Ann. Anim. Sci., Vol. 17, No. 3 (2017) 683–701 *DOI: 10.1515/aoas-2016-0086*

DE GRUYTER
OPEN

GENOMIC PREDICTION BY CONSIDERING GENOTYPE × ENVIRONMENT INTERACTION USING DIFFERENT GENOMIC ARCHITECTURES

Mehdi Bohlouli[1♦], Sadegh Alijani[1], Ardashir Nejati Javaremi[2], Sven König[3], Tong Yin[3]

[1]Department of Animal Science, University of Tabriz, Tabriz, East Azerbaijan, Iran
[2]Department of Animal Science, University College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran
[3]Institute of Animal Breeding and Genetics, University of Giessen, Giessen, Germany
♦Corresponding author: m.bohlouli@tabrizu.ac.ir

Abstract

In this study, accuracies of genomic prediction across various scenarios were compared using single-trait and multiple-trait animal models to detect genotype × environment (G × E) interaction based on REML method. The simulated high and low linkage disequilibrium (HLD and LLD) genome consisted of 15,000 and 50,000 SNP chip applications with 300 and 600 QTLs controlling the trait of interest. The simulation was done to create the genetic correlations between the traits in 4 environments and heritabilities of the traits were 0.20, 0.25, 0.30 and 0.35 in environments 1, 2, 3 and 4, respectively. Two strategies were used to predict the accuracy of genomic selection for cows without phenotypes. In the first strategy, phenotypes for cows in three environments were kept as a training set and breeding values for all animals were estimated using three-trait model. In the second one, only 25, 50 or 75% of records in the fourth environment and all the records in the other three environments were used to predict GBV for non-phenotyped cows in the environment 4. For the first strategy, the highest accuracy of 0.695 was realized in scenario HLD with 600 QTL and 50K SNP chip for the fourth environment and the lowest accuracy of 0.495 was obtained in scenario LLD with 600QTL and 15K SNP chips for the first environment. Generally, the accuracy of prediction increased significantly ($P<0.05$) with increasing the number of markers, heritability and the genetic correlation between the traits, but no significant difference was observed between scenarios with 300 and 600 QTL. In comparison with models without G × E interaction, accuracies of the GBV for all environments increased when using multi-trait models. The results showed that the level of LD, number of animals in training set and genetic correlation across environments play important roles if G × E interaction exists. In conclusion, G × E interaction contributes to understanding variations of quantitative trait and increasing accuracy of genomic prediction. Therefore, the interaction should be taken into account in conducting selection in various environments or across different genotypes.

Key words: genomic selection, genotype by environment interaction, linkage disequilibrium, simulation

Because of the widely used artificial insemination in dairy cattle breeding programs, progenies of bulls can be spread in environments with completely different climates, e.g. regions and countries. Breeding values of the bulls may not be constant across the environments (Calus et al., 2002; Hammami et al., 2009), which indicates the existence of genotype × environment (G × E) interaction. The G × E interaction occurs when genotypes react differently in different environments (Falconer and MacKay, 1996), implying that some genes might have different expressions according to the environments. These environment-dependent genes can be interesting and detected by including an interaction term between quantitative trait loci (QTL) and environment in the animal model (Lillehammer et al., 2008). Furthermore, including such interaction effect might increase the power of QTL detection (Lund et al., 2014), and consequently, benefit the selection progress in breeding process (Lillehammer et al., 2008). Many studies have reported the existence of G × E interaction for milk production traits in dairy cattle, while, in most of the studies, the environments were measured as discrete scales, e.g., the herd production levels (Calus et al., 2002; Kolmodin et al., 2002; Hammami et al., 2009) and different regions (Hammami et al., 2009; Bohlouli et al., 2014). However, Brügemann et al. (2011) and Bohlouli et al. (2013) also created continuous environmental descriptor, i.e. temperature-humidity index. Therefore, both multiple-trait and random regression models have been used to detect G × E interaction for milk production traits (Hammami et al., 2009; Brügemann et al., 2011; Bohlouli et al., 2014), by using phenotypic and pedigree information.

Genomic selection (GS), as introduced by Meuwissen et al. (2001), is a new approach for improving selection of quantitative traits in animal breeding. In GS, genome-wide marker data is used to trace all the QTL controlling the trait of interest and to predict genetic merit of animals by accounting for identity by state and the variation in relationship (Nejati-Javaremi et al., 1997). The acceptable accuracy of genomic prediction for young bulls with the shorter generation intervals due to early selection improves the genetic gain in dairy cattle (Dekkers, 2007; König et al., 2009). Genomic data can also be integrated into the models with G × E interaction, where the gene expression changes in the course of temperature-humidity index (Yin et al., 2014) or between different breeds and environments (Lillehammer et al., 2007; Lund et al., 2014). Yin et al. (2014) reported that applying random regression model with genotype data might be applicable to predict the genomic breeding values (GBV) for traits measured late in a dairy cow's life. Lillehammer et al. (2007) mentioned that the accuracy of genomic selection depends not only on its average effect but also on its interaction with the environment.

In GS, simulated data allow the researcher to explore the influences of the genetic architecture of the trait (Pimentel et al., 2013), the number of markers used for analysis (Solberg et al., 2008; Yin et al., 2014), and the data also allows for evaluating some sources of variability, such as drift, which cannot be assessed with the most of real data (Daetwyler et al., 2010). Recently, many strategies for simulation have been applied to compare accuracies of genomic predictions. Within this context, accuracies of genomic predictions strongly depend on the quantity and quality of phenotypic data (e.g. Yin et al., 2014), genetic parameter estimates, e.g. heritabil-

ity and genetic correlation (e.g. Hayes et al., 2009; Calus et al., 2013), the genomic architecture of the trait (e.g. Daetwyler et al., 2010) and population pedigree structures (e.g. Hickey and Gorjanc, 2012). Researchers have reported the accuracies of different (single- and multi-trait) genomic models using various scenarios (Calus and Veerkamp, 2011; Jiang et al., 2015). It has been shown that a multiple-trait genomic model had higher prediction accuracy than a single-trait genomic model (Jia and Jannink, 2012). Using multiple-trait genomic model to investigate G×E interaction, GS can accelerate genetic gains in predicted future climates (Hayes et al., 2016).

With regard to farms that are located in different environments, G × E interactions are not considered frequently in the simulation studies. Accordingly, the objectives of this study were 1) to simulate phenotypic data with quantifying the G × E interaction to follow the (co)variance matrices across different environments that were used by Bohlouli et al. (2014), and 2) to investigate the accuracy of genomic predictions for cows in the 4 environments and for non-phenotyped cows in an environment. For different data structure, the comparisons of accuracy of genomic prediction were carried out using different genomic architectures.

Material and methods

Simulation

Phenotypes and genotypes were simulated based on forward-in-time process using QMSim software (Sargolzaei and Schenkel, 2009). According to the simulation study from Yin et al. (2014), two different types of historical populations were simulated to create high linkage disequilibrium (HLD) or low LD (LLD). To create HLD between the markers and a QTL, 2500 generations were simulated with a constant size of 2000, followed by 70 generations with a gradual decrease in population size to 200 individuals. After 10 generations, the population size increased to 4040 in generation 2600 and then constant size of 4040 individuals to generation 2620. For scenarios with LLD, 1600 generations were considered with a constant size of 4000 and then increased to 4040 in generation 1620. Subsequently, constant size of 4040 individuals was simulated to generation 1640.

There were 40 sires in the last historical generation which were used as founders to create desirable population structure to mimic artificial insemination (AI) in dairy cattle population with many individuals but a low effective population size. In the second step, animals from the last historical generation were used as founders to simulate 10 recent generations for both HLD and LLD scenarios. The replacement rate was 50% for sires and 25% for dams and each mating produced a single progeny with 50% probability of being male. The simulated genome consisted of 30 chromosomes of 100 cM each. Each chromosome contained 500 or 1,667 bi-allelic markers to mimic 15000 (15k) or 50000 (50k) SNP chips. Therefore, two scenarios were simulated according to the number of markers on the genome. The number of QTLs was set to 10 or 20 per chromosome indicating 300 or 600 QTLs in whole genome. The positions of the markers and QTLs were randomized on the chromosomes and

for both of them, equal allele frequencies were assumed in the beginning of historical population and the mutation rate was assumed 2.5×10^{-5} per locus per generation. Effects of QTL alleles were drawn from a gamma distribution with a shape parameter 0.4 and the number of QTL alleles at each locus was randomly assigned. The parameters of simulation process are summarized in Table 1. QTL effects are restricted to being additive and sampled from normal or gamma distributions.

Table 1. Parameters of the simulation process

Population structure	Low LD[1]	High LD
Historical population		
no. of generations (size)	1640	2620
no. of animals in first generation	4040	2000
bottleneck	No	Yes[2]
no. of animals in last generation	4040	
Current population		
no. of founder males	40	
no. of founder females	4000	
no. of generations	10	
no. of offspring per mate	1	
probability for sex of the offspring	0.5	
selection and mating designs	Random	
replacement ratio for males	50%	
replacement ratio for females	25%	
culling criteria	Age	
Genome		
no. of chromosomes	30	
length of each chromosome (cM)	100	
no. of QTL per chromosome	10 or 20	
effects of QTL alleles	Gamma	
no. of bi-allelic markers per chromosome	500 or 1667	
marker and QTL mutation rate	2.5×10^{-5}	
marker and QTL allele frequencies	Equal	
crossover interference (cM)	25	
position of markers and QTL	Random	

[1]LD = linkage disequilibrium.
[2]Population size was 200 from generation 2570 to 2580.

Simulation environments

In the study by Bohlouli et al. (2014), common sires were selected with recorded daughters in different climatic conditions (warm, moderate, semi-cold and cold climates) to create genetic connectedness between environments. In order to follow the data structure that was created by Bohlouli et al. (2014), an R code written by Yin et al. (2014) was applied to modify the QMSim outputs. Genotypes of 2000 cows in the 10th recent generation were used in the analysis which they were progenies from 40 sires, with about half-sib group of 50 daughters per sire and the pedigree file included all animals in the 10 recent generations.

Following Yin et al. (2014), QTLs (300 or 600 QTLs) were randomly grouped into 10 classes. QTLs in classes 1 to 7, 2 to 8, 3 to 9 and 4 to 10 were expressed in environment 1, 2, 3 and 4, respectively, to mimic the different gene expressions in the 4 environments. Simultaneously, the simulation procedure could also establish the genetic correlations between environments by overlapping QTL groups in the 4 environments. True breeding values (TBV) of the animals in each environment equaled to the sum of the QTL effects of every cow in the corresponding environments. Phenotypes were created by adding a residual to TBV of the cows in the 4 environments. The heritabilities were 0.20, 0.25, 0.30 and 0.35 for the trait of interest in the 4 environments. 2000 phenotyped cows were assigned to the four environments, indicating that each environment contained 500 cows. At least 10 daughters per sire should be kept in each environment. Genetic correlations were obtained using correlations of TBVs of cows across environments and phenotypic correlations were calculated using average phenotypic records per sire in different environments. Heritabilities and average genetic and phenotypic correlations between 4 different environments are shown in Table 2.

Table 2. Heritabilities (diagonal and bold) and genetic (above diagonal) and phenotypic (below diagonal) correlations of the simulated data between different environments

Environment	1	2	3	4
1	**0.20**	0.84 (0.06)	0.67 (0.07)	0.50 (0.13)
2	0.38 (0.16)	**0.25**	0.82 (0.04)	0.63 (0.08)
3	0.31 (0.12)	0.40 (0.10)	**0.30**	0.81 (0.05)
4	0.22 (0.17)	0.30 (0.19)	0.48 (0.12)	**0.35**

Standard deviations are shown in parentheses.

For milk production traits, Bohlouli et al. (2014) reported that the lowest and the highest heritabilities were found in warm and cold regions, respectively, which were in line with the heritabilities in environment 1 and 4 of this study. Moreover, the lowest genetic correlation of 0.5 between two extreme environments reflects the physiological and practical background of this study. Usually, low genetic correlations across environments (<0.80) signify the existence of G × E interaction (Robertson, 1959).

Quality control

Quality Control (QC) was applied using preGSf90 program (Aguilar et al., 2011). SNPs with minor allele frequency (MAF) lower than 0.01 were deleted. To test the Hardy Weinberg Equilibrium (HWE), the SNP was excluded in cases the difference between observed and expected genotype frequencies was >0.15 (default value). For markers and QTLs, LLD scenarios had normal allele frequency distributions. But distributions of HLD scenarios were nearly U-shape, and after the QC, those became relatively normal distribution. Finally, for LLD and HLD scenarios, about 0.5% and 15% of SNPs were removed, respectively; and also the same ratios of non-segregating QTLs (fixed loci) were found when creating phenotypic records.

Linkage disequilibrium

The level of LD in the simulated scenarios can be assessed by calculating the squared correlation coefficient (r^2) between all possible pairs of markers (Hill and Robertson, 1968):

$$r^2 = \frac{D}{f(A)f(a)f(B)f(b)}$$

where:

$D = f(AB)\text{-}f(A)f(B)$, and $f(AB)$, $f(A)$, $f(a)$, $f(B)$, $f(b)$ are observed frequencies of haplotypes AB and of alleles A, a, B, b, respectively.

The *PLINK* software (Purcell et al., 2007) was used to estimate LD between marker pairs of 2000 cows in the last generation.

Statistical models

A multiple-trait animal model was applied to analyze records in the 4 environments by assuming the phenotypes measured in different environments are correlated traits (Ceron-Munoz et al., 2004; Hammami et al., 2009). The following four-trait genomic model was fitted to estimate variance components (Hayes et al., 2016):

$$\begin{bmatrix} y_1 \\ \vdots \\ y_4 \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \boldsymbol{I}_4 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_4 \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \boldsymbol{Z}_4 \end{bmatrix} \begin{bmatrix} g_1 \\ \vdots \\ g_4 \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_4 \end{bmatrix}$$

where:

y_i is observation for i^{th} trait (i=1 to 4 for the trait in the 4 environments),

$\boldsymbol{I}_i$ is the identity matrix,

μ_i is the overall mean for i^{th} trait,

$\boldsymbol{Z}_i$ is the design matrix for g_i,

g_i is the vector of genomic breeding values (GBV) of animals (phenotyped cows and their relatives) in i^{th} environment,

e_i is the vector of random residual effects.

For the first trait with 500 phenotypic records, the dimensions of I_i and Z_i were 500×500 and 500×number of animals in pedigree, respectively. It is assumed that $g_i \sim N(0,\mathbf{H}\otimes\mathbf{T})$, $e_i \sim N(0,\mathbf{H}\otimes\mathbf{R})$ and

$$\boldsymbol{T} = \begin{bmatrix} \sigma^2_{g1} & \cdots & \sigma_{g14} \\ \vdots & \ddots & \vdots \\ \sigma_{g41} & \cdots & \sigma^2_{g4} \end{bmatrix} \quad \text{and} \quad \boldsymbol{R} = \begin{bmatrix} \sigma^2_{e1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2_{e4} \end{bmatrix}$$

Under the single-step genomic best linear unbiased prediction model (Aguilar et al., 2010), the matrix **H** is the combination of the numerator relationship matrix (**A)** and the genomic relationship matrix (**G)** matrices to include non-genotyped animals, and **T** and **R** are the (co)variance matrices for additive genetic effect and residual of the four traits, respectively. The inverse of **H** was defined as:

$$\boldsymbol{H^{-1}} = \boldsymbol{A^{-1}} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G^{-1}} - \boldsymbol{A_{22}^{-1}} \end{bmatrix}$$

G matrix was calculated based on the approach of VanRaden (2008):

$$\boldsymbol{G^{*}} = \frac{\boldsymbol{ZZ'}}{2 \times \sum_{k=1}^{m} p_k (1 - p_k)}$$

where:

$\boldsymbol{Z} = (\boldsymbol{M} - \boldsymbol{2}_{pk})$ and $\boldsymbol{M}$ contained number of the second allele (0, 1, and 2) matrix with dimensions of the number of individual by the number of SNPs *(m)*,

p_k was frequency of the second allele in current population.

Matrix $\boldsymbol{G^*}$ was scaled so that the means of diagonals and off-diagonals are the same as in $\boldsymbol{A_{22}}$ (sub-matrix of $\boldsymbol{A}$ for genotyped animals) matrix and then combined with $\boldsymbol{A_{22}}$ to $\boldsymbol{G} = 0.95 \times \boldsymbol{G^*} + 0.05 \times \boldsymbol{A_{22}}$ in order to make invertible matrices.

Results of the multiple-trait animal model were compared to results obtained from a single-trait animal model, which consider the traits in different environments were still the same traits and only one more fixed effect indicated environment was added in the model. All the analyses were done using AIREMLF90 program (Misztal et al., 2002) via Average Information (AI) algorithm in the Restricted Maximum Likelihood (REML) method.

Accuracy of genomic prediction

Correlations between TBV and GBV estimated from the single-trait and the multiple-trait animal models for the 2000 cows were considered as the evaluation criteria. Two strategies were used to predict the accuracy of genomic selection for genotyped cows without phenotypes (validation set). In the first strategy, phenotypes for animals in three environments were kept as a training set and GBVs for all animals, including non-phenotyped but genotyped cows located in the removed environment, were estimated using three-trait genomic model. In the other words, for each animal in the validation set, three GBVs were estimated in the three environments. Then accuracies of genomic predictions were obtained using TBVs of 500 cows in the validation set and GBVs of those cows in the other three environments. In the second one, only 25, 50 or 75% of records in the fourth environment and all the records in the other three environments were used in the training set and GBVs were estimated for non-phenotyped cows at the fourth environment. The accuracies of genomic predictions were correlations between TBV and GBV for animals. The evaluation was done using 10 replicates for each scenario, and the means of different scenarios were compared using ANOVA procedure by the Duncan test at $P<0.05$ in SAS software (Statistical Analysis System, 2003).

Results

Linkage disequilibrium

The level of linkage disequilibrium was calculated as the average r^2 value for all possible pairs of markers for both LLD and HLD scenarios with 15K and 50K SNP chips across 10 replicates. Figure 1 plots the average r^2 with map distance up to 2 Megabase pair (*Mbp*) on the first chromosome. For all scenarios, the average r^2 decreased with the increasing marker distance. The r^2 for HLD scenarios were larger than for LLD scenarios when the distance between 2 SNP markers was small. However, with the increase of the distances between SNPs, especially for distances larger than 5.0 *Mbp* (not shown in the Figure), no difference in r^2 was found between HLD and LLD scenarios. The average r^2 was smaller than 0.05 when distances between the two markers were greater than 0.4 *Mbp* for LLD scenarios and 1.1 *Mbp* for HLD scenarios. The average r^2 for all scenarios was 0.01, when distances between two markers varied in the range of 5 and 100 *Mbp*.

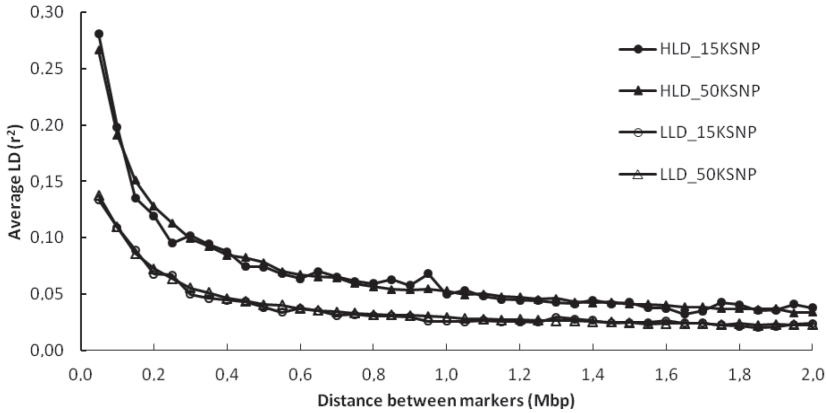

Figure 1. Average linkage disequilibrium (LD) measured by squared correlation coefficient (r^2) between SNP pairs against their map distance for different scenarios. High or low linkage disequilibrium (HLD or LLD) and different SNP density (15KSNP or 50KSNP)

Accuracies of genomic prediction

A: Without considering genotype by environment interaction

GBVs were predicted for cows with phenotypes and genotypes, which consider the traits in different environments were still the same trait and the single-trait models were used under 8 scenarios (HLD or LLD; 300QTL or 600QTL; 15KSNP or 50KSNP) listed in Table 3. Accuracies of GBVs were estimated using TBVs of cows in related environment using the averaged values from 10 replicates. Generally, accuracies of GBV increased with the increase of heritabilities. Accuracies for scenarios with HLD were higher than those for LLD scenarios. The accuracy of GBV was higher when marker density was increased from 15K SNP to 50K SNP. Doubling the number of simulated QTL from 300 to 600 had no significant impact ($P>0.05$) on the accuracy. Across all environments, the scenario with LLD, 600 QTLs and the low-density SNP chip (LLD_600Q_15KSNP) had the lowest accuracy.

Table 3. Accuracy and standard deviation (in parenthesis) of genomic predictions for animals in 4 environments via single-trait animal model. Scenarios include high or low linkage disequilibrium (HLD or LLD), different QTL number (300QTL or 600QTL) and different SNP density (15KSNP or 50KSNP)

Scenario	Environment (Heritability)			
	1 (0.20)	2 (0.25)	3 (0.30)	4 (0.35)
HLD_300QTL_15KSNP	0.541 ab B (0.061)	0.581 a A (0.026)	0.587 ab A (0.043)	0.557 abc AB (0.037)
HLD_300QTL_50KSNP	0.569 a B (0.030)	0.593 a AB (0.046)	0.602 a A (0.035)	0.606 a A (0.017)
HLD_600QTL_15KSNP	0.518 bc A (0.066)	0.528 bc A (0.067)	0.557 bc A (0.060)	0.548 bc A (0.049)
HLD_600QTL_50KSNP	0.541 ab A (0.060)	0.582 a A (0.059)	0.571 abc A (0.048)	0.575 ab A (0.087)
LLD_300QTL_15KSNP	0.514 bc A (0.030)	0.530 bc A (0.032)	0.532 cd A (0.032)	0.510 c A (0.058)
LLD_300QTL_50KSNP	0.488 cd B (0.039)	0.521 c AB (0.041)	0.556 bc A (0.055)	0.510 c A (0.034)
LLD_600QTL_15KSNP	0.460 d B (0.039)	0.493 c AB (0.033)	0.511 d A (0.051)	0.513 c A (0.059)
LLD_600QTL_50KSNP	0.535 abc A (0.048)	0.565 ab A (0.034)	0.558 bc A (0.028)	0.531 bc A (0.058)

Means without common letters (lowercase and uppercase letters stand for comparison within column and row means, respectively) are statistically different ($P<0.05$).

B: With considering genotype by environment interaction

Cows with phenotypes

The four-trait animal model was applied when considering cows in 4 environments as different traits. In comparison with accuracies from single-trait model (Table 3), accuracies of GBV estimated from four-trait model (Table 4) were increased in all scenarios and all environments. Generally, the accuracy of prediction decreased significantly when LLD instead of HLD. Traits with higher heritabilities also had higher accuracies ($P<0.05$). Therefore, for each scenario, accuracies of GBV were the highest for heritability of 0.35 in the fourth environment, because the highest heritability was assigned in this environment. In some cases, 50K SNP chips scenario was not significantly better than 15K SNP chips scenario, but, basically, GBVs were more accurate in scenarios with 50K SNP chips (Table 4). The highest accuracy of 0.695 was realized in the fourth environment of the scenario HLD_600QTL_50KSNP and the lowest accuracy of 0.495 was found in the first environment of the scenario LLD_600QTL_15KSNP.

Cows without phenotypes

In the first strategy, the records of all cows from one of the 4 environments were removed and consequently, data in the remaining 3 environments was analyzed via three-trait animal model. In the second strategy, 25, 50 and 75% of the records from the fourth environment were kept out, and prediction was done via four-trait animal model. In the two strategies, the GBVs were predicted from the information of their relatives in the same or the other environments. The average accuracies of genomic predictions for non-phenotyped cows in different environments are shown in Table 5. The accuracies of GBV estimated from the three-trait model (Table 5) were smaller than that estimated via four-trait models (Table 4 and Table 6). Generally, the accuracy of prediction increased significantly ($P<0.05$) when comparing results from LLD to HLD. Without phenotypic records for the environment (the first environment as the validation set), accuracies of predictions were generally low when the GBV of non-phenotyped cows were estimated using information of their relatives in the fourth environment, and the lowest accuracy (0.230) obtained from LLD_300QTL_15KSNP scenario. When animals located in the second environment or in the third environment were used as the validation set, the genomic accuracies of non-phenotyped cows generally were higher even when environments 1 and 4 were used in training set. These results are in line with accuracies of genomic predictions for the fourth environment when the third environment was used in training set. Generally, the accuracy of genomic prediction decreased when extreme environments were used as validation set. Among all scenarios, higher accuracies were achieved in scenarios with HLD and the 50KSNP chips.

Table 4. Accuracy and standard deviation (in parenthesis) of genomic predictions for animals in 4 environments via four-trait animal model. Scenarios include high or low linkage disequilibrium (HLD or LLD), different QTL number (300QTL or 600QTL) and different SNP density (15KSNP or 50KSNP)

Scenario	Environment (Heritability)			
	1 (0.20)	2 (0.25)	3 (0.30)	4 (0.35)
HLD_300QTL_15KSNP	0.551 ab C (0.047)	0.599 ab B (0.017)	0.639 a A (0.029)	0.644 b A (0.023)
HLD_300QTL_50KSNP	0.584 a C (0.046)	0.611 a C (0.038)	0.650 a B (0.015)	0.690 a A (0.021)
HLD_600QTL_15KSNP	0.540 bc C (0.038)	0.564 bc BC (0.056)	0.605 bc AB (0.043)	0.631 b A (0.043)
HLD_600QTL_50KSNP	0.553 ab C (0.051)	0.593 ab BC (0.041)	0.627 ab B (0.037)	0.695 a A (0.058)
LLD_300QTL_15KSNP	0.509 cd B (0.022)	0.563 bc A (0.047)	0.579 cd A (0.035)	0.582 d A (0.025)
LLD_300QTL_50KSNP	0.536 bc C (0.029)	0.578 abc B (0.038)	0.604 bc A (0.017)	0.625 bc A (0.019)
LLD_600QTL_15KSNP	0.495 d C (0.044)	0.536 c BC (0.051)	0.560 d AB (0.052)	0.595 cd A (0.039)
LLD_600QTL_50KSNP	0.537 bc C (0.031)	0.577 abc B (0.038)	0.596 bc AB (0.037)	0.623 bc A (0.034)

Means without common letters (lowercase and uppercase letters stand for comparison within column and row means, respectively) are statistically different ($P<0.05$).

Table 5. Accuracy and standard deviation (in parenthesis) of genomic predictions for non-phenotyped cows in an environment, validation set (VS), using the information of their relatives in the other environments, training set (TS), via three-trait animal model. Scenarios include high or low linkage disequilibrium (HLD or LLD), different QTL number (300QTL or 600QTL) and different SNP density (15KSNP or 50KSNP)

Scenario	VS	1			2			3			4		
	TS	2	3	4	1	3	4	1	2	4	1	2	3
HLD_300QTL_ 15KSNP		0.383abA (0.076)	0.366aA (0.068)	0.263abB (0.059)	0.385ab,A (0.098)	0.403ab,A (0.068)	0.367a,A (0.077)	0.321abc,B (0.055)	0.393abA (0.034)	0.388abcA (0.078)	0.257bC (0.018)	0.317abB (0.073)	0.380abcA (0.059)
HLD_300QTL_ 50KSNP		0.425aA (0.068)	0.396aA (0.050)	0.315aB (0.066)	0.393aA (0.055)	0.404abA (0.093)	0.387aA (0.061)	0.383aA (0.075)	0.393abA (0.080)	0.422abA (0.061)	0.343aB (0.069)	0.389aAB (0.065)	0.430abA (0.031)
HLD_600QTL_ 15KSNP		0.367abcA (0.120)	0.335abA (0.105)	0.307abA (0.082)	0.354abA (0.048)	0.403abA (0.083)	0.361aA (0.062)	0.292bcB (0.108)	0.334abAB (0.087)	0.417abcA (0.082)	0.264bB (0.127)	0.324abAB (0.054)	0.398abcA (0.092)
HLD_600QTL_ 50KSNP		0.393abA (0.088)	0.386aAB (0.076)	0.318aB (0.062)	0.392aA (0.094)	0.428aA (0.060)	0.384aA (0.069)	0.371abB (0.058)	0.412aAB (0.051)	0.446aA (0.049)	0.280abB (0.065)	0.338abB (0.068)	0.432aA (0.067)
LLD_300QTL_ 15KSNP		0.336abcA (0.157)	0.286bA (0.131)	0.230bA (0.153)	0.301bA (0.111)	0.338bA (0.105)	0.314abA (0.101)	0.244cA (0.131)	0.329bA (0.126)	0.348bcA (0.111)	0.233bB (0.117)	0.315bAB (0.100)	0.355cA (0.098)
LLD_300QTL_ 50KSNP		0.414abA (0.072)	0.379aA (0.050)	0.294abB (0.065)	0.351abA (0.051)	0.363abA (0.086)	0.308abA (0.091)	0.279cC (0.049)	0.343abB (0.047)	0.415abcA (0.057)	0.279abB (0.083)	0.322abAB (0.098)	0.390abcA (0.050)
LLD_600QTL_ 15KSNP		0.279cA (0.046)	0.279bA (0.065)	0.235abA (0.056)	0.311abA (0.087)	0.341bA (0.051)	0.286bA (0.081)	0.262cB (0.067)	0.315bAB (0.082)	0.367bcA (0.036)	0.251bB (0.050)	0.289bB (0.046)	0.374abcA (0.038)
LLD_600QTL_ 50KSNP		0.325bcA (0.070)	0.288bAB (0.041)	0.250abB (0.060)	0.349abA (0.088)	0.354abA (0.045)	0.345abA (0.075)	0.285cA (0.114)	0.335abA (0.076)	0.344cA (0.082)	0.248bC (0.054)	0.291bB (0.037)	0.365bcA (0.045)

Means without common letters (lowercase and uppercase letters stand for comparison within column and row means, respectively) are statistically different ($P<0.05$).

Average accuracies of genomic predictions for cows with 25, 50, and 75% of cows with phenotypic records in the fourth environment are given in Table 6. For all scenarios, accuracies of GBVs increased with increasing percentage of cows with phenotypes. The lowest accuracy of 0.315 was obtained for LLD_600QTL_15KSNP scenario, when 25% of phenotyped cows were available in the fourth environment, and the accuracy of prediction was the highest (0.604) for HLD_300QTL_50KSNP scenario, when 75% of cows had phenotypic records. Although, in some cases, the scenarios with 300 QTL had higher accuracies than scenarios with 600 QTL, no significant difference was observed between scenarios combined with 300 and 600 QTL.

Table 6. Accuracy and standard deviation (in parenthesis) of genomic predictions for non-phenotyped cows in the 4th environment with 25, 50, and 75% of phenotyped cows in the 4th environment via four-trait animal model. Scenarios include high or low linkage disequilibrium (HLD or LLD), different QTL number (300QTL or 600QTL) and different SNP density (15KSNP or 50KSNP)

Scenario	Phenotyped cows in the fourth environment (%)		
	25	50	75
HLD_300QTL_15KSNP	0.453 abC (0.044)	0.519 aB (0.054)	0.588 aA (0.031)
HLD_300QTL_50KSNP	0.463 aC (0.022)	0.568 aB (0.032)	0.604 aA (0.033)
HLD_600QTL_15KSNP	0.442 abcC (0.057)	0.520 abB (0.051)	0.587 aA (0.062)
HLD_600QTL_50KSNP	0.463 abB (0.063)	0.545 abA (0.067)	0.594 aA (0.028)
LLD_300QTL_15KSNP	0.367 edC (0.041)	0.493 bB (0.043)	0.538 abA (0.032)
LLD_300QTL_50KSNP	0.390 cdC (0.028)	0.511 abB (0.029)	0.570 aA (0.029)
LLD_600QTL_15KSNP	0.315 eB (0.121)	0.402 cAB (0.116)	0.451 cA (0.130)
LLD_600QTL_50KSNP	0.398 bcdB (0.037)	0.515 abA (0.042)	0.504 bcA (0.100)

Means without common letters (lowercase and uppercase letters stand for comparison within column and row means, respectively) are statistically different ($P<0.05$).

Discussion

The present study examined accuracies of genomic prediction across various scenarios via single-trait (to predict overall performance in the 4 environments) and multiple-trait (to predict GBV considering genotype by environment interaction) models. The LD levels and the SNP density that were used in different scenarios affected accuracies of genomic predictions. The trend of exponential decay of LD with the increase in physical distance was in agreement with other published studies (Jiménez-Montero et al., 2013; Yin et al., 2014). The HLD scenarios assumed a relatively small population size of 200 individuals in the historical population for 10 generations to create a bottleneck, which generates higher LD than the LLD scenarios without bottleneck (Figure 1). Because LD depends on the genetic structure of the population it can be arisen due to small effective population size (Brito et al., 2011). As previously reported in dairy cattle (e.g., Hayes et al., 2009; Brito et al., 2011), the extent of LD has a major impact on the accuracy of genomic prediction. Accuracies of prediction for high and low LD levels were consistent with reports from other studies (e.g. Pimentel et al., 2013; Yin et al., 2014). Therefore, the greater average accuracies were observed for scenarios with HLD than scenarios with LLD.

Increasing the number of markers without increasing the number of phenotypes can be counter-productive as co-linearity issues and it can reduce accuracy (Muir, 2007). However, Solberg et al. (2008) reported that the accuracies of prediction increased when the marker density was doubled. Yin et al. (2014) also stated that higher marker density would increase the level of linkage disequilibrium between markers and QTLs, consequently, more markers could capture a higher proportion of genetic variance of the trait of interest (Goddard, 2009; Muir, 2007). Therefore, average accuracies of scenarios with 50KSNP chip were higher than average accuracies of scenarios with 15KSNP chip.

Daetwyler et al. (2010) investigated the impact of genomic architecture on the accuracy of prediction using different methods and found that the GBLUP indicated relatively constant accuracies for traits with different numbers of QTL. But using Bayes B methodology, the accuracy was the highest with small number of QTL and decreased as number of QTL increased (Habier et al., 2009; Clark et al., 2011). Clark et al. (2011) demonstrated that in GBLUP, no significant difference was found when comparing different QTL numbers, unless number of QTL was very small. For example, accuracies of prediction were 0.37 and 0.38 for scenarios with 100 and 1000 QTLs, respectively (Clark et al., 2011). In agreement with the previous studies (Daetwyler et al., 2010; Clark et al., 2011; Wientjes et al., 2015), based on REML methodology, the constant accuracies of genomic prediction were observed for scenarios with 300 and 600 QTLs.

Recent studies (Daetwyler et al., 2008; Goddard, 2009; Hayes et al., 2009) have shown that the heritability of the trait of interest, as a factor underlying the genetic architecture of a trait, also can influence the accuracy of genomic prediction. Moser et al. (2012) reported that there was a strong relationship between the accuracy of genomic prediction and the heritability of the trait. For production traits with usually moderate heritabilities, the accuracy of prediction was better than that for traits with lower heritabilities, e.g. health traits. Calus et al. (2013) mentioned that the accuracies were 0.41 and 0.67 for traits with heritabilities of 0.05 and 0.30, respectively. In this study, using all phenotyped animals in the training set, among all scenarios, accuracies for the third and the fourth environments were higher compared with the results for the other environments, because heritabilities were increased from the first environment to the fourth environment (Table 4).

Higher accuracies in Table 4 than Table 3 indicated that multiple-trait models are more accurate than single-trait models. Multiple-trait animal models for genomic predictions have been reported recently (Calus and Veerkamp, 2011; Hayashi and Iwata, 2013; Guo et al., 2014), because more information can be considered simultaneously. As with the traditional genetic evaluation process, a multiple-trait model can increase the accuracy of genomic prediction using information from genetically correlated traits (Jia and Jannink, 2012; Guo et al., 2014; Jiang et al., 2015). In another study, Yin et al. (2014) reported that when using continuous environmental descriptor, correlated traits can be considered as longitudinal trait and genomic random regression model can be applied to predict the GBV. Hayashi and Iwata (2013) reported that, for a low heritable trait ($h^2 = 0.1$), accuracy of GBV estimated from multi-trait model was about 20% higher than accuracy from single-trait model.

With regard to benefits of multi-trait model over single-trait model (Calus and Veerkamp, 2011; Jia and Jannink, 2012; Guo et al., 2014), multiple-trait animal model can be useful to detect G × E interaction for milk production traits (Lillehammer et al., 2007), and use of information from phenotypes in different environments as different correlated traits can be the main advantage of the multiple-trait animal model to improve the accuracy of genomic prediction (Hayes et al., 2016).

The results (Table 5) indicated that accuracy of genomic prediction was lowest when the environments 1 and 4 were used in the training and validation sets respectively, due to the fact that there was less genetic correlation (0.50) between the environments 1 and 4 (Table 2). There was a high genetic correlation between the second and the third environments and the accuracies for non-phenotyped cows were higher, when using animals located in either of them as training set. Especially for HLD_600QTL_50KSNP scenario, average accuracies were 0.428 and 0.412 for the second and the third environments, respectively, when those environments were used in validation set. The highest average accuracy of 0.446 was obtained for non-phenotyped cows in the environment 3 when environment 4 was used in training set. Given genetic correlations between traits in environments 2 with 3 (0.82) and in environments 3 with 4 (0.81), a possible reason for the highest average accuracy (0.446) using environment 4 could be that the heritability of the environment 4 was the highest among the 4 environments.

As reported by de Roos et al. (2009) and Hayes et al. (2009), the accuracy of genomic prediction was determined by the number of phenotypic records used to estimate the GBV. Four-trait animal model for the trait with 25% of records in the fourth environment (Table 6) proved better than the three-trait animal model without any record in the fourth environment (Table 5). Because, cows in the validation set were genetically related to their half-sibs in the same and also in the other environments, and these close relationships might improve the accuracy of genomic prediction (Yin et al., 2014; Guo et al., 2014). Therefore, in practical breeding programs, accuracy of genomic selection can be improved even though only a small percentage of animals in the extreme environment have phenotypic records. The accuracy increased with increasing the phenotyped cows in the extreme environment. With increasing records, there will be more observations per SNP allele and then, the accuracy will be higher (Hayes et al., 2009). Using the 50% instead of 25% of phenotyped animals in training set, the increase in accuracy of genomic prediction was relatively higher for the LLD scenario than accuracy for HLD scenario (Table 6). For example, the increase in accuracies for HLD_300QTL_15KSNP and LLD_300QTL_15KSNP scenarios were 0.066 and 0.126, respectively. Therefore, to achieve a desirable accuracy in population with low LD, we should have relatively more phenotypes in the considered environment. Also Hozé et al. (2014) reported that it requires a large reference population per breed to increase accuracy of genomic prediction.

Genetic correlation between traits has previously been applied to improve accuracy of multi-trait (Calus and Veerkamp, 2011; Aguilar et al., 2011) and multi-breed (Olson et al., 2012; Karoui et al., 2012) genomic predictions. Olson et al. (2012) demonstrated that using a multi-trait model with SNP effects in different breeds

treated as correlated effects can increase accuracy of prediction. Karoui et al. (2012) found a slight increase in accuracies for a trait with high genetic correlation across breeds; however, no increase was found for fertility, because genetic correlations between breeds were low for this trait. Wientjes et al. (2015) mentioned that the accuracy of multi-breed genomic prediction might be influenced by factors such as the family relationships and level of LD. For reference population including animals from different environments or countries, the multi-trait genomic model is useful to investigate G × E interaction and to increase accuracy of genomic predictions (Hayes et al., 2016). For milk production trait of dairy cows, the genetic correlation of the same trait measured in Australia with most of the northern hemisphere countries is about 0.8 or less (Haile-Mariam et al., 2015). As the G × E interaction between Australia and other countries is considerable, Haile-Mariam et al. (2015) demonstrated that it will be better to include phenotypic and genotypic records of progenies of sires from other countries in the reference population to reduce biases of prediction via multi-trait analysis.

The simulated QTL effect was zero in an environment but it has an effect in the others. In the other words, G × E interaction can be caused either by alleles being expressed only in specific environment (Lillehammer et al., 2007) and when the QTL had an interaction with the environment, the model with the G × E interaction effect gave a higher power of QTL detection than model without interaction effect. Lillehammer et al. (2009) reported some QTLs that have not been reported earlier. They mentioned that gene by environment interaction is one possible cause of inconsistency between QTL mapping studies, such that QTL detected in one environment may not be detected in another, or if detected, may have different effects. Therefore, models that included G × E interaction can detect different gene expression and genetic variations across environments (Bastiaansen et al., 2014).

In total, our results without considering G × E interaction, like a population, agree with previous results from simulation studies, and the accuracy of genomic prediction depends on the level of LD, the number of markers and the heritability of trait of interest, and also no significant difference was observed between scenarios with 300 and 600 QTL. When including G × E interaction in the model, level of LD, number of animals in training set and the genetic correlation between the traits play the important roles to achieve a desirable accuracy of genomic prediction.

Conclusion

The multiple-trait genomic model, which simultaneously considers the same trait measured in different environments as different correlated traits, can improve the accuracy of genomic predictions. Therefore, G × E interaction should be taken into account to estimate GBVs more accurately and to select better genotypes. Applying a few phenotypic records in an environment gave more accuracy of prediction compared with an environment without any phenotypic records, and with regard to genetic correlations across environments, for LLD scenarios, it is necessary to have more phenotypic records to achieve a desirable accuracy. Furthermore, the G × E interaction analysis would contribute to understanding variation of quantitative trait,

and further studies are required to compare capability of different methods to detect gene expression across environments and find out the biological background behind the interaction.

Acknowledgement

The authors gratefully acknowledge funding from the German Federal Ministry of Education and Research (BMBF) and from the Förderverein Bioökonomieforschung e.V. (FBV) / German Holstein Association (DHV) for the collaborative project "KMU-innovativ-10: Kuh-L–cow calibration groups for the implementation of selection strategies based on high-density genotyping in dairy cattle", grant no. 031A416C.

References

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., Lawlor T.J. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci., 93: 743–752.

Aguilar I., Misztal I., Legarra A., Tsuruta S. (2011). Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. J. Anim. Breed. Genet., 128: 422–428.

Bastiaansen J.W.M., Bovenhuis H., Lopes M.S., Silva F.F., Megens H.J., Calus M.P.L. (2014). SNP effects depend on genetic and environmental context. Proc. 10th World Congress on Genetics Applied to Livestock Production, 17–22.08.2014, Vancouver, Canada.

Bohlouli M., Shodja J., Alijani S., Eghbal A. (2013). The relationship between temperature-humidity index and test-day milk yield of Iranian Holstein dairy cattle using random regression model. Livest. Sci., 157: 414–420.

Bohlouli M., Shodja J., Alijani S., Pirany N. (2014). Interaction between genotype and geographical region for milk production traits of Iranian Holstein dairy cattle. Livest. Sci., 169: 1–9.

Brito F.V., Neto J.B., Sargolzaei M., Cobuci J.A., Schenkel F.S. (2011). Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. BMC Genetics, 12: 80.

Brügemann K., Gernand E., von Borstel U.U., König S. (2011). Genetic analyses of protein yield in dairy cows applying random regression models with time-dependent and temperature × humidity-dependent covariates. J. Dairy Sci., 94: 4129–4139.

Calus M.P.L., Veerkamp R.F. (2011). Accuracy of multi-trait genomic selection using different methods. Genet. Sel. Evol., 43: 1–14.

Calus M.P.L., Groen A.F., De Jong G. (2002). Genotype by environment interaction for protein yield in Dutch dairy cattle as quantified by different models. J. Dairy Sci., 85: 3115–3123.

Calus M.P.L., de Haas Y., Pszczola M., Veerkamp R.F. (2013). Predicted accuracy of and response to genomic selection for new traits in dairy cattle. Animal, 7: 183–191.

Ceron-Munoz M., Tonhati F.H., Costa C.N., Rojas-Sarmiento D., Echeverri D.M. (2004). Factors that cause genotype by environment interaction and use of a multiple-trait herd-cluster model for milk yield of Holstein cattle from Brazil and Colombia. J. Dairy Sci., 87: 2687–2692.

Clark S.A., Hickey J.M., van der Werf J.H.J. (2011). Different models of genetic variation and their effect on genomic evaluation. Genet. Sel. Evol., 43: 18.

Daetwyler H.D., Villanueva B., Woolliams J.A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS ONE, 3: e3395.

Daetwyler H.D., Pong-Wong R., Villanueva B., Woolliams J.A. (2010). The impact of genetic architecture on genome-wide evaluation methods. Genetics, 185: 1021–1031.

De Roos A.P.W., Hayes B.J., Goddard M.E. (2009). Reliability of genomic predictions across multiple populations. Genetics, 183: 1545–1553.
Dekkers J.C.M. (2007). Prediction of response to marker-assisted and genomic selection using selection index theory. J. Anim. Breed. Genet., 124: 331–341.
Falconer D.S., MacKay T.F.C. (1996). Introduction to Quantitative Genetics, 4th ed., Longman Group, Essex, UK.
Goddard M. (2009). Genomic selection: Prediction of accuracy and maximisation of long term response. Genetica, 136: 245–257.
Guo G., Zhao F., Wang Y., Zhang Y., Du L., Su G. (2014). Comparison of single-trait and multiple-trait genomic prediction models. Genetics, 15: 30.
Habier D., Fernando R.L., Dekkers J.C.M. (2009). Genomic selection using low-density marker panels. Genetics, 182: 343–353.
Haile-Mariam M., Pryce J.E., Schrooten C., Hayes B.J. (2015). Including overseas performance information in genomic evaluations of Australian dairy cattle. J. Dairy Sci., 98: 1–17.
Hammami H., Rekik B., Bastin C., Soyeurt H., Bormann J., Stoll J., Gengler N. (2009). Environmental sensitivity for milk yield in Luxembourg and Tunisian Holsteins by herd management level. J. Dairy Sci., 92: 4604–4612.
Hayashi T., Iwata H. (2013). A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. BMC Bioinformatics, 14: 1–14.
Hayes B.J., Bowman P.J., Chamberlain A.J., Goddard M.E. (2009). Invited review: Genomic selection in dairy cattle: progress and challenges. J. Dairy Sci., 92: 433–443.
Hayes B.J., Daetwyler H.D., Goddard M.E. (2016). Models for genome × environment interaction: examples in livestock. Crop Sci., 56: 2251–2259.
Hickey J.M., Gorjanc G. (2012). Simulated data for genomic selection and genome-wide association studies using a combination of coalescent and gene drop methods. G3, 2: 425–427.
Hill W.G., Robertson A. (1968). Linkage disequilibrium in finite populations. Theor. Appl. Genet., 6: 226–231.
Hozé C., Fritz S., Phocas F., Boichard D., Ducrocq V., Croiseau P. (2014). Efficiency of multi-breed genomic selection for dairy cattle breeds with different sizes of reference population. J. Dairy Sci., 97: 3918–3929.
Jia Y., Jannink J.L. (2012). Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics, 192: 1513–1522.
Jiang J., Zhang Q., Ma L., Li J., Wang Z., Liu J.F. (2015). Joint prediction of multiple quantitative traits using a Bayesian multivariate antedependence model. Heredity, 115: 29–36.
Jiménez-Montero J.A., González-Recio O., Alenda R. (2013). Comparison of methods for the implementation of genome-assisted evaluation of Spanish dairy cattle. J. Dairy Sci., 96: 625–634.
Karoui S., Carabaño M.J., Díaz C., Legarra A. (2012). Joint genomic evaluation of French dairy cattle breeds using multiple-trait models. Genet. Sel. Evol., 44: 39.
Kolmodin R., Strandberg E., Madsen P., Jensen J., Jorjani H. (2002). Genotype by environment interaction in Nordic dairy cattle studied by use of reaction norms. Acta Agric. Scand. A Anim. Sci., 52: 11–24.
König S., Simianer H., Willam A. (2009). Economic evaluation of genomic breeding programs. J. Dairy Sci., 92: 382–391.
Lillehammer M., Ødegard J., Meuwissen T.H.E. (2007). Random regression models for detection of gene by environment interaction. Genet. Sel. Evol., 39: 105–121.
Lillehammer M., Goddard M.E., Nilsen H., Sehested E., Olsen H.G., Lien S., Meuwissen T.H.E. (2008). Quantitative trait locus-by-environment interaction for milk yield traits on *Bos taurus* autosome 6. Genetics, 179: 1539–1546.
Lillehammer M., Hayes B.J., Meuwissen T.H.E., Goddard M.E. (2009). Gene by environment interactions for production traits in Australian dairy cattle. J. Dairy Sci., 92: 4008–4017.
Lund M.S., Su G., Janss L., Guldbrandtsen B., Brøndum R.F. (2014). Genomic evaluation of cattle in a multi-breed context. Livest. Sci., 166: 101–110.
Meuwissen T.H.E., Hayes B., Goddard M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157: 1819–1829.

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., Lee D.H. (2002). BLUPF90 and related programs. Communication no. 28-07. Proc. 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France.

Moser G., Khatkar M.S., Hayes B.J., Raadsma H.W. (2012). Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. Genet. Sel. Evol., 42: 37.

Muir W.M. (2007). Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J. Anim. Breed. Genet., 124: 342–355.

Nejati-Javaremi A., Smith C., Gibson J. (1997). Effect of total allelic relationship on accuracy of evaluation and response to selection. J. Anim. Sci., 75: 1738–1745.

Olson K.M., VanRaden P.M., Tooker M.E. (2012). Multibreed genomic evaluations using purebred Holsteins, Jerseys, and Brown Swiss. J. Dairy Sci., 95: 5378–5383.

Pimentel E.C.G., Wensch-Dorendorf M., König S., Swalve H.H. (2013). Enlarging a training set for genomic selection by imputation of un-genotyped animals in populations of varying genetic architecture. Genet. Sel. Evol., 45: 12.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W., Daly M.J., Sham P.C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet., 81: 559–575.

Robertson A. (1959). The sampling variance of the genetic correlation coefficient. Biometrics, 15: 469–485.

Sargolzaei M., Schenkel F.S. (2009). QMSim: a large-scale genome simulator for livestock. Bioinformatics, 25: 680–681.

Solberg T.R., Sonesson A.K., Woolliams J.A., Meuwissen T.H.E. (2008). Genomic selection using different marker types and densities. J. Anim. Sci., 86: 2447–2454.

VanRaden P.M. (2008). Efficient methods to compute genomic predictions. J. Dairy Sci., 91: 4414–4423.

Wientjes Y.C.J., Calus M.P.L., Goddard M.E., Hayes B.J. (2015). Impact of QTL properties on the accuracy of multi-breed genomic prediction. Genet. Sel. Evol., 47: 42.

Yin T., Pimentel E.C.G., König U., Borstel V., König S. (2014). Strategy for the simulation and analysis of longitudinal phenotypic and genomic data in the context of a temperature × humidity-dependent covariate. J. Dairy Sci., 97: 2444–2454.

Received: 22 IX 2016
Accepted: 20 I 2017