2-1-2021

Assessment of alternative models for genetic analysis of worm and tick infestation in nellore cattle

T. L. Passafaro
University of Wisconsin-Madison

F. B. Lopes
Cobb Vantress Incorporated

Thomas W. Murphy
USDA ARS Roman L. Hruska U.S. Meat Animal Research Center, tom.murphy@usda.gov

B. D. Valente
Genus plc

R. C. Leite
Universidade Federal de Minas Gerais

See next page for additional authors

Authors

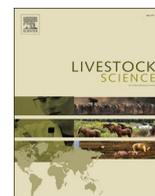T. L. Passafaro, F. B. Lopes, Thomas W. Murphy, B. D. Valente, R. C. Leite, G. J.M. Rosa, and F. L.B. Toral

Assessment of alternative models for genetic analysis of worm and tick infestation in nellore cattle☆

T.L. Passafaro [a], F.B. Lopes [b], T.W. Murphy [c], B.D. Valente [d], R.C. Leite [e], G.J.M. Rosa [a], F.L. B. Toral [e,*]

[a] Department of Animal Sciences, University of Wisconsin, 1675 Observatory Drive, 53706 Madison, WI, United States
[b] Cobb-Vantress Inc., Cobb, 4703 Highway 412 East, 72761 Siloam Springs, Arkansas, United States
[c] USDA, ARS, Genetics, Breeding, and Animal Health Research Unit, Roman L. Hruska U.S. Meat Animal Research Center, 68933 Clay Center, Nebraska, United States
[d] Pig Improvement Company, PIC, 100 Bluegrass Common Boulevard, 37075 Hendersonville, Tennessee, United States
[e] Departamento de Zootecnia, Escola de Veterinária, Universidade Federal de Minas Gerais, 6627 Avenida Presidente Antônio Carlos, São Luiz, 31270-901 Belo Horizonte, Minas Gerais, Brazil

HIGHLIGHTS

- Worms and ticks cause considerable economic and productive losses in beef cattle.
- The choice of statistical model can impact genetic analysis of parasite infestation.
- Heritability estimates were low to moderate for both worm and tick infestation.
- A better fit was observed for models employing non-Gaussian distributions.
- Genetic selection and parasite control practices should be carried out together.

ARTICLE INFO

ABSTRACT

Worms and ticks are important parasites in beef cattle, especially in tropical areas, causing significant economic and production losses. Understanding animal-to-animal variation on infestation for these parasites might guide genetic selection and improvement of management practices to attenuate its detrimental effects. Statistical models used to analyze such traits usually assume a Gaussian distribution for the observed data. However, this assumption is quite often inappropriate for counting data. Therefore, the objectives of this study were: 1) Estimate genetic parameters for worms and tick infestations in Nellore cattle, and 2) To compare the overall performance of six data analysis approaches for worm and tick infestation in Nellore cattle, using different specifications of generalized linear mixed models (GLMM) and response variables. Data consisted of presence/absence of parasites as well as counting observations for both worms and ticks in a Nellore herd in Brazil. The binary data were analysed with both Gaussian and Threshold models, whereas the counting data were studied using Gaussian models on the original and logarithmic scales, as well as Poisson and Zero-Inflated Poisson (ZIP) models. All models included the systematic effects of contemporary group and age, as well as the random additive genetic and residual effects. Models were compared using four criteria: Deviance Information Criterion (DIC), Spearman's correlation between predicted breeding values from different models, the agreement on the 5 and 50% top-ranked animals across models, and the Mean Squared Error of Prediction (MSEP) assessed via

Monte Carlo Cross-Validation (MCCV). The MCCV was performed using parallel computing through the Center for High Throughput Computing (CHTC) at the University of Wisconsin-Madison. The estimates of heritability ranged from 0.15 to 0.40 for worms and from 0.08 to 0.25 for ticks. According to the DIC, non-Gaussian models displayed the best goodness of fit compared to Gaussian models. DIC's results excluded Gaussian models on the logarithmic scale because fairer comparisons involve phenotypes on the same scale. Spearman's correlation and the percentage of agreement on the 5% and 50% top-ranked animals suggested some re-ranking of animals depending upon the model used. Monte Carlos Cross-Validation showed that all models presented similar MSEP with average values of 0.20 (binary data; worms), 0.18 (binary data; ticks), 15.69 (count data; worms), and 14.19 (count data; ticks). Moreover, performing MCCV in parallel had the benefit to deliver results for all models in about 2 days. Heritability estimates indicate that the selection of high merit animals for worms and tick infestation is possible feasible and can potentially contribute to the genetic progress. Furthermore, genetic selection should be performed concomitantly with traditional parasite control approaches. Overall, non-Gaussian models seem to be better suitable for genetic analysis of worm and tick infestation in beef cattle, because such models have lower DIC values with similar predictive performance compared to Gaussian models.

1. Introduction

Worms and ticks are responsible for significant production losses in beef and dairy cattle, causing severe economic impact to the livestock industry around the world. Annual global costs associated with cattle ticks range from 20 to 30 billion USD in 2016 (Lew-Tabor and Rodriguez Valle 2016), and such costs may be underestimated for the actual economic conjecture. Rashid et al. (2019) estimated financial costs of 50.67 USD animal^{-1} year^{-1} with the control, treatment and mortality related to worm infestations. Moreover, Grisi et al. (2014) assessed the annual cost of potential losses in dairy and beef cattle in Brazil and showed that parasites are one of the most important issues related to economic losses in dairy and beef cattle. According to these authors, potential losses are close to 7.11 and 3.24 billion USD per year for worms and ticks, respectively.

Parasite infections are typically controlled by chemical treatment. However, incorrect applications of such compounds might lead to parasite resistance (Fernández-Salas et al., 2012; De Graef et al., 2013), residues in milk and meat (Wolstenholme et al., 2004), and increase in production costs (Grisi et al., 2014). For these reasons, alternative methods to tackle parasite infestations might be beneficial, as they would allow a reduction in the use of chemicals and their associated problems. An alternative in this regard is the genetic selection for resistant animals. For instance, Frisch et al. (2000) showed a decrease in tick counts from 275 to 40 ticks/animal after fifteen years of selection in a population of Australian Belmont Adaptaur. Morris et al. (2005) reported a difference of 343 eggs gram^{-1} in the average breeding value of two lines of sheep divergently selected for fecal egg counts. These results support selection as an effective tool to decrease the number of parasites in dairy and beef cattle. Furthermore, genetic selection for low parasite loads offers an important advantage over conventional methods in that the improvement is expected to be more stable in a long term. According to Bishop (2012), parasite evolution takes more time to adapt to host resistance than chemical treatment. This author pointed out that host resistance is naturally polygenic and parasites would have to evolve against several genes, while for drug treatment only a few compounds must be suppressed.

Parasites infestation is often assessed by indicator traits such as the recorded number of ticks on the animal or the number of worm eggs present in feces. Thus, the resulting phenotype is either zero or a positive integer. Statistical models used for genetic evaluation and inference on genetic parameters for these traits frequently assume the Gaussian distribution of either raw or transformed data (Frisch et al., 2000; Burrow, 2001; Prayaga, 2003). Linear Gaussian models may offer only a poor approximation for count data as their sample space consists of non-negative integers only, and also the distribution of count variables is often right-skewed (Stroup, 2015; Villemereuil et al., 2016; Sae-Lim et al., 2017). An appropriate data transformation may help with the Gaussian approximation, but it still poses some challenges such as how to deal with zeros when using log transforms. Moreover, when using

transformations, the variance components and its respective genetic parameters are in a different scale from the original data which may bring some challenges in terms of a practical breeding . For instance, the back-transformation of estimate breeding values in a square root scale will generate only positive values which poses a challenge to select individuals in the original scale. Alternatively, generalized linear mixed models (GLMM) are theoretically more suitable to study non-Gaussian response variables, more specifically distributions that belong to the exponential family such as Bernoulli and Poisson (Nelder and Wedderburn, 1972; Tempelman, 1998). Accordingly, they have been applied to genetic analysis of worm and tick infestation in cattle using, for example, binary threshold models (Passafaro et al., 2015), Poisson models (Ayres et al., 2013), and Zero-Inflated Poisson models (Silva et al., 2011).

Although different models have been proposed for genetic evaluation of worm and tick infestations, there is no consensus on which one is the most appropriate for the prediction of future performance and identification of animals with the highest genetic merit. Choosing an appropriate model is crucial for breeding programs as it may improve selection decisions and enhance genetic gains, magnifying the positive impact on the aforementioned issues related to parasite infection and chemical treatment. The assessment of future performance can be investigated by cross-validation methods that indicates the model's ability to predict animal's genetic merit. Therefore, the objectives of this study were: 1) Estimate genetic parameters for worms and tick infestations in Nellore cattle, and 2) to compare the overall performance of six data analysis approaches for worm and tick infestation in Nellore cattle, using different specifications of generalized linear mixed models (GLMM) and response variables.

2. Material and methods

All experimental procedures used in the current study were approved by the Ethics and Animal Experimentation Committee of the Universidade Federal de Minas Gerais (Protocol 255/2010).

2.1. Dataset

The data were collected from April 2011 to June 2014 at the Mundo Novo Farm which is located in Uberaba, Minas Gerais, Brazil (19°24′33.3″S and 48°06′34.5″W). A description of this farm including climatic conditions, land area, nutrition management, health routines, and reproduction practices is provided in Passafaro et al. (2015). The animals used in this study were purebred Nellore cattle registered at the Brazilian Association of Zebu Breeders (ABCZ). Tick counts were recorded in naturally infected animals according to the method proposed by Wharton and Utech (1970), which consists of counting engorged female ticks $\geq$ 4 mm on one side of the animal. Fecal samples were collected from the rectum of the animals and used to count the number of worm eggs following the protocol of the modified MacMaster technique described by Ueno and Gonçalves (1998). To perform the

worm egg counts, 2 g of faeces were diluted with 28 ml of drinking water. Then, after sifting the mixture, a 2 ml aliquot was mixed with 2 ml of saturated Sheater's solution (500 g of sugar, 6.5 ml of phenol and 360 ml of water). The McMaster chamber, consisting of two slides separated by a 1.5 mm space, between which there are two count chambers of 1 cm^2, was filled with a 0.15 ml aliquot of the final solution and thereafter the worm egg counts were performed using a light microscope at a 10 x magnification. Worm egg and tick counts used in the present study are the actual counts observed at the McMaster chamber and at one side of the animal, respectively, without any multiplication by a specific constant. The final data included animals that were 500 to 597 days of age and had phenotypic information on both worm and tick loads. Contemporary groups were defined by the combination of sex, lot of management, and day of collection. Cohort groups with less than 5 records and/or no phenotypic variation were excluded for both traits. Furthermore, animals that received any kind of parasite treatment within 120 days prior to the day of collection were excluded from the data. After data editing, there remained 1665 animals have observations recorded for both worm eggs and tick counts. The pedigree file consisted of 5697 animals, including 601 and 3524 sires and dams, respectively. The pedigree was traced back to 17 generations, and this population has an average inbreeding of 0.033 with a maximum value of 0.312. A more detailed information for the whole population structure can be found in Oliveira et al. (2011).

2.2. Statistical models

Six statistical models were proposed to evaluate worm eggs and tick loads, including two specifications of the response variable for each parasite: presence/absence and the actual counting. The binary data (i.e. presence/absence of parasites) were studied with both Gaussian and threshold models, the later with two response categories. The actual counts of parasites (worms or ticks) were analyzed with Gaussian models on either the original or logarithmic scales, Poisson models, and Zero-Inflated Poisson (ZIP) models. The linear predictor function for all models can be expressed in matrix notation as follows:

$$\eta = X\beta + Zu$$

where η is the vector of the linear predictor associated with the 1665 observations; β is the vector containing the systematic effects including the overall intercept, the contemporary groups, and the linear covariate of age on the day of recording; u is the vector of additive genetic effects; X and Z are the incidence matrices of the systematic and additive genetic effects, respectively. The vector $u = [u_1 \ u_2 \ ... \ u_{5697}]^T$ of additive genetic effects was assumed to follow a normal distribution $u \sim N(0, A\sigma_u^2)$, where σ_u^2 is the additive genetic variance and A is the numerator relationship matrix related to the 5697 animals in the pedigree file. A short description of specific features of each model is provided in the following sub-sections.

2.2.1. Threshold model

The threshold model assumes that categorical traits are continuous and normally distributed on an underlying liability scale, with the observable sampling space divided into $(k + 1)$ intervals by k unobservable thresholds. These thresholds are displayed in increasing order and must satisfy the following condition — $\infty < t_0 < t_1 < t_2 < ... < t_k < \infty$. The number of intervals is equal to the number of categories, and the observed category for an individual depends on which interval contains its liability value. Similar to Gianola and Foulley (1983), the underlying linear model is $l = \eta + e$, where e represent a vector of independent and identically distributed standard normal random variables. A probit link function was considered for this analysis, and the residual variance was fixed at one to cope with identifiability issues. Worm eggs and tick count $\{w_i\}$ were classified as:

$$w_i = (1 \ if \ y_i > 0; \ 0 \ otherwise)$$

Thus, the density function of the conditional probability of presence/absence of parasites is:

$$P(w_i = 1|\eta_i) = \Phi(\eta_i)$$

where $\Phi(.)$ is the standard normal cumulative distribution function. Accordingly, the likelihood function is proportional to:

$$p(\mathbf{w}|\boldsymbol{\eta}) = \prod_{i=1}^{1665} \Phi(\eta_i)^{w_i} (1 - \Phi(\eta_i))^{1-w_i}$$

where $\mathbf{w} = (w_i)$ and $\boldsymbol{\eta} = (\eta_i)$ are the vectors of binary outcomes and linear predictors, respectively.

2.2.2. Gaussian model

Linear Gaussian models were also applied to the presence/absence of parasites (w_i), as well as to the raw counting data (y_i). In addition, the counting scores of worms and ticks were log-transformed as $z_i = log(y_i + 0.1)$. We tested different constants to add in the logarithm transformation (i.e., 0.1, 1, and 10). None of these constants were able to normalized the data, and the constant 0.1 was select because the variance components were more closed to the other methods. Such Gaussian models can be expressed as $t_i = \eta_i + e_i$, where t_i is either the binary, raw or log-transformed record, and e_i are independent and identically distributed residual terms with mean zero and variance σ_e^2. Therefore, the likelihood function becomes proportional to:

$$P(t_i|\eta_i) = \prod_{i=1}^{1665} N(t_i|\eta_i, \sigma_e^2)$$

2.2.3. Poisson model

A Poisson model was used to fit the recorded number of parasites, i.e. the counts (y_i). This model contains a single parameter λ expressing both the mean and the variance of the random variable. Such dependence between variance and mean makes overdispersion a common challenge for Poisson models. An approach to address this problem is by introducing a residual term that accommodates overdispersion. Therefore, the mixed Poisson model assumed that the recorded number of parasites, given the systematic and random effects, followed a Poisson distribution with parameter λ_i, where $log(\lambda_i) = \eta_i + e_i$, and e_i are independent and identically distributed Gaussian random variables with mean zero and variance σ_e^2. Thus, the likelihood function, in this case, is proportional to:

$$P(\mathbf{y}|\lambda) = \prod_{i=1}^{1665} \frac{\lambda_i^{y_i} exp(-\lambda_i)}{y_i!}$$

where $\mathbf{y} = (y_1, y_2, ..., y_n)$ and $\lambda = (\lambda_1, \lambda_2, ..., \lambda_n)$ are the vectors of counting outcomes, and the parameters of the Poisson model, respectively.

2.2.4. Zero-Inflated poisson model

A mixed effect Zero-Inflated Poisson (ZIP) model was also applied to study the recorded number of parasites (Lambert 1992; Rodrigues-Motta et al., 2007). The ZIP model is defined as a mixture distribution with two components. The first component expresses the "perfect state" which only produces zeros, with probability p_i, where $logit(p_i) = \eta_i + \varepsilon_i$, and ε_i are independent and identically distributed Gaussian random variables with mean zero and variance σ_ε^2. The "perfect state" represents the probability of an animal being resistant to parasites (no parasites at all) and it is useful to accommodate an excess of zeros in the dataset (Lambert 1992; Rodrigues-Motta et al., 2007; Naya et al., 2008). According to these authors, the second component of the ZIP model represents the "imperfect state" which is sampled with a probability $(1 - p_i)$ from a Poisson distribution with parameter λ_i, where $log(\lambda_i) =$

$\eta_i + e_i$, and e_i are independent and identically distributed Gaussian random variable with mean zero and variance $\sigma^2_{e_i}$. Therefore, the probability function of ZIP model is described as:

$$p(y_i|p_i, \lambda_i) = [\, p_i + (1-p_i)e^{-\lambda_i}\,]^{\,I(y_i=0)}[(1-p_i) \times (e^{-\lambda_i}\lambda_i^{y_i}/y_i!)]^{\,I(y_i>0)}$$

where y_i is the observed record, and $I(.)$ is an indicator function. Notice that the ZIP probability density function described above converts to a Poisson distribution whenever $p_i = 0$. The likelihood function of the ZIP model is proportional to:

$$p(\mathbf{y}|\mathbf{p}, \lambda) = \prod_{I(y_i=0)} [p_i + (1-p_i)e^{-\lambda_i}] \prod_{I(y_i>0)} [(1-p_i) \times (e^{-\lambda_i}\lambda_i^{y_i}/y_i!)]$$

where $\mathbf{y} = (y_i)$, $\mathbf{p} = (p_i)$ and $\lambda = (\lambda_i)$ are the vectors of counting outcomes, and the parameters of ZIP model, respectively.

Although the parameter p of the ZIP model can be fitted for each animal in the pedigree, in our study this parameter was assumed to be homogeneous in the population because of MCMC convergence issues. In this setting, p is the same for all animals and represents the percentage of the population that are in the perfect state. The λ component was estimated for all animals in the pedigree using a hierarchical Bayesian model, which can be represented as follows:

$$log(\lambda) = \mathbf{X}_\lambda \boldsymbol{\beta}_\lambda + \mathbf{Z}_\lambda \mathbf{u}_\lambda + \mathbf{e}_\lambda$$

where $\mathbf{X}_\lambda$ is the incidence matrix related to the systematic effects $\boldsymbol{\beta}_\lambda$, with dimensions (1665 × 52), $\mathbf{Z}_\lambda$ is the incidence matrix related to the additive random effects $\mathbf{u}_\lambda$, with dimensions (1665 × 5697), and $\mathbf{e}_\lambda$ is the vector of residual effects. The additive random effects were assumed to follow a normal distribution $\mathbf{u}_\lambda \sim N(0, \mathbf{A}\sigma^2_{u_\lambda})$, $\mathbf{A}$ is a known matrix of additive genetic relationships, and $\sigma^2_{u_\lambda}$ is the additive genetic variance related to λ. The residual effects were fit as a normal distribution $\mathbf{e}_\lambda \sim N(0, \mathbf{I_n}\sigma^2_{u_\lambda})$, $\mathbf{I_n}$ is an identity matrix of order (1665 × 1665), and $\sigma^2_{e_\lambda}$ is the residual variances related to λ.

2.3. Bayesian inference and MCMC

Statistical analyses were implemented under a Bayesian framework, for which prior distributions are specified for all unknown parameters. In all analyses, flat prior distributions were assigned for the systematic effects. The prior distribution for $\mathbf{u}$, as described before, was normal, given by $N(\mathbf{u}|0, \mathbf{A}\sigma^2_{\mathbf{u}})$. For the variance components $\sigma^2_{\mathbf{u}}$ and $\sigma^2_{\mathbf{e}}$, scaled inverse Chi-square distributions were assumed, i.e. $\chi^{-2}(\sigma^2_{\mathbf{u}}|V_{\mathbf{u}}, S_{\mathbf{u}})$ and $\chi^{-2}(\sigma^2_{\mathbf{e}}|V_{\mathbf{e}}, S_{\mathbf{e}})$, respectively, where V and S are the corresponding degrees of freedom and scale parameters. These parameters were the same for all models with values equal to $1/3$ for V and 1 for S. For the ZIP model, the parameter $\mathbf{p}$ was assigned a flat distribution.

Statistical analyses were implemented using the R package MCMCglmm (Hadfield, 2010). A chain of 1000,000 iterations was generated for each model. The first 200,000 iterations were discarded as burn-in and the remaining were sampled every 200 iterations. Therefore, 4000 samples were retained for statistical inference. Convergence was assessed by visual inspection of the trace plots and by the Geweke diagnostic (Geweke, 1992). A Gibbs sampler algorithm was considered for all Gaussian models, while the Metropolis-Hastings method was performed for threshold, Poisson, and ZIP models.

2.4. Model comparison

Statistical models were compared based on the following criteria:

2.4.1. Deviance information criterion (DIC)
The DIC was proposed by Spiegelhalter et al. (2002) as follows:

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

where $\overline{D}(\boldsymbol{\theta}) = E_{\theta|y}[D(\boldsymbol{\theta})]$ is the posterior expectation of the Bayesian deviance; and $D(\overline{\theta}) = 2log(\mathbf{y}|\boldsymbol{\theta})$ represents the goodness of fit of the model. The DIC penalizes model complexity based on the effective number of parameters $p_D = \overline{D}(\boldsymbol{\theta}) - D(\overline{\theta})$, where $\boldsymbol{\theta}$ is the vector of model parameters and $D(\overline{\theta})$ is the Bayesian deviance, as a function of the posterior mean of these parameters. Lower values of DIC are preferred, as they suggest a better model adjustment with a lower degree of complexity. DIC was used to compare models where response variables are on the same scale. Therefore, it was used only for model comparisons within each specification of response variable: binary outcome (Threshold vs. Gaussian), and counting data (Gaussian vs. Poisson vs. ZIP). Comparisons with the Gaussian model on the logarithmic scale were not considered.

2.4.2. Spearman rank correlations and the agreement on the 5 and 50% top-ranked animals
Spearman rank correlations of the posterior mean of breeding values were used to assess the agreement between models. The percentage of the agreement on the 5% and 50% top-ranked animals were used to investigate the difference in the proportion of the same individuals selected among different statistical models. These percentages were defined based on the replacement rate of sires and dams considered in the farm, respectively.

2.4.3. Predictive ability via monte-carlo cross-validation
The predictive ability of each model was assessed via Monte-Carlo Cross-Validation where the data was randomly split into two subsets that allocated approximately 80% and 20% of the observations to the training and testing sets, respectively. This procedure had the constraint that all records from contemporary groups with less than 36 observations were placed in the training set. The remaining cohort groups had their records split into the training and testing set. The Monte-Carlo Cross-Validation was repeated one hundred times, and for each iteration, a common folder was evaluated by all statistical models for both traits. Such analysis was performed in parallel, using computational resources available at University of Wisconsin-Madison through the Center for High Throughput Computing (CHTC). The loss function considered to assess predictive ability was the mean squared error of prediction (MSEP), which is given by:

$$\text{MSEP} = \frac{4}{n} \sum_{i=1}^{n} (t_i - \widehat{t})^2$$

where t_i is the response variable computed as the counting variable or as a binary outcome, $\widehat{t_i}$ is the mean of the conditional expectation function, and n is the number of observations in the testing set. The conditional expectation function for the threshold model (binary outcome), Gaussian model, Poisson model, and ZIP model corresponds to $\widehat{w}_i = \Phi(\widehat{\eta}_i)$, $\widehat{y}_i = \widehat{\eta}_i$, $\widehat{y}_i = exp(\widehat{\eta}_i)$, and $\widehat{y}_i = (1 - \widehat{p}_i)exp(\widehat{\lambda}_i)$, respectively. The Gaussian model on the logarithmic scale was back-transformed to original scale by $\widehat{y}_i = exp(\widehat{\eta}_i) - 0.1$, before computing the MSEP. Lower values of MSEP are preferred and indicate a better predictive ability. Comparison of statistical models using MSEP is only applied for models that the response variables are on the same scale (Vazquez et al., 2009).

3. Results

3.1. Descriptive statistics and heritability estimates

The mean of the actual worm eggs and tick counts were 3.01 (± 4.68) worm eggs animal^{-1} and 2.50 (± 4.04) ticks animal^{-1}, respectively (Fig. 1). Forty and forty-three percent of the animals in our study had counts equal to zero for worm eggs and ticks, respectively. The skewness coefficient for worm eggs count was 2.84, while for tick count it was 3.03. As expected, these results suggest that the worm and tick counts
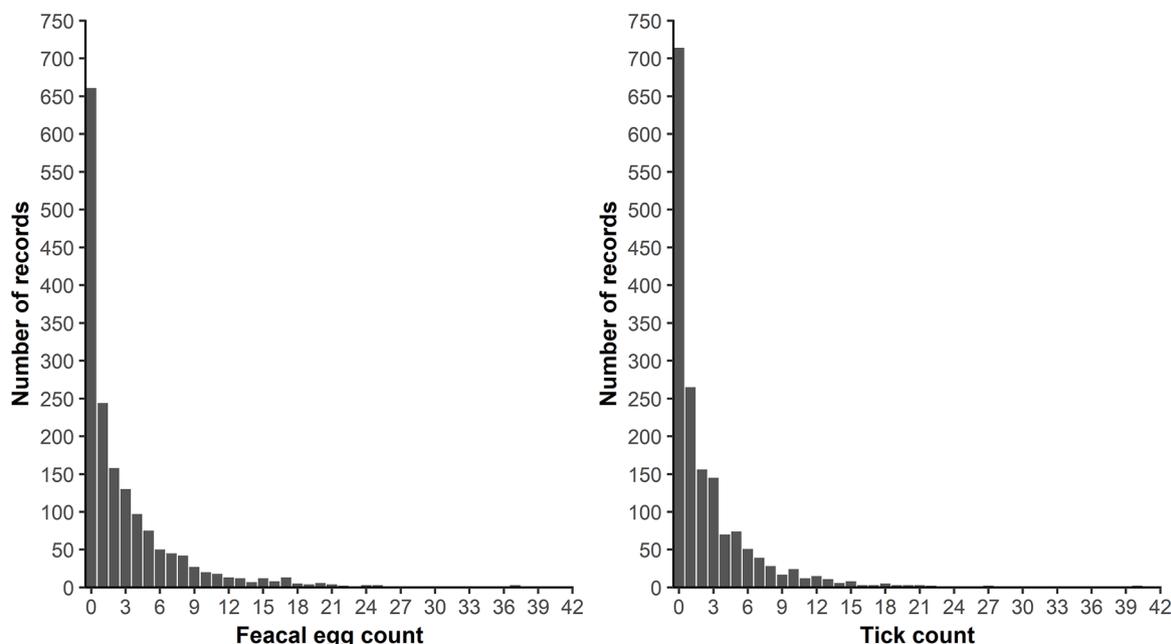
Fig. 1. Distribution of worm eggs (left) and tick count (right) in a Nellore beef cattle herd.

departure from a Gaussian distribution.

Posterior mean estimates of heritability (95% Highest Posterior Density) for worm infestations was 0.17 (0.09–0.25), 0.15 (0.05–0.25), 0.20 (0.09–0.33), 0.40 (0.21–0.62), and 0.20 (0.04–0.36) for the Gaussian model on binary scale, Gaussian model on original scale, Gaussian model on logarithm scale, Poisson model, and threshold model (Table 1), respectively. The posterior mean of heritability (95% Highest Posterior Density) for tick infestations was 0.18 (0.10–0.27), 0.08 (0.01–0.16), 0.12 (0.03–0.20), 0.21 (0.00–0.37), and 0.25 (0.09–0.41) for Gaussian model on binary scale, Gaussian model on original scale, Gaussian model on logarithm scale, Poisson model, and threshold model, respectively. Estimates of heritability were not obtained for the

ZIP model because there exists no such obvious estimator (Vazquez et al., 2012).

3.2. Model comparisons

The Poisson model showed the lowest DIC value (5753.93), followed by the ZIP model (5953.74) and Gaussian model on original scale (9452.78) for the recorded number of worm eggs. When treating phenotypes as binary for worm infestation, the threshold model displayed the lowest DIC (1920.31) followed by the Gaussian model (2004.37). Similarly, the Poisson model had the lowest value of DIC (5312.68), followed by the ZIP model (5478.18) and Gaussian model on original scale (8734.31) for the recorded number of ticks per animal. When treating phenotypes as binary for tick infestation, the threshold model had a lower DIC (1823.90) compared with the Gaussian model (1916.88).

Spearman rank correlations between the posterior mean of predicted breeding values among different models were all positive, with medium to high magnitude ranging from 0.61 to 0.99 and from 0.60 to 0.99 for worm egg and tick infestation (Table 2), respectively. The agreement on the 5% and 50% top-ranked animals ranged from 48.6% to 88.4% and from 64.7% to 92.9% for worm infestation, respectively (Table 3). Similarly, the agreement on the 5% and 50% top-ranked animals varied from 29.9% to 89.0% and from 70.7% to 93.4% for tick infestation, respectively. Overall, Spearman's correlation and the percentage of common selected animals were greater between models that treated

Table 1
Posterior mean (95% Highest Posterior Density intervals) of additive genetic variance (σ_a^2), residual variance (σ_e^2), and heritability (h^2) for genetic analysis of worm and tick infestation for Gaussian model on binary scale (BGAUS), Gaussian model on original scale (GAUS), Gaussian model on logarithm scale (LGAUS), Poisson model (POI), threshold model (THR), and Zero-Inflated Poisson model (ZIP).

Model	σ_a^2	σ_e^2	h^2
Worms			
[a]BINARY			
BGAUS	0.03 (0.02 - 0.05)	0.17 (0.15 - 0.18)	0.17 (0.09 - 0.25)
THR	0.26 (0.14 - 0.52)	1.00 (1.00 - 1.00)	0.20 (0.04 - 0.36)
[b]COUNT			
GAUS	2.53 (0.75 - 4.43)	14.81 (13.13 - 16.73)	0.15 (0.05 - 0.25)
LGAUS	0.54 (0.21 - 0.89)	2.13 (1.80 - 2.41)	0.20 (0.09 - 0.33)
POI	0.63 (0.29 - 0.99)	0.91 (0.61 - 1.21)	0.41 (0.21 - 0.62)
ZIP ([c]λ)	0.61 (0.28 - 0.99)	0.65 (0.34 - 0.97)	
Ticks			
[a]BINARY			
BGAUS	0.03 (0.02 - 0.05)	0.16 (0.14 - 0.17)	0.18 (0.10 - 0.27)
THR	0.34 (0.08 - 0.65)	1.00 (1.00 - 1.00)	0.25 (0.09 - 0.41)
[b]COUNT			
GAUS	0.82 (0.07 - 1.79)	10.15 (9.04 - 11.18)	0.07 (0.01 - 0.16)
LGAUS	0.25 (0.08 - 0.45)	1.93 (1.72 - 2.12)	0.12 (0.03 - 0.20)
POI	0.20 (0.00 - 0.37)	0.75 (0.55 - 0.94)	0.21 (0.00 - 0.37)
ZIP ([c]λ)	0.16 (0.00 - 0.32)	0.47 (0.29 - 0.64)	–

[a] BINARY is the presence/absence of parasites.
[b] COUNT is the recorded number of parasites.
[c] λ is the Poisson parameter of ZIP model.

Table 2
Spearman's rank correlation between predicted breeding values for genetic analysis of worm (below diagonal) and tick infestation (above diagonal) obtained with Gaussian model on binary scale (BGAUS), Gaussian model on original scale (GAUS), Gaussian model on logarithm scale (LGAUS), Poisson model (POI), threshold model (THR), and Zero-Inflated Poisson model (ZIP).

	THR	BGAUS	GAUS	LGAUS	POI	ZIP
THR	–	0.99	0.62	0.95	0.83	0.69
BGAUS	0.98	–	0.60	0.95	0.82	0.68
GAUS	0.64	0.61	–	0.79	0.89	0.90
LGAUS	0.94	0.94	0.81	–	0.95	0.85
POI	0.84	0.83	0.87	0.96	–	0.96
ZIP	0.77	0.76	0.87	0.92	0.99	–

Table 3

The agreement of the 5% (below diagonal) and 50% (above diagonal) top-ranked animals based on the predicted breeding values for genetic analysis of worm and tick infestations obtained with the Gaussian model on binary scale (BGAUS), Gaussian model on original scale (GAUS), Gaussian model on logarithm scale (LGAUS), Poisson model (POI), threshold model (THR), and Zero-Inflated Poisson model (ZIP).

Worms						
	THR	BGAUS	GAUS	LGAUS	POI	ZIP
THR	–	89.64	64.68	82.23	76.47	73.67
BGAUS	88.38	–	65.38	83.18	75.42	72.44
GAUS	49.30	48.59	–	77.18	79.46	79.25
LGAUS	80.63	80.99	62.68	–	88.06	84.52
POI	68.31	65.14	73.59	80.99	–	92.94
ZIP	61.62	58.10	76.41	72.18	88.38	–
Ticks						
	THR	BGAUS	GAUS	LGAUS	POI	ZIP
THR	–	93.36	72.54	90.20	81.36	74.32
BGAUS	89.08	–	70.72	89.04	80.16	72.82
GAUS	29.58	29.93	–	78.69	86.41	88.34
LGAUS	80.99	81.69	40.49	–	88.48	80.65
POI	59.86	60.56	55.28	74.65	–	90.62
ZIP	39.79	41.55	57.04	51.06	73.24	–

phenotypes in the same scale.

When phenotypes were treated as binary, the overall predictive performance of the Gaussian model was slightly better compared to the threshold model for either worm or tick infestation (Fig. 2). The Gaussian model on binary scale was the best to predict the absence of parasites, while the threshold model had a greater performance to evaluate the presence of parasites for both traits. When treating phenotypes as counting data, the ZIP model showed a slightly better overall performance for both traits, followed by the Gaussian model on the original scale, Poisson model, and Gaussian model on the logarithmic scale. The Gaussian model on the logarithmic scale presented the best-predicted values for the absence of parasites, while the Gaussian model on the original scale exhibited the highest performance when evaluating the presence of parasites for both traits.

4. Discussion

Overall, heritability estimates for worm and tick infestation was low to moderate, ranging from 0.08 to 0.40. These results are consistent with the literature, where the heritability estimates varied from 0.11 to 0.40 for worms and from 0.03 to 0.45 for ticks (Burrow, 2001; Prayaga and Henshall, 2005; Passafaro et al., 2015). The large range of heritability estimates may reflect differences in the statistical models used, population structure, type of infection (artificial vs. natural), and the season that records were collected. This result indicates that the phenotypic variance for worm and tick infestation is mainly accounted for by environmental factors, but that there is also an important genetic component for both infestations. Thus, genetic improvement through selective breeding is certainly possible. Nonetheless, the best practice may be to use genetic selection together with management practices (e. g., drug treatment, pasture rotation) to decrease parasite loads in a herd.

The DIC was used to compare models for which phenotypes were in the same scale, i.e. counts or binary. In our study, the Poisson model with a residual term displayed the lowest DIC compared to the other models evaluated. This result is most likely due to the fact that DIC penalizes models with more parameters. The ZIP model had more parameters compared to the Poisson model due to its binary component, partially justifying the observed result. Furthermore, Peñagaricano et al. (2011) and Ayres et al. (2013) reported that the addition of a residual term improved the performance of the Poisson model when compared to Gaussian models. Therefore, the inclusion of a residual term in the Poisson model may accommodate part of the overdispersion in the current study. Furthermore, in our study, the differences in DIC between non-Gaussian and Gaussian models were greater than 84. According to Spiegelhalter et al. (2002), differences in DIC greater than 7 are considered to be significant. Therefore, these results indicate that models which assume a non-Gaussian distributions showed a better goodness of fit than Gaussian models for genetic analysis of worm and tick infestation.

The results of Spearman's rank correlation and the agreement on the 5% and 50% of the top-ranked animals are suggestive of a re-ranking of
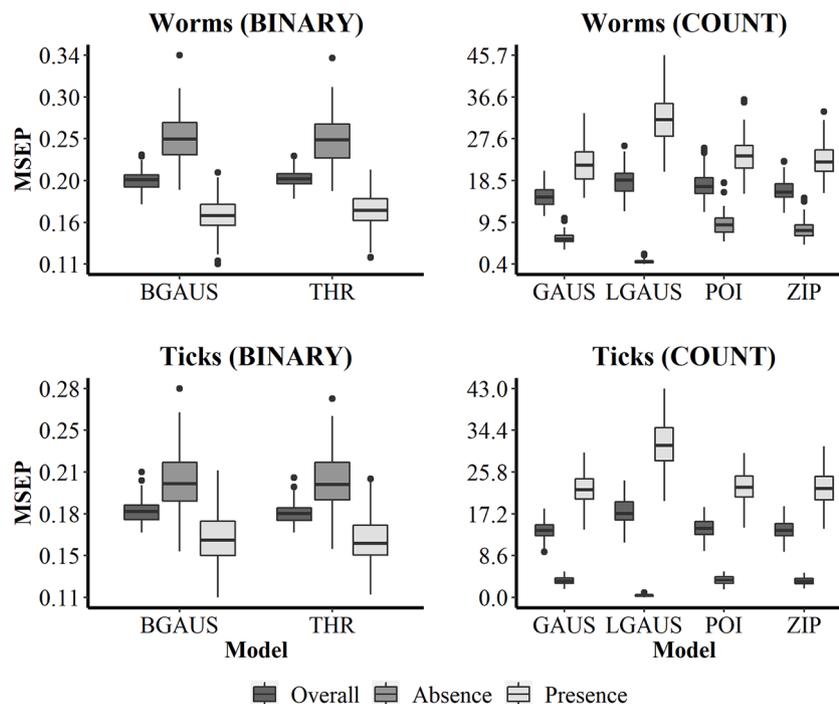

Fig. 2. Box plots of overall MSEP (dark grey), MSEP of absence of parasites (grey), and MSEP of presence of parasites (light grey) obtained in the cross-validation for Gaussian model on binary scale (BGAUS), Gaussian model on original scale (GAUS), Gaussian model on logarithmic scale (LGAUS), Poisson model (POI), threshold model (THR), and Zero-Inflated Poisson model (ZIP).

individuals depending upon the model used, especially between models for which the phenotypes were in different scales. On the other hand, models for which phenotypes are on the same scales tend to select the same individuals, leading to a similar genetic progress. Furthermore, sire replacement has a great impact in the selection gain over generations compared to dam replacement. Sires often produce more offspring during their productive life and the selection intensity in males is greater than in females. Therefore, the model used is an important factor for genetic improvement over time, especially when it identifies sires with the best genetic merit.

The predictive ability was very similar for the threshold model and Gaussian model when considering the binary data. In a simulation study, Meijering and Gianola (1985) reported that threshold models only outperformed Gaussian models when the trait's heritability was moderate to high in magnitude, and the incidence of the binary trait in the population was $\leq 10\%$. In our study, the heritability of both traits was low to moderate and the frequency of animals with worm egg and tick count equal to zero in the dataset were 40% and 43%, respectively. Therefore, our results are consistent with Meijering and Gianola (1985). The model's predictive ability was similar when considering the recorded number of worm egg and tick counts, except for the Gaussian model on the logarithmic scale. This result may be explained by the fact that 60% and 57% of the animals in the current study were infected with worms and ticks, respectively. The Gaussian model on the logarithmic scale poorly predicts counts above zero, and the overall MSEP performance was remarkably affected, resulting in the worst model predictive ability. The failure of models that assume non-Gaussian distributions to outperform Gaussian models in terms of predictive ability has been reported for categorical and counting data (Peñagaricano et al., 2011; Vazquez et al., 2012). Although non-Gaussian models have more appropriate properties to fit discrete data, Gaussian models have shown to be competitive with regard to predictive ability. In addition, the implementation of the Monte-Carlo Cross-Validation considering a parallel computing was crucial to deliver results in a feasible and reduced amount of time. For instance, testing 1200 models in the cross validation approach, took around 2 days to complete. If models were tested sequentially in a unique computer, the overall time could drastically increase, leading to months to complete the cross-validation process.

Genetic analysis of worm and tick infestation often considers the phenotype as counting data instead of treating it as categorical. Phenotypes expressed in classes have the disadvantages of introducing subjectivity into the analysis due to the difficulty of assigning the exact limits of adjacent categories, and also it is often associated with a reduction in the sample variance. The introduction of subjectivity into any genetic evaluation often produces outcomes that deviate from a Gaussian distribution (Varona et al., 2009). Furthermore, the reduction of phenotypic variance might affect the response to selection due to the lack of variability among animals. In the present study, when the phenotype was treated as binary the sample phenotypic variance decreased from 21.9 to 0.24 (counts per animal)^2 and from 16.3 to 0.24 (counts per animal)^2 for worms and ticks, respectively. An alternative to avoid this drastic reduction in variance is to define more classes for the outcome variable instead of only two categories. On the other hand, the main advantage of measuring phenotype into response categories is the practical impact in labor efficiency. Counting the number of parasites on animals is typically time-consuming, whereas scoring the level of infection for each animal is much faster and easier to record.

In this current study, Gaussian models were evaluated using both raw counts and the logarithmic transformation assuming a constant equal to 0.1. Both approaches often provide wrong interpretation of statistical results when they fail to align with any assumptions of the mixed model methodology including normality, independence, and linearity. For instance, Wilson et al. (1996) found that the log-transformation method is more likely to report spurious difference in parasite burdens in comparison to generalized linear models because

this transformation tends to be more associated with both errors of type I and II. In addition, Mair et al. (2015) reported that heritability estimates depend on the additional constant used in the log transformation. Therefore, when using Gaussian models with original or transformed phenotypes, such factors need to be investigated to avoid any misinterpretation of statistical results.

5. Conclusion

Cross-validation results presented no statistical difference in predictive ability between models. However, the DIC indicated a better goodness of fit for models which assumed a non-Gaussian distribution. Therefore, in our study, such models showed an advantage over Gaussian models and could be more explored in genetic analysis of worm and tick infestation. Furthermore, non-Gaussian models have another advantage over Gaussian models because they do not require data transformation, avoiding all its issues such as interpretability. Ultimately, heritability estimates were of low to moderate magnitude, indicating that genetic improvement is certainty possible through selective breeding.

Author statement

Tiago Luciano Passafaro: Conceptualization, Methodology, Data Curation, Formal Analysis, Writing – Original draft, Writing – Review & Editing. Fernando Brito Lopes: Conceptualization, Formal Analysis, Investigation, Writing – Review & Editing. Thomas Wayne Murphy: Writing – Review & Editing. Bruno Dourado Valente: Writing – Review & Editing. Romário Cerqueira Leite: Writing – Review & Editing. Guilherme Jordão de Magalhães Rosa: Writing – Review & Editing, Supervision. Fabio Luiz Buranelo Toral: Writing – Review & Editing, Supervision, Funding Acquisition.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.livsci.2020.104276.

References

Ayres, D.R., Pereira, R.J., Boligon, A.A., Silva, F.F., Schenkel, F.S., Roso, V.M., Albuquerque, L.G., 2013. Linear and Poisson models for genetic evaluation of tick resistance in cross-bred Hereford x Nellore cattle. J. Anim. Breed. Genet. 130, 417–424. https://doi.org/10.1111/jbg.12036.

Bishop, S.C., 2012. Possibilities to breed for resistance to nematode parasite infections in small ruminants in tropical production systems. Animal 6, 741–747. https://doi.org/10.1017/S1751731111000681.

Burrow, H.M., 2001. Variances and covariances between productive and adaptive traits and temperament in a composite breed of tropical beef cattle. Livest. Prod. Sci. 70, 213–233.

De Graef, J., Claerebout, E., Geldhof, P., 2013. Anthelmintic resistance of gastrointestinal cattle nematodes. Vlaams Diergeneeskd. Tijdschr. 82, 113–123.

Fernández-Salas, A., Rodríguez-Vivas, R.I., Alonso-Díaz, M.A., 2012. First report of a Rhipicephalus microplus tick population multi-resistant to acaricides and ivermectin

in the Mexican tropics. Vet. Parasitol. 183, 338–342. https://doi.org/10.1016/j. vetpar.2011.07.028.

Gianola, D., Foulley, J.L., 1983. Sire evaluation for ordered categorical data with a threshold model. Genet. Sel. Evol. 15, 201–224. https://doi.org/10.1186/1297-9686 -15-2-201.

Frisch, J.E., O Neill, C.J., Kelly, M.J., 2000. Using genetics to control cattle parasites - The Rockhampton experience. Int. J. Parasitol. 30, 253–264. https://doi.org/ 10.1016/S0020-7519(00)00010-2.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Bayesian Stat 4, 169–193, 1176289.

Grisi, L., Leite, R.C., Martins, J.R.D.S., Barros, A.T.M., De, Andreotti, R., Cançado, P.H.D., León, A.A.P., De, Pereira, J.B., Villela, H.S., 2014. Reassessment of the potential economic impact of cattle parasites in Brazil. Rev. Bras. Parasitol. Veterinária 23, 150–156. https://doi.org/10.1590/S1984-29612014042.

Hadfield, J.D., 2010. MCMC methods for multi-respoinse generalized linear mixed models: the MCMCglmm R package. J. Stat. Softw. 33, 1–22. https://doi.org/ 10.1002/ana.22635.

Lambert, D., 1992. Zero-inflated poisson regression, with an application to defects in manufacturing. Technometrics 34, 1–14. https://doi.org/10.2307/1269547.

Lew-Tabor, A.E., Rodriguez Valle, M., 2016. A review of reverse vaccinology approaches for the development of vaccines against ticks and tick-borne diseases. Ticks Tick Borne Dis. 7, 573–585. https://doi.org/10.1016/j.ttbdis.2015.12.012.

Mair, C., Stear, M., Johnson, P., Denwood, M., Jimenez de Cisneros, J.P., Stefan, T., Matthews, L., 2015. A Bayesian generalized random regression model for estimating heritability using overdispersed count data. Genet. Sel. Evol. 47, 51–64. https://doi. org/10.1186/s12711-015-0125-5.

Meijering, A., Gianola, D., 1985. Linear versus nonlinear methods of sire evaluation for categorical traits: a simulation study. Genet. Sel. Evol. 17, 115–132.

Morris, C.A., Wheeler, M., Watson, T.G., Hosking, B.C., Leathwick, D.M., 2005. Direct and correlated responses to selection for high or low faecal nematode egg count in Perendale sheep. New Zeal. J. Agric. Res. 48, 1–10. https://doi.org/10.1080/00288233.2005.9513625.

Naya, H., Urioste, J.I., Chang, Y.-.M., Rodrigues-Motta, M., Kremer, R., Gianola, D., 2008. A comparison between Poisson and zero-inflated Poisson regression models with an application to number of black spots in Corriedale sheep. Genet. Sel. Evol. 40, 379–394. https://doi.org/10.1051/gse:2008010.

Nelder, J.A., Wedderburn, R.W.M., 1972. Generalized Linear Models. J. R. Stat. Soc. A 135, 370–384. https://doi.org/10.2307/2344614.

Oliveira, P.S., Santana Junior, M.L., Pedrosa, V.B., Oliveira, E.C.M., Eler, J.P., Ferraz, J.B. S., 2011. Estrutura populacional de rebanho fechado da raça Nelore da linhagem Lemgruber. Pesq. Agropec. Bras. 46, 639–647, 10.1590/S0100-204×2011000600010.

Passafaro, T.L., Carrera, J.P.B., Santos, L.L., dos, Raidan, F.S.S., Santos, D.C.C., dos, Cardoso, E.P., Leite, R.C., Toral, F.L.B., 2015. Genetic analysis of resistance to ticks, gastrointestinal nematodes and Eimeria spp. in Nellore cattle. Vet. Parasitol. 210, 224–234. https://doi.org/10.1016/j.vetpar.2015.03.017.

Peñagaricano, F., Urioste, J.I., Naya, H., de los Campos, G., Gianola, D., 2011. Assessment of Poisson, Probit and linear models for genetic analysis of presence and number of black spots in Corriedale sheep. J. Anim. Breed. Genet. 128, 105–113. https://doi.org/10.1111/j.1439-0388.2010.00893.x.

Prayaga, K.C., 2003. Evaluation of beef cattle genotypes and estimation of direct and maternal genetic effects in a tropical environment. 2. Adaptive and temperament traits. Aust. J. Agric. Res 54, 1027–1038. https://doi.org/10.1071/AR03072.

Prayaga, K.C., Henshall, J.M., 2005. Adaptability in tropical beef cattle: genetic parameters of growth, adaptive and temperament traits in a crossbred population. Aust. J. Exp. Agric. 45, 971–983. https://doi.org/10.1071/EA05045.

Rashid, M., Rashid, I., Akbar, H., Ahmad, L., Hassan, M.A., Ashraf, K., Saeed, K., Ghardi, M., 2019. A systematic review on modelling approaches for economic losses studies caused by parasites and their associated diseases in cattle. Parasitology 146, 129–141. https://doi.org/10.1017/S0031182018001282.

Rodrigues-Motta, M., Gianola, D., Heringstad, B., Rosa, G.J.M., Chang, Y.M., 2007. A zero-inflated poisson model for genetic analysis of the number of mastitis cases in Norwegian Red cows. J. Dairy Sci. 90, 5306–5315. https://doi.org/10.3168/jds.2006-898.

Sae-Lim, P., Grova, L., Olesen, I., Varona, L., 2017. A comparison of nonlinear mixed models and response to selection of tick-infestation on lambs. PLoS ONE 12, e0172711. https://doi.org/10.1371/journal.pone.0172711.

Silva, F.F., Tunin, K.P., Rosa, G.J.M., da Silva, M.V.B., Azevedo, A.L.S., Verneque, R., da, S., Machado, M.A., Packer, I.U., 2011. Zero-inflated poisson regression models for QTL mapping applied to tick-resistance in a Gyr x Holstein F2 population. Genet. Mol. Biol. 34, 575–581. https://doi.org/10.1590/S1415-47572011005000049.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. Ser. B (Statistical Methodol.) 64, 583–639. https://doi.org/10.1111/1467-9868.00353.

Stroup, W.W., 2015. Rethinking the analysis of non-normal data in plant and soil science. Agron. J. 107, 811–827. https://doi.org/10.2134/agronj2013.0342.

Tempelman, R.J., 1998. Generalized linear mixed models in dairy cattle breeding. J. Dairy Sci. 81, 1428–1444. https://doi.org/10.3168/jds.S0022-0302(98)75707-8.

Ueno, H., Gonçalves, P.C., 1998. Manual Para Diagnóstico Das Helmintoses De Ruminantes, 4th ed. Japan International Coorporation Agency, Tokyo, Japan.

Varona, L., Moreno, C., Altarriba, J., 2009. A model with heterogeneous thresholds for subjective traits: fat cover and conformation score in the Pirenaica beef cattle. J. Anim. Sci. 87, 1210–1217. https://doi.org/10.2527/jas.2008-1361.

Vazquez, A.I., Perez-Cabal, M.A., Heringstad, B., Rodrigues-Motta, M., Rosa, G.J.M., Gianola, D., Weigel, K.A., 2012. Predictive ability of alternative models for genetic analysis of clinical mastitis. J. Anim. Breed. Genet. 129, 120–128. https://doi.org/10.1111/j.1439-0388.2011.00950.x.

Villemereuil, P., Schielzeth, H., Nakagawa, S., Morrissey, M., 2016. General methods for evolutionary quantitative genetic inference from generalized mixed models. Genetics 204, 1281–1294. https://doi.org/10.1534/genetics.115.186536.

Wharton, R.H., Utech, K.B.W., 1970. The Relation between engorgement and dropping of Boophilus microplus (Canestrini) (Ixodidade) to the assessment of tick numbers on cattle. J. Aust. Entomol. Soc. 9, 171–182.

Wilson, K., Grenfell, B.T., Shaw, D.J., 1996. Analysis of aggregated parasite distributions: a comparison of methods. Funct. Ecol. 10, 592–601.

Wolstenholme, A.J., Fairweather, I., Prichard, R., Von Samson-Himmelstjerna, G., Sangster, N.C., 2004. Drug resistance in veterinary helminths. Trends Parasitol 20, 469–476. https://doi.org/10.1016/j.pt.2004.07.010.