J. Dairy Sci. 101:1–12
https://doi.org/10.3168/jds.2017-13246

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

Received May 28, 2017.
Accepted September 13, 2017.
[1]Corresponding author: yangliguo2006@qq.com

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.