Genome-wide association studies to identify quantitative trait loci affecting milk production traits in water buffalo

J. J. Liu,*† A. X. Liang,*† G. Campanile,‡ G. Plastow,§ C. Zhang,§ Z. Wang,§ A. Salzano,‡ B. Gasparrini,‡ M. Cassandro,# and L. G. Yang*†[1]
*Key Lab of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, Huazhong Agriculture University, Wuhan, Hubei, China 430070
†Hubei Province's Engineering Research Center in Buffalo Breeding and Products, Wuhan, Hubei, China 430070
‡Department of Veterinary Medicine and Animal Productions, University of Naples "Federico II," Naples, Italy 80137
§Department of Agricultural, Food, and Nutritional Sciences, University of Alberta, Edmonton, AB, Canada T6G 2C8
#Department of Agronomy, Food, Natural Resources, Animal, and Environment, University of Padova, Agripolis, Legnaro, Italy 35020

ABSTRACT

Water buffalo is the second largest resource of milk supply around the world, and it is well known for its distinctive milk quality in terms of fat, protein, lactose, vitamin, and mineral contents. Understanding the genetic architecture of milk production traits is important for future improvement by the buffalo breeding industry. The advance of genome-wide association studies (GWAS) provides an opportunity to identify potential genetic variants affecting important economical traits. In the present study, GWAS was performed for 489 buffaloes with 1,424 lactation records using the 90K Affymetrix Buffalo SNP Array (Affymetrix/Thermo Fisher Scientific, Santa Clara, CA). Collectively, 4 candidate single nucleotide polymorphisms (SNP) in 2 genomic regions were found to associate with buffalo milk production traits. One region affecting milk fat and protein percentage was located on the equivalent of *Bos taurus* autosome (BTA)3, spanning 43.3 to 43.8 Mb, which harbored the most likely candidate genes *MFSD14A*, *SLC35A3*, and *PALMD.* The other region on the equivalent of BTA14 at 66.5 to 67.0 Mb contained candidate genes *RGS22* and *VPS13B* and influenced buffalo total milk yield, fat yield, and protein yield. Interestingly, both of the regions were reported to have quantitative trait loci affecting milk performance in dairy cattle. Furthermore, we suggest that buffaloes with the C allele at AX-85148558 and AX-85073877 loci and the G allele at AX-85106096 locus can be selected to improve milk fat yield in this buffalo-breeding program. Meanwhile, the G allele at AX-85063131 locus can be used as the favorable allele for improving milk protein percentage. Genomic prediction showed that the reliability of genomic estimated breeding values (GEBV) of 6 milk production traits ranged from 0.06 to 0.22, and the correlation between estimated breeding values and GEBV ranged from 0.23 to 0.35. These findings provide useful information to understand the genetic basis of buffalo milk properties and may play a role in accelerating buffalo breeding programs using genomic approaches.

Key words: genome-wide association study, quantitative trait loci, milk production, buffalo

INTRODUCTION

Water buffalo (*Bubalus bubalis*) is an important livestock species for the agricultural economy, supplying milk, meat, and draft power (Warriach et al., 2015). The global buffalo population was recently estimated to be 194 million, 97% of which were reared in Asia (FAOSTAT, 2014). Buffalo is well known for its high milk quality, with higher fat (6.4–8.0% vs. 4.1–5.0%) and protein (4.0–4.5% vs. 3.4–3.6%) contents than cow milk (Khedkar et al., 2016). Its compositional and functional properties make buffalo milk suitable for manufacture of dairy products, such as superior cream, butter, yogurt, and cheese, especially mozzarella cheese (Michelizzi et al., 2010). As one of the most famous dairy buffalo breeds, the Italian Mediterranean buffalo has reached a high productivity standard due to the intense work of selection and study by the Italian buffalo-breeding program. Italian Mediterranean buffaloes are of the river type, whereas in China and some southeast Asian countries, most buffalo breeds belong to the swamp type and have poor milk production. To improve buffalo milk performance, a crossbreeding strategy is often used by the traditional breeding industry. Although a remarkable improvement has been achieved over the years, milk production in buffalo is

still considerably lower than in cow. Presently, buffalo milk is ranked second for the world's total milk production but accounts for only about 13% of the total (FAOSTAT, 2014). Therefore, understanding the genetic architecture of milk properties is essential to accelerate the genetic improvement in water buffalo-breeding programs.

As is well known, QTL can be utilized to identify candidate genes and contribute to the dissection of genetic mechanisms underlying economic traits in animals. A large number of QTL have been detected for milk production traits in cattle, such as *DGAT1*, *ABCG2*, and *SCD1* genes (Lengi and Corl, 2007; Weller and Ron, 2011). Polymorphisms for *DGAT1* and *ABCG2* were detected in buffalo although the alleles were found to be fixed in some breeds (Tantia et al., 2006; Shi et al., 2012). This implies that these 2 genes may be responsible for the high milk fat and milk protein in buffalo. To more precisely identify markers and genomic regions associated with quantitative traits, genome-wide association studies (**GWAS**) are increasingly used and successfully incorporated into dairy cattle (Hayes et al., 2009), pig (Do et al., 2015), and poultry (Fulton, 2012) breeding programs. However, genomic research on buffalo is still very limited. Without an available complete buffalo genomic reference map, buffalo genomic studies use information from the most homologous species available—cattle (Amaral et al., 2008; Di Meo et al., 2008; Venturini et al., 2014). In our previous study (Wu at al., 2013), we investigated the transferability of Bovine SNP50 BeadChip (Illumina Inc., San Diego, CA) data from cattle to buffalo for GWAS and identified 7 SNP among 935 bovine SNP associating with buffalo milk performance. The development of a buffalo genotyping array opens new opportunities to explore key genes regulating buffalo milk properties and provides the possibility of utilizing genomic selection in the buffalo breeding industry (Iamartino et al., 2013). With the buffalo SNP array, 78 SNP (Iamartino et al., 2013; de Camargo et al., 2015; El-Halawany et al., 2015) have been found to influence milk production traits among different buffalo breeds. Furthermore, several suggestive genomic regions on BTA1, 5, 6, and 27 have been found to be associated with daily milk yield using GWAS in Egyptian buffalo (El-Halawany et al., 2017). Therefore, we hypothesize that a variety of novel genes and QTL may be identified to help with the genetic dissection of buffalo milk performance.

The present study aimed to detect important markers and genomic regions affecting milk production traits, and to investigate the feasibility of genomic selection as a potential selection strategy in buffalo breeding programs for accelerating the genetic improvement of buffalo economic traits.

MATERIALS AND METHODS

Ethics Statement

The data and samples of buffalo used in this study were provided by The Italian Buffalo Breeders Association (ANASB), which is responsible for the official herd book of buffalo population in Italy. The experimental design and animal treatments were approval by the Ethical Animal Care and Use Committee of University of Naples "Federico II." Moreover, the farmers were previously informed and in agreement with purpose and methods used.

Animal Resources and Phenotypic Data

A total of 1,424 lactation records were collected from 489 Italian Mediterranean buffaloes born from 2000 to 2011 and reared in 4 herds in southern Italy. Pedigree data consisted of 937 animals over 3 generations. Six milk production traits, peak milk yield (**PM**), total milk yield (**MY**), fat yield (**FY**), fat percentage (**FP**), protein yield (**PY**), and protein percentage (**PP**), were recorded. All milk production traits were adjusted to 270 d in milk (Baldi et al., 2011). The linear model used to adjust the records, including the factors herd-season (**HS**, 4 farms and 2 seasons), year of calving (<2005 and 2005–2014), parity (1 to 7 and ≥8), and calf sex (male and female), were tested through a fixed linear model. Only significant ($P < 0.05$) factors were included in the model as fixed effects to adjust the records of lactation length (**LL**) between 150 and 270 d using the LSM method. The records for LL >270 d were truncated at the 270-d milk production. Data from LL <150 d were excluded from this analysis. The adjusted formula for MY, FY, and PY was

$$Y270 = Yn \times \frac{LSM270}{LSMn},$$

where $Y270$ is the 270-d adjusted phenotypes MY270, FY270, and PY270; Yn is the observed phenotype at day n; $LSM270$ and $LSMn$ are the least squares means of the observed phenotypes at d 270 and day n; n (150 < n < 270) is the days of LL. Then, FP270 and PP270 were adjusted as follows:

$$FP270 = \frac{FY270}{MY270} \text{ and } PP270 = \frac{PY270}{MY270}.$$

Genotyping and Quality Control

Genomic DNA was extracted from whole blood using a standard phenol-chloroform extraction protocol.

Genotyping was conducted at Delta Genomics (Edmonton AB, Canada) using the 90K Axiom Buffalo SNP Array (Affymetrix/Thermo Fisher Scientific, Santa Clara, CA). Quality control (**QC**) for genotypes was performed using PLINK1.9 (Purcell and Chang, 2015) for individuals with a call rate $\geq 97\%$. The SNP were selected with the criteria of genotyping call rate $\geq 95\%$, minor allele frequency ≥ 0.05, P-value of chi-squared test of Hardy-Weinberg equilibrium $\geq 10^{-5}$, and with map information.

Breeding Values and Estimation of Genetic and Phenotypic Parameters

The EBV for the 6 milk production traits (PM, MY270, FY270, PY270, FP270, and PP270) were estimated with a univariate animal model suing ASReml 3.0 (Gilmour et al., 2009):

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{p} + \mathbf{e},$$

where $\mathbf{y}$ is a vector of phenotypes adjusted to 270-d records for all traits except PM; $\mathbf{X}$ is a incidence matrix associated with the fixed effects; $\mathbf{b}$ is a fixed vector containing HS, calving year, parity, and calf sex for 270-d MY270, FY270, and PY270, and containing HS, calving year, and parity for FP270, PP270, and PM traits; $\mathbf{Z}_1$ is a incidence matrix associated with individual additive genetic effects; $\mathbf{a}$ is a random vector of animal's breeding values or additive genetic effects $\left[\mathbf{a} \sim \left(\mathbf{0}, \mathbf{A}\sigma_a^2\right)\right]$; $\mathbf{Z}_2$ is a incidence matrix associated with animals' permanent environmental effects; $\mathbf{p}$ is a vector of permanent environmental effects of individual animal $\left[\mathbf{p} \sim \left(\mathbf{0}, \mathbf{I}\sigma_p^2\right)\right]$; and $\mathbf{e}$ is a random vector of residual errors $\left[\mathbf{e} \sim \left(\mathbf{0}, \mathbf{I}\sigma_e^2\right)\right]$. The σ_a^2, σ_p^2, and σ_e^2 terms are the additive genetic, permanent environment, and residual error variances, respectively; and $\mathbf{A}$ and $\mathbf{I}$ are the additive genetic relationship and identity matrices, respectively.

The genetic and phenotypic parameters, including heritability and genetic and phenotypic correlations, were estimated with the following pairwise bivariate animal model using ASReml 3.0 (Gilmour et al., 2009):

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & 0 \\ 0 & \mathbf{X}_2 \end{bmatrix}\begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & 0 \\ 0 & \mathbf{Z}_2 \end{bmatrix}\begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_3 & 0 \\ 0 & \mathbf{Z}_4 \end{bmatrix}\begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix},$$

where $\mathbf{y}_1$ and $\mathbf{y}_2$ are the vectors of phenotypes for traits 1 and 2, respectively; $\mathbf{X}_1$ and $\mathbf{X}_2$ are the incidence matrices associating the fixed effects to vectors $\mathbf{y}_1$ and $\mathbf{y}_2$, respectively; $\mathbf{b}_1$ and $\mathbf{b}_2$ are the vectors of fixed effects for traits 1 and 2, respectively; $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are the incidence matrices associating the random additive genetic effect of $\mathbf{a}_1$ and $\mathbf{a}_2$ to vectors of $\mathbf{y}_1$ and $\mathbf{y}_2$, respectively; $\mathbf{Z}_3$ and $\mathbf{Z}_4$ are the incidence matrices associating the random permanent effects $\mathbf{p}_1$ and $\mathbf{p}_2$ to vectors of $\mathbf{y}_1$ and $\mathbf{y}_2$, respectively; and $\mathbf{e}_1$ and $\mathbf{e}_2$ are the random vectors of residual errors for traits 1 and 2, respectively. The variance components definition for $\mathbf{a}$, $\mathbf{p}$, and $\mathbf{e}$ remained as the same as in the univariate model. The heritability was an averaged estimates using variance components obtained from the corresponding pairwise bivariate analyses (Miar et al., 2014).

Genome-Wide Association Analyses

Deregressed EBV (**DEBV**) were calculated according to the method reported by Garrick et al. (2009). The deregression procedure was performed by weighted EBV with the weight $w_i = \left(1 - h^2\right)\Big/\left\{\left[c + \left(1 - r_i^2\right)\Big/r_i^2\right]h^2\right\}$, where c was the part of genetic variance not explained by markers, which was assumed to be 0.4 (Saatchi et al., 2014); h^2 was the heritability of the trait; and r_i^2 was the reliability of EBV of the ith animal. The DEBV were considered a new phenotype for subsequent genomic association and prediction analyses.

A total of 412 animals with both DEBV and genotype were used to perform the genome-wide association analysis using the ridge regression BLUP (rrBLUP) model in R (Endelman, 2011):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} + \mathbf{e},$$

where $\mathbf{y}$ is the vector of DEBV, $\mathbf{X}$ is a vector of coded SNP genotypes (-1, 0, or 1), $\boldsymbol{\beta}$ is the fixed additive genetic value attributed to SNP under evaluation, $\mathbf{u}$ is the vector of the background polygenic effects with normal distribution $\left[\mathbf{u} \sim \left(\mathbf{0}, \mathbf{K}\sigma_u^2\right)\right]$, where $\mathbf{K}$ is the genomic relationship matrix obtained from pedigree information (Endelman, 2011) and σ_u^2 is the genetic variance, and $\mathbf{e}$ is the vector of residual errors with normal distribution $\left[\mathbf{e} \sim \left(\mathbf{0}, \mathbf{I}\sigma_e^2\right)\right]$, where $\mathbf{I}$ is the identity matrix and σ_e^2 is the residual variance. False discovery rate (**FDR**) was adopted to adjust P-value for all detected SNP, and the genome-wide significant threshold of SNP was defined as FDR <0.10 and the suggestive threshold as $P < 10^{-4}$. The LSM of EBV for the 3 genotypes affecting milk production traits of significant SNP were calculated by a general liner model using R package lsmeans (Lenth, 2016), and the significant threshold was set at $P < 0.05$.

Haplotype Analyses

The 0.5-Mb genomic window around significant SNP was identified to construct the haplotype blocks using Haploview 4.2 (Barrett et al., 2005). Association between each haplotype combination and 6 milk production traits were evaluated with Bonferroni *t*-test in SAS 9.4 (SAS Institute Inc., Cary, NC) as described by Yang et al. (2015). The genomic region that covered haplotype blocks affecting ($P < 0.05$) buffalo milk production traits was identified as a candidate region.

Bioinformatics Analyses of Candidate Genes

Because the 90K Affymetrix Buffalo SNP Array was aligned to the bovine genome, identification of genes within candidate regions was based on the *Bos taurus* UMD3.1 genomic assembly (http://bovinegenome.org/?q=node/61). Genes were submitted to the database Bovinemine (http://bovinegenome.org/bovinemine) for pathway and gene ontology (**GO**) term annotation. The functions and related information about the genes were summarized using the available database of GeneCards (http://www.genecards.org/).

Genomic Prediction

A 5-fold cross validation was used to evaluate the accuracy of predicted genomic EBV (**GEBV**; Lukić et al., 2015). Briefly, the 412 buffaloes were randomly grouped into 5 folds. Each time, 4 folds (330 buffaloes) were set as the training population and the remaining fold (82 buffaloes) was used as the validation population. For each validation, the EBV and reliability were used as pseudo-phenotypes in a genomic BLUP model, and GEBV and its reliability were evaluated through the genomic relationship matrix using the gebv software (Sargolzaei et al., 2009). Both the reliability of GEBV and the correlation between EBV and GEBV were used to evaluate the prediction accuracy (Moser et al., 2009). The average reliability value of the 5 cross-fold validation was finally used to evaluate the feasibility of genomic prediction.

RESULTS AND DISCUSSION

Phenotype and Genotype Data Description

Italian Mediterranean buffalo lactation usually lasts around 270 d. Hence, we adjusted the buffalo milk production traits for 270 d to estimate more reliable EBV (Baldi et al., 2011). Six 270-d adjusted traits including PM, MY270, FY270, FP270, PY270, and PP270 were used for GWAS. In Table 1, the phenotypic descriptive statistics of analyzed samples are reported. Overall, the Italian Mediterranean buffalo in this study can be considered a highly selected breed with a higher milk production yield (2,965.6 ± 516.1 kg) than that reported in Murrah (1,712.5 kg; Tonhati et al., 2004) and Nili-Ravi (1,984.4 kg; Khan et al., 2005) buffaloes. The peak milk yield of total lactation showed an average of 14.9 ± 2.6 kg. Buffalo milk is usually rich in fat and protein (Rosati and Van Vleck, 2002). The average fat and protein percentages in our population were 8.3 and 4.7%, respectively, similar to that of Murrah and Nili-Ravi buffalo (Tonhati et al., 2004; Khan et al., 2005) but higher than that reported by Khedkar et al. (2016). The buffalo milk fat and protein yields were 244.9 ± 47.3 and 137.5 ± 22.9 kg, respectively. In the future, it could be interesting to assess coagulation ability of buffalo milk to improve cheese yield genetically.

The estimated genetic parameters for each trait are shown in Table 2. We found the heritability of buffalo milk production traits to be moderate (0.19–0.38), and the h^2 values for FY270 (0.35) and PP270 (0.38) were higher than those of other traits (0.19–0.33). In agreement with previous studies (Rosati and Van Vleck, 2002; de Camargo et al., 2015), we also detected high and positive genetic correlations among PM, MY270, FY270, and PY270, ranging from 0.80 to 0.98, and a moderate genetic correlation between FP270 and PP270 (0.58). As the genetic parameters of studied traits underpin further GWAS (de Camargo et al., 2015), there would be more power to detect SNP associated with FY270 and PP270 because they have higher heritability values.

We filtered genotype data under a QC scenario, and several 462 individuals and 60,387 SNP remained. The

Table 1. Descriptive statistics of the 270-d adjusted milk production traits in water buffalo

Trait[1]	n	Mean	SD	Minimum	Maximum	CV (%)
PM (kg)	1,408	14.9	2.6	6.6	28.0	17.7
MY270 (kg)	1,408	2,965.6	516.1	1,394.0	5,160.0	17.4
FY270 (kg)	1,401	244.9	47.3	102.0	429	19.3
FP270 (%)	1,401	8.3	0.9	5.5	11.6	11.2
PY270 (kg)	1,408	137.5	22.9	63.0	259.0	16.7
PP270 (%)	1,408	4.7	0.3	3.8	5.5	5.6

[1]PM = peak milk yield; MY270 = 270-d total milk yield; FY270 = 270-d fat yield; FP270 = 270-d fat percentage; PY270 = 270-d protein yield; PP270 = 270-d protein percentage.

Table 2. Estimates of heritability (bold, in diagonal), genetic correlation (above diagonal), and phenotypic correlation (below diagonal) among the studied traits

Trait[1]	PM	MY270	FY270	FP270	PY270	PP270
PM	**0.28**	0.91	0.81	−0.33	0.81	−0.31
MY270	0.88	**0.33**	0.82	−0.46	0.95	−0.49
FY270	0.77	0.83	**0.35**	0.49	0.98	0.29
FP270	−0.25	−0.28	0.40	**0.27**	0.13	0.58
PY270	0.84	0.96	0.87	−0.10	**0.19**	0.17
PP270	−0.16	−0.22	0.15	0.50	0.15	**0.38**

[1]PM = peak milk yield; MY270 = 270-d total milk yield; FY270 = 270-d fat yield; FP270 = 270-d fat percentage; PY270 = 270-d protein yield; PP270 = 270-d protein percentage.

SNP that passed QC were distributed uniformly across bovine chromosomes based on the bovine genome (Supplemental Table S1; https://doi.org/10.3168/jds .2017-13246) with an average density of 23 SNP/Mb on each chromosome, which provided powerful information for further GWAS and genomic prediction.

GWAS

Because of the special phenotypic data structure due to unbalanced repeated lactation records, DEBV of buffalo individuals were calculated and selected as a new response variable for GWAS analysis. The DEBV make good use of available information from genotyped animals as well as from their relatives, which can appropriately avoid bias introduced by simply pooling or averaging data information and account for heterogeneous variance (Garrick et al., 2009). Moreover, use of DEBV is needed to generate a higher reliability of genomic breeding values than EBV and are widely accepted in genomic studies (Ostersen et al., 2011). With single marker and single trait association between DEBV and SNP for each trait (Figure 1), we identified 26 suggestive SNP ($P < 10^{-4}$) associated with at least 1 of the 6 milk production traits (Supplemental Table S2; https://doi.org/10.3168/jds.2017-13246). After FDR adjustment, 4 SNP (AX-85148558, FDR = 0.006; AX-85106096, FDR = 0.006; AX-85073877, FDR = 0.08; AX-85063131, FDR = 0.09) were significantly associated with buffalo milk fat yield and protein percentage. According to the map information, the 4 SNP were located on BTA3 and BTA14, homologous to buffalo chromosomes 6 (BBU6) and 15 (BBU15), respectively (El Nahas et al., 2001).

For the significant SNP, we calculated the LSM of EBV for the 3 genotypes affecting the trait to investigate their genetic contribution shown in Figure 2. Three loci, AX-85148558, AX-85106096, and AX-85073877, were associated with PP270, and the individuals with CC or GG homozygous genotypes showed higher ($P < 0.01$ or $P < 0.05$) protein percentage among all 3 genotypes. Hence, we suggest that buffaloes with the C allele at AX-85148558 and AX-85073877 and the G allele at AX-85106096 could be selected to improve protein percentage in buffalo milk. Meanwhile, AX-85063131, affecting FY270, showed that the individuals with the GG genotype had higher ($P < 0.01$ or $P < 0.05$) fat yield compared with animals with AA or AG genotypes. Therefore, the G allele at the AX-85063131 locus could be selected as a favorable allele to improve milk fat yield in this buffalo-breeding program. The low frequency of the favorable allele provides more opportunity for improvement. However, it is also possible that the significant association was due to the small number of observations for the favorable allele. Compared with dairy cattle, dairy buffalo selection breeding programs started only recently. The lack of selection pressure in the buffalo industry and different buffalo breeds may also contribute to the low frequency of the favorable allele and genotype. We therefore recommend verifying the favorable allele effect for milk production traits in a larger sample size or by testing them in an independent buffalo population.

In previous studies, several suggestive SNP ($P < 10^{-4}$) were identified and showed associations with milk production traits in Mediterranean buffalo (9 SNP; Iamartino et al., 2013), Murrah buffalo (22 SNP; de Camargo et al., 2015), and Egyptian buffalo (8 SNP; El-Halawany et al., 2017) populations. Unfortunately, overlapping SNP were not found among the 4 GWAS studies. The most likely reasons are the differences of genetic composition of breeds, different selection pressure in the populations, varying environmental conditions, inbreeding, effective population size, and allelic frequencies (El-Halawany et al., 2017). In addition, buffalo SNP arrays are not specific for different species of buffalo. Furthermore, we adjusted all milk production traits to 270 d to account for different lactation lengths as a selection criterion (Rosati and Van Vleck, 2002; Tonhati et al., 2008) and used multiple testing (Genovese et al., 2002) to control the FDR of significant SNP to make them more conservative.

Candidate Genomic Regions

To further detect QTL associated with milk production traits, we used a 0.5-Mb window around 4 significant SNP to determine the linkage disequilibrium (**LD**) relationships (Figure 3). Three identified SNP AX-85148558, AX-85106096, and AX-85073877 for PP270 were located on BTA3 at 43 Mb, and within the detected window, from 43,084,940 to 43,795,880 bp, 2 haplotype blocks were recognized. The first block (**3_block1**), spanning 149 kb, involved 5 SNP, and the second block (**3_block2**), spanning 14 kb, consisted of only 2 SNP. Moreover, 3 SNP with almost complete LD (0.97–1.00) were harbored in 3_block1. For AX-85063131, which had an effect on FY270, the region from 66,462,566 to 66,962,566 bp of BTA14 was identified, and 2 adjacent haplotype blocks were built among 13 SNP. The SNP AX-85063131 showed strong LD

(0.96–1.00) with another 4 SNP to construct the first haplotype block on BTA14 (**14_block1**). The second block (**14_block2**), with a length of 219 kb, covered 8 SNP based on the 90K buffalo SNP panel.

For each block, we performed association between each haplotype and 6 buffalo milk production traits (Table 3). Compared with the single-marker association, analysis of haplotype association is more powerful for detecting significant effects on traits (Yang et al., 2015). Hence, 2 blocks within the 3_43 region were shown to be associated with buffalo milk protein percentage ($P = 0.001$ and 0.003) as well as fat percentage ($P = 0.001$ and 0.03). Two adjacent blocks on the 14_66 region were found to have an effect on fat yield ($P = 0.002$ and 0.0006), protein yield ($P = 0.02$ and 0.0004), and total milk yield ($P = 0.007$ and 0.0005) in buffalo. The results are consistent with the positive genetic correlations among these traits. According to
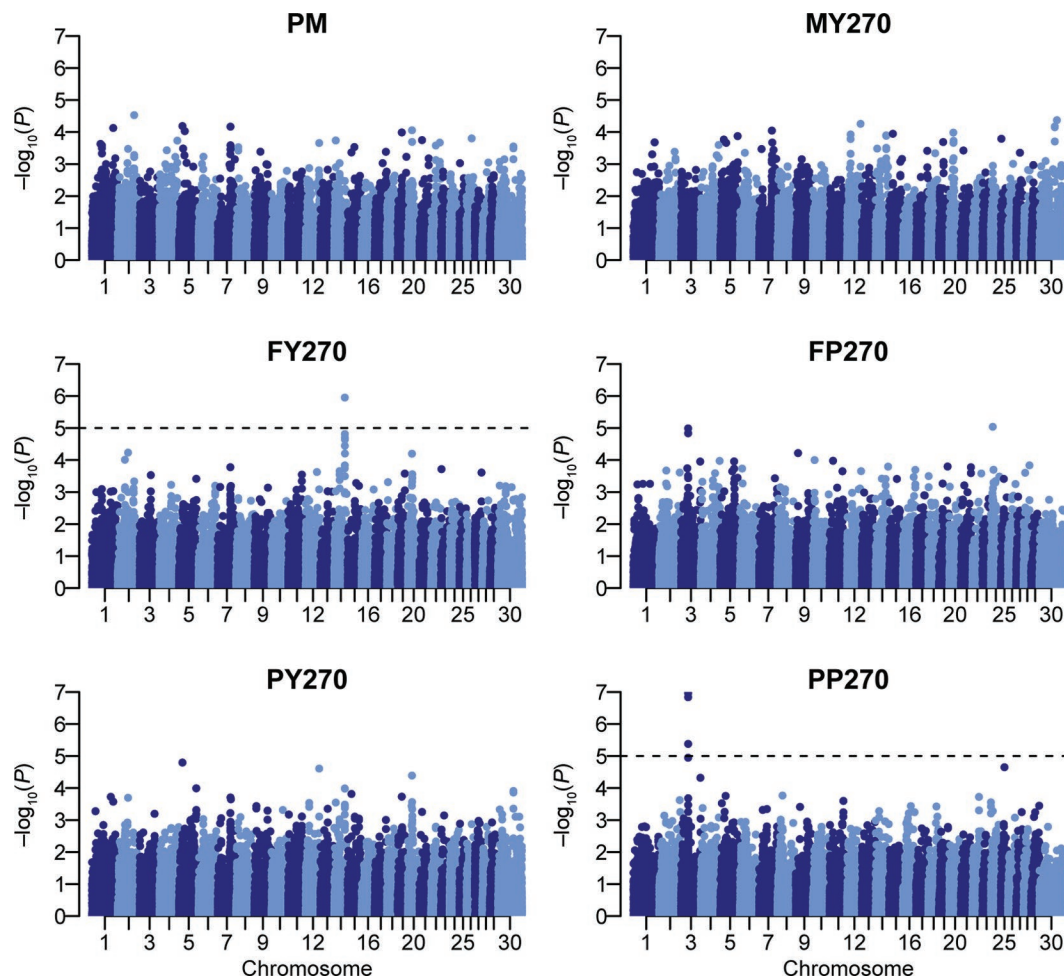

Figure 1. Manhattan plots for genome-wide association studies of 6 milk production traits in water buffalo: PM = peak milk yield; MY270 = 270-d total milk yield; FY270 = 270-d fat yield; FP270 = 270-d fat percentage; PY270 = 270-d protein yield; PP270 = 270-d protein percentage. On the y-axis are $-\log_{10}$ (P-values) and the horizontal line represents $P = 1.0 \times 10^{-5}$. On the x-axis are the physical positions of SNP by chromosome based on the *Bos taurus* UMD3.1 genome assembly (http://bovinegenome.org/?q=node/61). Color version available online.

GWAS and haplotype analyses, we identified 2 regions at 43.3–43.8 Mb on BTA3 (**3_43**) and 66.5–67.0 Mb on BTA14 (**14_66**) as QTL affecting milk production traits of water buffalo. Interestingly, the region on BTA3 at 43.29 Mb was also identified (Hu et al., 216) as a bovine QTL affecting milk fat yield, fat percentage, protein yield, and protein percentage. From the BovineMine database (Elsik et al., 2016), we discovered 2 milk protein yield–associated QTL on BTA14 at 66.5 and 66.9 Mb, respectively, which overlapped with the 14_66 region in the present study. In addition, previous studies have demonstrated that the region of 66.02–66.15 Mb on BTA14 was linked to a QTL mainly affecting milk fat yield in dairy cattle (Harder et al., 2006; Wibowo et al., 2008). de Camargo et al. (2015) detected 1 SNP (AX-85154407) on BTA14 close to the 14_66 region associated with buffalo milk production in Murrah buffalo. From GWAS analyses, El-Halawany et al. (2017) discovered several suggestive genomic regions on BTA1, 5, 6, and 27 affecting daily milk yield in 94 Egyptian buffaloes, which have also been recognized in some dairy cattle breeds. The identification of same genomic regions in both cattle and buffalo using different experimental designs and analysis methods increase the confidence that 3_43 and 14_66 could be promising QTL associated with milk performance.

Candidate Genes Within Identified Genomic Regions

Within the candidate region, we discovered a total of 13 genes and 1 small nucleolar RNA, as listed in Table 4. Among the genes, *MFSD14A*, *SLC35A3*, and *PALMD* in the 3_43 region, and *RGS22* and *VPS13B* in 14_66 region, were considered the most interesting candidate genes for buffalo milk production traits.

On the 3_44 region, *MFSD14A* was mapped to 43,332,801–43,386,388 bp on BTA3; *MFSD14A* is a protein-coding gene with product homology to the solute carrier protein family (SLC; Sreedharan et al., 2011; Doran et al., 2016). Gene ontology analysis showed that *MFSD14A* plays roles in the biological process of transmembrane transport (GO:0055085), molecular function of transporter activity (GO:0005215) and cellular components including GO:0016020 and GO:0016021. Whitworth et al. (2011) found that *MFSD14A* downregulated transcripts of the nuclear transfer in pig blastocyst stage, and Doran et al. (2016) reported that the gene might be involved in mice spermatogenesis. Although the functional study of *MFSD14A* is limited, we identified 2 significant SNP (AX-85106096 and AX-85148558) located in the gene, which implied *MFSD14A* may be a novel candidate gene and had genetic effects on buffalo milk fat yield.
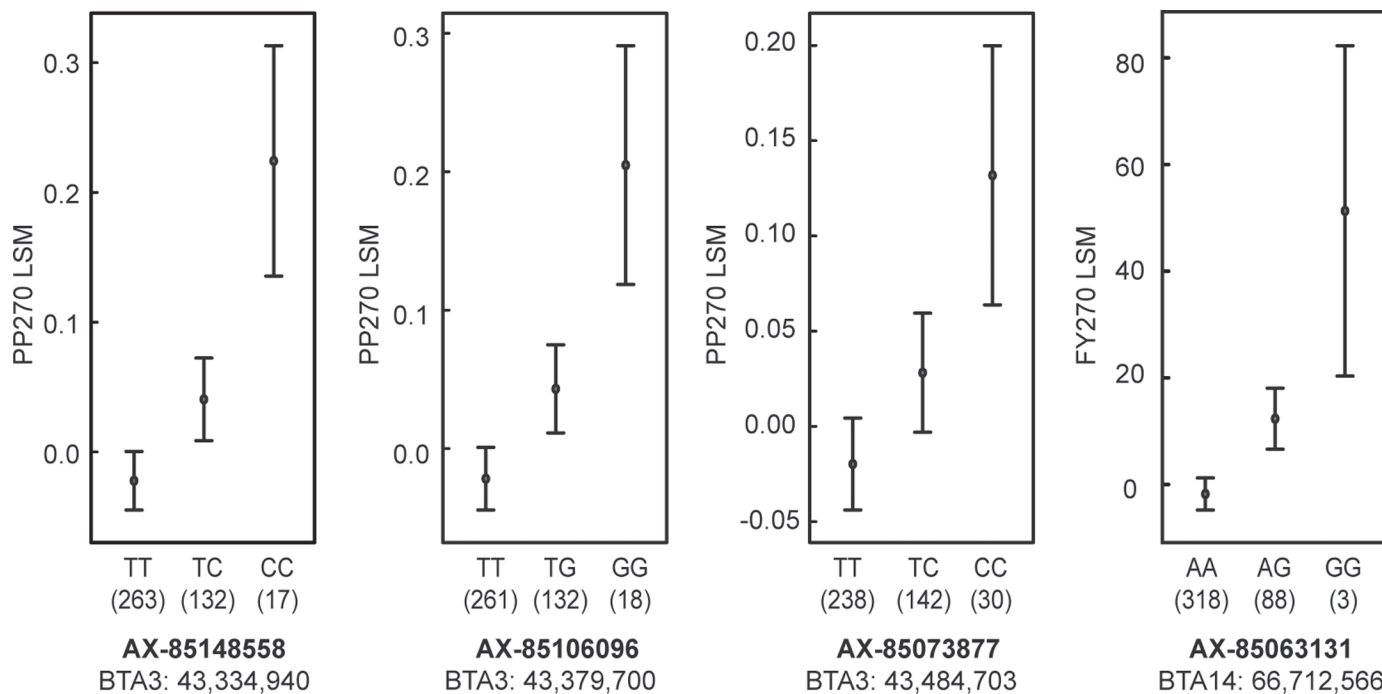

Figure 2. Least squares means (and 95% CI) of EBV for the 3 genotypes affecting the trait of significant SNP (false discovery rate <0.10) detected from genome-wide association studies. PP270 = 270-d protein percentage; FY270 = 270-d fat yield. Individuals with different genotypes are shown for each SNP showing significant (*P* < 0.05 or *P* < 0.01) association with PP270 or FY270. Numbers under each genotype represent the total number of animals with that genotype; the SNP position under each SNP is based on the *Bos taurus* UMD 3.1 genome assembly (http://bovinegenome.org/?q=node/61).
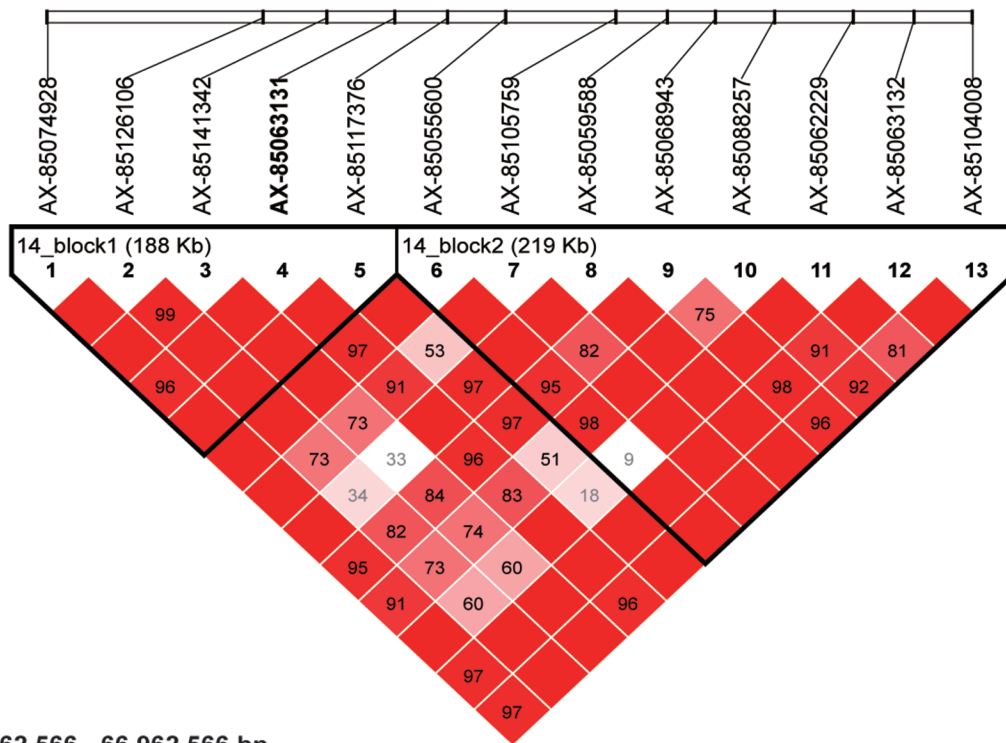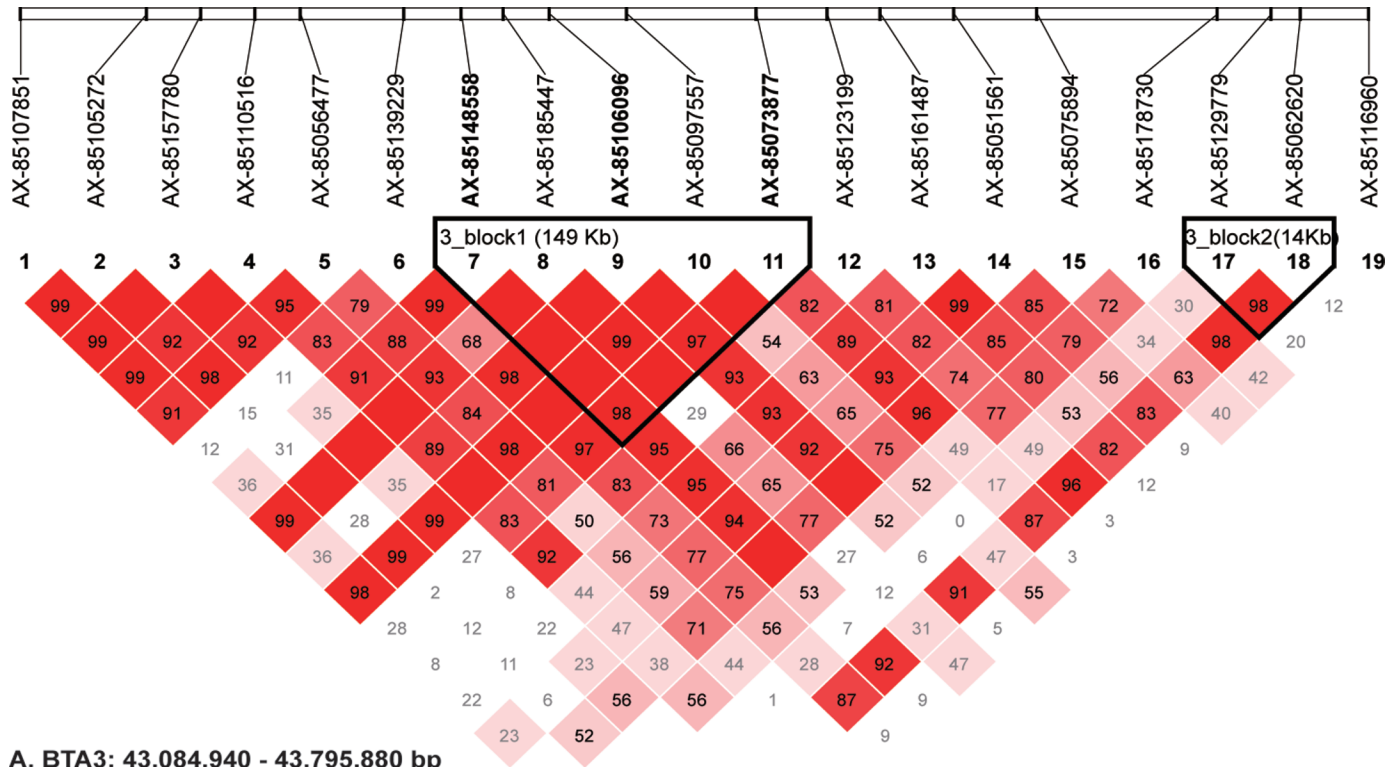
Figure 3. Haplotype block patterns for the significant SNP (false discovery rate <0.10) based on linkage disequilibrium (LD) within detected regions. The numbers on the top indicate the SNP order in the region; SNP in bold are significant SNP detected from genome-wide association studies. The SNP grouped in each triangle box mean they are grouped in one block based on LD (squared correlation coefficient, r^2). Color version available online.

Table 3. Haplotype association analyses for 6 milk production traits in water buffalo

Block (BTA_block)	Haplotype	Frequency[1] (%)	PM	MY270	FY270	FP270	PY270	PP270
3_block1	TGTGT	41.94	0.69	0.21	0.36	0.0001***	0.56	0.0001***
	TATAT	30.28						
	CGGGC	19.85						
3_block2	CG	59.82	0.33	0.05	0.27	0.03*	0.05	0.0026***
	AA	21.96						
	CA	18.09						
14_block1	GGACT	40.28	0.21	0.0071**	0.0022**	0.69	0.02*	0.57
	GAGAC	18.62						
	AAGAT	18.00						
	GAGAT	10.41						
14_block2	AGATGAAC	27.85	0.04*	0.0005***	0.0006***	0.99	0.0004***	0.73
	GGATGAAC	27.59						
	GGGTAAGC	21.16						
	GGATGAAT	7.55						

[1]Frequency of individuals of each haplotype among population.

[2]PM = peak milk yield; MY270 = 270-d total milk yield; FY270 = 270-d fat yield; FP270 = 270-d fat percentage; PY270 = 270-d protein yield; PP270 = 270-d protein percentage.

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$: significant association after Bonferroni multiple test.

The *SLC35A3* gene has been identified as the causative gene of vertebral malformation (Thomsen et al., 2006), which was reported to negatively affect milk production in dairy cattle (Kadri et al., 2014). Chu et al. (2010) studied association analyses between dairy performance traits and polymorphism of *SLC35A3* and provided evidence that *SLC35A3* could influence milk performance. The *SLC35A3* gene encodes a protein that participates in molecular pathways involved mainly in transport of nucleotide sugars (R-HSA-727802) and transport of vitamins, nucleosides, and related molecules (R-BTA-425397). The GO annotations related to this gene including sugar:proton symporter activity (GO:0005351) and UDP-*N*-acetylglucosamine trans-membrane transporter activity (GO:0015788). Hence, *SLC35A3* should be an interesting candidate gene for milk performance.

The palmdelphin (*PALMD*) gene, also called *HIAT1*, codes a cytosolic protein implicated in p53 phosphorylation (Moioli et al., 2013a). The *PALMD* gene is thought to be involved in the variation of milk yield. It is differentially expressed between high-yielding and low-yielding Holsten cattle (Seo et al., 2016) and Moioli et al. (2013b) showed that a mutation in exon 5 (c.547 T>A) of *PALMD* caused a reduction in the milk-producing ability of sheep. In a study to identify potential causal mutations related to pig production traits, *PALMD* polymorphisms had significant effects

Table 4. Details of candidate genes within the 3_43 and 14_66 genomic regions

Region[1]	Within gene[2]	Gene start (bp)	Gene end (bp)	Description
3_43	*RTCA*	43,157,188	43,185,261	RNA 3′-terminal phosphate cyclase
	DBT	43,189,905	43,230,271	Dihydrolipoamide branched chain transacylase E2
	LRRC39	43,251,251	43,272,503	Leucine rich repeat containing 39
	TRMT13	43,270,874	43,291,189	tRNA methyltransferase 13 homolog
	SASS6	43,291,333	43,331,499	Spindle assembly 6 homolog
	MFSD14A	43,332,801	43,386,388	Major facilitator superfamily domain containing 14A
	SLC35A3	43,400,346	43,444,844	Solute carrier family 35 member A3
	AGL	43,504,601	43,585,149	Amylo-α-1, 6-glucosidase, 4-α-glucanotransferase
	FRRS1	43,619,570	43,629,406	Ferric chelate reductase 1
	PALMD	43,688,103	43,748,131	Palmdelphin
14_66	***RGS22***	66,472,379	66,593,433	Regulator of G-protein signaling 22
	COX6C	66,637,801	66,647,721	Cytochrome c oxidase subunit 6C
	VPS13B	66,648,395	67,461,111	Vacuolar protein sorting 13 homolog B
	SNORA70	66,777,378	66,777,512	Small nucleolar RNA

[1]Regions: 3_43, the region on BTA3 spanning 43.3 to 43.8 Mb; 14_66, the region on BTA14 spanning 66.02 to 66.15 Mb.

[2]Within genes, genes within candidate region are based on *Bos taurus* UMD 3.1 genome assembly (http://bovinegenome.org/?q=node/61); those in bold are the most interesting candidate genes based on their effect on milk production performance in dairy cattle or identified as harboring significant SNP associated with milk production traits of buffalo.

on back fat thickness (Martínez-Montes et al., 2017). These functional studies of *PALMD* indicate it could be an important target gene with an important role in the regulation of milk production.

On the 14_66 region, *RGS22* was identified as one of the most interesting genes because it was previously reported as a candidate gene influencing protein yield in Holstein cattle (Marques et al., 2011). From the GeneCards database, we discovered that *RGS22* was more overexpressed in breast (42.6) than in other tissues. Similarly, some functional genes controlling of milk properties, likes *DGAT1* (Cases et al., 1998), *ABCG2* (Bionaz and Loor, 2008), and *SCD1* (Lengi and Corl, 2007), also show high expression patterns in lactating mammary tissue. Among its related pathways are signaling by GPCR (R-HSA-372790) and peptide ligand-binding receptors (R-HSA-418594). The GO annotations related to this gene include GTPase activator activity (GO:0005096).

The *VPS13B* gene on BTA14 was located at 66,648,395 to 67,461,111 bp. One significant SNP (AX-85063131) and 4 suggestive SNP (AX-85063132, AX-85104008, AX-85104009, and AX-85079883) detected for FY270 were harbored in this gene. The *VPS13B* gene encodes a potential transmembrane protein, which plays a role in the development and function of the eye, hematological system, and central nervous system. The main biological process described for *VPS13B* is protein transport (GO:0015031), and it may function in vesicle-mediated transport and protein sorting within cells. In a previous study, *VPS13B* was detected within a QTL associated with leg morphology in dairy cattle (van den Berg et al., 2014). In the review by Capitan et al. (2014), it was suggested that the genomic region on BTA14 covering *RGS22*, *COX6C*, and *VPS13B* was associated with female fertility as well as milk production traits in Holstein cattle. Accordingly, all findings indicate that the 14_66 genomic region may be an important QTL, with *RGS22* as a key candidate gene related to milk production or milk content in dairy buffalo.

Genomic Prediction

For the identified regions and markers affecting buffalo milk performance, it is essential to verify that the association can be generalized to other populations. Hence, we used a 5-fold cross validation to evaluate the reliability of GEBV based on genotype data from validation sets (Table 5). The average reliability of GEBV of 6 buffalo milk production traits ranged from 0.06 (PY270) to 0.22 (PP270). The accuracy of GEBV was also assessed by the strength of correlation between GEBV and EBV (Gondro et al., 2013), which varied

from 0.23 to 0.35 in the present validation set. The observed reliabilities of GEBV in our population were at a similar level to those of Nordic Red (0.19–0.23) using a single-country reference population (Brøndum et al., 2011). The correlations between GEBV and published EBV ranged from 0.25 to 0.70 in Holstein, which were higher than those in our study (Su et al., 2010). The number of available phenotypes, pedigree information, and genotypes in training sets contribute to the accuracy of genomic prediction (Pryce et al., 2012). Generally, reliability is improved by including a greater number of SNP and increasing the reference population size as reported in dairy (Goddard and Hayes, 2009) and beef (Brito et al., 2011). The heritability of the trait is another factor that influences genomic prediction accuracy (Hayes et al., 2009, Brito et al., 2011). Indeed, traits with higher heritability had higher reliability of GEBV in our prediction. The best accuracy was observed for PP270, which had the highest heritability (0.38) among the studied traits. In addition, the method used to calculate the reliability of GEBV also influenced results. Therefore, a more precise buffalo genomic map, a larger sample size, and higher SNP density are required to improve the detection power and prediction accuracy.

CONCLUSIONS

The aim of the present study was to investigate important genomic regions and genes associated with milk production traits in water buffalo. Genomic prediction was conducted to explore the feasibility of genomic selection as a potential strategy in buffalo breeding programs. Four SNP (AX-85148558, AX-85106096, AX-85073877, and AX-85063131) and 2 genomic region (3_43 and 14_66) were found to be associated with buffalo milk performance. Five genes, including *MFSD14A*, *SLC35A3*, *PALMD*, *RGS22*, and *VPS13B*, were identified as novel candidate genes for milk production in buffaloes. The reliability of genomic prediction for all traits ranged from 0.06 to 0.22. These findings pro-

Table 5. The reliability of genomic EBV (GEBV) and correlation between EBV and GEBV for the validation sets (mean ± SEM)

Trait[1]	Reliability	Correlation
PM	0.14 ± 0.003	0.23 ± 0.05
MY270	0.08 ± 0.002	0.26 ± 0.03
FY270	0.12 ± 0.003	0.30 ± 0.03
FP270	0.19 ± 0.004	0.32 ± 0.04
PY270	0.06 ± 0.002	0.27 ± 0.03
PP270	0.22 ± 0.004	0.35 ± 0.03

[1]PM = peak milk yield; MY270 = 270-d total milk yield; FY270 = 270-d fat yield; FP270 = 270-d fat percentage; PY270 = 270-d protein yield; PP270 = 270-d protein percentage.

vide useful preliminary information for buffalo breeding programs to adopt genomic selection. However, a more precise buffalo genomic map is required to identify the genes affecting the traits, and a larger sample size may help to improve the detection power and prediction accuracy. Furthermore, studies are needed to evaluate other interesting milk traits, such as coagulation ability and titratable acidity, which are important technological traits in producing cheese.

ACKNOWLEDGMENTS

REFERENCES

Amaral, M. E. J., J. R. Grant, P. K. Riggs, N. B. Stafuzza, A. Edson Filho, T. Goldammer, R. Weikard, R. M. Brunner, K. J. Kochan, and A. J. Greco. 2008. A first generation whole genome RH map of the river buffalo with comparison to domestic cattle. BMC Genomics 9:631.

Baldi, F., M. M. M. Laureano, D. G. M. Gordo, A. B. Bignardi, R. R. A. Borquis, L. G. d. Albuquerque, and H. Tonhati. 2011. Effect of lactation length adjustment procedures on genetic parameter estimates for buffalo milk yield. Genet. Mol. Biol. 34:62–67.

Barrett, J. C., B. Fry, J. Maller, and M. J. Daly. 2005. Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265.

Bionaz, M., and J. J. Loor. 2008. Gene networks driving bovine milk fat synthesis during the lactation cycle. BMC Genomics 9:366.

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

Brøndum, R. F., E. Rius-Vilarrasa, I. Strandén, G. Su, B. Guldbrandtsen, W. Fikse, and M. S. Lund. 2011. Reliabilities of genomic prediction using combined reference data of the Nordic Red dairy cattle populations. J. Dairy Sci. 94:4700–4707.

Capitan, A., P. Michot, A. Baur, R. Saintilan, C. Hoze, D. Valour, F. Guillaume, D. Boichon, A. Barbat, and D. Boichard. 2014. Genetic tools to improve reproduction traits in dairy cattle. Reprod. Fertil. Dev. 27:14–21.

Cases, S., S. J. Smith, Y.-W. Zheng, H. M. Myers, S. R. Lear, E. Sande, S. Novak, C. Collins, C. B. Welch, and A. J. Lusis. 1998. Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. Proc. Natl. Acad. Sci. USA 95:13018–13023.

Chu, Q., Y. Zhang, D. Sun, Y. Yu, and Y. Wang. 2010. Identification of the complex vertebral malformation gene in Chinese Holstein and its association with dairy performance traits. Hereditas 32:732–736.

de Camargo, G. M., R. Aspilcueta-Borquis, M. Fortes, R. Porto-Neto, D. Cardoso, D. Santos, S. Lehnert, A. Reverter, S. Moore, and H. Tonhati. 2015. Prospecting major genes in dairy buffaloes. BMC Genomics 16:872.

Di Meo, G. P., A. Perucatti, S. Floriot, H. Hayes, L. Schibler, D. Incarnato, D. Di Berardino, J. Williams, E. Cribiu, and A. Eggen. 2008. An extended river buffalo (*Bubalus bubalis*, 2n= 50) cytogenetic map: assignment of 68 autosomal loci by FISH-mapping and R-banding and comparison with human chromosomes. Chromosome Res. 16:827–837.

Do, D. N., L. L. G. Janss, J. Jensen, and H. N. Kadarmideen. 2015. SNP annotation-based whole genomic prediction and selection: An application to feed efficiency and its component traits in pigs. J. Anim. Sci. 93:2056–2063.

Doran, J., C. Walters, V. Kyle, P. Wooding, R. Hammett-Burke, and W. H. Colledge. 2016. Mfsd14a (Hiat1) gene disruption causes globozoospermia and infertility in male mice. Reproduction 152:91–99.

El-Halawany, N., H. Abdel-Shafy, A. S. Abd-El-Monsif, M. A. Abdel-Latif, A. F. Al-Tohamy, and O. M. A. El-Moneim. 2017. Genome-wide association study for milk production in Egyptian buffalo. Livest. Sci. 198:10–16.

El-Halawany, N., H. Abdel-Shafy, A. E. M. A. Shawky, M. A. Abdel-Latif, A. F. M. Al-Tohamy, and O. M. A. El-Moneim. 2015. Genome-wide association study for milk production in Egyptian buffalo. Pages 10–16 in Proc. Int. Symp. Animal Functional Genomics. The 6th International Symposium on Animal Functional Genomics, Piacenza, Italy.

El Nahas, S. M., H. A. de Hondt, and J. E. Womack. 2001. Current status of the river buffalo (*Bubalus bubalis* L.) gene map. J. Hered. 92:221–225.

Elsik, C. G., D. R. Unni, C. M. Diesh, A. Tayal, M. L. Emery, H. N. Nguyen, and D. E. Hagen. 2016. Bovine Genome Database: New tools for gleaning function from the *Bos taurus* genome. Nucleic Acids Res. 44(D1):D834–D839.

Endelman, J. B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4:250–255.

FAOSTAT. 2014. Food and Agriculture Organization the United Nations Statistics Division. http://faostat3.fao.org/compare/E.

Fulton, J. E. 2012. Genomic selection for poultry breeding. Anim. Front. 2:30–36.

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.

Genovese, C. R., N. A. Lazar, and T. Nichols. 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15:870–878.

Gilmour, A. R., B. Gogel, B. Cullis, R. Thompson, and D. Butler. 2009. ASReml user guide release 3.0. VSN International Ltd., Hemel Hempstead, UK.

Goddard, M. E., and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10:381–391.

Gondro, C., J. Van der Werf, and B. Hayes. 2013. Genome-wide association studies and genomic prediction. Springer, Amsterdam, the Netherlands.

Harder, B., J. Bennewitz, N. Reinsch, G. Thaller, H. Thomsen, C. Kühn, M. Schwerin, G. Erhardt, M. Förster, and F. Reinhardt. 2006. Mapping of quantitative trait loci for lactation persistency traits in German Holstein dairy cattle. J. Anim. Breed. Genet. 123:89–96.

Hayes, B. J., P. Bowman, A. Chamberlain, and M. Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92:433–443.

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

Iamartino, D., J. L. Williams, T. Sonstegard, J. Reecy, C. v. Tassell, E. L. Nicolazzi, S. Biffani, F. Biscarini, S. Schroeder, and D. A. de Oliveira. 2013. The buffalo genome and the application of genomics in animal management and improvement. Pages 151–158 in Proc. Buffalo Bulletin. International Buffalo Information Center, Bangkok, Thailand.

Kadri, N. K., G. Sahana, C. Charlier, T. Iso-Touru, B. Guldbrandtsen, L. Karim, U. S. Nielsen, F. Panitz, G. P. Aamand, and N. Schulman. 2014. A 660-Kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic Red cattle: Additional evidence for the common occurrence of balancing selection in livestock. PLoS Genet. 10:e1004049.

Khan, M., A. Hyder, I. Bajwa, M. Rehman, and F. Hassan. 2005. Prediction of lactation yield from last-record day and average daily yield in Nili-Ravi buffaloes. Pak. Vet. J. 25:175.

Khedkar, C., S. Kalyankar, and S. Deosarkar. 2016. Buffalo milk. Pages 522–528 in Encyclopedia of Food and Health. B. Caballero, P. M. Finglas, and F. Toldrá, ed. Academic Press, New York, NY.

Lengi, A. J., and B. A. Corl. 2007. Identification and characterization of a novel bovine stearoyl-CoA desaturase isoform with homology to human SCD5. Lipids 42:499–508.

Lenth, R. V. 2016. Least squares means: The R package lsmeans. J. Stat. Softw. 69:1–33.

Lukić, B., R. Pong-Wong, S. Rowe, D. Koning, I. Velander, C. Haley, A. Archibald, and J. Woolliams. 2015. Efficiency of genomic prediction for boar taint reduction in Danish Landrace pigs. Anim. Genet. 46:607–616.

Marques, E., J. Grant, Z. Wang, D. Kolbehdari, P. Stothard, G. Plastow, and S. Moore. 2011. Identification of candidate markers on bovine chromosome 14 (BTA14) under milk production trait quantitative trait loci in Holstein. J. Anim. Breed. Genet. 128:305–313.

Martínez-Montes, A. M., A. Fernandez, D. Pérez-Montarelo, E. Alves, R. Benitez, Y. Nunez, C. Ovilo, N. Ibañez-Escriche, J. Folch, and A. Fernandez. 2017. Using RNA-Seq SNP data to reveal potential causal mutations related to pig production traits and RNA editing. Anim. Genet. 48:151–165.

Miar, Y., G. Plastow, S. Moore, G. Manafiazar, P. Charagu, R. Kemp, B. Van Haandel, A. Huisman, C. Zhang, and R. McKay. 2014. Genetic and phenotypic parameters for carcass and meat quality traits in commercial crossbred pigs. J. Anim. Sci. 92:2869–2884.

Michelizzi, V. N., M. V. Dodson, Z. Pan, M. E. J. Amaral, J. J. Michal, D. J. McLean, J. E. Womack, and Z. Jiang. 2010. Water buffalo genome science comes of age. Int. J. Biol. Sci. 6:333–349.

Moioli, B., M. Scatà, G. Catillo, F. Napolitano, and F. Pilla. 2013a. Whole-genome genotyping for detecting mutations in the genes that are directly responsible for the trait variation. Page P0621 in Proc. Plant and Animal Genome XXI, San Diego, CA.

Moioli, B., M. C. Scatà, R. Steri, F. Napolitano, and G. Catillo. 2013b. Signatures of selection identify loci associated with milk yield in sheep. BMC Genet. 14:76.

Moser, G., B. Tier, R. E. Crump, M. S. Khatkar, and H. W. Raadsma. 2009. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. Genet. Sel. Evol. 41:56.

Ostersen, T., O. F. Christensen, M. Henryon, B. Nielsen, G. Su, and P. Madsen. 2011. Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in purebred pigs. Genet. Sel. Evol. 43:38.

Pryce, J. E., J. Arias, P. Bowman, S. Davis, K. Macdonald, G. Waghorn, W. Wales, Y. Williams, R. Spelman, and B. Hayes. 2012. Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. J. Dairy Sci. 95:2108–2119.

Purcell, S., and C. Chang. 2015. PLINK 1.9. https://www.cog-genomics.org/plink2.

Rosati, A., and L. D. Van Vleck. 2002. Estimation of genetic parameters for milk, fat, protein and mozzarella cheese production for the Italian river buffalo *Bubalus bubalis* population. Livest. Prod. Sci. 74:185–190.

Saatchi, M., R. D. Schnabel, J. F. Taylor, and D. J. Garrick. 2014. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. BMC Genomics 15:442.

Sargolzaei, M., F. S. Schenkel, and P. M. Vanraden. 2009. gebv: Genomic breeding value estimator for livestock. Technical report to the Dairy Cattle Breeding and Genetics Committee. University of Guelph.

Seo, M., H.-J. Lee, K. Kim, K. Caetano-Anolles, J. Y. Jeong, S. Park, Y. K. Oh, S. Cho, and H. Kim. 2016. Characterizing milk production related genes in Holstein using RNA-seq. Asian-australas. J. Anim. Sci. 29:343.

Shi, D. S., J. Wang, Y. Yang, F. H. Lu, X. P. Li, and Q. Y. Liu. 2012. DGAT1, GH, GHR, PRL and PRLR polymorphism in water buffalo (*Bubalus bubalis*). Reprod. Domest. Anim. 47:328–334.

Sreedharan, S., O. Stephansson, H. B. Schiöth, and R. Fredriksson. 2011. Long evolutionary conservation and considerable tissue specificity of several atypical solute carrier transporters. Gene 478:11–18.

Su, G., B. Guldbrandtsen, V. Gregersen, and M. Lund. 2010. Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population. J. Dairy Sci. 93:1175–1183.

Tantia, M. S., R. K. Vijh, B. P. Mishra, B. Mishra, S. B. Kumar, and M. Sodhi. 2006. DGAT1 and ABCG2 polymorphism in Indian cattle (*Bos indicus*) and buffalo (*Bubalus bubalis*) breeds. BMC Vet. Res. 2:32.

Thomsen, B., P. Horn, F. Panitz, E. Bendixen, A. H. Petersen, L.-E. Holm, V. H. Nielsen, J. S. Agerholm, J. Arnbjerg, and C. Bendixen. 2006. A missense mutation in the bovine SLC35A3 gene, encoding a UDP-N-acetylglucosamine transporter, causes complex vertebral malformation. Genome Res. 16:97–105.

Tonhati, H., M. F. Cerón-Muñoz, J. A. d. Oliveira, L. El Faro, A. L. F. Lima, and L. G. d. Albuquerque. 2008. Test-day milk yield as a selection criterion for dairy buffaloes (*Bubalus bubalis* Artiodactyla, Bovidae). Genet. Mol. Biol. 31:674–679.

Tonhati, H., M. Muñoz, J. Duarte, R. Reichert, J. Oliveira, and A. Lima. 2004. Estimates of correction factors for lactation length and genetic parameters for milk yield in buffaloes. Arq. Bras. Med. Vet. Zootec. 56:251–257.

van den Berg, I., S. Fritz, S. Rodriguez, D. Rocha, M. Boussaha, M. S. Lund, and D. Boichard. 2014. Concordance analysis for QTL detection in dairy cattle: A case study of leg morphology. Genet. Sel. Evol. 46:31.

Venturini, G. C., D. Cardoso, F. Baldi, A. Freitas, R. Aspilcueta-Borquis, D. Santos, and H. Tonhati. 2014. Association between single-nucleotide polymorphisms and milk production traits in buffalo. Genet. Mol. Res. 13:10256.

Warriach, H. M., D. M. Mcgill, R. D. Bush, P. C. Wynn, and K. R. Chohan. 2015. A review of recent developments in buffalo reproduction - a review. Asian-australas. J. Anim. Sci. 28:451–455.

Weller, J. I., and M. Ron. 2011. Invited review: quantitative trait nucleotide determination in the era of genomic selection. J. Dairy Sci. 94:1082–1090.

Whitworth, K. M., J. Zhao, L. D. Spate, R. Li, and R. S. Prather. 2011. Scriptaid corrects gene expression of a few aberrantly reprogrammed transcripts in nuclear transfer pig blastocyst stage embryos. Cell. Reprogram. 13:191–204.

Wibowo, T. A., C. T. Gaskins, R. C. Newberry, G. H. Thorgaard, J. J. Michal, and Z. Jiang. 2008. Genome assembly anchored QTL map of bovine chromosome 14. Int. J. Biol. Sci. 4:406.

Wu, J. J., L. J. Song, F. J. Wu, X. W. Liang, B. Z. Yang, D. C. Wathes, G. E. Pollott, Z. Cheng, D. S. Shi, and Q. Y. Liu. 2013. Investigation of transferability of BovineSNP50 BeadChip from cattle to water buffalo for genome wide association study. Mol. Biol. Rep. 40:743–750.

Yang, S.-H., X.-J. Bi, Y. Xie, C. Li, S.-L. Zhang, Q. Zhang, and D.-X. Sun. 2015. Validation of PDE9A gene identified in GWAS showing strong association with milk production traits in Chinese Holstein. Int. J. Mol. Sci. 16:26530–26542.