(-\beta_3 t)}$

y_t is the test day (t) milk yield, measured from the beginning of lactation; β_1, β_2, β_3 and β_4 are the parameters that define the scale and shape of the lactation curve, varying according to the equation.

Once the function used to fit the curves has been established, important information regarding lactation can be extracted: the time to peak yield (tp), which is the time (X axis) that the animal takes to reach the maximum point in milk yield; the milk yield at peak (Y_{max}), that corresponds to the maximum point of the milk yield curve (Y axis); total production (TP), which is the absolute milk yield of each individual during the lactation period; and persistence (PS), corresponding to the capacity of each individual to keep the production level after the peak yield.

Considering the Wood's equation (Wood, 1967), these measures are represented by (Centoducati *et. al*., 2012):

$$tp_i = \frac{\hat{\beta}_{2i}}{\hat{\beta}_{3i}}, \qquad (\text{Eq. 3})$$

$$TP_i = \int_0^{305} \hat{\beta}_{1i} t^{\hat{\beta}_{2i}} e^{-\hat{\beta}_{3i} t} dt, \qquad (\text{Eq. 4})$$

$$Y_{max.i} = \hat{\beta}_{1i} (\hat{\beta}_{2i}/\hat{\beta}_{3i})^{\hat{\beta}_{2i}} e^{-\hat{\beta}_{2i}}, \qquad (\text{Eq. 5})$$

$$PS_i = -(\hat{\beta}_{2i} + 1) \ln(\hat{\beta}_{3i}), \qquad (\text{Eq. 6})$$

wherein $\hat{\beta}_{1i}$, $\hat{\beta}_{2i}$ and $\hat{\beta}_{3i}$ are the individual estimates of the parameters of Wood's equation (Wood, 1967) presented in Table 1.

The estimation of the parameters of the nonlinear mixed models was performed using the algorithm proposed by Lindstrom and Bates (1990), which consists of alternating

between two steps: 1. PNLS (penalized nonlinear least squares) step; and 2. LME (Linear Mixed Effects) step.

Fit comparison

The equations described in Table 1 were compared according to the AIC and BIC criteria; the residual standard deviation and the RMSE. Durbin-Watson statistics and test were also used to assess the residual autocorrelation in the models. These measures are described below.

The AIC (Akaike,1973) measures the suitability of a model through the Kullback-Leibler distance (K-L), according to the following equation:

$$\text{AIC} = -2\log L\left(\widehat{\phi}\right) + 2n_{par}, \tag{Eq. 7}$$

wherein $L\left(\widehat{\phi}\right)$ is the maximum of likelihood function and n_{par} is the number of parameters considered in the model.

The BIC, proposed by Schwarz (1978), is defined as the statistic that maximizes the probability of identifying the true model among those evaluated and can be defined as:

$$\text{BIC} = -2\log L\left(\widehat{\phi}\right) + n_{par}\log(M), \tag{Eq. 8}$$

wherein $L\left(\widehat{\phi}\right)$ is the maximum of likelihood function, n_{par} is the number of parameters considered in the model and M is the total number of observations. According to Pinheiro and Bates (2000), under these definitions, the lower the BIC value, the better the model. The same interpretation extends to AIC.

The RMSE functions as a generalized standard deviation (Ghavi Hossein-Zadeh, 2015) and can be calculated by the following expression:

$$\text{RMSE} = \sqrt{\frac{SSE}{M - n_{par} - 1}}, \tag{Eq. 9}$$

wherein SSE is the error sum of squares of the model, M is the number of observations and n_{par} is the number of fixed effect parameters in the model. As this measure is associated with the error sum of squares, the best model will be the one with the lowest RMSE.

The Durbin-Watson statistic (Durbin and Watson, 1971), also evaluated by statistical test, aims to identify the presence of autocorrelation between the residuals of the specified model. This statistic is given by:

$$\text{DW} = \frac{\sum_{k=1}^{M}(e_k - e_{k-1})^2}{\sum_{k=1}^{M} e_k^2}, \tag{Eq. 10}$$

wherein e_k is the residue of the k-th observation. This measure can vary between 0 and 4, and there will be signs of autocorrelation when it is close to 2. The assumption of the test is that the errors of the regression model are generated by a first-order autoregressive process, i.e.:

$$e_k = \rho e_k + a_k, \tag{Eq. 11}$$

wherein $a_k \sim N(0, \sigma^2)$ and ρ is the correlation parameter, complying $-1 < \rho < 1$, on which the test is based. The hypotheses of the Durbin-Watson test are given by

$$\begin{cases} H_0: \rho = 0 \\ H_1: \rho \neq 0 \end{cases}$$

The null hypothesis of the test assumes that the errors are independent, i.e., if it is concluded that $\rho = 0$, the error e_k will be normally distributed with mean 0 and constant variance. On the other hand, considering the alternative hypothesis, the conclusion is that $\rho \neq 0$, i.e., there is dependence between residues according to a first-order autoregressive structure (Piccardi et al., 2017).

Genome-Wide Selection

The estimates of the parameters from the best model among those evaluated were used as pseudo-phenotypes for calculating the SNP effects using the statistical methods applied to GWS. Subsequently, the estimates of SNP effects were used to predict the individuals GEBVs. Considering each lactation variable (initial production, rate of increase, rate of decline, lactation peak, time to peak yield, persistence and total production) obtained in the previous step, the general model of genomic selection associated with the i-th individual ($i = 1, \ldots, 223$) and considering 37,673 SNPs is given by (Meuwissen et al., 2001):

$$y = 1\mu + X\beta + e, \tag{Eq. 12}$$

wherein y is the vector (223×1) containing the estimates of the abovementioned variables; 1 is the vector of ones (223×1); μ is the general mean; X is the incidence matrix of markers ($223 \times 37,673$); β is the vector of marker effects ($37,673 \times 1$); and e is the vector of errors (223×1). In our study, the BLASSO method (Bayesian Least Absolute Shrinkage and Selection Operator) (DE LOS CAMPOS *et al.*, 2009) was used to estimate marker effects and predict the GEBVs.

Based on the genotypic and phenotypic variances, the heritabilities of the traits were estimated, which are calculated as the proportion of the total (phenotypic) variance explained by the genotypic variance, i.e.:

$$h_i^2 = \frac{V_{gen.i}}{V_{fen.i}}, \tag{Eq. 13}$$

wherein V_{gen} is the genotypic variance and V_{fen} is the phenotypic variance.

The GEBVs of the animals for the seven traits of the lactation curve (initial production, rate of increase, rate of decline, lactation peak, time to peak yield, persistence, and total production) were obtained as follows:

$$GEBV = X\hat{\beta}, \tag{Eq. 14}$$

wherein X is the incidence matrix previously described and $\hat{\beta}$ is the estimated vector of markers effects.

Kappa coefficient

The selection of individuals based on the different genomic predictions of the lactation parameters was compared using the Kappa coefficient of agreement (Cohen, 1960). This method assesses, considering the probability of random selection, the agreement

between the individuals selected according to the different predictions. The calculation of this coefficient can be done as follows:

$$\hat{k} = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}, \qquad \text{(Eq. 15)}$$

wherein the numerator $\Pr(a) - \Pr(e)$ represents the proportion of observations in which the agreement occurred beyond what was randomly expected, and the denominator $1 - \Pr(e)$ denotes the proportion of observations in which there was no agreement, also considering the information at random. This coefficient varies from 0 to 1 and, as the simple coefficient of agreement, the agreement increases as it approaches 1.

Computational methods

All the statistical analyses were performed in R software (R Core Team, 2018). The fit of the nonlinear mixed models was performed using the Linstrom and Bates (1990) algorithm, implemented in the *nlme* package (Pinheiro et al., 2017). The cluster analysis to obtain the fixed effect factors of the model was performed using the *cluster* package (Maechler et al., 2017). Genomic Selection analysis was carried out in *BGLR* package (De Los Campos and Rodriguez, 2016). In total, 100,000 iterations, burn-in of 20,000 and thin of 10 were used. The descriptive measures were obtained using the *dplyr* package (Wickham et al., 2017).

RESULTS

For the estimation of fixed effects, the individuals were separated into four groups of production levels by grouping the variables previously described. The groups and descriptive measures are shown in Table 2.

Table 2. Means, standard deviations (sd) and amplitude of milk production records according to production levels.

Production levels	N[1]	Average production±sd (kg)	Minimum (kg)	Maximum (kg)
1 (n=143)	1,12	14.18±7.00	1.80	47.20
2 (n=43)	378	18.33±8.87	2.50	53.00
3 (n=10)	81	19.11±9.55	3.80	48.40
4 (n=30)	248	23.06±9.04	3.90	48.40

1: Total number of observations per group.

The fourth group presented the highest average production among all groups (23.06 kg day), which was formed by 248 observations from 30 animals. The largest group was composed of 143 animals (approximately 63% of the sample) and had the lowest average production.

The values of AIC and BIC criteria ranged, respectively, from 10,013.79 to 12,625.04 and from 10,101.92 to 12,713.16 (Table 3). In general, four models stood out regarding the criteria analyzed: Wood (1967), Cobby and Le Du (1978), Nelder (1966) and Rook (1993).

Wood's model was the most relevant regarding the AIC and BIC criteria (10,013.79 and 10,101.92, respectively), and also presented low values for the residual standard deviation ($\hat{\sigma}$ = 2.71) and RMSE (2.42). Considering the residual standard deviation ($\hat{\sigma}$) and the RMSE criteria, the best model was Cobby and Le Du ($\hat{\sigma}$ = 2.70; RMSE = 2.36), followed by Wood ($\hat{\sigma}$ = 2.71; RMSE = 2.42). Other equations, such as Nelder (1966) and Rook (1993), also fitted the data well (Table 3). The highest values of AIC and BIC corresponded to the models of Brody et al. (1923), Papajcsik and Bodero (1988) and Dhanoa (1981). The other results for the goodness-of-fit measures are detailed in Table 3.

Table 3. Goodness-of-fit measures of the nine nonlinear models for the genomic prediction of lactation curves.

Author (s)	AIC[1]	BIC[2]	$\hat{\sigma}$[3]	RMSE[4]	DW[5]
Brody et al. (1923)	11,390.06	11,450.64	4.66	4.40	1.05*
Sikka (1950)	10,887.88	10,976.00	3.40	3.06	1.61*
Nelder (1966)	10,127.15	10,215.27	2.76	2.46	1.92ns
Wood (1967)	10,013.79	10,101.92	2.71	2.42	2.04ns
Cobby and Le Du (1978)	10,198.85	10,286.98	2.70	2.36	2.05ns
Dhanoa (1981)	12,625.04	12,713.16	7.67	7.66	0.35*
Papajcsik and Bodero (1988)	11,782.34	11,842.92	5.17	4.86	1.36*
Rook (1993)	10,100.85	10,216.51	2.81	2.52	2.00*
Cappio-Borlino et al. (1995)	10,637.92	10,726.05	3.66	3.37	1.45*

1: Akaike Information Criterion; 2: Bayesian Information Criterion; 3: Residual Standard Deviation; 4: Root Mean Square Error; 5: Durbin-Watson Statistics; *: Significant autocorrelation at the 5% level; ns: Non-significant autocorrelation at the 5% level.

Regarding the presence of residual autocorrelation, the Durbin-Watson statistic ranged from 1.05 to 2.05, with values close to 2 for most models. According to the Durbin-Watson statistical test, Nelder, Wood and Cobby and Le Du models did not show significant residual correlations. As the Wood's model presented the lowest AIC and BIC values, in addition to suitable results for the other goodness-of-fit measures, the subsequent analyses of the herd lactation traits were based on this model.

After estimating the parameters of the lactation curve ($\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$, corresponding to the initial production, rate of increase and rate of decline, respectively) of the 223 animals by the Wood's model, these estimates were submitted to statistical analysis applied to GWS using the BLASSO method. The descriptive statistics of the GEBVs estimated by BLASSO based on the markers effects is shown in Table 4.

Table 4. Descriptive statistics of the genomic breeding values of the animals for the lactation traits estimated by BLASSO.

Traits	Mean	SD[1]	CV(%)[2]	Minimum	Maximum
$\hat{\beta}_1$ – Initial production (kg day)	8.56	0.22	2.58	8.09	9.21
$\hat{\beta}_2$ – Rate of increase	0.2358	0.0033	1.39	0.2283	0.2485
$\hat{\beta}_3$ – Rate of decline	0.0054	0.0001	1.61	0.0052	0.0057
Total production (kg/lactation)	4,680.00	515.80	11.02	3,837.00	7,306.00
Time to peak yield (days)	67.55	6.25	9.24	58.59	107.96
Lactation peak (kg day)	18.77	0.91	4.84	17.21	23.29
Persistence	6.91	0.12	1.78	6.66	7.38

1: Standard deviation; 2: Coefficient of variation.

The coefficient of variation for the GEBVs ranged from 1.39% for the rate of increase to 11.02% for the total production. The behavior of the genomic prediction of lactation curves can be seen in Figures 1 and 2.
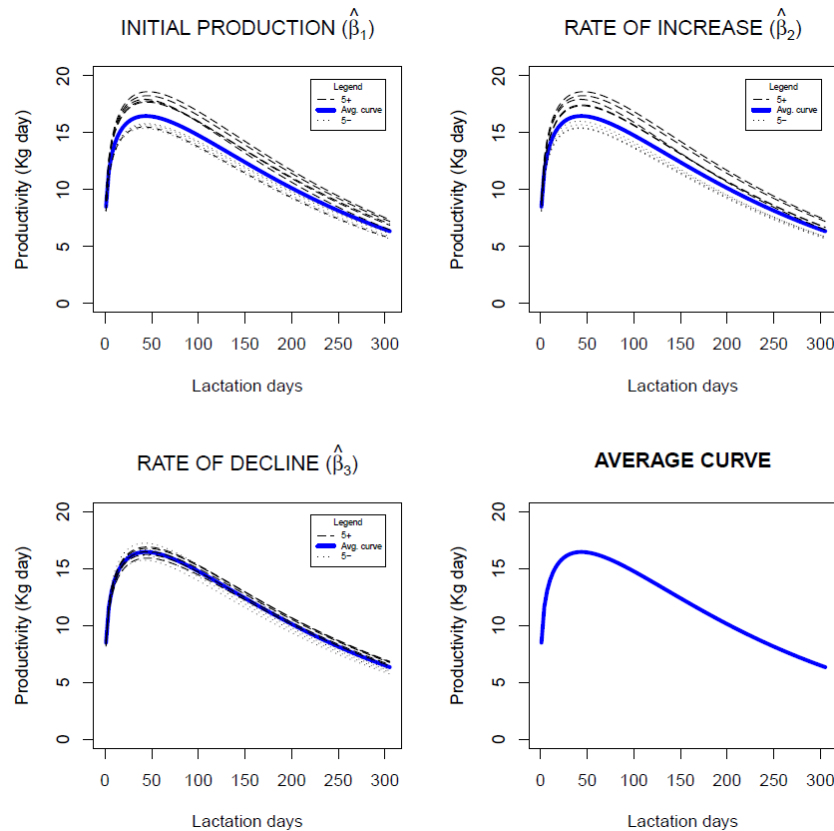

Figure 1. Genomic prediction of lactation curves for the 5 best (5+) and worst (5-) individuals based on the genomic breeding values of the animals for $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$, in addition to the average curve of the herd (223 individuals).

The initial production, represented by $\hat{\beta}_1$, showed an average of 8.56 kg, with a coefficient of variation of only 2.58%, and a range of 8.09 to 9.21 kg. Total production was the trait that showed the highest coefficient of variation (11.02%). In addition, there was an amplitude of 3,469 kg/lactation between the most and least productive individuals. Persistence showed similar results among animals, with a mean and standard deviation of 6.91 and 0.12, respectively, resulting in a coefficient of variation of 1.78%. Based on the estimates, the lactation peak occurs in approximately 67 days, in which the average production is 18.77 kg as the maximum point of the lactation curve. Both measures had low standard deviations and coefficients of variation (Table 4). The rate of decline ($\hat{\beta}_3$) also showed a low standard deviation (0.0001) and a low coefficient of variation (1.61%).

The lactation curves of the 5 best (5+) and worst (5-) individuals based on the genomic breeding values of the animals for the total production, lactation peak, persistence and time to peak yield are shown in Figure 2.
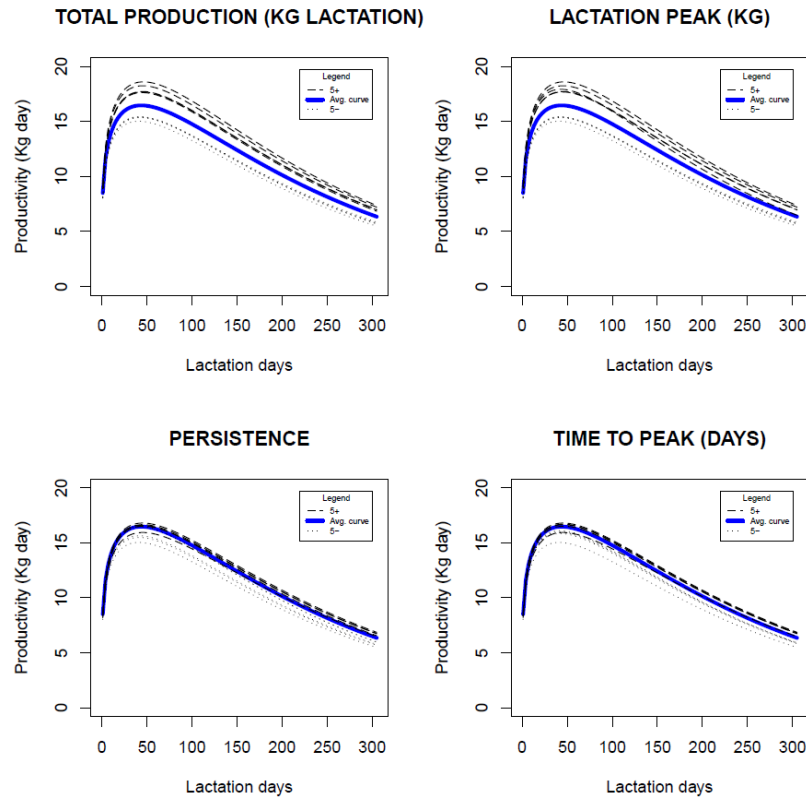
TOTAL PRODUCTION (KG LACTATION)

LACTATION PEAK (KG)

PERSISTENCE

TIME TO PEAK (DAYS)

Figure 2. Lactation curves for the 5 best (5+) and worst (5-) individuals based on genomic breeding values of the animals for total production (kg/lactation), lactation peak (kg day), persistence and time to peak yield (days) of the herd (223 individuals).

Heritabilities and correlations among GEBVs of the animals for the studied traits were estimated from the genotypic and phenotypic variances (Table 5). Heritabilities ranged from 0.09 for the rate of increase to 0.29 for persistence, which was the trait with the highest heritability. The genomic heritabilities of the lactation curve parameters obtained by the Wood's model showed similar results, being 0.10 for $\hat{\beta}_1$ and $\hat{\beta}_2$ and 0.09 for $\hat{\beta}_3$. Lactation peak showed low heritability compared to most of the other variables (0.12).

Table 5. Heritabilities (main diagonal) and correlations among genomic estimated breeding values (upper triangular) of the lactation traits.

Variables	$\hat{\beta}_1$[1]	$\hat{\beta}_2$[2]	$\hat{\beta}_3$[3]	TP[4]	Peak[5]	Pers.[6]	tp[7]
$\hat{\beta}_1$	**0.10**	0.70	0.04	0.84	0.91	0.25	0.17
$\hat{\beta}_2$		**0.10**	-0.20	0.87	0.88	0.54	0.40
$\hat{\beta}_3$			**0.09**	-0.37	-0.19	-0.90	-0.85
TP				**0.27**	0.98	0.65	0.54
Peak					**0.12**	0.51	0.41
Pers.						**0.29**	0.95
tp							**0.22**

1: initial production (kg day); 2: rate of increase; 3: rate of decline; 4: total production (kg/lactation); 5: lactation peak (kg day); 6: persistence; 7: time to peak yield (days).

The GEBVs of the animals for lactation peak showed a high positive correlation with the GEBVs for $\hat{\beta}_1$ (0.91) and $\hat{\beta}_2$ (0.88), and the correlation between the GEBVs for these two parameters was also high and positive (0.70). The correlation between the GEBVs for persistence and time to peak yield was also high (0.95). The GEBVs for rate of decline ($\hat{\beta}_3$) showed a high negative correlation with persistence and time to peak yield (Table 5).

The Kappa coefficients of agreement between the individuals selected according to each variable, in addition to the correlations between GEBVs for the evaluated traits are shown in Figure 3.
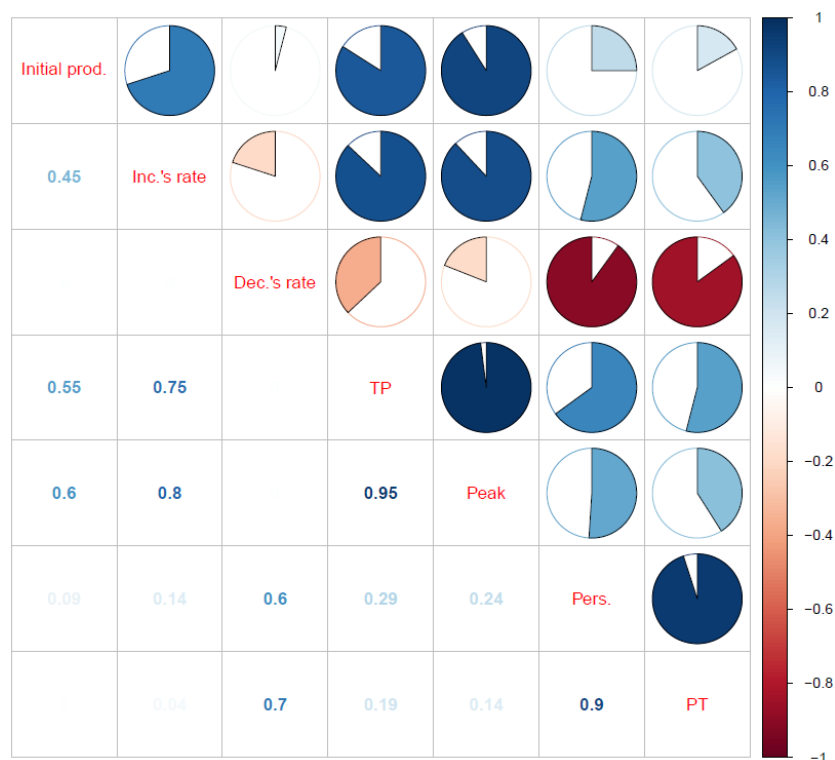

Figure 3. Correlations between genomic estimated breeding values for the lactation traits (upper triangular) and Kappa coefficients of agreement of the 10% best animals (lower triangular). The diagonal represents the seven lactation traits: initial production (Initial prod.), rates of increase and decrease (Inc's rate and Dec's rate, respectively), total production (TP), peak of lactation (Peak), persistence (Pers.) and peak time (PT).

DISCUSSION

Genomic prediction of lactation curves represents the genetic behavior of the milk production of the animals in the herd throughout lactation. According to Bangar and Verma (2017), the lactation curve cannot be fitted through a linear model, since its trend is not linear with time. Therefore, several mathematical models have been applied to explain the flow of milk production throughout lactation in dairy cows (Macciotta et al., 2011; Bangar and Verma, 2017; Piccardi et al., 2017).

The parameters of the lactation curve models are usually estimated as fixed effects, i.e., without considering the environmental factors that may also affect individual productivity. One way to overcome this issue is to use a two-step procedure. In the first step, regression models are fitted to remove environmental effects. In the second step, the milk production measures corrected for environmental effects are used to fit nonlinear models considering each animal individually.

An alternative to the two-step procedure is to use the NLME models, which allow the insertion of fixed and random effects in the model corresponding to effects of different traits and individuals, respectively, in a single fit. In the current study, the NLME model of Wood (1967) was the most efficient to fit the Girolando data based on the AIC and BIC criteria (10,013.79 and 10,101.92, respectively), which is in agreement with the result of Bangar and Verma (2017), who also aimed to compare models based on the same goodness-of-fit measures using information from Gir cattle. Ferreira et al. (2015) and Piccardi et al. (2017) also considered the Wood's equation as the most efficient to fit, respectively, data from Holstein cattle reared in southwest Paraná and to describe the lactation of dairy cows in the Santa Fé and Córdoba region, Argentina. Wood's model showed the lowest standard deviation in the comparison of models for fitting the lactation of Jersey cattle in the study of Cankaya et al. (2011), who also considered this model as the best for describing the lactation curve of this breed. In addition, in the present study, Wood's equation showed an absence of residual autocorrelation, which was also verified by Ghavi Hossein-Zadeh (2015) for buffalo data, and by Piccardi et al. (2017), who fitted this equation for Holstein cattle. The better suitability of the Wood's equation for different breeds and species justifies the wide use of this model, which was highlighted by Ghavi Hossein-Zadeh (2015) and Macciotta et al. (2011). Wood's equation using the nonlinear mixed models approach allowed the genetic identification of superior individuals according to their respective lactation traits without the effect of production levels.

In Figures 1 and 2, which show the genomic curves of the best and worst animals in the herd, the genomic behavior of the individuals' milk production can be identified. Considering the total production, for example, it can be observed that the animals with greater production reached the lactation peak shortly after (approximately 50 days) the animals with lower production (approximately 45 days), and the rate of decline was similar between the best and worst individuals. This variable was the one that presented the biggest graphic differences and the highest coefficient of variation (11.02% - Table 4), which may explain the great differences between animals.

The heritability estimates for the parameters of Wood's model (0.10 for $\hat{\beta}_1$ and $\hat{\beta}_2$ and 0.09 for $\hat{\beta}_3$) were lower than those reported by Yilmaz *et al.* (2011) and Rekaya et al. (2000), which estimates varied from 0.14 to 0.26 for $\hat{\beta}_1$, 0.18 to 0.32 for $\hat{\beta}_2$ and 0.15 to 0.19 for $\hat{\beta}_3$. In both studies, the Wood's equation was used to estimate the parameters using information from Brown Swiss cattle and crosses of the Holstein and Friesian breeds, respectively. On the other hand, the heritabilities from the current study were higher than those described by Saghanezhad et al. (2017), who reported heritabilities of 0.017, 0.022 and 0.06 for $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$, respectively, using data from Holstein cows and the Wood's model. The heritabilities from the current study were similar to the results of Shanks et al. (1981), who reported estimates of 0.10 for $\ln(\hat{\beta}_1)$, using a standardization for the initial

production, 0.06 for $\hat{\beta}_2$ and 0.14 for $\hat{\beta}_3$ in the first lactation, considering data from the California Dairy Herd Improvement Association.

Persistence and total production (kg/lactation) were the traits that showed the highest heritabilities. Canaza-Cayo et al. (2015) estimated heritabilities for nine different persistence estimates proposed by different authors, and found results that varied from 0.18 to 0.33, corroborating with the heritability estimated for persistence in the present study (0.29). These authors also analyzed Girolando cattle, which justifies the similar results. On the other hand, the heritability estimate of 0.29 for persistence in our study was higher than those found by Yilmaz *et al.* (2011), Muir et al. (2004) and Saghanezhad et al. (2017), who used information from Brown Swiss cattle and reported heritabilities ranging from 0.05 to 0.23.

The heritability for total production (0.27) was higher than that found by Yilmaz et al. (2011) and Pereira et al. (2012), who analyzed data from Swiss and Gir cattle and reported results of 0.18 and 0.21, respectively. On the other hand, it was similar to the result of Canaza-Cayo et al. (2015), who used the random regression models approach, and it was lower than the estimates found by Jakobsen et al. (2002) and Cobuci et al. (2006), who reported heritabilities ranging from 0.31 to 0.42 for Holstein cattle from different locations.

The lactation peak showed low heritability (0.12) when compared to the studies of Shanks et al. (1981), Saghanezhad et al. (2017), Yilmaz et al. (2011), and Muir et al. (2004), who reported heritabilities ranging from 0.16 to 0.42 for this trait. Regarding the time to peak yield, a higher heritability (0.22) was estimated compared to the abovementioned studies, in which heritabilities ranging from 0.013 to 0.13 for the same trait were reported.

Tekerli et al. (2000) evaluated the relationships between the lactation traits through their phenotypic correlations. Genetic correlations between lactation traits via mixed models were estimated by Boujenane and Hilal (2012), El-Awady (2013), and Farhangfar and Rowlinson (2007). In our study, the simple correlations between the GEBVs of the animals for each trait were used.

When a high correlation between the GEBVs for two traits was observed, the Kappa coefficient for selecting the best individuals according to these traits was also high (Figure 3). This was expected, since a high correlation between two traits indicates that the behavior of one trait in relation to the other is similar, also resulting in the similarity between the best individuals for the same traits. The high correlations between the GEBVs for initial production ($\hat{\beta}_1$) and the rate of increase ($\hat{\beta}_2$) with the lactation peak (0.91 and 0.88, respectively) indicate that the higher the initial production and rate of increase, the later will be the lactation peak. The GEBVs for lactation peak were also strongly correlated with the GEBVs for total production (correlation of 0.98), which shows that lactation peak is the trait most associated with the total productivity of the animals under study, indicating that genetically, individuals who have a high lactation peak tend to be more productive.

Figure 3 also shows that the Kappa coefficient of agreement between the individuals selected according to the lactation peak and the total production was 0.95, which is the highest value among the coefficients of agreement evaluated. This result corroborates with the high genetic correlation between the same variables found by Boujenane and Hilal (2012), El-Awady (2013), Farhangfar and Rowlinson (2007), Muir et al. (2004), Saghanezhad et al. (2017) and Rekaya et al. (2000), who reported genetic correlations ranging from 0.87 to 0.98. Therefore, we have an indication that these traits show a high genetic correlation for different breeds and species.

The GEBVs for initial production ($\hat{\beta}_1$) and the rate of increase ($\hat{\beta}_2$) were also highly correlated with the GEBVS of the animals for total production (0.84 and 0.87, respectively), in addition to high Kappa coefficients (0.55 and 0.75, respectively). Therefore, individuals who have high rate of increase and initial production tend to have a high total production. Similarly, the selection of the best individuals according to these two variables and the total production also showed high concordances, which may confirm a genetic relationship between them. The same can be concluded for the correlation between the GEBVS for the variables $\hat{\beta}_1$ and $\hat{\beta}_2$ with the GEBVs for lactation peak. The correlation between the GEBVs for initial and total production (0.84) estimated in the current study was higher than those found by Boujenane and Hilal (2012) and Rekaya et al. (2000), who reported correlations of 0.38 and 0.23, respectively, for Holstein-Friesian cattle.

The GEBVs for persistence of lactation showed moderate correlation with the GEBVs for total production and presented correlation close to 1 with the GEBVs for time to peak yield, showing that animals that take longer time to reach the maximum production are those that present lower persistence of lactation. Using this measure to estimate lactation persistence, Boujenane and Hilal (2012) found a genetic correlation of 0.77 and Saghanezhad et al. (2017) also found a genomic correlation close to 1 (0.99) between the persistence of lactation and the time to peak yield.

CONCLUSIONS

Among the nine equations compared using the nonlinear mixed models' approach, the Wood's model showed the best performance, since it was the one that presented the lowest AIC and BIC values and did not present residual autocorrelation. Using the estimates of Wood's equation parameters as response variables, the application of Genome Wide Selection through BLASSO method allowed the construction of the lactation curves of Girolando cattle, the breed with the greatest milk production in Brazil. In addition, it allowed the visualization of genetic differences among the animals in the herd. Such differences showed the variation between genetic measures with no influence of external factors, ensured by the fit of nonlinear mixed models. Moreover, this methodology allowed accessing the heritabilities and correlations among genomic breeding values of the individuals for the selected traits. The identification of the best individuals was carried out based on the genomic breeding values for the different traits; analyzing the Kappa coefficient of agreement for the individuals selected, a high coefficient was found in the comparisons that presented high correlation between genomic breeding values. Finally, the application of genomic selection for variables without the influence of environmental effects allows a more precise estimation of the correlation between the GEBVs for the lactation traits, since environmental effects can considerably influence the estimates and selection of animals.

ACKNOWLEDGMENTS

AUTHORS' CONTRIBUTIONS

Conceptualization: Teixeira, F. R. F.; Nascimento, M. Data Acquisition: Silva, M. V. G. B. Data analysis: Teixeira, F. R. F.; Nascimento, M. Design of Methodology: Teixeira, F. R. F.; Nascimento, M.; Software development: Teixeira, F. R. F.; Nascimento, M.; Azevedo, C. F. Writing and editing: Teixeira, F. R. F.; Nascimento, M.; Cecon, P. R.; Cruz, C. D.; SILVA, F. F.; Nascimento, A. C. C.; Azevedo, C. F.; Paixão, D. M.; Marques, D. B. D.; Carneiro, A. P. S.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Akaike H (1974). A new look at the statistical model identification. *IEEE T. Automat. Contr.* 19: 716-723.

Bangar YC and Verma MR (2017). Non-linear modelling to describe lactation curve in Gir crossbred cows. *J. Anim. Sci. Technol.* 59(3): 1-7.

Boujenane I and Hilal B (2012). Genetic and non-genetic effects for lactation curve traits in Holstein-Friesian cows. *Arch. Tierz.* 5: 450-457.

Brody S, Ragsdale AC and Turner CW (1923). The rate of decline of milk secretion with the advance of the period of lactation. *J. Gen. Physiol.* 5: 441-444.

Canaza-Cayo AW, Lopes PS, Silva MVGV, Torres RA, et al. (2015). Genetic Parameters for Milk Yield Lactation Persistence Using Random Regression Models in Girolando Cattle. *Asian Australas. J. Anim. Sci.* 28: 1407-1418.

Cankaya S, Unalan A and Soydan E (2011). Selection of a mathematical model to describe the lactation curves of Jersey cattle. *Archiv Tierzucht.* 54: 27-35.

Cappio-Borlino A, Pulina G and Rossi G (1995). A non-linear modification of Wood's equation fitted to lactation curves of Sardinian dairy ewes. *Small Ruminant Res.* 18: 75-79.

Cardona SJC, Álvarez JDC, Sarmento JLR, Herrera LGG, et al. (2015). Association of SNPs in the genes for κ -casein and β-lactoglobulin with lactation curves in dairy goats. *Pesqui. Agropecu.* 50: 224-232.

Centoducati P, Maggiolino A, De Palo P and Tateo A (2012). Application of Wood's model to lactation curve of Italian Heavy Draft horse mares. *J. Dairy Sci.* 95: 5770-5775.

Cobby JM and Le Du YLP (1978). On fitting curves to lactation data. *Anim. Prod.* 26: 127-133.

Cobuci JA, Costa CN, Teixeira NM and Freitas AF (2006). Use of Lagendre polynomials and Wilmink functions in genetic evaluations for persistence of lactation in Holstein cows. *Arq. Bras. Med. Vet. Zoo.* 58: 614-623.

Cohen J (1960). A coeficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20: 37-46.

Daltro DS, Padilha AH, Silva MCGB, Kern EL, et al. (2019). Heterosis in the lactation curves of Girolando cows with emphasis on variations of the individual curves. *J. Appl. Anim. Sci.* 47: 85-95.

de Los Campos G and Rodrigues PP (2016). Bayesian Generalized Linear Regression. R package version 1.0.5. Available at [https://CRAN.R-project.org/package=BGLR].

de Los Campos G, Naya H, Gianola D, Crossa J, et al. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics.* 182: 375-385.

Dhanoa MS (1981). A note on an alternative form of the lactation curve model of Wood. *Anim. Prod.* 32: 349-351.

Durbin J and Watson GS (1971). Testing for serial correlation in least squares regression III. *Biometrika.* 58: 1-19.

El-Awady HG (2013). Genetic aspects of lactation curve traits and persistence indices in Friesian cows. *Arch. Zootech.* 16: 15-29.

Farhangfar H and Rowlinson P (2007). Genetic Analysis of Wood's Lactation Curve for Iranian Holstein Heifers. *J. Biol. Sci.* 7: 127-135.

Ferreira AGT, Henrique DS, Vieira RAM, Maeda EM, et al. (2015). Fitting mathematical models to lactation curves from Holstein cows in the southwestern region of the state of Parana, Brazil. *An. Acad. Bras. Ciênc.* 87: 503-517.

Gower JC (1971). A general coefficient of similarity and some of its properties. *Biometrics.* 27: 857-54.

Hossein-Zadeh GN (2015). Comparison of non-linear models to describe the lactation curves for milk yield and composition in buffaloes (*Bubalus bubalis*). *Animal.* 10: 248-261.

Jakobsen JH, Madsen P, Jensen J, Pedersen J, et al. (2002). Genetic parameters for milk production and persistence for Danish Holsteins estimated in random regression models using REML. *J. Dairy Sci.* 85: 1607-1616.

Lindstrom MJ and Bates DM (1990). Nonlinear Mixed Effects Models for Repeated Measures Data. *Biometrics.* 46: 673-687.

Macciotta NPP, Dimauro C, Rassu SPG, Steri, R, et al. (2011). The mathematical description of lactation curves in dairy cattle. *Ital. J. Anim. Sci.* 10: 213-223.

Macciotta NPP, Gaspa G, Bomba L, Vicario D, et al. (2015). Genome-wide association analysis in Italian Simmental cows for lactation curve traits using low-density (7K) SNP panel. *J. Dairy Sci.* 98: 8175-8185.

Maechler M, Rousseeuw P, Struyf A, Hubert M, et al. (2018). Cluster: Cluster Analysis Basics and Extensions. R package version 2.1.0. Available at [https://cran.r-project.org/web/packages/cluster/index.html].

Meuwissen THE, Hayes BJ and Goddard ME (2001). Prediction of total genetic value using genome wide dense marker maps. *Genetics.* 157: 1819-1829.

Muir BL, Fatehi J and Schaeffer LR (2004). Genetic Relationships Between Persistence and Reproductive Performance in First-Lactation Canadian Holsteins. *J. Dairy Sci.* 87: 3029-3037.

Nelder JA (1966). Inverse polynomials, a useful group of multi-factor response functions. *Biometrics.* 22: 128-141.

Papajcsik I and Bodero J (1988). Modeling lactation curves of friesian cows in subtropical climate. *Anim. Prod.* 47: 201-207.

Pereira RJ, Verneque RS, Lopes PS, Santana JR, et al. (2012). Milk yield persistence in Brazilian Gyr cattle based on a random regression model. *Genet. Mol. Res.* 11: 1599-1609.

Piccardi M, Macchiavelli E, Funes AC, BÓ GA, et al. (2017). Fitting milk production curves through nonlinear mixed models. *J. Dairy Res.* 84: 146-153.

Pinheiro JC and Bates DM (2000). Mixed-Effect Models in S and S-PLUS. Springer, New York, FL, USA.

Pinheiro J, Bates D, Debroy S, Sarkar D, et al. (2017). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-131. Available at [https://CRAN.R-project.org/package=nlme].

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at [https://www.R-project.org/].

Rekaya R, Carabaño MJ and Toro MA (2000). Bayesian Analysis of Lactation Curves in Holstein-Friesian Cattle. *J. Dairy Sci.* 83: 2691-2701.

Rook AJ, France J and Dhanoa MS (1993). On mathematical description of lactation curves. *J. Agric. Sci.* 121: 97-102.

Saghanezhad F, Atashi H, Dadpasand M, Zamiri MS, et al. (2017). Estimation of Genetic Parameters for Lactation Curve Traits in Holstein Dairy Cows in Iran. *Iran. J. Appl. Anim. Sci.* 7: 559-566.

Schwarz G (1978). Estimating the dimension of a model. *Ann. Stat.* 6: 461-464.

Shanks RD, Berger PJ, Freeman AE and Dickinson FN (1981). Genetic Aspects of Lactation Curves. *J. Dairy Sci.* 64: 1852-1860.

Sikka LC (1950). A study of lactation as affected by heredity and environment. *J. Dairy Res.* 17: 231-252.

Silva GMGB, Canaza-Cayo AW, Lopes PS, Cobuci JA, et al. (2015). Programa de Melhoramento Genético da Raça Girolando: do teste de progênie às avaliações genômicas. *Inf. Agropec.* 36: 35-40.

Tekerli M, Akinci Z, Dogan I1950 Akcan A (2000). Factor Affecting the Shape of Lactation Curves of Holstein Cows from the Balikesir Province of Turkey. *J. Dairy Sci.* 83: 1381-1386.

Wickham H, Francois R, Henry L, Muller K (2017). dplyr: A Grammar of Data Manipulation. R package version 0.7.3. Available at [https://CRAN.R-project.org/package=dplyr].

Wood PDP (1967). Algebraic model of the lactation curve in cattle. *Nature.* 216: 164-165.

Yilmaz I, Eyduran E, Kaygisiz A and Javed K (2011). Estimates of Genetic Parameters for Lactation Shape Parameters with Multivariate Statistical Technique in Brown Swiss Cattle. *Int J. Agr. Biol.* 13: 174-178.