Accuracy of Genomic Prediction for Tick Resistance in Braford and Hereford Cattle

F.F. Cardoso[*]**, B.P. Sollero**[*]**, C. C. G. Gomes**[*]**, H. B. Comin**[*]**, V. M. Roso**[†]**, R. H. Higa**[‡]**, M. J. Yokoo**[*]**, A. R. Caetano**[§]**,
I. Aguilar**[#]

[*]Embrapa Southern Region Animal Husbandry, Bagé, [†]Gensys Associated Consultants, Porto Alegre, [‡]Embrapa Agriculture Informatics, Campinas, [§]Embrapa Genetic Resources & Biotechnology, Brasilia, Brazil, [#]National Agricultural Research Institute, Canelones, Uruguay

ABSTRACT: This work aimed to determine the accuracy and bias of genomic predictions of Braford (BO) and Hereford (HH) cattle genetic resistance to ticks. Repeated 10,673 tick counts were obtained from 3,435 BO and 928 HH cattle from Delta G Connection breeding program. A subset of 2,803 BO and 652 HH samples were genotyped and 41,045 markers remained after quality control. Log transformed records were adjusted by a pedigree repeatability model to estimate breeding values (EBV), subsequently used to obtained deregressed EBV. Data were split into five subsets for cross-validation using k-means and random clustering. Genomic predictions with moderate accuracies (0.38 to 0.60) were obtained by best unbiased linear prediction (GBLUP), BayesB and single step GBLUP indicating that, despite some bias, genomic selection could be used as practical tool to improve cattle genetic resistance to ticks.

Keywords: beef cattle, genomic selection, health, tick resistance

Introduction

One of the main animal health problems in tropical and subtropical cattle production is the *Rhipicephalus microplus* tick, which causes decreased performance, hide devaluation, and increased production costs with acaricide treatments and transmission of infectious diseases. Estimated losses due to ticks in Brazil reach two billion dollars annually (Grisi et al. (2002)). Higher susceptibility to this parasite also prevents the wide use of British breeds in crossbreeding systems to improve beef quality of zebu cattle in tropical regions. Nevertheless, the existence of genetic variability for tick resistance indicates that there is potential for improving this trait in cattle (Frisch (1999)). Breeding more resistant cattle would lead to lower use of acaricides and therefore reduce potential presence of chemical residues in beef, as well as the often observed development of parasite resistance to high cost acaricide treatments.

The availability of technologies for large scale genotyping and of methodologies to incorporate dense molecular marker information into genetic evaluation (Meuwissen et al. (2001); Aguilar et al. (2010)) has opened the possibility to select cattle for tick resistance without having to expose selection candidates to risky parasitism. The aim of this work was to determine the accuracy and bias of genomic predictions for genetic resistance to ticks in Braford (BO) and Hereford (HH) cattle using cluster cross-validation, and therefore to assess the utility of genomic selection to identify Braford and Hereford cattle more resistant to this parasite.

Materials and Methods

Animals and tick count data. Two or three subsequent tick counts at the whole left side were obtained between 2010 and 2013 from 3,435 BO and 928 HH cattle from eight herds of the Delta G Connection breeding program, totaling 10,673 records. During post-weaning animals were kept under the same feeding and sanitary management conditions, observed daily in their paddocks, and tick counts performed when average individual infestation exceeded 20 engorged tick females with at least 4.5 mm in diameter. Average age during the evaluation period was 524 ± 65 days and average count was 35 ± 42 ticks. Pedigree information recovered from historical breeding records comprised 12,056 animals and was highly incomplete due to use of multiple-sires matings, which resulted in 65% of the animals with tick count observations having unknown paternity.

Genotypes. A subset of 3,545 phenotyped animals was genotyped with the Illumina BovineSNP50 BeadChip (50K), and 131 of their sires were genotyped with the Illumina High-Density Bovine BeadChip Array (HD). Genotype quality control was implemented using R/SNPStats package (Clayton (2012)) to remove samples with call rates <0.90, heterozygosity deviations >3.0 standard deviations, mismatching sex and duplicated records. Only single nucleotide polymorphisms (SNPs) mapped to autosomes, with call rates >0.98, minor allele frequencies (MAF) >0.03 or not highly significant deviations from Hardy-Weinberg equilibrium (P>10^{-7}) were considered for further analyses. Moreover, for SNPs in the same position or highly correlated (r>0.98), only the one with highest MAF was retained. The HD panel was filtered to select only SNPs that were also present in the 50K panel. After editing, a total of 41,045 SNP markers (78%) and 3,586 samples (98%) remained, including 131 sires, 2,803 BO and 652 HH yearling bulls, steers and heifers with tick count records. Sporadically missing genotypes were imputed using FImpute software (Sargolzaei et al. (2011)).

Statistical analyses. Initially, estimated breeding values (EBV) were obtained by adjusting a pedigree based repeatability model to base 10 logarithm of tick counts + 1, considering the fixed effect of contemporary groups (CG - same farm, sex, year and season of birth, management, and tick count date); linear coefficients for zebu breed

proportion and heterozygosity; linear and quadratic coefficients for animal age at tick counting, and random animal additive genetic, permanent environment and residual effects. Contemporary groups with less than five individuals and animals that were 3.5 standard deviation above or below the mean tick count of their CG were excluded from the analyzes. Covariance components and genetic parameters for tick counts were estimated using Bayesian inference through the Gibbs2f90 software (Misztal et al. (2002)). The approach proposed by Garrick et al. (2009) was used to calculate deregressed EBV (DEBV) and corresponding weights, which were then used to estimate marker effects by best unbiased linear prediction (GBLUP) and BayesB methods (Meuwissen et al. (2001)). BayesB method was implemented with R/BGLR package (de los Campos & Perez (2013)) setting the fraction of marker with null effect at 95% (π=0.95). Direct genomic breeding values (DGV) were obtained by the sum over all markers of the genotype covariate value times the estimated SNP effect, represented by their posterior means in BayesB. Alternatively, genomic enhanced breeding values (GEBV) were obtained by directly combining phenotypic, genomic and pedigree information in the single step method (ssGBLUP) of Misztal et al. (2009), which together with GBLUP were implemented using the Blupf90 software family (Misztal et al. (2002)). Missing parentage information was handled in ssGBLUP by treating the sire of animals with uncertain paternity as unknown.

Cross-validation and prediction accuracy. Utility of genomic predictions to select young animals without phenotypes for tick resistance was assessed by cross-validation. The 3,455 animals with genotype and phenotype for tick counts were divided into five groups by two strategies using R (R Core Team, 2013): by K-means clustering based on marker relationship distance or at random, to respectively characterize genomic selection scenarios where target animals are closely or more distantly related to the training set. Average genomic relationship (g_{ij}) based on the VanRaden et al. (2009) standardization of each animal within and between groups were calculated to characterize relatedness between training and validation sets (Saatchi et al. (2011)). For each grouping strategy, five-fold cross-validation was carried out by alternatively using records of four groups as training set to derive genomic predictions for the fifth (validation) group whose data was omitted in the analysis. Prediction accuracies were derived from genetic correlations between tick counts and DGV/GEBV of genotyped animals from the five validation sets, estimated in a bivariate animal model using a pedigree-based numerator relationship matrix, except that covariances between individuals in different groups were zeroed out (Saatchi et al. (2013)). Prediction abilities were also assessed by Pearson correlations between EBV on DGV/GEBV and regression coefficients of EBV on DGV/GEBV.

Results and Discussion

Clustering. K-means clustering based on genomic relationships (Table 1) yielded highly unbalanced group sizes and, as expected, average g_{ij} was larger within than between groups. Group 1 contained almost exclusively Hereford cattle and had the greatest inbreeding coefficient and genetic dissimilarity to the other four groups, which in turn essentially included outbred Braford animals with about 1/3 of zebu proportion. These results indicated that the K-means clustering partitioned the genotyped populations by two factors: breed composition first (Group 1 vs. Groups 2 to 5) and then Braford animals into more related subgroups also with different degrees of inbreeding (Table 1). On the other hand, random groups were balanced with 691 animals each, had similar average inbreeding (0.01 ± 0.05) and breed composition (30% zebu proportion), and average relationship within and between groups close to zero, due to centering of the genomic relationship matrix (VanRaden et al. (2009)).

Table 1. Number of individuals (N) and averages (± SD) of zebu proportion, inbreeding coefficient, and within and between group genomic relationship (g_{ij}) for K-means clustering groups.

Group	1	2	3	4	5
N	629	230	1211	471	914
Zebu proportion	0.02	0.37	0.35	0.34	0.35
Inbreeding coefficient	0.09 ±0.03	0.03 ±0.03	-0.01 ±0.03	-0.03 ±0.04	0.00 ±0.03
g_{ij} within group	0.14 ±0.04	0.07 ±0.05	0.004 ±0.03	0.01 ±0.04	0.02 ±0.03
g_{ij} between group	-0.03 ±0.04	-0.005 ±0.05	-0.003 ±0.03	-0.002 ±0.03	-0.01 ±0.04

Accuracy and bias of genomic selection. Genetic correlations between EBV for tick counts and cross-validation genomic prediction were of moderate magnitude, ranging between 0.38 and 0.48 for K-means clustering and between 0.44 and 0.60 for random clustering, depending on the genomic prediction method (Table 2). All genomic methods were superior to pedigree BLUP (PBLUP), with highest increase in genetic correlation obtained by ssGBLUP (71% for K-means and 114% for random groups), which combines genomic and pedigree information, followed by BayesB (39% for K-means and 89% for random clustering) and GBLUP (36% for K-means and 57% for random clustering). Conversely, GBLUP and ssGBLUP yield over dispersed predictions with EBV regression coefficients below 1, while BayesB had under dispersed DGVs in K-means cross-validation (Table 2). This indicates potential bias on genomic selection for tick resistance based on these methods. Heritabilities of genomic and pedigree cross-validation predictions were all above 0.90, except for BayesB in the K-means clustering scenario (Table 2). Heritabilities below perfect genetic determination value of 1 were also observed in U.S. Angus (Saatchi et al. (2011)) and Hereford (Saatchi et al. (2013)) cattle and may be due to differences between genomic and

pedigree relationships within cluster, especially in our incomplete pedigree data.

Table 2. Heritability (h²), genetic correlation with tick counts (r_g), and regression coefficients (β)§ of cross-validation genomic predictions¥ using different methods.

Method	K-means clustering			Random clustering		
	h^2	r_g	β	h^2	r_g	β
BayesB	0.75	0.39	1.46	0.92	0.53	0.83
	±0.03	±0.06		±0.01	±0.06	
GBLUP	0.90	0.38	0.29	0.93	0.44	0.36
	±0.02	±0.07		±0.02	±0.07	
PBLUP	0.92	0.28	1.03	0.92	0.28	0.99
	±0.01	±0.07		±0.01	±0.07	
ssGBLUP	0.91	0.48	0.34	0.96	0.60	0.25
	±0.02	±0.07		±0.01	±0.08	

§Estimated regression slope of pedigree estimated breeding value (EBV) using full data on cross-validation genomic predictions.
¥Direct genomic values for BayesB and genomic BLUP (GBLUP), EBV for pedigree BLUP (PBLUP) and genomic enhanced breeding values for single step genomic BLUP (ssGBLUP).

None of the studied methods excels in all evaluated criteria; therefore, further research to reduce bias of ssGBLUP GEBV from multi-breed populations, for example, using population specific allele frequencies to scale genomic relationship (Simeone et al., 2012) or to develop blending strategies to combine BayesB DGV with PBLUP parent averages, could yield more accurate genomic predictions to select young Hereford and Braford cattle for tick resistance.

Pearson correlations between EBV and genomic prediction were very low for Group 1 under the K-means clustering scenario for all tested genomic methods (Figure 1). This indicates that a training set composed of just Braford cattle would not yield accurate prediction for Hereford candidates, despite the latter breed contribution an average 62.5% of the Braford genome.

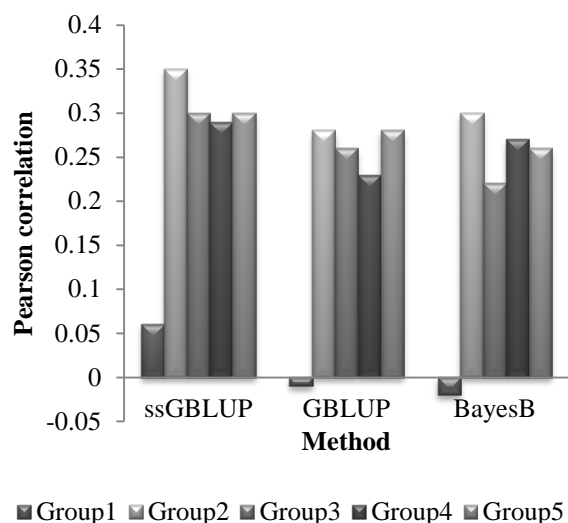■ Group1 ■ Group2 ■ Group3 ■ Group4 ■ Group5

Figure 1. Pearson correlations between estimated breeding values and genomic predictions obtained by genomic best unbiased linear prediction (GBLUP), single step GBLUP (ssGBLUP) and BayesB methods for each K-means clustering cross-validation group.

Conclusion

Moderate accuracy values found in this study indicate that genomic predictions could be used as a practical tool to improve genetic resistance to ticks and in the development of resistant lines of Braford and Hereford cattle.

Literature Cited

Aguilar, I., Misztal, I., Johnson, D. L. et al. (2010) J. Dairy Sci., 93: 743-752.

Clayton, D. (2012). R/snpStats package version 1.11.0.

de los Campos, G., Perez, P. (2013). R/BGLR package version 1.0.2.

Frisch, J. E. (1999) Int. J. Parasitol. 29:57-71.

Garrick, J.D., Taylor, J. F. and Fernando, R. L. (2009). Genet. Sel. Evol. 41:55.

Grisi , L.; Massard, L. C.; Borja, G. E. M.; et al. (2002). Hora Vet. 125:8-10.

Meuwissen, T. H. E., Hayes, B. J. and Goddard, M. E. (2001). Genetics 157:1819–1829.

Misztal, I., Legarra, A. and Aguilar, I. (2009). J. Dairy Sci. 92:4648–4655.

Misztal, I., Tsuruta, S., Strabel, T., et al. (2002). in Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France. Comm. 28–07

R Core Team (2013). R Found. Stat. Comp. http://www.R-project.org/.

Saatchi, M., Ward, J. and Garrick, D. J. (2013) J. Anim. Sci. 91:1538-1551.

Saatchi, M., McClure, M. C., McKay, S. D. et al. (2011). Genet. Sel. Evol. 43:40.

Sargolzaei, M., Chesnais, J. P., and Schenkel, F. S. (2011). J. Anim. Sci. 89, E-Suppl. 1 / J. Dairy Sci. 94, E-Suppl. 1: 421 (333).

Simeone, R., Misztal, I., Aguilar, I. et al. (2012) J Anim Breed Genet. 129:3-10.

VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R. et al. (2009). J. Dairy Sci. 92:16–24.