BIOMETRY, MODELLING AND STATISTIC

Statistical modeling of vigor ratings in ruzigrass breeding

Jales Mendes Oliveira Fonseca[1], Flávia Maria Avelar Gonçalves[2], Fausto Souza Sobrinho[3], Júlio Sílvio de Sousa Bueno Filho[4], Flávio Rodrigo Gandolfi Benites[3], Davi Henrique Lima Teixeira[5] and José Airton Rodrigues Nunes[2*]

[1]Bayer Crop Sciences, 55018, Stanton, Minnesota, United States of America. [2]Departamento de Biologia, Instituto de Ciências Naturais, Universidade Federal de Lavras, Lavras, Minas Gerais, Brazil. [3]Empresa Brasileira de Pesquisa Agropecuária, Embrapa Gado de Leite, Coronel Pacheco, Minas Gerais, Brazil. [4]Departamento de Estatística, Instituto de Ciências Exatas e Tecnológicas, Universidade Federal de Lavras, Lavras, Minas Gerais, Brazil. [5]Universidade Federal Rural da Amazonia, Capitão Poço, Pará, Brazil. *Author for correspondence. E-mail: jarnunes@ufla.br

ABSTRACT. Ruzigass (*Urochloa ruziziensis*) is a forage crop with high agronomic and nutritional value. Plant breeders often assess ruzigrass phenotypic traits via vigor ratings. The analyses of these categorical data often fail to meet the usual statistical assumptions. In this study, we compared four fittings of linear models for vigor rating analyses: i) a linear mixed model for the original scale (LMM), ii) a linear mixed model for a Box–Cox transformed scale (BCLMM), iii) a multinomial generalized mixed model using a probit link function, also known as threshold model (GLMM), and iv) a hierarchical Bayesian model, also referred to as a Bayesian threshold model (HBM). Additionally, biomass yield was assessed, and the indirect selection of high-performing genotypes was evaluated. The experimental design included 2,204 ruzigrass genotypes randomized into augmented blocks. Six graders visually assessed each plot using a rating scale. Fitting methods were sampled from three scenarios, employing one, three, or six graders. A nonnull genetic variance component was detected for vigor and biomass yield traits. Except for BCLMM, the methods for analyzing vigor ratings were correlated. The correlations and coincidence indices for selecting genotypes increased with the number of graders. The analysis of vigor ratings under Gaussian approximations is riskier when a single grader is used to evaluate genotypes. The GLMM and HBM perform similarly and are more recommended and suitable analyses of vigor ratings when selecting high-performing ruzigrass genotypes.

Keywords: *Urochloa ruziziensis*; selection efficiency; Box-Cox transformation; threshold model; mixed model approach; forage breeding.

Introduction[1]

Ruzigass [*Urochloa ruziziensis* (R. Germ. & CM Evrard) Crins.], syn. *Brachiaria ruziziensis*, is a crop widely used in forage breeding programs due to its high agronomic and nutritional value (Marcelino et al., 2020; Nouhoun et al., 2022). It is an obligate sexual species under ploidy level (2n = 2x = 18). In a tetraploid level (2n = 4x = 36), it has been used to obtain interspecific hybrids (Pereira et al., 2005; Timbó et al., 2014).

The development of a new ruzigrass cultivar requires selecting and recombining the best-performing genotypes across multiple breeding cycles to accumulate favorable alleles and produce desirable agronomic traits. Assessing these traits in forage crops can be costly, time-consuming, and labor-intensive, depending on the stage and size of the trials. Forage breeders have often assigned categorical ratings to discriminate genotypes in field evaluations for several important traits, such as resistance to pests (Silva et al., 2013), regrowth capacity (Gouveia et al., 2022), turf quality (Whitman et al., 2022), flowering (Marcon et al., 2021), and biomass production (Price & Casler, 2014; Teixeira et al., 2020; Dias et al., 2022). The Embrapa Dairy Cattle Research Center has used vigor ratings to select from many genotypes at the early stages of the ruzigrass breeding program. The vigor rating was devised to adjust the biomass rating by considering various plant aspects, such as health, size of leaves and stalks, height, regrowth, and tillering. However, the selection of genotypes has been based solely on realized field ratings, with no further statistical analysis.

Some statistical analyses have been used to address ratings or scores provided by plant breeders. Ratings are ordinal categorical variables (Agresti, 2002); however, average ratings have often been successfully

---

[1] The article was made available as a pre-print under the DOI number https://doi.org/10.21203/rs.3.rs-3970804/v1

approximated by Gaussian distributions through the central limit theorem (Mood et al., 1974). Sometimes, this assumption does not hold, and nonlinear data transformations have been applied (e.g., Box & Cox, 1964). The search for a model that accurately represents the data may improve the interpretation of parameters and predictive ability, thus avoiding spurious results (Jaeger, 2008; Bolker et al., 2009; Stroup, 2015). For breeding purposes, this could avoid erroneous genotype rankings and lead to more efficient selection.

Plant breeders are continuously searching for flexible methods to analyze data. New statistical tools available in free software (R packages) have been used to avoid transforming data into count data or proportions (Venables & Ripley, 2002; Bates et al., 2015; Christensen, 2019; Hadfield, 2010; R Core Team, 2020). Generalized linear mixed models can be fit to different distributions of the exponential family and are an extension of linear mixed models (McCulloch & Searle, 2004). The exponential family unifies several random variables, among which a multinomial distribution can be employed for ordinal categorical data (Nelder & Wedderburn, 1972). Those models could be fit either in a frequentist (McCulloch & Searle, 2004) or a Bayesian approach (Gianola & Foulley, 1983; Gianola & Fernando, 1986). In the Bayesian approach, a hierarchical model formulation is needed to construct prior distributions, from which joint posterior distributions for parameters and data can be determined. This method provides a convenient way to infer marginal distributions for parameters. Gouy et al. (2013) emphasized the suitability of threshold models under a hierarchical Bayesian framework for analyzing rust resistance ratings to select sugarcane clones. Correa et al. (2016) used the usual hierarchical formulation for threshold models from animal and plant breeding purposes.

In this study, we aim to find flexible models to analyze vigor ratings for efficiently selecting high-performing ruzigrass genotypes by comparing four methods: i) a linear mixed model at the original scale (LMM), ii) a linear mixed model at the transformed Box–Cox scale (BCLMM), iii) a multinomial generalized linear mixed model (GLMM), and iv) a hierarchical Bayesian model (HBM). The analyses use R packages. Additionally, the effect of having multiple graders in the final selection of ruzigrass genotypes using different methods is investigated.

Material and methods

Location

The study was conducted at the Embrapa Dairy Cattle experimental station, located at 21°33' S latitude and 43°06' W longitude, at 410 m above sea level in Coronel Pacheco, Minas Gerais State, Brazil. The soil is predominantly classified as red yellow Argisol (Santos et al., 2006). According to Köppen, the climate is classified as humid subtropical (Cwa type mesothermic), with an average annual temperature of 19°C. Winters are commonly dry and cold, while summers are rainy with moderately high temperatures. The average annual rainfall is 1,536 mm.

Experimental Design and Germplasm

In August 2011, the experiment was laid out as an augmented block design with 51 blocks, each containing between 28 to 72 genotypes. The nonreplicated treatments consisted of 2,204 ruzigrass genotypes derived from selected seeds during the second recurrent intraspecific forage breeding program cycle at Embrapa Dairy Cattle. Marandu (*U. brizantha*) and Basilisk (*U. decumbens*) cultivars were used as controls and were thus replicated in all blocks. The plants were transplanted, and the plots consisted of a single plant with an average area of 1.5 m^2, spaced one meter apart.

A total of 350 kg of the formula 8-28-16 (nitrogen–phosphorus–potassium) was applied per hectare. One ton of the formula 20-05-20 (nitrogen–phosphorus–potassium) was top-dressed per hectare, partitioned across each cut during the rainy season. Manual weeding was performed as needed, and a homogenizing cut was carried out at the beginning of December 2011. In 2012, two evaluation cuts were performed, one in January and a second approximately 40 days later. For this study, only the first evaluation cut was analyzed.

Traits Evaluated

Vigor rating and biomass yield were assessed per plot. Visual ratings of plant vigor ranged from 1 to 5 as follows: 1 being very bad, 2 bad, 3 regular, 4 good, and 5 very good (Fonseca et al., 2020). Six different graders performed visual assessments immediately before the cutting. To evaluate biomass yields, the plots were cut at 5 cm above

the soil level, and biomass was weighed using a portable suspension scale, measured in grams.

Statistical analysis

Biomass yield

Biomass yield data were checked for outliers, and their residuals were tested for normality. The residuals were inspected using both the Shapiro–Wilk test (Shapiro & Wilk, 1965) and diagnostic plots (Q–Q plot) via the *shapiro.test* and *qqnorm* functions of the *stats* R package (R Core Team, 2020). The model presented in Fonseca et al. (2020) was fitted as follows:

$$y = X\mu + Z_1 b + Z_2 g + e, \tag{1}$$

where y is the vector of biomass yield; μ is the intercept; b is the vector of block effects, $b \sim N(0, I\sigma_b^2)$; g is the vector of genotype effects, $g \sim N(0, I\sigma_g^2)$; e is the vector of errors, $e \sim N(0, I\sigma_e^2)$; and X, Z_1, and Z_2 are the design matrices of the effects μ, b, and g, respectively. The estimation of fixed effects (best linear unbiased estimator - BLUE) and the prediction of random effects (best linear unbiased prediction - BLUP) were performed using Henderson's system of equations (Henderson, 1984). Variance components were estimated using the restricted maximum likelihood (REML) method (Patterson & Thompson, 1971). The significance of variance components was evaluated by the likelihood ratio test at a significance level of 5%, using the *lmerTest* package (Kuznetsova et al., 2017) available in R (R Core Team, 2020). The *lmer* function of the *lme4* R package was used to fit the linear mixed model (Bates et al., 2015). Repeatability (R) was calculated following Cullis et al. (2006), as presented in Schmidt et al. (2019): $R_{Cullis}^2 = 1 - \left(\overline{v}_{\Delta..}^{BLUP} / 2\sigma_g^2\right)$, where $\overline{v}_{\Delta..}^{BLUP}$ is the mean variance of the difference between two BLUPs and σ_g^2 is the genotypic variance based on biomass.

Vigor rating

Vigor ratings were analyzed using four approaches: i) a linear mixed model in the original scale (LMM), ii) a linear mixed model in the transformed Box–Cox scale (BCLMM), iii) a multinomial generalized linear mixed model (GLMM), and iv) a multinomial hierarchical Bayesian model (HBM) across three distinct scenarios as follows:

Scenario 1 – One grader

To analyze the vigor rating via the LMM, Model (1) was applied, replacing the response variable biomass with a vigor rating. Similarly, for the BCLMM, Model (1) was applied to adjust vigor ratings under the Box–Cox transformation scale using the following expression:

$$y_t(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & if \lambda \neq 0, \\ log\, y, & if \lambda = 0 \end{cases},$$

where λ is the Box–Cox transformation parameter; y is the vector of vigor ratings; and $y_t(\lambda)$ is the vector of transformed vigor ratings. The λ estimate was calculated by maximizing the likelihood function (Box & Cox, 1964) of Model (1), assuming that all effects were fixed. For this purpose, the *boxcox* function available in the *MASS* R package was used (Venables & Ripley, 2002). The normality of the residuals was checked using the Shapiro–Wilk test (Shapiro & Wilk, 1965) and the Q–Q plot procedure via the respective functions *shapiro.test* and *qqnorm* functions from *stats* R package (R Core Team, 2020).

To investigate the vigor ratings via the GLMM under a frequentist approach, the data were analyzed via a threshold model assuming a multinomial distribution (Agresti, 2002; Agresti, 2007) and a probit link function as follows:

$$\Phi^{-1}(\pi_c | \mu, b, g) = X\mu + Z_1 b + Z_2 g, \tag{2}$$

where Φ^{-1} is the probit link function; π_c is the probability associated with each category; c is the number of categories (i.e., c=5 in this study); μ is the intercept; b is the vector of block effects, $b \sim N(0, I\sigma_b^2)$; g is the vector of genotype effects, $g \sim N(0, I\sigma_g^2)$; and X, Z_1, and Z_2 are the design matrices for the effects μ, b and g, respectively. Maximum likelihood estimates of parameters and BLUPs of genotypic effects were obtained via Laplace approximation to compute the likelihood function (Pinheiro & Chao, 2006). The *clmm* function of the *ordinal* R package was used (Christensen, 2019). The significance values of variance components were assessed via the likelihood ratio test. Repeatability of vigor ratings under the LMM, BCLMM, and GLMM was calculated following Cullis et al. (2006), as previously described.

For the analysis of vigor ratings via the HBM, i.e., the GLMM under a Bayesian approach, Model (2) was extended to incorporate priors for each effect. An inverse Gamma *prior* was used for block and genotype effect, a noninformative *prior* for fixed effects, and a fixed *prior* of 1 for the residual (Hadfield, 2010). The model ran for 500,000 Markov chain Monte Carlo (MCMC) simulation iterations with a burn-in of 100,000 samples, thinning every 200 iterations, resulting in a posterior distribution of 2,000 samples of estimated parameters. The mean and highest posterior density (HPD) intervals were determined on the basis of the samples. The *MCMCglmm* function of the *MCMCglmm* R package was used (Hadfield, 2010). Repeatability was calculated following the standard repeatability procedure as $R = \dot{\sigma}_g^2 / (\dot{\sigma}_g^2 + \dot{\sigma}_\Phi^2)$, where $\dot{\sigma}_g^2$ is the genetic variance estimate and where $\dot{\sigma}_\Phi^2$ is the probit distribution variance, which equals 1. All methods used to assess vigor ratings were applied for each grader.

Scenario 2 – Three graders

To analyze vigor ratings via the LMM and BCLMM, Model (1) was applied, and the average vigor rating was used as the response variable. Alternatively, for the GLMM and HBM, vigor ratings were analyzed via an extension of Model (2) to include the random effect of graders. Thus, these models allowed the vigor rating to be assessed under its original scale instead of the average of the vigor ratings. For the HBM, an inverse Gamma *prior* was used to model the grader effect. In this scenario, the same three randomly selected graders were used in all methods.

b3) Scenario 3 – Six graders

Similar models presented in scenario 2 were developed, considering all six graders involved in this study. Therefore, the only difference between scenarios 2 and 3 is the number of graders used to calculate the average vigor ratings for the LMM and BCLMM. For the GLMM and HBM, the difference is the inclusion of three graders within their grader effect.

Comparison between vigor ratings and scenarios and their correspondence with biomass yield

Correlation and selection efficiency

Spearman correlations between biomass yields and vigor ratings were obtained from the BLUPs of the genotypes for all the scenarios and methods. The *cor* function from the *stats* R package was used (R Core Team, 2020). The selection efficiency was calculated following a modified version of the coincidence index (*CI*) proposed by Hamblin and Zimmermann (1986). Originally, the CI aimed to assess the proportion of coincident genotypes selected under two different planting systems, assuming a specific selection intensity (SI). In this study, CI was used to calculate the selection efficiency of each model to select the top 5% high-performing genotypes via grades as follows: $CI = (A - C)/(B - C)$, where A is the number of genotypes selected on the basis of both grade and biomass simultaneously; B is the top 5% of high-performing genotypes on the basis of biomass; and C is the expected number of coincident genotypes selected by chance, i.e., a proportion of B due to chance (C = B×0.05).

Response to selection simulation

Expected genetic gains based on direct and indirect selection were calculated considering the top 5% BLUP of genotypes. Responses to selection were presented on both normal and categorical scales.

Results

The genetic variance was significant for biomass yield and vigor ratings across all methods and scenarios (Table 1). The distribution of vigor ratings varied across graders (Figure 1), impacting the proportion of variance explained among different graders and methods (Figure 2). The repeatability for biomass yield was 0.46, whereas for vigor ratings, such magnitudes varied from 0.35 (HBM) to 0.85 (BCLMM) in scenario 1 (one grader). For scenario 2 (three graders), substantial differences in repeatability, latent variables, and threshold estimates occurred between the GLMM and HBM methods. Alternatively, the LMM and BCLMM methods presented slight changes in terms of repeatability, although the genetic and residual variances varied in magnitude. There were minor variations from scenarios 2 to 3 when all six graders were assessed, with the exception of the range of values for the latent variable in the HBM.

The normality of the vigor rating residuals was assessed before and after performing a Box–Cox transformation via both a statistical test (Shapiro–Wilk normality test) and diagnostic plots (Q–Q plot). The results from both procedures revealed that the vigor rating residuals on the original scale did not follow a

normal distribution, regardless of the number of graders considered in the analysis. Additionally, the application of the Box–Cox transformation was not effective in ensuring normality of the residuals (data not shown).

Table 1. Estimates of the means (μ), genetic variances ($\hat{\sigma}_g^2$), error variances ($\hat{\sigma}_e^2$), repeatability (R^2), latent variables, and thresholds for biomass yields and vigor ratings across three scenarios.

Parameter	Biomass	Vigor ratings					
		LMM	BCLMM	GLMM	HBM		
					LI 2.5%	Posterior mean	LS 2.5%
Scenario 1 (grader A)							
μ	1.2 kg plant^{-1}	2.83	14.24	-	-	-	-
$\hat{\sigma}_g^2$	0.1418*	0.32*	180.92*	0.55*	0.25	0.56	0.92
$\hat{\sigma}_e^2$	0.1579	0.50	28.03	1.00	-	1.00	-
R^2	0.46	0.38	0.85	0.34	0.20	0.35	0.48
Latent variable	-	-	-	-	1.62	1.92	2.21
Threshold 1	-	-	-	-1.93	1.32	1.49	1.69
Threshold 2	-	-	-	-0.43	2.58	2.89	3.23
Threshold 3	-	-	-	0.97	3.75	4.22	4.72
Threshold 4	-	-	-	2.30	-	-	-
Scenario 2 (graders A, B, and C)							
μ		3.37	71.7	-	-	-	-
$\hat{\sigma}_g^2$		0.21*	3205.0*	3.00*	2.71	2.98	3.29
$\hat{\sigma}_e^2$		0.39	660.3	1.00	-	1.00	-
R^2		0.35	0.82	0.88	0.73	0.75	0.76
Latent variable		-	-	-	0.53	4.16	7.99
Threshold 1		-	-	-4.25	2.14	2.25	2.38
Threshold 2		-	-	-1.98	4.39	4.55	4.71
Threshold 3		-	-	0.32	6.30	6.49	6.70
Threshold 4		-	-	2.26	-	-	-
Scenario 3 (all graders)							
μ		3.07	63.18	-	-	-	-
$\hat{\sigma}_g^2$		0.21*	3354.8*	2.72*	0.54	0.94	1.37
$\hat{\sigma}_e^2$		0.34	484.4	1.00	1.00	1.00	1.00
R^2		0.38	0.86	0.93	0.22	0.51	0.79
Latent variable		-	-	-	2.07	3.49	5.08
Threshold 1		-	-	-3.49	2.28	2.35	2.43
Threshold 2		-	-	-1.14	4.39	4.49	4.59
Threshold 3		-	-	0.99	6.17	6.30	6.43
Threshold 4		-	-	2.80	-	-	-

*Significant by the likelihood ratio test at a p value<0.001; LMM: vigor ratings fitted by a linear mixed model; BCLMM: vigor ratings fitted by Box–Cox transformations; GLMM: vigor ratings fitted by a generalized linear mixed model; HBM: vigor ratings fitted by a hierarchical Bayesian model; IL: inferior limit; SL: superior limit.
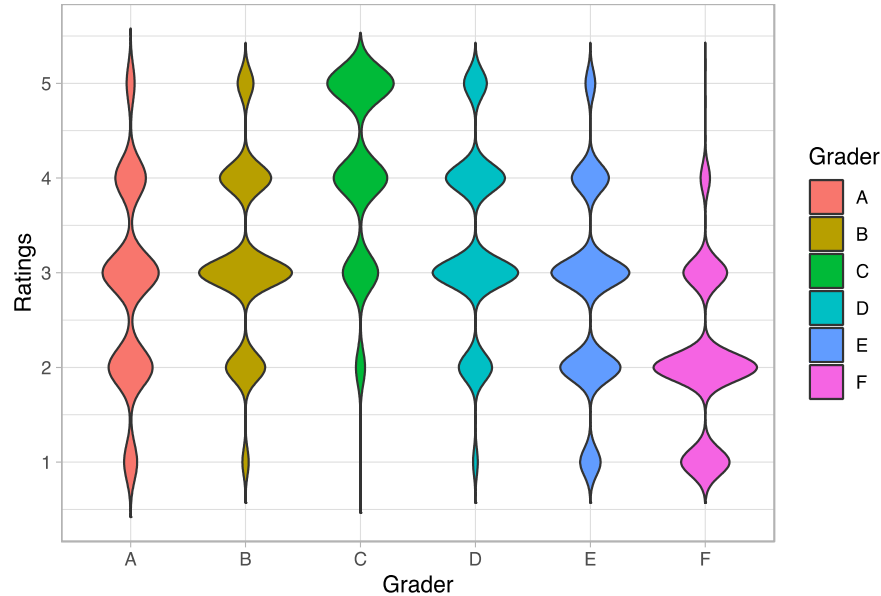

Figure 1. Distribution of vigor ratings across graders when assessing ruzigrass genotypes visually.
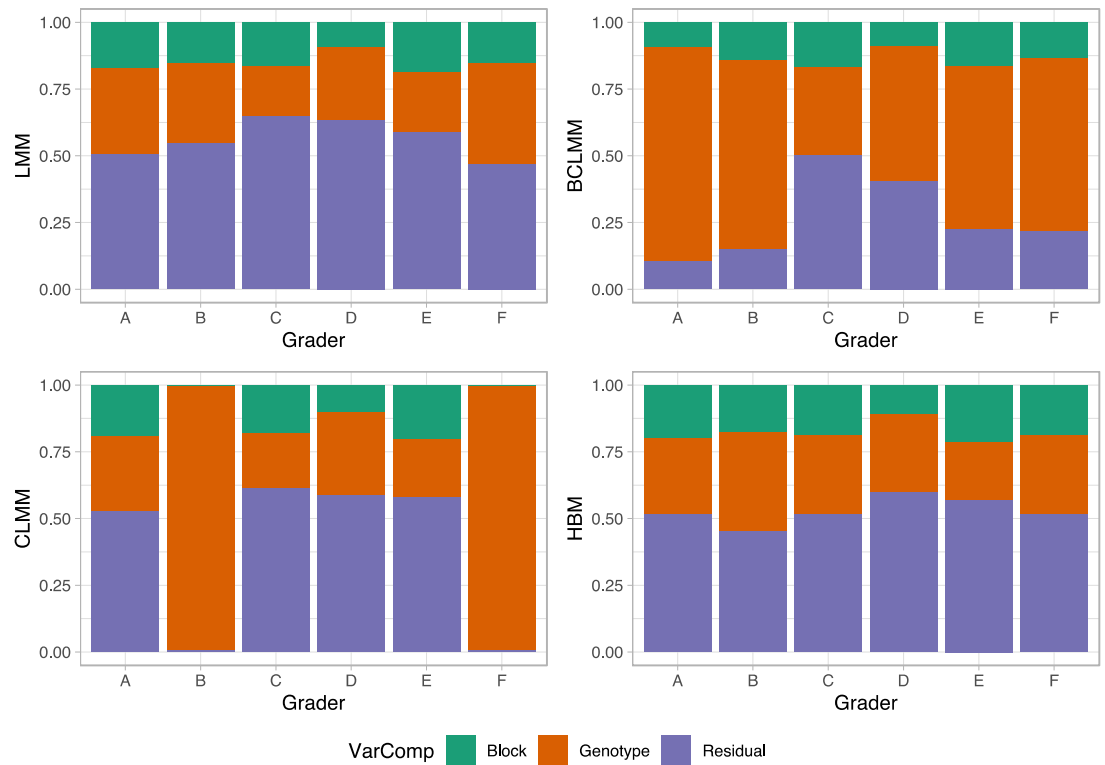
Figure 2. The proportion of variance components based on different graders and methods.

The correlations (r) between the BLUPs of genotypes for vigor ratings obtained from different methods were high overall, ranging from 0.85 to 0.99 across the scenarios (Figures 3, 4, and 5). While the inclusion of graders slightly decreased the correlation between the BLUPs of genotypes for vigor ratings, the correlation between vigor ratings and biomass increased. This increase reflected the coincidence index estimates, whose values were highest for the GLMM (0.44) and HBM (0.43) (Figure 5). The coincidence index between vigor ratings across different methods was high overall, regardless of the scenario, except for the BCLMM (Figures 3, 4, and 5). These results indicate the potential benefit of selecting high-performing genotypes at earlier stages of ruzigrass breeding via vigor ratings. Nonetheless, the method and the number of graders can limit the benefits of indirectly selecting genotypes.
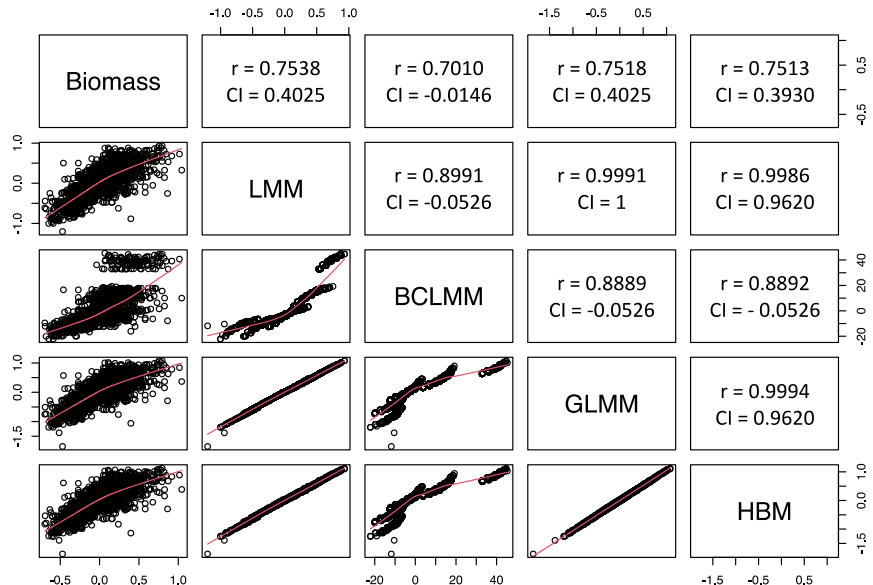

Figure 3. Spearman correlations and coincidence indices between the BLUPs of various genotypes for the biomass yields and vigor ratings using different methods in scenario 1 (one grader).
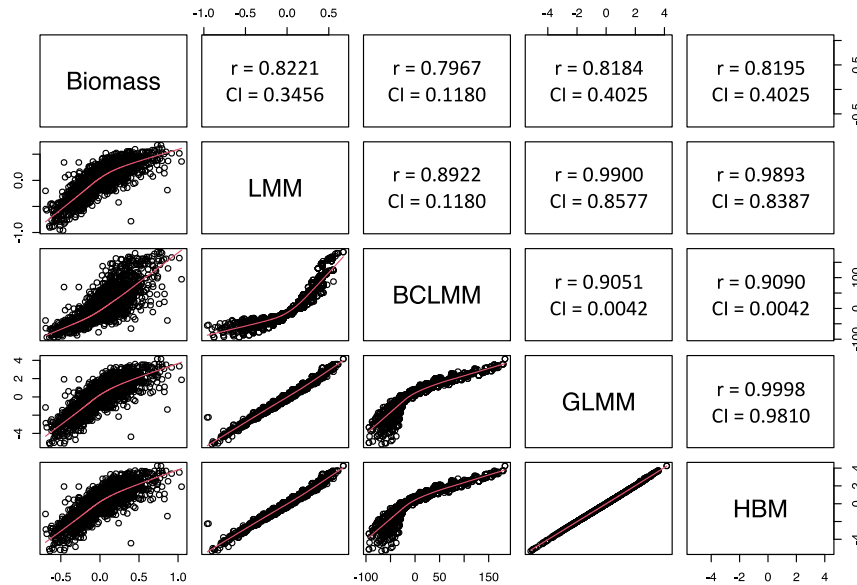
Figure 4. Spearman correlations and coincidence indices between the BLUPs of biomass yields and the BLUPS of average vigor ratings across different methods in scenario 2 (three graders).
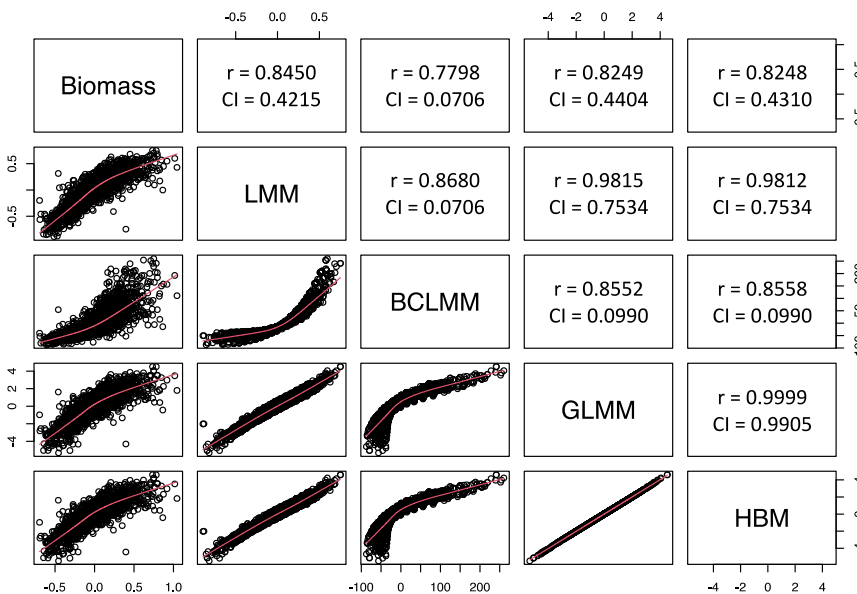

Figure 5. Spearman correlations and coincidence indices between the BLUPs of the biomass yields and the BLUPS of the vigor ratings across different methods for scenario 3 (six graders).

Simulations based on the direct response to selection using the top 5% BLUP of genotypes revealed a 50% increase in biomass (Table 2), whereas the indirect response to selection reported relatively lower values. In scenario 1, vigor ratings resulted in an indirect response with approximately a 35% increase in biomass via the LMM, GLMM, and HBM, while the BCLMM resulted in an approximately 12% increase. Similar results were observed in scenarios 2 and 3, with a significant increase only when the BCLMM was applied.

Table 2. Simulations based on direct and indirect responses to selection, considering the top 5% BLUPs of genotypes.

	Response to selection (kg plot^{-1})		
Biomass	0.603		
	Scenario 1	Scenario 2	Scenario 3
LMM	0.418	0.423	0.444
BCLMM	0.140	0.273	0.232
GLMM	0.418	0.442	0.435
HBM	0.418	0.445	0.433

The application of generalized models using the probit link function provided estimates of thresholds between categories, allowing the representation of the response to selection based on the vigor ratings' ordinal categorical scale (Figure 6). Moreover, these threshold estimates enabled estimation of the probability of a genotype being found in a given category. For example, approximately 80% of the genotypes were likely to be found in categories 2 and 3, whereas 2.71 and 1.05% of the genotypes were likely to be present in categories 1 and 5, respectively (red graph). After simulating the selection of the top 5% BLUP of genotypes on the basis of vigor ratings, there was an increase in the likelihood of finding genotypes in higher categories. For example, the probabilities of finding genotypes in categories 4 and 5 were 40.04 and 8.45%, respectively (blue graph), whereas these probabilities were significantly lower in the original population (red graph).
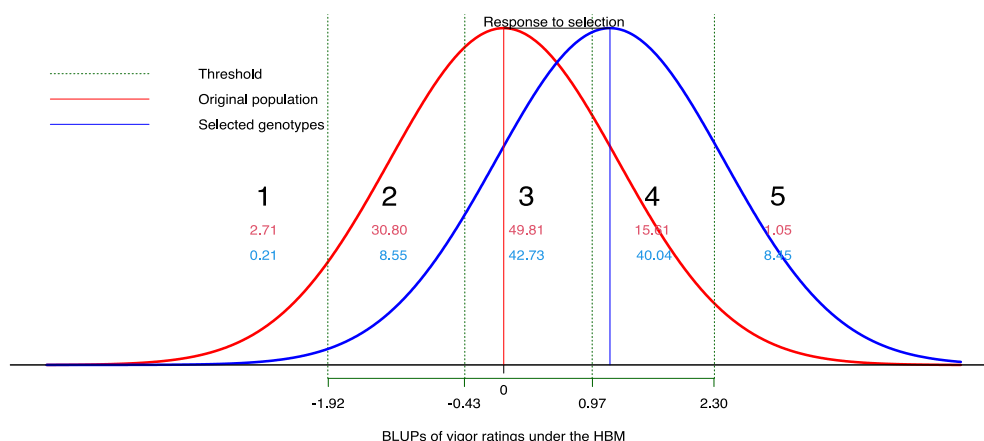

Figure 6. Predictions of genetic gains based on the rating in scenario 1 for the HBM.

Discussion

The use of visual ratings in forage breeding is justified due to their feasibility and efficiency in indirectly selecting superior genotypes. However, the genetic architecture of each trait should be considered, as it influences selection efficiency. For example, qualitative traits, controlled by fewer loci and less influenced by the environment, can be efficiently assessed through ratings (i.e., grade scales). In contrast, for quantitative traits, ratings can also be useful if the correlations between the trait of interest and their corresponding ratings are relatively high. For example, Silva et al. (2013) used ratings to assess the genetic value of ruzigrass genotypes resistant to *Collaria oleosa*. Souza Sobrinho et al. (2010) indirectly selected ruzigrass genotypes that were resistant to two species of spittlebugs (*Mahanarva spectabilis* and *Deois schach*) via visual ratings. Riday (2009) reported the efficiency of visual ratings in selecting high-yielding red clover genotypes. In addition to their efficiency in selecting superior genotypes, visual ratings can also assist breeders in eliminating undesirable genotypes (Atkins, 1964; Burton, 1982).

Notably, the vigor rating analyses presented herein in three different scenarios have specific uses based on the populations under investigation and the selection phase of the breeding pipeline. Breeding programs have distinct evaluation/selection phases that ultimately lead to new cultivars. As these phases progress and selections are made, genetic variances among genotypes diminish. The use of multiple graders enhances the method's power to detect real differences between similar genotypes, thus increasing genetic gains (Fonseca et al., 2020). However, a single grader might be considered for selecting high-performing genotypes (or discarding poor ones) in the early stages of selection cycles where the genetic variances between genotypes are highest. This study focused on the evaluation of ruzigrass genotypes at early stages of a selection cycle following the forage breeding pipeline at Embrapa Dairy Cattle. Broad genetic variation in biomass yields and vigor ratings was observed, enabling the selection of superior genotypes. However, low repeatability estimates for both traits indicated challenges in assessing true genotypic merit, either due to i) pronounced environmental effect associated with the experimental design used (augmented block design) or ii) the quantitative genetic architecture of the traits under field evaluation (biomass yields and plant vigor ratings). Therefore, forage breeding programs that follow Embrapa Dairy Cattle should avoid the use of a single grader to perform biomass selection via vigor ratings. Although visual screening of ruzigrass during field evaluations

is possible, the effect of a single grader can vary considerably, ultimately impacting decisions. Fonseca et al. (2020) evaluated the number of graders for visual assessment, recommending at least three graders for ruzigrass.

In addition to the variation related to the number of graders, other factors can influence important parameters estimates. For example, the variations in the repeatability estimates across different methods and scenarios reflected the scale used in the analysis. While the GLMM and HBM presented similar magnitudes due to both using the probit scale, the LMM assumed a normal scale, and the BCLMM assumed a scale derived from its lambda, thus resulting in different magnitudes. The high-repeatability estimates for the BCLMM in scenarios 1 and 2 should be interpreted cautiously. Besbes et al. (1993) also reported an increase in heritability when Box–Cox transformations were applied to transform data describing egg production in laying hens. The high magnitudes presented here become questionable when the model adjustments between methods are compared against each other. The BCLMM under scenarios 1 and 2 showed noticeably high deviance (data not shown), reinforcing the idea that data transformation can lead to spurious results (Jaeger, 2008; Bolker et al., 2009; Stroup, 2015).

The correlations between the BLUPs of biomass yields and the BLUPs of vigor ratings showed a strong positive relationship, supporting the efficiency of indirect selection (Figures 3, 4, and 5). Moreover, these positive relationships increased in scenarios 2 and 3 for all methods. A deeper analysis of the BCLMM scatterplot in scenario 1 (Figure 1) reveals the discretization of its BLUPs, which corresponds to the same number of categories (i.e., five) used for visually assessing genotypes. However, this discretization dissolved in the subsequent scenarios as the number of graders increased, leading to an approximation of the normal distribution of the vigor ratings, as suggested by the central limit theorem (Mood et al., 1974). Nevertheless, the approximation derived from the Box–Cox transformation did not assure normality in any scenario; thus, its application should be carried out with caution. Alternatively, the LMM provided comparable results when contrasted with the more flexible models (i.e., GLMM and HBM). This is likely due to the large number of genotypes in this trial (2004 unreplicated genotypes plus two checks), indicating that the central limit theorem is a robust principle and can support the analysis of vigor ratings via LMM even though some assumptions are violated. Furthermore, the LMM is simpler and less computationally demanding. However, it is expected that the GLMM and HBM provide better estimates when trials investigate fewer genotypes.

Although the correlations between methods followed similar positive results, the CI estimates showed that BCLMM did not identify the top 5% high-performing ruzigrass genotypes. Therefore, the simulated increase in biomass derived from that analysis is likely due to random sampling. However, the LMM, GLMM, and HBM generated a significant response to selection (~37%) when compared to the response to selection on based on biomass per se (~50%). The biomass increases presented in this study were derived from simulations and should not be compared with actual selection responses documented elsewhere. These results were used to compare the effects of different methods to analyze vigor ratings and assess the benefits of using grade scales to evaluate forage biomass.

Although different models may provide similar responses to selection, vigor rating adjustments via generalized models are more appropriate for statistical judgment (Jaeger, 2008; Bolker et al., 2009; Stroup, 2015). Generalized models are more flexible because of their ability to model response variables via distributions that share similar properties unified by the exponential family (Nelder & Wedderburn, 1972). However, because of their intrinsic complexities, such models are often overlooked in plant breeding. This study demonstrated the application of GLMM and HBM by incorporating a multinomial distribution and a probit link function via packages available in R (R Core Team, 2020). Compared with other distributions, multinomial distributions better describe ordinal categorical variables (e.g., vigor ratings) (Jansen, 1991; Agresti, 2002; Agresti, 2007) and enable threshold estimates between categories (Agresti, 2002; Gianola & Fernando, 1986); enabling results to be reported on their original scale. Falconer and Mackay (1996) highlighted the benefits of using a threshold model to classify genotypes according to their genetic merits. As shown in Figure 6, there was an approximately 80% chance of finding a genotype in categories 2 and 3, while this probability was only 1% for category 5 (red graph). After simulating the selection of the top 5% of the BLUPs in terms of vigor ratings, the probabilities of finding genotypes in categories 4 (40.04%) and 5 (8.45%) increased (blue graph). Thus, selecting genotypes with higher ratings might increase the probability of finding genotypes in higher categories in future breeding cycles.

Conclusion

There are statistical alternatives for analyzing vigor ratings, including linear mixed models and generalized linear mixed models. However, some models can generate spurious results and should be used cautiously. The number of graders can affect decisions and ultimately limit the response to selection. Although there were no marked differences among the LMM, GLMM, and HBM, generalized models are most appropriate and informative for reporting vigor ratings due to their flexibility in incorporating distributions that effectively describe ordinal categorical scales. Vigor ratings can be used to select the best ruzigrass genotypes to optimize biomass yields during the early stages of a breeding program.

Data availability

There is no data to provide.

Acknowledgements

References

Atkins, R. E. (1964). Visual selection for grain yield in barley. *Crop Science*, *4*(5), 494-497. https://doi.org/10.2135/cropsci1964.0011183X000400050018x

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). John Wiley & Sons.

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). John Wiley & Sons.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Besbes, B., Ducrocq, V., Foulley, J.-L., Protais, M., Tavernier, A., Tixierboichard, M., & Beautnont, C. (1993). Box-Cox transformation of egg-production traits of laying hens to improve genetic parameter estimation and breeding evaluation. *Livestock Production Science*, *33*(3/4), 313-326. https://doi.org/10.1016/0301-6226(93)90010-F

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology Evolution*, *24*(3), 127-135. https://doi.org/10.1016/j.tree.2008.10.008

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, *26*(2), 211-243. https://doi.org/10.1111/J.2517-6161.1964.TB00553.X

Burton, G. W. (1982). Improved recurrent restricted phenotypic selection increases Bahiagrass forage yields. *Crop Science*, *22*(5), 1058–1061. https://doi.org/10.2135/cropsci1982.0011183x002200050040x

Christensen, R. H. B. (2019). *Ordinal - Regression models for ordinal data*. R package version 2019.12-10.

Corrêa, F. M., Silva, J. W., Ferreira, D. F., & Bueno Filho, J. S. S. (2016). Bayesian algorithms for analysis of categorical ordinal data. *Brazilian Journal of Biometrics*, *34*(4), 597-620.

Cullis, B. R., Smith, A. B., & Coombes, N. E. (2006). On the design of early generation variety trials with correlated data. *Journal of Agricultural Biological and Environmental Statistics*, *11*, 381-393. https://doi.org/10.1198/108571106X154443

Dias, J. A., Rosado, L. R., Benites, F., Souza Sobrinho, F., Nunes, J. A. R., & Gonçalves, F. M. A. (2022). Efficiency of indirect selection for green biomass production of *Urochloa ruziziensis*. *Crop Breeding and Applied Biotechnology*, *22*(2), 1-8. https://doi.org/10.1590/1984-70332022v22n2a22

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (4th ed.). Longman.

Fonseca, J. M. O., Nunes, J. A. R., Gonçalves, F. M. A., Souza Sobrinho, F., Benites, F. R. G., & Teixeira, D. H. L. (2020). Predictive approach to optimize the number of visual graders for indirect selection of high-yielding *Urochloa ruziziensis* genotypes. *Crop Breeding and Applied Biotechnology*, *20*(3), 1-7. https://doi.org/10.1590/1984-70332020v20n3a48

Gianola, D., & Fernando, R. L. (1986). Bayesian methods in animal breeding Theory. *Journal of Animal Science*, *63*(1), 217-244. https://doi.org/10.2527/jas1986.631217x

Gianola, D., & Foulley, J. (1983). Sire evaluation for ordered categorical data with a threshold model. *Genetics Selection Evolution*, *15*(2), 201-224. https://doi.org/10.1186/1297-9686-15-2-201

Gouveia, B. T., Mateus, R. G., Barrios, S. C. L., Valle, C. B., Bueno Filho, J. S. S., Rios, E. F., Dias, A. M., & Nunes, J. A. R. (2022). Combining ability and selection for agronomic and nutritional traits in *Urochloa* spp. hybrids. *Grass and Forage Science*, *77*(1), 33-44. https://doi.org/10.1111/gfs.12555

Gouy, M., Rousselle, Y., Bastianelli, D., Lecomte, P., Bonnal, L., Roques, D., Efile, J.-C., Rocher, S., Daugrois, J., Toubi , L., Nabeneza , S., Hervouet, C., Telismart, T., Denis, M., Thong-Chane, A., Glaszmann, J. C., Hoarau, J.-Y., Nibouche, S., & Costet, L. (2013). Experimental assessment of the accuracy of genomic selection in sugarcane. *Theoretical and Applied Genetics*, *126*, 2575-2586. https://doi.org/10.1007/s00122-013-2156-z

Hadfield, J. D. (2010). MCMC Methods for multi-response generalized linear mixed models: The MCMCglmm R Package. *Journal of Statistical Software*, *33*(2), 1-22. https://doi.org/10.18637/jss.v033.i02

Hamblin, J., & Zimmermann, M. J. O. (1986). Breeding common bean for yield in mixtures. In J. Janick (Ed.), *Plant Breeding Reviews* (v. 4, pp. 245-272). Wiley.

Henderson, C. R. (1984). *Applications of linear models in animal breeding*. University of Guelph.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434-446. https://doi.org/10.1016/j.jml.2007.11.007

Jansen, J. (1991). Fitting regression models to ordinal data. *Biometrical Journal*, *33*(7), 807-815. https://doi.org/10.1002/bimj.4710330707

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1-26. https://doi.org/10.18637/jss.v082.i13

Marcelino, L. L., Moreira, G. R., Souza Sobrinho, F., Almeida, M. I. V., Cóser, A. C., Cunha, G. M., & Benites, F. R. G. (2020). Nutritive value of improved populations *Brachiaria ruziziensis*. *Semina: Ciências Agrárias*, *41*(1), 323-334. https://doi.org/10.5433/1679-0359.2020v41n1p323

Marcon, F., Brugnoli, E. A., Nunes, J. A. R., Gutierrez, V. A., Martínez, E. J., & Acuña, C. A. (2021). Evaluating general combining ability for agromorphological traits in tetraploid bahiagrass. *Euphytica*, *217*, 1-11. https://doi.org/10.1007/s10681-021-02942-5

McCulloch, C. E., & Searle, S. R. (2004). *Generalized, linear, and mixed models*. John Wiley & Sons.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, *135*(3), 370-384. https://doi.org/10.2307/2344614

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics* (3rd ed.). McGraw Hill.

Nouhoun, Z., Traoré, T. C., Sawadogo, E. T. B. P., Ayantunde, A., Prasad, K. V. S. V., Blummel, M., Balehegn, M., Rios, E., Dubeux, J. C., Boote, K. J., & Adesogan, A. T. (2022). Herbage accumulation and nutritive value of cultivar Mulato II, Congo grass, and Guinea grass cultivar C1 in a subhumid zone of West Africa. *Agronomy Journal*, *114*(1), 138-147. https://doi.org/10.1002/agj2.20861

Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, *58*(3), 545-554. https://doi.org/10.1093/biomet/58.3.545

Pereira, A. V., Souza Sobrinho, F., Valle, C. B., Lédo, F. J. S., Botrel, M. A., Oliveira, J. S., & Xavier, D. F. (2005) Selection of interspecific *Brachiaria* hybrids to intensify milk production on pastures. *Crop Breeding Applied Biotechnology*, *5*(1), 99-104. https://doi.org/10.12702/1984-7033.v05n01a13

Pinheiro, J. C., & Chao, E. C. (2006). Efficient laplacian and adaptive gaussian quadrature algorithms for multilevel generalized linear mixed models. *Journal of Computational and Graphical Statistics*, *15*(1), 58-81. https://doi.org/10.1198/106186006X96962

Price, D. L., & Casler, M. D. (2014). Divergent selection for secondary traits in upland tetraploid switchgrass and effects on sward biomass yield. *Bioenergy Research*, *7*, 329-337. https://doi.org/10.1007/s12155-013-9374-8

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Riday, H. (2009). Correlations between visual biomass scores and forage yield in space planted red clover (*Trifolium pratense* L.) breeding nurseries. *Euphytica*, *170*(3), 339-345. https://doi.org/10.1007/s10681-009-9991-7

Schmidt, P., Hartung, J., Bennewitz, J., & Piepho, H.-P. (2019). Heritability in plant breeding on a genotype-difference basis. *Genetics*, *212*(4), 991-1008. https://doi.org/10.1534/genetics.119.302134

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (Complete Samples). *Biometrika*, *52*(3/4), 591-611. https://doi.org/10.2307/2333709

Silva, D. M., Moraes, J. C., Auad, A. M., Fonseca, M. G., & Souza Sobrinho, F. (2013). Genetic variability of *Brachiaria ruziziensis* clones to *Collaria oleosa* (Hemiptera: Miridae) based on leaf injuries. *American Journal of Plant Sciences*, *4*(12), 2418-2424. https://doi.org/10.4236/ajps.2013.412300

Santos, H. G., Jacomine, P. K. T., Anjos, L. H. C., Oliveira, V. A., Oliveira, J. B., Coelho, M. R., ... Cunha, T. J. F. (2006). *Sistema brasileiro de classificação de solos*. Embrapa.

Souza Sobrinho, F., Auad, A. M., & Lédo, F. J. S. (2010). Genetic variability in *Brachiaria ruziziensis* for resistance to spittlebugs. *Crop Breeding and Applied Biotechnology*, *10*(1), 83-88.

Stroup, W. W. (2015). Rethinking the analysis of non-normal data in plant and soil science. *Agronomy Journal*, *107*(2), 811-827. https://doi.org/10.2134/agronj2013.0342

Teixeira, D. H. L., Gonçalves, F. M. A., Nunes, J. A. R., Souza Sobrinho, F., Benites, F. R. G., & Dias, K. O. G. (2020). Visual selection of *Urochloa ruziziensis* genotypes for green biomass yield. *Acta Scientiarum. Agronomy*, *42*(1), 1-8. https://doi.org/10.4025/actasciagron.v42i1.42444

Timbó, A. L. O., Souza, P. N. C., Pereira, R. C., Nunes, J. D., Pinto, J. E. B. P., Souza Sobrinho, F., & Davide, L. C. (2014). Obtaining tetraploid plants of *Urochloa ruziziensis* (*Brachiaria ruziziensis*). *Revista Brasileira Zootecnia*, *43*(3), 127-131. https://doi.org/10.1590/S1516-35982014000300004

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. Springer.

Whitman, B., Iannone, B. V., Kruse, J. K., Unruh, J. B., & Dale, A. G. (2022). Cultivar blends: A strategy for creating more resilient warm season turfgrass lawns. *Urban Ecosystems*, *25*, 797-810. https://doi.org/10.1007/s11252-021-01195-3