J. Dairy Sci. 102:1–6
https://doi.org/10.3168/jds.2018-15298

Short communication: Replication of genome-wide association studies for milk production traits in Chinese Holstein by an efficient rotated linear mixed model

Dan Wang,* Chao Ning,* Jian-Feng Liu, Qin Zhang, and Li Jiang†
National Engineering Laboratory for Animal Breeding, Ministry of Agricultural Key Laboratory of Animal Genetics, Breeding and Reproduction, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

ABSTRACT

Milk is regarded as an important nutrient for humans, and Chinese Holstein cows provide high-quality milk for billions of Chinese people. Therefore, detecting quantitative trait nucleotides (QTN) or candidate genes for milk production traits in Chinese Holstein is important. In this study, we performed genome-wide association studies (GWAS) in a Chinese Holstein population of 6,675 cows and 71,633 single nucleotide polymorphisms (SNP) using deregressed proofs (DRP) as phenotypes to replicate our previous study in a population of 1,815 cows and 39,163 SNP using estimated breeding values (EBV) as phenotypes. The associations between 3 milk production traits—milk yield (MY), fat percentage (FP), and protein percentage (PP)—and the SNP were determined by using an efficient rotated linear mixed model, which benefits from linear transformations of genomic estimated values and Eigen decomposition of the genomic relationship matrix algorithm. In total, we detected 94 SNP that were significantly associated with one or more milk production traits, including 7 SNP for MY, 76 for FP, and 36 for PP; 87% of these SNP were distributed across *Bos taurus* autosomes 14 and 20. In total, 83 SNP were found to be located within the reported quantitative trait loci (QTL) regions, and one novel segment (between 1.41 and 1.49 Mb) on chromosome 14 was significantly associated with FP, which could be an important candidate QTL region. In addition, the detected intervals were narrowed down from the reported regions harboring causal variants. The top significant SNP for the 3 traits was ARS-BFGL-NGS-4939, which is located within the *DGAT1* gene. Five detected genes (*CYHR1*, *FOXH1*, *OPLAH*, *PLEC*, *VPS28*) have effects on all 3 traits. Our study provides a suite of QTN, candidate genes, and a novel QTL associated with milk production traits, and thus forms a solid basis for genomic selection and molecular breeding for milk production traits in Chinese Holstein.
Key words: Chinese Holstein, milk production traits, genome-wide association study (GWAS), quantitative trait nucleotides (QTN)

Short Communication

Milk is an important source of nutrients for humans, and milk production traits are the most important economic traits in dairy cattle. Over the last decades, several genome-wide association studies (**GWAS**) focusing on identifying genes for milk production traits have been performed in many different dairy cattle populations, and important candidate genes associated with milk production traits have been identified (Heyen et al., 1999; Cohen-Zinder et al., 2005; Mai et al., 2010; Pryce et al., 2010; Maxa et al., 2012; Meredith et al., 2012; Raven et al., 2014; Nayeri et al., 2016). In our previous GWAS study, we identified 105 significant SNP associated with milk production traits using the Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA) and the EBV pseudo-phenotype (Jiang et al., 2010). In the current study, more high-density SNP markers were used in the analyses. In addition, some studies have shown that EBV should not be used for association analysis (Garrick et al., 2009). Therefore, deregressed proofs (**DRP**) were derived using the method proposed by Garrick et al. (2009) based on the EBV and used as the response variable in the current study. We then carried out GWAS in a larger Chinese Holstein population by rotated linear mixed model algorithm, which takes advantage of linear transformations of genomic estimated values (Wang et al., 2014; Ning et al., 2018a) and Eigen decomposition of the genomic relationship matrix algorithm (Lippert et al., 2011). We further validated the results by comparing the detected quantitative trait nucleotides (**QTN**) to the QTLdb database (http://www.animalgenome.org/QTLdb; Hu et al., 2016) for each trait of interest, respectively.

Received June 28, 2018.
Accepted November 15, 2018.
*These authors contributed equally to this work.
†Corresponding author: lijiang@cau.edu.cn

In our study, a Chinese Holstein population of 6,675 cows were genotyped with the BovineSNP50v1 BeadChip (54,001 SNP; Illumina Inc.), the BovineSNP50v2 BeadChip (54,609 SNP; Illumina Inc.), and the GeneSeek Genomic Profiler HD (76,879 SNP; Neogen Corp., Lansing, MI), which were described in our previous longitudinal GWAS (Ning et al., 2018b). The individuals genotyped with the GeneSeek Genomic Profiler HD chip were used as reference, and all others were imputed to the high density of 70k SNP by using FImputev2.2 (Sargolzaei et al., 2014). After removing SNP with a minor allele frequency <0.03 and *P*-value of Hardy-Weinberg equilibrium $<10^{-6}$, 71,633 SNP remained and were utilized. We obtained the EBV of the 6,675 cows and their parents from the routine national genetic evaluation and estimated DRP by using the method proposed by Garrick et al. (2009). The summary statistics of the EBV and DRP for the 3 traits of the cows are presented in Supplemental Table S1 (https://doi.org/10.3168/jds.2018-15298).

A rapid genome-wide mixed-model association analysis method by linear transformations of genomic estimated values was used in the present study. The method has been described in the previous published articles (Wang et al., 2014; Ning et al., 2018a). We focus on the additive test here, and further improve the computational efficiency by Eigen decomposition of the genomic relationship matrix. A brief derivation of the method is shown below.

The whole additive SNP effect model is given as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}, \qquad [1]$$

where we assume that there are n individuals and m SNP. Thus, $\mathbf{y}$ is an $n \times 1$ vector of DRP values; $\mathbf{b}$ is a vector of fixed effects, which contains only population mean in our study; $\mathbf{u}$ is an $m \times 1$ vector of additive SNP effects; $\mathbf{e}$ is a vector of residual errors; $\mathbf{X}$ is the design matrix for the fixed effects; $\mathbf{Z}$ are standardized design matrices for additive SNP effects. Matrix $\mathbf{Z}$ is constructed as follows:

$$\mathbf{Z} = \left(\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_j, ..., \mathbf{z}_m\right) \Big/ \sqrt{2\sum p_j\left(1-p_j\right)}, \qquad [2]$$

where p_j is the allele frequency of allele A for the jth SNP, and $\mathbf{z}_j$ is the jth SNP vector with elements defined as

$$\mathbf{z}_j = \begin{cases} 2-2p_j & AA \\ 1-2p_j & Aa \\ 0-2p_j & aa \end{cases}. \qquad [3]$$

For Eq. [1], we define the following variance matrices of all random effects,

$$\mathbf{u} \sim \left(\mathbf{0}, \mathbf{I}\sigma_a^2\right),\ \mathbf{e} \sim \left(\mathbf{0}, \mathbf{I}\sigma_e^2\right), \qquad [4]$$

where $\mathbf{I}$ is the identity matrix, and σ_a^2 and σ_e^2 are additive and residual variances, respectively. We set $\mathbf{a} = \mathbf{Zu}$, and $\mathbf{a}$ is defined as individuals' additive effect vector. Then, Eq. [1] can be rewritten as

$$\mathbf{y} = \mathbf{Xb} + \mathbf{a} + \mathbf{e}. \qquad [5]$$

Then, the phenotypic (co)variances matrix is

$$\begin{aligned} \mathbf{V} &= \mathrm{var}(\mathbf{y}) = \mathrm{var}(\mathbf{a}+\mathbf{e}) = \mathbf{Z}\,\mathrm{var}(\mathbf{u})\mathbf{Z}' + \mathrm{var}(\mathbf{e}) \\ &= \mathbf{ZZ}'\sigma_a^2 + \mathbf{I}\sigma_e^2, \end{aligned} \qquad [6]$$

where we define $\mathbf{K} = \mathbf{ZZ}'$. Matrix $\mathbf{K}$ is the genomic relationship matrix identical to the first method of VanRaden (2008). Eigen decomposition of $\mathbf{K}$ gives $\mathbf{K} = \mathbf{UDU}'$, where $\mathbf{D}$ is a diagonal matrix containing the eigenvalues and $\mathbf{U}$ is the matrix of eigenvectors in the order of the corresponding eigenvalues ($\mathbf{UU}' = \mathbf{I}$). We rotate Eq. [3] with $\mathbf{U}'$, and the mixed model is rewritten as

$$\mathbf{U}'\mathbf{y} = \mathbf{U}'\mathbf{Xb} + \mathbf{U}'\mathbf{Zu} + \mathbf{U}'\mathbf{e}. \qquad [7]$$

Defining $\mathbf{y}^* = \mathbf{U}'\mathbf{y}$, $\mathbf{X}^* = \mathbf{U}'\mathbf{X}$, $\mathbf{Z}^* = \mathbf{U}'\mathbf{Z}$, $\mathbf{e}^* = \mathbf{U}'\mathbf{e}$, then

$$\mathbf{y}^* = \mathbf{X}^*\mathbf{b} + \mathbf{Z}^*\mathbf{u} + \mathbf{e}^*. \qquad [8]$$

The rotated phenotypic (co)variances matrix is

$$\begin{aligned} \mathbf{V}^* &= \mathrm{var}(\mathbf{y}^*) \\ &= \mathrm{var}\left(\mathbf{Z}^*\mathbf{u} + \mathbf{e}^*\right) \\ &= \mathbf{Z}^*\mathbf{Z}^{*\prime}\sigma_a^2 + \mathbf{UU}'\sigma_e^2 \\ &= \mathbf{U}'\mathbf{ZZU}\sigma_a^2 + \mathbf{UU}'\sigma_e^2 \\ &= \mathbf{U}'\mathbf{UDU}'\mathbf{U}\sigma_a^2 + \mathbf{I}\sigma_e^2 \\ &= \mathbf{D}\sigma_a^2 + \mathbf{I}\sigma_e^2. \end{aligned} \qquad [9]$$

The matrix $\mathbf{V}^*$ is a diagonal matrix, and its inversion can be obtained by replacing each element in the diagonal with its reciprocal. It is similar to the variance estimation method proposed by Lippert et al. (2011).

According to the BLUP method proposed by Henderson (1949), the random SNP effects can be predicted by

$$\hat{\mathbf{u}} = \left(\mathbf{I}\hat{\sigma}_a^2\right)\mathbf{Z}^{*\prime}\,\mathbf{P}^*\,\mathbf{y}^*, \qquad [10]$$

where

$$\mathbf{P}^* = \mathbf{V}^{*-1} - \mathbf{V}^{*-1}\mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\mathbf{V}^{*-1}.$$

The product of $\mathbf{P}^*$ and $\mathbf{y}^*$ can be obtained as follows:

$$\begin{aligned}\mathbf{P}^*\mathbf{y}^* &= \mathbf{V}^{*-1}\mathbf{y}^* - \mathbf{V}^{*-1}\mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{y}^* \\ &= \mathbf{V}^{*-1}\left(\mathbf{y}^* - \mathbf{X}^*\hat{\mathbf{b}}\right).\end{aligned} \quad [11]$$

The (co)variance matrix of estimated SNP effects is

$$\begin{aligned}\mathrm{var}(\hat{\mathbf{u}}) &= \left(\mathbf{I}\hat{\sigma}_a^2\right)\mathbf{Z}^{*\prime}\mathbf{P}^*\mathrm{var}\left(\mathbf{y}^*\right)\mathbf{P}^*\mathbf{Z}^*\left(\mathbf{I}\hat{\sigma}_a^2\right) \\ &= \left(\hat{\sigma}_a^2\right)^2\mathbf{Z}^{*\prime}\mathbf{P}^*\mathbf{V}^*\mathbf{P}^*\mathbf{Z}^*.\end{aligned} \quad [12]$$

We can prove that

$$\begin{aligned}&\mathbf{P}^*\mathbf{V}^*\mathbf{P}^* = \left[\mathbf{V}^{*-1} - \mathbf{V}^{*-1}\mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\right]\mathbf{V}^* \\ &\left[\mathbf{V}^{*-1} - \mathbf{V}^{*-1}\mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\right] \\ &= \left[\mathbf{I} - \mathbf{V}^{*-1}\mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\right] \\ &\left[\mathbf{V}^{*-1} - \mathbf{V}^{*-1}\mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\right] \\ &= \mathbf{V}^{*-1} - \mathbf{V}^{*-1}\mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^*\right)^{-1} \\ &\mathbf{X}^{*\prime}\mathbf{V}^{*-1} - \mathbf{V}^{*-1}\mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\mathbf{V}^{*-1} \\ &+ \mathbf{V}^{*-1}\mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^* \\ &\left(\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\mathbf{V}^{*-1} \\ &= \mathbf{V}^{*-1} - \mathbf{V}^{*-1}\mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\mathbf{V}^{*-1} \\ &= \mathbf{P}^*.\end{aligned} \quad [13]$$

Therefore, Eq. [11] can be further simplified as

$$\mathrm{var}\left(\hat{\mathbf{u}}\right) = \left(\hat{\sigma}_a^2\right)^2\mathbf{Z}^{*\prime}\mathbf{P}^*\mathbf{Z}^*. \quad [14]$$

For the jth SNP, the estimated effect and corresponding variance are

$$\hat{u}_j = \left(\mathbf{I}\hat{\sigma}_a^2\right)\mathbf{z}_j{}^{*\prime}\mathbf{P}^*\mathbf{y}^*, \quad [15]$$

$$\begin{aligned}\mathrm{var}\left(\hat{\mathbf{u}}\right)_{jj} &= \left(\hat{\sigma}_a^2\right)^2\mathbf{z}_j{}^{*\prime}\mathbf{P}^*\mathbf{z}_j{}^* \\ &= \left(\hat{\sigma}_a^2\right)^2\begin{bmatrix}\mathbf{z}_j{}^{*\prime}\mathbf{V}^{*-1}\mathbf{z}_j{}^* - \mathbf{z}_j{}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^*\left(\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^*\right)^{-1} \\ \mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{z}_j{}^*\end{bmatrix},\end{aligned} \quad [16]$$

where $\mathbf{z}_j$ is the jth SNP vector of the rotated SNP matrix $\mathbf{Z}^*$. With matrix $\left(\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^*\right)^{-1}$ precalculated, it requires only $O(nt)$, which is the time complexity, because V^{*-1} is a simple diagonal matrix. Here, t is the rank of $\mathbf{X}^*$, and $t = 1$ as the only intercept is included in the model.

The Wald chi-squared test for jth SNP effect is

$$\frac{\hat{u}_j^2}{\mathrm{var}\left(\hat{\mathbf{u}}\right)_{jj}} \sim \chi^2(1). \quad [17]$$

The source code can be freely accessed at https://github.com/chaoning/rotatedGWAS.

Before scanning markers in the GWAS, we first estimated additive variances, residual variances, and heritability for the 3 milk traits with DRP used as phenotypes (Supplemental Table S2; https://doi.org/10.3168/jds.2018-15298). The heritability of DRP can reflect the accuracy of DRP; that is, higher heritability indicates higher accuracy. Compared with the study of Jensen et al. (2012) in the Nordic Holstein population ($h^2 \sim 0.9$), the heritability of DRP in our Chinese Holstein population was much lower. This is because our population of Chinese Holsteins included mostly cows, whereas the Nordic Holstein population in their analysis comprised bulls with a large group of daughters.

To control false-positive rates, the Bonferroni correction was adopted to adjust for multiple testing, and the threshold for genome-wide significance was $0.05/N$, where N was the number of effective SNP calculated by the PLINK "-indep-pairwise 50 5 0.5" command (Purcell et al., 2007). This command calculates linkage disequilibrium (**LD**) between each pair of SNP within a window of 50 SNP, removes 1 SNP of a pair with LD >0.2 until there were no such pairs, and then moves the window forward by 5 SNP each time until it passes through the whole genome. Finally, 45,727 SNP passed this control filter and thus the genome-wise significant threshold was 0.05/45,727 = 1.093E-6.

A quantile-quantile (Q-Q) plot was used to examine whether population stratification existed in our experimental population (Pearson and Manolio, 2008). From Figure 1, it is apparent that the results were not threat-

ened by systematic bias for MY, FP, and PP. The association analysis showed that 94 SNP were significantly associated with milk production traits, including 7 SNP for MY, 76 for FP, and 36 for PP (Supplemental Tables S3, S4, and S5, respectively; https://doi.org/10.3168/jds.2018-15298). Genome-wide Manhattan plots of $-\log_{10}(P)$ are shown in Figure 2. Several SNP were associated with more than one milk production trait, which might be explained by genetic correlations among them (Toghiani, 2012; Zhao et al., 2015). The top significant SNP for the 3 traits was ARS-BFGL-NGS-4939, which is located within the *DGAT1* gene on BTA14; *DGAT1* is a functional causal gene affecting milk yield and milk composition traits, which encodes the key rate-limiting enzyme catalyzing the last step in triglyceride synthesis (Smith et al., 2000; Grisart et al., 2002, 2004). In addition, 5 genes (*CYHR1*, *FOXH1*, *OPLAH*, *PLEC*, *VPS28*) were detected to have potential effects on all 3 traits, which could be useful for molecular breeding for milk production in Chinese Holsteins. Furthermore, 6 common significant SNP were identified for the 3 traits and they showed effects for MY in the opposite direction to those for FP and PP (Supplemental Table S6; https://doi.org/10.3168/jds.2018-15298).

The significant SNP were mainly located within a 0.6-Mb segment (between 1.5 and 2.1 Mb) on BTA14 for MY; a 1.1-Mb segment (93.5–94.6 Mb) on BTA5 and a 4.3-Mb segment (1.4–5.7 Mb) on BTA14 for FP; and a 1.2-Mb segment (1.4–2.6 Mb) on BTA14 and a 3.4-Mb segment (31.2–34.6 Mb) on BTA20 for PP. Most of these significant SNP were distributed across BTA14 and 20, with 71% (67/94) on BTA14, which is consistent with the findings of previous studies (Jiang et al., 2010; Mai et al., 2010; Sermyagin et al., 2018), and 83 SNP fell in reported QTL regions (http://www.animalgenome.org/QTLdb). Furthermore, one novel segment (between 1.41 and 1.49 Mb) on BTA14 was significantly associated with FP and could be an important potential candidate QTL region.

Compared with our previous GWAS in Chinese Holsteins by Jiang et al. (2010), we extended the population size from 1,815 to 6,675 cows and increased the marker density from 39,163 to 71,633 SNP. The present study detected narrower genomic intervals harboring causal variants. A marker-derived genomic relationship matrix replaced the pedigree-based numerator relationship matrix to model the polygenic effect, which might be an important cause of the smaller QTL intervals. Additionally, DRP obtained from the routine national genetic evaluation of millions of individuals, instead of EBV, were used as the response variable, which benefited from correcting confounding environmental factors and taking advantage of population pedigree information. Our results were similar to the outcome of Ning et al. (2018b), who validated their novel longitudinal GWAS method using test-day records with the same chip data and conducted a GWAS of MY, FP, and PP in the first 3 parities.

The idea of implementing GWAS by linear transformation of genomic estimated values originated with Strandén and Garrick (2009), who derived the equivalency of SNP-BLUP and genomic (**G**)BLUP for genomic predictions. Taking advantage of their derivation, Wang et al. (2012) proved that BLUP of SNP effects could be obtained by back-solving SNP solutions after GBLUP. To test the significance of SNP effects,

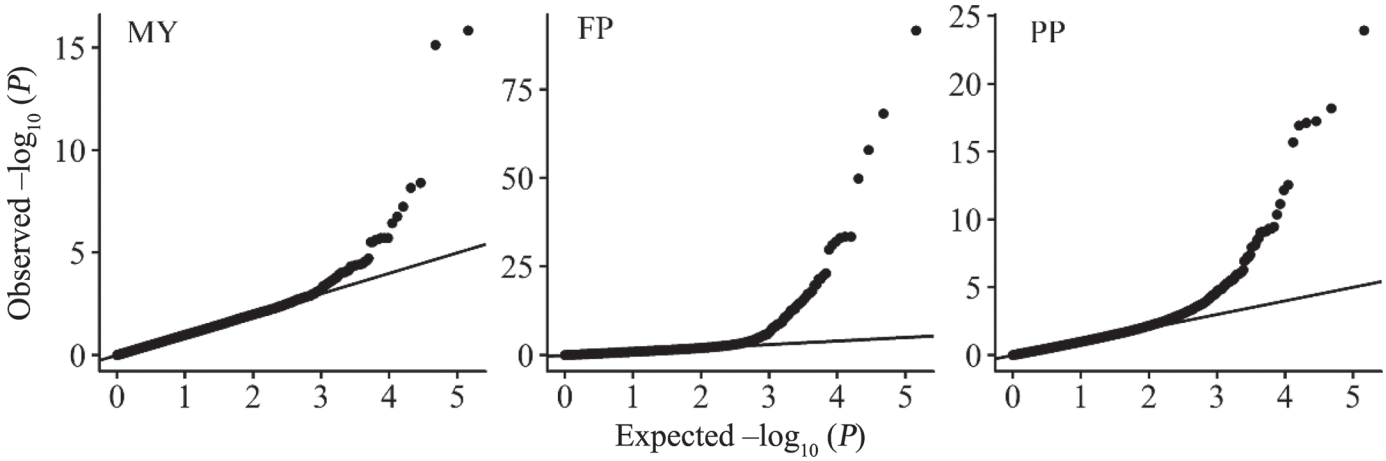

Figure 1. Quantile-quantile (Q-Q) plots of genome-wide association results for milk yield (MY), fat percentage (FP), and protein percentage (PP), respectively.

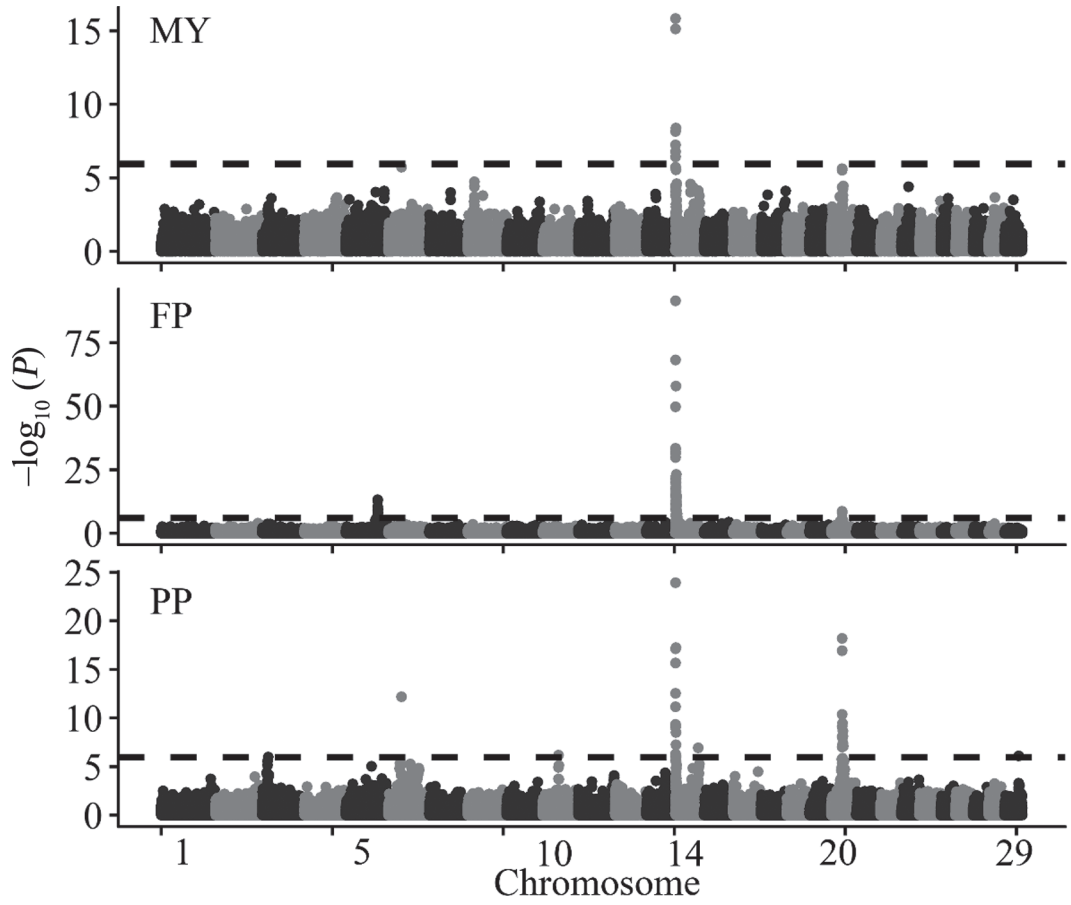

Figure 2. Genome-wide Manhattan plot of $-\log_{10}(P)$ for association of SNP loci with milk yield (MY), fat percentage (FP), and protein percentage (PP), respectively. Chromosomes 1 to 29 are shown separated by color. The horizontal dashed lines indicate the genome-wise significance threshold [$-\log_{10}$(1.093E-6)].

Gualdrón Duarte et al. (2014) built test statistics according to the estimated SNP effects and corresponding variances. Ning et al. (2018a) extended the algorithm to epistatic association analysis. In their studies, they found that their GBLUP-based method had almost identical *P*-values with better-known "one-SNP-at-a-time" methods, such as EMMA (efficient mixed model association; Kang et al., 2008) and FaST-LMM (factored spectrally transformed linear mixed models; Lippert et al., 2011). However, the time complexity of variance component estimation is $O(n^3)$ for each iteration. In this study, we first performed eigenvalue decomposition of genomic relationship matrix and then rotated the linear mixed model with the eigenvectors [time complexity of $O(n^2)$], which reduced the time complexity of variance component estimation to $O(n)$. The time complexity for GWAS is also reduced to $O(tn)$ in the rotated model compared with $O(tn^2)$ of previous GBLUP-based GWAS, where t is the rank of matrix $\mathbf{X}$. Because of these advantages, our optimized program took only 33 min for each trait analysis on a single core of Intel Xeon E5 2.2 GHz CPU, and it could significantly expedite variance component estimation and GWAS analysis compared with previous GBLUP-based GWAS.

In summary, we revealed 94 significant SNP distributed in 23 genes associated with MY, FP, and PP with our optimized program. These findings promote a better understanding of the genetic architecture of milk production traits and provide important information on potential markers for genomic selection in Chinese Holsteins.

ACKNOWLEDGMENTS

The study was financed by the National Major Development Program of Transgenic Breeding (2014ZX0800953B, China), the Natural Science Foundation of Beijing (6182023, China), the Education Foundation of China Agricultural University (2017DKB001,

China), and the Program for Changjiang Scholars and Innovative Research Team in University (IRT_15R62, China).

REFERENCES

Cohen-Zinder, M., E. Seroussi, D. M. Larkin, J. J. Loor, A. Everts-van der Wind, J. H. Lee, J. K. Drackley, M. R. Band, A. G. Hernandez, M. Shani, H. A. Lewin, J. I. Weller, and M. Ron. 2005. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Res. 15:936–944.

Garrick, D. J., J. F. Taylor, and R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41:55.

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell. 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res. 12:222–231.

Grisart, B., F. Farnir, L. Karim, N. Cambisano, J. J. Kim, A. Kvasz, M. Mni, P. Simon, J. M. Frere, W. Coppieters, and M. Georges. 2004. Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. Proc. Natl. Acad. Sci. USA 101:2398–2403.

Gualdrón Duarte, J. L., R. J. Cantet, R. O. Bates, C. W. Ernst, N. E. Raney, and J. P. Steibel. 2014. Rapid screening for phenotype-genotype associations by linear transformations of genomic evaluations. BMC Bioinformatics 15:246.

Henderson, C. 1949. Estimation of changes in herd environment. J. Dairy Sci. 32:706–715.

Heyen, D. W., J. I. Weller, M. Ron, M. Band, J. E. Beever, E. Feldmesser, Y. Da, G. R. Wiggans, P. M. VanRaden, and H. A. Lewin. 1999. A genome scan for QTL influencing milk production and health traits in dairy cattle. Physiol. Genomics 1:165–175.

Hu, Z. L., C. A. Park, and J. M. Reecy. 2016. Developmental progress and current status of the animal QTLdb. Nucleic Acids Res. 44(D1):D827–D833.

Jensen, J., G. S. Su, and P. Madsen. 2012. Partitioning additive genetic variance into genomic and remaining polygenic components for complex traits in dairy cattle. BMC Genet. 13:44.

Jiang, L., J. Liu, D. Sun, P. Ma, X. Ding, Y. Yu, and Q. Zhang. 2010. Genome wide association studies for milk production traits in Chinese Holstein population. PLoS One 5:e13661.

Kang, H. M., N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman, M. J. Daly, and E. Eskin. 2008. Efficient control of population structure in model organism association mapping. Genetics 178:1709–1723.

Lippert, C., J. Listgarten, Y. Liu, C. M. Kadie, R. I. Davidson, and D. Heckerman. 2011. FaST linear mixed models for genome-wide association studies. Nat. Methods 8:833–835.

Mai, M. D., G. Sahana, F. B. Christiansen, and B. Guldbrandtsen. 2010. A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. J. Anim. Sci. 88:3522–3528.

Maxa, J., M. Neuditschko, I. Russ, M. Forster, and I. Medugorac. 2012. Genome-wide association mapping of milk production traits in Braunvieh cattle. J. Dairy Sci. 95:5357–5364.

Meredith, B. K., F. J. Kearney, E. K. Finlay, D. G. Bradley, A. G. Fahey, D. P. Berry, and D. J. Lynn. 2012. Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. BMC Genet. 13:21.

Nayeri, S., M. Sargolzaei, M. K. Abo-Ismail, N. May, S. P. Miller, F. Schenkel, S. S. Moore, and P. Stothard. 2016. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. BMC Genet. 17:75.

Ning, C., D. Wang, H. Kang, R. Mrode, L. Zhou, S. Xu, and J. F. Liu. 2018a. A rapid epistatic mixed-model association analysis by linear retransformations of genomic estimated values. Bioinformatics 34:1817–1825.

Ning, C., D. Wang, X. Zheng, Q. Zhang, S. Zhang, R. Mrode, and J. F. Liu. 2018b. Eigen decomposition expedites longitudinal genome-wide association studies for milk production traits in Chinese Holstein. Genet. Sel. Evol. 50:12.

Pearson, T. A., and T. A. Manolio. 2008. How to interpret a genome-wide association study. JAMA 299:1335–1344.

Pryce, J. E., S. Bolormaa, A. J. Chamberlain, P. J. Bowman, K. Savin, M. E. Goddard, and B. J. Hayes. 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. J. Dairy Sci. 93:3331–3345.

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

Raven, L. A., B. G. Cocks, and B. J. Hayes. 2014. Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. BMC Genomics 15:62.

Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics 15:478.

Sermyagin, A. A., E. A. Gladyr, K. V. Plemyashov, A. A. Kudinov, A. V. Dotsev, T. E. Deniskova, and N. A. Zinovieva. 2018. Genome-wide association studies for milk production traits in Russian population of Holstein and black-and-white cattle. Pages 591–599 in Proceedings of the Scientific-Practical Conference "Research and Development - 2016." K. Anisimov et al., ed. Springer, Cambridge, UK.

Smith, S. J., S. Cases, D. R. Jensen, H. C. Chen, E. Sande, B. Tow, D. A. Sanan, J. Raber, R. H. Eckel, and R. V. Farese Jr.. 2000. Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking Dgat. Nat. Genet. 25:87–90.

Strandén, I., and D. J. Garrick. 2009. Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. J. Dairy Sci. 92:2971–2975.

Toghiani, S. 2012. Genetic relationships between production traits and reproductive performance in Holstein dairy cows. Arch. Tierzucht 55:458–468.

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

Wang, H., I. Misztal, I. Aguilar, A. Legarra, R. L. Fernando, Z. Vitezica, R. Okimoto, T. Wing, R. Hawken, and W. M. Muir. 2014. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. Front. Genet. 5:134.

Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. Genet. Res. (Camb.) 94:73–83.

Zhao, F. P., G. Guo, Y. C. Wang, X. Y. Guo, Y. Zhang, and L. X. Du. 2015. Genetic parameters for somatic cell score and production traits in the first three lactations of Chinese Holstein cows. J. Integr. Agric. 14:125–130.