G3: Genes|Genomes|Genetics Early Online, published on April 8, 2017 as doi:10.1534/g3.117.041442

Genome-enabled prediction of breeding values for feedlot average daily weight gain in Nelore cattle[1]

Adriana L. Somavilla[*], Luciana C. A. Regitano[†], Guilherme J. M. Rosa[§*], Fabiana B. Mokry[‡], Mauricio A. Mudadu[†], Polyana C. Tizioto[‡], Priscila S. N. Oliveira[‡], Marcela M. Souza[‡], Luiz L. Coutinho[I], Danísio P. Munari[*,2]

[*]Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal, Brazil; [†]Embrapa Pecuária Sudeste, São Carlos, Brasil; [§]University of Wisconsin-Madison, Madison, WI, USA; [‡]Departmento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, Brasil; [I]Departmento de Zootecnia, Universidade de São Paulo, Piracicaba, Brasil

Abstract: Nelore is the most economically important cattle breed in Brazil, and the use of genetically improved animals has contributed to increase beef production efficiency. The Brazilian beef feedlot industry has grown considerably in the last decade, so the selection of animals with higher growth rates on feedlot has become quite important. Genomic selection could be used to reduce generation intervals and improve the rate of genetic gains. The aim of this study was to evaluate the prediction of genomic estimated breeding values for average daily gain in 718 feedlot-finished Nelore steers. Analyses of three Bayesian model specifications (Bayesian GBLUP, BayesA, and BayesCπ) were performed with four genotype panels (Illumina BovineHD BeadChip, TagSNPs, GeneSeek High and Low-density indicus). Estimates of Pearson correlations, regression coefficients, and mean squared errors were used to assess accuracy and bias of predictions. Overall, the BayesCπ

[1] This study was funded by the Brazilian Agricultural Research Corporation (EMBRAPA) and Sao Paulo State Research Foundation (FAPESP) scholarship no. 2012/23702-8, scholarship no. 2013/21644-3, and grant no 2012/23638-8

[2] Corresponding author: danisio@fcav.unesp.br

model resulted in less biased predictions. Accuracies ranged from 0.18 to 0.27, which are reasonable values given the heritability estimates (from 0.40 to 0.44) and sample size (568 animals in the training population). Furthermore, results from *Bos taurus indicus* panels were as informative as those from Illumina BovineHD, indicating that they could be used to implement genomic selection at lower costs.

Keywords: Genomic selection, *Bos taurus indicus*, growth, feedlot performance

Introduction

Brazil has the world's second largest cattle herd with over 200 million heads (Instituto Brasileiro de Geografia e Estatística 2013), with the Nelore (*Bos taurus indicus*) being the most widespread and economically important breed. As the total pasture area in Brazil has decreased over the decades, productivity gains have become an important factor for beef production (Martha *et al.* 2012). The Nelore breed has been selected for growth rate traits on pasture based on traditional pedigree and phenotypes analysis, however, the Brazilian beef feedlot industry has grown about 50% in the last decade (Millen *et al.* 2011), and novel breeding objectives and criteria are required.

In this context, the application of technologies to improve animal performance and thus to supply genetically improved animals for both pasture and feedlot systems are a critical factor to overcome the challenge of increasing the Brazilian beef production efficiency.

Nowadays, exploring the availability of technology to genotype thousands of single nucleotide polymorphisms (SNP) distributed across the genome, allows the application of genomic selection (GS). Phenotypic and SNP data information are then combined to predict genomic estimated breeding values (GEBV) earlier in the life of the

47 animals (Meuwissen *et al.* 2001). It has been argued that GS could lead to a decrease in
48 generation interval, and improvement of the rate of genetic gain (Schaeffer 2006), and also
49 assist the better control of inbreeding rates (Daetwyler *et al.* 2007).

50 Based on the importance of the Nelore cattle in Brazil and the increasing use of
51 feedlot systems, it is necessary to identify appropriate methodologies that allow genomic
52 selection of animals with higher growth rates on feedlots. The aim of the current study was
53 to compare different regression models and SNP panels in terms of accuracy, bias and
54 precision of genomic estimated breeding values for average daily weight gain (ADG) in
55 feedlot-finished Nelore steers.

56

57 **Material and methods**

58 Samples

59 During the mating seasons of 2006/07 through 2008/09, 804 steers, offspring of 34
60 Nelore bulls from 17 lineages, chosen to represent the genealogies of the Nelore breed in
61 Brazil, were generated through fixed-time artificial insemination in five farms. They were
62 raised to 21 months of age and then moved to either the Embrapa Southeast Livestock
63 (São Carlos - SP, Brazil) or the Embrapa National Beef Cattle Center (Campo Grande -
64 MS, Brazil) during three seasons in feedlot experiment periods (2009, 2010 and 2011).
65 Animals were fed with a total mixed ration (TMR) diet with 13% crude protein and 71%
66 total digestible nutrients (dry matter basis, corn or sorghum, soybean meal, soybean hull,
67 cotton seed, limestone, mineral mixture, urea, and monensin). The diet was provided twice
68 a day in which the feed offered (total mixture composed by concentrate:silage, 40:60 ratio)
69 was adjusted daily *ad libitum*. The animals were weighed every 14 days without fasting,
70 for an average period of 91 days. Steer rearing and sample collection protocols were

71 approved by Animal Care and Use Committee from the Embrapa Southeast Livestock
72 (São Carlos, Brazil).
73
74 Phenotype and genotype datasets
75 The initial dataset consisted of 7,236 weighting records from the 804 steers, but
76 only those from the 15th up to 77th days in feedlot were considered to estimate ADG, to
77 disregard the first weight and also because after this period more than 30% of the animals
78 had already been slaughtered. A linear regression analysis of live weight over time was
79 performed using the remained 3,523 records from 803 steers, using the lm function of the
80 R software (R Development Core Team 2014). The slope was used as the ADG during the
81 feedlot period for the purpose of considering only the linear weight gain and avoiding
82 comparison with different feedlot period lengths.
83 Steers were assigned to 39 contemporary groups (CG) containing from 5 to 42
84 animals, which combined information on mating season (3 levels), experimental feedlot (2
85 levels) and slaughter group (32 levels of animals slaughtered in the same week). After
86 that, the phenotype and genotype datasets were merged to ensure that they had the same
87 individuals. The summary of age at feedlot entry, starting weight, ADG and days in
88 feedlot on the remaining animals are presented in Table 1.
89 There were in total 780 steers and 34 bulls genotyped with the Illumina BovineHD
90 BeadChip (Illumina, San Diego, CA). The initial dataset contained 742,906 markers, in
91 which unplaced, mitochondrial and sex-linked SNP were first discarded, as well as
92 duplicated markers (e.g. two different names and positions for the same SNP). SNP were
93 also filtered based on two other panels: GeneSeek Genomic Profiler (GGP) HDi 80K and
94 GGP LDi 20K (Gene Seek Inc., Lincoln, NE). The panels were built specifically for *Bos*

95 *taurus indicus* breeds. Originally, the GGP HDi 80k/LDi 20k contained 74,085/19,721
96 markers, of which 69,942/18,464 were available in the primary dataset.

97 Paternity correction and quality control (QC) were performed to improve results.
98 To deal with SNP presenting significant deviation from the Hardy-Weinberg Proportions
99 (HWP) deviation, we checked plots of HWP versus percentage of heterozygous, and 17
100 SNP with more than 80% of heterozygous were excluded from the three datasets because
101 they could reflect an error during the genotyping procedure (Ziegler 2009). Quality control
102 was performed using the R package SNPtats (Clayton 2012). SNPs were kept for further
103 analysis only if they had call rate > 98% and minor allele frequency (MAF) > 1%. The
104 MAF filter excluded 20.0, 1.9 and 7.3% of the total SNP from the 770k, HDi, and LDi
105 panels, respectively.

106 After QC, the Beagle v.3.3.2 (Browning and Browning 2009) software was used
107 for phase inference and imputation of missing genotypes for each SNP panel. Finally, to
108 constitute a fourth SNP panel scenario, Tagger (Bakker *et al.* 2005), which is based on
109 linkage disequilibrium (LD) between markers (r^2), was used. This tool estimates the r^2
110 between all SNP pair and then selects a minimal set (TagSNPs) of markers with a $r^2 \geq 0.3$
111 with at least one another marker on the same chromosome. We have chosen this threshold
112 because it is the overall average r^2 at the distance of 10kb to 25kb, obtained in a previous
113 analysis of the same animals (Mudadu *et al.* 2016). The final number of SNP was 15,863;
114 63,945; 82,933 and 534,787 for the LDi, HDi, TagSNP and 770k panels, respectively.
115
116 Fixed effects modeling and adjusted phenotypes

117 The adjusted phenotype ($\hat{y}$) was represented as $\hat{y} = y - 1\hat{\mu} - W\hat{\alpha}$, in which y is
118 the vector of observations, $\hat{\mu}$ is the overall mean, W is an incidence matrix for fixed

119 effects (CG and animal age at feedlot entry) and $\hat{\alpha}$ is the vector of fixed effects estimates.
120 A residual analysis was performed at this point and animals with the normalized residuals
121 with absolute values larger than 3.5 were removed, thus 718 steers remained into the
122 dataset.
123
124 Models for genomic-enabled prediction
125 Three specifications were considered for building genome-enabled prediction
126 models: BayesA, BayesCπ and Bayesian GBLUP. The R package BGLR (de los Campos
127 and Rodriguez 2014) was used to fit the models, a flat (non-informative) prior was
128 assigned to the intercept. For the BayesA method, a normal distribution was assigned to
129 the marker effects, $\beta_j \sim N\left(0, \sigma^2_{\beta j}\right)$, where $j = (1, \ldots, p)$, p is the number of SNPs, and $\sigma^2_{\beta j}$
130 is the individual variance for the SNP effect. In a second level of hierarchy, each $\sigma^2_{\beta j}$ was
131 assigned independent and identically distributed (iid) Scaled-inverse Chi-square density,
132 with degrees of freedom (df_β) set to 5 and scale parameter (S_β) treated as unknown,
133 following a Gamma distribution with shape (s) and rate (r) parameters. The parameter s
134 was set to s=1.1 and r was solved so that 80% of proportion of the variance of the response
135 was attributed the linear predictor. On this model, the prior marginal distribution of marker
136 effects is a scaled-t density, with parameters df_β and S_β (Rosa *et al.* 2003).
137 For the BayesCπ model, the prior for each marker effect was an iid mixture of
138 point of mass (1- π) at zero (spike) and a slab that follows a Gaussian distribution,
139 $\beta_j \sim N\left(0, \sigma^2_\beta\right)\pi$, where σ^2_β is the common variance for the SNP effects. The additional
140 parameter π represents the prior proportion of non-zero effects and was treated as an
141 unknown, with a Beta prior distribution $\pi \sim Beta(p_0, \pi_0)$, with $p_0 > 0$ and $\pi_0 \in [0,1]$.
142 The parameters were set to $p_0 = 2$ and $\pi_0 = 0.5$, which give a uniform prior in the

143 interval $[0,1]$. Thus, differently from BayesA, BayesCπ sets some SNP effects to zero,
144 within a variable selection framework.

145 The Bayesian GBLUP (BGBLUP) model was implemented as a Bayesian
146 Reproducing Kernel Hilbert Spaces (RKHS) regression (de los Campos *et al.* 2009), using
147 a single kernel, user-defined (co)variance matrix K. The vectors of additive random effects
148 were assigned multivariate normal priors, $u \sim N(0, K\sigma_u^2)$, in which $\sigma_u^2 \sim \chi^{-2}(S, df)$ and
149 K was set as a marker-derived relationship matrix G, built as the first method proposed by
150 VanRaden (2008). Briefly, let M_{nxm} be a genotype matrix with n (number of samples)
151 rows and m (number of SNPs) columns, Z_{nxm} be the centered M matrix, and $G =$
152 $\frac{zz'}{2\sum p_j(1-p_j)}$, where the denominator is the total variance across loci. The degrees of
153 freedom (df) was set to 5 and the scale parameter (S) was solved so that 80% of
154 proportion of the variance of the response was attributed the linear predictor.

155 The number of iterations, burn-in and thinning interval parameters were
156 graphically evaluated and were different for each model (Table 2), and the length of the
157 chain used to compute posterior statistics was 25,000, 20,000, and 10,000 for BayesA,
158 BayesCπ, and BGBLUP, respectively. For BayesA and BayesCπ, the marker-based
159 genetic variance (σ_g^2) was computed as the sum of the variance explained by each SNP
160 marker $(\sigma_{\beta j}^2)$, while for BGBLUP the genetic variance was equal to σ_u^2. For the three
161 models, the narrow sense heritability was estimated as: $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_e^2)$, where σ_e^2 is
162 the residual variance.

163

164 Validation

The dataset was divided into training (animals from seasons 1 and 2) and testing (animals from season 3) subgroups, which contained 568 and 150 animals, respectively. For the BayesA and BayesCπ models, the GEBV on the testing set was defined as $GEBV_{i(tst)} = \sum_{j=1}^{p} g_{ij}\hat{\beta}_{trn}$, where g_{ij} is the genotype of the j[th] SNP on the i[th] animal and $\hat{\beta}_{trn}$ is the vector of the SNP marker effect estimated on the training set. For Bayesian GBLUP, phenotypes of testing subgroup were set as missing and samples of u were obtained in each iteration from the posterior distribution $[u, \sigma_u^2, \sigma_e^2|\hat{y}]$.

The correlation between GEBV and adjusted phenotype of animals on testing subgroup, $r(GEBV_{i(tst)}, \widehat{y_{i(tst)}})$, was used as an estimation of prediction accuracy. The slopes of regressing adjusted phenotypes on GEBV for animals in testing subgroup $\left(b_{\widehat{y_{tst}},GEBV_{tst}}\right)$ were evaluated as a measure of bias, which can be used to verify whether genomic predictions are inflated or deflated. The last comparison criterion was the mean square error, $MSE = \frac{\sum_1^{ntst}(GEBV_i - \hat{y}_i)^2}{n_{tst}}$, where n_{tst} is the size of testing dataset, that was used as a measure of precision and bias of the point estimator.

Data availability

The phenotypic and genotypic data are available at figshare repository and their description and accession numbers are listed in File S1. File S2 contains a custom R script used in the analysis.

Results and discussion

Accuracy of genomic-enabled breeding values

187 Pearson correlation coefficients between adjusted phenotypes and GEBV were
188 used as a proxy of genome-enabled prediction accuracies (Table 3). All estimates were
189 quite similar, ranging from 0.24 to 0.27. Bolormaa *et al.* (2013) reported even lower
190 accuracies (from 0.13 to 0.24) of GEBV for ADG in feedlot using GBLUP estimates in
191 *Bos taurus taurus* and *Bos taurus indicus* animals. When analyzing ADG of almost 4,000
192 Nelore young bulls in pasture using traditional BLUP, Fragomeni *et al.* (2013) reported an
193 EBV accuracy of 0.56, which suggests we could achieve higher accuracies than we found
194 in the present study.

195 It is known that the success of genomic selection depends on the accuracy of
196 GEBV, which in turn is a function of heritability, size of training population and effective
197 population size (N_e) (Goddard and Hayes 2009). Based on the simulation presented by van
198 der Werf (2013), who considered a population with N_e=250 (estimated N_e of Nelore
199 cattle=214 (Mudadu *et al.* 2016)) and a trait with h^2=0.5, a training population of 500
200 animals would reach an accuracy of 0.2, similar to our results. Moreover, the authors
201 showed that a training population of more than 2,000 individuals would be required to
202 achieve an accuracy of 0.4. Another key factor is the level of relationship among animals
203 in the training and testing sets. The present study evaluated half-sib families and according
204 to Hayes *et al.* (2009), this structure allows estimating only the effects of paternal alleles
205 with high accuracies, decreasing the reliability of the GEBVs.

206 Taking into account the above-mentioned factors, we point out that the crucial
207 points would be to increase the number of reference animals, and to include animals with
208 different levels of relationship, thus the SNP markers effects could be better estimated.
209 Since ADG in feedlot-finished steers could be viewed as a new selection criterion for

Nelore cattle, it is important to estimate the GEBVs with high accuracies in order to allow selection of young animals and genetic gains at a reduced genotyping costs.

Bias and precision measures of genomic-enabled breeding values

Regression coefficients of adjusted phenotypes on GEBV (Table 4) were used to measure the extent of prediction bias, since values greater or lower than 1 are related to deflated or inflated GEBV, respectively. For the 770k panel, only the results from BayesCπ models were not considered biased. Also, it is clear that estimates from BayesA models (except for TagSNP) were deflated, which means the GEBVs were not in the same scale as the adjusted phenotypes. The opposite was observed for all models applied to TagSNP dataset, thus it seems that selecting markers based only on their pairwise r^2 resulted in overestimated predictors.

Differences among prediction accuracies were negligible, thus information on slopes and MSE (Table 4) were combined and the models resulting in less biased GEBV were 770k-BayesCπ, HDi-BayesCπ and LDi-BayesCπ. The current average cost of genotyping can easily reach $150.00, $100.00 and $50.00 per animal, for 770k, HDi and LDi, respectively. Therefore, if it would be possible to predict accurate GEBV using less dense panels of SNP at lower costs, the implementation and application of genomic selection would be better accepted by the beef cattle industry.

Estimates of variance components

The divergences in the variance components (Table 5) were expected, since the markers included in each models captures different proportions of the genetic variance. For example, the marker-based genetic variance estimated using BGBLUP was the lowest

(about 0.02) in this study. For BayesA and BayesCπ the genetic variance is a function of SNP effects and their uncertainty variances and allelic frequencies (Gianola *et al.* 2009). Results from BayesA models were not consistent among SNP panels and, we hypothesized that by fitting a great number of markers, larger is the captured marker-based genetic variance (Table 5).

BayesCπ models resulted in less biased GEBVs, and its coefficients of heritability ranged from 0.41 to 0.44 (Table 5). This was similar to the coefficient reported by Olivieri *et al.* (2016) for ADG in Nelore cattle in post-weaning feedlot performance test ($h^2 = 0.43$). Although heritability is a population parameter, it is known that magnitudes of heritability estimates of similar traits are often similar across populations.

Conclusion

For the purpose of comparing GEBV estimates using different SNP panels and Bayesian models, we considered some of the most common criteria used to evaluate the quality of the genome-enabled predictions. Overall, all SNP panels and models provided similar accuracies, however *Bos taurus indicus* SNP chips (HDi and LDi) and methods that zero a proportion of the SNP effects, such as BayesCπ, seem to result in less biased predictions. Furthermore, results from less dense marker panels based on *Bos taurus indicus* were as good as the high-density panel, and at lower genotyping costs.

Acknowledgments

This study was funded by the Brazilian Agricultural Research Corporation (EMBRAPA) and Sao Paulo State Research Foundation (FAPESP) scholarships nº. 2012/23702-8 and 2013/21644-3, and grant nº 2012/23638-8. We thank Dante Pazzanese

Lanna, Michele Nascimento, Amália Chaves, Andrea Souza, Rymer Tullio, Sérgio Medeiros and Antônio Rosa for data collection, help and technical assistance. The authors would like to acknowledge the collaborative efforts among EMBRAPA, Sao Paulo State University, FAPESP and University of Wisconsin – Madison.

Literature Cited

Bakker, P. I. W., R. Yelensky, I. Pe'er, S. B. Gabriel, M. J. Daly *et al.*, 2005 Efficiency and power in genetic association studies. Nat. Genet. 37: 1217–1223.

Bolormaa, S., J. E. Pryce, K. Kemper, K. Savin, B. J. Hayes *et al.*, 2013 Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in Bos taurus, Bos indicus, and composite beef cattle. J. Anim. Sci. 91: 3088–3104.

Browning, B. L., and S. R. Browning, 2009 A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. Am. J. Hum. Genet. 84: 210–223.

de los Campos, G., D. Gianola, and G. J. M. Rosa, 2009 Reproducing kernel Hilbert spaces regression: A general framework for genetic evaluation. J. Anim. Sci. 87: 1883–1887.

de los Campos, G., and P. P. Rodriguez, 2014 *Bglr: Bayesian generalized linear regression*.

Clayton, D., 2012 *snpStats: SnpMatrix and XSnpMatrix classes and methods*.

Daetwyler, H. D., B. Villanueva, P. Bijma, and J. A. Woolliams, 2007 Inbreeding in genome-wide selection: Inbreeding in genome-wide selection. J. Anim. Breed. Genet. 124: 369–376.

282 Fragomeni, B. de O., D. C. B. Scalez, F. L. B. Toral, J. A. G. Bergmann, I. G. Pereira *et*
283 *al.*, 2013 Genetic parameters and alternatives for evaluation and ranking of Nellore
284 young bulls in pasture performance tests. Rev. Bras. Zootec. 42: 559–564.
285 Gianola, D., G. de los Campos, W. G. Hill, E. Manfredi, and R. Fernando, 2009 Additive
286 Genetic Variability and the Bayesian Alphabet. Genetics 183: 347–363.
287 Goddard, M. E., and B. J. Hayes, 2009 Mapping genes for complex traits in domestic
288 animals and their use in breeding programmes. Nat. Rev. Genet. 10: 381–391.
289 Hayes, B. J., P. M. Visscher, and M. E. Goddard, 2009 Increased accuracy of artificial
290 selection by using the realized relationship matrix. Genet. Res. 91: 47.
291 Instituto Brasileiro de Geografia e Estatística, 2013 Anuário Estatístico do Brasil.
292 Martha, G. B., E. Alves, and E. Contini, 2012 Land-saving approaches and beef
293 production growth in Brazil. Agric. Syst. 110: 173–177.
294 Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard, 2001 Prediction of Total Genetic
295 Value Using Genome-Wide Dense Marker Maps. Genetics 157: 1819–1829.
296 Millen, D. D., R. D. L. Pacheco, P. M. Meyer, P. H. M. Rodrigues, and M. De Beni
297 Arrigoni, 2011 Current outlook and future perspectives of beef production in Brazil.
298 Anim. Front. 1: 46–52.
299 Mudadu, M. A., L. R. Porto-Neto, F. B. Mokry, P. C. Tizioto, P. S. N. Oliveira *et al.*, 2016
300 Genomic structure and marker-derived gene networks for growth and meat quality
301 traits of Brazilian Nelore beef cattle. BMC Genomics 17.:
302 Olivieri, B. F., M. E. Z. Mercadante, J. N. dos S. G. Cyrillo, R. H. Branco, S. F. M.
303 Bonilha *et al.*, 2016 Genomic Regions Associated with Feed Efficiency Indicator
304 Traits in an Experimental Nellore Cattle Population (R. N. PENA i SUBIRÀ, Ed.).
305 PLOS ONE 11: e0164390.

R Development Core Team, 2014 *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rosa, G. J. M., C. R. Padovani, and D. Gianola, 2003 Robust Linear Mixed Models with Normal/Independent Distributions and Bayesian MCMC Implementation. Biom. J. 45: 573–590.

Schaeffer, L. R., 2006 Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 123: 218–223.

VanRaden, P. M., 2008 Efficient Methods to Compute Genomic Predictions. J. Dairy Sci. 91: 4414–4423.

van der Werf, J., 2013 Genomic Selection in Animal Breeding Programs, pp. 543–561 in *Genome-Wide Association Studies and Genomic Prediction*, edited by C. Gondro, J. van der Werf, and B. Hayes. Humana Press, Totowa, NJ.

Ziegler, A., 2009 Genome-wide association studies: quality control and population-based measures. Genet. Epidemiol. 33: S45–S50.

330 Table 1. Summary of age and weight at feedlot entry, ADG and days in feedlot for the 718

331 Nelore steers

	Age (d)	Weight (kg)	ADG (kg/d)	Days in feedlot
Minimum	542	226	0.193	48
Mean (sd)	649 (45)	361 (51)	1.235 (0.407)	92 (20)
Maximum	745	510	2.457	119

332

333 Table 2. Parameters of Gibbs sampler for each model

	Model		
MCMC samples	BayesA	BayesCπ	BGBLUP
Total	400,000	600,000	160,000
Burn-in	150,000	200,000	60,000
Thinning	10	20	10
Posterior*	25,000	20,000	10,000

334 *Final number of samples used to calculate features of posterior distributions.

335

336

337 Table 3. Pearson correlation coefficients used as proxy estimates of prediction accuracies
338 of genomic estimated breeding values for ADG of the 150 animals in testing subgroup

	SNP panel[1]			
Model	770k	TagSNP	HDi	LDi
BGBLUP	0.26	0.24	0.25	0.26
BayesA	0.26	0.25	0.26	0.27
BayesCπ	0.26	0.25	0.25	0.26

339 [1]actual number of SNPs included in the analysis: 770k - 534,787; TagSNP - 82,933; HDi -
340 63,945; LDi - 15,863.

341

342 Table 4. Regression coefficients (b) of GEBV on adjusted phenotype and mean squared
343 errors (MSE) of predictions for the 150 animals in testing subgroup

	SNP panel[1]							
Model	770k		TagSNP		HDi		LDi	
	b	MSE	b	MSE	b	MSE	b	MSE
BGBLUP	1.15	1.58	0.46	1.59	1.10	1.58	1.11	1.59
BayesA	1.29	1.09	0.69	1.24	1.68	1.32	1.99	1.37
BayesCπ	0.98	1.12	0.45	1.12	0.94	0.94	0.93	0.94

344 [1]actual number of SNPs included in the analysis: 770k - 534,787; TagSNP - 82,933; HDi -
345 63,945; LDi - 15,863.

346

Table 5. Estimates of residual ($\boldsymbol{\sigma_e^2}$) and genetic ($\boldsymbol{\sigma_g^2}$) variance components, heritability ($\mathbf{h^2}$) and proportion of non-zero effects ($\boldsymbol{\pi}$) for all models

SNP panel[1]	Parameter	BGBLUP[2]	BayesA[2,3]	BayesCπ[2,3]
770k	σ_e^2	0.05 (0.04-0.06)	0.06 (0.05-0.07)	0.05 (0.04-0.06)
	σ_g^2	0.02 (0.01-0.04)	0.06	0.03
	h^2	0.31 (0.19-0.45)	0.53 (0.49-0.58)	0.41 (0.36-0.47)
	π	—	—	0.98 (0.96-1.00)
TagSNP	σ_e^2	0.05 (0.04-0.06)	0.06 (0.05-0.07)	0.05 (0.04-0.06)
	σ_g^2	0.02 (0.01-0.04)	0.04	0.03
	h^2	0.32 (0.19-0.46)	0.40 (0.36-0.45)	0.42 (0.37-0.48)
	π	—	—	0.98 (0.96-1.00)
HDi	σ_e^2	0.05 (0.04-0.06)	0.06 (0.05-0.07)	0.05 (0.04-0.06)
	σ_g^2	0.02 (0.01-0.04)	0.03	0.03
	h^2	0.32 (0.19-0.46)	0.31 (0.28-0.35)	0.42 (0.37-0.48)
	π	—	—	0.98 (0.96-1.00)
LDi	σ_e^2	0.05 (0.04-0.06)	0.06 (0.05-0.07)	0.05 (0.03-0.06)
	σ_g^2	0.02 (0.01-0.04)	0.02	0.04
	h^2	0.32 (0.19-0.45)	0.28 (0.25-0.32)	0.44 (0.36-0.47)
	π	—	—	0.98 (0.96-1.00)

[1]actual number of SNPs included in the analysis: 770k - 534,787; TagSNP - 82,933; HDi - 63,945; LDi - 15,863; [2]numbers in brackets refers to the highest posterior density intervals (HPD) at 95% (lower bound–upper bound). [3]HPD for σ_g^2 for models BayesA and BayesCπ could not be estimated.