CrossMark

The impact of genomic relatedness between populations on the genomic estimated breeding values

Peipei Ma[1,2], Ju Huang[1], Weijia Gong[3], Xiujin Li[1], Hongding Gao[4], Qin Zhang[1], Xiangdong Ding[1*] and Chonglong Wang[5*]

Abstract

In genomic selection, prediction accuracy is highly driven by the size of animals in the reference population (RP). Combining related populations from different countries and regions or using a related population with large size of RP has been considered to be viable strategies in cattle breeding. The genetic relationship between related populations is important for improving the genomic predictive ability. In this study, we used 122 French bulls as test individuals. The genomic estimated breeding values (GEBVs) evaluated using French RP, America RP and Chinese RP were compared. The results showed that the GEBVs were in higher concordance using French RP and American RP compared with using Chinese population. The persistence analysis, kinship analysis and the principal component analysis (PCA) were performed for 270 French bulls, 270 American bulls and 270 Chinese bulls to interpret the results. All the analyses illustrated that the genetic relationship between French bulls and American bulls was closer compared with Chinese bulls. Another reason could be the size of RP in China was smaller than the other two RPs. In conclusion, using RP of a related population to predict GEBVs of the animals in a target population is feasible when these two populations have a close genetic relationship and the related population is large.

Keywords: Genomic prediction, Genomic relationship, Joint population prediction

Short communication

Since genomic selection (GS) was first described by Meuwissen et al. [1], with the constantly decreasing genotyping cost, this technology has revolutionized breeding of both livestock and crops in the last few years. The size of reference population (RP) and the relationship between the reference and candidate population were reported to be the important factors affecting accuracy of genomic prediction [2–4].

The advantage of using GS has been limited due to limited size of RP. Firstly, a low number of progeny-test proven bulls were available in each country especially in countries which mainly relied on importing bull semen from the other countries, e.g. China [5]. Secondly, it is not economically feasible to genotype all the animals as RP since the contribution of the cows may be less than the cost for genotyping them [6]. To gain accuracy of GEBV, two strategies were used in practice. One strategy is to combine the reference data from several countries. The other one is to use the RP from a commercial institute e.g. CDCB (https://www.uscdcb.com/what-we-do/genomics). However, it was reported that the relationship between RP and candidate individual was a crucial factor for prediction accuracy in genomic prediction [7, 8]. Therefore, it is necessary to investigate the relationship between populations before applying these strategies.

The objectives of this study were 1) to investigate the correlations between genomic estimated breeding values (GEBVs) for French bulls using Chinese, French and American RP separately; 2) to explore the reasons led to different GEBVs, by analyzing the linkage disequilibrium (LD) phase persistence, genetic

* Correspondence: xding@cau.edu.cn; ahwchl@163.com
[1]National Engineering Laboratory for Animal Breeding, Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China
[5]Key Laboratory of Pig Molecular Quantitative Genetics, Institute of Animal Husbandry and Veterinary Medicine, Anhui Academy of Agricultural Sciences, Hefei 230031, China
Full list of author information is available at the end of the article

Ma et al. Journal of Animal Science and Biotechnology (2018) 9:64

Page 2 of 6

relatedness, and population structure among French, American and Chinese populations.

Materials and methods
Data
A total of 122 French bulls were used as test set in this study. The GEBVs of milk yield, fat percentage, protein percentage, confirmation and feet_legs evaluated using American RP and French RP separately was provided by Gènes DIFFSUION. The GEBVs of these 122 French bulls using Chinese RPs were estimated in this study. The Chinese RP consisted of 1,568 Chinese cows with both genotype and phenotype. De-regressed proof (DRP) was used as the response variable for genomic prediction in this study. Genotypes of 270 French bulls, 270 American bulls and 270 Chinese bulls were used to compare the relationship among three populations. These 270 French bulls were the progenies of the imported French bulls and cows. So did the American bulls. The Chinese bulls were randomly selected from the native population. All the animals were genotyped using Illumina Bovine SNP50 BeadChip (Illumina, San Diego, CA, USA). After deleting SNPs with a minor allele frequency smaller than 0.01, 45,404 SNPs on 29 autosomes were retained.

Model
GBLUP [9] was used for prediction of GEBV using Chinese RP. The model is as follows:

$$y = 1\mu + Zg + e$$

where y is a vector of DRP from Chinese population, μ is the overall mean, g is a vector of GEBV, 1 is a vector of ones, Z is the design matrix for linking g to y, and e is a vector of the random residuals. Random effects were assumed to be normally distributed as $g \sim N(0, G\sigma_g^2)$ and $e \sim N(0, I\sigma_e^2)$, where σ_g^2 is the additive genetic variance, σ_e^2 is the residual variance, G is the genomic relationship matrix constructed with all the markers using the formula $G = MM' / \sum 2p_i(1 - p_i)$ [9]. The genotypes in M were coded as 0, 1, and 2 for A_1A_1, A_1A_2 and A_2A_2 and then centralized by subtracting $2p_i$ [9], where p_i was the allele frequency of A_2 and was calculated based on the genotypes from the individuals used in the model. DMU package [10] was used to estimate variance components and obtain solutions of the mixed model equations.

Validation of genomic predictive ability
The Spearman's rank correlation coefficient between GEBVs predicted using different RPs was used as a measurement of concordance of GEBVs. The correlation coefficient between GEBVs evaluated from Chinese RPs and from French RPs was named as COR_{CF}. Accordingly, COR_{CA} was used for that between Chinese RPs and American RPs and COR_{FA} for that between French RPs and American RPs.

The measurement of relatedness between different populations
To examine the genetic relatedness between different RPs, three measurements of genetic distance were performed for 270 French bulls, 270 American bulls and 270 Chinese bulls: 1) the persistence of LD phase between two populations. It was calculated as the correlation of linkage disequilibrium (r^2) of adjacent marker pairs on each autosome [11, 12]. The persistence of LD phase between each pair of these three populations was named PER_{CF}, PER_{CA}, PER_{FA}. 2) the number of pair of related individuals between different populations. All pair-wise relationship can be classified as monozygotic twins, 1st-, 2nd- or 3rd- degree relatives by the estimation of kinship coefficients using Kinship-based INference for Genome-wide association study (KING) software package [13]. 3) the principal components (PCs) of marker genotype data. Principal components analysis (PCA) was performed on genotype using KING [13]. We used the plot of PC2 against PC1 as the description of genetic similarity among three populations.

Results
The comparison of genomic prediction using different RPs
The spearman's rank correlation coefficient between GEBVs using RP from two of three countries is shown in Table 1. For all traits, the correlation between GEBVs using French RP and using American RP (COR_{FA}) is much larger than the correlation between GEBVs using Chinese RP and using American RP (COR_{CA}) or French RP (COR_{CF}). COR_{FA} for fat percentage achieved the highest (0.862) while COR_{CA} for milk yield was the

Table 1 Spearman's rank correlation coefficient between GEBVs evaluated using different RP

Trait	COR_{CF}[a]	COR_{CA}[b]	COR_{FA}[c]
Milk yield	0.157	0.160	0.752
Fat percentage	0.162	0.167	0.862
Protein percentage	0.403	0.420	0.805
Conformation	0.442	0.448	0.643
Mammary system	0.359	0.402	0.765
Feet_legs	0.233	0.308	0.472

[a]The correlation between GEBVs estimated using Chinese RP and using American RP
[b]The correlation between GEBVs estimated using Chinese RP and using French RP
[c]The correlation between GEBVs estimated using French RP and using American RP

Ma et al. Journal of Animal Science and Biotechnology (2018) 9:64

Page 3 of 6

lowest (0.060). COR_{CF} ranged from 0.133 (for feet_legs) to 0.442 (for conformation). COR_{CA} was similar as COR_{CF} and ranged from 0.060 (for milk yield) to 0.420 (for protein percentage). COR_{FA} was much larger than COR_{CF} and COR_{CA} and ranged from 0.472 (for feet_legs) to 0.862 (for fat percentage).

The plot of GEBVs of milk yield using different RPs is presented in Fig. 1. The trends of GEBVs using American RP and French RP are similar while the trends of GEBVs of using Chinese RP are relative different from the GEBVs using the other two RPs.

LD and persistence of LD phase

The LD of each chromosome from each population and persistence of LD phase (PER) between populations are shown in Table 2. The mean r^2 of adjacent SNP pairs within each chromosome ranged from 0.13 (chromosomes 27 and 28) to 0.19 (chromosomes 6 and 14) for Chinese RP, 0.14 (chromosomes 27 and 28) to 0.20 (chromosomes 6 and 14) in both France and USA RPs. The mean r^2 across all chromosomes were 0.16 in China and 0.17 in France and USA. The persistence of LD phase between France and USA RPs was apparently higher than that between China and the other two countries. The PER_{CF} ranged from 0.893 of chromosome 28 to 0.959 of chromosome 14. The PER_{CA} ranged from 0.931 of chromosome 9 to 0.973 of chromosome 14. The PER_{FA} ranged from 0.942 of chromosome 19 to 0.974 of chromosome 29.

The kinship coefficients and classification of all pair-wise relationship

The number of pairs of related individuals in each relationship group which was determined by KING software was listed in Table 3. There was 1 pair of individuals in 1st-degree, 1 in 2nd-degree and 596 in 3rd-degree based on genomic relationship between Chinese population and French population. Based on genomic relationship between Chinese population and American population, there were 2 pairs of individuals in 1st-degree, 0 in 2nd-degree and 1,174 in 3rd-degree. Compared with genomic relationship between Chinese population and French population or American population, there were much more pairs of individuals in 1st, 2nd and 3rd degrees based on genomic relationship between French population and American population, which meant there were more related individuals in these two populations.

The principal component analysis (PCA)

Figure 2 illustrates that the relationship between French population and American population was closer than the relationship between them and Chinese population.

Discussion

In this study, we investigated the difference on GEBVs for French Holstein bulls using references from different countries. The genomic relatedness between different populations were investigated to illustrate the results. The results showed that the correlation between GEBVs estimated using French RP and using American RP was higher than the correlation between GEBVs estimated using Chinese RP and French/American RP. The LD phase persistence analysis, kinship coefficients and the PCA showed that the relationship between French population and American population was closer than that between Chinese population and American or French population.
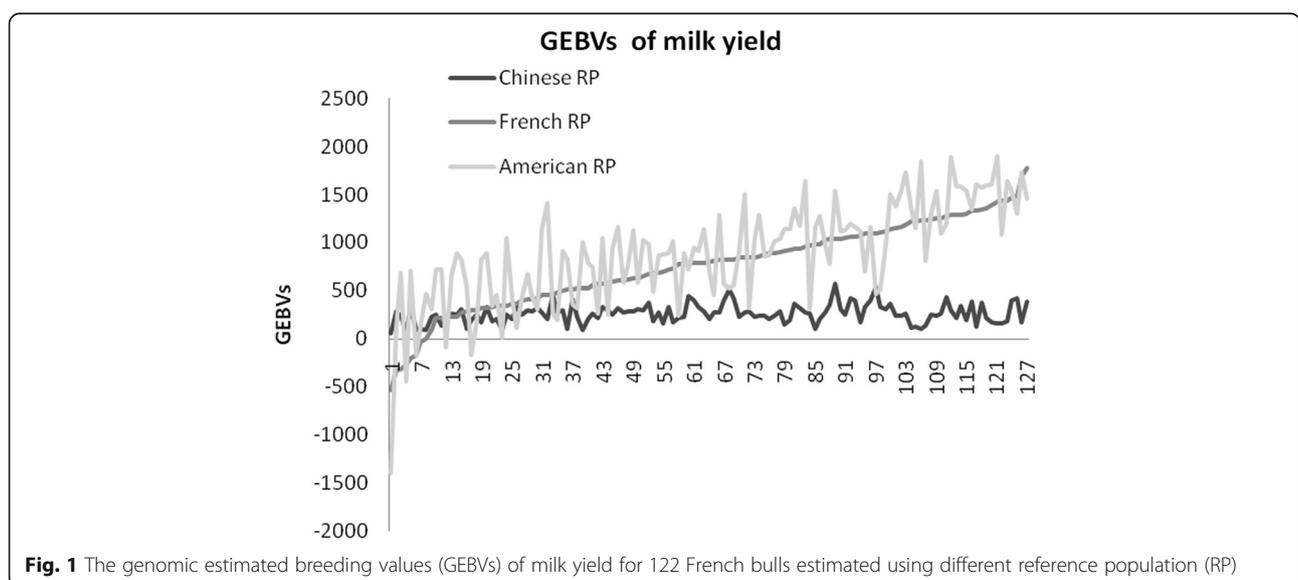
Fig. 1 The genomic estimated breeding values (GEBVs) of milk yield for 122 French bulls estimated using different reference population (RP)

Ma et al. Journal of Animal Science and Biotechnology (2018) 9:64

Page 4 of 6

Table 2 Linkage disequilibrium (LD) of adjacent markers for each *Bos Taurus* autosome (BTA)

BTA	No. of SNPs	Mean r^2			Persistence[a]		
		China	France	USA	China- France	China-USA	France-USA
1	3431	0.18	0.19	0.19	0.934	0.952	0.968
2	2829	0.17	0.18	0.17	0.930	0.952	0.962
3	2550	0.17	0.19	0.18	0.920	0.946	0.950
4	2570	0.16	0.17	0.17	0.909	0.933	0.948
5	2271	0.15	0.16	0.16	0.926	0.953	0.968
6	2575	0.19	0.20	0.20	0.934	0.954	0.966
7	2352	0.17	0.19	0.18	0.924	0.944	0.961
8	2430	0.17	0.18	0.18	0.912	0.944	0.949
9	2095	0.16	0.17	0.17	0.898	0.931	0.960
10	2206	0.17	0.19	0.19	0.927	0.938	0.955
11	2295	0.16	0.17	0.16	0.930	0.957	0.966
12	1773	0.15	0.16	0.16	0.920	0.947	0.958
13	1850	0.16	0.17	0.17	0.907	0.954	0.960
14	1831	0.19	0.20	0.20	0.959	0.973	0.972
15	1762	0.15	0.16	0.15	0.923	0.941	0.967
16	1726	0.17	0.18	0.19	0.926	0.954	0.948
17	1600	0.16	0.17	0.17	0.922	0.944	0.957
18	1376	0.16	0.17	0.16	0.930	0.944	0.965
19	1420	0.15	0.16	0.16	0.905	0.954	0.942
20	1568	0.17	0.19	0.18	0.919	0.949	0.966
21	1483	0.16	0.17	0.17	0.942	0.964	0.966
22	1324	0.15	0.16	0.15	0.934	0.959	0.968
23	1092	0.14	0.16	0.15	0.921	0.949	0.968
24	1312	0.16	0.17	0.16	0.932	0.959	0.969
25	1004	0.15	0.16	0.16	0.926	0.933	0.972
26	1116	0.15	0.17	0.16	0.904	0.943	0.963
27	981	0.13	0.14	0.14	0.932	0.945	0.967
28	981	0.13	0.14	0.14	0.893	0.939	0.955
29	1087	0.14	0.15	0.15	0.945	0.965	0.974
Mean	52,890[b]	0.16	0.17	0.17	0.924	0.949	0.962

[a]The correlation of r^2 of adjacent SNP pairs between two populations
[b]Sum over 29 autosomes

Table 3 The number of pairs of related individuals between different populations

Relationship	Criteria	No. of pairs between CF[a]	No. of pairs between CA[b]	No. of pairs between FA[c]
MZ twin	[0.354,]	0	0	0
1st-degree	[0.177,0.354]	1	2	32
2nd-degree	[0.0884,0.177]	1	0	1057
3rd-degree	[0.0442,0.0884]	596	1174	4447
Unrelated	[,0.0442]	71,492	71,724	66,554
Total		72,090	72,900	72,090

[a]Chinese population and French population
[b]Chinese population and American population
[c]French population and American population

Ma *et al. Journal of Animal Science and Biotechnology* (2018) 9:64

Page 5 of 6

Fig. 2 The principal components of marker genotype data. The first principle component (PC1) versus the second principle component (PC2) calculated using marker genotype data
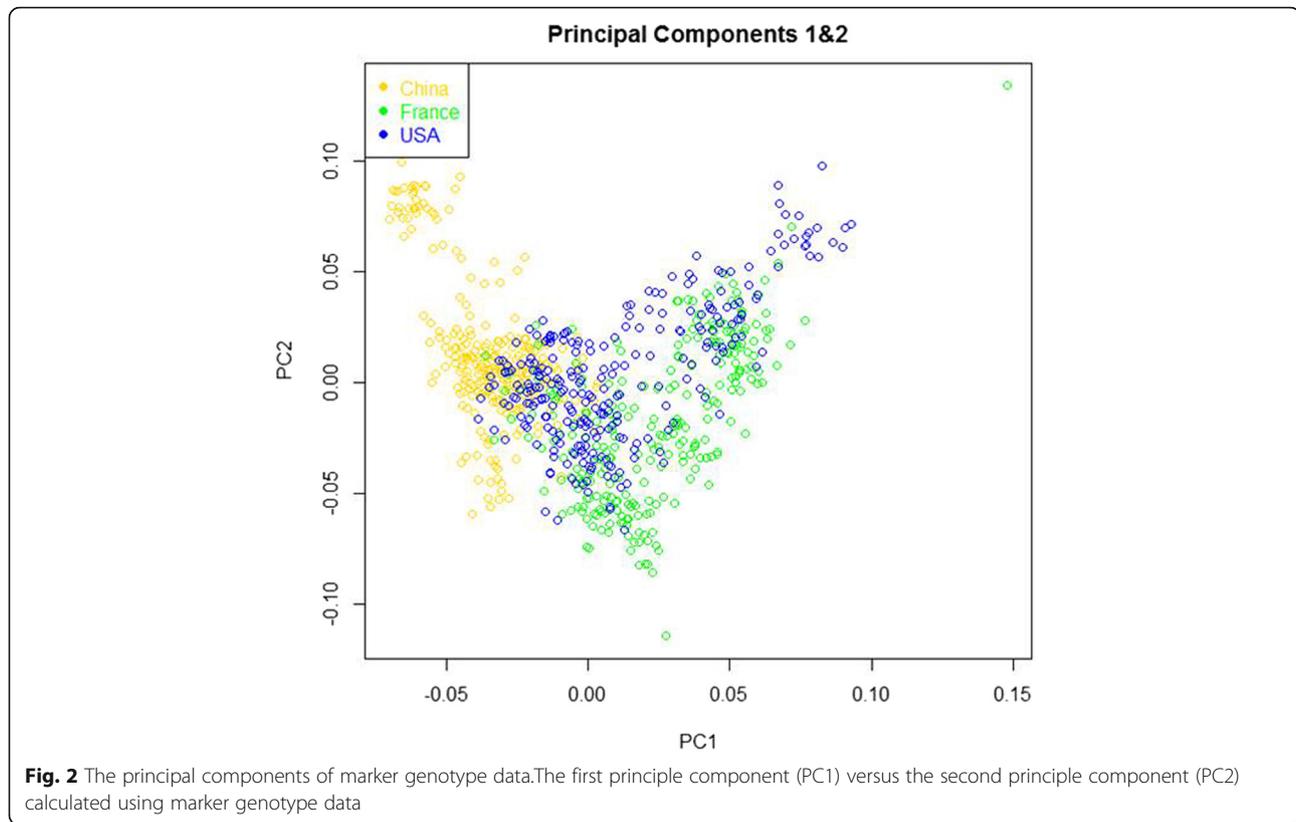
For combined RP, a close relationship between populations reflects a similar LD structures among populations which enabled the joint prediction feasible. Lund et al. [14] used European Holsteins as joint reference to predict Nordic Holstein, Dutch Holstein, French Holstein and German Holstein and found reliability improved by up to 10% compared with using separate RP. A joint Nordic Red dairy cattle RP was intended to improve the accuracy of genomic prediction in the previous study [15]. However, the results showed that the prediction for Swedish and Finnish population was improved slightly when the Danish Red dairy cattle were added into the RP since the relationship between Finnish Red and Swedish Red was closer compared with the relationship between Danish Red and the other two populations [15]. Similar pattern was observed when G matrix was used to measure the relationship among three countries in our study and the report from Brøndum et al. [15]. Higher related individuals were observed between Swedish and Finnish Red in their study and between American population and French population in our study. It is consisted with the conclusion from previous studies that the prediction ability was improved by including related individuals in the RP [16, 17]. The average of kinship among individuals from different countries was calculated, and the results showed

the average relationship of any two countries was similar with the others (data not shown). One of the reasons could be that too many small values diluted the close relationship, which illustrated that the average of kinship matrix was not suitable as the criterion to measure the relationship between populations.

Another reason leading to the spearman's rank correlation coefficient between GEBVs using Chinese RP and using American/French RP smaller than the other two correlations could be that the size of RP was different. Since Chinese RP in this study only included 1568 individuals, which may be not as informative as proven bulls from the other two countries. The combined RP between Nordic Holstein population, which is one member of Eurogenomics, and Chinese Holstein population had been utilized to investigate the improvement of reliability of genomic prediction in previous studies [5, 18]. The results showed the reliability of genomic prediction for Chinese population was improved greatly while little improvement for Nordic population [5]. Therefore, the size of RP should be considered when joint-population prediction was conducted besides taking the relationship between different populations into account. There is possibility to improve the genomic prediction ability for populations with a small number of RP even if the relationship between the added population and target population is distant.

Ma *et al. Journal of Animal Science and Biotechnology* (2018) 9:64

Page 6 of 6

Conclusions

Information from the other related populations was applied to improve the predictive ability. However, our results showed that the GEBVs were in different rank when a loose related population was used as RP. Integrating results from previous studies, we concluded that it was feasible to predict the GEBVs of a target population using RP of a related population in the condition that there was a close genetic relationship between these two populations and the size of related population is large.

Abbreviations
DRP: De-regressed proof; GEBVs: Genomic breeding values; LD: Linkage disequilibrium; PCA: Principal component analysis

Availability of data and materials
The datasets generated and/or analyzed during the current study are not publicly available due to data subject to Dairy Association of China but are available from the corresponding author on reasonable request.

Authors' contributions
PM, XL, HG, QZ, XD, CW conceived and designed this study. PM, JH, XL,WG did the analysis. PM, JH, HG contributed to the writing of manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]National Engineering Laboratory for Animal Breeding, Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China. [2]Department of Animal Science, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai 200240, China. [3]Gènes DIFFSUION, 3595 Route de Tourna, 59500 Douai, France. [4]Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, Tjele, Denmark. [5]Key Laboratory of Pig Molecular Quantitative Genetics, Institute of Animal Husbandry and Veterinary Medicine, Anhui Academy of Agricultural Sciences, Hefei 230031, China.

References
1. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157:1819–29.
2. Goddard ME, Hayes BJ. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet Nature Publishing Group. 2009;10:381–91.
3. Gao H, Christensen OF, Madsen P, Nielsen US, Zhang Y, Lund MS, et al. 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.
4. Pszczola M, Strabel T. Mulder H a, Calus MPL. Reliability of direct genomic values for animals with different relationships within and to the reference population. J Dairy Sci. 2012;95:389–400.
5. Zhou L, Ding X, Zhang Q, Wang Y, Lund MS, Su G. Consistency of linkage disequilibrium between Chinese and Nordic Holsteins and genomic prediction for Chinese Holsteins using a joint reference population. Genet Sel Evol. 2013;45:7.
6. Pryce J, Hayes B. A review of how dairy farmers can use and pro fit from genomic technologies. Anim Prod Sci. 2012;52:180–4.
7. Gao H, Su G, Janss L, Zhang Y, Lund MS. Model comparison on genomic predictions using high-density markers for different groups of bulls in the Nordic Holstein population. J Dairy Sci. 2013;96:4678–87.
8. Habier D, Tetens J, Seefried F-R, Lichtner P, Thaller G. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet Sel Evol. 2010;42:5.
9. VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008;91:4414–23.
10. Madsen P, Sørensen P, Su G, Damgaard LH, Thomsen H, Labouriau, R. DMU - a package for analyzing multivariate mixed models. In: Proceedings of the 8th World Congress on Genetics Applied to Livestock Production. Minas Gerais: Instituto Prociência. 2006;11–27.
11. Sargolzaei M, Schenkel FS, Jansen GB. Schaeffer LR. Extent of Linkage Disequilibrium in Holstein Cattle in North America. 2008:2106–17.
12. de Roos APW, Hayes BJ, Spelman RJ, Goddard ME. Linkage disequilibrium and persistence of phase in Holstein–Friesian. Jersey and Angus Cattle Genet. 2008;179:1503–12.
13. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010;26:2867–73.
14. Lund MS, Roos APW De, Vries AG De, Druet T, Ducrocq V, Guillaume F, et al. Improving genomic prediction by EuroGenomics collaboration. Proc WCGALP 2010, Leipzig. 2010;7–10.
15. Brøndum RF, Rius-Vilarrasa E, Strandén I, Su G, Guldbrandtsen B, Fikse WF, et al. Reliabilities of genomic prediction using combined reference data of the Nordic red dairy cattle populations. J Dairy Sci. 2011;94:4700–7.
16. Habier D, Fernando RL, Dekkers JCM. The impact of genetic relationship information on genome-assisted breeding values. Genetics. 2007;177:2389–97.
17. Wu X, Lund MS, Sun D, Zhang Q, Su G. Impact of relationships between test and training animals and among training animals on reliability of genomic prediction. J Anim Breed Genet. 2015;132:366–75.
18. Ma P, Lund MS, Ding X, Zhang Q, Su G. Increasing imputation and prediction accuracy for Chinese Holsteins using joint Chinese-Nordic reference population. J Anim Breed Genet. 2014;131:462–72.