

RESEARCH ARTICLE

Open Access

# Comparison of single-trait and multiple-trait genomic prediction models

Gang Guo<sup>1,2,3,4†</sup>, Fuping Zhao<sup>1†</sup>, Yachun Wang<sup>3</sup>, Yuan Zhang<sup>3</sup>, Lixin Du<sup>1\*</sup> and Guosheng Su<sup>4\*</sup>

## Abstract

**Background:** In this study, a single-trait genomic model (STGM) is compared with a multiple-trait genomic model (MTGM) for genomic prediction using conventional estimated breeding values (EBVs) calculated using a conventional single-trait and multiple-trait linear mixed models as the response variables. Three scenarios with and without missing data were simulated; no missing data, 90% missing data in a trait with high heritability, and 90% missing data in a trait with low heritability. The simulated genome had a length of 500 cM with 5000 equally spaced single nucleotide polymorphism markers and 300 randomly distributed quantitative trait loci (QTL). The true breeding values of each trait were determined using 200 of the QTLs, and the remaining 100 QTLs were assumed to affect both the high (trait I with heritability of 0.3) and the low (trait II with heritability of 0.05) heritability traits. The genetic correlation between traits I and II was 0.5, and the residual correlation was zero.

**Results:** The results showed that when there were no missing records, MTGM and STGM gave the same reliability for the genomic predictions for trait I while, for trait II, MTGM performed better than STGM. When there were missing records for one of the two traits, MTGM performed much better than STGM. In general, the difference in reliability of genomic EBVs predicted using the EBV response variables estimated from either the multiple-trait or single-trait models was relatively small for the trait without missing data. However, for the trait with missing data, the EBV response variable obtained from the multiple-trait model gave a more reliable genomic prediction than the EBV response variable from the single-trait model.

**Conclusions:** These results indicate that MTGM performed better than STGM for the trait with low heritability and for the trait with a limited number of records. Even when the EBV response variable was obtained using the multiple-trait model, the genomic prediction using MTGM was more reliable than the prediction using the STGM.

**Keywords:** Genomic selection, Reliability, Multiple-trait model, Single-trait model, Heritability

## Background

The availability of genome-wide markers, such as single nucleotide polymorphism (SNP) markers, has made it possible to predict breeding values of candidate animals using genomic information. The genomic prediction principle was first proposed by Meuwissen *et al.* [1]. A typical genomic prediction procedure is to estimate simultaneously the effects of all the SNPs available in the genotype data, and then to sum up all the predicted SNP effects as the

genomic estimated breeding value (GEBV). Selection based on the GEBV is called genomic selection. Because GEBV is calculated based on genetic marker information rather than on phenotypic information, genomic selection can shorten the generation interval, while maintaining the accuracy of the estimated breeding value (EBV) at an acceptable level [1,2]. Genomic selection is especially useful for low heritability traits, sex-limited traits, and traits that are difficult or expensive to measure, such as carcass, health, longevity, and fertility traits. The advantages of genomic selection have been corroborated by simulation and empirical studies [1-5]. Recently, genomic selection has been successfully implemented in dairy cattle breeding programs in many countries to accelerate the genetic progress and reduce the cost of progeny testing [6-9].

\* Correspondence: lxdu@263.net; guosheng.su@agrsci.dk

†Equal contributors

<sup>1</sup>National Center for Molecular Genetics and Breeding of Animal, Institute of Animal Sciences, Chinese academy of Agricultural Sciences, Beijing 100193, China

<sup>4</sup>Department of Genetics and Biotechnology, Faculty of Agricultural Sciences, Aarhus University, Tjele DK-8830, Denmark

Full list of author information is available at the end of the article

Numerous genomic selection studies have focused on single-trait analyses. However, many traits are genetically correlated, such as the reproductive and milk yield traits in dairy cattle and these traits may have different heritabilities. Some traits, such as feed efficiency, may be recorded only in a small number of animals because of the difficulty of measuring them. Like traditional genetic evaluation, a multiple-trait model is expected to increase the accuracy of the GEBV by making use of information from genetically correlated traits. The benefit of using a multiple-trait model will be more profound for traits with low heritability and a small number of phenotypic records. Multiple-trait models for genomic prediction have been reported recently [10-14]. It has been shown that a multiple-trait genomic model (MTGM) had higher prediction accuracy than a single-trait genomic model (STGM).

Bayesian variable selection methods generally outperform linear mixed models; often called the genomic best linear unbiased prediction (GBLUP) method [1,15]. However, the advantage of Bayesian methods over GBLUP is dependent on the genetic architecture [16], such as the number of quantitative trait loci (QTLs) and the density of the markers. Clark *et al.* [17] found that GBLUP produced slightly better prediction accuracy than the BayesB model when a trait was affected by a large number of QTLs with small effects. Similar trends were observed by Coster *et al.* [18] and Li *et al.* [19]. Many studies using real data have reported that GBLUP performs as well as Bayesian variable selection models for most traits [20-22]. Ober *et al.* [23] showed that BayesB was less accurate than GBLUP in predicting phenotypes of QTLs based on the genomic sequence data of *Drosophila melanogaster*. The main advantages of GBLUP over Bayesian methods are, its implementation is straightforward using existing residual maximum likelihood (REML) and BLUP programs, and it requires less computation time, which can be an important factor in the practical application of a genomic prediction method. Although some studies have shown that computing time for Bayesian models can be reduced greatly using the expectation-maximization (EM) and variational Bayes algorithms [19,24], GBLUP models (at either the SNP or individual animal levels) are still attractive approaches in practical genomic evaluations [6,7,25].

Three types of response variables that have been used widely to predict the GEBV are EBV, daughter yield deviation (DYD), and deregressed proof (DRP) [9,13,14,26]. The EBV of a bull is calculated from the information of all available relatives including the daughters. The DYD of a bull is the average of the daughters' actual performances adjusted for fixed and non-genetic random effects and genetic effects of the daughters' dams. The DRP is derived from the EBV [27] and can be considered as an analogue of DYD. Because EBV is estimated from

the information of all relatives, the reliability of EBV is higher than that of DYD or DRP. Furthermore, EBV can be obtained directly from a database of routine genetic evaluations. Our previous simulation study showed that, in genomic prediction, using the conventional EBV as the response variable gave slightly better results than using DYD in most scenarios [26]. In practical routine genetic evaluation, EBV (and also DRP or DYD) is usually calculated using multiple-trait models. This poses an important question: are MTGMs needed if a multiple-trait model is used to derive the response variables?

The objective of this study was to compare a STGM and a MTGM for genomic prediction using conventional EBVs estimated with a conventional single-trait linear mixed model and a conventional multiple-trait linear mixed model as response variables. The comparison was carried out using data from various simulation scenarios considering heritability of two genetically correlated traits and the proportion of missing records in the data for the two traits.

## Materials and methods

### Simulation schemes

Genomic predictions were obtained using both a STGM and MTGM with simulated data for two genetically correlated traits. Trait I was assumed to have high heritability ( $h^2 = 0.3$ ) and trait II was assumed to have low heritability ( $h^2 = 0.05$ ). The genetic correlation was set as 0.5 and the residual error correlation was 0.

In the simulation scheme, the initial population comprised 50 sires and 50 dams, and this structure was kept constant for 50 historical generations. Then the population was extended to 1,000 sires and 200,000 dams. Thereafter, four generations (G1-G4) were generated to obtain the data used for the analysis. The population was assumed to be under random mating conditions with no overlapping generations. In G1-G4, all the bulls were genotyped and all the cows had a phenotypic record. The G1-G3 bulls were used as "reference animals" and their EBVs were used as the response variables for the genomic predictions; the G4 bulls were used as "validation animals".

The simulations also generated reference populations with a small amount of data for one of the two traits. To simplify the simulation and analysis without losing the generality of the data, traits with a small number of records were handled by masking the EBVs of some individuals in the reference population, instead of by generating incomplete phenotype data. In this way, three response variable datasets were generated: (1) no missing EBVs for either of the traits; (2) no missing EBVs for trait II, but 90% of the EBVs for trait I were missed at random (*i.e.* the EBVs of only 300 of the 3000 bulls in the reference population were used for genomic prediction); and (3)

no missing EBVs of trait I, but 90% of the EBVs for trait II were missed at random.

The simulated genome consisted of five chromosomes (100 cM each). A total of 5000 biallelic SNP markers were spaced equally across the whole genome at distances of 0.1 cM. Three hundred biallelic QTLs were generated and divided randomly into three groups (100 each). One hundred of the QTLs affected only trait I, 100 affected only trait II, and 100 affected traits I and II. The QTL positions were located randomly according to the gene distribution in the first 500 cM of the standard mouse genome (NCBI 2005). The QTL effects were drawn from a gamma distribution with a shape parameter of 0.84 and scale parameter of 5.4, and assigned positive or negative by equal chance. Hayes and Goddard [28] noted that published estimates of QTL effects resembled a gamma distribution with shape parameter of 0.4. However, generally, only the QTLs that are statistically significant are reported in the literature. This conditional reporting can lead to a marked upward bias of the shape parameter [29]. Therefore, in the present study, a more representative shape parameter of 0.64 was chosen. A scale parameter of 5.4 was chosen arbitrarily because the variance of the resulting EBVs was standardized before use. The 100 pleiotropic QTLs were assumed to have the same effect for both traits; therefore, the expected genetic correlation between the two traits is 0.5. All QTL effects were assumed to be additive. True breeding values (TBV) were calculated by summing all the QTL effects and subsequently scaling them to a realized genetic variance of 1. Phenotypic value was generated as the sum of the TBVs and a random residual sampled from a normal distribution  $N(0, (1-h^2)/h^2)$ .

### Statistical models

The GEBVs were predicted using the performance and genotype information of the bulls in the reference population. The genomic prediction model used in this study was GBLUP, which is a linear mixed model with a genomic relationship matrix.

The STGM is defined as:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{g} + \mathbf{e} \quad (1)$$

where  $\mathbf{y}$  is the vector of response variables (the conventional EBV was used in this study),  $\mathbf{1}$  is the vector with elements of 1,  $\mu$  is the intercept,  $\mathbf{g}$  is the vector of genomic breeding values,  $\mathbf{Z}$  is the design matrix that associates genomic breeding values with response variables, and  $\mathbf{e}$  is the vector of random residuals. It is assumed that  $\mathbf{g} \sim N(\mathbf{0}, \mathbf{G}\sigma_g^2)$ , where  $\sigma_g^2$  is additive genetic variance and  $\mathbf{G}$  is the realized relationship matrix calculated from the SNP marker information, and  $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$ , where  $\sigma_e^2$

is residual variance and  $\mathbf{I}$  is the  $n \times n$  identity matrix. A detailed description of how  $\mathbf{G}$  is computed can be found in VanRaden [30] and Hayes *et al.* [6].

The MTGM is defined as:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 \end{bmatrix} \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} \quad (2)$$

where  $\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$  is the vector of response variables of traits I and II,  $\mathbf{I}_1$  and  $\mathbf{I}_2$  are the identity matrices,  $\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$  is the vector of intercepts of traits I and II,  $\begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix}$  is the vector of genomic breeding values of the two traits,  $\mathbf{Z}_1$  and  $\mathbf{Z}_2$  are the design matrix that associate genomic breeding values with response variables, and  $\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}$  is the vector of random residuals of the two traits. It is assumed that  $\begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix} \sim N(\mathbf{0}, \mathbf{G}\otimes\mathbf{H})$ , where  $\mathbf{H} = \begin{bmatrix} \sigma_{g_1}^2 & \sigma_{g_{12}} \\ \sigma_{g_{12}} & \sigma_{g_2}^2 \end{bmatrix}$  is the variance and covariance matrix of the genomic breeding values of the two traits, and  $\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} \sim N(\mathbf{0}, \mathbf{I}\otimes\mathbf{R})$ , where  $\mathbf{R} = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_{12}} \\ \sigma_{e_{12}} & \sigma_{e_2}^2 \end{bmatrix}$  is the residual variance and covariance matrix of the two traits.

Although DRP is usually used as the response variable for genomic predictions, in this study, conventional EBVs were used to omit the extra calculation of DRP but without losing the generality of the comparisons between the different scenarios. The EBVs of the animals were estimated from phenotypic data of all the dams in G1–G4 using both a single-trait and a multiple-trait animal model. The models for estimating EBVs were consistent with the genomic prediction models described above, but incorporating a pedigree-based genetic relationship matrix, instead of a genomic relationship matrix. The variance components were estimated from the data and used to calculate the EBVs and GEBVs using the average information REML algorithm (AI-REML). The analysis was executed using the DMU package [31].

### Evaluation of genomic prediction

The evaluation was based on 10 replicates for each scenario and the average of the results was reported. The positions and effects of the QTLs were randomized, and the initial population was generated separately for each

replicate. For each scenario, the reliability of the genomic predictions was measured as the squared correlation between the GEBVs and the TBVs ( $R^2_{GEBV}$ ). Unbiasedness was assessed by regression of the TBVs on the GEBVs. Because EBV is a regressed variable, using EBV as the response variable for genomic prediction will deflate the GEBV in proportion to the reliability of the EBV. GEBV in the real scale can be obtained by scaling GEBV with  $1/r^2_{EBV}$ , where  $r^2_{EBV}$  is the average reliability of EBVs of the animals in the reference population[26]. Therefore, the regression coefficients were calculated based on the original GEBV and the rescaled GEBV. A Hotelling-Williams *t* test [32,33] was used to determine the difference between the validation correlations obtained from the single-trait and multi-trait models.

## Results

### Reliability of EBV and regression coefficient of TBV on EBV

For the animals with no records, the EBVs were predicted from the information of their relatives. For the validation animals, the EBV for trait I ( $h^2 = 0.30$ ) had much higher reliability than the EBV for trait II ( $h^2 = 0.05$ ), as shown in Table 1. The multiple-trait model did not improve the reliability of the EBV for trait I, but significantly increased the prediction accuracy of the EVB for trait II. In addition, the multiple-trait model slightly improved unbiasedness for trait II because, for trait II, the regression coefficient was close to 1; however, the regression coefficients between the single-trait and multiple-trait models were not statistically different.

### Reliability of GEBV

The reliabilities of the GEBVs for animals in the validation population are presented in Table 2. When no trait records were missing, MTGM and STGM generated the same reliability for trait I, while for trait II, MTGM increased the reliability of the GEBV by 0.007 when using the EBV response variable generated by the multi-trait model (EBV\_m) and by 0.033 when using the EBV response variable generated by the single-trait model

(EBV\_s). When there were missing records for trait I, the reliability of GEBV was 0.142 higher with MTGM than with STGM using EBV\_m and 0.111 higher using EBV\_s. When there were missing records in trait II, the reliability of GEBV was 0.181 higher with MTGM than with STGM using EBV\_m and 0.211 higher using EBV\_s. Moreover, the reliability of the GEBVs generated with the same model using EBV\_m as the response variable was higher than their reliability using EBV\_s in all scenarios with missing records.

The Hotelling Williams *t* test showed that the accuracy of genomic prediction using EBV\_m was significantly higher than the accuracy using EBV\_s for traits with missing records (Table 3). The differences in accuracy between the single-trait and multi-trait models were not statistically significant in the scenarios with no missing records.

### Regression of TBV on GEBV

Table 4 shows the regression coefficients of TBV on GEBV in the validation data. For all scenarios, the intercept was close to 0 (in the range  $-0.018$  to  $0.012$ , data not shown). Before rescaling the GEBVs the regression coefficients ( $b_{GEBV\_m}$  and  $b_{GEBV\_s}$ ) were greater than 1 for all scenarios, and this was more serious for trait II (low heritability) than for trait I (high heritability). After rescaling the GEBVs (by dividing the GEBVs by the average reliability of the EBVs), the newly calculated regression coefficients ( $b_{GEBV\_mc}$  and  $b_{GEBV\_sc}$ ) ranged from 0.832 to 1.002 (with an average of 0.907) using EBV\_m, and from 0.964 to 1.069 (with an average of 1.021) using EBV\_s. In general, the differences between the regression coefficients for STGM and MTGM using the same response variables were small.

## Discussion

The main advantage of MTGM over STGM is that MTGM uses information from genetically correlated traits [10,11]. The present study showed that MTGM gave more accurate GEBVs than STGM for the trait with low heritability and for the trait that had missing data when the data for the genetically correlated traits were

**Table 1 Reliability of estimated breeding values (EBVs) and regression coefficients of true breeding value on EBV (The subscripts are the standard deviations of 10 replicates)**

| Trait | Reference animals      |                        | Validation animals     |                        |                        |                        |
|-------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
|       | Single-trait           | Multiple-trait         | Single-trait           |                        | Multiple-trait         |                        |
|       | $R^2_{EBV_s}$          | $R^2_{EBV_m}$          | $R^2_{EBV_s}$          | $b_{EBV\_s}$           | $R^2_{EBV_m}$          | $b_{EBV\_m}$           |
| I     | 0.951 <sub>0.001</sub> | 0.952 <sub>0.001</sub> | 0.348 <sub>0.023</sub> | 0.970 <sub>0.034</sub> | 0.350 <sub>0.024</sub> | 0.971 <sub>0.037</sub> |
| II    | 0.792 <sub>0.004</sub> | 0.822 <sub>0.014</sub> | 0.182 <sub>0.009</sub> | 0.905 <sub>0.074</sub> | 0.229 <sub>0.019</sub> | 0.919 <sub>0.071</sub> |

$R^2_{EBV}$ : the squared correlation between conventional EBVs and true breeding value (TBV) for the reference and validation animals;

$b_{EBV}$ : the regression coefficient of TBV on the EBV for the validation animals;

EBV\_s: EBV response variable estimated using the conventional single-trait linear mixed model;

EBV\_m: EBV response variable estimated using the conventional multi-trait linear mixed model.

**Table 2 Reliability of genomic estimated breeding values (GEBVs) for the validation animals (The subscripts are the standard deviations 10 replicates)**

| Trait | Data type                  | Single-trait           |                        | Multiple-trait         |                        |
|-------|----------------------------|------------------------|------------------------|------------------------|------------------------|
|       |                            | $R^2_{GEBV_m}$         | $R^2_{GEBV_s}$         | $R^2_{GEBV_m}$         | $R^2_{GEBV_s}$         |
| I     | No missing for both traits | 0.874 <sub>0,003</sub> | 0.873 <sub>0,002</sub> | 0.874 <sub>0,001</sub> | 0.873 <sub>0,002</sub> |
|       | Missing for trait I        | 0.474 <sub>0,006</sub> | 0.473 <sub>0,008</sub> | 0.616 <sub>0,004</sub> | 0.584 <sub>0,004</sub> |
| II    | No missing for both traits | 0.718 <sub>0,007</sub> | 0.723 <sub>0,006</sub> | 0.725 <sub>0,005</sub> | 0.756 <sub>0,006</sub> |
|       | Missing for trait II       | 0.373 <sub>0,010</sub> | 0.338 <sub>0,013</sub> | 0.554 <sub>0,009</sub> | 0.549 <sub>0,011</sub> |

$R^2_{GEBV}$ : the squared correlation between GEBV and true breeding value (TBV) for the validation animals using the single-trait and multi-trait models;

$G_{EBV_m}$ : GEBV predicted using the EBV response variable estimated using the conventional multi-trait linear mixed model;

$G_{EBV_s}$ : GEBV predicted using the EBV response variable estimated using the conventional single-trait linear mixed model.

complete. Hayashi and Iwata [12] reported that, compared to single-trait analysis, accuracy was increased with multi-trait analysis for a low heritability trait ( $h^2 = 0.1$ ) that had a high genetic correlation (0.7) with a high heritability trait ( $h^2 = 0.8$ ), using Bayesian variable selection models. In the present study, MTGM was favorable for the trait with a small number of records, which is very important in practical breeding programs because phenotype information for all traits of interest is often not available for all the animals in a reference population. For example, there is usually a limited amount of data for traits that are difficult or expensive to measure, such as carcass, feed efficiency, and disease traits. The accuracies of GEBVs obtained using a STGM will be low for traits with limited phenotypic data. By using information from the correlated and more easily measured traits, a MTGM will improve the accuracies of the GEBVs that are obtained. However, a MTGM will not be distinctly better than a STGM for traits with high heritability and for traits whose complete phenotypic data are available. This result was congruent with the findings of

**Table 3 t values of Hotelling Williams t test in difference between correlations (correlation between genomic prediction and true breeding value) from the single-trait and the multiple-trait models**

| Trait | Data type                  | Hotelling Williams t value |         |
|-------|----------------------------|----------------------------|---------|
|       |                            | GEBV_m                     | GEBV_s  |
| I     | No missing for both traits | 0.000                      | 0.000   |
|       | Missing for trait I        | 7.156**                    | 6.237** |
| II    | No missing for both traits | 0.718                      | 2.701   |
|       | Missing for trait II       | 8.079**                    | 8.883** |

\*\*Significantly different ( $P > 0.01$ ).

GEBV\_m: the genomic estimated breeding value (GEBV) predicted using the EBV response variable estimated from the conventional multi-trait linear mixed model.

GEBV\_s: the GEBV predicted using the EBV response variable estimated from the conventional single-trait linear mixed model.

Hayashi and Iwata [12] who reported that multiple-trait and single-trait analyses made no difference to the prediction accuracy for a trait with high heritability ( $h^2 = 0.8$ ).

In this study, for the validation animals with no missing data, the reliability of GEBVs for the high heritability trait ( $h^2 = 0.3$ ) ranged from 0.873 to 0.874, and for low heritability trait (0.05) the reliability of GEBVs ranged from 0.718 to 0.756. However, the reliability of EBVs for the same animals ranged from 0.348 to 0.350 for the high heritability trait and from 0.182 to 0.229 for low heritability trait. These results showed that the reliability of GEBV was much higher than the reliability of EBV, and there were smaller differences in the reliability of GEBV between the low heritability and high heritability traits compared with the differences in the reliability of EBV. Su *et al.* [9] also reported relatively small difference in reliability of GEBV between milk (high heritability) and fertility (low heritability) in a Danish Holstein population. The small difference in reliability of GEBV for the validation animals between low and high heritability traits indicated that genetic evaluation using genomic prediction is relatively more beneficial for the trait with low heritability, particularly when no records are available for the candidate animals and their offspring (e.g., the pre-selection of young bulls for progeny testing). Therefore, compared with selection based on conventional EBV, genomic selection makes it relatively easier to improve functional traits such as udder health and fertility by reducing the cost of inputs [34], and to obtain a balanced genetic progress between functional traits and production traits.

In the present study, EBVs estimated from a single-trait or multiple-trait model (based on pedigree information) were used as response variables for genomic prediction. The results indicated that the reliability of GEBVs using EBV\_m as the response variable were slightly higher than their reliability using EBV\_s, except for trait II with no missing data. The reliability of GEBVs using EBV\_m improved by 0.1% to 3.5% points compared with their reliability using EBV\_s, because, in the multi-trait model, information about the correlated trait was used. Because the correlated trait data were used to estimate EBV\_m, it can be argued that a multiple-trait model will be better than a single-trait model for genomic prediction when EBV\_m is used as the response variable. The results from this study showed that even when the response variable was obtained from the multiple-trait model, the reliability of genomic prediction using MTGM was further improved over the reliability of STGM. A possible reason could be that the use of information from the correlated trait may not be exactly the same in the conventional BLUP and GBLUP methods. Thus, additional information from the correlated trait may be available for genomic prediction of the target trait, even though the response

**Table 4 Regression coefficients of true breeding values on genomic estimated breeding values for the validation animals (The subscripts are the standard deviations of 10 replicates)**

| Trait | Data type                  | Single-trait           |                        |                        |                        | Multiple-trait         |                        |                        |                        |
|-------|----------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
|       |                            | $b_{\text{GEBV}_m}$    | $b_{\text{GEBV}_{mc}}$ | $b_{\text{GEBV}_s}$    | $b_{\text{GEBV}_{sc}}$ | $b_{\text{GEBV}_m}$    | $b_{\text{GEBV}_{mc}}$ | $b_{\text{GEBV}_s}$    | $b_{\text{GEBV}_{sc}}$ |
| I     | No missing for both traits | 1.105 <sub>0,016</sub> | 1.002 <sub>0,015</sub> | 1.108 <sub>0,017</sub> | 1.006 <sub>0,016</sub> | 1.105 <sub>0,016</sub> | 1.003 <sub>0,015</sub> | 1.112 <sub>0,017</sub> | 1.008 <sub>0,016</sub> |
|       | Missing for trait I        | 1.066 <sub>0,029</sub> | 0.959 <sub>0,032</sub> | 1.071 <sub>0,034</sub> | 0.964 <sub>0,037</sub> | 1.089 <sub>0,029</sub> | 0.980 <sub>0,031</sub> | 1.084 <sub>0,046</sub> | 0.975 <sub>0,048</sub> |
| II    | No missing for both traits | 1.240 <sub>0,049</sub> | 0.875 <sub>0,038</sub> | 1.486 <sub>0,060</sub> | 1.048 <sub>0,054</sub> | 1.250 <sub>0,049</sub> | 0.880 <sub>0,038</sub> | 1.516 <sub>0,059</sub> | 1.069 <sub>0,055</sub> |
|       | Missing for trait II       | 1.212 <sub>0,116</sub> | 0.848 <sub>0,089</sub> | 1.444 <sub>0,209</sub> | 1.010 <sub>0,156</sub> | 1.190 <sub>0,067</sub> | 0.832 <sub>0,042</sub> | 1.567 <sub>0,121</sub> | 1.095 <sub>0,082</sub> |

$b_{\text{GEBV}}$ : the regression coefficient of true breeding values (TBV) on the estimated breeding values (EBVs) for the validation animals.

$\text{GEBV}_m$ : the genomic estimated breeding value (GEBV) predicted using the EBV response variable calculated from the conventional multi-trait linear mixed model.

$\text{GEBV}_s$ : the GEBV predicted using the EBV response variable calculated from the conventional single-trait linear mixed model.

$\text{GEBV}_{mc}$ :  $\text{GEBV}_m$  divided by the average reliability of EBV for the reference animals.

$\text{GEBV}_{sc}$ :  $\text{GEBV}_s$  divided by the average reliability of EBV for the reference animals.

variable was obtained using the information of the correlated trait. A more important reason is that the prediction of GEBVs for the validation animals can use the information of the correlated trait directly through the covariance structure among the traits.

In the simulations, all the phenotypic records were available for the estimation of EBVs. The data with missing records for genomic prediction were generated by masking some of the EBVs. In a real world application however, missing phenotypes cannot be used to estimate EBVs. Therefore, in the scenarios with missing records, the prediction accuracy that was obtained in this study could be higher than the accuracy expected in the real world. On the other hand, the gain from using a multi-trait model in a real scenario with missing records might be larger than the gain from the simulated scenario, because the amount of information of the related trait in relation to the trait of interest would be relatively larger.

In this study, a GBLUP model was used for multiple-trait genomic prediction. Compared with the Bayesian variable selection models, the advantages of the multiple-trait GBLUP model are its low computational demand and its straightforward implementation using existing methodologies and standard linear mixed model software. However, the GBLUP model assumes that the effects of all the SNPs have the same normal distribution for a trait, and that the covariance between traits is the same for all SNP effects. This assumption may be not very appropriate and could limit the advantage of the GBLUP multiple-trait genomic prediction method. Hayashi and Iwata [12] used Bayesian variable selection models for multiple-trait genomic prediction. The Bayesian models assume a proportion of the SNP markers have an effect while others have a null effect, which describes the property of SNP effects more appropriately. However, their models assume a SNP either has an effect on all traits in the model or has no effect on any of the traits, and that the covariance between traits is the same for all SNP effects. In a real life scenario, different traits are affected by different sets of genes, and some genes have effects on more than one trait. More

sophisticated models that can account for these features are required to fully exploit the advantages of multiple-trait genomic prediction.

## Conclusions

The results reported here suggest that a MTGM can improve the accuracy of genomic predictions, especially for low heritability traits and for traits with only a small amount of phenotype data. The EBV response variables derived from the multiple-trait and single-trait models had a relatively small influence on the reliability of GEBVs for the trait without missing data. However, for the trait with missing data, the response variable obtained from the multiple-trait model gave better genomic predictions than the response variable obtained from the single-trait model. Even when response variables derived from the multiple-trait model were used, the genomic prediction using MTGM still generated GEBVs with higher reliability than the GEBVs generated using STGM.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

GG and FZ performed the data analyses and drafted the manuscript. YW assisted in the study design. GS and YZ reviewed the manuscript. LD and GS conceived and designed the study as well as co-supervised the work. All authors read and approved the final manuscript.

## Acknowledgements

We acknowledge the three anonymous reviewers for insightful comments on an earlier version of the manuscript. This work was supported by the National Natural Science Foundation of China (Grant No. 31200927), the Chinese National Modern Agricultural Industry Technology Fund for Scientists in Sheep Industry System (Grant No. CARS-39-04B), the Chinese National Key Project (Grant Nos. 2011BAD28B02, 2012BAD12B06), the Chinese National Nonprofit Institute Research Grant (Grant No. 2012cj-2), and the Green Development and Demonstration Programme, Denmark for the "Genomic Selection — From function to efficient utilization in cattle breeding" project (Grant No. 3405-10-0137).

## Author details

<sup>1</sup>National Center for Molecular Genetics and Breeding of Animal, Institute of Animal Sciences, Chinese academy of Agricultural Sciences, Beijing 100193, China. <sup>2</sup>Beijing Sanyuan Lvhe Dairy Cattle Breeding Center, Beijing 100076,

China. <sup>3</sup>College of Animal Science and Technology, China Agricultural University, Beijing 100193, China. <sup>4</sup>Department of Genetics and Biotechnology, Faculty of Agricultural Sciences, Aarhus University, Tjele DK-8830, Denmark.

Received: 19 November 2013 Accepted: 26 February 2014  
Published: 4 March 2014

## References

1. Meuwissen THE, Hayes BJ, Goddard ME: **Prediction of total genetic value using genome-wide dense marker maps.** *Genetics* 2001, **157**(4):1819–1829.
2. Schaeffer L: **Strategy for applying genome-wide selection in dairy cattle.** *J Anim Breed Genet* 2006, **123**(4):218–223.
3. Weigel K, De Los CG, González-Recio O, Naya H, Wu X, Long N, Rosa G, Gianola D: **Predictive ability of direct genomic values for lifetime net merit of Holstein sires using selected subsets of single nucleotide polymorphism markers.** *J Dairy Sci* 2009, **92**(10):5248.
4. Vazquez A, Rosa G, Weigel K, De Los Campos G, Gianola D, Allison D: **Predictive ability of subsets of single nucleotide polymorphisms with and without parent average in US Holsteins.** *J Dairy Sci* 2010, **93**(12):5942.
5. Makowsky R, Pajewski NM, Klimentidis YC, Vazquez AI, Duarte CW, Allison DB, De Los Campos G: **Beyond missing heritability: prediction of complex traits.** *PLoS genetics* 2011, **7**(4):e1002051.
6. Hayes B, Bowman P, Chamberlain A, Goddard M: **Invited review: genomic selection in dairy cattle: progress and challenges.** *J Dairy Sci* 2009, **92**(2):433–443.
7. VanRaden P, Van Tassell C, Wiggans G, Sonstegard T, Schnabel R, Taylor J, Schenkel F: **Invited review: reliability of genomic predictions for North American Holstein bulls.** *J Dairy Sci* 2009, **92**(1):16–24.
8. Harris B, Johnson D: **Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation.** *J Dairy Sci* 2010, **93**(3):1243–1252.
9. Su G, Guldbandsen B, Gregersen V, Lund M: **Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population.** *J Dairy Sci* 2010, **93**(3):1175–1183.
10. Calus MPL, Veerkamp RF: **Accuracy of multi-trait genomic selection using different methods.** *Genet Sel Evol* 2011, **43**(1):1–14.
11. Jia Y, Jannink JL: **Multiple-trait genomic selection methods increase genetic value prediction accuracy.** *Genetics* 2012, **192**(4):1513–1522.
12. Hayashi T, Iwata H: **A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits.** *BMC bioinformatics* 2013, **14**(1):1–14.
13. Aguilar I, Misztal I, Tsuruta S, Wiggans G, Lawlor T: **Multiple trait genomic evaluation of conception rate in Holsteins.** *J Dairy Sci* 2011, **94**(5):2621–2624.
14. Tsuruta S, Misztal I, Aguilar I, Lawlor T: **Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins.** *J Dairy Sci* 2011, **94**(8):4198–4204.
15. Meuwissen T, Goddard M: **Accurate prediction of genetic values for complex traits by whole-genome resequencing.** *Genetics* 2010, **185**(2):623–631.
16. Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA: **The impact of genetic architecture on genome-wide evaluation methods.** *Genetics* 2010, **185**(3):1021–1031.
17. Clark SA, Hickey JM, van der Werf JH: **Different models of genetic variation and their effect on genomic evaluation.** *Genet Sel Evol* 2011, **43**:18.
18. Coster A, Bastiaansen JW, Calus MP, Van Arendonk JA, Bovenhuis H: **Sensitivity of methods for estimating breeding values using genetic markers to the number of QTL and distribution of QTL variance.** *Genet Sel Evol* 2010, **42**:9.
19. Li Z, Sillanpää MJ: **Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection.** *Theor Appl Genet* 2012, **125**(3):419–435.
20. Zhong S, Dekkers JCM, Fernando RL, Jannink JL: **Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study.** *Genetics* 2009, **182**(1):355–364.
21. Rius-Vilarrasa E, Brøndum R, Strandén I, Guldbandsen B, Strandberg E, Lund M, Fikse W: **Influence of model specifications on the reliabilities of genomic prediction in a Swedish–Finnish red breed cattle population.** *J Anim Breed Genet* 2012, **295**(5):369–397.
22. De Los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MP: **Whole-genome regression and prediction methods applied to plant and animal breeding.** *Genetics* 2013, **193**(2):327–345.
23. Ober U, Ayroles JF, Stone EA, Richards S, Zhu D, Gibbs RA, Stricker C, Gianola D, Schlather M, Mackay TFC: **Using whole-genome sequence data to predict quantitative trait phenotypes in *Drosophila melanogaster*.** *PLoS genetics* 2012, **8**(5):e1002685.
24. Li Z, Sillanpää MJ: **Estimation of quantitative trait locus effects with epistasis by variational Bayes algorithms.** *Genetics* 2012, **190**(1):231–249.
25. Su G, Madsen P, Nielsen US, Mäntysaari EA, Aamand GP, Christensen OF, Lund MS: **Genomic prediction for Nordic Red Cattle using one-step and selection index blending.** *J Dairy Sci* 2012, **95**(2):909–917.
26. Guo G, Lund M, Zhang Y, Su G: **Comparison between genomic predictions using daughter yield deviation and conventional estimated breeding value as response variables.** *J Anim Breed Genet* 2010, **127**(6):423–432.
27. Jairath L, Dekkers J, Schaeffer L, Liu Z, Burnside E, Kolstad B: **Genetic evaluation for herd life in Canada.** *J Dairy Sci* 1998, **81**(2):550–562.
28. Hayes B, Goddard ME: **The distribution of the effects of genes affecting quantitative traits in livestock.** *Genet Sel Evol* 2001, **33**(3):209–230.
29. Allison DB, Fernandez JR, Heo M, Zhu S, Etzel C, Beasley TM, Amos CI: **Bias in estimates of quantitative-trait–locus effect in genome scans: demonstration of the phenomenon and a method-of-moments procedure for reducing bias.** *Am J Hum Genet* 2002, **70**(3):575–585.
30. VanRaden P: **Efficient methods to compute genomic predictions.** *J Dairy Sci* 2008, **91**(11):4414–4423.
31. Madsen P, Jensen J: **DMU: A user's guide. A package for analysing multivariate mixed models. Version 6, release 4.7.** 2007. [http://dmu.agrsci.dk/DMU/Doc/Current/dmuv6\\_guide.5.2.pdf](http://dmu.agrsci.dk/DMU/Doc/Current/dmuv6_guide.5.2.pdf) Accessed Nov. 15, 2007.
32. Steiger JH: **Tests for comparing elements of a correlation matrix.** *Psychol Bull* 1980, **87**(2):245.
33. Dunn OJ, Clark V: **Comparison of tests of the equality of dependent correlation coefficients.** *J Am Stat Assoc* 1971, **66**(336):904–908.
34. Coren A, Steine T, Colleau J, Pedersen J, Pribyl J, Reinsch N: **Economic values in dairy cattle breeding, with special reference to functional traits. Report of an EAAP-working group.** *Livest Prod Sci* 1997, **49**:1–21.

doi:10.1186/1471-2156-15-30

Cite this article as: Guo et al.: Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genetics* 2014 **15**:30.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

