Tag-SNP selection using Bayesian genome-wide association study for growth and adaptation traits in Hereford and Braford cattle

G. S. Campos[1], F. A. Reimann[1], V. S. Junqueira[2], J. Braccini[3], L. L. Cardoso[1,4], M. J. I. Yokoo[4], B. P. Sollero[4], C. C. Gulias-Gomes[4], A. A. Boligon[1], A. R. Caetano[5], F. F. Cardoso[1,4]

[1]Departmento Zootecnia, Universidade Federal de Pelotas, Pelotas, RS, Brazil
[2]Departmento de Zootecnia, Universidade Federal de Viçosa, Viçosa, MG, Brazil
[3]Departmento de Zootecnia, Universidade Federal do Rio Grande do Sul, RS, Brazil
[4]Embrapa Pecuária Sul, Bagé, RS, Brazil.
fernando.cardoso@embrapa.br (Corresponding Author)
[5]Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, Brazil

Summary

The aim of this study was to perform a Bayesian genome-wide association study (GWAS) to identify genomic regions associated with growth and adaptation traits in Hereford and Braford cattle, and to select Tag-SNPs to represent these regions in low-density panels useful for genomic predictions. Phenotypic data from 126,290 animals from the Conexão Delta G genetic breeding program and a set of 3,545 animals and 131 sires genotyped with the Illumina BovineSNP50 and HD chips, respectively, were used. Using BayesB (π=0.995) method, Tag-SNPs were selected according to parameters such as window variance, model frequency, t-like statistic, linkage disequilibrium and minor allele frequency to compose low-density panels. Estimated cross-validation accuracies for growth traits were obtained by calculating genetic correlations between observed phenotypes and direct genomic values. Based on BayesA and Tag-SNPs, these accuracy values ranged from 0.13 to 0.30 for k-means and 0.36 to 0.65 for random clustering of animals to compose validation groups. For adaptation traits, observed genetic correlations ranged from 0.18 to 0.42 and 0.33 to 0.61 for k-means clustering and random, respectively. Although genomic prediction accuracies were higher with the full marker panel, predictions with low-density panels retained on average 76% of the accuracy obtained with BayesB for growth traits and 64% for adaptation traits. The proposed Tag-SNP panels may be useful for future functional enrichment and fine mapping studies and for lower-cost commercial genomic predictions.

Keywords: beef cattle, genomic prediction, gwas, low-density panel

Introduction

Genome-wide association studies (GWAS) enable the identification of single nucleotide polymorphisms (SNPs) associated with traits of interest and to explore regions of the genome where these SNPs are located in search for quantitative trait loci (QTL) and/or genes related to the phenotypic expression of the respective trait. Bayesian regression methods, which combine a priori information on marker effects with phenotypes related to traits of interest, can be used to concurrently estimate all effects from available markers (Meuwissen et al., 2001). The identification of the most informative markers associated with traits of interest may lead to the design of low-density marker panels that explain a substantial part of the genetic variance for these traits. If a good predictive performance is proven, these panels can

become highly desirable as lower-cost solutions for commercial applications (Van Eenennaam et al., 2014).

The aim of this study was to perform a GWAS using Bayesian methodology to identify genomic regions associated with growth and adaptation traits in Hereford and Braford cattle and to select Tag-SNPs representative of these regions. Another objective was to verify the prediction accuracies of the proposed low-density panels and compare them with results from the complete set of SNP markers.

Material and methods

Phenotype, Genotype and Pedigree data. Phenotypic records from 126,290 Hereford and Braford animals from the Conexão Delta G breeding program were used. The studied traits were: birth weight (BW), weaning weight adjusted to 205 days of age (WW205), yearling weight adjusted to 550 days age (YW550), post weaning weight gain adjusted for 345 days of age (PWG345), eye pigmentation (EP), and hair coat at weaning (HCW) and at yearling (HCY). Data from 3,545 animals genotyped with the Illumina Bovine50K panel and 131 bulls genotyped with the BovineHD chip were used. After quality control, 41,045 SNPs and 3,592 samples were used, including 2,934 Braford and 658 Hereford animals.

Genome wide association study (GWAS) using Bayesian inference. The approach proposed by Garrick et al. (2009) was used to calculate deregressed breeding values (DEBV). These DEBVs were then analyzed in a mixed effects model with SNP allelic substitution effects based on BayesA and BayesB methods (Meuwissen et al., 2001) using GenSel software version 4.0 (Fernando & Garrick, 2009). For BayesB the parameter π was assumed to be equal to 0.995.

The most informative markers representing windows that explained the largest proportions of genetic variance (potential QTLs) related to the trait were first selected according to parameters that reflect: 1) the chains proportion that included a certain marker in the model (model frequency) and 2) the marker mean posterior effect based only on the chains that included the particular marker in the model divided by the standard deviation of these effects (t-like statistics). Subsequently, the final selection considered the linkage disequilibrium (LD) between pairs of SNPs and minor allele frequency (MAF) values. See Sollero et al. (2017) for additional details of this method.

Prediction accuracy of low-density SNP panels. Genotyped animals were divided into four or five groups by two cross-validation strategies using the R software: according to genomic relationships by k-means clustering and randomly. Combined prediction accuracies across all groups were derived as the genetic correlation between observed phenotypes (y) and direct genomic values (DGV) for a given trait, estimated in a bivariate animal model, using a numerator relationship matrix where animals were assumed to be unrelated between clusters (A*). DGV prediction accuracies for the low-density panels and the full set of markers were estimated using BayesA and BayesB ($\pi = 0.995$), respectively.

Results and discussion

Genome wide association study and Tag-SNP. The BayesB method including all animals and markers was used in the analysis with the parameter π of 0.995 (Zare et al., 2014), which corresponds to 0.5% of the SNPs receiving a non-zero effect in the model at each iteration. A

large π value was chosen to allow only regions with stronger associations to be identified (Zare et al., 2014). Estimated heritabilities (h^2) for each trait were: 0.22 (BW), 0.14 (WW205), 0.28 (YW550), 0.16 (PWG345), 0.58 (HCW), 0.66 (HCY) and 0.81 (EP), respectively. The most representative windows observed in the GWAS results (BayesB π = 0.995) were selected considering a threshold of 0.2% of the genetic variance explained by each window.

For growth traits, the number of windows selected ranged from 55 to 75 and the number of SNPs ranged from 1,008 to 1,369 in all selected windows, which accounted from 28.8% to 37.43% of the total genetic variance. For adaptation traits, the number of windows varied from 27 to 104. The traits with the lowest and highest number of markers in the select windows were BC (468 SNPs) and EP (1,841 SNPs), respectively. Application of additional selection criteria to identify Tag-SNPs for each window resulted in low-density panels with between 78 and 103 SNPs for growth traits and 40 to 159 for adaptation traits.

Prediction accuracy of Tag-SNP panels. Cross-validation accuracies derived from estimated genetic correlations between each phenotype and DGVs obtained with BayesA using Tag-SNPs (BayesA_Tag-SNP) and the full marker panel (41,045) with BayesB (π = 0.995), were greater for all traits when groups were formed randomly in relation to k-means (Table 1).

Table 1. Genetic correlations for growth and adaptation traits between observed phenotypes and direct genomic value predictions from cross-validation using different methods.

		Methods[2]		Tag-SNP *vs.* all
Traits[1]	Cluster	BayesA_Tag-SNP	BayesB	SNP[3]
BW	k-means	0.30±0.08	0.36±0.09	83%
	random	0.65±0.20	0.74±0.16	88%
WW205	k-means	0.20±0.06	0.25±0.05	80%
	random	0.30±0.08	0.42±0.10	71%
YW550	k-means	0.21±0.07	0.30±0.09	70%
	random	0.50±0.23	0.67±0.15	75%
PWG345	k-means	0.13±0.09	0.20±0.11	65%
	random	0.36±0.24	0.46±0.23	78%
HCW	k-means	0.18±0.03	0.33±0.04	55%
	random	0.33±0.06	0.66±0.06	50%
HCY	k-means	0.26±0.10	0.39±0.15	67%
	random	0.55±0.09	0.68±0.11	81%
EP	k-means	0.42±0.11	0.75±0.10	56%
	random	0.61±0.12	0.78±0.10	78%

[1]BW: birth weight; WW205: weaning weight adjusted for 205 days of age; YW550: yearling weight adjusted for 550 days of age; PWG345: post-weaning gain adjusted for 345 days of age; HCW: hair coat at weaning; HCY: hair coat at yearling; and EP: eye pigmentation.
[2]BayesA_Tag-SNP: Bayesian model with t-distribution and probability (π = 0), using only more informative markers; BayesB: Bayesian mixture of t-distribution and point mass on zero with probability (π = 0.995), using all markers (41,045).
[3]Tag-SNP *vs.* all SNP: % of the full SNP panel accuracy retained by the tag-SNP panel.

For growth traits, genetic correlations ranged from 0.13 to 0.30 for groups generated by *k*-means clustering and from 0.36 to 0.65 for groups formed at random, using Tag-SNPs

(Table 1). The greatest genetic correlation values were observed for BW: 0.30 for *k*-means and 0.65 for random clustering. For adaptation traits, values ranged from 0.18 to 0.42 for k-means and from 0.33 to 0.61 for random clustering (Table 1). The largest observed value was for EP, of 0.42 for k-means and 0.61 for random clustering. For most traits, higher accuracy values were found using the full marker panel vs Tag-SNPs. Prediction accuracies are expected to decrease as marker panel density decreases because of the lower expected linkage disequilibrium between highly dispersed markers and QTLs affecting the trait of interest (Habier et al., 2009).

Accuracies observed for the proposed low-density panels for growth traits ranged from 65% to 83% (average: 76%)in comparison to accuracies obtained with the full marker panel. For BC, Tag-SNP accuracy corresponded on average to 64% (between 50 and 81%) in comparison to accuracies obtained with the full marker panel, including k-means and random cross-validations.

Obtained results indicate the applied strategy can be successfully used to select informative SNP markers which could be used to design low-density panels and eventually reduce genotyping costs associated with genomic predictions, while retaining suitable accuracies in comparison to full density marker panels. Low-density panels may provide a viable alternative for use of genomic selection for genetic improvement of profitability-associated traits in commercial beef cattle herds.

Conclusion

The BayesB method was observed to be adequate for the selection of SNPs based on the model frequency and t-like parameters. Genomic predictions obtained with the proposed low-density panels using BayesA presented moderate prediction accuracies for growth and adaptation traits. These panels may be useful for future functional enrichment and fine mapping studies and for lower-cost commercial applications of genomic prediction.

List of References

Fernando, R. L. & D. J. Garrick, 2009. GenSel-user manual. Mapping Genes for Complex Traits in Domestic Animals and Their Use in Breeding Programmes, 3rd Edn. Version 2.

Garrick, D. J., J. F. Taylor & R.L. Fernando, 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet Sel Evol, 41:55.

Habier, D., R. L. Fernando & J. C. M Dekkers, 2009. Genomic selection using low-density marker panels. Genetics, 182: 343-353.

Meuwissen, T. H. E., B. J. Hayes & M. E. Goddard, 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157:1819-1829.

Sollero, B. P., V. S. Junqueira, C. C. Gomes, A. R. Caetano & F. F. Cardoso, 2017. Tag SNP selection for prediction of tick resistance in Brazilian Braford and Hereford cattle breeds using Bayesian methods. Genet Sel Evol, 49(1):49.

Van Eenennaam, A. L., K. A. Weigel, A. E. Young, M. A. Cleveland & J. C. Dekkers, 2014. Applied Animal Genomics: Results from the Field. Annu Rev Anim Biosci, 2(1): 105-139.

Habier, R. Fernando, D. J. Garrick, W. G. Hill & J. C. M. Dekkers, 2012. Genome-wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens. Anim Genet, 43(1): 87-96.

Zare, Y., G. E. Shook, M. T. Collins & B. W. Kirkpatrick, 2014. Genome-wide association analysis and genomic prediction of Mycobacterium avium subspecies paratuberculosis

infection in US Jersey cattle. PLoS One, 9(2):e88380.