

SHORT COMMUNICATION

Open Access

# Improving the accuracy of genomic prediction in Chinese Holstein cattle by using one-step blending

Xiujin Li<sup>†</sup>, Sheng Wang<sup>†</sup>, Ju Huang, Leyi Li, Qin Zhang and Xiangdong Ding<sup>\*</sup>

## Abstract

**Background:** The one-step blending approach has been suggested for genomic prediction in dairy cattle. The core of this approach is to incorporate pedigree and phenotypic information of non-genotyped animals. The objective of this study was to investigate the improvement of the accuracy of genomic prediction using the one-step blending method in Chinese Holstein cattle.

**Findings:** Three methods, GBLUP (genomic best linear unbiased prediction), original one-step blending with a genomic relationship matrix, and adjusted one-step blending with an adjusted genomic relationship matrix, were compared with respect to the accuracy of genomic prediction for five milk production traits in Chinese Holstein. For the two one-step blending methods, de-regressed proofs of 17 509 non-genotyped cows, including 424 dams and 17 085 half-sisters of the validation cows, were incorporated in the prediction model. The results showed that, averaged over the five milk production traits, the one-step blending increased the accuracy of genomic prediction by about 0.12 compared to GBLUP. No further improvement in accuracies was obtained from the adjusted one-step blending over the original one-step blending in our situation. Improvements in accuracies obtained with both one-step blending methods were almost completely contributed by the non-genotyped dams.

**Conclusions:** Compared with GBLUP, the one-step blending approach can significantly improve the accuracy of genomic prediction for milk production traits in Chinese Holstein cattle. Thus, the one-step blending is a promising approach for practical genomic selection in Chinese Holstein cattle, where the reference population mainly consists of cows.

## Background

A reference population with sufficient size is essential in genomic selection (GS) [1-3]. For dairy cattle, in almost all countries with developed dairy industry, thousands of progeny-tested bulls with highly reliable estimated breeding value (EBV) are used to form the national reference population. However, constituting such a reference population is not feasible in some countries, e.g. China, where the number of bulls with highly reliable EBV is limited. As an alternative, cows can be used to form the reference population. Ding et al. [4] investigated the accuracy of genomic prediction using a reference population consisting

of cows, and showed that genomic selection using cows is feasible. However, a larger population of reference cows was required to obtain comparable accuracies of genomic prediction than when progeny-tested bulls are used as reference population, because cow EBV are generally less reliable than bull EBV [4]. Further efforts are needed to improve the accuracy of genomic prediction in such a situation.

The term “one-step blending” was used to distinguish it from the original single-step approach using DPR (de-regressed proofs) instead of raw phenotypes [5]. In the present study, we investigated the possible improvements in the accuracy of genomic prediction by applying the one-step blending approach to Chinese Holsteins, for which the reference population consists primarily of cows. In addition, the influence of the relationship between the non-genotyped animals and genotyped selection candidates

\* Correspondence: xding@cau.edu.cn

<sup>†</sup>Equal contributors

Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture of China, National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

on the prediction accuracy of one-step blending was also investigated.

## Methods

### Data

The data consisted of 4917 Chinese Holstein cows born from 1998 to 2009 and 240 progeny-tested bulls born from 1984 to 2005, all of which had official EBV on five milk production traits (milk yield, fat yield, fat percentage, protein yield, and protein percentage). These official EBV were obtained based on a multiple-trait random regression test-day model [6]. DRP of all animals were derived from their EBV according to VanRaden and Wiggans [7] and used as response variables for genomic prediction. Reliabilities of the DRP were calculated according to Liu et al. [8]. All animals had reliabilities of DRP greater than 0.40 (for cows) or 0.80 (for bulls). Out of the 4917 cows, 4106 born before 2008, together with the 240 bulls, were taken as the reference population, and the remaining 811 cows born in or after 2008 were used as the validation population.

All individuals in the reference and validation populations were genotyped with the Illumina BovineSNP50 BeadChip (Illumina, San Diego, CA). Missing genotypes of single nucleotide polymorphisms (SNPs) with known chromosomal positions were imputed by BEAGLE [9], and those with unknown chromosomal positions were discarded. After imputation, SNPs with minor allele frequency (MAF) less than 0.01 were removed, leaving 46 422 SNPs for genomic prediction.

To implement the one-step blending approach, all non-genotyped dams and half-sisters of the validation cows that had DRP with reliabilities greater than 0.40, were considered. Of the 811 validation cows, 425 had non-genotyped dams (424 in total) and all had non-genotyped half-sisters (17 085 in total, ranging from 154 to 2672).

Blood samples were collected from Chinese Holstein cattle when the regular quarantine inspection of the farms was conducted. The procedure for collecting the blood samples was carried out in strict accordance with the protocol approved by the Animal Welfare Committee of China Agricultural University (Permit Number: DK996).

### Statistical models

Three methods, GBLUP, the original one-step blending and the adjusted one-step blending, were implemented for genomic prediction of animals in the validation population.

#### GBLUP

The following genomic BLUP model [10] was used to predict genomic breeding values:

$$\mathbf{y} = 1\mu + \mathbf{Zg} + \mathbf{e},$$

where  $\mathbf{y}$  is the vector of DRP of the reference animals,  $\mathbf{g}$  is the vector of additive genetic effects, which assumed to follows a normal distribution  $N(0, \mathbf{G}\sigma_g^2)$ , with  $\mathbf{G}$  being the genomic relationship matrix constructed using the first method of VanRaden [10], and  $\mathbf{e}$  is the vector of random errors, assumed to follow a normal distribution  $N(0, \mathbf{D}\sigma_e^2)$ , with  $\mathbf{D}$  being a diagonal matrix with  $d_{ii} = 1/w_i$ , where  $w_i = r_{DRP_i}^2 / (1 - r_{DRP_i}^2)$  ( $r_{DRP_i}^2$  is the reliability of DRP of individual  $i$ ) [10,11]. The estimates in  $\mathbf{g}$  based on this model are termed direct genomic breeding values (DGV).

#### Original one-step blending

Following Legarra et al. [12], Aguilar et al. [13], and Christensen and Lund [14], the one-step blending method has the same model as GBLUP, except that the vector  $\mathbf{y}$  also contains the DRP of non-genotyped animals and vector  $\mathbf{g}$  is assumed to follow a normal distribution  $N(0, \mathbf{H}\sigma_g^2)$ , where  $\mathbf{H}$  is defined as:

$$\mathbf{H} = \begin{bmatrix} \mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}'_{12} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G} \\ \mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{12} & \mathbf{G} \end{bmatrix},$$

with  $\mathbf{A}_{11}$ ,  $\mathbf{A}_{12}$ , and  $\mathbf{A}_{22}$  sub-matrices of  $\mathbf{A}$  (the pedigree-based relationship matrix), and subscripts 1 and 2 refer to non-genotyped and genotyped animals, respectively. The estimates in  $\mathbf{g}$  based on this model are termed the genomic enhanced breeding values (GEBV).

#### Adjusted one-step blending

To avoid the potential incompatibility in scale between the coefficients of  $\mathbf{G}$  and  $\mathbf{A}_{22}$  involved in the  $\mathbf{H}$  matrix, which could lead to incorrect weighting of the pedigree and genomic information, as pointed out by Forni et al. [15], the  $\mathbf{G}$  matrix was adjusted following Gao et al. [16], i.e.,

$$\mathbf{G}_a = \mathbf{G}\beta + \alpha,$$

where  $\beta$  and  $\alpha$  are obtained from the following equations:

$$\begin{aligned} \text{Avg}(\text{diag}(\mathbf{G}))\beta + \alpha &= \text{Avg}(\text{diag}(\mathbf{A}_{22})), \\ \text{Avg}(\text{offdiag}(\mathbf{G}))\beta + \alpha &= \text{Avg}(\text{offdiag}(\mathbf{A}_{22})), \end{aligned}$$

Where  $\text{Avg}(\text{diag}(*))$  means the average value of diagonal elements of matrix  $*$ ;  $\text{Avg}(\text{offdiag}(*))$  means the average value of non-diagonal elements of matrix  $*$ .

The variance components ( $\sigma_g^2$  and  $\sigma_e^2$ ) involved in the three models were estimated using AI-REML, as implemented in the software DMU [17].

### Evaluation of the accuracy of genomic prediction

The accuracy of genomic predictions was evaluated as  $r_v = \frac{r_{\hat{g},DRP}}{r_{DRP}}$  [5], where  $r_{\hat{g},DRP}$  is the correlation between the estimated  $\mathbf{g}$  (DGV or GEBV) and the DRP in the validation population and  $r_{DRP}$  is the average of the square root of the reliability of the DRP of the validation cows.

In addition, the theoretical accuracy of the DGV or GEBV was calculated for each individual in the same way as in conventional BLUP, following Henderson [18] from the diagonal of the inverse of the mixed model equation (MME), and the average theoretical accuracy over validation animals was also used to evaluate the accuracy of genomic predictions.

### Results and discussion

As shown in Table 1, for the 811 validation cows,  $r_v$  and average theoretical accuracies from the original one-step blending increased by 0.12 and 0.02, respectively, compared with the accuracies from GBLUP averaged over the five traits. Accuracies from the adjusted one-step blending approach were almost the same as those from the original one-step blending. Theoretical accuracies

were much higher than  $r_v$ , which was also observed in other studies [3,19-21]. The theoretical accuracy may also be overestimated owing to sampling errors in elements of the genomic relationship matrix as pointed out by Goddard et al. [22]. In comparison with GBLUP, the one-step blending approach can significantly improve the accuracy of genomic prediction by incorporating the phenotypes (DRP) of non-genotyped relatives of the selection candidates. However, the adjusted one-step blending did not result in further improvements in accuracy compared with the original one-step blending, probably because the original  $\mathbf{G}$  matrix was little adjusted in our situation, since the estimates of  $\beta$  and  $\alpha$  were 0.992 (close to 1) and 0.017 (close to 0), respectively, while they were 0.859 and 0.298 in the study of Christensen et al. [23]. Similar results were also observed by Gao et al. [16] in the Nordic Holstein population, where the adjusted one-step blending resulted in little improvement in the prediction accuracy and estimates of  $\beta$  and  $\alpha$  were 0.976 and 0.085, respectively.

Among the 811 validation cows, 425 had both non-genotyped dams and half-sisters, while 386 with genotyped dams had only non-genotyped half-sisters. For

**Table 1 Accuracies of genomic prediction for the validation cows**

Trait	GBLUP		Original one-step		Adjusted one-step	
	$r_v^*$	Theoretical accuracy	$r_v^*$	Theoretical accuracy	$r_v^*$	Theoretical accuracy
<b>All validation cows</b>						
Milk yield	0.36	0.73	0.46	0.75	0.46	0.75
Fat yield	0.44	0.73	0.60	0.75	0.60	0.75
Fat %	0.52	0.67	0.60	0.69	0.60	0.69
Protein yield	0.37	0.72	0.51	0.74	0.51	0.75
Protein %	0.51	0.67	0.62	0.69	0.62	0.69
Average	0.44	0.70	0.56	0.72	0.56	0.73
<b>386 validation cows with genotyped dams</b>						
Milk yield	0.34	0.73	0.36	0.73	0.37	0.74
Fat yield	0.42	0.73	0.44	0.74	0.44	0.74
Fat %	0.50	0.66	0.52	0.67	0.52	0.67
Protein yield	0.36	0.72	0.37	0.73	0.37	0.73
Protein %	0.48	0.67	0.50	0.67	0.50	0.67
Average	0.42	0.70	0.44	0.71	0.44	0.71
<b>425 validation cows with both non-genotyped dams and half-sisters</b>						
Milk yield	0.36	0.73	0.55	0.76	0.55	0.76
Fat yield	0.46	0.73	0.73	0.76	0.72	0.76
Fat %	0.54	0.67	0.69	0.70	0.69	0.70
Protein yield	0.39	0.75	0.62	0.76	0.62	0.76
Protein %	0.54	0.67	0.73	0.70	0.73	0.70
Average	0.46	0.71	0.66	0.74	0.66	0.74

\* $r_v$  is equal to the correlation between the estimated breeding and the DRP divided by the average of the square root of the reliability of the DRP in the validation cows.

**Table 2 Accuracies of genomic prediction for 425 validation cows when their non-genotyped dams and not their non-genotyped half-sisters were used in the one-step blending approach**

Trait	GBLUP		Original one-step	
	$r_v^*$	Theoretical accuracy	$r_v^*$	Theoretical accuracy
Milk yield	0.36	0.73	0.54	0.76
Fat yield	0.46	0.73	0.73	0.76
Fat percentage	0.54	0.67	0.70	0.69
Protein yield	0.39	0.75	0.61	0.76
Protein percentage	0.54	0.67	0.74	0.69
Average	0.46	0.71	0.66	0.73

\* $r_v$  is equal to the correlation between the estimated breeding and the DRP divided by the average of the square root of the reliability of the DRP in the validation cows.

validation cows with genotyped dams,  $r_v$  and the theoretical accuracies obtained from both one-step blending approaches were nearly the same as those from GBLUP (Table 1), while for validation cows with both non-genotyped dams and half-sisters,  $r_v$  were improved by 15 to 26 percentage points and 1 to 3 percentage points for the theoretical accuracy, when using the one-step blending approach (Table 1). Again, in all these cases, the adjusted one-step blending did not perform better than the original one-step blending. These results suggest that, compared with GBLUP, improvements in accuracies from the one-step blending approach were almost completely contributed by the non-genotyped dams. To further prove this, we discarded all non-genotyped half-sisters and only included the non-genotyped dams of 425 validation cows in the one-step blending approach. As expected,  $r_v$  and the theoretical accuracies of the 425 validation cows from the original one-step blending approach (Table 2) were almost the same as those in the scenario when both non-genotyped dams and half-sisters were included in the one-step blending approach (Table 1). The reason for this is that all non-genotyped half-sisters were daughters of 19 genotyped sires in the reference population and the information from these daughters was part of the DRP of the sires. Therefore, these half-sisters contributed little extra information for genomic prediction.

## Conclusions

Averaged over the five milk production traits, both one-step blending methods increased  $r_v$  and the average theoretical accuracy by about 0.12 and 0.02, respectively, compared to GBLUP. However, the adjusted one-step blending did not perform better than the original one-step blending in our situation. In our situation, improvements in accuracies from both one-step blending approaches were almost completely contributed by the non-genotyped dams of the validation animals.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contribution

XJL performed statistical analysis and wrote the manuscript. JH, LYL and SW prepared data. XDD and QZ jointly conceived the design of the study, made substantial contribution to the results interpretation and revised the manuscript. All authors read and approved the manuscript.

## Acknowledgements

This work was supported by the National '948' Project (2011-G2A), Ph.D. Programs Foundation of Ministry of Education of China (20110008110001), National Natural Science Foundation of China (31272418, 31372158), the earmarked fund for CARS-37, National Dairy Industry System in Beijing Team, Scientific Research Foundation for Returned Scholars, Ministry of Education of China, Program for Changjiang Scholar and Innovation Research Team in University (Grant No. IRT1191). We thank the Dairy Association of China for supplying the official EBV and Beijing Dairy Cattle Center for providing blood and semen samples.

Received: 6 April 2014 Accepted: 14 September 2014

Published online: 14 October 2014

## References

- Goddard ME, Hayes BJ: Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat Rev Genet* 2009, **10**:381–391.
- Hayes B, Bowman P, Chamberlain A, Goddard ME: Invited review: genomic selection in dairy cattle: progress and challenges. *J Dairy Sci* 2009, **92**:433–443.
- 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**:16–24.
- Ding X, Zhang Z, Li X, Wang S, Wu X, Sun D, Yu Y, Liu J, Wang Y, Zhang Y, Zhang S, Zhang Y, Zhang Q: Accuracy of genomic prediction for milk production traits in Chinese Holstein. Population using a reference population consisting of cows. *J Dairy Sci* 2013, **96**:5315–5323.
- 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**:909–917.
- Schaeffer LR, Jamrozik J, Kistemaker GJ, Van Doormaal BJ: Experience with a test-day model. *J Dairy Sci* 2000, **83**:1135–1144.
- VanRaden PM, Wiggans GR: Derivation, calculation, and use of national animal model information. *J Dairy Sci* 1991, **74**:2737–2746.
- Liu Z, Reinhardt F, Bünger A, Reents R: Derivation and calculation of approximate reliabilities and daughter yield-deviations of a random regression test-day model for genetic evaluation of dairy cattle. *J Dairy Sci* 2004, **87**:1896–1907.
- Browning BL, Browning SR: A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 2009, **84**:210–223.
- VanRaden PM: Efficient methods to compute genomic predictions. *J Dairy Sci* 2008, **91**:4414–4423.
- Garrick DJ, Taylor JF, Fernando RL: Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol* 2009, **41**:55.
- Legarra A, Aguilar I, Misztal I: A relationship matrix including full pedigree and genomic information. *J Dairy Sci* 2009, **92**:4656–4663.
- Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ: Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J Dairy Sci* 2010, **93**:743–752.
- Christensen OF, Lund MS: Genomic prediction when some animals are not genotyped. *Genet Sel Evol* 2010, **42**:2.
- Forni S, Aguilar I, Misztal I: Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet Sel Evol* 2011, **43**:1.
- Gao H, Christensen OF, Madsen P, Nielsen US, Zhang Y, Lund MS, Su G: Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. *Genet Sel Evol* 2012, **44**:8.

17. Madsen P, Sørensen P, Su G, Damgaard LH, Thomsen H, Labouriau R: *DMU-a Package for Analyzing Multivariate Mixed Models*, In *Proceedings of the 8<sup>th</sup> World Congress on Genetics Applied to Livestock Production: 13-18 August 2006; Belo Horizonte*. 2006.
18. Henderson CR: **Best linear unbiased estimation and prediction under a selection model**. *Biometrics* 1975, **31**:423–447.
19. Bijma P: **Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations**. *J Anim Breed Genet* 2012, **129**:345–358.
20. Edel C, Neuner S, Emmerling R, Goetz K-U: **A note on using 'forward prediction' to assess precision and bias of genomic predictions**. *Interbull Bulletin* 2012, **46**:16–19.
21. Su G, Guldbrandtsen B, Gregersen V, Lund M: **Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population**. *J Dairy Sci* 2010, **93**:1175–1183.
22. Goddard M, Hayes B, Meuwissen T: **Using the genomic relationship to matrix predict the accuracy of genomic selection**. *J Anim Breed Genet* 2011, **128**:409–421.
23. Christensen OF, Madsen P, Nielsen B, Ostersen T, Su G: **Single-step methods for genomic evaluation in pigs**. *Animal* 2012, **6**:1565–1571.

doi:10.1186/s12711-014-0066-4

**Cite this article as:** Li *et al.*: Improving the accuracy of genomic prediction in Chinese Holstein cattle by using one-step blending. *Genetics Selection Evolution* 2014 **46**:66.

**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)

