Genome-wide analysis reveals genetic diversity, linkage disequilibrium, and selection for milk production traits in Chinese buffalo breeds

X. R. Lu,^{1*} A. Q. Duan,^{1*} W. Q. Li,^{2*} H. Abdel-Shafy,³ H. E. Rushdi,³ S. S. Liang,¹ X. Y. Ma,¹ X. W. Liang,¹ and T. X. Deng^{1†}

¹Key Laboratory of Buffalo Genetics, Breeding and Reproduction Technology, Buffalo Research Institute, Chinese Academy of Agricultural Sciences, Nanning 530001, China

²BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China

³Department of Animal Production, Faculty of Agriculture, Cairo University, 12613 Giza, Egypt

ABSTRACT

The water buffalo is an important dual-purpose livestock that is widespread throughout central and southern China. However, there has been no characterization of the population genetics of Chinese buffalo. Using an Axiom buffalo genotyping array (Thermo Fisher Scientific, Wilmington, DE), we analyzed the genetic diversity, linkage disequilibrium pattern, and signature of selection in 176 Chinese buffaloes from 13 breeds. A total of 35,547 SNP passed quality control and were used for further analyses. Population genetic analysis revealed a clear separation between swamp and river types. Ten Chinese indigenous breeds were clustered into the swamp group, the Murrah and Nili-Ravi breeds were clustered into the river group, and the crossbred breed was closer to the river group. Genetic diversity analysis showed that the swamp group had a lower average expected heterozygosity. Linkage disequilibrium decay distance was much shorter in the swamp group compared with the river group, with an average square of correlation coefficient value of 0.2 of approximately 50 kb. Analysis of runs of homozygosity indicated extensive remote and recent inbreeding within swamp and river groups, respectively. Moreover, one genomic region under selection was detected between the river and swamp groups. Our findings contribute to our understanding of the characterization of population genetics in Chinese buffaloes, which in turn may be used in buffalo breeding programs.

Key words: buffalo, linkage disequilibrium, population analysis, selection signature

INTRODUCTION

The water buffalo (*Bubalus bubalis*), primarily raised for milk, meat, and draft power, is an essential part of the agricultural economy of many countries around the world. Buffaloes mainly consist of 2 types, river (*B. bubalis bubalis*; chromosome number $2n = 50$) and swamp (*B. bubalis carabanesis*; $2n = 48$), which differ in morphology and chromosomal karyotypes (Mishra et al., 2015). The river-type buffalo is classified as a dual-purpose animal with considerable potential for milk and meat production, whereas the swamp-type buffalo is essentially a draft animal with lower milk production.

The largest buffalo population is mainly confined to Asia, where China ranks third in the world's total production of buffalo milk, preceded only by India and Pakistan (FAO, 2019). Interestingly, the largest number of swamp buffaloes across the world is found in China because the Chinese indigenous buffalo breeds are mainly referred to as swamp type. Like in other Asian countries, Chinese swamp-type buffalo cows were traditionally crossed with river-type buffalo bulls of the Murrah and Nili-Ravi breeds to improve milk production performance. Although such crossbred animals are fertile, they may show lower reproductive performance compared with the swamp-type buffalo (Harisah et al., 1989; Degrandi et al., 2014). The crossbred animals produce considerably higher milk yield than the swamp-type buffaloes (~1,700 vs. ~650 kg/lactation). Even after practicing continuous crossbreeding for several decades, the milk yield of crossbred buffaloes is still lower than that of river-type buffalo breeds (~2,200 kg/lactation; Deng et al., 2019). Therefore, the dairy buffalo industry prioritizes the improvement of milk production in terms of quantity and quality together. The implementation of new strategies, such as investigating the genetic architecture of milk and meat production, could facilitate a better understanding of economically important traits. Subsequently, these strategies may be used in buffalo breeding programs.

Received July 30, 2019.

Accepted January 13, 2020.

*These authors contributed equally to this work.

†Corresponding author: dtx282000@163.com

Knowledge of genetic diversity within and among breeds and populations is crucial for their domestication, conservation, and management. In the past few decades, 2 independent domestication events of the river and swamp buffaloes were confirmed using Y-chromosomal (Yindee et al., 2010) and mitochondrial DNA (Lei et al., 2007) technologies. The river buffalo was domesticated in the western region of the Indian subcontinent around 6,300 yr before present (Kumar et al., 2007; Nagarajan et al., 2015), and the swamp buffalo was domesticated in the border between Indochina and southwestern China around 4,000 yr before present (Zhang et al., 2016; Colli et al., 2018). More important, robust geographic differentiation was found within swamp buffaloes (Zhang et al., 2016), and the southwestern buffalo populations in China had the highest genetic diversity compared with other domestication centers (central and southeastern China; Yue et al., 2013). Such domestication events, including natural and artificial selection, assist in the formation and stabilization of buffalo breed characteristics. With the acceleration of urbanization, however, the total swamp buffalo population in China has shown a rapidly declining trend in recent years, especially for some endangered swamp breeds such as the Haizi buffalo. It is well established that a reduction in the effective population size causes a decline in genetic diversity by increasing both genetic drift and inbreeding rate (Wu et al., 2014). Because genetic diversity underpins an animal population's resilience and persistence, it is highly important to investigate the genetic diversity of swamp buffalo breeds in China. These results play a vital role in the future of buffalo genetic improvement programs and support sustainability of the indigenous buffalo breeds in Chinese agriculture.

In the last decade, genome- and transcriptome-wide molecular markers have been identified using high-throughput sequencing (Blanca et al., 2012; Wheat, 2012; Bertolini et al., 2014). These markers can be used for animal breeding and genetic studies. The first SNP genotyping platform (Axiom buffalo genotyping array; 90K Affymetrix; Thermo Fisher Scientific, Wilmington, DE) designed explicitly for the river buffalo has recently been developed (Iamartino et al., 2017) and used for genome-wide analysis study in different buffalo breeds (de Camargo et al., 2015; El-Halawany et al., 2017; Liu et al., 2018). This SNP 90K array provides a viable choice for buffalo scientific research in areas such as molecular breeding, complex traits, conservation, and biodiversity.

Several methods have been developed to detect signatures of recent selection in domesticated animals (Biswas and Akey, 2006). These methods are based mainly on linkage disequilibrium (**LD**) metrics, spectra

of allele frequencies, and characteristics of haplotype structures in the studied populations (Helyar et al., 2011). Notably, identifying signatures of selection could provide insight into the genomic response to domestication and selection for production traits, which may help in the design of more efficient selection schemes (de Simoni Gouveia et al., 2014). During the last few years, selective sweep regions in domesticated animals associated with distinct traits have been identified, including milk production traits in Holstein cattle (Qanbari et al., 2010), reproductive traits in dairy cattle (Qanbari et al., 2011), feed efficiency in different breeds of beef and dairy cattle (Barendse et al., 2009), elongation of the back in pigs (Rubin et al., 2012), and lack of horns in sheep (Kijas et al., 2012). For the Azeri and Khuzestani buffalo breeds, Mokhber et al. (2018) identified 13 selective sweep regions that are potentially related to economically important traits using the genome-wide SNP data. To date, however, the selection footprints among Chinese buffalo breeds remain unexplored. The present study had 2 main aims. The first was to investigate the genetic diversity, LD extent, and genome-wide signals of selection in 13 indigenous buffalo breeds of the southern region of China using the first SNP genotyping platform specific for buffalo (SNP 90K array). The second was to explore the relationship between signatures of selection and associations with milk production traits between river and swamp buffalo breeds.

MATERIALS AND METHODS

Ethics Statement

All animal work, experimental protocols, and animal care were approved by the Animal Ethics Committee of the Buffalo Research Institute, Chinese Academy of Agricultural Sciences, Nanning, China (approval code GXBRI-06–2019).

Sampling and Genotyping

A total of 176 unrelated buffaloes representing 13 breeds were included in the present study (Table 1). Blood samples were collected from 102 animals as follows: 23 river-type buffaloes, 20 crossbred buffaloes from the Buffalo Research Institute, and 59 swamp buffaloes sampled from different villages in southwest China. In addition, publicly available genotypic data on 74 buffalo individuals from 5 Chinese swamp breeds, provided by Colli et al. (2018), were involved in this study. The geographical locations of all animals used in the present study are shown in Figure 1. Maptools and ggplot2 packages implemented in R software (R Core

Table 1. Genetic variability for the analyzed buffalo populations in the southern China region¹

Group and breed	Code	N	N _{SNP}	Heterozygosity ($\pm$ SD)		A _R
				Observed	Expected	
River		43	10,750	0.418 $\pm$ 0.01	0.422 $\pm$ 0.01	1.999
Murrah	MUR	12	10,156	0.410 $\pm$ 0.03	0.391 $\pm$ 0.02	1.952
Nili-Ravi	NIR	11	10,423	0.368 $\pm$ 0.04	0.356 $\pm$ 0.03	1.913
Crossbred	CRO	20	10,431	0.449 $\pm$ 0.02	0.445 $\pm$ 0.01	1.993
Swamp		133	10,819	0.344 $\pm$ 0.02	0.356 $\pm$ 0.02	1.950
Hunan	HUN	15	10,615	0.352 $\pm$ 0.04	0.352 $\pm$ 0.03	1.888
Ensi	ENS	15	10,605	0.338 $\pm$ 0.04	0.346 $\pm$ 0.03	1.871
Fuling	FUL	15	10,613	0.338 $\pm$ 0.04	0.345 $\pm$ 0.03	1.868
Yibin	YIB	15	10,513	0.339 $\pm$ 0.04	0.347 $\pm$ 0.03	1.874
Yangzhou	YAN	14	9,530	0.356 $\pm$ 0.04	0.348 $\pm$ 0.03	1.878
Dechang	DEC	12	10,434	0.357 $\pm$ 0.04	0.358 $\pm$ 0.02	1.916
Dehong	DEH	12	10,443	0.346 $\pm$ 0.04	0.353 $\pm$ 0.03	1.890
Fuzhong	FUZ	12	10,224	0.348 $\pm$ 0.04	0.350 $\pm$ 0.03	1.882
Xilin	XIL	11	10,437	0.333 $\pm$ 0.04	0.345 $\pm$ 0.03	1.866
Guizhou	GUI	12	10,445	0.330 $\pm$ 0.04	0.351 $\pm$ 0.03	1.880

¹N = sample size after quality control; N_{SNP} = the number of SNP with minor allele frequency >0.2 in the 10,821-SNP subset; A_R = allelic richness.

Team, 2018) were used to create the map with information provided by the National Geomatics Center of China (<http://www.ngcc.cn/>).

Genomic DNA for each blood sample was isolated using the TIANamp Blood DNA Kit [Tiangen Biotech (Beijing) Co. Ltd., Beijing, China]. The quality and quantity of the isolated DNA were evaluated using a NanoDrop2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE) and 1.5% gel electrophoresis, respectively. The SNP genotyping was performed using an Axiom buffalo genotyping array (Thermo Fisher Scientific). The SNP positions were assigned to the latest buffalo genome assembly (UOA_WB_1: GCA_003121395.1). For quality control, SNP with a minor allele frequency (**MAF**) ≤ 0.05 or a call rate $\leq 95\%$ were excluded; SNP with unknown or duplicate chromosomal positions were also eliminated. Likewise, individuals with a call rate $\leq 95\%$ were omitted. The filtering options were performed using PLINK 1.9 (Chang et al., 2015). After applying the filtering process, a total of 35,547 SNP remained and were used for subsequent analyses.

Population Genetics Analysis

Principal components analysis was performed using Eigensoft 7.2 software with the Tracy-Widom test (Patterson et al., 2006) and visualized by in-house R scripts. A free computational phylogenetics package of programs for inferring evolutionary trees and estimating genetic diversity was applied. Among them, PHYLogeny Inference Package (version 3.697) was conducted to calculate the distance matrix (Felsenstein, 2016). Moreover, Molecular Evolutionary Genetics Analysis

(MEGA7) was used to illustrate the neighbor-joining tree, as proposed by Kumar et al. (2016). Population structure among the analyzed buffaloes was determined using Admixture 1.3.0 software (Alexander et al., 2009) with the cluster number set from 2 to 4. Admixture plots were further visualized using Excel 2016 (Microsoft Corp., Redmond, WA).

Genetic Diversity

A common informative subset of 10,821 SNP with MAF ≥ 0.2 in both the river and swamp groups was used to estimate 3 measures (Ai et al., 2013) of genome-wide genetic diversity with the “hierfstat” R package (Goudet, 2005): observed heterozygosity (**Ho**), expected heterozygosity (**He**), and allelic richness (**A_R**). Pairwise genetic differentiation of fixation index (**F_{ST}**; Tills, 1977) among populations was estimated using the “adeigenet” R package (Jombart, 2008).

LD Decay Assay

Genome-wide LD decay (**LDD**) was calculated within the river and swamp groups. To make the statistics more meaningful, we randomly selected the same number ($n = 43$) of animals in each population. We further performed an analysis of LD and visualized the LDD for the river and swamp groups using PopLD-decay software (Zhang et al., 2019).

Identification of Runs of Homozygosity

To detect runs of homozygosity (**ROH**), SNP that deviated from Hardy-Weinberg proportion ($P <$

0.0001) were first excluded using PLINK 1.9 (Chang et al., 2015). In this regard, only 22,478 SNP were retained for ROH analysis (Peripolli et al., 2017). The ROH for river and swamp groups were detected using the sliding window-based method implemented in the detectRUNS package of R (Biscarini et al., 2018). In this issue, the following conditions were applied: (1) the size of the sliding window was 15 loci; (2) the threshold of overlapping windows of the same state (homozygous or heterozygous) to call an SNP in a run was set to 0.05; (3) the minimum number of SNP in a run was 15; (4) 1 heterozygous genotype and only 1 missing SNP per window were allowed; (5) the maximum gap between consecutive SNP to still be considered a potential run was 1 Mb; (6) the minimum length of the run was 1 Mb; (7) and the minimum SNP density was set to 1 SNP every 50 kb. The minimum ROH length of 1 Mb was set to exclude short and very common ROH. The program allows 1 heterozygous and 1 missing call per window, which was used to minimize the occasional genotyping error or missing genotype occurring in an otherwise unbroken homozygous segment that could result in the underestimation of ROH. The ROH-based

inbreeding (F_{ROH}) for river and swamp groups was calculated using the following equation:

$$F_{ROHj} = \sum_k \text{length}(\text{ROH}_k) / L,$$

where ROH_k = the k th ROH in individual j 's genome and L = the total length of the genome (or X chromosome).

Identification of Selection Signatures

Selection signature analysis on the contrast model was performed using hapFLK 1.4 software (Fariello et al., 2013). Briefly, we first performed LD pruning analysis and Admixture analysis to obtain the population IDs required for hapFLK analysis. These were done using PLINK1.9 (Chang et al., 2015) and Admixture 1.3.0 (Alexander et al., 2009), respectively. The main conditions for LD pruning analysis were as follows: a window size of 50 SNP, a step size of 5 bp, and a correlation coefficient (r) threshold of 0.5. In the hapFLK analysis, the number of clusters was set to 2 and the number of expectation-maximization iterations ($-nfit$) was speci-

Figure 1. Geographic locations of the analyzed buffalo breeds in the southern China region. CRO = crossbred; DEC = Dechang; DEH = Dehong; ENS = Ensi; FUL = Fuling; FUZ = Fuzhong; GUI = Guizhou; HUN = Hunan; MUR = Murrah; NIR = Nili-Ravi; XIL = Xilin; YAN = Yangzhou; YIB = Yibin.

fied as 20. Referring to a previous report by Yuan et al. (2017), the hapFLK values were adjusted (adj) using the following equation:

$$\text{hapFLK}_{\text{adj}} = (\text{hapFLK} - \text{mean hapFLK}) / \text{SD hapFLK},$$

where robust estimates of the mean and SD values of the hapFLK statistics were calculated using the “MASS” R package (Venables and Ripley, 2002). The *P*-values of $\text{hapFLK}_{\text{adj}}$ were computed using a chi-squared test with in-house R scripts. The threshold value was set at $-\log P = 3$ as previously suggested by Yuan et al. (2017). The Manhattan plot for the selective sweeps was visualized using the “qqman” R package (Turner, 2014).

Gene Annotation

We annotated the identified regions under significant selection pressure using the National Center for Biotechnology Information's Genome Data Viewer on the buffalo genome (UOA_WB_1; Low et al., 2019). Genes within a region spanning 50 kb upstream and downstream of the candidate selection signatures were annotated.

RESULTS

SNP Characteristics

After adopting the filtering options, 35,547 SNP remained. This subset of SNP covered 1.31 Gb, with an average distance of 36.8 kb between adjacent SNP (Supplemental Table S1; <https://doi.org/10.3168/jds.2019-17364>). The length of buffalo chromosomes in the current study ranged from 21.97 Mb on chromosome (Chr) 24 to 102.63 Mb on Chr1. Mean distances between adjacent SNP per chromosome ranged from 32.69 kb on Chr19 to 54.63 kb on ChrX (Supplemental Table S1). Because almost 20% of the current buffalo array was polymorphic in swamp buffalo, 10,821 informative SNP were used in both the river and swamp groups for further genetic diversity analyses. Using these markers, statistical information on the 176 buffalo along with genetic variability are summarized in Table 1.

Population Analysis

Principal components analysis showed a distinct separation between the river and swamp types (Figure 2A). The crossbred buffaloes originated from the swamp- and river-type buffaloes and were concordantly located

between them. A similar genetic relationship among analyzed breeds was also supported by phylogenetic analysis (Figure 2B) and population structure analysis (Figure 2C). As shown in Figure 2C, cluster number = 2 represented the most appropriate population number for the present data set, indicating differentiation between river and swamp types. This suggested that the studied buffalo breeds can be divided into 2 groups (river and swamp) for further analysis.

Genetic Diversity

Distribution of 5 MAF classes in the river and swamp groups is presented in Figure 3. Compared with the river group, swamp buffalo breeds had the highest proportion of SNP with lower MAF category (0, 0.1). The river buffalo breeds had a relatively high proportion of SNP with high MAF [mostly MAF values of (0.2, 0.3), (0.3, 0.4), and (0.4, 0.5)].

Genome-wide genetic diversity metrics within breeds were measured using the H_o , H_e , and A_R parameters (Table 1). Overall, buffaloes in the river group had higher genetic diversity than the swamp breeds as measured by the H_o (0.418 ± 0.01 vs. 0.344 ± 0.02 , respectively) and H_e (0.422 ± 0.01 vs. 0.352 ± 0.03 , respectively). The average heterozygosity estimates were highest for the crossbred group within the river group (H_o : 0.449 ± 0.02 ; H_e : 0.445 ± 0.01) and lowest for the Xilin breed within the swamp group (H_o : 0.333 ± 0.04 ; H_e : 0.345 ± 0.03). The allelic richness for the swamp group ($A_R = 1.950$) was lower than that for the river group ($A_R = 1.999$). Notably, within the swamp group, the Dechang breed was observed to have the highest A_R ($A_R = 1.916$), whereas the Xilin breed revealed the lowest A_R ($A_R = 1.866$).

Population differentiation estimates showed that the pairwise F_{ST} values ranged from 0.0076 (Fuzhong-Yangzhou) to 0.7535 (Fuzhong-Nili-Ravi) among buffalo breeds in the current investigation (Supplemental Table S2; <https://doi.org/10.3168/jds.2019-17364>). For the swamp group, Hunan was most closely related to Fuzhong ($F_{ST} = 0.0087$), Guizhou ($F_{ST} = 0.0092$), Yibin ($F_{ST} = 0.0100$), Ensi ($F_{ST} = 0.0128$), Fuling ($F_{ST} = 0.0129$), Yangzhou ($F_{ST} = 0.0148$), Dechang ($F_{ST} = 0.0189$), Xilin ($F_{ST} = 0.0229$), and Dehong ($F_{ST} = 0.0498$). Regarding the river group, crossbred individuals showed a closer relationship with the Murrah ($F_{ST} = 0.0542$) than the Nili-Ravi ($F_{ST} = 0.0724$) breeds.

LD and Autozygosity Segments

Overall estimated LD within each of the river and swamp groups was different in the present study (Fig-

Figure 2. Population analysis of 13 buffalo breeds in southern China. (A) Principal components analysis plots display the individual relationships of 176 buffaloes. PC = principal component. (B) Neighbor-joining representation of the pairwise Nei's D genetic distances among populations. (C) Population structure of 176 buffaloes inferred by model-based clustering using Admixture (Alexander et al., 2009). K = cluster number.

ure 4). Compared with the river group ($r^2 = 0.61$), the highest maximum value of average LD ($r^2 = 0.88$) was found within the swamp group. As expected, LD estimates between pairwise SNP generally declined with increasing physical distance. The LDD in the swamp group declined faster than within the river group with increasing distance. The average distance when LDD decreased to a value of 0.2 was approximately 15 kb for the swamp group and 50 kb for the river group.

To estimate recent inbreeding, we performed a genome-wide autozygosity analysis with ROH between the river and swamp groups. The result showed that buffaloes in the swamp group had higher overall levels of F_{ROH} than buffaloes in the river group (Figure 5A). Moreover, we also estimated the ROH distribution by the length between the river and swamp groups (Figure 5B). The results showed a difference in genetic diversity between the river and swamp groups. Generally, buffaloes in the swamp group presented a larger average fraction of ROH than did buffaloes in the river group. However, the swamp group had a lower fraction of the short ROH (0–2 Mb) and long ROH (>16 Mb) segments compared with the river group.

Signatures of Selection

The hapFLK analysis that accounted for the haplotype information and hierarchical structure (Fariello et al., 2013; Walugembe et al., 2019) was used to identify selection footprints on the contrast model (river vs. swamp). As a result, the genomic region between 70.81 and 71.63 Mb on Chr19 was identified (Figure 6). Table 2 summarizes the annotated information of outlier SNP within the contrast model. The length of the candidate region was approximately 0.82 Mb. A total of 15 SNP within the selected region was found, corresponding to 13 candidate genes.

DISCUSSION

A large variety of buffalo genetic resources are found in China. Most of them are swamp type. They are mainly distributed in 18 provinces in the central and south regions of the country. Each region has different types of buffaloes, which are mainly kept for draft. Therefore, they produce less milk compared with the river buffalo breeds. To improve milk production in

Chinese buffalo breeds, Murrah and Nili-Ravi buffaloes have been imported to China in the last century and intensively used for crossbreeding. The application of crossbreeding has remarkably improved milk production in the native breeds by around 30% and 115% more than that of local swamp buffaloes in the F_2 crosses of Murrah and Nili-Ravi, respectively (Yang et al., 2013). Consequently, the risk of losing part of the genetic pool or diversity possibly existed. Recently, the Axiom buffalo 90K genotyping array was developed to facilitate genome-wide studies of diversity or associations with economically important traits in buffalo species. Although this chip was mainly for river buffalo breeds, approximately 25% of the SNP were polymorphic when tested in the swamp buffalo (Iamartino et al., 2017). In this work, we performed population analyses of 13 buffalo breeds in southern China. Principal components analysis and phylogenetic analyses both showed that 2 distinct clusters formed without overlapping among the studied breeds. In this regard, 10 indigenous buffaloes were clustered into the swamp group; Murrah and Nili-Ravi breeds were grouped into the river group, whereas the crossbred buffalo was concordantly located between the 2 groups. Population structure analysis further showed that the crossbred buffaloes could be grouped into the river group. In this regard, we further investigated the genetic diversity, LDD, and ROH, signatures of selection. These analyses help in the understanding

of the genetic basis of milk production traits in dairy buffaloes, which promotes the development of the dairy buffalo industry.

Using the medium-density buffalo SNP chip data, we found that buffalo breeds in the swamp group had lower genetic diversity than buffalo breeds in the river group based on heterozygosity measures. Similar results have been reported by Colli et al. (2018). However, Chinese indigenous breeds (swamp breeds) sustain higher levels of genetic variability than river breeds. These results are supported by previous reports that used microsatellite markers (Barker et al., 1997; Zhang et al., 2007, 2008) and mitochondrial DNA (Kathiravan et al., 2011; Zhang et al., 2016). However, these differences can also be explained by the current buffalo 90K array, although it is optimized for use in 4 river buffalo breeds (Mediterranean, Murrah, Nili-Ravi, and Jaffarabadi) and has a low representation of swamp breeds (Iamartino et al., 2017). The disproportionate distribution of MAF between the river and swamp groups shows the SNP ascertainment bias. For example, swamp buffalo breeds had the highest proportion of SNP with low MAF compared with the river buffalo group. The relatively high proportion of SNP with high MAF in the Chinese crossbred buffalo can be attributed to the fact that these buffaloes were recently crossbred with a significantly high inheritance of Murrah and Nili-Ravi ancestry. Moreover, we observed differences in the genetic

Figure 3. Distributions of minor allele frequency (MAF) of genotype SNP in the river and swamp buffalo populations.

differentiation among the studied breeds. As expected, F_{ST} among the swamp breeds was lower than among the river and crossbred breeds, ranging from 0.0076 to 0.7535, suggesting a lack of differentiation among Chinese swamp breeds. This was lower than the value of 0.52 observed among the African buffalo (*Syncerus caffer*) breeds by Smits et al. (2013) but higher than the F_{ST} estimates reported in previous studies (Vijh et al., 2008; Zhang et al., 2011; Mishra et al., 2015). These differences may be caused by small sample sizes or differences in buffalo breeds, geographical distribution, methods, or marker density.

Analysis of the genome-wide LDD plays a vital role in the genome-wide analysis study mapping of loci associated with economically important traits in livestock animals. Our previous studies have demonstrated that the LD extent was different between purebred and crossbred buffalo populations, with purebreds having the highest levels of LDD (Deng et al., 2019). In the current study, the LDD declined rapidly with increasing physical distance between markers for the swamp group compared with the river group. This result suggested that the swamp buffalo breeds had high genetic diversity. Of note, LD extent across populations is much shorter in

Chinese swamp buffaloes than in river buffaloes. With a threshold of $r^2 = 0.2$, LD extends to only 15 kb across swamp buffaloes, which is in stark contrast to 50 kb across river buffaloes. This finding was also supported by previous studies (Deng et al., 2019). Moreover, it is well known that ROH could be used to assist with the interpretation of the inbreeding coefficient and give insights about population history (Purfield et al., 2012b; Bjelland et al., 2013); the short and long ROH can reflect remote and recent inbreeding activity, respectively (Ferenčaković et al., 2013). Our results showed that the swamp buffalo had a lower fraction in the short (0–2 Mb) and long (>16 Mb) ROH classes than did the river buffalo. This finding suggested that extensive remote and recent inbreeding activity was found within the swamp and river groups, respectively. However, some ascertainment bias may exist in the swamp buffalo ROH levels because a low amount of polymorphic markers may cause an overestimation of ROH amounts. Similar points of view have been supported by Gautier et al. (2010) and Purfield et al. (2012a).

Understanding the signatures of selection detected within livestock breeds contributes to the identification of genomic regions that are or have been targeted by

Figure 4. Linkage disequilibrium decay analysis within the river and swamp buffalo populations.

Figure 5. Autozygosity frequency distribution of runs of homozygosity (ROH) within the river and swamp buffalo populations. (A) Violin plot of inbreeding coefficient (F_{ROH}) detected within river and swamp buffaloes. The upper edge of the box indicates the upper quartile; the lower edge of the box indicates the lower quartile; the midline of the box indicates the median; the dots represent outliers; and the shape of the violin shows the density at any position. (B) Distribution of the fraction of ROH across length classes.

selection (Makina et al., 2015). Here, we performed selective sweep analysis between the river and swamp groups using hapFLK analysis. The result showed that only 1 genomic region under selection was identified with a length of approximately 0.82 Mb. Clearly, limited genomic regions were identified in the present study using the Axiom buffalo genotyping 90 K array. On the other hand, Mokhber et al. (2018) identified 13 genomic regions under selection within the 2 river buffalo breeds (Azeri and Khuzestani) using the same platform. The reason for the discrepancy may be the design of the Axiom buffalo SNP array, which mainly targeted the genomic study of river buffalo breeds. Out of the 90K SNP, about 24K markers were also polymorphic in swamp buffalo (Iamartino et al., 2017; Deng et al., 2019). In other words, a lower polymorphic SNP for swamp breeds may result in SNP ascertainment bias. Moreover, gene annotation analysis showed that a total of 15 significant SNP in the candidate region were detected within Chr19, corresponding to 13 candidate genes. Some of these genes have a vital role in mammary gland development and lactation. For example, 3 solute carrier family genes [solute carrier family 6 member 3 (*SLC6A3*), solute carrier family 9 member A3 (*SLC9A3*), and solute carrier family 12 member 7 (*SLC12A7*)] identified in the current study were previously reported to be associated with the regulation of milk protein. The solute carrier family proteins play a functional role in signal transduction as well as amino acid and glucose transporters in the bovine mammary gland (Bionaz et al., 2011). Notably, Basham et al. (2015) reported that aryl-hydrocarbon receptor repressor (*AHRR*) could mediate the AHR signal pathway to block milk production in the mammary epithelial cell

Table 2. Summary of the selective sweep regions detected by hapFLK (Fariello et al., 2013) analysis between river and swamp groups

Chromosome	SNP ¹	Position (bp)	Minor allele frequency		P-value	Gene
			River	Swamp		
19	AX-85140042	70,808,082	0.3605	0.0038	0.00073	<i>LPCAT1</i>
19	AX-85057565	70,856,547	0.1279	0.0076	0.00050	<i>SLC6A3</i>
19	AX-85061562	70,926,080	0.1047	0.0189	0.00049	<i>CLPTM1L</i>
19	AX-85065919	70,940,065	0.2791	0.3220	0.00050	<i>CLPTM1L</i>
19	AX-85094947	70,953,630	0.3605	0.0038	0.00050	<i>TERT</i>
19	AX-85092356	71,043,581	0.3605	0.0414	0.00050	<i>SLC12A7</i>
19	AX-85083255	71,121,031	0.0349	0.2748	0.00050	<i>NKD2</i>
19	AX-85048805	71,138,472	0.0833	0.4848	0.00049	<i>NKD2</i>
19	AX-85088305	71,150,069	0.2674	0.0038	0.00049	<i>LOC112580655</i>
19	AX-85123109	71,206,965	0.1860	0.1515	0.00048	<i>BRD9</i>
19	AX-85048806	71,236,414	0.3095	0.0075	0.00044	<i>LOC102398132</i>
19	AX-85108518	71,273,917	0.0238	0.1203	0.00038	<i>TPPP</i>
19	AX-85089633	71,425,140	0.2442	0.0000	0.00041	<i>SLC9A3</i>
19	AX-85100202	71,538,136	0.4535	0.4621	0.00045	<i>AHRR</i>
19	AX-85135336	71,631,407	0.2442	0.0038	0.00049	<i>SDHA</i>

¹Positions are given according to the latest reference assembly of the buffalo genome (UOA_WB_1: GCA_003121395.1).

Figure 6. Whole-genome scan for selective sweeps using hapFLK statistic (Fariello et al., 2013). The horizontal blue line represents the significance threshold. The number 25 represents the Chr X. TPPP = tubulin polymerization promoting protein.

and block transcription of the milk β -casein gene. This information suggested that the analysis results of the selection signature in the present study were somewhat reliable. However, further research on these candidate genes is required to explore the molecular mechanisms underlying phenotypes of important traits in buffaloes.

CONCLUSIONS

We made comprehensive analyses of molecular variability of the buffalo genome in the buffalo populations in southwest China. Our findings confirmed the divergent evolution and distinct population structure among Chinese buffaloes. We presented the LD pattern at the genome level in Chinese buffalo breeds and found that the swamp breeds underwent extensive remote inbreeding activity. Moreover, we identified promising candidate genes that are under directional selection for production performance in Chinese buffalo breeds. These results enrich our understanding of the genome biology of Chinese buffaloes.

ACKNOWLEDGMENTS

We thank the smallholders for providing us with buffalo samples. This study was funded by the Major Science and Technology Projects in Guangxi (AA16450002), Central Guidance for Local Science and Technology Development Projects (ZY18164003), and Natural Science Foundation of Guangxi (2017GXNSFBA198191 and 2017GXNSFBA198022). All of the funds are from the Department of Science and Technology of Guangxi Zhuang Autonomous Region (Nanning, China). We thank Angela Christine Oki for editing the language

of this paper. The authors have stated no conflicts of interest.

REFERENCES

- Ai, H., L. Huang, and J. Ren. 2013. Genetic diversity, linkage disequilibrium and selection signatures in Chinese and Western pigs revealed by genome-wide Snp markers. *PLoS One* 8:e56001. <https://doi.org/10.1371/journal.pone.0056001>.
- Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19:1655–1664. <https://doi.org/10.1101/gr.094052.109>.
- Barendse, W., B. E. Harrison, R. J. Bunch, M. B. Thomas, and L. B. Turner. 2009. Genome wide signatures of positive selection: The comparison of independent samples and the identification of regions associated to traits. *BMC Genomics* 10:178. <https://doi.org/10.1186/1471-2164-10-178>.
- Barker, J. S. F., S. S. Moore, D. J. S. Hetzel, D. Evans, and K. Byrne. 1997. Genetic diversity of Asian water buffalo (*Bubalus bubalis*): Microsatellite variation and a comparison with protein-coding loci. *Anim. Genet.* 28:103–115. <https://doi.org/10.1111/j.1365-2052.1997.00085.x>.
- Basham, K. J., C. J. Leonard, C. Kieffer, D. N. Shelton, M. E. McDowell, V. R. Bhonde, R. E. Looper, and B. E. Welm. 2015. Dioxin exposure blocks lactation through a direct effect on mammary epithelial cells mediated by the aryl hydrocarbon receptor repressor. *Toxicol. Sci.* 143:36–45. <https://doi.org/10.1093/toxsci/kfu203>.
- Bertolini, F., G. Schiavo, E. Scotti, A. Ribani, P. L. Martelli, R. Casadio, and L. Fontanesi. 2014. High-throughput SNP discovery in the rabbit (*Oryctolagus cuniculus*) genome by next-generation semiconductor-based sequencing. *Anim. Genet.* 45:304–307. <https://doi.org/10.1111/age.12121>.
- Bionaz, M., and J. J. Loo. 2011. Gene networks driving bovine mammary protein synthesis during the lactation cycle. *Bioinform. Biol. Insights* 5:83–98.
- Biscarini, F., P. Cozzi, G. Gaspa, and G. Marras. 2018. detectRUNS: Detect Runs of Homozygosity and Runs of Heterozygosity in Diploid Genomes. Accessed Feb. 2020. <https://github.com/bioinformatics-ptp/detectRUNS/tree/master/detectRUNS>.
- Biswas, S., and J. M. Akey. 2006. Genomic insights into positive selection. *Trends Genet.* 22:437–446. <https://doi.org/10.1016/j.tig.2006.06.005>.
- Bjelland, D. W., K. A. Weigel, N. Vukasinovic, and J. D. Nkrumah. 2013. Evaluation of inbreeding depression in Holstein cattle using

- whole-genome SNP markers and alternative measures of genomic inbreeding. *J. Dairy Sci.* 96:4697–4706. <https://doi.org/10.3168/jds.2012-6435>.
- Blanca, J., C. Esteras, P. Ziaresolo, D. Pérez, V. Fernández-Pedrosa, C. Collado, R. Rodrá Guez de Pablos, A. Ballester, C. Roig, J. Cañizares, and B. Picó. 2012. Transcriptome sequencing for SNP discovery across *Cucumis melo*. *BMC Genomics* 13:280. <https://doi.org/10.1186/1471-2164-13-280>.
- Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4:7. <https://doi.org/10.1186/s13742-015-0047-8>.
- Colli, L., M. Milanese, E. Vajana, D. Iamartino, L. Bomba, F. Puglisi, M. Del Corvo, E. L. Nicolazzi, S. S. Ahmed, J. R. Herrera, L. Cruz, S. Zhang, A. Liang, G. Hua, L. Yang, X. Hao, F. Zuo, S.-J. Lai, S. Wang, R. Liu, Y. Gong, M. Mokhber, Y. Mao, F. Guan, A. Vlaic, B. Vlaic, L. Ramunno, G. Cosenza, A. Ahmad, I. Soysal, E. Ö. Ünal, M. Ketudat-Cairns, J. F. Garcia, Y. T. Utsunomiya, P. S. Baruselli, M. E. J. Amaral, R. Pampai, M. G. Drummond, P. Galbusera, J. Burton, E. Hoal, Y. Yusnizar, C. Sumantri, B. Moidli, A. Valentini, A. Stella, J. L. Williams, and P. Ajmone-Marsan. 2018. New insights on water buffalo genomic diversity and post-domestication migration routes from medium density SNP chip data. *Front. Genet.* 9:53. <https://doi.org/10.3389/fgene.2018.00053>.
- de Camargo, G. M., R. R. Aspilueta-Borquis, M. R. Fortes, R. Porto-Neto, D. F. Cardoso, D. J. Santos, S. A. Lehnert, A. Reverter, S. S. Moore, and H. Tonhati. 2015. Prospecting major genes in dairy buffaloes. *BMC Genomics* 16:872. <https://doi.org/10.1186/s12864-015-1986-2>.
- de Simoni Gouveia, J. J., M. V. da Silva, S. R. Paiva, and S. M. de Oliveira. 2014. Identification of selection signatures in livestock species. *Genet. Mol. Biol.* 37:330–342. <https://doi.org/10.1590/S1415-47572014000300004>.
- Degradandi, T. M., S. Pita, Y. Panzera, E. H. C. D. Oliveira, J. R. F. Marques, M. R. Figueiró, L. C. Marques, L. Vinadé, R. J. Gunski, and A. D. V. Garnerio. 2014. Karyotypic evolution of ribosomal sites in buffalo subspecies and their crossbreed. *Genet. Mol. Biol.* 37:375–380. <https://doi.org/10.1590/S1415-47572014000300009>.
- Deng, T., A. Liang, J. Liu, G. Hua, T. Ye, S. Liu, G. Campanile, G. Plastow, C. Zhang, Z. Wang, A. Salzano, B. Gasparrini, M. Cassandro, H. Riaz, X. Liang, and L. Yang. 2019. Genome-wide SNP data revealed the extent of linkage disequilibrium, persistence of phase and effective population size in purebred and crossbred buffalo populations. *Front. Genet.* 9:688. <https://doi.org/10.3389/fgene.2018.00688>.
- 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. 2017. Genome-wide association study for milk production in Egyptian buffalo. *Livest. Sci.* 198:10–16. <https://doi.org/10.1016/j.livsci.2017.01.019>.
- FAO (Food and Agriculture Organization of the United Nations). 2019. Accessed Feb. 2020. <http://faostat.fao.org>.
- Fariello, M. I., S. Boitard, H. Naya, M. SanCristobal, and B. Servin. 2013. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics* 193:929–941. <https://doi.org/10.1534/genetics.112.147231>.
- Felsenstein, J. 2016. PHYLIP (Phylogeny Inference Package). Department of Genome Sciences, University of Washington, Seattle, WA.
- Ferenčaković, M., J. Sölkner, and I. Curik. 2013. Estimating autozygosity from high-throughput information: Effects of SNP density and genotyping errors. *Genet. Sel. Evol.* 45:42. <https://doi.org/10.1186/1297-9686-45-42>.
- Gautier, M., D. Laloë, and K. Moazami-Goudarzi. 2010. Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PLoS One* 5:e13038. <https://doi.org/10.1371/journal.pone.0013038>.
- Goudet, J. 2005. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Resour.* 5:184–186. <https://doi.org/10.1111/j.1471-8286.2004.00828.x>.
- Harisah, M., T. Azmi, M. Hilmi, M. Vidyadaran, T. Bongso, Z. Nava, V. Momongan, and P. Basrur. 1989. Identification of crossbred buffalo genotypes and their chromosome segregation patterns. *Genome* 32:999–1002. <https://doi.org/10.1139/g89-544>.
- Helyar, S. J., J. Hemmer-Hansen, D. Bekkevold, M. Taylor, R. Ogden, M. Limborg, A. Cariani, G. Maes, E. Diopere, G. Carvalho, and E. E. Nielsen. 2011. Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Mol. Ecol. Resour.* 11:123–136. <https://doi.org/10.1111/j.1755-0998.2010.02943.x>.
- Iamartino, D., E. L. Nicolazzi, C. P. Van Tassell, J. M. Reecy, E. R. Fritz-Waters, J. E. Koltes, S. Biffani, T. S. Sonstegard, S. G. Schroeder, P. Ajmone-Marsan, R. Negrini, R. Pasquariello, P. Ramelli, A. Coletta, J. F. Garcia, A. Ali, L. Ramunno, G. Cosenza, D. A. A. de Oliveira, M. G. Drummond, E. Bastianetto, A. Davassi, A. Pirani, F. Brew, and J. L. Williams. 2017. Design and validation of a 90K SNP genotyping assay for the water buffalo (*Bubalus bubalis*). *PLoS One* 12:e0185220. <https://doi.org/10.1371/journal.pone.0185220>.
- Jombart, T. 2008. adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>.
- Kathiravan, P., R. Kataria, B. Mishra, P. Dubey, D. Sadana, and B. Joshi. 2011. Population structure and phylogeography of Toda buffalo in Nilgiris throw light on possible origin of aboriginal Toda tribe of South India. *J. Anim. Breed. Genet.* 128:295–304. <https://doi.org/10.1111/j.1439-0388.2011.00921.x>.
- Kijas, J. W., J. A. Lenstra, B. Hayes, S. Boitard, L. R. Porto Neto, M. San Cristobal, B. Servin, R. McCulloch, V. Whan, K. Gietzen, S. Paiva, W. Barendse, E. Ciani, H. Raadsma, J. McEwan, and B. Dalrymple. 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol.* 10:e1001258. <https://doi.org/10.1371/journal.pbio.1001258>.
- Kumar, S., M. Nagarajan, J. S. Sandhu, N. Kumar, V. Behl, and G. Nishanth. 2007. Mitochondrial DNA analyses of Indian water buffalo support a distinct genetic origin of river and swamp buffalo. *Anim. Genet.* 38:227–232. <https://doi.org/10.1111/j.1365-2052.2007.01602.x>.
- Kumar, S., G. Stecher, and K. Tamura. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33:1870–1874. <https://doi.org/10.1093/molbev/msw054>.
- Lei, C. Z., W. Zhang, H. Chen, F. Lu, R. Liu, X. Yang, H. Zhang, Z. Liu, L. Yao, Z. Lu, and Z. L. Zhao. 2007. Independent maternal origin of Chinese swamp buffalo (*Bubalus bubalis*). *Anim. Genet.* 38:97–102. <https://doi.org/10.1111/j.1365-2052.2007.01567.x>.
- Liu, J. J., A. Liang, G. Campanile, G. Plastow, C. Zhang, Z. Wang, A. Salzano, B. Gasparrini, M. Cassandro, and L. Yang. 2018. Genome-wide association studies to identify quantitative trait loci affecting milk production traits in water buffalo. *J. Dairy Sci.* 101:433–444. <https://doi.org/10.3168/jds.2017-13246>.
- Low, W. Y., R. Tearle, D. M. Bickhart, B. D. Rosen, S. B. Kingan, T. Swale, F. Thibaud-Nissen, T. D. Murphy, R. Young, L. Lefevre, D. A. Hume, A. Collins, P. Ajmone-Marsan, T. P. L. Smith, and J. L. Williams. 2019. Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity. *Nat. Commun.* 10:260. <https://doi.org/10.1038/s41467-018-08260-0>.
- Makina, S. O., F. C. Muchadeyi, E. van Marle-Koster, J. F. Taylor, M. L. Makgahlela, and A. Maiwashe. 2015. Genome-wide scan for selection signatures in six cattle breeds in South Africa. *Genet. Sel. Evol.* 47:92.
- Mishra, B. P., P. Dubey, B. Prakash, P. Kathiravan, S. Goyal, D. Sadana, G. Das, R. Goswami, V. Bhasin, B. Joshi, and R. S. Kataria. 2015. Genetic analysis of river, swamp and hybrid buffaloes of north-east India throw new light on phylogeography of water buffalo (*Bubalus bubalis*). *J. Anim. Breed. Genet.* 132:454–466. <https://doi.org/10.1111/jbg.12141>.
- Mokhber, M., M. Moradi-Shahrababak, M. Sadeghi, H. Moradi-Shahrababak, A. Stella, E. Nicolazzi, J. Rahmaninia, and J. L. Williams.

2018. A genome-wide scan for signatures of selection in Azeri and Khuzestani buffalo breeds. *BMC Genomics* 19:449. <https://doi.org/10.1186/s12864-018-4759-x>.
- Nagarajan, M., K. Nimisha, and S. Kumar. 2015. Mitochondrial DNA variability of domestic river buffalo (*Bubalus bubalis*) populations: Genetic evidence for domestication of river buffalo in Indian sub-continent. *Genome Biol. Evol.* 7:1252–1259. <https://doi.org/10.1093/gbe/evv067>.
- Patterson, N., A. L. Price, and D. Reich. 2006. Population structure and eigenanalysis. *PLoS Genet.* 2:e190. <https://doi.org/10.1371/journal.pgen.0020190>.
- Peripolli, E., D. P. Munari, M. V. Silva, A. L. Lima, R. Irgang, and F. Baldi. 2017. Runs of homozygosity: Current knowledge and applications in livestock. *Anim. Genet.* 48:255–271.
- Purfield, D. C., D. P. Berry, S. McParland, and D. G. Bradley. 2012a. Runs of homozygosity and population history in cattle. *BMC Genet.* 13:70. <https://doi.org/10.1186/1471-2156-13-70>.
- Purfield, D. C., D. P. Berry, S. McParland, and D. G. Bradley. 2012b. Runs of homozygosity and population history in cattle. *BMC Genet.* 13:70. <https://doi.org/10.1186/1471-2156-13-70>.
- Qanbari, S., D. Gianola, B. Hayes, F. Schenkel, S. Miller, S. Moore, G. Thaller, and H. Simianer. 2011. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics* 12:318. <https://doi.org/10.1186/1471-2164-12-318>.
- Qanbari, S., E. Pimentel, J. Tetens, G. Thaller, P. Lichtner, A. Sharifi, and H. Simianer. 2010. A genome-wide scan for signatures of recent selection in Holstein cattle. *Anim. Genet.* 41:377–389. <https://doi.org/10.1111/j.1365-2052.2009.02016.x>.
- R Core Team. 2018. R: A language and environment for statistical computing. *Computing* 1:12–21.
- Rubin, C.-J., H.-J. Megens, A. M. Barrio, K. Maqbool, S. Sayyab, D. Schwochow, C. Wang, Ö. Carlborg, P. Jern, C. B. Jørgensen, A. L. Archibald, M. Fredholm, M. A. M. Groenen, and L. Andersson. 2012. Strong signatures of selection in the domestic pig genome. *Proc. Natl. Acad. Sci. USA* 109:19529–19536. <https://doi.org/10.1073/pnas.1217149109>.
- Smits, N., C. Berthouly, D. Cornelis, R. Heller, P. Van Hooft, P. Chardonnet, A. Caron, H. Prins, B. J. van Vuuren, H. De Jongh, and J. Michaux. 2013. Pan-African genetic structure in the African buffalo (*Syncerus caffer*): Investigating intraspecific divergence. *PLoS One* 8:e56235. <https://doi.org/10.1371/journal.pone.0056235>.
- Tills, D. 1977. The use of the F_{ST} statistic of Wright for estimating the effects of genetic drift, selection and migration in populations, with special reference to Ireland. *Hum. Hered.* 27:153–159. <https://doi.org/10.1159/000152865>.
- Turner, S. D. 2014. qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv* <https://doi.org/10.1101/005165>.
- Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with S*. 4th ed. Springer, New York, NY.
- Vijh, R. K., M. S. Tania, B. Mishra, and S. T. Bharani Kumar. 2008. Genetic relationship and diversity analysis of Indian water buffalo (*Bubalus bubalis*). *J. Anim. Sci.* 86:1495–1502. <https://doi.org/10.2527/jas.2007-0321>.
- Walugembe, M., F. Bertolini, C. M. B. Dematawewa, M. P. Reis, A. R. Elbeltagy, C. J. Schmidt, S. J. Lamont, and M. F. Rothschild. 2019. Detection of selection signatures among Brazilian, Sri Lankan, and Egyptian chicken populations under different environmental conditions. *Front. Genet.* 9:737. <https://doi.org/10.3389/fgene.2018.00737>.
- Wheat, C. W. 2012. SNP discovery in non-model organisms using 454 next generation sequencing. *Methods Mol. Biol.* 888:33–53. https://doi.org/10.1007/978-1-61779-870-2_3.
- Wu, F. Q., S. K. Shen, X. J. Zhang, Y. H. Wang, and W. B. Sun. 2014. Genetic diversity and population structure of an extremely endangered species: The world's largest Rhododendron. *AoB Plants* 7:plu082. <https://doi.org/10.1093/aobpla/plu082>.
- Yang, B. Z., X. W. Liang, J. Qin, C. J. Yang, and J. H. Shang. 2013. Brief introduction to the development of Chinese dairy buffalo industry. *Buffalo Bull.* 32:111–120.
- Yindee, M., B. Vlamings, W. Wajjwalku, M. Techakumphu, C. Lo-hachit, S. Sirivaidyapong, C. Thitaram, A. Amarasinghe, P. Alexander, B. Colenbrander, and J. A. Lenstra. 2010. Y-chromosomal variation confirms independent domestications of swamp and river buffalo. *Anim. Genet.* 41:433–435. <https://doi.org/10.1111/j.1365-2052.2010.02020.x>.
- Yuan, Z., E. Liu, Z. Liu, J. W. Kijas, C. Zhu, S. Hu, X. Ma, L. Zhang, L. Du, H. Wang, and C. Wei. 2017. Selection signature analysis reveals genes associated with tail type in Chinese indigenous sheep. *Anim. Genet.* 48:55–66. <https://doi.org/10.1111/age.12477>.
- Yue, X.-P., R. Li, W.-M. Xie, P. Xu, T.-C. Chang, L. Liu, F. Cheng, R.-F. Zhang, X.-Y. Lan, H. Chen, and C.-Z. Lei. 2013. Phylogeography and domestication of Chinese swamp buffalo. *PLoS One* 8:e56552. <https://doi.org/10.1371/journal.pone.0056552>.
- Zhang, C., S. S. Dong, J. Y. Xu, W. M. He, and T. L. Yang. 2019. PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35:1786–1788. <https://doi.org/10.1093/bioinformatics/bty875>.
- Zhang, Y., Y. Lu, M. Yindee, K. Y. Li, H. Y. Kuo, Y. T. Ju, S. Ye, M. O. Faruque, Q. Li, Y. Wang, V. C. Cuong, L. D. Pham, B. Bouahom, B. Yang, X. Liang, Z. Cai, D. Vankan, W. Manatchaiworakul, N. Kowlim, S. Duangchantrasiri, W. Wajjwalku, B. Colenbrander, Y. Zhang, P. Beerli, J. A. Lenstra, and J. S. F. Barker. 2016. Strong and stable geographic differentiation of swamp buffalo maternal and paternal lineages indicates domestication in the China/Indochina border region. *Mol. Ecol.* 25:1530–1550. <https://doi.org/10.1111/mec.13518>.
- Zhang, Y., D. Sun, Y. Yu, and Y. Zhang. 2007. Genetic diversity and differentiation of Chinese domestic buffalo based on 30 microsatellite markers. *Anim. Genet.* 38:569–575. <https://doi.org/10.1111/j.1365-2052.2007.01648.x>.
- Zhang, Y., D. Sun, Y. Yu, and Y. Zhang. 2008. Genetic variation and divergence among swamp buffalo, river buffalo and cattle: A microsatellite survey on five populations in China. *Asian-Australas. J. Anim. Sci.* 21:1238–1243. <https://doi.org/10.5713/ajas.2008.70746>.
- Zhang, Y., D. Vankan, Y. Zhang, and J. S. Barker. 2011. Genetic differentiation of water buffalo (*Bubalus bubalis*) populations in China, Nepal and south-east Asia: Inferences on the region of domestication of the swamp buffalo. *Anim. Genet.* 42:366–377. <https://doi.org/10.1111/j.1365-2052.2010.02166.x>.

ORCIDS

- X. R. Lu <https://orcid.org/0000-0002-7822-6953>
H. Abdel-Shafy <https://orcid.org/0000-0003-3282-3899>
H. E. Rushdi <https://orcid.org/0000-0002-8516-0366>
T. X. Deng <https://orcid.org/0000-0002-7442-6739>