Veterinary Parasitology 270 (2019) 31–39

Contents lists available at ScienceDirect

Veterinary Parasitology

journal homepage: www.elsevier.com/locate/vetpar

Research paper

Quantification of differences in resistance to gastrointestinal nematode infections in sheep using a multivariate blood parameter

Amy Bell[a], Jody McNally[a], Daniel V. Smith[b], Ashfaqur Rahman[b], Peter Hunt[a], Andrew C. Kotze[c], Sonja Dominik[a], Aaron Ingham[c,*]

[a] CSIRO Agriculture and Food, Armidale, New South Wales, 2350, Australia
[b] CSIRO Data61, Sandy Bay, Tasmania, 7005, Australia
[c] CSIRO Agriculture and Food, St Lucia, Queensland, 4067, Australia

ARTICLE INFO

Keywords:
Blood parameters
Haemonchus contortus
Resistance
Sex differences
Sheep-nematoda

ABSTRACT

Breeding for resistance to gastrointestinal nematodes (GIN) in sheep relies largely on the use of worm egg counts (WEC) to identify animals that are able to resist infection. As an alternative to such measures of parasite load we aimed to develop a method to identify animals showing resistance to GIN infection based on the impact of the infection on blood parameters. We hypothesized that blood parameters may provide a measure of infection level with a blood-feeding parasite through perturbation of red blood cell parameters due to feeding behaviour of the parasite, and white blood cell parameters through the mounting of an immune response in the host animal. We measured a set of blood parameters in 390 sheep that had been exposed to an artificial regime of repeated challenges with *Trichostrongylus colubriformis* followed by *Haemonchus contortus*. A simple analysis revealed strong relationships between single blood parameters and WECs with correlation coefficients −0.54 to −0.60. We then used more complex multi-variate methods based on supervised classifier models (including Bayesian Network) as well as regression models (Lasso and Elastic Net) to study the relationships between WECs and blood parameters, and derived algorithms describing the relationships. The ability of these algorithms to classify sheep GIN resistance status was tested using the WEC and blood parameters collected from a different group of 418 sheep that had acquired natural infections of *H. contortus* from pasture. We identified the most resistant and most susceptible animals (10% percentiles) of this group based on WECs, and then compared the identities of these animals to the identities of animals that were predicted to be most resistant and most susceptible by our algorithms. The models showed varying abilities to predict susceptible and resistant sheep, with up to 65% of the most susceptible animals and 30% of the most resistant animals identified by the Elastic Net model algorithms. The prediction algorithms derived from female sheep data performed better than those for male sheep in some cases, with the predicted animals accounting for up to 50–60% of the actual resistant and susceptible female animals. Heritability values were calculated for blood parameters and the aggregate trait descriptions defined by the novel prediction algorithms. The aggregate trait descriptions were moderately heritable and may therefore be suitable for use in genetic selection strategies. The present study indicates that multivariate models based on blood parameter data showed some ability to predict the resistance status of sheep to infection with *H. contortus*.

1. Introduction

Parasitic nematodes are a major cost and constraint to livestock production, with particular relevance for sheep and cattle (Charlier et al., 2014; Mavrot et al., 2015). Producers commonly mitigate this impact through selective breeding programs involving individuals with high breeding values for resistance to internal parasites. Such parasite resistant individuals are identified by breeding values based on phenotypes such as measures of parasite load, for example counts of eggs in faeces for gastrointestinal parasites (Berton et al., 2017; Kim et al., 2014; Williams et al., 2010). The heritability for worm (or faecal) egg count (WEC) in sheep has been reported to range from 0.2 to 0.4 (Bishop, 2012; Brown and Fogarty, 2016), in goats from 0.1 to 0.35 (Burrow and Henshall, 2014; Bishop, 2012) and in cattle from 0.04 to 0.36 (Barlow and Piper, 1985; Burrow, 2001; May et al., 2017).

An alternative phenotype for the identification of animals that are able to resist parasitic disease may be determined after measuring the impact of infection on the host, rather than measuring parasitic load

* Corresponding author.
E-mail address: Aaron.Ingham@csiro.au (A. Ingham).

https://doi.org/10.1016/j.vetpar.2019.05.007
Received 1 March 2019; Received in revised form 16 May 2019; Accepted 17 May 2019

directly. A number of different studies have shown differences in the level of haematological parameters in the blood of sheep infected with haematophagous gastrointestinal nematodes (GIN). For example, erythrocyte or red blood cell levels as measured by Haematocrit (HCT), also known as packed cell volume (PCV), which is a ratio of the volume of red blood cells over the total volume of blood, has been shown to have a relationship with parasitic load and a heritability of 0.35 to 0.45 (Albers et al., 1990). In other studies, characterisation and comparison of the blood of populations of sheep selected for an ability to resist or succumb to nematode infection have also shown differences in levels of eosinophils, serum antibody, and CD4 T cells (Douch et al., 1996; Windon, 1996).

Extending from these early studies, our laboratory has previously shown that a multivariate analysis of a set of blood parameters, rather than individual parameters as described above, could be used to define algorithms that predicted an animal's ability to resist infection by the parasitic nematode *Haemonchus contortus* (Andronicos et al., 2014). In the study, we tested for correlations between blood parameters and worm burden, as measured by WEC, in sheep that had been selected for resistance or susceptibility to nematodes. The blood parameters tested included red blood cells, haemoglobin, haematocrit, reticulocytes, white blood cells, neutrophils, eosinophils. We then derived two functions that related the variation in these parameters to nematode infection and tested their ability to predict the resistance status of individual animals. The resulting algorithms achieved a binary classification system that provides a ranking of 0/1 (resistant or susceptible) to the animals. Initial results showed the functions could discriminate resistant (with 80% accuracy) and susceptible sheep (100% accuracy) at later time points of infection (from 21 days), but not prior to, or during, the early stages. However, when the functions were applied to a population of outbred sheep the ability to classify the resistance status was poor.

Here we aimed to use more sophisticated mathematical models to produce a continuous scale of resistance, enabling extreme elite performers to be identified (rather than a binary classification that generated groups), and obtain better classification accuracy, particularly in outbred contemporary populations of sheep. The models were developed in an initial nematode infection trial using outbred rather than selection line sheep, and then these models were tested for an ability to rank sheep for nematode resistance in a second independent trial. We assessed the potential of these models to be used in breeding programs by calculating the heritability and compared the result to that of worm egg count (WEC), the current industry standard for ranking resistance to GIN.

2. Materials and methods

2.1. Animal trials

2.1.1. Trial 1: artificial worm challenge

Trial 1 data set was used to develop the novel blood phenotypes. Sampling and handling of sheep born in late spring of 2003 and 2004 (Dominik et al., 2010) was carried out in accordance to methods approved in the CSIRO animal ethics committee applications ARA 03/12, 03/20, 03/74, 04/07, 04/51, 06/01. The population was a Romney x Merino cross and comprised 390 sheep, with 192 males and 198 females. After weaning at approximately 5 months of age, all animals were faecal sampled to assess any pre-existing parasite burden. Around 50% of lambs had low worm egg counts (WEC) (whole flock median WEC = 300 eggs/g). Irrespective of infection status all lambs were treated with anthelmintic drugs (fenbendazole, levamisole and naphthalaphos at the manufacturer's recommended dose). Lambs were then challenged with 3 orally administered parasite infections: a primary and secondary challenge with 20 000 McMaster isolate *Trichostrongylus colubriformis* larvae (*Tc*) and a third challenge with 5000 Kirby1981 isolate *H. contortus* larvae (*Hc*). Each infection proceeded for 5 weeks before an anthelmintic treatment (as described above) was used to clear the infection, followed by a 1-week break before the next challenge. At the end of the infection regime, animals received a further anthelmintic treatment. The infection and sampling protocol was designed to resemble that used by Beh et al. (2002). Each parasite challenge was assessed using larvae cultured from pooled faeces and was found to consist predominantly of the intended parasite species. Faecal samples were taken from the rectum on days 21, 28 and 35 of each infection and WEC assessed using the McMaster egg flotation technique for each sample (Gordon and Whitlock, 1939). WEC was not adjusted for faecal moisture content as samples were assumed to be equivalent. For each infection, the mean of the WEC at the 3 sampling times was analysed (mean WEC 1, 2 and 3). Peripheral blood was collected from the jugular vein into potassium EDTA Vacutainer tubes. Five ml of blood was sampled immediately prior to the primary infection for all animals and following the tertiary infection on days 0, 4, 15 and 28 (for 2003 born sheep) and days 0, 4, 14 and 29 (for 2004). Only the data derived from analysis of the blood samples taken prior to the first infection, and at the final time point of the tertiary infection (days 28 and 29), were examined in the present study.

2.1.2. Trial 2: natural field H. contortus challenge

Data from Trial 2 was used for testing of the algorithms and to estimate variance components and calculate heritabilities. All sampling and animal handling performed in this trial was carried out in accordance to methods approved in CSIRO animal ethics committee application ARA 14/29. Blood was sampled from, and WEC determined in, 418 Merino sheep (216 male and 202 female) that were exposed to a natural *H. contortus* challenge between February and March of 2015. The sheep were not administered with long lasting drenches prior to the sampling period to ensure each animal was subject to infection from the pasture. Random faecal samples were collected to monitor WEC levels 1–2 times per week. At approximately 28 days post peak infection (during March 2015), corresponding to mean WEC levels above 800 eggs per gram (epg), blood and faeces were sampled from each sheep. At the conclusion of sampling, all sheep were treated with an appropriate anthelmintic to terminate the infection.

2.2. Blood cell profiling

The whole blood samples were analysed for a standard haematological panel using a Cell Dyn 3700 (Abbott Laboratories, USA) according to the manufacturer's instructions and cell values recorded at 10^9/l. The blood was profiled for 13 parameters (white blood cells (WBC), neutrophils (NEU), lymphocytes (LYM),monocytes (MONO), eosinophils (EOS), basophils (BASO), red blood cells (RBC), haemoglobin (HGB), haematocrit (HCT), mean corpuscular volume (MCV), mean corpuscular haemoglobin (MCH), mean corpuscular haemoglobin concentration (MCHC), and platelets (PLT)).

2.3. Phenotype algorithm development

The relationships between the blood parameters and WECs were examined using two different models:

i) a supervised classification approach; based on two discrete classes of infection status (animals either resistant or susceptible to infection)
ii) a linear regression approach, based upon the penalised least squares optimisation using a continuous scale of infection status (rather than two discrete classes)

For both models, the parameters were defined as follows:

Let T be a $\boldsymbol{n\ x\ m}$ matrix representing blood samples collected from m animals $T = [\boldsymbol{x}_1 \ldots \boldsymbol{x}_m]$ and $\boldsymbol{L}$ be the corresponding vector of infection levels L=[l_1, ...,l_m], where the infection level was estimated from the observed Worm Egg Count (WEC) in the faecal sample. The blood sample of each animal was represented by a n x 1 vector of the blood parameters, $\boldsymbol{x_i} = [x_1, ..., x_n]^T$. Both models were trained upon the data set of blood samples T in order to minimise the estimation error of the infection measure. Each model was trained using values obtained from the samples of artificially challenged sheep in Trial 1 and then the accuracy of the model was validated from the blood samples of sheep naturally exposed to *H. contortus* in Trial 2. Blood parameters were used as collected (without adjustment) in all experiments. Animals were ranked (i.e. resistance scores) based on egg counts for both classification and regression. Functions were then developed to transform blood parameters of an animal to resistance score.

The two modelling approaches are described below:

i) Supervised classification based model:
The supervised classifier was defined such that:

$$l_i \approx f_\theta(x_1, \ ..., x_n)$$

where the classifier function f maps the vector of blood parameters ($\boldsymbol{x_i}$) of each sheep to two discrete classes of infection status $l_i \in \{r, s\}$ where r is the resistance class and s is the susceptibility class. The class label l_i associated with each sheep was estimated from it's observed WEC_i. Sheep with $WEC_i > S$ were classed as susceptible and $WEC_i < S$ were classed as resistant. Here S is a threshold. During training, the classifier was optimised through adaptation of its parameters θ to minimise the number of sheep in the training set that had their infection status incorrectly estimated by it.
Once the classifier was trained, values obtained from cell profiling of the blood samples were applied to the classifier producing posterior probabilities for the two infection classes p_r and p_s such that $p_r + p_s = 1$ where $0 \leq p_r, p_s \leq 1$. Here p_r represents the probability of the sample belonging to the resistance class and p_s represents the probability of the sample belonging to the susceptible class. Commonly in a classification task, the sheep would be classified as resistant if $p_r > p_s$, and susceptible if $p_r \leq p_s$. In this study, however, we used the p_s estimates from the classifier to generate a continuous scale of resistance to *H. contortus* infection, where a value closer to 0 is indicative of higher resistance and a value closer to 1 is indicative of higher susceptibility to infection.
A set of classifiers were trained to evaluate the accuracy in which the resistant/susceptible animals could be classified. The performance of a number of classifiers including the Decision Tree, AdaBoost, Naïve Bayes and Bayesian Network were evaluated using WEKA Data Mining software. (Hall et al., 2009). To categorise each animal's infection status as either susceptible ($e_i > S$) or resistant ($e_i < S$), the classification threshold S was set to a WEC of 10,000.

ii) Regression based models for ranking nematode resistance in male, female or mixed cohorts:
The infection resistance metric of the sheep ($\boldsymbol{L}$) was computed by normalising the set of observed WEC values, so that the metric values remained in the range, $0 \leq l_i \leq 1$. The square root transform of metric $\boldsymbol{L}$ was then applied in order to improve the fit of data to a Gaussian distribution. The standard linear regression function was defined as:

$$\boldsymbol{L} = BT + \boldsymbol{\varepsilon}$$

iii) where T is the matrix composed of m blood samples, B represents a 1 x n vector of blood parameter coefficients and ε is the model error. This function was solved using standard least squares optimisation. Some of the blood parameters in T may decrease the ability to estimate the infection resistance metric, however, before the analysis is conducted, it is unclear which subset of the blood parameters these are. Whilst blood parameters can be independently examined and selected before modelling, in this study we applied variable selection methods. That is, the blood parameter selection is implicitly incorporated into the least squared optimisation used in model construction.

Two regularised linear regression models were used to incorporate parameter penalisation into the least squares solution:

i) LASSO (Least Absolute Shrinkage and Selection Operator); incorporates the l1-norm (sum of the absolute coefficients) into the least squares formulation in order to shrink parameter coefficient values during optimisation, and hence, automate blood parameter selection by forcing blood parameters with minimal influence (i.e. small coefficient values) to zero. When LASSO is used upon a group of correlated parameters that are each relevant to the dependent variable (i.e. this case the infection status), only one of the parameters will be selected.
ii) Elastic net; adds an additional l2-norm (the sum of the squared coefficients) into the parameter penalisation term of LASSO (l1-norm) to allow groups of correlated blood parameters to contribute to the model when informative.

The LASSO and Elastic Net models were trained using the data set from Trial 1. The original set of 13 blood parameters were combined with log10 and square root transformations of the original parameters to generate a data-set of 39 features. The square root transform of metric $\boldsymbol{L}$ was then applied in order to improve the fit of data to a Gaussian distribution.

Ten-fold cross-validation was used to evaluate the performance of the LASSO and Elastic Net regression models for sheep from the same flock in Trial 1. All the animals were randomly split into 10 similar sized groups and each animal was represented in one group only. In an iterative fashion each group became a test set and the remaining nine groups were combined and used as the training set. This ensured that all sheep were only tested on a single occasion, and the blood sample from a single sheep was not incorporated into the test and training sets simultaneously. The average performance across the test sets is reported.

The approach was initially applied to all of the animals in the trial, but was subsequently used to develop models for male only or female only groups of animals.

2.4. Heritability estimates

Pedigree was available for all animals from Trial 2 for the estimation of heritabilities. Animals originated from a flock with distinct features that need to be considered in the linear model for the estimation of variance components and phenotype heritability. The flock comprised two lines that were resistant and susceptible to fly strike (Smith et al., 2009). Base animals in the pedigree represented three genetic groups, defining their wool type (Ultrafine/Superfine, Fine/Fine Medium and Medium/Strong). Of the 436 sheep in this flock, 216 animals were from the flystrike susceptible selection line and 220 from the flystrike resistant line (Line), and 229 were males and 207 females (sex). A total of 210 animals were born as single and 226 as multiple (birth type). Of these 240 were reared as singles and 195 reared as twins (rear type).

A univariate animal model with genetic groups of wool type was fitted in ASReml 3 (Gilmour et al., 2009) testing fixed effects sex (male and female), birth type and rear type and selection line for breechstrike resistance. Significant effects were retained in the model. Phenotypic and genetic variance components were estimated and the heritability as the ratio of the genetic over the phenotypic variance.

3. Results

3.1. Analysis of blood parameters in the trial 1 data-set

The 13 blood parameters of the sheep challenged in Trial 1 were

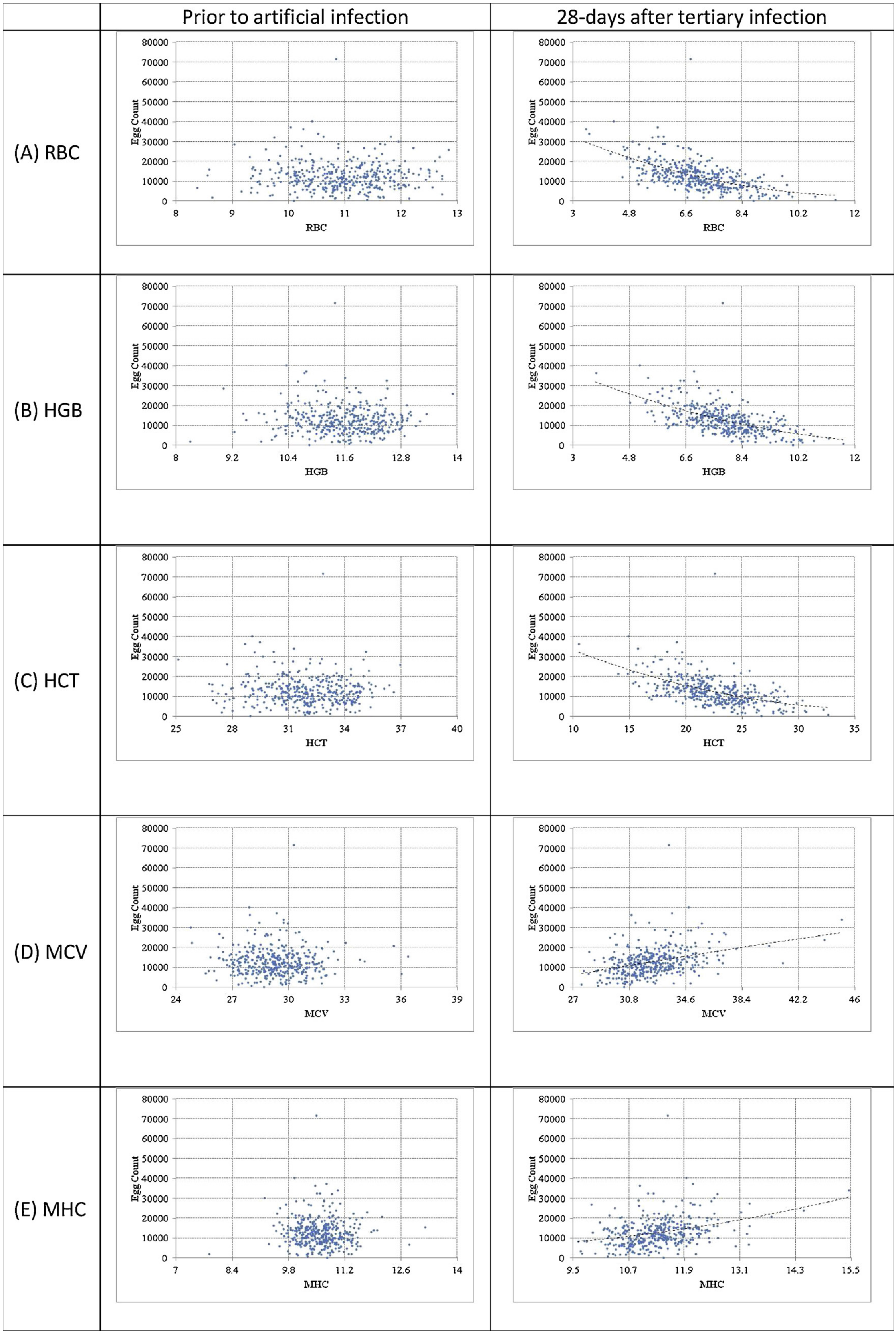

Fig. 1. Relationship between blood parameters and egg count. Blood parameters measured in individual animals prior to artificial infection (left-hand panels) and 28-days after tertiary infection (right-hand panels); parameters are RBC (A), HGB (B), HCT (C), MCV (D) and MHC (E).

examined prior to infection and then 28 or 29 days following the tertiary *H. contortus* infection. The RBC, HGB, HCT, MCV, MHC blood parameters were found to be most useful for discriminating the samples collected at the two time points. As shown in Fig. 1 images A to E the measured range for each blood parameter was RBC (no infection: 9–13, infection: 4.8 to 10.2), HGB (no infection: 9.5–13, infection: 4.8 to 10.2), HCT (no infection: 28–36, infection: 15 to 30), MCV (no infection: 27–32, infection: 27.5 to 37), MHC (no infection: 9.8–11.5, infection: 9.5 to 13). These blood parameters did not show any correlation with low WEC from natural infection prior to artificial infection, whereas the correlations were present at 28–29 days after the *H. contortus* infection; the correlation coefficients for RBC, HGB, HCT, MCV, MHC with WEC were –0.594, –0.566, –0.540, 0.370, and 0.333 respectively.

We ran the experiments using data from samples collected on day 28, the earliest day when the differences between blood parameters become detectable in trial 1. On that day, the egg count range was 1000 to 28,000 in trial 1 and 0 to 59,600 trial 2. We considered a classification threshold of WEC 10,000 epg on day 28 or 29 in order to categorise each animal as either susceptible (> 10,000) or resistant (< 10,000), as this was the mid value of the distribution. We then examined the relationship between animals in these discrete groups and the three blood parameters that had shown the highest correlation coefficients, as described above (that is, RBC, HGB and HCT). Fig. 2 presents a scatter plot of these three parameters in animals labelled as resistant (i.e. WEC < 10,000) or susceptible (i.e. WEC > 10,000). Two relationships are apparent: (i) Highly resistant and highly susceptible animals are at two opposite extremes of the distribution, and (ii) the relationship between ability to resist nematode infection as measured by WEC and the three blood parameters is linear in nature.

3.2. Supervised classifier model

Table 1 shows the prediction accuracies that were obtained from the different classifiers using a tenfold cross validation on data generated from Trial 1. Cross validation of the data set ensured that the models were not trained and tested upon the same sheep. The highest classification accuracy of 76.2% was obtained with Bayesian Network (BN). Consequently, the BN was selected to estimate the infection resistance metric (p_r, p_s) for the naturally exposed sheep in Trial 2 (as described in Section 3.4).

3.3. Regression models

Regression equations were generated for the Elastic net and LASSO models using data from Trial 1 for male or female sheep separately, or from all of the sheep. The LASSO equations are shown in Table 2.

The performance of the LASSO and Elastic Net regression models that were trained upon blood samples from Trial 1 were tested,(i.e. used to predict infection status, $\hat{\boldsymbol{L}}$) upon the sheep in Trial 1 or the blood samples of a different flock of sheep in Trial 2 is shown in Table 3. The Root Mean Squared Error (RMSE) and correlation coefficient metrics (R) were used to compare the infection resistance metric estimated by each model and the infection resistance metric computed from the WEC.

The regression performance was almost equivalent between the LASSO and Elastic Net approaches (identical R values for all model/gender comparisons with the exception of the all sheep group for Trial 1, 0.73 vs 0.72). The results indicate that the regression models were able to estimate the infection status scores within the same flock of sheep (that is, Trial 1-based models tested on Trial 1 sheep) with a reasonable level of accuracy: average estimation errors for $\boldsymbol{L}$ ranged between 8.4% and 9.3%. It is clear, however, that the performance of each regression model decreased when testing was performed upon a different flock of sheep that were naturally exposed to an infection in the field (that is, Trial 1-based models tested on Trial 2 sheep). Table 3 shows the average RMSE increased by between 27% and 53% for the Trial 2 estimations when compared to Trial 1. The female model was most effective at generalising to the sheep in Trial 2, producing an average estimation error of 12.3%. The all sheep model has an average estimation error that was 25% higher than the female model; whilst the male model was the least effective at generalising across trials with a significant estimation error of 20%. This contrast between the use of the model on the trial data set that it had been derived from (Trial 1) compared to its use on a subsequent trial (Trial 2) is illustrated in the scatter plots of actual versus estimated infection metric, shown in Fig. 3.

3.4. Prediction of sheep as nematode susceptible or resistant using the supervised classifier (Bayesian Network) or the regression models (Lasso and Elastic net)

The animals in Trial 2 were sorted in ascending order based upon their WEC, and the animals identified in the 0–10% percentile band were categorised as resistant and those in the 90–100% percentile band as susceptible. The same set of animals were also sorted in ascending order using their predicted infection status values from the gender specific and gender independent Elastic Net and LASSO regression models ($\hat{\boldsymbol{L}}$), and the Bayesian Network classifier (p_s). In a similar fashion the identities of the animals in the 0–10% and 90–100% percentile bands were categorised as resistant or susceptible, respectively. The set of sheep represented in either the resistance or susceptible groups for both the actual WEC and predicted infection values for each of the models is shown in Table 4. A model with set intersection of higher cardinality (number of animals in the set) in the percentile bands was more effective at identifying sheep in the flock associated with the most extreme infection resistance or susceptibility.

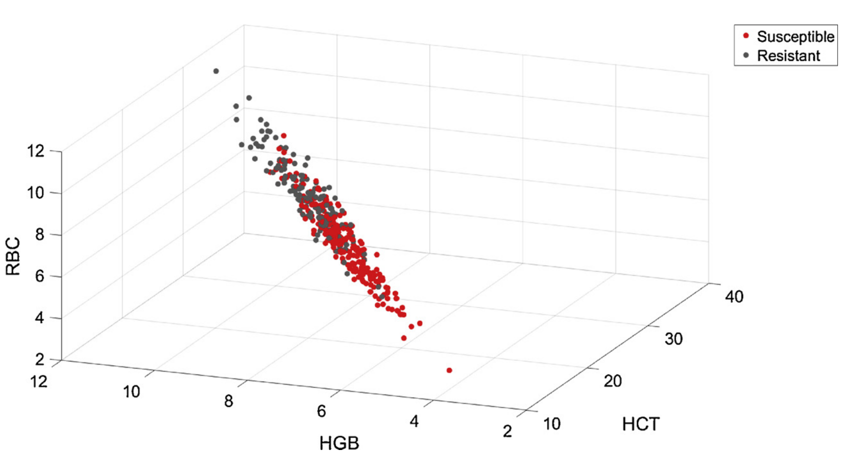

Fig. 2. Scatter plot of individual sheep RBC, HGB and HCT values. Sheep were classified as resistant or susceptible based on egg counts; resistant sheep are represented by grey points and susceptible sheep by red (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Table 1
Prediction accuracies of different classifiers.

Classifier	Precision (NR)	Precision (R)	Recall (NR)	Recall (R)	% Correctly Classified Instances
Decision Tree (J48)	0.857	0.582	0.675	0.801	72.1
Decision Tree (Rep Tree)	0.812	0.609	0.747	0.695	72.8
AdaBoost.M1	0.865	0.600	0.695	0.809	73.6
Naïve Bayes Classifier (NB)	0.814	0.649	0.791	0.681	75.1
Bayesian Network (BN)	0.839	0.650	0.775	0.738	76.2

3.5. Heritabilities

Heritabilities (h^2) for WEC and blood parameters were calculated and the results are presented in Table 5. Heritability values obtained ranged from low to high, with no heritability for neutrophil count (NEU), a low value (0.10) for basophil count (BASO), and the highest for mean corpuscular volume (MCV), with $h^2 = 0.66$. Overall, the heritabilities of blood parameters associated with red blood cells, i.e. RBC, HGB, MCV, MCH and MCHC ranged from $h^2 = 0.35 – 0.66$, whereas the parameters associated with other cell types ranged from $h^2 = 0 – 0.37$. WEC had a heritability of $h^2 = 0.33$ and the heritabilities of the phenotypes produced by the models ranged from $h^2 = 0.37$ - 0.44.

4. Discussion

The present study has examined the use of blood parameter measurements to predict the ability of sheep to resist infection by *H. contortus*. We have used the data from a large trial involving artificial worm challenges to develop models describing the relationships between the blood parameters and worm egg counts. We have then tested these models on both the data set from which they were derived, and, more importantly, on a separate data set from a natural worm challenge experiment. The models showed some ability to predict the resistance status of the most resistant or susceptible animals (as determined by WEC) in the natural field challenge population. We compared the prediction abilities of two types of models: one based on a classifier for each animal using a Bayesian approach, called 'scale' which ranks the resistivity of each animal, 0 indicating highly resistant and 1 indicating highly susceptible, while the other models used regression approaches to predict the infection resistance metric for each animal using the Elastic Net or Lasso techniques. The regression approaches were shown to offer a performance advantage over the supervised classifiers in predicting the top 10% most resistant and most susceptible animals (based on WEC) to infection from the natural worm challenge experiment.

This study also demonstrated that these novel phenotypes had a sufficiently high heritability to be used successfully in a breeding program. Heritabilities were in the range of common sheep production traits (Safari et al., 2005) which indicates that reasonable genetic gains can be expected when these would be incorporated in a breeding program. However, in order for these novel traits to be used in a breeding program, the correlations with the relevant production traits need to be established.

The classifier algorithms developed by the Lasso model provide an insight into the blood cellular populations that enabled sheep to resist GIN infection. These algorithms are composed of blood parameter values, each with a coefficient that scales with the predictive power of that parameter. Here we produced algorithms that classified GIN resistance status for the entire population of sheep as well as for male and female specific groups. The only blood parameter common to each algorithm was the number of red blood cells, which might be expected given that *H. contortus* is a blood feeding parasite. A possible interpretation is that following infection, a resistant animal is likely to have higher numbers of red blood cells than a susceptible sheep.

Sex specific differences were observed in the regression algorithms of male and female sheep, which suggests that they may differ in the mechanism by which they resist, or respond to, GIN infection. In male sheep only parameters associated with red blood cells were represented in the algorithm suggesting a mechanism associated with this cell type that boosts or prevents loss of red blood cell numbers. Such a mechanism is more indicative of a resilience response, but in this instance it is specifically the resilience of the sheep with regard to erythropoiesis. The result may also indicate that male sheep generate a lower level of immune response than do female sheep thereby making the response more difficult to detect. In the female sheep cell types associated with immune function such as white blood cells, eosinophils, basophils and neutrophils were all associated with an ability to resist GIN infection, as measured by lower WEC. Indeed such cell populations have been

Table 3
Performance of Elastic Net and Lasso regression models when tested on the Trial 1 or 2 sheep; analyses were performed using samples from all of the sheep, or males and females separately. Each model was trained upon the samples of sheep in Trial 1; the 'All' model was trained upon all of the sheep, whilst the 'Female' and 'Male' models were trained upon gender specific groups of sheep separately.

The data-set used for testing each model	Models	RMSE $\sqrt{E((L-\hat{L})^2)}$	R
Trial 1	Lasso - All	0.084	0.73
	Elastic Net - All	0.085	0.72
	Lasso - Female	0.089	0.69
	Elastic Net - Female	0.088	0.69
	Lasso - Male	0.093	0.72
	Elastic Net - Male	0.091	0.72
Trial 2	Lasso - All	0.153	0.60
	Elastic Net - All	0.153	0.60
	Lasso - Female	0.123	0.76
	Elastic Net - Female	0.122	0.76
	Lasso - Male	0.200	0.15
	Elastic Net - Male	0.200	0.15

Table 2
The blood parameter algorithms associated with the male, female and all sheep regression models used to estimate *H. contortus* infection resistance when cross-validated using blood samples from the Trial 1 data-set.

Model	Algorithm
Lasso - Male	−0.25RBC_sqrt + 0.02MCV_log10 − 0.35RBC_log10 + 0.86
Lasso - Female	0.05NEU + 0.03MONO − 0.13MCHC − 0.12HGB − 0.39RBC + 0.03NEU_sqrt − 0.06EOS_log10 + 0.80
Lasso	0.09WBC + 0.08BASO + 0.26EOS − 0.13MCHC − 0.40RBC − 0.28EOS_sqrt − 0.06BASO_log10 − 0.22RBC_log10 + 0.93

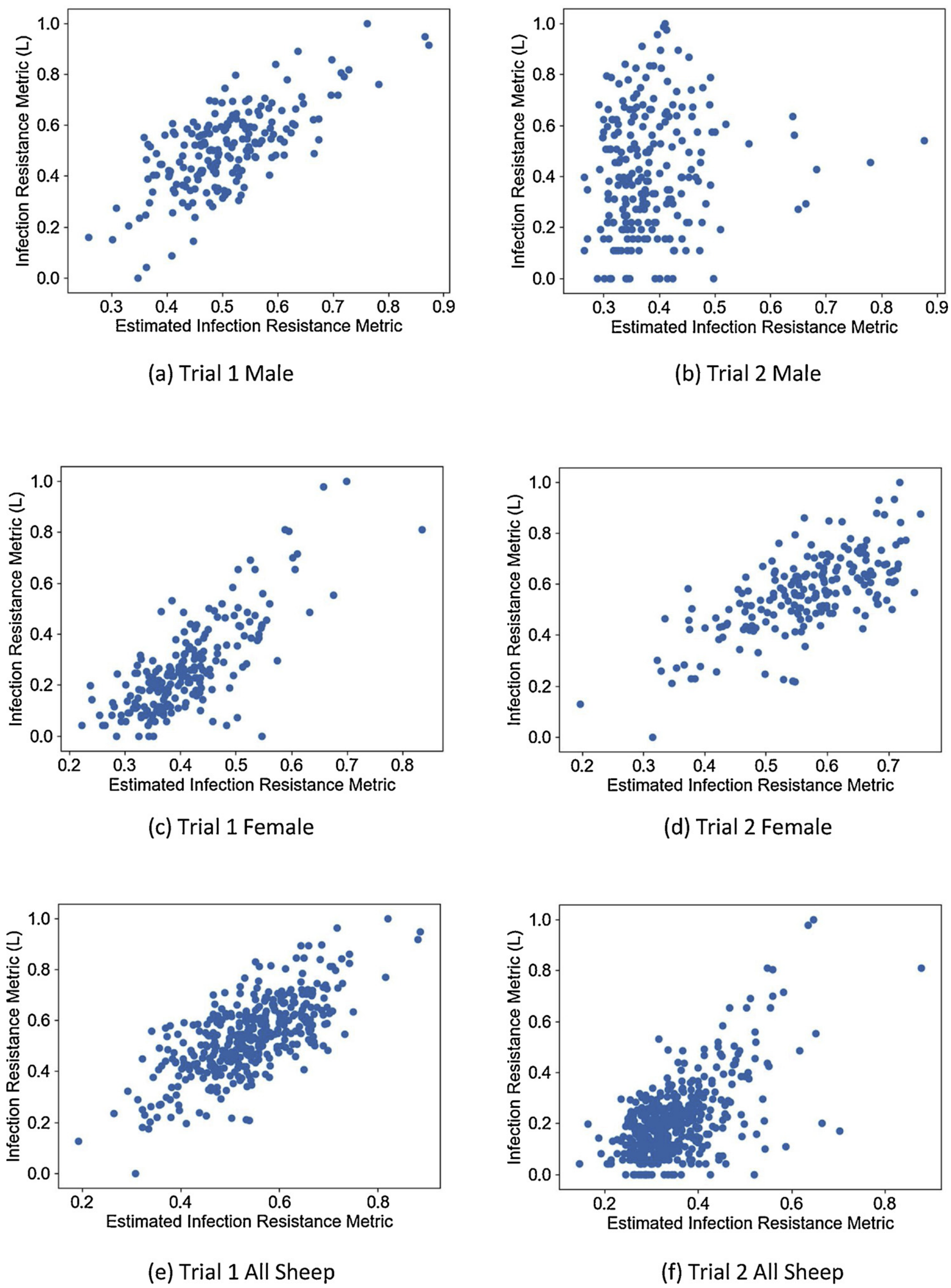

Fig. 3. A comparison of the infection resistance metric (computed from WEC) and the infection resistance metric estimated from Lasso models that were trained using the samples of animals from Trial 1 and then tested either on the samples of animals from the Trial 1 (left-hand panels) or tested on the samples of animals from Trial 2 (right-hand panels). The analyses were performed on male sheep only (top row), female sheep only (middle row), or all sheep (bottom row). A 10 fold cross-validation was performed to test each model on the blood samples of each sheep (in their respective gender group).

associated with the elimination of GIN in sheep previously (Garza et al., 2018; Meeusen et al., 2005; Ortolani et al., 2013; Terefe et al., 2007), and suggest that the mechanism of resistance operating in female sheep involved active expulsion or moderation of the invading nematodes. The influence of sex on the ability of sheep to resist GIN infection has also been reported previously with female sheep found to respond more strongly than male sheep (Abuargob and Stear, 2014; Barger, 1993). Classic genetic methodology relies heavily on sire selection and as such the ability to progress a sheep nematode resistance trait could be hampered if the underlying biology of the phenotype differs between sexes or if the distribution of performance is much narrower in male sheep. Selection on both, the male and female trait might lead to more effective immune responses, that enhance genetic improvement for nematode resistance but this would increase costs associated with collection of the trait. Whereas in the latter case a highly sensitive phenotypic test would be required to accurately rank males for the nematode resistance trait. Further investigation would be required to demonstrate these claims and to fully understand the implications this, or other factors such as age of animal, may have for current approaches aimed at advancing nematode resistance.

The regression classifier models based on blood parameters showed varying abilities to identify sheep classed as susceptible or resistant based on WEC. The Elastic Net model algorithm identified 65% of the most susceptible animals but only 30% of the most resistant animals. The observation that the algorithms identified a higher proportion of sheep at the susceptible end of the nematode response distribution is

Table 4
Intersection of sheep in the 0–10% percentile and 90–100% percentile WEC bands when ranked with Elastic Net and Lasso regression models, and the Bayesian Network supervised classifier model. Gender-specific and gender-independent models were used.

Model	Percentile band	
	0-10% (Resistant)	90-100% (Susceptible)
Elastic Net	**Sheep classified Resistant by algorithm and WEC / Total**	**Sheep classified Susceptible by algorithm and WEC / Total**
All Sex Combined animals	12 / 43	28 / 43
Male animals	4 / 22	4 / 22
Female animals	11 / 21	12 / 21
Lasso		
All Sex Combined animals	10 / 43	28 / 43
Male animals	3 / 22	2 / 22
Female animals	9 / 21	11 / 21
Bayesian Network		
All Sex Combined animals	6 / 43	23 / 43
Male animals	4 / 22	4 / 22
Female animals	4 / 21	11 / 21

Table 5
Heritabilities for WEC, blood parameters, and models.

Trait	Heritability
WEC	0.33 ± 0.12
WBC	0.15 ± 0.10
NEU	0.00 ± 0.05
LYM	0.37 ± 0.16
MONO	0.28 ± 0.14
EOS	0.33 ± 0.16
BASO	0.10 ± 0.09
RBC	0.47 ± 0.15
HGB	0.35 ± 0.14
HCT	0.27 ± 0.12
MCV	0.66 ± 0.16
MCH	0.58 ± 0.17
MCHC	0.63 ± 0.18
PLT	0.14 ± 0.12
BN scale	0.37 ± 0.15
Log_WEC	0.25 ± 0.12
Elastic	0.41 ± 0.14
Lasso	0.44 ± 0.14

consistent with our earlier study where the functions discriminated susceptible sheep at 100% accuracy and resistant sheep with 80% accuracy (Andronicos et al., 2014). The higher classification accuracy achieved by the earlier study may be attributed to the uniformity of response in the selection line sheep. More generally, the higher classification accuracy of susceptible sheep may be due to differences in the complexity of the response types. A susceptible animal can be accurately detected through loss of blood and anaemia whereas multiple cell types operating in a temporal fashion are required to successfully detect and eliminate the GIN burden in resistant sheep. Overall the combination of results suggests these blood based tests could be used as phenotypes in genetic evaluation in identifying and eliminating susceptible sheep from a breeding population.

5. Conclusion

In this study we have demonstrated that it is possible to develop machine learning algorithms that classify the ability of sheep to resist GIN infections based on blood parameters. Significantly GIN resistance status could be determined in outbred contemporary sheep. The parameters represented in the algorithms provide an insight into the biology of the resistance response and the observation that the resistance response differs between sexes is important. Further studies could determine if attempts to improve GIN resistance in sheep based solely on selection of WEC are impacted by this sex specific difference or if the approach could be enhanced by including a red blood cell based metric. The algorithms gave rise to aggregate trait descriptions, which were moderately heritable, and may therefore be suitable for use in genetic selection strategies. Further refinement is required before this type of test could be applied more broadly but the results are encouraging. It would also be desirable to validate this type of approach for other significant animal diseases.

Declaration of interest

The present study was funded by the CSIRO.

All sampling and animal handling performed in this trial was carried out in accordance to methods approved in CSIRO animal ethics committee applications EC 03/12, 03/20, 03/74, 04/07, 04/51, 06/01, and ARA 14/29.

CRediT authorship contribution statement

Amy Bell: Data curation, Formal analysis, Methodology, Writing - original draft, Writing - review & editing. **Jody McNally:** Data curation, Investigation, Methodology, Project administration, Writing - original draft, Writing - review & editing. **Daniel V. Smith:** Formal analysis, Validation, Visualization, Writing - original draft, Writing - review & editing. **Ashfaqur Rahman:** Formal analysis, Validation, Visualization, Writing - original draft, Writing - review & editing. **Peter Hunt:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing - original draft, Writing - review & editing. **Andrew C. Kotze:** Data curation, Formal analysis, Methodology, Writing - original draft, Writing - review & editing. **Sonja Dominik:** Formal analysis, Methodology, Validation, Visualization, Writing - original draft, Writing - review & editing. **Aaron Ingham:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing.

Acknowledgements

The authors would like to thank Yutao Li for review of the manuscript and Jen Smith for provision of sheep genotypic data from Trial 2. The data generated in Trial 1 was supported in part by SheepGenomics which is an initiative of Australian Wool Innovation Limited and Meat & Livestock Australia.

References

Abuargob, O., Stear, M.J., 2014. Differences between female and castrated male lambs in susceptibility to natural, predominantly Teladorsagia circumcincta infection. Vet. Parasitol. 205, 588–594.
Albers, G.A.A., Gray, G.D., Lejambre, L.F., Barger, I.A., Barker, J.S.F., 1990. The effect of hemonchus-contortus infection on hematological parameters in young merino sheep and its significance for productivity. Anim. Prod. 50, 99–109.
Andronicos, N.M., Henshall, J.M., Le Jambre, L.F., Hunt, P.W., Ingham, A.B., 2014. A one shot blood phenotype can identify sheep that resist Haemonchus contortus challenge. Vet. Parasitol. 205, 595–605.
Barger, I.A., 1993. Influence of sex and reproductive status on susceptibility of ruminants to nematode parasitism. Int. J. Parasitol. 23, 463–469.
Barlow, R., Piper, L.R., 1985. Genetic analyses of nematode egg counts in Hereford and crossbred hereford cattle in the subtropics of New South-Wales. Livest. Prod. Sci. 12, 79–84.
Beh, K.J., Hulme, D.J., Callaghan, M.J., Leish, Z., Lenane, I., Windon, R.G., Maddox, J.F., 2002. A genome scan for quantitative trait loci affecting resistance to

Trichostrongylus colubriformis in sheep. Anim. Genet. 33, 97–106.
Berton, M.P., Silva, R.M.D., Peripolli, E., Stafuzza, N.B., Martin, J.F., Alvarez, M.S., Gavina, B.V., Toro, M.A., Banchero, G., Oliveira, P.S., Eler, J.P., Baldi, F., Ferraz, J.B.S., 2017. Genomic regions and pathways associated with gastrointestinal parasites resistance in Santa Ines breed adapted to tropical climate. J. Anim. Sci. Biotechnol. 8.
Bishop, S.C., 2012. Possibilities to breed for resistance to nematode parasite infections in small ruminants in tropical production systems. Animal 6, 741–747.
Brown, D.J., Fogarty, N.M., 2016. Genetic relationships between internal parasite resistance and production traits in Merino sheep. Anim. Prod. Sci. 57, 209–215.
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.
Burrow, H.M., Henshall, J.M., 2014. Relationships between adaptive and productive traits in cattle, goats and sheep in tropical environments. Proceedings, 10th world congress of genetics applied to livestock production Vancouver Canada.
Charlier, J., van der Voort, M., Kenyon, F., Skuce, P., Vercruysse, J., 2014. Chasing helminths and their economic impact on farmed ruminants. Trends Parasitol. 30, 361–367.
Dominik, S., Hunt, P.W., McNally, J., Murrell, A., Hall, A., Purvis, I.W., 2010. Detection of quantitative trait loci for internal parasite resistance in sheep. I. Linkage analysis in a Romney x Merino sheep backcross population. Parasitology 137, 1275–1282.
Douch, P.G., Green, R.S., Morris, C.A., McEewan, J.C., Windon, R.G., 1996. Phenotypic markers for selection of nematode-resistant sheep. Int. J. Parasitol. 26, 899–911.
Garza, J.J., Greiner, S.P., Bowdridge, S.A., 2018. Ovine vital neutrophil extracellular traps bind and impair Haemonchus contortus L3 in a breed-dependent manner. Parasite Immunol. 40.
Gilmour, A.R., Gogel, B.J., Cullis, B.R., Thompson, R., 2009. ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.
Gordon, H.M., Whitlock, H.V., 1939. A new technique for counting nematode eggs in sheep faeces. J. Counc. Sci. Ind. Res. 12, 50–52.
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. ACM SIGKDD Explor. Newsl. 11, 10–18.
Kim, E.S., Sonstegard, T.S., Silva, M.V.G.B., Gasbarre, L.C., Van Tassell, C.P., 2014. Identification of quantitative trait loci affecting gastrointestinal parasite resistance in an experimental Angus population. Anim. Genet. 45, 117–121.
Mavrot, F., Hertzberg, H., Torgerson, P., 2015. Effect of gastro-intestinal nematode infection on sheep performance: a systematic review and meta-analysis. Parasit. Vectors 8.
May, K., Brugemann, K., Yin, T., Scheper, C., Strube, C., Konig, S., 2017. Genetic line comparisons and genetic parameters for endoparasite infections and test-day milk production traits. J. Dairy Sci. 100, 7330–7344.
Meeusen, E.N.T., Balic, A., Bowles, V., 2005. Cells, cytokines and other molecules associated with rejection of gastrointestinal nematode parasites. Vet. Immunol. Immunopathol. 108, 121–125.
Ortolani, E.L., Leal, M.L.D., Minervino, A.H.H., Aires, A.R., Coop, R.L., Jackson, F., Suttle, N.F., 2013. Effects of parasitism on cellular immune response in sheep experimentally infected with Haemonchus contortus. Vet. Parasitol. 196, 230–234.
Safari, E., Fogarty, N.M., Gilmour, A.R., 2005. A review of genetic parameter estimates for wool, growth, meat and reproduction traits in sheep. Livest. Prod. Sci. 92, 271–289.
Smith, J.L., Brewer, H.G., Dyall, T., 2009. Heritability and phenotypic correlations for breech strike and breech strike resistance indicators in merinos. AAABG (Barossa Valley). pp. 334–337.
Terefe, G., Grisez, C., Prevot, F., Bergeaud, J.P., Dorchies, P., Brunel, J.C., Francois, D., Fourquaux, I., Jacquiet, P., 2007. In vitro pre-exposure of Haemonchus contortus L3 to blood eosinophils reduces their establishment potential in sheep. Vet. Res. 38, 647–654.
Williams, A.R., Greeff, J.C., Vercoe, P.E., Dobson, R.J., Karlsson, L.J.E., 2010. Merino ewes bred for parasite resistance reduce larval contamination onto pasture during the peri-parturient period. Animal 4, 122–127.
Windon, R.G., 1996. Genetic control of resistance to helminths in sheep. Vet. Immunol. Immunopathol. 54, 245–254.