

Optimising bias and accuracy in genomic prediction of breeding values

G. R. Gowane¹, Sang Hong Lee^{2,3}, Sam Clark², Nasir Moghaddar², Hawlader A Al- Mamun² & Julius H. J. van der Werf²

¹ Animal Genetics & Breeding Division, ICAR-Central Sheep & Wool research Institute, Avikanagar 304501 Rajasthan, India

gopalgowane@gmail.com <mailto:Riberc@univut.com> (Corresponding Author)

²School of Environmental and Rural Sciences, University of New England, Armidale 2351 NSW, Australia

³Centre for Population Health Research, School of Health Sciences and Sansom Institute of Health Research, University of South Australia, Adelaide, Australia

Summary

Reference populations used for genomic selection (GS) usually involve highly selected genotyped individuals which may result in biased prediction of genomic estimated breeding values (GEBV). Bias and accuracy of GEBV in animal breeding programs was explored for various prediction methods. The data was simulated to compare Best Linear Unbiased Prediction of breeding values using pedigree based relationships (PBLUP), genomic relationships for genotyped animals only (GBLUP) and a Single Step approach (SSGBLUP), where information on genotyped individuals was used to infer realised relationship among all available genotyped and non-genotyped individuals that were linked through pedigree. In the SSGBLUP, varying weights ($\alpha=0.95, 0.50$) for the genomic relationship matrix (**G**) relative to the **A**-matrix weights ($1-\alpha$) were applied to construct an **H** matrix. Different selection and mating designs with various heritabilities (h^2) and QTL models were tested to compare the methods. Results showed that the accuracy of the GEBV prediction increased linearly with an increase in the number of animals selected for genotyping in the reference data. For a random mating design with no selection (RR), all prediction methods were unbiased. Prediction bias was evident in GBLUP, when a smaller proportion was more intensely selected for genotyping but bias was smaller when the proportion of selectively genotyped animals was 20% or higher. The SSGBLUP ($\alpha=0.95$) showed more accuracy compared to GBLUP and there was less bias with selective genotyping. However, PBLUP and SSGBLUP did show some bias with selection and assortative mating, probably due to not fully accounting for allele frequency changes due to selection of QTL with larger effects. This bias was larger in SSGBLUP than in PBLUP, likely due to the **G**- and **A**-matrices not being coherently scaled with allele frequency changes. SSGBLUP required lower values of α to decrease bias and increase accuracy of GEBV with selection and positive assortative mating. Models with a higher h^2 were more accurate and less biased in the prediction, compared to those with a lower h^2 . Results suggest that selective genotyping in a breeding programme can lead to significant bias in prediction of GEBV when only evaluating genotyped individuals. The SSGBLUP method can provide more accurate and less biased estimates but more attention needs to be paid to appropriate scaling of **A** and **G** matrices in selected populations.

Key Words: Genomic Selection, GBLUP, Bias of GEBV Prediction, Single Step methods

Introduction

Best Linear Unbiased Prediction (BLUP) provides unbiased estimates of breeding values in populations under selection, conditional on the inclusion of all information used in the selection decisions (Henderson 1975; Sorensen and Kennedy 1984). However, in the case of genomic prediction based on a selected reference population, this condition might not be met, in particular when only genotyped individuals are evaluated in genomic BLUP (GBLUP). A number of studies (VanRaden et al., 2009a, 2009b, Patry and Ducrocq 2011, Vitezica et al. 2011) reported decreased accuracy of genomic estimated breeding value (GEBV) and increased bias due to selective genotyping of sires. Single Step genomic BLUP (SSGBLUP) (Legarra et al. 2009; Christensen and Lund 2010) combines genomic relationships from genotyped individuals with pedigree relationships with non-genotyped individuals and this integration should allow information on unselected animals to be included, with all relationships tracing back to a conceptual unselected base population. However, SSGBLUP requires the genomic relationship matrix (**G**) and pedigree-based relationship matrix (**A**) to refer to the same base population as otherwise new bias could be introduced. Some studies have discussed this issue and proposed a slight modification in the SSGBLUP procedure (Forni et al. 2011, Vitezica et al. 2011, Christensen et al. 2012).

The present study aimed at exploring the effect of several parameters on accuracy and bias of genomic prediction in an animal breeding program. Genetic evaluations based on pedigree BLUP (PBLUP), GBLUP and SSGBLUP were compared. We investigated a number of factors affecting bias and accuracy of genetic evaluations, including (1) the proportion of individuals selected to have genotype information; (2) the genetic model of the trait as determined by the heritability and the number of QTLs explaining the variance in the trait; and (3) scenarios with and without selection and assortative mating.

Material and Methods

Population and genotype simulation

Data were simulated using QMSim (Sargolzaei & Schenkel, 2009). A historical population with effective population size (N_e) of 100, with 50 males and 50 females producing 2 progeny each by random union of

gametes, was generated for 95 generations and thereafter the number of progeny was gradually expanded to 1000 offspring until the 100th generation. In the subsequent 10 generations (101-110), 50 males were mated to 500 females who produced 1000 progeny following either random mating or positive assortative mating. The genomic structure consisted of 30 chromosomes of equal length (1 Morgan) with biallelic markers (60,000) that were randomly distributed across the genome with an equal frequency (0.5) in the first generation of the historical population. The mutation rate of the markers and QTL was 2.5×10^{-8} per locus per generation. Two genetic models were simulated that differed in the number and distribution of QTL effects (90 and 990 QTLs). The QTL allele effects were sampled from a Gamma distribution with a shape and scale parameter of 0.4 and 1.0, respectively. Genotype effects at individual QTL were added to true breeding values (TBV) and these were re-scaled to match the input value for the additive genetic variance. For individuals in the 101st generation these were normally distributed with mean 0 and variance h^2 . Residual effects on phenotype were independent and normally distributed with mean 0 and variance $(1-h^2)$. Therefore, the mean and variance for the simulated phenotypes were zero and one, respectively. Phenotypes were recorded for all individuals in the last 10 generations. Two different heritabilities ($h^2=0.3$ and 0.5) were tested for each scenario. Selection of sires was either random or based on estimated breeding values (EBV) obtained by PBLUP. Positive assortative mating was also applied, giving rise to three different scenarios: I: Random selection and random mating (RR), II: Selection on EBV and Random mating (SR), III: Selection on EBV and positive assortative mating (SA). Under each of these 3 scenarios, predictions of EBV were obtained by each of the different methods and modelling parameters. Scenario PBLUP involved BLUP prediction based on all phenotype and pedigree information from 9 generations to predict EBV in the 10th generation. For each of the last 4 generations (6th to 9th generation), either 125 (25%), 250 (50%) or 500 (100%) males were selected for genotyping to form a reference population consisting of 500, 1000 or 2000 animals (scenario G500, G1000 or G2000). In scenario G9550, all the 9550 animals from all 9 preceding generations were genotyped and used as a reference population. In scenario SS500, SS1000 and SS2000, a SSGBLUP was used to combine information from pedigree (9550 individual's pedigree) and genomic relationships from either 500, 1000 or 2000 genotyped individuals in the reference, respectively. For each scenario, we conducted 25 replicates.

Analysis and breeding value prediction

In total 1000 selection candidates in the 10th generation were used as a validation population to determine the bias and accuracy of predicted breeding values.

PBLUP analysis was based on pedigree information from the 9 preceding generations that include 9550 animals. The following linear model was used

(1)

where $\mathbf{y}$ is a vector of observations; $\mathbf{b}$ is a vector of fixed effects (sex); $\mathbf{a}$ is a vector of direct additive genetic effects of individual animals; $\mathbf{e}$ is a vector of residual errors; and $\mathbf{X}$ and $\mathbf{Z}$, are known incidence matrices. Assumptions in the model were $\mathbf{V}(\mathbf{a}) = \mathbf{A}\sigma_a^2$ and $\mathbf{V}(\mathbf{e}) = \mathbf{I}\sigma_e^2$; where $\mathbf{I}$ is an identity matrix, $\mathbf{A}$ is the numerator relationship matrix between animals derived from pedigree and σ_a^2 and σ_e^2 are additive genetic and residual variances, respectively.

GBLUP analysis was used to estimate genomic breeding values. A genomic relationship matrix ($\mathbf{G}$) was constructed by method of Yang et al. (2010) using PLINK 1.9 (Chang et al., 2015). The genomic analysis was done using MTG2 (Lee and van der Werf, 2016) to predict GEBVs for animals in generation 10. The model used for analysis was

(2)

where $\mathbf{g}$ is a vector of additive genetic effects of the individual animal, with $\text{var}(\mathbf{g}) = \mathbf{G}\sigma_g^2$ and other terms are defined as above.

SSGBLUP analysis combined pedigree and genomic information in a single step extending the animal model to include marker genotypes. The model for evaluation was:

(3)

where, $\mathbf{u}$ is the vector of direct additive genetic effects of individual animals with a distribution of $N(\mathbf{0}, \mathbf{H}\sigma_u^2)$. The $\mathbf{H}$ matrix involves non-genotyped and genotyped individuals. Other terms are already defined as above.

The tuning of $\mathbf{G}$ matrix was done in the $\mathbf{H}^{-1}$ matrix component $\mathbf{A}^{22} + (\mathbf{G}^{-1} - \mathbf{A}^{22})^{-1}$ with $\mathbf{G}$ being $\alpha\mathbf{G} + (1-\alpha)\mathbf{A}^{22}$ and two values of α were compared; 0.95 and 0.50. Apart from this, the modification of $\mathbf{H}$ matrix was tried as suggested by Vitezica et al. (2011) that involved adding a small constant to all elements of the $\mathbf{G}$. The value for this constant was derived by equating the sum of the elements of the $\mathbf{G}$ to the sum of the elements of the $\mathbf{A}^{22}$. The family of BLUPF90 programs (Mitszal, 2008; Aguillar et al. 2014) were used to analyse the data.

Accuracy of prediction of breeding values was obtained as the Pearson correlation between TBV and GEBV of all individuals in generation 10. Bias was estimated as regression of TBV on GEBV.

Results and Discussion

Effect of selection and mating design on accuracy of (G)EBV prediction

Increasing number of individuals in reference increases accuracy of GEBV prediction

A larger proportion of genotyped animals led to a larger sample size in the reference population which

increased the accuracy of GBLUP as well as for SSGBLUP prediction of GEBV (Table 1). The increase in the accuracy was approximately linear with sample size across the different scenarios where the number of genotyped males increased from 25% to 100%.

For the RR scenario and a trait with heritability of 0.3 controlled by 990 QTLs, PBLUP gave an accuracy of 0.48 ± 0.01 . GBLUP analysis gave a similar accuracy with G1000, but this increased further with the addition of more genotyped individuals in the reference. Accuracy with SSGBLUP was much higher; *i.e.* 0.59 ± 0.01 with SS500 that increased to 0.68 ± 0.01 for SS2000. Using genotype information on all animals (G9550) gave the highest prediction accuracy, as expected (0.79 ± 0.01).

In the SR scenario, it was observed that the accuracy of both EBV and GEBV prediction was lower than in RR. For PBLUP, the accuracy was 0.34 ± 0.01 . A similar value for the prediction accuracy was obtained by GBLUP with G1000 (0.31 ± 0.01). SSGBLUP with SS500 gave an accuracy of 0.46 ± 0.01 , and with SS2000 it was 0.60 ± 0.01 . For the SA scenario, the accuracy of prediction from PBLUP was 0.45 ± 0.01 . A similar prediction accuracy was obtained with G1000 (0.45 ± 0.02). SSGBLUP predicted GEBV again with more accuracy, *e.g.* accuracy was 0.55 with SS1000. The lower accuracies in GBLUP as compared to SSGBLUP were also reported by Vitezica et al. (2011) and can be likely attributed to SSGBLUP using information also on pedigree as well as better accounting for selection. The lower accuracy in the SR and SA scenario compared to RR is likely due to the Bulmer effect, *i.e.* the variation between families is reduced due to selection, leading to a lower correlation between EBV and TBV (Bijma, 2012). Assortative mating increases the variance among offspring and shows therefore higher correlations. With smaller reference population, GBLUP predictions seemed equally affected by selection as PBLUP predictions, suggesting that in this case the genomic information captures mainly the between family variance. With a larger reference population the effect of selection is smaller, suggesting more of the within family information is captured in that case. Another interesting feature with SA scenario is, decreasing the weights of α from 0.95 to 0.5 in Single Step improved the accuracy (Table 3), although the difference is statistically non-significant.

Effect of different heritability estimates and QTL models on prediction accuracy

For PBLUP, change of h^2 from 0.3 to 0.5 increased the accuracy by 18.8%, 8.8% and 20.0% for RR, SR and SA scenarios, respectively. Similarly, a gain in accuracy was observed for GBLUP and SSGBLUP across different methods when h^2 was increased. The increase in accuracy was higher for GBLUP when a smaller proportion was genotyped. The increase in accuracy for SSGBLUP was less compared to GBLUP with higher heritability. Non-random selection and mating designs (SR and SA) showed higher gains in accuracy compared to the random mating (RR) design with increased heritability.

Variation in the number of QTLs did not affect the accuracy of prediction of GEBV for the RR scenario. Similarly, accuracy for EBV obtained by PBLUP were unaffected by variation in QTL model for RR, SR and SA scenarios. This was probably due to the fact that the gamma distribution employed in the simulation resulted in a very few large QTLs. Prediction accuracy was affected by QTL number for SR and SA scenarios when using GBLUP and SSGBLUP, where the prediction accuracy improved when 90-QTL model was changed to a 990-QTL model, with more gains for SA than the SR scenario (Table 1). The 90-QTL model had few QTL with larger effect and the number of QTL with large effect decreased as the total number of QTL increased. This was thus responsible for improving prediction accuracy in the 990-QTL model.

Effect of selection and mating designs on bias of GEBV prediction

In the case of no selection, all methods showed no bias with regression coefficients around one. In the case of selection, (SR scenario), bias in GEBV prediction was evident. G500 and G1000 were biased with regression coefficients of 0.78 ± 0.06 and 0.85 ± 0.03 , respectively. GBLUP with G2000 was less biased (0.96 ± 0.02). For SSGBLUP, there was no bias in GEBV prediction in the SR scenario under the 990-QTL model but some bias was observed under the 90-QTL model. However, bias with SSGBLUP was significantly smaller than with GBLUP, especially when only a highly selected proportion was genotyped. For the SA design, the bias in the prediction of breeding value was very high for all methods of prediction, including PBLUP, where a regression coefficient of 0.88 ± 0.02 was observed. The GBLUP approach gave slightly under-dispersed GEBVs (with the regression coefficient equal to 1.11 ± 0.07 and 1.13 ± 0.05 for G500 and G1000, respectively). However, for G2000 and G9550, the estimated regression coefficient was 1.01 ± 0.03 and 0.96 ± 0.01 , respectively, indicating less bias when genomic information is available on more or all animals. The SSGBLUP methods using $\alpha=0.95$ was biased in the SA scenario. For the large group of non-genotyped animals the genetic values are, a priori, conditioned on genetic values of genotyped animals (Legarra et al. 2009) that are actually based on current genotypic frequencies of recent generations of selected animals, where significant changes in allelic frequency took place due to selection and assortative mating design. Even for SS2000 model, the prediction bias was 0.70 ± 0.04 and bias was larger with SSGBLUP than with GBLUP. It seems likely that the bias with the SSGBLUP methods is caused by an inconsistent scaling of the **A** and **G**-matrices, likely due to changing allele frequencies due to selection. For example, the **G** matrix in the SSGBLUP method uses average allele frequency of genotyped animals, not those of the base population. In the SA scenario, allele frequencies changed at a faster rate, inbreeding accumulated more rapidly and the genetic variance decreased for most of all selection scenarios. An inappropriate scaling of **G** versus **A** is also evident from the decreased bias that was observed when the α -value was decreased from 0.95 to 0.5 (Table 3). The improvement in regression of TBV on GEBV estimates for 990-

QTL model was 27.9% for SS500, 25% for SS1000 and 24.3% for SS2000 by shifting α from 0.95 to 0.50. The fact that allele frequency changes cause bias in the SSGBLUP method and to some extent in the PBLUP method is evident from the fact that this bias is smaller when a 990-QTL model is used rather than a 90-QTL model. Allele frequency changes were generally lower in the 990-QTL model.

GBLUP estimates with upward as well as downward bias were also reported by Vitezica et al. (2011) for a scenario where strong selection was simulated. We observed that by keeping $\alpha=0.95$ and by using the methods of Vitezica et al. (2011), the bias existed for the SA scenario in the current study. In animal model BLUP, the base population is always considered to be unselected and the mean genetic value of the base population is estimated to be zero, with any selection and genetic trend accounted for via the **A**-matrix, provided that information on unselected individuals is included in the analysis and assuming an infinitesimal model. However, in GBLUP where only information is used on selected individuals with genotype, this condition is not met.

Effect of heritability and QTL number variation on bias of prediction

Increasing the heritability from 0.3 to 0.5 resulted in relatively less biased estimates of prediction across prediction methods and mating designs (Table 2). For the RR scenario, with a change in the number of QTLs from 90 to 990-QTL, there was not much change in bias and the trend was pretty similar. However, when selection was involved (SR and SA scenarios), the 90-QTL model showed bias for all the methods of prediction. BLUP (also SSGBLUP and G9550 in the current study) should be robust against deviation from infinitesimal model (Maki-Tanila and Kennedy 1986); however with higher allele frequency changes for 90-QTL model, where there are a few QTLs with large effects, the prediction of breeding value was biased across all the methods. For the 90-QTL model, the allele frequency was drifting with high magnitude from generation to generation. There were smaller allele frequency changes for the 990-QTL model and also the prediction bias was significantly lower.

Conclusion

Our study indicated that using only highly selected genotyped individuals for genomic prediction results in considerable bias. The Single Step approach resulted in more accurate and less biased estimates of breeding value. However, with selection and assortative mating, some bias was also observed with the Single Step method, likely due to inappropriate merging of the **A** and **G**-matrices due to changing allele frequencies at large QTL. Such changes depended on the assumed QTL model and further work is needed to check whether such allele frequency changes can also cause such problems in real data.

Table 1. Estimates of accuracy of (G)EBV prediction using different approaches.

h²	QTLs	Design	PBLUP (0.01-0.02)	G500 (0.01-0.02)	G1000 (0.01-0.02)	G2000 (0.01-0.02)	SS500 (0.01-0.03)	SS1000 (0.00-0.03)	SS2000 (0.00-0.03)	G9550 (0.00-0.01)
0.30	90	RR	0.49	0.38	0.50	0.62	0.59	0.63	0.68	0.78
		SR	0.33	0.22	0.28	0.49	0.44	0.48	0.56	0.63
		SA	0.43	0.32	0.40	0.56	0.41	0.46	0.57	0.65
	990	RR	0.48	0.38	0.49	0.61	0.59	0.63	0.68	0.79
		SR	0.34	0.24	0.31	0.52	0.46	0.51	0.60	0.73
		SA	0.45	0.39	0.45	0.62	0.50	0.55	0.62	0.77
0.50	90	RR	0.51	0.44	0.57	0.69	0.64	0.68	0.74	0.85
		SR	0.37	0.27	0.33	0.53	0.46	0.50	0.59	0.65
		SA	0.49	0.41	0.47	0.62	0.54	0.58	0.65	0.68
	990	RR	0.57	0.47	0.59	0.70	0.67	0.71	0.76	0.86
		SR	0.37	0.34	0.40	0.61	0.54	0.59	0.68	0.81
		SA	0.54	0.51	0.56	0.71	0.63	0.66	0.75	0.83

Numbers in the parentheses are range for estimates of Standard Error (SE)

RR: No selection and random mating design; SR: Selection on the basis of EBV and random mating design; SA: Selection on the basis of EBV and assortative mating design

PBLUP: Pedigree (10550 with 9550 pedigree and 1000 validation population) based best linear unbiased prediction; G500: GBLUP with 25% close male relatives (125) from 6th to 9th generations in reference; G1000: GBLUP with 50% close male relatives (250) from 6th to 9th generations in reference; G2000: GBLUP with 100% close male relatives (500) from 6th to 9th generations in reference; SS500: Single-Step GBLUP (with $\alpha=0.95$) with pedigree information from 9550 relatives from 9 preceding generations and 25% close male genotyped relatives (125) from 6th to 9th generations in G matrix; SS1000: Single-Step GBLUP (with $\alpha=0.95$) with pedigree information from 9550 relatives from 9 preceding generations and 50% close male genotyped relatives (250) from 6th to 9th generations in G matrix; SS2000: Single-Step GBLUP (with $\alpha=0.95$) with pedigree information from 9550 relatives from 9 preceding generations and 100% close male genotyped relatives (125) from 6th to 9th generations in G matrix; G9550: GBLUP with 9550 relatives from complete pedigree genotyped and used in reference

Table 2. Estimates of bias of (G)EBV prediction using different approaches.

h²	QTLs	Design	PBLUP (0.01-0.03)	G500 (0.03-0.08)	G1000 (0.02-0.06)	G2000 (0.01-0.03)	SS500 (0.01-0.05)	SS1000 (0.01-0.05)	SS2000 (0.01-0.04)	G9550 (0.01-0.02)
----------------------	-------------	---------------	-----------------------------	----------------------------	-----------------------------	-----------------------------	-----------------------------	------------------------------	------------------------------	-----------------------------

0.30	90	RR	1.04	0.97	1.03	1.02	1.00	0.99	1.00	1.00
		SR	0.87	0.67	0.73	0.84	0.85	0.86	0.80	0.80
		SA	0.72	0.81	0.96	0.90	0.49	0.53	0.56	0.78
	990	RR	0.95	1.16	1.01	1.01	1.01	1.00	1.00	1.00
		SR	1.00	0.78	0.85	0.96	1.00	1.02	0.96	0.98
		SA	0.88	1.11	1.13	1.01	0.68	0.72	0.70	0.96
	90	RR	0.93	0.97	0.98	0.98	0.95	0.96	0.97	0.99
		SR	0.75	0.67	0.74	0.84	0.76	0.79	0.75	0.75
		SA	0.78	0.85	0.94	0.87	0.63	0.67	0.67	0.75
0.50	990	RR	1.02	1.07	1.04	1.04	1.00	1.00	1.00	1.01
		SR	0.91	0.85	0.92	0.97	1.02	1.06	0.97	0.98
		SA	0.93	1.09	1.13	1.01	0.86	0.88	0.85	0.96

Table 3. Estimates of accuracy and bias of GEBV prediction in Single Step approach with tuning of **H** matrix

Criteria	QTLs	Design	SS500 [$\alpha=0.95$, \$] (0.01-0.05)	SS1000 [$\alpha=0.95$, \$] (0.01-0.05)	SS2000 [$\alpha=0.95$, \$] (0.01-0.04)	SS500 [$\alpha=0.50$] (0.01-0.05)	SS1000 [$\alpha=0.50$] (0.01-0.05)	SS2000 [$\alpha=0.50$] (0.01-0.04)
Accuracy	90	RR	0.59	0.63	0.68	0.58	0.61	0.66
		SR	0.44	0.48	0.56	0.42	0.45	0.55
		SA	0.41	0.46	0.57	0.45	0.50	0.58
	990	RR	0.59	0.63	0.68	0.57	0.61	0.65
		SR	0.46	0.51	0.60	0.43	0.47	0.57
		SA	0.50	0.55	0.62	0.51	0.56	0.64
Bias	90	RR	1.00	0.99	1.00	1.10	1.11	1.13
		SR	0.85	0.86	0.80	0.96	1.00	1.00
		SA	0.49	0.53	0.56	0.65	0.69	0.70
	990	RR	1.01	1.00	1.00	1.09	1.11	1.23
		SR	1.00	1.02	0.96	1.12	1.16	1.18
		SA	0.68	0.72	0.70	0.87	0.90	0.87

The estimate of h^2 was assumed to be 0.30; \$: modification of SSGBLUP as suggested by Vitezica et al. (2011), $\alpha=0.50$: shrinking information from **G** matrix to 50%, $\alpha=0.95$: shrinking information from **G** matrix to 95%

List of References

- Aguilar, I., I. Misztal, S. Tsuruta, A. Legarra & H. Wang, 2014. PREGSF90 – POSTGSF90: Computational Tools for the Implementation of Single-step Genomic Selection and Genome-wide Association with Ungenotyped Individuals in BLUPF90 Programs. Proceedings, 10th World Congress of Genetics Applied to Livestock Production
- Bijma, P., 2012. Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations. J Anim Breed Genet. 5:345-358.
- Chang, C. C., C. C. Chow, L.C.A.M. Tellier, S. Vattikuti, S. M. Purcell & J. J. Lee, 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Giga Science, 4. <https://www.cog-genomics.org/plink2>
- Christensen, O. F. & M. S. Lund, 2010 Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42: 2–2.
- Christensen, O. F., P. Madsen, B. Nielsen, T. Ostensen, & G. Su, 2012 Single-step methods for genomic evaluation in pigs. Animal 6: 1565–1571.
- Forni, S., I. Aguilar, & I. Misztal, 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genet. Sel. Evol. 43:1
- Goddard, M. E. & B. J. Hayes, 2007. Genomic Selection. J. Anim. Breed. Genet. 124: 323-330.
- Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423
- Lee, S. H. & J. H. J. van der Werf, 2016. MTG2: An efficient algorithm for multivariate linear mixed model analysis based on genomic information. Bioinformatics 32: 1420-1422.
- Legarra, A., I. Aguilar, & I. Misztal, 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92: 4656–4663.
- Maki-Tanila, A. & B.W. Kennedy, 1986. Mixed-model methodology under genetic models with a small

- number of additive and non-additive loci. *Proc. 3rd World Congr. Genet. Appl. Livest. Prod.*, vol. 12, p. 443
12. 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
 13. Misztal, I. 2008. BLUPF90 - a flexible mixed model program in Fortran 90. *Animal and Dairy Science*, University of Georgia, August 2008.
 14. Patry, C. & V. Ducrocq, 2011. Evidence of biases in genetic evaluations due to genomic preselection in dairy cattle. *J. Dairy Sci.* 94: 1011-1020.
 15. Sargolzaei, M. & F. S. Schenkel, 2009. QMSim: a large-scale genome simulator for livestock. *Bioinformatics* 25 (5): 680-681.
 16. Sorensen, D. A. & B. W. Kennedy, 1984. Estimation of response to selection using least squares and mixed model methodology. *J. Anim. Sci.* 58: 1097–1103.
 17. VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor & F. S. Schenkel, 2009a. Invited review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92: 16–24.
 18. VanRaden, P. M., M. E. Tooker, & J. B. Cole, 2009b. Can you believe those genomic evaluations for young bulls? *J. Dairy Sci.* 92(E-Suppl. 1): 314 (Abstr.)
 19. Vitezica, Z. G., I. Aguilar, I. Misztal, & A. Legarra, 2011. Bias in genomic predictions for populations under selection. *Genet Res (Camb)* 93: 357–366.
 20. Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard, P. M. Visscher, 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature Genet* 42:565-571.