

PROCEEDINGS

Open Access

# Estimating genomic breeding values and detecting QTL using univariate and bivariate models

Mario PL Calus\*, Han A Mulder, Roel F Veerkamp

From 14th QTL-MAS Workshop  
Poznan, Poland. 17-18 May 2010

## Abstract

**Background:** Genomic selection is particularly beneficial for difficult or expensive to measure traits. Since multi-trait selection is an important tool to deal with such cases, an important question is what the added value is of multi-trait genomic selection.

**Methods:** The simulated dataset, including a quantitative and binary trait, was analyzed with four univariate and bivariate linear models to predict breeding values for juvenile animals. Two models estimated variance components with REML using a numerator (A), or SNP based relationship matrix (G). Two SNP based Bayesian models included one (BayesA) or two distributions (BayesC) for estimated SNP effects. The bivariate BayesC model sampled QTL probabilities for each SNP conditional on both traits. Genotypes were permuted 2,000 times against phenotypes and pedigree, to obtain significance thresholds for posterior QTL probabilities. Genotypes were permuted rather than phenotypes, to retain relationships between pedigree and phenotypes, such that polygenic effects could still be estimated.

**Results:** Correlations between estimated breeding values (EBV) of different SNP based models, for juvenile animals, were greater than 0.93 (0.87) for the quantitative (binary) trait. Estimated genetic correlation was 0.71 (0.66) for model G (A). Accuracies of breeding values of SNP based models were for both traits highest for BayesC and lowest for G. Accuracies of breeding values of bivariate models were up to 0.08 higher than for univariate models. The bivariate BayesC model detected 14 out of 32 QTL for the quantitative trait, and 8 out of 22 for the binary trait.

**Conclusions:** Accuracy of EBV clearly improved for both traits using bivariate compared to univariate models. BayesC achieved highest accuracies of EBV and was also one of the methods that found most QTL. Permuting genotypes against phenotypes and pedigree in BayesC provided an effective way to derive significance thresholds for posterior QTL probabilities.

## Background

Genomic selection is particularly beneficial for difficult or expensive to measure traits [1]. One strategy to partly tackle these issues in breeding schemes previously, without using genotypic information, was multi-trait selection [e.g. 2]. An important question is therefore what

the added value is of multi-trait genomic selection. VanRaden and Sullivan [3] showed some benefit using this approach in international dairy cattle evaluations. There are, however, no other reports so far on applications of multi-trait genomic selection. The objective of this study was to present methods to apply multi-trait genomic breeding value prediction, and to evaluate their performance and impact on accuracy of prediction compared to single trait applications. In addition, the ability of one model to detect QTL was investigated.

\* Correspondence: [mario.calus@wur.nl](mailto:mario.calus@wur.nl)  
Animal Breeding and Genomics Centre, Wageningen UR Livestock Research,  
Lelystad, Netherlands  
Full list of author information is available at the end of the article

## Methods

### Estimation of breeding values

Simulated data of the 14<sup>th</sup> QTL-MAS workshop was analyzed with univariate and bivariate applications of four different models to predict breeding values for juvenile animals without phenotypes. A linear model was assumed for both the quantitative and binary trait. Using a linear model for binary traits is expected to give breeding values that are highly related to those obtained from a threshold model, when trait incidence is moderate [e.g. 4], which is the case here with a value of 0.30. The first two models used ASREML to estimate variance components:

$$y_{ij} = \mu_j + animal_{ij} + e_{ij}$$

where  $y_{ij}$  is the phenotypic record of animal  $i$ ,  $\mu_j$  is the overall mean for trait  $j$ ,  $animal_{ij}$  is the random polygenic effect of animal  $i$  for trait  $j$ , and  $e_{ij}$  is a random residual for animal  $i$ . Model A used a numerator relationship matrix for polygenic effects, while model G used a SNP based genomic relationship matrix. For G, matrix  $\mathbf{G}$  was calculated as [5]:

$$\mathbf{G} = \frac{\mathbf{Z}\mathbf{Z}'}{2 \sum p_i(1 - p_i)}$$

where  $\mathbf{Z}$  contained marker genotypes for all animals across loci, being -1 and 1 for either homozygote and 0 for the heterozygote genotype, corrected for allele frequency per locus in the current population.

The third and fourth model were based on Gibbs sampling and included SNP effects, next to the pedigree based relationship matrix:

$$y_{ij} = \mu_j + animal_{ij} + \sum_{k=1}^{nloc} \sum_{l=1}^2 SNP_{ijkl} + e_{ij}$$

where  $SNP_{ijkl}$  is a random effect for allele  $l$  on trait  $j$  at locus  $k$  of animal  $i$ . The difference between those two models is that 1 (BayesA) or 2 (BayesC) distributions for SNP effects are considered, respectively.

SNP effects, denoted as  $SNP_{ijkb}$ , were estimated in BayesA and BayesC as  $q_{ijkl} \times v_{.jk}$  [6], where  $q_{ijkl}$  is the effect size of allele  $l$  at locus  $k$  and  $v_{.jk}$  is the direction vector for locus  $k$  that scales the effect at locus  $k$  for trait  $j$ . In the original implementation [6], variance of the direction vector  $v_{.jk}$ , denoted as  $\mathbf{V}$ , is sampled for each trait  $j$  separately, without considering covariances between traits across loci. Here, both in BayesA and BayesC, in  $\mathbf{V}$  covariances between traits across loci are considered.

### QTL mapping

BayesC, also known as Bayesian stochastic search variable selection (BSSVS) [7], involved sampling presence of a QTL at each SNP position from a Bernoulli distribution with probability equal to

$$\frac{P(\mathbf{v}_j | \mathbf{0}, \mathbf{V}) \times Pr_j}{P(\mathbf{v}_j | \mathbf{0}, \mathbf{V}) \times Pr_j + P(\mathbf{v}_j | \mathbf{0}, \mathbf{V} / 100) \times (1 - Pr_j)}$$

where  $P(\mathbf{v}_j | \mathbf{0}, \mathbf{V})$  is the probability of sampling  $\mathbf{v}_j$  from  $N(\mathbf{0}, \mathbf{V})$ , and  $Pr_j$  is the prior probability of presence of a QTL at SNP position  $j$ .  $Pr_j$  was calculated per locus as 50 divided by the total number of SNPs, reflecting that 50 QTL were expected. Posterior QTL probabilities were calculated as proportions of cycles after burn-in that a locus was placed in the distribution with large effects and therefore was sampled from  $N(\mathbf{0}, \mathbf{V})$ . For more details on prior distributions and fully conditional distributions, see Meuwissen and Goddard [6].

To obtain significance thresholds for posterior QTL probabilities for the bivariate BayesC model, genotypes were permuted 2,000 times against phenotypes and pedigree.

## Results

### Variance components

Estimated variance components obtained from bivariate models A and G were used to calculate heritabilities and genetic correlations (Table 1). Estimated heritabilities and genetic correlations were not significantly different between bivariate models A and G. Standard errors of heritabilities and genetic correlations were lower for G than for A.

### Breeding values

Correlations were calculated among EBVs of all models for juvenile animals (Table 2). EBV of model A had correlations with SNP based models of ~0.6 for both traits. EBV were highly correlated between G and BayesA and BayesA and BayesC, but the correlation dropped to 0.94-0.95 when comparing G and BayesC. EBV of univariate and bivariate models had correlations of  $\geq 0.98$  for the quantitative trait, and  $\geq 0.93$  for the binary trait.

**Table 1 Estimated heritabilities and genetic correlations**

Model	$h^2$					
	Quantitative	s.e.	Binary	s.e.	$r_g$	s.e.
A	0.53	0.06	0.22	0.04	0.66	0.09
G	0.46	0.03	0.29	0.03	0.71	0.06

Estimated heritabilities and genetic correlations ( $r_g$ ), and standard errors (s.e.), obtained with models with an additive genetic (A) or SNP based relationship matrix (G).

**Table 2 Correlations between predicted breeding values of juvenile animals**

		Univariate				Bivariate			
		A	G	BayesA	BayesC	A	G	BayesA	BayesC
Uni	A		0.60	0.67	0.63	0.99	0.62	0.61	0.58
	G	0.60		0.98	0.94	0.60	0.99	0.99	0.94
	BayesA	0.62	1.00		0.98	0.66	0.98	0.99	0.96
	BayesC	0.56	0.95	0.96		0.63	0.94	0.96	0.98
Biv	A	0.93	0.62	0.64	0.60		0.63	0.61	0.58
	G	0.60	0.95	0.95	0.94	0.64		0.99	0.95
	BayesA	0.58	0.94	0.95	0.96	0.63	0.99		0.98
	BayesC	0.50	0.88	0.88	0.95	0.57	0.94	0.96	

Correlations between breeding values predicted using univariate (Uni) and bivariate (Biv) models with an additive genetic (A) or SNP based relationship matrix (G), and a Bayesian model with one (BayesA) or two distributions (BayesC) for SNP effects. Correlations above (below) the diagonal are for the quantitative (binary) trait.

Accuracies increased markedly going from A to G, and less so going from G to BayesA and from BayesA to BayesC (Table 3). Accuracies were always higher for bivariate than for univariate models. The relatively small increase in accuracy for bivariate compared to univariate models, especially for the quantitative trait, indicates that added power in the bivariate model was limited for this trait. Coefficients of regressing true breeding values on estimated breeding values were <1.0 for all models, indicating that variance of EBV was generally overestimated (Table 3). This overestimation was smaller for SNP based models, compared to model A.

### QTL detection

Detection of QTL was considered for univariate and bivariate BayesC models, while significance thresholds were only derived for the bivariate BayesC model. Therefore, only detected QTL from the bivariate BayesC model were used in the comparison of QTL detection methods.

For the quantitative trait 14 out of 32 QTL were detected, while for the binary trait 8 out of 22 were

**Table 3 Accuracies and regressions of true on estimated breeding values for juvenile animals**

Model	Accuracy				Regression coefficient			
	Quantitative trait		Binary trait		Quantitative trait		Binary trait	
	Uni.	Biv.	Uni.	Biv.	Uni.	Biv.	Uni.	Biv.
A	0.39	0.39	0.47	0.52	0.84	0.84	0.71	0.75
G	0.61	0.62	0.72	0.79	0.96	0.96	0.83	0.88
BayesA	0.63	0.64	0.73	0.81	0.96	0.96	0.84	0.91
BayesC	0.66	0.67	0.79	0.85	0.93	0.93	0.91	0.95

Correlations between true and estimated breeding values, and coefficients of regressions of true on estimated breeding values, predicted using univariate (Uni.) and bivariate (Biv.) models with an additive genetic (A) or SNP based relationship matrix (G), and a Bayesian model with one (BayesA) or two distributions (BayesC) for SNP effects

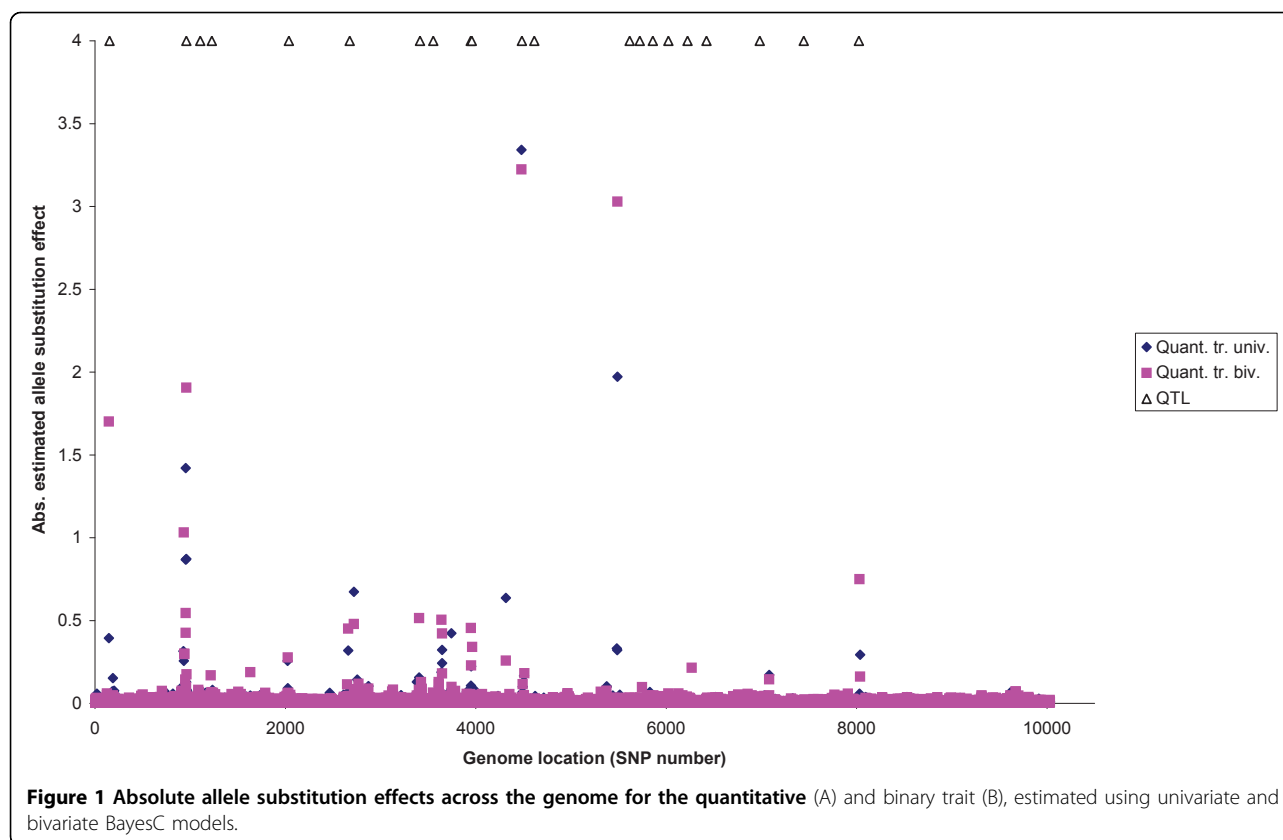
detected [8]. SNP that were declared significant together explained 35.0% and 22.6% of the genetic variance of the quantitative and binary trait, respectively. Polygenic effects explained only 4.3 and 1.1% of the genetic variance. This indicates that most of the genetic variance (i.e. 60.7 and 76.2% for the quantitative and binary trait, respectively) in the bivariate BayesC model was explained by effects of SNP that were not declared significant.

Absolute allele substitution effects for both traits estimated with univariate and bivariate BayesC models were plotted across the genome, together with positions of all additive QTL (Figure 1). Generally, allele substitution effects from the bivariate model were higher than from the univariate model. Since most additive QTL were pleiotropic, the bivariate model indeed seemed able to use information from both traits to estimate SNP effects per trait more accurately.

### Discussion

This study aimed to present methods to apply multi-trait genomic breeding value prediction, to evaluate impact on accuracy of prediction compared to single trait genomic breeding value prediction, and to detect QTL with one of the models. Results clearly indicated that accuracy of EBV increased when model complexity increased to allow better modeling of the genetic architecture. First, accuracy increased going from model A, to SNP based models with increasing flexibility to model SNP effects (in the order: G, BayesA, BayesC). Second, accuracy of EBV for both traits increased more for all SNP based models when using bivariate instead of univariate applications, compared to model A. This confirms results of a simulation study for dairy cattle showing that model G yields higher accuracies when using data of multiple countries compared to one country [3]. Third, considering that few QTL had relatively large effects, it was expected that the model best able to give more weight to loci with large effect – BayesC – fits the data best. Results are fully in agreement with this expectation. This suggests that SNP based models were better able to capture pleiotropic effects of QTL.

The model that achieved highest EBV accuracy, i.e. BayesC, was also one of the presented models that detected most QTL. The model that is best able to detect the position of QTL, however, is not always the model that is best able to predict total genetic merit of animals [9]. Permuting genotypes against phenotypes and pedigree in model BayesC provided an effective way to derive significance thresholds for posterior QTL probabilities. Note that SNP genotypes after the permutation no longer followed Mendelian inheritance. Lack of Mendelian inheritance probably results in fewer associations, since SNPs are less likely to capture pedigree effects, and therefore a lower threshold. In the present



data, however, polygenic effects only captured a very small fraction of the variance, indicating that the applied permutation strategy will have had a minor impact on significance thresholds.

## Conclusions

The EBV accuracy clearly improved for both traits for all bivariate models compared to their univariate counterparts. BayesC achieved highest EBV accuracies and was also one of the methods presented at the workshop that found most QTL.

## Acknowledgements

MPLC and RFV were funded by the EU RobustMilk project that is financially supported by the European Commission under the Seventh Research Framework Programme, Grant Agreement KBBE-211708, and HM by the EU SABRE project that is financially supported by the European Commission under the Sixth Research Framework Programme, contract No. FOOD-CT-2006-016250. The content of this paper is the sole responsibility of the authors, and it does not necessarily represent the views of the Commission or its services.

This article has been published as part of *BMC Proceedings* Volume 5 Supplement 3, 2011: Proceedings of the 14th QTL-MAS Workshop. The full contents of the supplement are available online at <http://www.biomedcentral.com/1753-6561/5?issue=S3>.

## Authors' contributions

MPLC developed software for the Bayesian models, carried out analyses and drafted the manuscript. HAM and RFV helped to interpret results and write the manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

Published: 27 May 2011

## References

- Haley CS, Visscher PM: Strategies to utilize marker-quantitative trait loci associations. *J Dairy Sci* 1998, **81**(Suppl. 2):85-97.
- Apiolaza LA: Very early selection for solid wood quality: screening for early winners. *Ann Forest Sci* 2009, **66**(6).
- VanRaden P, Sullivan P: International genomic evaluation methods for dairy cattle. *Genet Sel Evol* 2010, **42**(1):7.
- Weller JI, Misztal I, Gianola D: Genetic analysis of dystocia and calf mortality in Israeli Holsteins by threshold and linear models. 1988, **71**(9):2491-2501.
- VanRaden PM: Efficient methods to compute genomic predictions. *J Dairy Sci* 2008, **91**(11):4414-4423.
- Meuwissen THE, Goddard ME: Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genet Sel Evol* 2004, **36**(3):261-279.
- Verbyla KL, Hayes BJ, Bowman PJ, Goddard ME: Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. *Genet Res* 2009, **91**(5):307-311.
- Mucha S, Pszczola M, Strabel T, Wolc A, Paczyńska P, Szydlowski M: Comparison of analyses of the QTLMAS XIV common dataset. II: QTL analysis. *BMC Proceedings* 2011, **5**(Suppl 3).
- Calus MPL, Meuwissen THE, Windig JJ, Knol EF, Schrooten C, Vereijken ALJ, Veerkamp RF: Effects of the number of markers per haplotype and clustering of haplotypes on the accuracy of QTL mapping and prediction of genomic breeding values. *Genet Sel Evol* 2009, **41**:11.

doi:10.1186/1753-6561-5-S3-S5

Cite this article as: Calus et al.: Estimating genomic breeding values and detecting QTL using univariate and bivariate models. *BMC Proceedings* 2011 **5**(Suppl 3):S5.