Acta Scientiarum
http://www.periodicos.uem.br/ojs/
ISSN on-line: 1807-8621
Doi: 10.4025/actasciagron.v46i1.65081

Quantile regression for genomic selection of growth curves

Ana Carolina Campana Nascimento, Camila Ferreira Azevedo, Cynthia Aparecida Valiati Barreto, Gabriela França Oliveira and Moysés Nascimento*

Departamento de Estatística, Universidade Federal de Viçosa, Av. Peter Henry Rolfs, s/n, 36570-900, Campus Universitário, Viçosa, Minas Gerais, Brazil. *Author for correspondence. E-mail: moysesnascim@ufv.br

ABSTRACT. This study evaluated the efficiency of genome-wide selection (GWS) based on regularized quantile regression (RQR) to obtain genomic growth curves based on genomic estimated breeding values (GEBV) of individuals with different probability distributions. The data were simulated and composed of 2,025 individuals from two generations and 435 markers randomly distributed across five chromosomes. The simulated phenotypes presented symmetrical, skewed, positive, and negative distributions. Data were analyzed using RQR considering nine quantiles (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9) and traditional methods of genomic selection (specifically, RR-BLUP, BLASSO, BayesA, and BayesB). In general, RQR-based estimation of the GEBV was efficient—at least for a quantile model, the results obtained were more accurate than those obtained by the other evaluated methodologies. Specifically, in the symmetrical-distribution scenario, the highest accuracy values were obtained for the parameters with the models RQR0.4, RQR0.3, and RQR0.4. For positive skewness, the models RQR0.2, RQR0.3, and RQR0.1 presented higher accuracy values, whereas for negative skewness, the best model was RQR0.9. Finally, the GEBV vectors obtained by RQR facilitated the construction of genomic growth curves at different levels of interest (quantiles), illustrating the weight–age relationship.

Keywords: conditional quantiles; genomic prediction; GWS; genetic breeding.

Received on September 20, 2022.
Accepted on December 9, 2022.

Introduction

Growth curves help summarize the weight-age (or yield-time) relationship in animals and plants based on a few interpretable parameters, such as mature weight (asymptotic weight) and maturity rate (growth rate). Growth-curve models aim to explain the growth process over time. This information regarding the demand and care needed for each plant or animal development stage facilitates effective economic decision-making.

Pong-Wong and Hadjipavlou (2010) used a two-step method to analyze growth trajectories using a Genomic Selection (GS) approach in the current post-genomic era. In this framework, growth models fit weight–age (or yield–time) data in the first step, whereas growth curve parameter estimates are used as the dependent variables in GS methods in the second step. Ibáñez-Escriche and Blasco (2011) showed that GS modified the growth curve with markers acting simultaneously on the three parameters of the Gompertz curve. Using a genome-wide association study (GWAS) framework, Howard et al. (2015) performed a study based on polynomial coefficients to identify genomic regions affecting growth and feed intake curves in Duroc boars. Silva et al. (2017) identified GWAS-based candidate genes, whose biological functions can be useful in explaining the genetic basis of postnatal growth in pigs.

These genomic methods focus on estimating single nucleotide polymorphism (SNP) marker effects in the mean body weight (BW) over time; the function is defined for the expected value of BW conditional on X (age), or simply E(BW|X). Mosteller and Tukey (1977) showed that regression curves can be estimated for different quantiles of the response variable distribution, providing a complete picture of the regression (Cade & Noon, 2003). This method, called Quantile Regression (QR) (Koenker & Bassett Jr., 1978), can estimate parameters for all portions of the probability distribution of the response variable (e.g., BW). This enabled us to better understand the relationship between response (BW) and explanatory variables (age). In addition, QR does not require assumptions regarding the error distribution and is robust to outliers (Oliveira et al., 2021b).

To deal with high-dimensional problems, such as a large number of parameters and a small number of observations, methods combining shrinkage estimation and variable selection (BLASSO, BayesA, BayesB, etc.) have been proposed for GS. Under a QR framework, this method is called regularized QR (RQR) (Li & Zhu, 2008). Nascimento et al. (2017) proposed the use of RQR in GS studies was proposed by Nascimento et al. (2017) using simulated data. This approach has been used in plant and animal breeding studies and has shown to be a promising technique. Nascimento et al. (2019a) estimated the Genomic Estimated Breeding Values (GEBV) for different levels (quantiles) of the probability distribution associated with flowering time in common bean. Oliveira et al. (2021b) evaluated the efficiency of RQR throughout the breeding process using simulated data from autogamous plants. Barroso et al. (2017) used the same methodology to estimate the effects of SNP markers on growth curves in pigs.

However, few studies have evaluated phenotypes that assume non-normal distributions, such as skewed distributions. However, breeding programs may lead to phenotypes with skewed distributions of flowering time (Maurer et al., 2015; Nascimento et al., 2019a), order of parity (Varona, Ibañez-Escriche, Quintanilla, Niguera, & Casellas, 2008), and hormone concentrations (de Campos et al., 2015; Mathur et al., 2012). Thus, treating phenotypes inappropriately by assuming symmetrical distributions can substantially affect the accuracy of genomic prediction (Maurer et al., 2015).

In this study, we propose an RQR methodology for the genomic prediction of growth curves using simulated data with distributions with different levels of asymmetries and compare its prediction accuracy values with those of traditional GS methods (namely, RR-BLUP, BLASSO, BayesA, and BayesB).

Material and methods

Simulated population

Publicly available simulated data from the QTLMAS2009 (Coster, Bastiaansen, Calus, van Arendonk, & Bovenhuis, 2010) were used in this study. The dataset consisted of 2,025 individuals from two generations with complete information on 453 SNP markers that were randomly distributed over five chromosomes. Individuals in this dataset consisted of 25 parents (20 females and 5 males) and 2,000 progenies from 100 full-sib families (a combination of male and female parents). Each family included 20 full-sibs.

Of the 100 families, 50 (1,000 individuals) had phenotypic records (BW) generated at five time points (t= 0, 132, 265, 397, and 530 days), according to the logistic growth curve:

$$y(t) = \frac{\alpha_1}{1+\exp\left[\frac{(\alpha_2 - t_j)}{\alpha_3}\right]} \quad (1)$$

where α_1 is the asymptotic weight; α_2 is the inflection point of the curve; and α_3 is the slope of the curve.

Genomic breeding values and skewness

True genomic breeding values (TGBVs) for each curve parameter were simulated for each individual as the sum of additive effects based on six quantitative trait locus). One large QTL and five small QTLs were assigned to each of the three curve parameters, assuming the same heritability ($h^2 = 0.50$) (Coster et al., 2010). The original dataset was modified by inserting positive and negative skewness to evaluate the impact of skewness (non-normal phenotypic distributions) on the genomic prediction accuracy of growth curves. We also evaluated normal phenotypic distribution (the original simulated dataset). Random errors were added to the TGBVs based on three distributions: (1) normal error as $N\{0, \sigma_e^2 = sd[y_i(t)]\}$, (2) positive skewed error as $\exp\left(\lambda = {}^{1}/_{\sigma_e^2}\right)$, and (3) negative skewed error as $-\exp\left(\lambda = {}^{1}/_{\sigma_e^2}\right)$. Thus, the distribution of phenotypic values (growth curve parameters) was simulated as symmetric, positively skewed, and negatively skewed. The modified datasets are freely accessible at https://github.com/licaeufv.

Estimation of growth curve parameters

A logistic nonlinear regression model (model 1) was used to estimate the growth curve parameters based on individual weight–age data. The estimated growth curve parameters (α_1: asymptotic weight, α_2: inflection point of the curve, and α_3: slope of the curve) were used as phenotypes in subsequent analyses.

Genomic prediction analyses

Genomic prediction (GP) analyses were performed using the data from all individuals with phenotypic and genotypic records. Different methods were used in this study.

Traditional GS methods

The methods used in this study include BLASSO (De Los Campos et al., 2009), RR-BLUP, BayesA, and BayesB (Meuwissen, Hayes, & Goddard, 2001). These methods assume normal distributions for the phenotype values. Bayesian methods were implemented using 100,000 MCMC (Markov chain Monte Carlo) iterations, with a burn-in and thin layer of 10,000 and five iterations, respectively.

Regularized quantile regression. Data were analyzed using regularized quantile regression (RQR) (Li & Zhu, 2008) based on nine quantiles (τ): 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9. This method consists of obtaining the vector of marker effect estimates (β_τ) that solves the following optimization problem:

$$\beta_\tau = \text{argmin}_{\boldsymbol{\beta}_\tau}\{\sum_{i=1}^{n}\rho_\tau[\widehat{\alpha}_{Si} - (\mu + \sum_{k=1}^{m} x_{ik}\beta_{\tau k})] + \lambda\sum_{k=1}^{m}|\beta_{\tau k}|\},$$

where $\tau \in]0,1[$ indicates the quantile, μ is the mean, x_{ik} is the incidence of the ith individual and the kth marker, $\beta_{\tau k}$ is the effect of the ktb marker, $\sum_{k=1}^{m}|\beta_{\tau k}|$ is the sum of the absolute values of the regression coefficients, and λ is the regularization parameter. The parameter $\rho_\tau(.)$ is denoted as a check function (Koenker & Bassett Jr., 1978), and is defined by

$$\rho_\tau\left[\widehat{\alpha}_{Si} - \left(\mu + \sum_{k=1}^{m} x_{ik}\beta_{\tau k}\right)\right] = \begin{cases} \tau \cdot \left[\widehat{\alpha}_{Si} - \left(\mu + \sum_{k=1}^{m} x_{ik}\beta_{\tau k}\right)\right], \text{if } \widehat{\alpha}_{Si} - \left(\mu + \sum_{k=1}^{m} x_{ik}\beta_{\tau k}\right) > 0, \\ -(1-\tau) \cdot \left[\widehat{\alpha}_{Si} - \left(\mu + \sum_{k=1}^{m} x_{ik}\beta_{\tau k}\right)\right], \quad \text{otherwise.} \end{cases}$$

Thus, the values of $\beta_{\tau k}$ represent marker effects in the τ^{th} quantile of interest.

A grid of shrinkage parameter values (λ) ranging from 0 to the posterior λ estimate from BLASSO is used. However, we reported results that yielded the highest prediction accuracy.

The genomic estimated breeding values (GEBVs) were obtained for each quantile as $GEBV_{(\tau)} = \sum_{k=1}^{m} x_{ik}\hat{\beta}_{\tau k}$, where τ represents the τ^{th} quantile of interest. Subsequently, the mean genomic growth curve for the ith animal and the τ^{th} quantile was estimated as

$$\hat{y}_{(\tau)ij} = \frac{\hat{\mu}_{(\tau)\hat{\alpha}_1} + \hat{u}_{(\tau)\hat{\alpha}_{1i}}}{\left\{1+\exp\left[\left(\hat{\mu}_{(\tau)\hat{\alpha}_2} + \hat{u}_{(\tau)\hat{\alpha}_{2i}}\right) - \left(\hat{\mu}_{(\tau)\hat{\alpha}_3} + \hat{u}_{(\tau)\hat{\alpha}_{3i}}\right)t_{ij}\right]\right\}} \quad (3)$$

where $\hat{y}_{(\tau)ij}$ is the predicted BW for animal i at age j (t_{ij}) and quantile τ; $\hat{\mu}_{(\tau)\hat{\alpha}_1}$, $\hat{\mu}_{(\tau)\hat{\alpha}_2}$ and $\hat{\mu}_{(\tau)\hat{\alpha}_3}$ are the adjusted trait means (parameter estimates for the logistic model) for α_1, α_2 and α_3, respectively. Mean genomic growth curves were also obtained using other methods evaluated in this study (BLASSO, RR-BLUP, BayesA, and BayesB).

In addition, for each genomic selection method and scenario (symmetric, positive, and negative skewness), the slope of the regression of GEBVs on TGBVs was calculated as a measure of prediction bias.

Comparison of methodologies under a GS approach

Genomic prediction analyses were performed using a two-fold cross-validation approach, with each fold defined using Ward's hierarchical clustering method (Ward Jr., 1963) based on genotypes. The accuracy of genomic prediction for the 13 different genomic selection models (BLASSO, RR-BLUP, BayesA, BayesB, $RQR_{0.1}$, $RQR_{0.2}$, ..., $RQR_{0.9}$) was calculated as the average (from each fold) Pearson correlation coefficient between the pre-adjusted phenotypes and GEBVs divided by the square root of heritability. Subsequently, bar graphs were generated using accuracy values and their respective standard errors.

In addition, Cohen's kappa coefficient (Cohen, 1960) was used to calculate the percentage of the top 10% of individuals with the highest GEBVs across 13 different genomic selection models (BLASSO, RR-BLUP, BayesA, BayesB, $RQR_{0.1}$, $RQR_{0.2}$, ..., $RQR_{0.9}$). Cohen's Kappa coefficient is given by, $C=NC-C_{Random}/(1-C_{Random})$, where NC is the relative observed agreement among methodologies, and C_{Random} is the hypothetical probability of random agreement.

Computational features

Data analysis was carried out on a computer with a 2.60GHz core i7 processor and 16GB of RAM. Analyses were performed using the functions *nls* (nonlinear logistic model fitting), *rq* (regularized quantile regression), and *BGLR* (Bayesian LASSO, RR-BLUP, BayesA, and B) (de Los Campos & Pérez-Rodríguez, 2014) from the R package (R Core Team, 2021) statistics, quantreg (Koenker, 2015) and BLGR (Pérez, de Los Campos, Crossa, & Gianola, 2010). The plausibility of these values was assessed separately for each MCMC chain using the Raftery–Lewis and Geweke convergence diagnostics in the boa (Smith, 2007) R package. All data and R codes used in this study are freely accessible on the Web: https://github.com/licaeufv.

Results

Figures 1, 2, and 3 show the distributions of phenotypic and genotypic (Breeding Values - BV) values for all parameters α_1 - mature weight, α_2 - inflection point, and α_3 - slope considering symmetric, positive, and negative skewness phenotypic distributions, respectively. As expected, compared with the distributions of the genotypic values, the distributions of the phenotypic values presented larger ranges for all the simulated scenarios (Figures 1, 2, and 3). The distributions of genotypic values were symmetrical for all scenarios and parameters. These distributions were concentrated in the middle, lower, and high quantiles of phenotypic distributions for the symmetric, positive, and negative skewness scenarios, respectively (Figures 1, 2, and 3).

In the symmetric distribution scenario, the highest accuracy values obtained were 0.62, 0.28, and 0.58, respectively, for α_1, α_2, and α_3 parameters considering $RQR_{0.4}$, $RQR_{0.3}$, and $RQR_{0.4}$, respectively (Figure 1). These values were higher than those obtained from the BLASSO (0.57, 0.21, and 0.49), RR-BLUP (0.51, 0.21, and 0.47), BayesA (0.60, 0.21, and 0.55), and BayesB (0.60, 0.21, and 0.54) models (Figure 1). For α_3 the accuracy of the $RQR_{0.5}$ model (0.54) was slightly lower than that obtained for BayesA (0.55), which considered the symmetric distribution and expected value in the estimation.

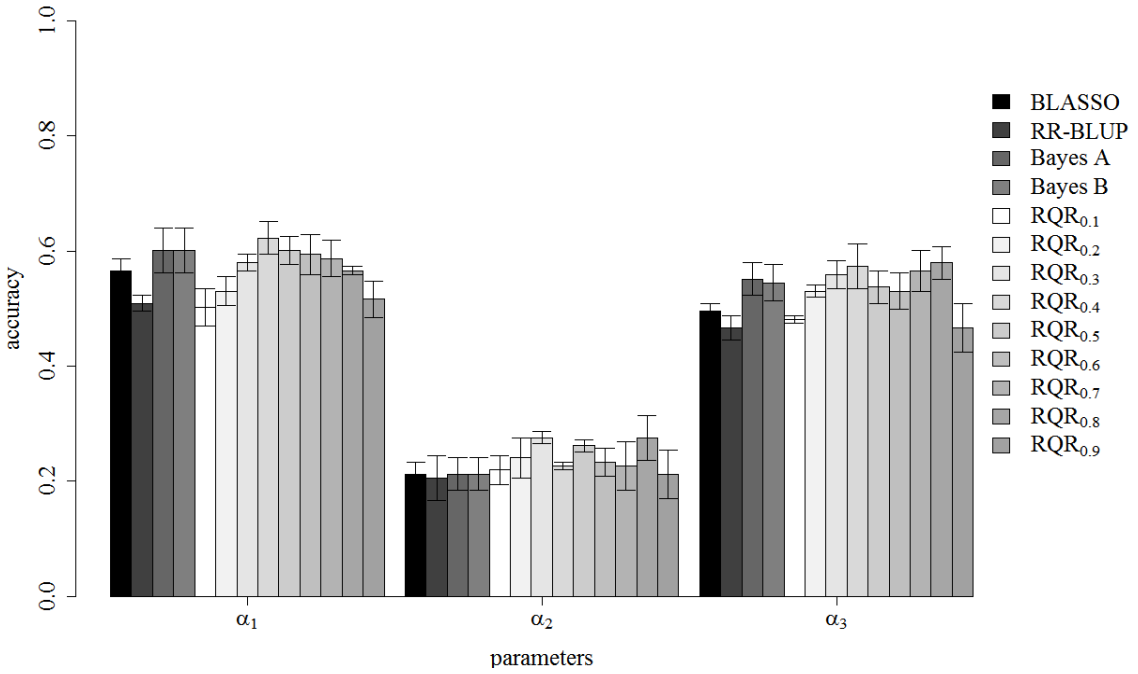

Figure 1. Estimated accuracy and standard error values using thirteen different genomic selection models (BLASSO, RR-BLUP, BayesA, BayesB, $RQR_{0.1}$, $RQR_{0.2}$, ..., $RQR_{0.9}$) considering the symmetrical phenotypic distribution scenario. Y-axes show average estimated accuracy values for each parameter (X-axis).

Considering the positive skewness, the $RQR_{0.2}$, $RQR_{0.3}$, and $RQR_{0.1}$ models presented higher accuracy values (0.70, 0.34, and 0.52) than the BLASSO (0.60, 0.20, and 0.40), RR-BLUP (0.52, 0.18, and 0.37), BayesA (0.66, 0.18, and 0.40), and BayesB (0.66, 0.18, and 0.40) models (Figure 2).

For negative skewness, the $RQR_{0.9}$ model presented higher accuracy values (0.70, 0.57, and 0.57) for all parameters when compared to BLASSO (0.64, 0.37, and 0.37), RR-BLUP (0.55, 0.30, and 0.30), BayesA (0.67, 0.40, and 0.41), and BayesB (0.67, 0.41, and 0.41) models (Figure 3).

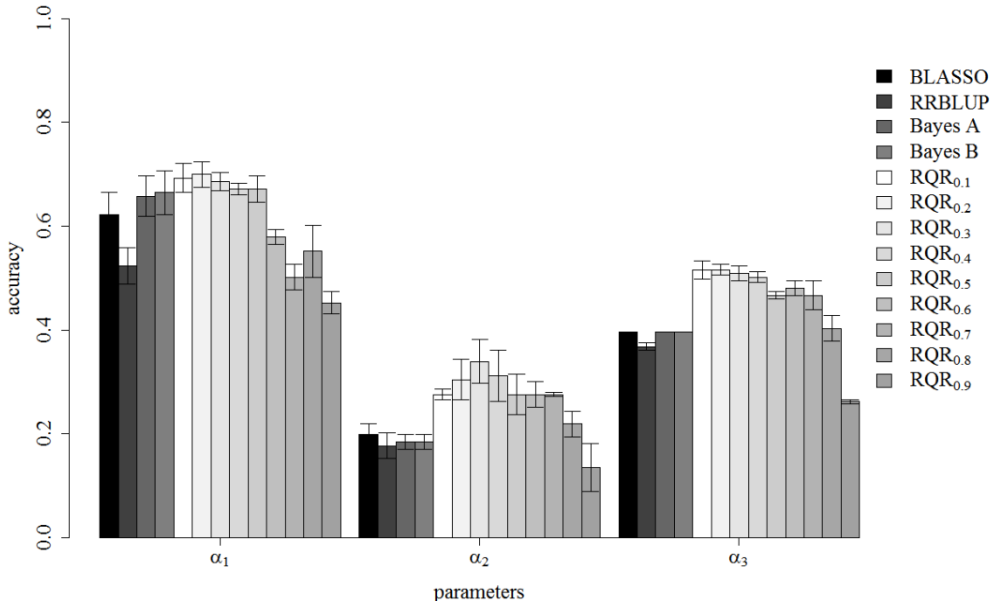

Figure 2. Estimated accuracy and standard error values using thirteen different genomic selection models (BLASSO, RR-BLUP, BayesA, BayesB, $RQR_{0.1}$, $RQR_{0.2}$, ..., $RQR_{0.9}$) considering the positive skewness phenotypic distribution scenario. Y-axes show average estimated accuracy values for each parameter (X-axis).

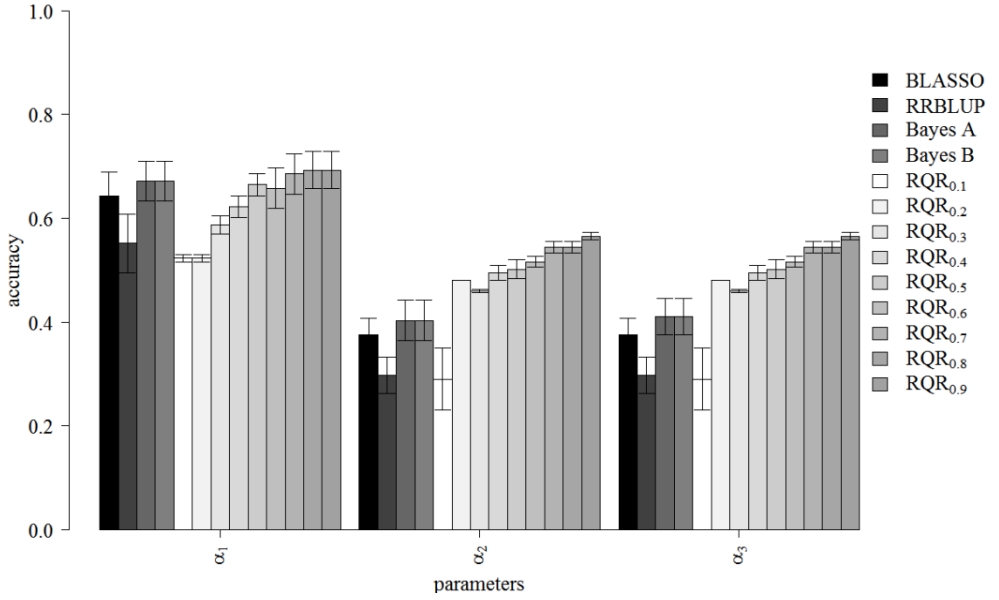

Figure 3. Estimated accuracy and standard error values using thirteen different genomic selection models (BLASSO, RR-BLUP, BayesA, BayesB, $RQR_{0.1}$, $RQR_{0.2}$, ..., $RQR_{0.9}$) considering the negative skewness phenotypic distribution scenario**.** Y-axes show average estimated accuracy values for each parameter (X-axis).

The average genomic growth curves using the GEBVs obtained from six different genomic selection models (BLASSO, RR-BLUP, BayesA, BayesB, $RQR_{0.5}$, and RQR_{best}) and the true values for each parameter of the logistic growth curve are shown in Figures 4, 5, and 6. The RQR_{best} model was obtained by considering quantile regression fits that presented higher prediction accuracy values. Thus, were considered $RQR_{0.4}$, $RQR_{0.3}$, and $RQR_{0.4}$, for the, α_1, α_2, and α_3 parameters in the symmetric scenario; $RQR_{0.2}$, $RQR_{0.3}$, and $RQR_{0.1}$, for the, α_1, α_2, and α_3 parameters in the positive skewness scenario; and $RQR_{0.9}$, for all parameters (α_1, α_2, and α_3) in the negative skewness scenario.

Thirteen different genomic prediction models (BLASSO, RR-BLUP, BayesA, BayesB, $RQR_{0.1}$, $RQR_{0.2}$, ..., $RQR_{0.9}$) were fit considering the cross-validation approach, where the two subsamples were obtained using Ward hierarchical clustering analysis (Figure S1 available at https://det.ufv.br/moyses-nascimento/). The accuracy estimates for all parameters and scenarios ranging from 0.09 to 0.70 are presented in Figures 1, 2 and 3. According to these values, the RQR models showed better results than those obtained from BLASSO, RR-BLUP, and BayesA and BayesB (Figures 1, 2, and 3).

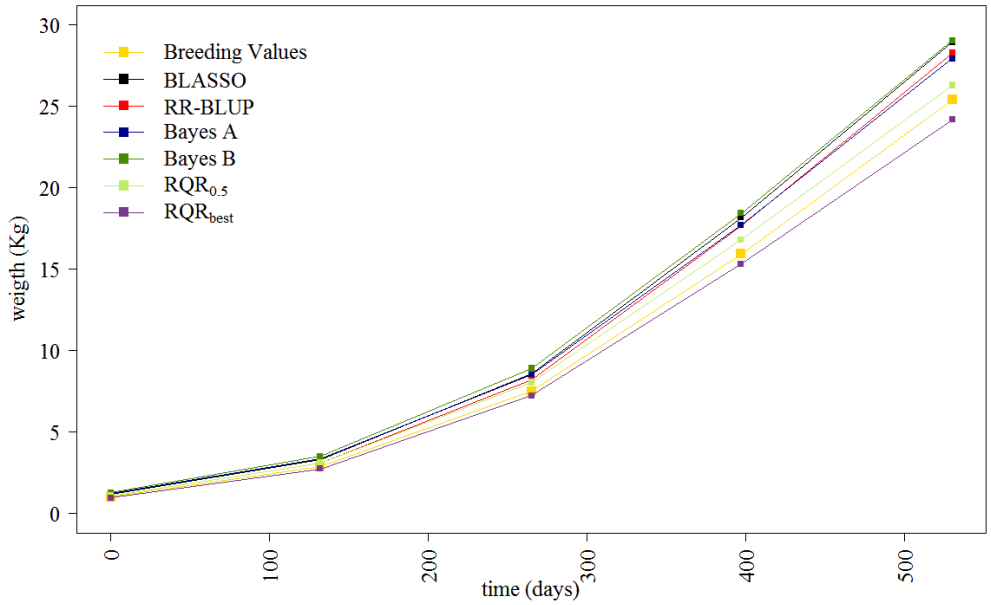

Figure 4. Mean genomic growth curves using the genomic estimated breeding values (GEBVs) obtained by 6 different genomic selection models (BLASSO, RR-BLUP, BayesA, BayesB, $RQR_{0.5}$, and RQR_{best}) and the breeding values for each parameter of a logistic growth curve considering the symmetric scenario.

Although evidence of significant bias was detected (Table 1), considering the mean genomic growth curves (Figures 4, 5, and 6), RQR showed the least bias compared to BLASSO, RR-BLUP, BayesA, and BayesB for all parameters and scenarios. When the distribution of phenotypic values is symmetric, $RQR_{0.5}$ and RQR_{best} fits are the nearest curves obtained using the breeding values for each parameter of the logistic growth model (Figure 4). For the positive and negative scenarios, RQR_{best} yielded estimates with the least bias (Figures 5 and 6).

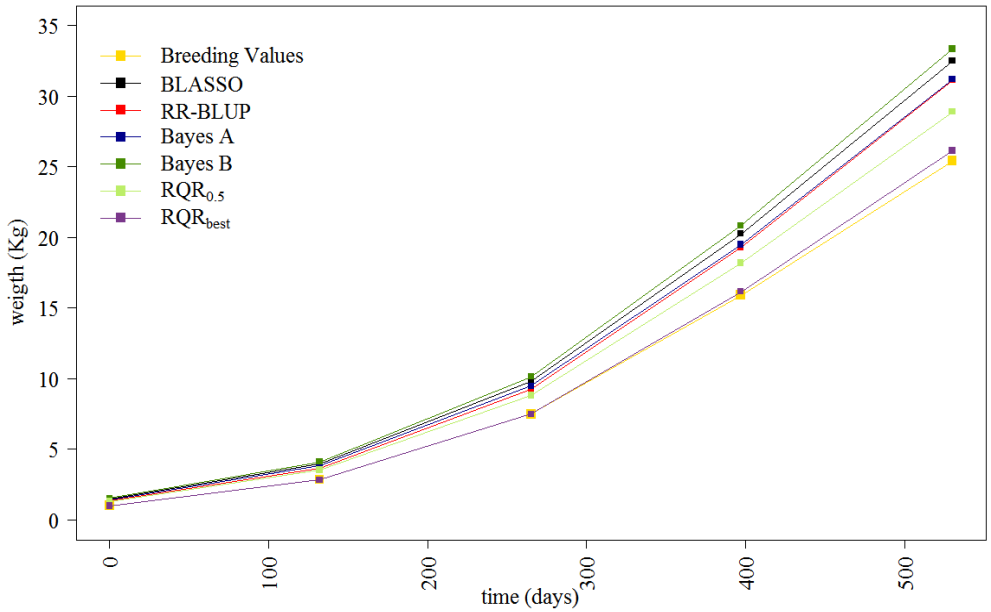

Figure 5. Mean genomic growth curves using the genomic estimated breeding values (GEBVs) obtained by 6 different genomic selection models (BLASSO, RR-BLUP, BayesA, BayesB, $RQR_{0.5}$, and RQR_{best}) and the breeding values for each parameter of a logistic growth curve considering the positive skewness scenario.

In general, when compared to the mean curve based on true values, the curve obtained by RQR_{best} outperformed those from the other models (BLASSO, RR-BLUP, BayesA, BayesB) (Figures 4, 5, and 6).

The curves generated by BLASSO and BayesB presented similar behaviors for RR-BLUP and BayesA. These curves, when compared with those obtained from quantile models $RQR_{0.5}$ and RQR_{best}, presented higher growth over the time range (from 0 to 530 days); the curve behavior was overestimated for the symmetric and positive skewness scenarios

(Figures 4 and 5). In the negative skewness scenario, compared with the RQR_{best} model, these models added to $RQR_{0.5}$ underestimated the curve behavior (i.e., presented lower growth over the range in the study) (Figure 6).

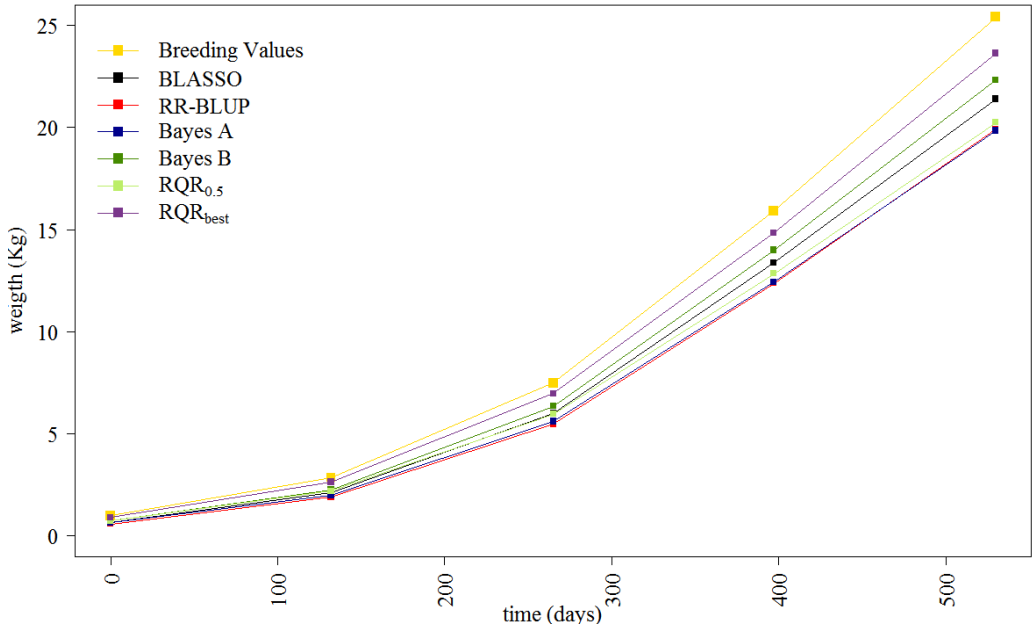

Figure 6. Mean genomic growth curves using the genomic estimated breeding values (GEBVs) obtained by 6 different genomic selection models (BLASSO, RR-BLUP, BayesA, BayesB, $RQR_{0.5}$, and RQR_{best}) and the breeding values for each parameter of a logistic growth curve considering the negative skewness scenario.

The average genomic growth curves (based on GEBVs) obtained using 13 different genomic selection models (BLASSO, RR-BLUP, BayesA, BayesB, $RQR_{0.1}$, $RQR_{0.2}$, ..., $RQR_{0.9}$) in each scenario are shown in Figures S2, S3, and S4 (available at https://det.ufv.br/moyses-nascimento/). As expected, for higher quantile regressions, the curve behavior shows higher growth over time when compared to the lower quantile regressions (Figures S2, S3, and S4 - available at https://det.ufv.br/moyses-nascimento/). The curves based on BLASSO, BayesB, RR-BLUP, and BayesA exhibit similar behaviors.

In general, the slopes between breeding values and GEBV of all models and scenarios were significantly different from the unit ($p < 0.01$), indicating a significant bias in the prediction. The RQR_{best} was not significantly different from that in the positive skewness scenario. Although evidence of significant bias was detected, the slope values derived from the RQR_{best} were slightly lower for all traits and scenarios (Table 1).

Table 1. Regression coefficient estimates of parameters regressed on breeding values for all scenarios and 6 different genomic selection models (BLASSO, RR-BLUP, BayesA, BayesB, $RQR_{0.5}$, and RQR_{best}).

	Scenarios		
Models	Symmetric	Positive	Negative
BLASSO	0.87*	0.79*	1.17*
RR-BLUP	0.89*	0.82*	1.25*
Bayes A	0.91*	0.82*	1.26*
Bayes B	0.87*	0.77*	1.12*
$RQR_{0.5}$	0.96*	0.89*	1.24*
RQR_{best}	1.05*	0.97^{ns}	1.07*

*Significant at 1% probability by *t*-test; ^{ns} = not significant.

Spearman's correlation (upper diagonal) and Cohen's Kappa concordance coefficients (lower diagonal) between GEBVs were obtained using six different genomic selection models (BLASSO, RR-BLUP, BayesA, BayesB, $RQR_{0.5}$, and RQR_{best}) considering the symmetric, positive, and negative skewness phenotypic distribution scenarios, as shown in Tables 2, 3, and 4.

In general, except for the $RQR_{0.5}$ model in scenario 1, Spearman's correlations varied from moderate to high positive values. The lowest Spearman's correlation was observed between the RR-BLUP and RQR_{best} models (0.52) in the negative-skewness scenario. The highest Spearman's correlation coefficient was observed for BayesA and BayesB (1.00) in the symmetric scenario (Tables 2, 3, and 4).

After ranking the individuals according to the GEBVs obtained from six different genomic selection models (BLASSO, RR-BLUP, BayesA, BayesB, RQR0.5, and RQR_{best}), considering all parameters and scenarios, the percentage of selected individuals in common was calculated using Cohen's kappa coefficient based on GEBVs for the first 100 individuals (lower diagonal Tables 2, 3, and 4). The RQR models present a Cohen's kappa concordance coefficient lower than 0.5. Specifically, the lowest value was observed between BLASSO and $RQR_{0.5}$ (0.14), and the highest value between BLASSO and RQR_{best} (0.82) in the symmetric scenario (Table 2).

Table 2. Comparison of estimates of Spearman's correlation (upper diagonal) and Cohen's Kappa concordance (lower diagonal) coefficients between GEBV values obtained using 6 different genomic selection models (BLASSO, RR-BLUP, BayesA, BayesB, $RQR_{0.5}$, and RQR_{best}) for each parameter of a logistic growth curve considering the symmetric scenario.

Parameters	Models	BLASSO	RR-BLUP	Bayes A	Bayes B	RQR_{best}	$RQR_{0.5}$
α_1	BLASSO	1.00	0.93	0.95	0.94	0.86	0.88
	RR-BLUP	0.69	1.00	0.88	0.88	0.87	0.87
	Bayes A	0.70	0.57	1.00	1.00	0.88	0.87
	Bayes B	0.70	0.57	0.99	1.00	0.88	0.87
	RQR_{best}	0.55	0.60	0.57	0.57	1.00	0.93
	$RQR_{0.5}$	0.62	0.58	0.63	0.63	0.73	1.00
α_2	BLASSO	1.00	0.93	1.00	0.99	0.88	-0.08
	RR-BLUP	0.74	1.00	0.93	0.95	0.73	0.14
	Bayes A	0.96	0.74	1.00	1.00	0.87	-0.09
	Bayes B	0.96	0.75	0.99	1.00	0.86	-0.06
	RQR_{best}	0.82	0.61	0.81	0.81	1.00	-0.13
	$RQR_{0.5}$	0.14	0.25	0.16	0.16	0.15	1.00
α_3	BLASSO	1.00	0.56	0.92	0.93	0.69	0.66
	RR-BLUP	0.49	1.00	0.68	0.65	0.55	0.67
	Bayes A	0.85	0.6	1.00	1.00	0.83	0.84
	Bayes B	0.87	0.57	0.97	1.00	0.82	0.82
	RQR_{best}	0.65	0.37	0.69	0.70	1.00	0.89
	$RQR_{0.5}$	0.60	0.55	0.72	0.71	0.68	1.00

RQR_{best}: $RQR_{0.4}$, $RQR_{0.3}$, and $RQR_{0.4}$ fit models for, respectively, parameters in the symmetric scenario; $RQR_{0.2}$, $RQR_{0.3}$, and $RQR_{0.1}$ fit models for the, α_1, α_2, and α_3 parameters in the positive skewness scenario and $RQR_{0.9}$ fit models for all parameters (α_1, α_2, and α_3) in the negative skewness scenario.

Table 3. Comparison of estimates of Spearman's correlation (upper diagonal) and Cohen's Kappa concordance (lower diagonal) coefficients between GEBV values obtained using 6 different genomic selection models (BLASSO, RR-BLUP, BayesA, BayesB, $RQR_{0.5}$, and RQR_{best}) for each parameter of a logistic growth curve considering the positive skewness scenario.

Parameters	Models	BLASSO	RR-BLUP	Bayes A	Bayes B	RQR_{best}	$RQR_{0.5}$
α_1	BLASSO	1.00	0.96	0.96	0.96	0.79	0.81
	RR-BLUP	0.74	1.00	0.85	0.85	0.65	0.73
	Bayes A	0.77	0.58	1.00	0.99	0.81	0.85
	Bayes B	0.81	0.58	0.95	1.00	0.85	0.85
	RQR_{best}	0.63	0.46	0.64	0.67	1.00	0.77
	$RQR_{0.5}$	0.59	0.45	0.69	0.68	0.57	1.00
α_2	BLASSO	1.00	0.98	1.00	0.90	0.56	0.37
	RR-BLUP	0.79	1.00	0.96	0.81	0.52	0.33
	Bayes A	0.95	0.74	1.00	0.93	0.57	0.39
	Bayes B	0.66	0.49	0.70	1.00	0.50	0.45
	RQR_{best}	0.31	0.33	0.28	0.22	1.00	0.60
	$RQR_{0.5}$	0.33	0.36	0.31	0.22	0.41	1.00
α_3	BLASSO	1.00	0.92	0.97	0.96	0.65	0.69
	RR-BLUP	0.69	1.00	0.82	0.79	0.53	0.52
	Bayes A	0.76	0.46	1.00	1.00	0.68	0.75
	Bayes B	0.76	0.46	0.99	1.00	0.67	0.76
	RQR_{best}	0.30	0.23	0.31	0.31	1.00	0.82
	$RQR_{0.5}$	0.39	0.23	0.48	0.49	0.45	1.00

RQR_{best}: $RQR_{0.4}$, $RQR_{0.3}$, and $RQR_{0.4}$ fit models for, respectively, parameters in the symmetric scenario; $RQR_{0.2}$, $RQR_{0.3}$, and $RQR_{0.1}$ fit models for, the, α_1, α_2, and α_3 parameters in the positive skewness scenario and $RQR_{0.9}$ fit models for all parameters (α_1, α_2, and α_3) in the negative skewness scenario.

Table 4. Comparison of estimates of Spearman's correlation (upper diagonal) and Cohen's Kappa concordance (lower diagonal) coefficients between GEBV values obtained using 6 different genomic selection models (BLASSO, RR-BLUP, BayesA, BayesB, $RQR_{0.5}$, and RQR_{best}) for each parameter of a logistic growth curve considering the negative skewness scenario.

Parameter	Models	BLASSO	RR-BLUP	Bayes A	Bayes B	$RQR_{0.5}$	RQR_{best}
α_1	BLASSO	1.00	0.91	0.92	0.90	0.78	0.84
	RR-BLUP	0.79	1.00	0.85	0.86	0.78	0.75
	Bayes A	0.85	0.71	1.00	1.00	0.81	0.89
	Bayes B	0.82	0.72	0.95	1.00	0.81	0.88
	$RQR_{0.5}$	0.64	0.61	0.66	0.67	1.00	0.79
	RQR_{best}	0.66	0.60	0.72	0.72	0.63	1.00
α_2	BLASSO	1.00	0.96	0.99	1.00	0.71	0.59
	RR-BLUP	0.86	1.00	0.97	0.97	0.73	0.54
	Bayes A	0.94	0.90	1.00	1.00	0.70	0.60
	Bayes B	0.95	0.91	0.99	1.00	0.71	0.60
	$RQR_{0.5}$	0.56	0.52	0.56	0.55	1.00	0.43
	RQR_{best}	0.53	0.43	0.49	0.51	0.35	1.00
α_3	BLASSO	1.00	0.98	0.96	0.97	0.80	0.64
	RR-BLUP	0.90	1.00	0.91	0.93	0.71	0.53
	Bayes A	0.71	0.70	1.00	1.00	0.87	0.76
	Bayes B	0.80	0.77	0.91	1.00	0.87	0.74
	$RQR_{0.5}$	0.47	0.42	0.56	0.55	1.00	0.82
	RQR_{best}	0.38	0.34	0.45	0.44	0.49	1.00

RQR_{best}: $RQR_{0.4}$, $RQR_{0.3}$, and $RQR_{0.4}$ fit models for, respectively, parameters in the symmetric scenario; $RQR_{0.2}$, $RQR_{0.3}$, and $RQR_{0.1}$ fit models for, the, α_1, α_2, and α_3 parameters in the positive skewness scenario and $RQR_{0.9}$ fit models for all parameters (α_1, α_2, and α_3) in the negative skewness scenario.

Discussion

We proposed genomic selection (GS) for growth curves based on RQR to obtain curves for different parts (quantiles) of the BW distribution in the presence of skew. The use of RQR to estimate genomic breeding value was efficient because, at least for one quantile model, the accuracy values presented better results when compared to those obtained from BLASSO, RR-BLUP, BayesA, and BayesB for all scenarios (Figures 1, 2, and 3). These results are reasonable because unlike traditional methods based on conditional expectations, E(Y|X), RQR allows fitting regression models on different parts of the distribution of the variable response, enabling a complete understanding of the phenomenon under study (Barroso et al., 2017; Cade & Noon, 2003; Koenker & Bassett Jr., 1978; Nascimento et al., 2017; Oliveira et al., 2021b). Therefore, it seems possible to find the best model to represent the relationship between the dependent (phenotype) and independent (marker effects) variables, thereby increasing the predictive performance of the model.

For the symmetric scenario, the best models for predicting α_1 (asymptotic weight), α_2 (inflection point), and α_3 (slope of the curve) genetic values were $RQR_{0.4}$, $RQR_{0.3}$, and $RQR_{0.4}$, respectively. These results make sense because the best-regularized quantiles are around the center of the distribution of the phenotypic values, which, in symmetrical situations, concentrate on the major mass of probability. As expected, under the skewness situation, the RQR models with lower ($RQR_{0.2}$, $RQR_{0.3}$, and $RQR_{0.1}$) and higher ($RQR_{0.9}$) quantiles presented better results for predicting GBVs for positive and negative skewness phenotypic distribution scenarios, respectively. In these situations, the mass of probability is concentrated on the lower and higher quantiles for positive and negative skewness phenotypic distributions, respectively.

Specifically, as with several traits, growth curve parameters can present different skewness levels, and RQR fit can improve model accuracy. In these cases, a functional relationship defined as higher (> 0.50) or lower quantiles (< 0.50) can improve GWS studies and subsequently improve the selection process of individuals in breeding programs. However, due to the infinite number of quantiles in RQR, finding the "best" one to explain the functional relationship is still a challenge.

In general, all models presented moderate to high positive Spearman correlations (Tables 1, 2, and 3). However, considering the agreement of Cohen's kappa coefficient, the classification agreement between RQR fit models and non-quantile models (RR-BLUP, BLASSO, BayesA, and BayesB) varied from moderate (0.60-0.79) to minimal (0.21-0.39) (McHugh, 2012). This result suggests differences between the quantile and non-quantile model classifications. The difference between the results of Spearman's correlations and Cohen's kappa coefficient occurs because kappa coefficients consider the possibility of random concordance.

Altogether, these concordance results indicate that using quantile regression to obtain curves on different parts of the BW distribution and even combining this information to set an RQR_{best} model is an interesting

and promising approach. In addition, the advantages of RQR are combined with those of the two-step approach. The two-step approach allows obtaining the GEBVs for any time in the observed range, ranking the animals using the GEBVs values directly for the parameter estimates and by the GEBVs of the evaluated trait.

Nonlinear QR to describe growth curves in plant breeding programs has already been used to evaluate dry matter accumulation in garlic plants by Puiatti et al. (2018; 2020), and the length and width of the fruit of pepper genotypes by Oliveira et al. (2021a). In all of these studies, nonlinear QR was efficient in fitting models at different levels and classifying genotypes. In the context of genomic selection, RQR has already been successfully applied by Nascimento et al. (2019a) to predict the genetic value of individuals in traits associated with the flowering time of the common bean, showing a very promising technique for the traditional techniques of genomic selection.

In animal breeding, nonlinear RQ has been successfully used to fit the lactation curve of dairy cows and growth curves in pigs at different established quantiles (Younesi, Shariati, Zerehdaran, Nooghabi, & Løvendahl, 2019; Nascimento et al., 2019b). In genomic data, Barroso et al. (2017), using the RQR, built genomic growth curves using RQR, which allowed the identification of genetically superior individuals in relation to growth efficiency. Furthermore, RQR enables us to find, in different quantiles, the most relevant markers for each trait evaluated and their respective chromosomal positions.

RQR is a promising and efficient technique in plant and animal breeding. However, more studies using different sizes of datasets (individuals and markers) are needed to address the efficiency of RQR. Other issues are related to the shrinkage parameter, which can be defined using a grid of values, cross-validation, or a Bayesian approach (Alhamzawi, Yu, & Benoit, 2012) and the definition of the best quantile fit model considering different levels of skewness.

Conclusion

The proposed model based on Quantile regression (QR) provided more accurate values than BLASSO, RR-BLUP, BayesA, and BayesB for all the simulated phenotypes with different skewness levels. The GEBV vectors obtained by RQR enabled the construction of genomic growth curves at different levels of interest (quantiles), comprehensively revealing the weight–age relationship.

References

Alhamzawi, R., Yu, K., & Benoit, D. F. (2012). Bayesian adaptive Lasso quantile regression. *Statistical Modelling*, *12*(3), 279-297. DOI: https://doi.org/10.1177/1471082X1101200304

Barroso, L. M. A., Nascimento, M., Nascimento, A. C. C., Silva, F. F., Serão, N. V. L., Cruz, C. D., ... Guimarães, S. E. F.. (2017). Regularized quantile regression for SNP marker estimation of pig growth curves. *Journal of Animal Science and Biotechnology*, *8*(59), 1-9. DOI: https://doi.org/10.1186/s40104-017-0187-z

Cade, B. S., & Noon, B. R. (2003). A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, *1*(8), 412-420. DOI: https://doi.org/10.1890/1540-9295(2003)001[0412:AGITQR]2.0.CO;2

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46. DOI: https://doi.org/10.1177/001316446002000104

Coster, A., Bastiaansen, J. W. M., Calus, M. P. L., van Arendonk, J. A. M., & Bovenhuis, H. (2010). Sensitivity of methods for estimating breeding values using genetic markers to the number of QTL and distribution of QTL variance. *Genetics Selection Evolution*, *42*(9), 1-11. DOI: https://doi.org/10.1186/1297-9686-42-9

Campos, C. F., Lopes, M. S., Silva, F. F., Veroneze, R., Knol, E. F., Sávio Lopes, P., & Guimarães, S. E. F. (2015). Genomic selection for boar taint compounds and carcass traits in a commercial pig population. *Livestock Science*, *174*, 10-17. DOI: https://doi.org/10.1016/j.livsci.2015.01.018

De Los Campos, G, & Pérez-Rodríguez, P. (2014). *Bayesian generalized linear regression. R package version 1.0. 4*. Vienna, AT: The R Foundation. Retrieved on July 16, 2018 from http://CRAN. R-Project. Org/Package= BGLR

de los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., ... Cotes, J. M. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics*, *182*(1), 375-385. DOI: https://doi.org/10.1534/genetics.109.101501

Howard, J. T., Jiao, S., Tiezzi, F., Huang, Y., Gray, K. A., & Maltecca, C. (2015). Genome-wide association study on legendre random regression coefficients for the growth and feed intake trajectory on Duroc Boars. *BMC Genetics*, *16*(59), 1-11. DOI: https://doi.org/10.1186/s12863-015-0218-8

Ibáñez-Escriche, N., & Blasco, A. (2011). Modifying growth curve parameters by multitrait genomic selection. *Journal of Animal Science*, *89*(3), 661-668. DOI: https://doi.org/10.2527/jas.2010-2984

Koenker, R. (2015). *Quantile Regression in R: a Vignette*. Retrieved on Feb. 28, 2018 from https://Cran. r-Project.Org/Web/Packages/Quantreg/Vignettes/Rq

Koenker, R., & Bassett Jr., G. (1978). Regression quantiles. *Econometrica*, *46*(1), 33-50. DOI: https://doi.org/10.2307/1913643

Li, Y., & Zhu, J. (2008). L 1-norm quantile regression. *Journal of Computational and Graphical Statistics*, *17*(1), 163-185. DOI: https://doi.org/10.1198/106186008X289155

Mathur, P. K., ten Napel, J., Bloemhof, S., Heres, L., Knol, E. F., & Mulder, H. A. (2012). A human nose scoring system for boar taint and its relationship with androstenone and skatole. *Meat Science*, *91*(4), 414-422. DOI: https://doi.org/10.1016/j.meatsci.2012.02.025

Maurer, A., Draba, V., Jiang, Y., Schnaithmann, F., Sharma, R., Schumann, E., ... Pillen, K. (2015). Modelling the genetic architecture of flowering time control in barley through nested association mapping. *BMC Genomics*, *16*(1), 1-12. DOI: https://doi.org/10.1186/s12864-015-1459-7

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276-282. DOI: https://doi.org/10.11613/BM.2012.031

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819-1829. DOI: https://doi.org/10.1093/genetics/157.4.1819

Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression: a second course in statistics*. London, UK: Pearson.

Younesi, H. N., Shariati, M. M., Zerehdaran, S., Nooghabi, M. J., & Løvendahl, P. (2019). Using quantile regression for fitting lactation curve in dairy cows. *Journal of Dairy Research*, *86*(1), 19-24. DOI: https://doi.org/10.1017/S0022029919000013

Nascimento, M., Silva, F. F., Resende, M. D.V., Cruz, D, C., Nascimento, A. C.C., Viana, J. M. S., Azevedo, C. F., & Barroso, L. M. A. (2017). Regularized quantile regression applied to genome-enabled prediction of quantitative traits. *Genetics and Molecular Research*, *16*(1), 1-12. DOI: https://doi.org/10.4238/gmr16019538

Nascimento, A. C., Nascimento, M., Azevedo, C., Silva, F., Barili, L., Vale, N., ... Serão, N. (2019a). Quantile regression applied to genome-enabled prediction of traits related to flowering time in the common bean. *Agronomy*, *9*(12), 1-10. DOI: https://doi.org/10.3390/agronomy9120796

Nascimento, M., Nascimento, A. C. C., Dekkers, J. C. M., & Serão, N. V. L. (2019b). Using quantile regression methodology to evaluate changes in the shape of growth curves in pigs selected for increased feed efficiency based on residual feed intake. *Animal*, *13*(5), 1009-1019. DOI: https://doi.org/10.1017/S1751731118002616

Oliveira, A. C. R., Cecon, P. R., Puiatti, G. A., Guimarães, M. E. S., Cruz, C. D., Finger, F. L., ... Lacerda, M. S. (2021a). Nonlinear models based on quantiles in the fitting of growth curves of pepper genotypes. *Revista Brasileira de Biometria*, *39*(3), 447-459. DOI: https://doi.org/10.28951/rbb.v39i3.505

Oliveira, G. F., Nascimento, A. C. C., Nascimento, M., Sant'Anna, I. C., Romero, J. V., Azevedo, C. F., ... Moura, E. T. C. (2021b). Quantile regression in genomic selection for oligogenic traits in autogamous plants: A simulation study. *PLoS ONE*, *16*(1), 1-12. DOI: https://doi.org/10.1371/journal.pone.0243666

Pérez, P., de Los Campos, G., Crossa, J., & Gianola, D. (2010). Genomic-enabled prediction based on molecular markers and pedigree using the Bayesian linear regression package in R. *The Plant Genome*, *3*(2), 106-116. DOI: https://doi.org/10.3835/plantgenome2010.04.0005

Pong-Wong, R., & Hadjipavlou, G. (2010). A two-step approach combining the Gompertz growth model with genomic selection for longitudinal data. *BMC Proceedings*, *4*(1), 1-5. DOI: https://doi.org/10.1186/1753-6561-4-S1-S4

Puiatti, G. A., Cecon, P. R., Nascimento, M., Nascimento, A. C. C., Carneiro, A. P. S., Silva, F. F., ... Oliveira, A. C. R. (2018). Quantile regression of nonlinear models to describe different levels of dry matter accumulation in garlic plants. *Ciência Rural*, *48*(1), 1-6. DOI: https://doi.org/10.1590/0103-8478cr20170322

Puiatti, G. A., Cecon, P. R., Nascimento, M., Nascimento, A. C. C., Carneiro, A. P. S., Silva, F. F., ... Cruz, C. D. (2020). Nonlinear quantile regression to describe the dry matter accumulation of garlic plants. *Ciência Rural*, *50*(1), 1-8. DOI: https://doi.org/10.1590/0103-8478cr20180385

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, AT: R Foundation for Statistical Computing.

Silva, G. N., Nascimento, M., Sant'Anna, I. C., Cruz, C. D., Caixeta, E. T., Carneiro, P. C. S., ... Oliveira, M. S. (2017). Artificial neural networks compared with Bayesian generalized linear regression for leaf rust resistance prediction in Arabica coffee. *Pesquisa Agropecuária Brasileira*, *52*(3), 186-193. DOI: https://doi.org/10.1590/s0100-204x2017000300009

Smith, B. J. (2007). boa: An R package for MCMC output convergence assessment and posterior inference. *Journal of Statistical Software*, *21*(11), 1-37. DOI: https://doi.org/10.18637/jss.v021.i11

Varona, L., Ibañez-Escriche, N., Quintanilla, R., Niguera, J. L., & Casellas, J. (2008). Bayesian analysis of quantitative traits using skewed distributions. *Genetics Research*, *90*(2), 179-190. DOI: https://doi.org/10.1017/S0016672308009233

Ward Jr., J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*, 236-244. DOI: https://doi.org/10.1080/01621459.1963.10500845