BUFFALO GENETICS AND GENOMICS

EDITED BY: Guohua Hua, Hamdy Abdel-Shafy, Tingxian Deng, Yang Zhou and Wai Yee Low

PUBLISHED IN: Frontiers in Genetics

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88974-542-5
DOI 10.3389/978-2-88974-542-5

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

BUFFALO GENETICS AND GENOMICS

Topic Editors:
Guohua Hua, Huazhong Agricultural University, China
Hamdy Abdel-Shafy, Cairo University, Egypt
Tingxian Deng, Buffalo Research Institute, Chinese Academy of Agricultural Sciences, China
Yang Zhou, Huazhong Agricultural University, China
Wai Yee Low, University of Adelaide, Australia

Citation: Hua, G., Abdel-Shafy, H., Deng, T., Zhou, Y., Low, W. Y., eds. (2022). Buffalo Genetics and Genomics. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88974-542-5

Table of Contents

frontiers in Genetics

EDITORIAL
published: 28 January 2022
doi: 10.3389/fgene.2021.820627

Editorial: Buffalo Genetics and Genomics

Hamdy Abdel-Shafy[1]*, Tingxian Deng[2], Yang Zhou[3], Wai Yee Low[4] and Guohua Hua[3]

[1]Department of Animal Production, Faculty of Agriculture, Cairo University, Giza, Egypt, [2]Key Laboratory of Buffalo Genetics, Breeding and Reproduction Technology, Buffalo Research Institute, Chinese Academy of Agricultural Sciences, Nanning, China, [3]Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, College of Animal Science and Technology, Huazhong Agricultural University, Wuhan, China, [4]The Davies Livestock Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Adelaide, Australia

Keywords: buffalo, evolutionary biology, population genetics, molecular genetics, omics

Editorial on the Research Topic

Buffalo Genetics and Genomics

Buffalo (*Bubalus bubalis*) are important livestock species with significant contribution to food security for thousands of years as a source of milk, meat, leather, dung, hide, horns, traction power, etc. Buffalo production is almost doubled during the last decades due to the improvement in management and nutrition practices along with advanced breeding approaches. To ensure more food security, it is important to sustain the improvement and efficiency of buffalo production to meet the current and upcoming human needs. Genetic improvement is usually used to achieve this goal by selecting the best individuals and breeding them to pass down their favorable genetic materials to the next generations. In this regard, the merit of an animal is predicted in terms of its estimated breeding values (EBVs) even without knowledge of the genetic control of the relevant traits.

With the release of buffalo genome assemblies such as the upgraded reference with long read sequencing (Low et al., 2019), the revolution of high-throughput genotyping technologies has opened the field of buffalo breeding to use omics information to increase the efficiency of selection, including but not limited to genomic prediction, genome-wide association studies (GWASs), evolutionary biology, and functional genomics. These approaches are showed the potential to significantly alter our understanding of the genetic basis of economically important traits in buffalo and enable the scientists to draw a complete picture that previously had major gaps. In this regard, our research topic yielded eleven publications covering diverse approaches and ideas, e.g., classical breeding, genomic prediction, candidate genes, and molecular characterization of different buffalo breeds.

The increased efficiency of production during the last decades is commendable. Although persistency for milk production traits has economic importance, limited studies have been performed so far to determine their genetic parameters in buffalo. Therefore, Nazari et al. estimated the genetic parameters of different persistency measures for milk production traits in Iranian buffalo. They proposed persistency measures of fat production that had favorable low genetic correlations with total milk yield; hence it has an additional benefit when designing breeding schemes. However, the implementation of successful breeding programs based on classical prediction in buffalo is hindered by the lack of sufficient pedigree information traced back many generations ago. This is partially due to natural mating in buffalo, which is still a common reproductive approach used in most farms. A possible solution is to use genetic groups during estimation for variance component and EBV. However, as the percent of missing genealogies increased, the accuracy of prediction is going to decrease regardless the genetic grouping strategies and trait analyzed (Gómez et al.). Another possible solution to overcome the missing pedigree information is to use genomic data. Even with the availability of pedigree information, genomic methods can provide more accurate prediction than those of traditional estimations. For example,

OPEN ACCESS

Edited and reviewed by:
Johann Sölkner,
University of Natural Resources and Life Sciences Vienna, Austria

***Correspondence:**
Hamdy Abdel-Shafy
hamdyabdelshafy@agr.cu.edu.eg

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 23 November 2021
Accepted: 16 December 2021
Published: 28 January 2022

Citation:
Abdel-Shafy H, Deng T, Zhou Y, Low WY and Hua G (2022) Editorial: Buffalo Genetics and Genomics. Front. Genet. 12:820627. doi: 10.3389/fgene.2021.820627

the average accuracies for GBLUP and ssGBLUP were increased by 0.03 and 0.08 units over pBLUP (0.21), respectively for milk production traits in Philippine buffalo (Herrera et al.). Although these results are promising, the advantage of using genomic information for genetic improvement in buffalo is still lower than what was expected. It would be attribute to the small number of genotyped animals, using animal own performance, and small sample size (Abdel-Shafy et al., 2020a). One possible solution is to establish a multi-breed reference population (Bolormaa et al., 2013). In this case, it is very important to ensure that the target breed is presented in the multi-breed reference population; otherwise, the accuracy of prediction will be very low due to the inconsistence of linkage disequilibrium (LD) among breeds. In this regard, Rahimmadar, et al. studied the LD structure among different buffalo breeds. They found that the LD measure among SNPs is decreased by increasing the physical distance from 100 Kb to 1 Mb. They also reported that the LD patterns were almost similar among studied breeds. Therefore, the multi-breed reference population for buffalo would be established to increase the accuracy of prediction.

Recently, it has been reported that incorporating biological information and pre-selected genetic markers can increase the accuracy of prediction (Hayes and Daetwyler, 2019). Detection of these loci can be achieved by GWASs, as it has been previously reported for milk production traits in different buffalo breeds (de Camargo et al., 2015; El-Halawany et al., 2017; Iamartino et al., 2017; Mokhber, 2017; da Costa Barros et al., 2018; Liu et al., 2018; Herrera et al., 2018; Lu et al., 2020; Abdel-Shafy et al., 2020b; Awad et al., 2020). However, none of the detected regions was overlapped among different populations and/or validated. In this case, candidate gene approaches would be a complementary method to accurately identify genetic markers and/or causative mutations associated with the relevant trait (Wilkening et al., 2009). In this regard, Tyagi et al. suggested several promising genes for milk production and immunity to be considered for further studies in Indian Murrah buffalo. Likewise, Cosenza et al. and Rehman et al. have intensively studied the evolutionary relationship, comparative genomic, physiochemical properties, and association analysis of casein gene family in different buffalo breeds. They provided useful information about the roles of casein gene family for the variation in milk production traits. In addition, Zhu et al. and Zhang et al. investigated the long non-coding RNAs (lncRNAs) profiles of adipose and muscle tissues in buffalo. They have been identified and verified several differentially expressed lncRNAs in adipose and muscle tissues revealing the importance of lncSAMM50 in lipid accumulation of buffalo adipocytes.

Since cattle and buffalo are closely related species, it is common to compare the findings from buffalo studies with their relevant ones from previous cattle studies. In this regards, Shao et al. compile the genetic parameters and GWASs for different reproductive traits in both cattle and buffalo populations and highlighted possible options to be implemented for improving buffalo breeding. Recently, the research priorities and strategic plans in developing countries have focused on improving the performance of local breeds to face climate change. Swamp buffalo, which are mainly used for agricultural operations in China and Southeast Asian countries, are currently facing additional challenge of being neglected due to rising farm mechanization. This subspecies can be developed for milk and/or meat production under harsh environments and can be used as a strategic option to secure the income of smallholders. Therefore, the challenges and possible opportunities for improving the productivity of swamp buffalo in the Southeastern Asia are comprehensively discussed by Pineda et al.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

ACKNOWLEDGMENTS

We wish to thank all the authors and reviewers for their valuable contributions to ensure high quality articles for this research topic and we hope this collection will be of interest to the research community.

REFERENCES

Abdel-Shafy, H., Awad, M. A. A., El-Regalaty, H., Ismael, A., El-Assal, S. E.-D., and Abou-Bakr, S. (2020). A Single-Step Genomic Evaluation for Milk Production in Egyptian buffalo. *Livestock Sci.* 234, 103977. doi:10.1016/j.livsci.2020.103977

Abdel-Shafy, H., Awad, M. A. A., El-Regalaty, H., El-Assal, S. E., and Abou-Bakr, S. (2020). Prospecting Genomic Regions Associated with Milk Production Traits in Egyptian Buffalo. *J. Dairy Res.* 87 (4), 389–396. doi:10.1017/S0022029920000953

Awad, M. A. A., Abou-Bakr, S., El-Regalaty, H., El-Assal, S. E-D., and Abdel-Shafy, H. (2020). Determination of Potential Candidate Genes Associated with Milk Lactose in Egyptian buffalo. *World's Vet. J.* 10 (1), 35–42. doi:10.36380/scil.2020.wvj5

Bolormaa, S., Pryce, J. E., Kemper, K., Savin, K., Hayes, B. J., Barendse, W., et al. (2013). Accuracy of Prediction of Genomic Breeding Values for Residual Feed Intake and Carcass and Meat Quality Traits in *Bos T*, *Bos I*, and Composite Beef Cattle1. *J. Anim. Sci.* 91 (7), 3088–3104. doi:10.2527/jas.2012-5827

da Costa Barros, C., de Abreu Santos, D. J., Aspilcueta-Borquis, R. R., de Camargo, G. M. F., de Araújo Neto, F. R., and Tonhati, H. (2018). Use of Single-Step Genome-Wide Association Studies for Prospecting Genomic Regions Related to Milk Production and Milk Quality of Buffalo. *J. Dairy Res.* 85 (4), 402–406. doi:10.1017/s0022029918000766

de Camargo, G., Aspilcueta-Borquis, R., Fortes, M., Porto-Neto, R., Cardoso, D., Santos, D., et al. (2015). Prospecting Major Genes in Dairy Buffaloes. *BMC Genomics* 16, 872. doi:10.1186/s12864-015-1986-2

El-Halawany, N., Abdel-Shafy, H., Shawky, A.-E. -M. A., Abdel-Latif, M. A., Al-Tohamy, A. F. M., and Abd El-Moneim, O. M. (2017). Genome-Wide Association Study for Milk Production in Egyptian Buffalo. *Livestock Sci.* 198, 10–16. doi:10.1016/j.livsci.2017.01.019

Hayes, B. J., and Daetwyler, H. D. (2019). 1000 Bull Genomes Project to Map Simple and Complex Genetic Traits in Cattle: Applications and Outcomes. *Annu. Rev. Anim. Biosci.* 7, 89–102. doi:10.1146/annurev-animal-020518-115024

Herrera, J. R. V., Flores, E. B., Duijvesteijn, N., Gondro, C., and van der Werf, J. H.J. (2018). "Genome-Wide Association Study for Milk Traits in Philippine Dairy Buffaloes," in

Proceedings, 11th World Congress on Genetics Applied to Livestock Production, New Zealand, February 11–16, 2018.

Iamartino, D., Nicolazzi, E. L., Van Tassell, C. P., Reecy, J. M., Fritz-Waters, E. R., Koltes, J. E., et al. (2017). Design and Validation of a 90K SNP Genotyping Assay for the Water Buffalo (Bubalus Bubalis). *PLoS One* 12 (10), e0185220. doi:10.1371/journal.pone.0185220

Liu, J. J., Liang, A. X., Campanile, G., Plastow, G., Zhang, C., Wang, Z., et al. (2018). Genome-Wide Association Studies to Identify Quantitative Trait Loci Affecting Milk Production Traits in Water buffalo. *J. Dairy Sci.* 101 (1), 433–444. doi:10.3168/jds.2017-13246

Low, W. Y., Tearle, R., Bickhart, D. M., Rosen, B. D., Kingan, S. B., Swale, T., et al. (2019). Chromosome-Level Assembly of the Water Buffalo Genome Surpasses Human and Goat Genomes in Sequence Contiguity. *Nat. Commun.* 10 (1), 260. doi:10.1038/s41467-018-08260-0

Lu, X. R., Duan, A. Q., Li, W. Q., Abdel-Shafy, H., Rushdi, H. E., Liang, S. S., et al. (2020). Genome-Wide Analysis Reveals Genetic Diversity, Linkage Disequilibrium, and Selection for Milk Production Traits in Chinese Buffalo Breeds. *J. Dairy Sci.* 103 (5), 4545–4556. doi:10.3168/jds.2019-17364

Mokhber, M. (2017). "Genome-Wide Association Study for Milk Production in Iranian Buffalo," in 1st International and 5th National Conference on Organic vs. Conventional Agriculture, Ardabil, Iran, August 16–17, 2017.

Wilkening, S., Chen, B., Bermejo, J. L., and Canzian, F. (2009). Is There Still a Need for Candidate Gene Approaches in the Era of Genome-Wide Association Studies? *Genomics* 93 (5), 415–419. doi:10.1016/j.ygeno.2008.12.011

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 04 February 2021
doi: 10.3389/fgene.2021.625335

Accounting for Genetic Differences Among Unknown Parents in *Bubalus bubalis*: A Case Study From the Italian Mediterranean Buffalo

Mayra Gómez[1], Dario Rossi[1], Roberta Cimmino[1], Gianluigi Zullo[1], Yuri Gombia[1], Damiano Altieri[1], Rossella Di Palo[2] and Stefano Biffani[3]*

[1] Italian National Association of Buffalo Breeders, Caserta, Italy, [2] Department of Veterinary Medicine and Animal Production, University of Federico II, Naples, Italy, [3] Institute of Agricultural Biology and Biotechnology, National Research Council, Milan, Italy

The use of genetic evaluations in the Water Buffalo by means of a Best Linear Unbiased Prediction (BLUP) animal model has been increased over the last two-decades across several countries. However, natural mating is still a common reproductive strategy that can increase the proportion of missing pedigree information. The inclusion of genetic groups in variance component (VC) and breeding value (EBV) estimation is a possible solution. The aim of this study was to evaluate two different genetic grouping strategies and their effects on VC and EBV for composite (n = 5) and linear (n = 10) type traits in the Italian Mediterranean Buffalo (IMB) population. Type traits data from 7,714 buffalo cows plus a pedigree file including 18,831 individuals were provided by the Italian National Association of Buffalo Breeders. VCs and EBVs were estimated for each trait fitting a single-trait animal model and using the official DNA-verified pedigree. Successively, EBVs were re-estimated using modified pedigrees with two different proportion of missing genealogies (30 or 60% of buffalo with records), and two different grouping strategies, year of birth (Y30/Y60) or genetic clustering (GC30, GC60). The different set of VCs, estimated EBVs and their standard errors were compared with the results obtained using the original pedigree. Results were also compared in terms of efficiency of selection. Differences among VCs varied according to the trait and the scenario considered. The largest effect was observed for two traits, udder teat and body depth in the GC60 genetic cluster, whose heritability decreased by −0.07 and increased by +0.04, respectively. Considering buffalo cows with record, the average correlation across traits between official EBVs and EBVs from different scenarios was 0.91, 0.88, 0.84, and 0.79 for Y30, CG30, Y60, and CG60, respectively. In bulls the correlations between EBVs ranged from 0.90 for fore udder attachment and udder depth to 0.96 for stature and body length in the GC30 scenario and from 0.75 for udder depth to 0.90 for stature in the GC60 scenario. When a variable proportion of missing pedigree is present using the appropriate strategy to define genetic groups and including them in VC and EBV is a worth-while and low-demanding solution.

Keywords: buffalo, breeding values, unknown parent groups, type traits, heritability

Edited by:
Tingxian Deng,
Institute of Buffalo (CAAS), China

Reviewed by:
Faiz-ul Hassan,
University of Agriculture, Faisalabad, Pakistan
Hossam Eldin Rushdi Ahmed Ali Osman,
Cairo University, Egypt

Correspondence:
Stefano Biffani
stefano.biffani@ibba.cnr.it;
biffani@ibba.cnr.it

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *02 November 2020*
Accepted: *13 January 2021*
Published: *04 February 2021*

Citation:
Gómez M, Rossi D, Cimmino R, Zullo G, Gombia Y, Altieri D, Di Palo R and Biffani S (2021) Accounting for Genetic Differences Among Unknown Parents in Bubalus bubalis: A Case Study From the Italian Mediterranean Buffalo. Front. Genet. 12:625335. doi: 10.3389/fgene.2021.625335

INTRODUCTION

The Water Buffalo (*Bubalus bubalis*) is a large bovid mainly distributed in the Asian continent where the 97% of its world population is concentrated [Food and Agriculture Organization (FAO), 2020]. The name "water buffalo" is due to its adaptation to flooded or swampy areas, where it partially submerges and walks on the bottom mud without difficulty. The rest of the water buffalo world population (3%) is raised in the Mediterranean area historically characterized by the same optimal rearing conditions. In the European continent only the 0.2% of its world population is found and about 93% of these animals are located in south-central Italy (Neglia et al., 2020). The total census in Italy has increased considerably over the last decade, making it one of the most important dairy species in the country. In 2019, 34,990 lactating buffaloes have been registered to the official herd book. Moreover, 666,960 controlled lactations and 9,953 type traits evaluations are available and officially recorded [Associazione Nazionale Allevatori Specie Bufalina (ANASB), 2020]. Thanks to the physical-chemical properties of its milk—high concentration in protein and fat (FC $\sim$ 8%) and favorable coagulation (Costa et al., 2020b)—the main zootechnical interest of the Italian Mediterranean Buffalo (IMB) is the production of the iconic traditional dairy products like the Mozzarella di Bufala Campana (Boselli et al., 2020), which has a great economic impact on the Italian food industry (ISMEA, 2020). Costa et al. (2020a,b) refers to the outstanding increase of IMB population size observed in the last 15 years, as well as the increase in terms of kilos of cheese produced, the larger herd size, the constant expansion in registered herds and the increment in milk price. Therefore, the economic interest in this specie makes it necessary to develop new innovative tools to improve the breeding process.

The implementation of genetic evaluations in the Water Buffalo based on a BLUP animal model has been increasing over the last decade across several countries (Agudelo-Gómez et al., 2015; Safari et al., 2018; Abdel-Shafy et al., 2020). The prediction of breeding values (EBVs) constitutes an integral part of most breeding programs which are based on two fundamental pillars: phenotypic data (e.g., milk production%, fat%, protein, or morphological trait) and genealogical information (i.e., a pedigree). However, if animals with unknown parents are present in the pedigree, bias in the prediction of both variance component (VC) and EBV is expected (Peškovičová et al., 2004; Petrini et al., 2015). BLUP methodology allows for the simultaneous estimation of fixed and random effects but gaps in the relationship matrix may jeopardize its unbiasedness due to the inability of correctly estimating and disentangling genetic and environmental components (Postma, 2006; Gómez et al., 2016; Wolak and Reid, 2017). Indeed, incomplete pedigree information can lead to inaccurate prediction of animal genetic potential, overestimating or underestimating animal breeding value and hampering decisions based on the selection eventually causing economic losses (Raoul et al., 2016; Carneiro et al., 2017; Abdel-Shafy et al., 2020).

One of the reason behind incomplete pedigree information is the use of natural mating, still common in the buffalo herds, which makes parentage assignment more complex. Indeed, in IMB the use of the artificial insemination (AI) is still moderate (Parlato and Van Vleck, 2012). According to official data [Associazione Nazionale Allevatori Specie Bufalina (ANASB), 2020] and following a worldwide tendency (Singh and Balhara, 2016; Purohit et al., 2019), the proportion of natural mating in IMB decreased from around 76 to 62% from 2010 to 2019 [Associazione Nazionale Allevatori Specie Bufalina (ANASB), 2020]. These values, even if promising, are still lower than what it is observed in other species such as in dairy cattle, where the use of artificial insemination is close to 100% (Rodríguez-Martínez and Peña Vega, 2013; Ugur et al., 2019). Among the reasons why natural mating is still the most common reproduction technology for water buffalo there are physiological and reproductive aspects, herd management, breeding techniques, and organization (Neglia et al., 2020).

Despite being a routine analysis, it is almost impossible for the farmer to bear the total cost of parentage verification and to have his entire herd genotyped. In detail, in 2019 approximately 10,000 individuals have received a type trait evaluation in Italy but only 4,671 were DNA tested [Associazione Nazionale Allevatori Specie Bufalina (ANASB), 2020]. Hence, we are in a situation where phenotypic data are available for many animals, but a large proportion of these animals do not have complete pedigree information. Despite this limitation, the number of paternity tests in IMB in year 2019 showed a two-fold increase compared to year 2018.

Moreover, parentage testing is often reserved only for the best animals causing additional biases in the genetic evaluation being eventually based on a selected and non-random sample of the effective population. Furthermore, the possibility of using a larger number of data, albeit with incomplete pedigree, allows to observe all the variability of the trait of interest and therefore to obtain more accurate estimates.

The problem of incomplete pedigree has existed for many years and continues to be one of the main issues in genetic evaluations. Several researchers have worked on possible statistical approaches in order to correct for the presence of gaps in the pedigree (Peškovičová et al., 2004; Carneiro et al., 2017; Tonussi et al., 2017; Shiotsuki et al., 2018; Nwogwugwu et al., 2020; Macedo et al., 2020). The implementation of new technologies such as high-throughput single-nucleotide polymorphism (SNP) genotyping will certainly solve most of the problems linked to uncertain paternity but this is true only for individuals who are still alive or whose biological samples are available. Moreover, although genomics is the new standard in breeding and genetics, there are still some problems that need to be solved regarding how to cope with missing pedigree information (Tonussi et al., 2017; Misztal et al., 2020).

One suggested solution when dealing with an incomplete pedigree is the use of "Genetic Groups" approach, suggested over 30 years ago by Westell et al. (1988). This approach is based on the concept that subjects born in a certain period or coming from a certain area are the result of specific selective choices and therefore "genetically different" from other subjects born in other periods or from other areas.

The inclusion of genetic groups in VC and EBV is a method that has been adopted and extensively validated, as an example, in beef and dairy cattle (Perez-Enciso and Fernando, 1992; Sullivan, 1995; Theron et al., 2002; Peškovičová et al., 2004; Phocas and Laloë, 2004; Petrini et al., 2015; Wolak and Reid, 2017). The assignment of genetics groups to animals with uncertain genealogy represents a simple and effective solution to increase the accuracy of genetic evaluations (Henderson, 1988; Cardoso and Tempelman, 2003).

However, a crucial aspect is the strategy used to define the genetic groups. Therefore, the aim of this study was to evaluate the use of different genetic grouping strategies and its effects on VC and EBV estimation for 5 composite and 10 linear traits in the IMB population.

MATERIALS AND METHODS

Ethics Statement

Animal welfare and use committee approval was not needed for this study as datasets were obtained from pre-existing databases based on routine animal recording procedures.

Data Description

Data for the present study were provided by the Italian National Association of Buffalo Breeders (ANASB) and consisted of linear appraisal records from years 2004 to 2020. The initial data set included 79,342 IMB cows from 464 herds phenotyped for fifteen type traits. The type traits were five composite traits, namely, final score (FS), structure (ST), feet and legs (FL), yield potential (YP) and udder teat (UT), and 10 linear traits, namely, stature (STAT), body depth (BD), body length (BL), foot angle (FA), fore udder attachment (FUA), rear udder width (RUW), udder depth (UD), teat placement (TP), teat length (TL), and body condition score (BCS). The median age at evaluation was 46 months. The scale used for scoring varied according to the set of observed traits. Composite traits were scored on a 65–100 scale, linear traits were scored on a 1–50 scale and BCS was scored on a 4.5–9.5 scale. Overall 17 official classifiers were enrolled in the scoring procedures. Data editing consisted of retaining only cows from herds with at least two contemporaries (i.e., individuals classified by the same classifier in the same round of classification) and whose ascendants were confirmed by a DNA parentage test. Finally, 7,714 buffalo cows belonging to 194 herd with a pedigree containing 18,831 individuals were used in the analysis. Descriptive statistics are in **Table 1**.

TABLE 1 | Mean, standard deviation (SD), minimum (Min), maximum (Max), and coefficient of variation (CV) for traits evaluated in the IMB.

Type	Trait	Mean	SD	Min	Max	CV
Composite	Final score (FS)	81.34	1.82	65	87	0.02
	Structure (ST)	82.50	2.38	69	91	0.03
	Feet and legs (FL)	80.19	2.59	65	89	0.03
	Under teat (UT)	80.30	2.64	65	90	0.03
	Yield potential (YP)	83.44	2.14	71	90	0.03
Linear	Stature (STAT)	30.57	6.56	8	50	0.21
	Body depth (BD)	29.48	6.00	7	50	0.20
	Body length (BL)	31.50	6.56	10	50	0.21
	Foot angle (FA)	22.65	6.14	3	50	0.27
	Fore udder attachment (FUA)	22.39	6.84	2	46	0.31
	Rear udder width (RUW)	24.20	6.12	2	50	0.25
	Udder depth (UD)	27.69	6.33	2	50	0.23
	Teat placement (TP)	21.30	4.74	1	50	0.22
	Teat length (TL)	23.85	7.04	2	50	0.30
	Body condition score (BCS)	7.34	0.47	4.5	9.5	0.06

Alteration of Genetic Relationships and Grouping Strategies

The impact of different genetic grouping strategies on VC, EBV, and their accuracies (ACC) was investigated using the original pedigree and a modified pedigree where two different proportion of missing genealogies, namely, 30% (30) and 60%, (60) were randomly introduced. The choice of using these two thresholds was based not only on the need to mimic the real situation observed across ANASB farms but also to investigate the effect of moderate or massive pedigree gaps. After introducing the missing genealogy, the individual was assigned to a specific genetic group. Genetic groups (GG) were created following two clustering methods.

The first method (Y) was based on the year of birth and on an average generation interval, which for the IMB was defined (based on an estimation on actual IMB data) as 6 years. Individuals born before 1985 was considered as base animals and assigned to group 1. The remainder of the buffaloes was assigned to six different groups.

The second grouping strategy (GC) was based on the genetic distances estimated from the original pedigree. The procedure consisted of two steps. In the first step the pedigree-based additive relationship matrix was calculated and used as input for a hierarchical cluster analysis using a complete-linkage clustering method (Kaufman and Rousseeuw, 2009). This method works in a bottom-up manner. Each object is initially considered as a single-element cluster (leaf). At each step of the algorithm, the two clusters that are the most similar are combined into a new bigger cluster (nodes).

This procedure is iterated until all points are member of just one single big cluster (root). The result is a tree that can be plotted as a dendrogram. In the second step, the dendrogram is visually evaluated to define *a priori* the cut-off level that will identify the number of clusters (i.e., genetic groups). Each individual is then assigned to a particular cluster. Following the above mentioned procedure, fourteen different genetic groups were created (**Supplementary Figure 1**).

In detail at the end of the procedures, four scenarios were created according to the grouping strategy (Y or GC) and the proportion of missing genealogies (30 or 60%).

Successively, VC, EBV, and ACC were estimated for each trait presented in **Table 1** fitting a single-trait animal model and using the original pedigree (GOLD) and the four scenarios, namely Y30, Y60, GC30, and GC60. Estimates from GOLD were considered as *gold standard*. The estimation of VC, EBV, and ACC was repeated 10 times per each scenario (Y30, Y60, GC30, and GC60). The average number of animals and its standard deviation per scenario are shown in **Table 2**.

TABLE 2 | Average number of animals (and standard deviation) by genetic grouping strategy (GG) and proportion of missing genealogies.

GG	Level	Proportion of missing genealogies	
		30%	60%
Y[a]	1	43 (0)	43 (0)
	2	456 (1)	456 (1)
	3	1,798 (87)	1,800 (89)
	4	2,524 (432)	2694 (607)
	5	2,394 (412)	2,876 (906)
	6	1,001 (271)	1,435 (715)
	7	148 (35)	218 (106)
GC[b]	1	5,973 (656)	6,279 (656)
	2	695 (52)	985 (54)
	3	369 (22)	559 (23)
	4	279 (53)	450 (55)
	5	291 (81)	468 (81)
	6	345 (96)	556 (96)
	7	134 (35)	218 (35)
	8	356 (90)	579 (90)
	9	206 (58)	330 (57)
	10	101 (31)	162 (30)
	11	219 (61)	353 (61)
	12	238 (56)	393 (56)
	13	249 (65)	397 (66)
	14	69 (21)	109 (22)

[a]*Grouping strategy based on the year of birth and on an average generation interval set to 6 years.*
[b]*Grouping strategy based on the genetic distances estimated from the original pedigree.*

Genetic Analysis

The following single-trait animal model with groups was used to estimate VC, their corresponding heritability, and breeding value for each considered trait:

$$y_{ijklm} = \mu + hyc_i + PA_j + DIM_k + NM_l + a_m + \sum_{n=1}^{p} t_{mn} g_n + e_{ijklm}$$

where y_{ijklm} is the score of each trait for a given buffalo cow; μ is the overall mean; hyc_i is the fixed effect of the ith herd-year of evaluation-classifier (i = 1,. . .957); PA_j is the fixed effect of the jth age nested within parity (j = 1,. . .173); DIM_k is the fixed effect of the kth days in milk (k = 1,. . .30); NM_l is the fixed effect of the lth number of milking (l = 1,. . .3); a_m is the random additive genetic effect of the mth buffalo; g_n is the fixed group effect based on Y or GG and containing the nth ancestor; t_{mn} is the additive relationship between the nth and mth animals and the summation is over all p ancestors of animal m; and e_{ijklm} is the random residual effect.

In matrix notation, the model can be written as:

$$y = Xb + Z_a Q_a g_a + Z_a a + e$$

where matrix X is an incidence matrix relating phenotypic records in vector y to fixed effects in vector b, matrix Z_a is an incidence matrix relating phenotypic records in vector y to animal additive genetic effects in vector a, matrix Q_a is an incidence matrix relating animals in vector a to unknown parent groups in vector g_a. Vectors a and e have means 0 and variances $A\sigma_a^2$ and σ_e^2, respectively.

The corresponding mixed-model equations were:

$$\begin{bmatrix} X'X & X'Z & X'ZQ \\ Z'X & Z'Z + A^{-}\alpha & Z'ZQ \\ Q'Z'X & Q'Z'Z & Q'Z'ZQ \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ Q'Z'y \end{bmatrix}$$

Solving the equations the breeding value of an animal m will be:

$$a_{m^*} = Q\hat{g} + \hat{a}_m$$

The accuracy of EBV was calculated as recommended by Aguilar et al. (2020):

$$\text{Accuracy}_{ij} = 1 - \frac{SE^2}{(1 + fx)\, v_a}$$

where SE is the standard error for the animal solution i in trait j, fx corresponds to individual inbreeding and v_a is the additive variance σ_a^2 .

Comparison of Analysis

Results from different scenarios were compared based on descriptive statistics (i.e., mean and standard errors) of VC, Pearson's correlations between EBVs grouped by animal status (i.e., bulls with at least 10 daughters, buffalo cows with or without progeny), re-rankings of first 10 bulls, efficiency of selection (SEf) as defined later and genetic trends, estimated by the linear regression of EBVs on year of birth.

The SEf was calculated as proposed by Petrini et al. (2015) and Peškovičová et al. (2004), which defined SEf as the ratio between EBVs excluding ($\bar{x}_{gg\_GG}$) and including genetic groups ($\bar{x}_{GG\_GG}$):

$$\text{SEf } (\%) = 100 \times \bar{x}_{gg\_GG} / \bar{x}_{GG\_GG}$$

The SEf was calculated for the best 10, 30, and 50% animals, respectively.

Softwares

Data preparation and editing, and all statistical analysis were performed using the R programming environment v.3.6.1 (R Core Team, 2019), except VC which were estimated using AIREMLF90 (Misztal et al., 2002) and EBV which were obtained using BLUPF90 (Misztal et al., 2018). The R package *optiSel* (Wellmann, 2019) was used to calculate the pedigree-based additive relationship matrix and the package *stats* for the hierarchical cluster analysis (R Core Team, 2019). The analyses were run on the ANASB server[1] using an Intel® Pentium® CPU G3220 @ 3.00GHz, with 2 CPUs and 16 Gb of RAM.

[1]http://www.anasb.it

RESULTS

Data Overview

Descriptive statistics for the analyzed traits are shown in **Table 1**. The deviation from the normal distribution was moderate, with kurtosis values ranging from 0.03 to 2.07. Traits distribution was skewed to the right (**Supplementary Figure 2**) and the average coefficient of variation was 2.8 and 24.4% for composite and linear traits, respectively.

Variance Components and Heritability

The VC and heritability estimates from the different scenarios are shown in the **Tables 3**, **4** for composite and linear traits, respectively. The estimated genetic variance was highest for five linear traits (STAT, FUA, RUW, UD, and TL), intermediate for BD, BL, FA, and TP, while the lowest were for composite traits and BCS. On average, the estimates of the additive variances from the GOLD scenario were the highest, observing largest differences

TABLE 3 | Component of variance and hereditability for the composite traits obtained in the different pedigree scenario in the IMB.

Scenario	Parameter	FS	ST	FL	UT	YP
GOLD	$\sigma^2 a$	0.55	0.98	0.74	1.02	0.58
	$\sigma^2 e$	2.02	2.90	4.67	5.96	2.39
	$\sigma^2 p$	2.57	3.88	5.41	6.98	2.98
	h^2 ± s.e.	**0.22 ± 0.03**	**0.25 ± 0.03**	**0.14 ± 0.03**	**0.15 ± 0.03**	**0.20 ± 0.04**
Y30	$\sigma^2 a$	0.54	0.89	0.73	0.98	0.55
	$\sigma^2 e$	2.03	2.98	4.67	6.01	2.43
	$\sigma^2 p$	2.57	3.87	5.40	6.98	2.97
	h^2 ± s.e.	**0.21 ± 0.03**	**0.23 ± 0.03**	**0.14 ± 0.03**	**0.14 ± 0.03**	**0.18 ± 0.04**
Y60	$\sigma^2 a$	0.55	0.87	0.74	0.99	0.50
	$\sigma^2 e$	2.02	3.00	4.65	5.99	2.48
	$\sigma^2 p$	2.56	3.86	5.39	6.98	2.97
	h^2 ± s.e.	**0.21 ± 0.04**	**0.22 ± 0.04**	**0.14 ± 0.03**	**0.14 ± 0.03**	**0.17 ± 0.05**
GC30	$\sigma^2 a$	0.51	0.93	0.78	1.17	0.52
	$\sigma^2 e$	2.06	2.94	4.62	5.83	2.45
	$\sigma^2 p$	2.56	3.86	5.40	6.99	2.97
	h^2 ± s.e.	**0.20 ± 0.04**	**0.24 ± 0.04**	**0.14 ± 0.03**	**0.17 ± 0.03**	**0.18 ± 0.04**
GC60	$\sigma^2 a$	0.48	0.83	0.84	1.52	0.51
	$\sigma^2 e$	2.08	3.02	4.55	5.48	2.46
	$\sigma^2 p$	2.56	3.85	5.40	7.00	2.97
	h^2 ± s.e.	**0.19 ± 0.05**	**0.22 ± 0.05**	**0.16 ± 0.05**	**0.22 ± 0.05**	**0.17 ± 0.06**

$\sigma^2 a$ = additive genetic variance; $\sigma^2 e$ = residual variance; $\sigma^2 p$ = phenotypic variance; h2 = hereditability; s.e. = standard error.

TABLE 4 | Component of variance and hereditability for the linear traits obtained in the different pedigree scenario in the IMB.

Scenario	Parameter	STAT	BD	BL	FA	FUA	RUW	UD	TP	TL	BCS
GOLD	$\sigma^2 a$	9.33	4.44	4.90	2.89	6.64	6.21	7.69	2.53	10.46	0.030
	$\sigma^2 e$	17.01	19.19	16.20	28.37	31.34	23.47	22.64	16.62	29.35	0.159
	$\sigma^2 p$	26.34	23.63	21.10	31.26	37.98	29.68	30.33	19.16	39.81	0.189
	h^2 ± s.e.	**0.35 ± 0.03**	**0.19 ± 0.03**	**0.23 ± 0.03**	**0.09 ± 0.02**	**0.17 ± 0.03**	**0.21 ± 0.03**	**0.25 ± 0.03**	**0.13 ± 0.03**	**0.26 ± 0.03**	**0.16 ± 0.03**
Y30	$\sigma^2 a$	8.82	4.21	4.92	3.14	6.18	6.11	6.76	2.25	10.32	0.025
	$\sigma^2 e$	17.54	19.32	16.23	28.08	31.67	23.51	23.34	16.88	29.31	0.163
	$\sigma^2 p$	26.36	23.53	21.15	31.22	37.85	29.62	30.10	19.12	39.63	0.188
	h^2 ± s.e.	**0.33 ± 0.03**	**0.18 ± 0.03**	**0.23 ± 0.03**	**0.10 ± 0.03**	**0.16 ± 0.03**	**0.21 ± 0.03**	**0.22 ± 0.03**	**0.12 ± 0.03**	**0.26 ± 0.03**	**0.13 ± 0.03**
Y60	$\sigma^2 a$	8.50	4.20	4.92	3.14	5.98	5.65	6.63	2.39	10.25	0.026
	$\sigma^2 e$	17.96	19.26	16.24	28.07	31.75	23.89	23.29	16.71	29.20	0.162
	$\sigma^2 p$	26.45	23.46	21.15	31.21	37.73	29.55	29.92	19.11	39.45	0.188
	h^2 ± s.e.	**0.32 ± 0.04**	**0.18 ± 0.04**	**0.23 ± 0.04**	**0.10 ± 0.03**	**0.16 ± 0.04**	**0.19 ± 0.04**	**0.22 ± 0.04**	**0.13 ± 0.03**	**0.26 ± 0.04**	**0.14 ± 0.04**
GC30	$\sigma^2 a$	9.10	4.10	4.89	2.99	6.03	5.41	6.97	2.45	9.80	0.028
	$\sigma^2 e$	17.27	19.40	16.22	28.24	31.77	23.85	23.08	16.66	29.77	0.158
	$\sigma^2 p$	26.37	23.50	21.12	31.23	37.80	29.26	30.05	19.12	39.57	0.188
	h^2 ± s.e.	**0.35 ± 0.04**	**0.17 ± 0.03**	**0.23 ± 0.04**	**0.10 ± 0.03**	**0.16 ± 0.03**	**0.18 ± 0.04**	**0.23 ± 0.04**	**0.13 ± 0.03**	**0.25 ± 0.04**	**0.15 ± 0.04**
GC60	$\sigma^2 a$	9.93	3.48	5.41	3.09	5.33	5.39	6.61	2.60	9.72	0.026
	$\sigma^2 e$	16.55	19.95	15.75	28.10	32.38	24.10	23.30	16.51	29.70	0.161
	$\sigma^2 p$	26.48	23.42	21.16	31.19	37.71	29.49	29.91	19.12	39.43	0.187
	h^2 ± s.e.	**0.38 ± 0.05**	**0.15 ± 0.05**	**0.26 ± 0.05**	**0.10 ± 0.04**	**0.14 ± 0.05**	**0.18 ± 0.05**	**0.22 ± 0.05**	**0.14 ± 0.04**	**0.25 ± 0.05**	**0.14 ± 0.05**

$\sigma^2 a$ = additive genetic variance; $\sigma^2 e$ = residual variance; $\sigma^2 p$ = phenotypic variance; h2 = hereditability; s.e. = standard error.

with GC60 for the STAT (−0.60) and FUA (+1.31) traits, respectively.

Differences among heritability estimates varied according to the trait and the scenario considered and are presented in **Figure 1**. The green line identifies the heritability from the GOLD scenario. The largest differences were observed in the scenario GC60 for trait UT (0.22 vs. 0.15) and for trait BD (0.15 vs. 0.19). Moreover, GC60 showed the highest within-trait variability, with maximum differences for UT, BCS, and FS (0.39, 0.21, and 0.18, respectively), and minimum differences of 0.08 for RUW and STAT (result not show). Standard errors of heritabilities for all traits were low, ranging from 0.03 (GOLD) to 0.05 (GC60).

Correlations Between Breeding Values

The correlations between EBVs in the different scenarios are shown in **Table 5**. Results differed depending on sex: higher estimates were observed in the female population when using a grouping strategy based on the year of birth (Y), while for the bulls higher estimates were observed with the genetic clustering strategy (GC). On average, the correlations were

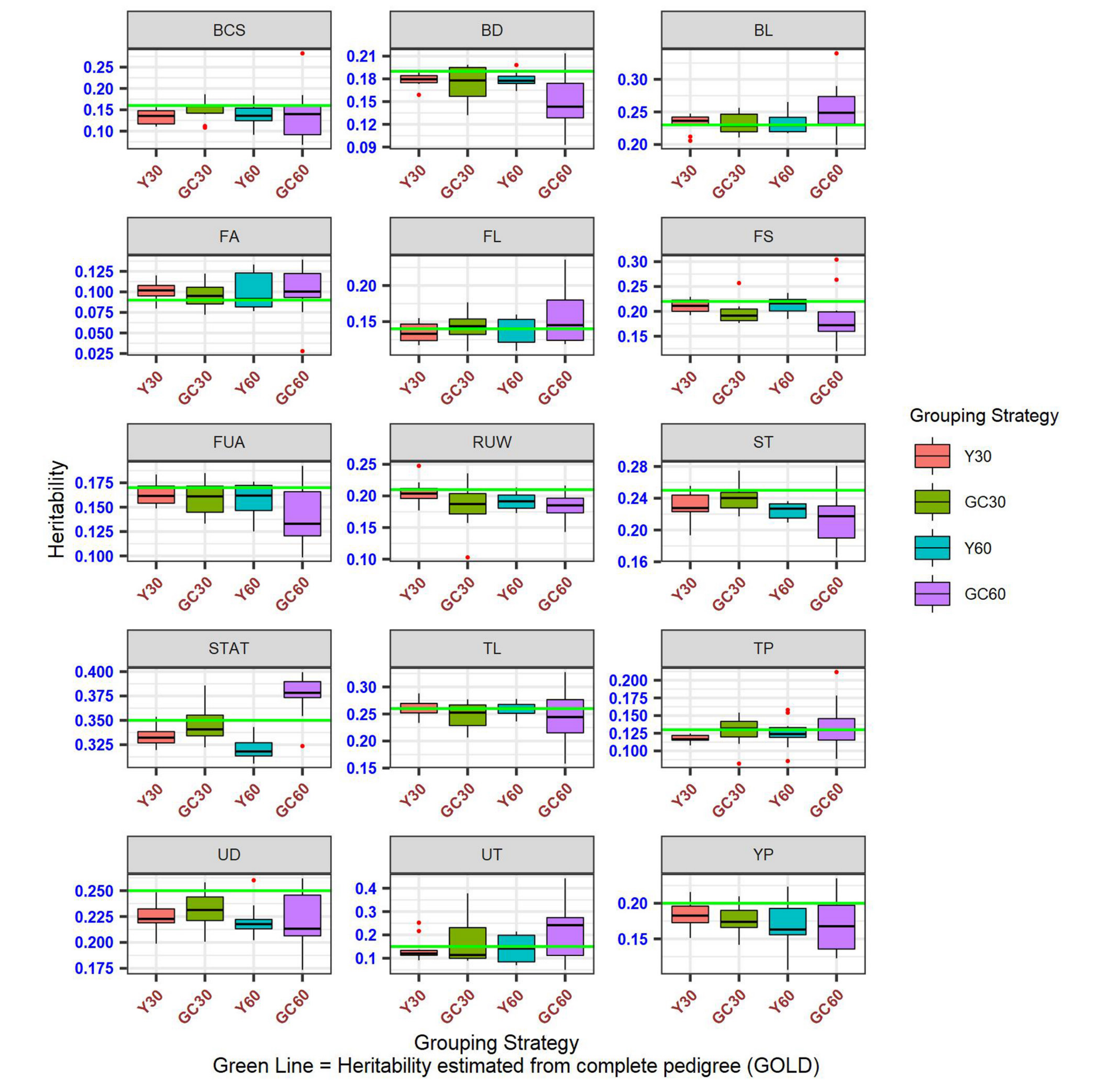

FIGURE 1 | Box plot of the hereditability for composite and linear traits obtained in the different pedigree scenario in the IMB.

TABLE 5 | Average correlations for buffalo cows and bulls' EBVs for the composite and linear traits obtained in the different pedigree scenario in the IMB.

Trait[a]	Female with records				Bulls			
	Y30	Y60	GC30	GC60	Y30	Y60	GC30	GC60
FS	0.92	0.85	0.89	0.80	0.94	0.86	0.93	0.85
ST	0.93	0.87	0.90	0.81	0.89	0.77	0.94	0.84
FL	0.88	0.80	0.88	0.77	0.90	0.70	0.93	0.77
UT	0.89	0.80	0.85	0.73	0.92	0.79	0.92	0.80
YP	0.88	0.78	0.84	0.71	0.91	0.75	0.92	0.85
STAT	0.95	0.89	0.93	0.87	0.95	0.87	0.95	0.89
BD	0.92	0.85	0.89	0.79	0.91	0.77	0.92	0.79
BL	0.92	0.84	0.90	0.81	0.94	0.85	0.95	0.87
FA	0.85	0.75	0.83	0.68	0.76	0.63	0.90	0.77
FUA	0.92	0.85	0.88	0.78	0.90	0.74	0.89	0.76
RUW	0.92	0.86	0.90	0.80	0.92	0.82	0.93	0.81
UD	0.95	0.90	0.92	0.86	0.87	0.74	0.90	0.74
TP	0.88	0.80	0.86	0.75	0.85	0.72	0.91	0.81
TL	0.95	0.90	0.92	0.86	0.88	0.77	0.92	0.77
BCS	0.89	0.82	0.87	0.77	0.76	0.59	0.90	0.78
Average	**0.91**	**0.84**	**0.88**	**0.79**	**0.89**	**0.76**	**0.92**	**0.81**

[a] *See* ***Table 1*** *for trait acronym.*

positive and high. Considering buffalo cows with records, the average correlation across traits between official EBVs and EBVs from different scenarios were 0.91, 0.88, 0.84, and 0.79 for Y30, GC30, Y60, and GC60, respectively. The best results were observed for STAT, UD, and TL (average $r = 0.91$) while the most affected trait was FA in the scenario GC60 ($r = 0.68$).

In the case of bulls, the correlation between EBVs in the grouping GC30 ranged from 0.90 for FUA to 0.96 for STAT and BL, while, in the GC60 scenario the values range between 0.75 for UD to 0.90 for STAT (**Table 5**). As expected, the highest correlations occurred in scenarios where the proportion of missing pedigree was lower (i.e., Y30 and GC30).

Accuracy of Breeding Values

The accuracy of breeding values across traits and scenarios for bulls with at least 10 daughters and buffalo cows with own record are presented in **Table 6**. The drop in accuracy for bulls ranged from 0.06 for stature in the scenario GC30 to 0.24 for YP in the scenario Y60. Similar pattern was observed in buffalo cows, with higher accuracies in the Y30 and GC30 scenarios. On average the best results were shown by GC30 (average accuracy = 0.43) and Y30 (average accuracy = 0.42), while the worst results were in the scenario GC60 (average accuracy = 0.34) and Y60 (average accuracy = 0.32) (**Figure 2**).

Selection Efficiency

The result of the average selection efficiency for the three different selection intensities (top 10, 30, and 50%) for composite and linear trait are summarized in the **Table 7**. Average of SEf ranged from 22.12 (Top 50 for FL in GC60 scenario) to 85.94% (Top 10 for FS in GC30 scenario) for the composite trait, and from 17.09 (Top 50 for FA in Y60 scenario) to 88.80% (Top 10 for STAT in GC30) for linear traits.

Observing the average intensity of selection across scenarios, the highest value was in GC30 (81.27%), followed by 78.75, 67.41, and 65.22% in Y30, GC60, and Y60, respectively. The average intensity of selections for the best 10, 30, and 50% were 73.16, 60.40, and 42.31%, respectively.

Within each scenario, selection efficiency in composite traits was more effective than in linear traits. When the best 10% of individuals were selected, four out of five composite traits had a

TABLE 6 | Average accuracy buffalo cows and bulls' EBVs for the composite and linear traits obtained in the different genetic group in the IMB.

Trait[a]	GOLD		Y30		Y60		GC30		GC60	
	Bulls	Female	Bulls	Female	Bulls	Female	Bulls	Female	Bulls	Female
FS	0.55	0.29	0.47	0.24	0.37	0.21	0.46	0.21	0.33	0.17
ST	0.58	0.32	0.47	0.24	0.38	0.22	0.50	0.26	0.39	0.20
FL	0.47	0.22	0.39	0.20	0.28	0.13	0.40	0.16	0.30	0.13
UT	0.48	0.23	0.33	0.14	0.28	0.13	0.37	0.17	0.35	0.19
YP	0.46	0.21	0.34	0.15	0.22	0.10	0.36	0.13	0.28	0.11
STAT	0.64	0.39	0.56	0.34	0.45	0.29	0.58	0.35	0.49	0.33
BD	0.52	0.26	0.42	0.20	0.34	0.17	0.42	0.18	0.31	0.13
BL	0.56	0.30	0.47	0.24	0.38	0.21	0.48	0.23	0.42	0.24
FA	0.39	0.17	0.30	0.11	0.21	0.08	0.3	0.10	0.26	0.09
FUA	0.51	0.25	0.41	0.19	0.30	0.14	0.41	0.17	0.30	0.12
RUW	0.54	0.28	0.44	0.22	0.35	0.18	0.45	0.21	0.33	0.16
UD	0.58	0.32	0.46	0.24	0.37	0.21	0.50	0.25	0.39	0.21
TP	0.46	0.21	0.34	0.14	0.23	0.10	0.37	0.14	0.31	0.13
TL	0.59	0.33	0.49	0.26	0.42	0.25	0.51	0.26	0.41	0.23
BCS	0.49	0.24	0.33	0.13	0.28	0.13	0.38	0.14	0.28	0.11
Average	**0.52**	**0.27**	**0.42**	**0.20**	**0.32**	**0.17**	**0.43**	**0.20**	**0.34**	**0.17**

[a] *See* ***Table 1*** *for trait acronym.*

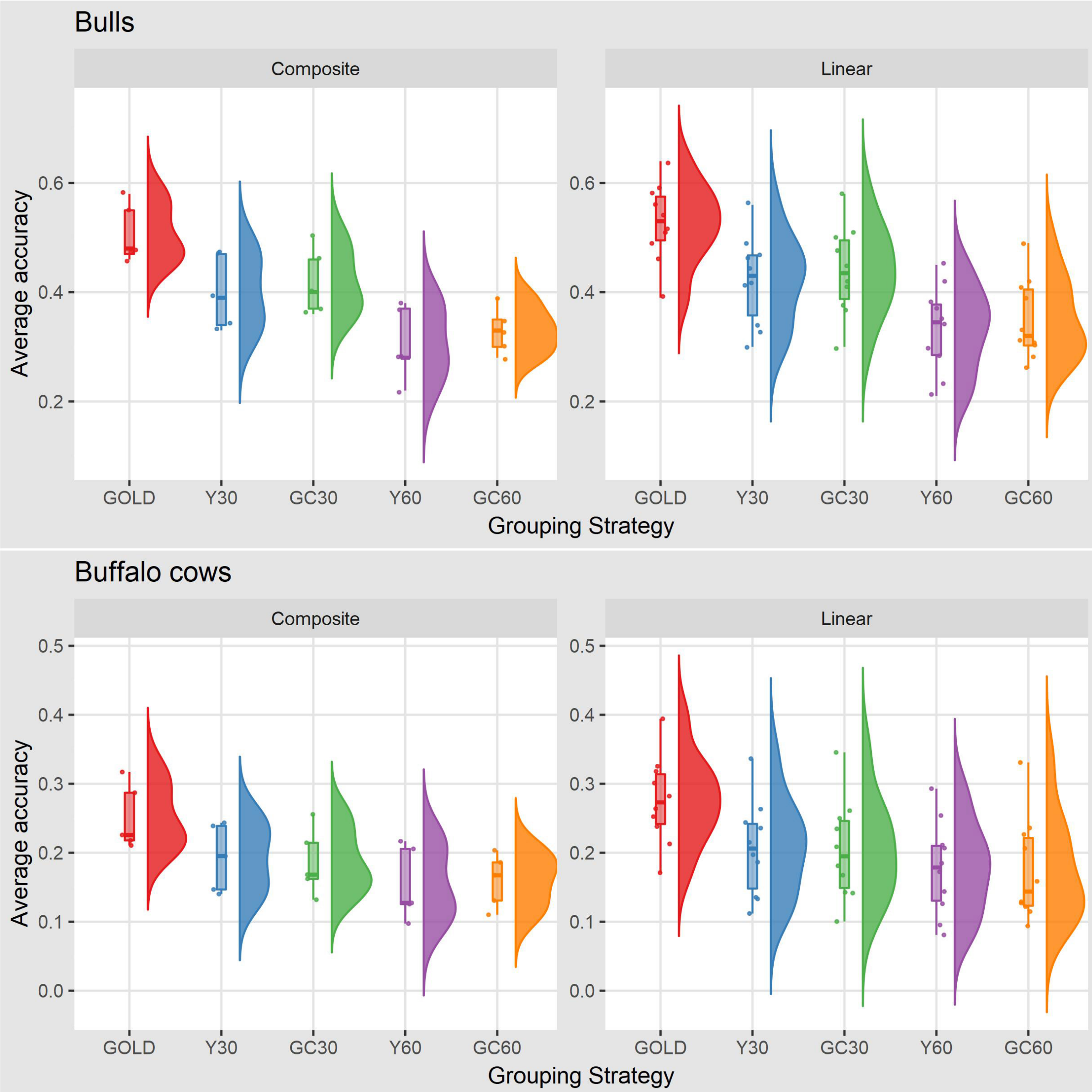

FIGURE 2 | Box plot and histogram of average accuracy for the composite and linear traits by sex obtained in the different genetic group in the IMB.

selection efficiency higher than 60%, while only three out of 10 linear traits exceeded such a threshold (**Table 7**). A similar trend was observed selecting 30% (3/5; 4/10 ≥ 50.01%) or 50% (3/5; 4/10 ≥ 32.91%).

In terms of standard deviation, the GC30 scenario showed the lowest standard deviation (average = 4.61), while the values obtained from GC60 and Y60 tend to be higher, with an average SD of 7.94 and 7.82, respectively.

Re-Ranking

The effect of the different genetic grouping strategies on the ranking of the bulls was explored using only three linear traits, with high, medium, and low heritability, namely STAT (h^2 = 0.35), UD (h^2 = 0.23), and FA (h^2 = 0.10). Spearman's rank correlation calculated on 111 bulls in STAT-UD-FA were 0.921–0.884–0.842, 0.913–0.852–0.728, 0.846-0.695-0.659, and 0.811-0.690-0.587 for GC30, Y30, GC60, and Y60, respectively. The consistency of ranking across grouping strategy can also be effectively visualized with a target plot (Biscarini et al., 2016). The rankings of the first 10 bulls across replicates and grouping strategy for STAT, UD, and FA are presented in **Figures 3–5**, respectively. Each cloud of points represents the ranking of the bull across replicates and within grouping strategy. When the points within the clouds are more dispersed, a larger re-ranking was observed (e.g., BULL9 for STAT trait).

TABLE 7 | Mean (SD) of efficiency (%) in the selection of the best animals for the composite and linear traits obtained in the different pedigree scenario in the IMB.

Trait[a]	Best	Y30	Y60	GC30	GC60
FS	*10%*	85.11 (4.39)	76.59 (5.96)	85.94 (3.64)	78.64 (5.26)
	30%	79.40 (2.77)	67.42 (4.21)	78.47 (2.59)	65.03 (6.70)
	50%	58.02 (3.89)	49.24 (5.10)	58.36 (5.12)	47.36 (4.51)
ST	*10%*	85.90 (2.94)	71.76 (5.51)	85.56 (2.31)	74.16 (6.45)
	30%	75.29 (3.37)	60.79 (6.84)	76.88 (4.62)	59.54 (4.54)
	50%	61.40 (4.61)	47.40 (4.23)	61.42 (5.33)	45.17 (4.57)
FL	*10%*	77.56 (7.12)	52.94 (7.27)	81.86 (4.94)	59.99 (9.58)
	30%	65.24 (10.81)	39.40 (8.31)	69.17 (5.18)	39.45 (11.85)
	50%	45.27 (12.22)	23.61 (6.33)	51.17 (3.24)	22.12 (8.49)
UT	*10%*	83.86 (4.50)	75.82 (7.74)	82.08 (7.25)	66.32 (9.99)
	30%	75.59 (2.78)	59.06 (8.66)	72.53 (5.06)	53.05 (10.68)
	50%	55.33 (3.93)	42.65 (8.12)	50.11 (6.21)	36.07 (5.85)
YP	*10%*	81.20 (6.07)	65.42 (9.91)	81.62 (4.73)	73.81 (5.19)
	30%	67.99 (6.35)	48.52 (8.86)	61.37 (6.64)	54.01 (8.20)
	50%	47.39 (7.15)	32.15 (5.96)	43.55 (4.82)	36.29 (7.72)
STAT	*10%*	88.20 (3.48)	78.17 (4.52)	88.80 (2.79)	75.53 (8.75)
	30%	78.07 (2.95)	65.43 (6.32)	77.66 (5.48)	64.95 (5.46)
	50%	57.01 (3.31)	45.98 (6.61)	53.91 (4.30)	43.90 (6.23)
BD	*10%*	79.56 (4.03)	65.32 (11.87)	80.48 (4.37)	63.23 (9.87)
	30%	66.58 (4.16)	45.51 (12.36)	64.56 (6.59)	45.16 (6.68)
	50%	47.75 (3.74)	30.42 (8.87)	42.74 (6.22)	26.65 (5.52)
BL	*10%*	87.32 (5.02)	75.34 (6.01)	87.00 (3.77)	78.22 (5.52)
	30%	76.58 (5.90)	63.58 (5.40)	75.95 (4.08)	65.03 (9.02)
	50%	51.46 (4.69)	40.51 (5.15)	52.71 (5.65)	43.08 (7.51)
FA	*10%*	72.68 (7.42)	57.76 (10.66)	77.91 (6.54)	58.96 (13.58)
	30%	41.43 (6.45)	28.49 (12.88)	58.62 (7.53)	38.32 (12.55)
	50%	26.53 (6.16)	17.09 (11.12)	42.98 (7.74)	22.99 (7.66)
FUA	*10%*	78.37 (5.20)	63.29 (9.79)	75.38 (2.78)	66.12 (7.49)
	30%	69.73 (6.10)	51.93 (7.88)	67.30 (5.10)	52.37 (8.48)
	50%	49.77 (8.29)	35.40 (6.14)	51.03 (2.96)	37.75 (7.66)
RUW	*10%*	78.26 (7.49)	66.44 (7.90)	83.74 (3.74)	68.61 (5.45)
	30%	74.16 (7.34)	59.74 (4.37)	74.46 (4.55)	58.60 (4.38)
	50%	50.57 (9.91)	38.05 (6.88)	58.88 (4.28)	40.27 (4.11)
UD	*10%*	74.12 (4.78)	61.15 (4.64)	78.28 (5.91)	60.84 (7.90)
	30%	62.03 (7.14)	45.55 (5.72)	66.55 (5.47)	45.43 (8.68)
	50%	41.64 (4.63)	25.30 (6.08)	47.77 (6.73)	31.05 (5.62)
TP	*10%*	74.87 (4.63)	63.73 (9.03)	79.06 (3.56)	71.11 (9.27)
	30%	59.99 (4.50)	43.66 (11.04)	69.87 (4.91)	58.61 (7.78)
	50%	40.42 (3.30)	30.73 (9.10)	51.78 (4.33)	41.76 (3.32)
TL	*10%*	76.27 (4.41)	57.14 (7.85)	75.23 (4.48)	57.49 (7.14)
	30%	66.10 (3.88)	46.45 (7.27)	66.49 (4.77)	44.98 (7.09)
	50%	45.97 (4.30)	29.81 (8.24)	48.94 (4.28)	30.68 (5.25)
BCS	*10%*	58.02 (6.73)	47.45 (8.57)	76.07 (8.29)	58.15 (7.70)
	30%	59.16 (5.72)	37.89 (8.57)	67.05 (4.03)	51.64 (7.30)
	50%	40.14 (7.26)	23.36 (8.73)	51.47 (5.14)	36.17 (6.06)

[a]*See **Table 1** for trait acronym.*

Genetic Trend

The genetic trends for both composite and linear traits are presented in **Figures 6**, **7**. Overall a flat trend was observed until year 2013 for all traits. After this year, positive trends were observed and differences among years were enhanced on including genetic groups. For composite traits, an underestimation of the genetic trend was observed when the GC30 and GC60 grouping strategies were used.

Specific behaviors were detected across linear traits. Genetic trends for STAT, FUA, and TL showed the same pattern as the composite traits. BD and BL showed an uneven trend, with a clear positive trend from year 2014. However, when using GC30 and GC60 grouping strategies, EBVs were more regressed than when EBVs were estimated using a grouping strategy based on the year of birth. Similar results were observed for FA and UD where, particularly for recent years, Y30 and Y60 EBVs were higher than GC30 and GC60 EBVs. Finally, BCS showed a flat trend until 2014 followed by a slight decrease, a pattern common to all grouping strategies.

The different grouping strategies have had an impact on the EBVs scale. From year 2000 the average increase in the scenario without genetic groups (GOLD) was +0.032 for composite traits and +0.014 for linear traits (**Figure 8**). The average increase in composite traits was +0.046, +0.042, +0.026, and +0.020 when the Y30/Y60/GC30/GC60 genetic group was used, respectively. The same order was observed in the linear trait set with an average increase of +0.020, +0.018, +0.009, and +0.006.

DISCUSSION

In this study, the effect of two genetic grouping strategies on the estimation of VC and EBV for type traits in a parentage-tested IMB sub-population was evaluated. In the last three years the IMB has experienced an exponential increase in term of registered animals in the Herd Book. As a consequence, IMB is facing a situation where phenotypic data are available for many animals, but some animals lack complete genealogical data. Records from individuals without pedigree information has been excluded from the genetic evaluation or assumed to have an unknown sire. Such practice results in loss of information and potentially could compromise expected genetic gain (Sapp et al., 2007). To mitigate this undesirable effect, several statistical methods have been developed over the years. The use of genetic grouping, parentage probabilities, use of phenotypic information to increase the probability of defining the paternity, iterative empirical Bayesian model (ITER), Bayesian hierarchical model (HIER), and model based on the average relationship matrix (ANRM), have been applied to account for uncertain paternity (Henderson, 1988; Peškovičová et al., 2004; Sapp et al., 2007; Petrini et al., 2015; Carneiro et al., 2017; Shiotsuki et al., 2018; Macedo et al., 2020).

Genetic groups are normally created according to different criteria, for example on the basis of origin, sex, herd, or year of birth of the individual. The creation of the GG is not a simple procedure and can sometimes present some practical problems. Genetic groups modeling must be balanced as groups with few animals might impair the estimation of the GG effect (Rodriguez et al., 1996; Peškovičová et al., 2004; Petrini et al., 2015). At the same time, very large groups are not able to capture the actual differences which exist among individuals. However, (Quaas, 1988) warned about potential bias in defining a determinate

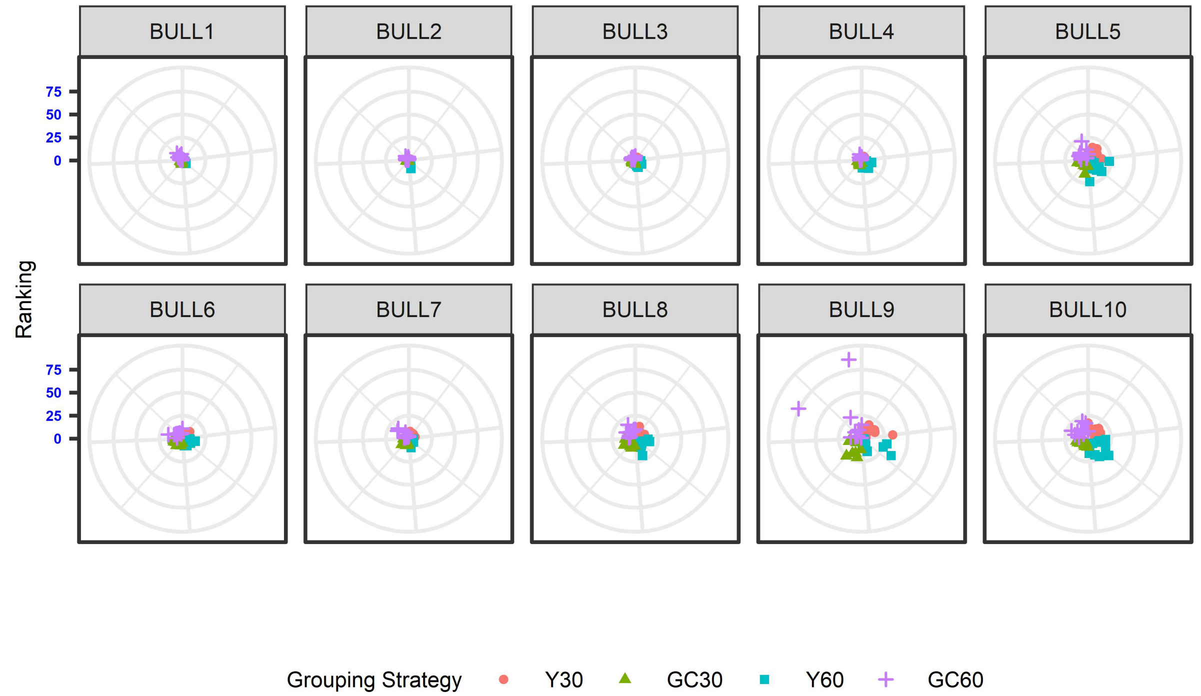

FIGURE 3 | Ten best ranked bulls for the Stature trait according to the different genetic group in the IMB. When the points within the clouds are more dispersed, a larger re-ranking was observed.

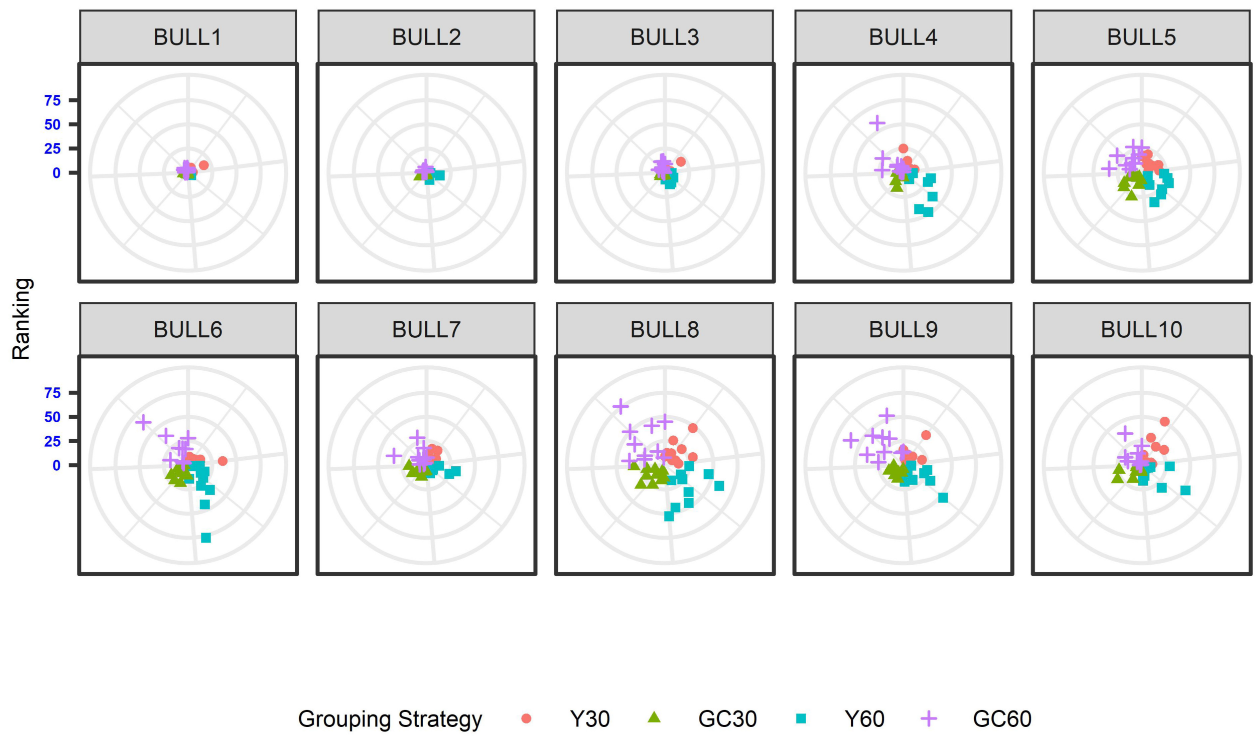

FIGURE 4 | Ten best ranked bulls for the Udder depth trait according to the different genetic group in the IMB. When the points within the clouds are more dispersed, a larger re-ranking was observed.

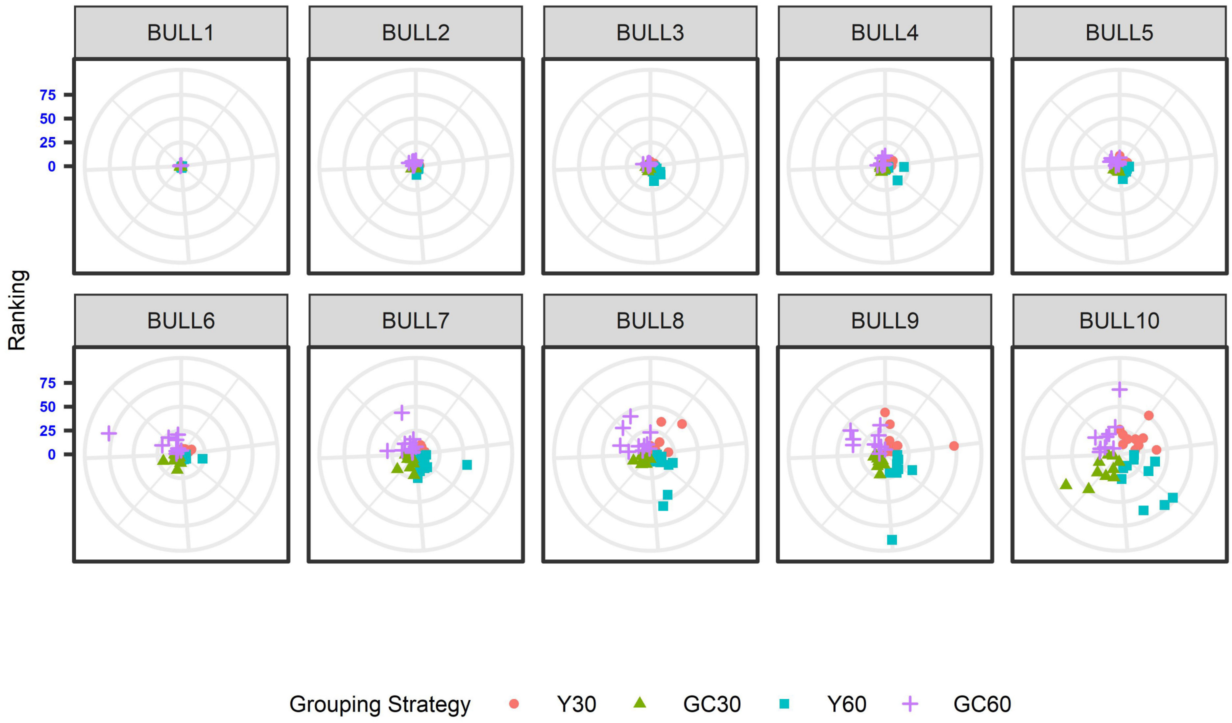

FIGURE 5 | Ten best ranked bulls for the Foot angle trait according to the different genetic group in the IMB. When the points within the clouds are more dispersed, a larger re-ranking was observed.

grouping strategy due to the effects of confusion between groups. In our case, the "phantom" parents of an individual are always assigned to the same group, because the grouping is based on animal itself, not on its parents, as shown by other studies (Peškovičová et al., 2004; Shiotsuki et al., 2013; Petrini et al., 2015; Wolak and Reid, 2017).

Results have shown that including GG in the mixed model equation had an effect on the estimates of both VC, which can be observed in **Tables 3**, **4**, and EBV (**Table 5**). Pieramati and Van Vleck (Pieramati and Van Vleck, 1993) obtained lower estimates of additive genetic variance with models that included genetic group. However, we have found that the estimates of VC and EBV with the Y30 and GC30 genetic groups are quite close to the GOLD estimates. These results support the efficiency of the methodology to estimate the true parameters. According to the magnitude of heritability estimates, the GC60 scenario was the one that showed the largest discrepancy with GOLD, confirmed by the highest SE (0.05). Petrini et al. (2015) suggested that such result may be caused by the structure of the group itself. Indeed, the size of GG should be homogeneous and well balanced. In the present study, when a genetic clustering strategy was used, a greater number of groups with a more heterogeneous size was observed. These results depend on the pedigree structure of the IMB, because its completeness is mainly related to the use of artificial insemination. Bulls used for AI have a more complete pedigree both on paternal and maternal side. The fourteen groups used in the GC strategy (**Table 2**) are based on the relationship matrix and hence are strictly related to the completeness of the paternal line. Indeed in the GC scenario we had a particular group – namely group 1 – which basically included all individuals with no pedigree information and whose size was from 10 to 20-fold larger than the others. Those evidences matched results from Santana et al. (2013) and Shiotsuki et al. (2013) who stressed the importance of the structure of the groups, especially in terms of their number and size (Petrini et al., 2015).

As expected, EBVs accuracy decreased when an increased proportion of missing pedigree was simulated (**Table 6**). However, when the proportion of missing pedigree was 30%, the average percentage point drop in accuracy was 10 and 7 for bulls and buffalo cows, respectively. We can therefore hypothesize that the contemporary use of the available pedigree information and of the most appropriate GG strategy will mitigate the loss in accuracy of the EBV due to missing pedigree information. Sullivan (1995) suggested the importance of the inclusion of genetic groups in EBV estimation and that data should not be discarded due to the uncertainty of the paternities. Surely, the problem of uncertain paternities might possibly be mitigated by the use of genomic selection (Abdel-Shafy et al., 2020; Macedo et al., 2020; Misztal et al., 2020), however, the genotyping of all animals in a herd might still be too expensive. In the case of IMB, the use of GG is a practical and no cost solution to integrate all the available information into the genetic evaluations process eventually not compromising the accuracy of the results.

On the other hand, Pearson's correlations between EBVs were generally high in all clustering scenarios. However, Y30 and GC30 scenarios showed the highest correlations. Several studies have

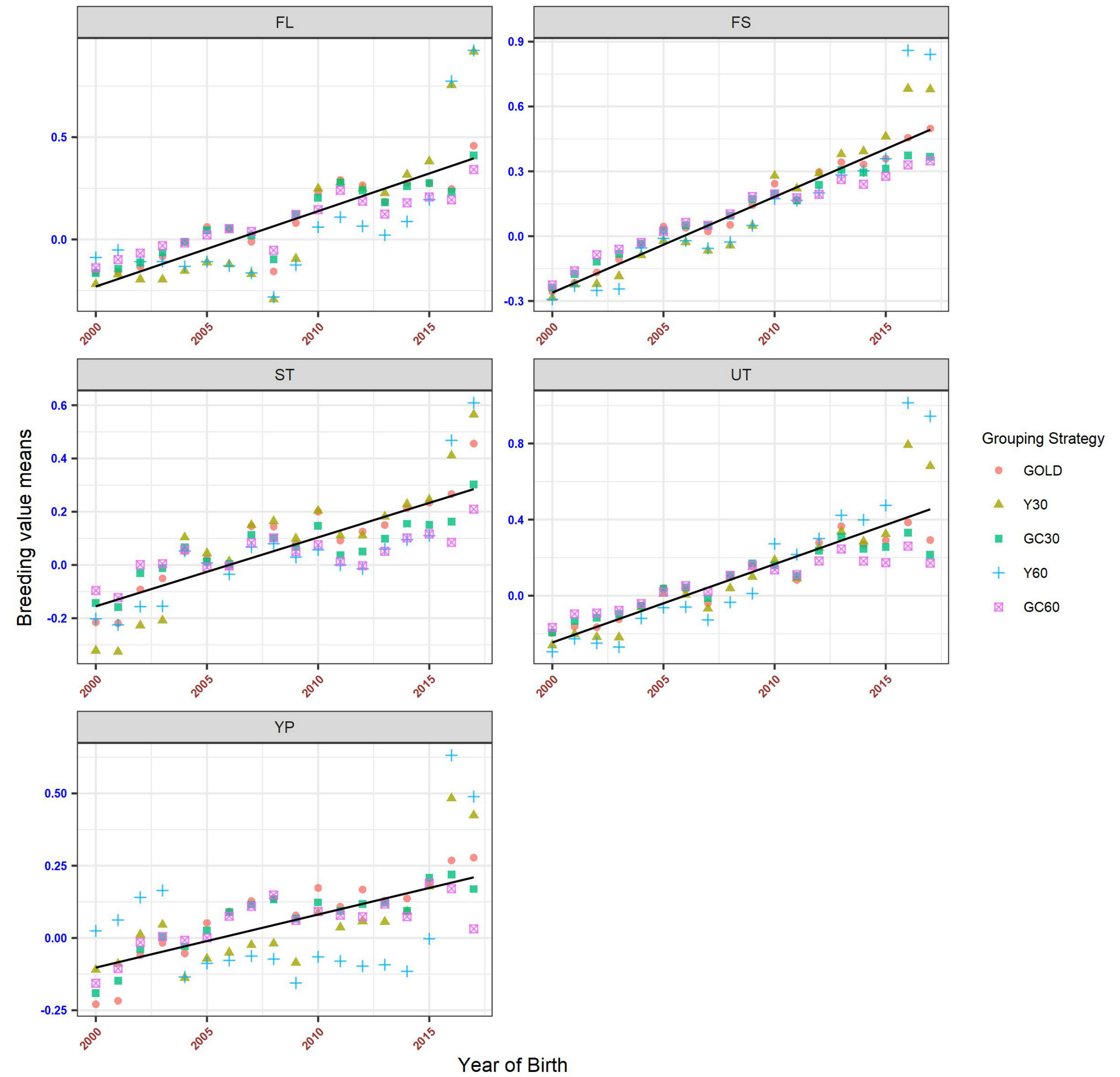

FIGURE 6 | Genetic trend by year of birth for the composite trait, according to the different genetic group.

shown that correlation coefficients between EBVs lower than 0.70 could suggest changes in the classification of animals (Crews and Franke, 1998; Petrini et al., 2015). Moreover, if we analyze results within traits, we can observe a relationship with heritability value. In our case, the trait that had the lowest correlation coefficient (r = 0.68) was FA, whose h^2 was 0.10. In addition, observing the correlations within sex, the Y30/Y60 genetic group strategy showed the highest coefficients for buffalo cows, while for bulls GC30 was the most appropriate for the data. This result was somewhat expected because the strategy based on the hierarchical clustering is strictly related to the relationship matrix, i.e., on the pedigree information. The number of AI bulls in the IMB population is limited (n < 100) and most of them have common ancestors. This means that grouping based on the relationship matrix will be possibly biased by the sire's pedigree. Actually, all individuals with both parents missing have been assigned to group 1 (**Table 2**), possibly regressing their breeding value. On the other hand, the year of birth has a more balanced behavior and it is less linked to the pedigree. Therefore, our results suggest that the EBV and consequently the ranking of the animals, will be closely influenced by the nature of the trait and by the structure and type of grouping adopted (Shiotsuki et al., 2018).

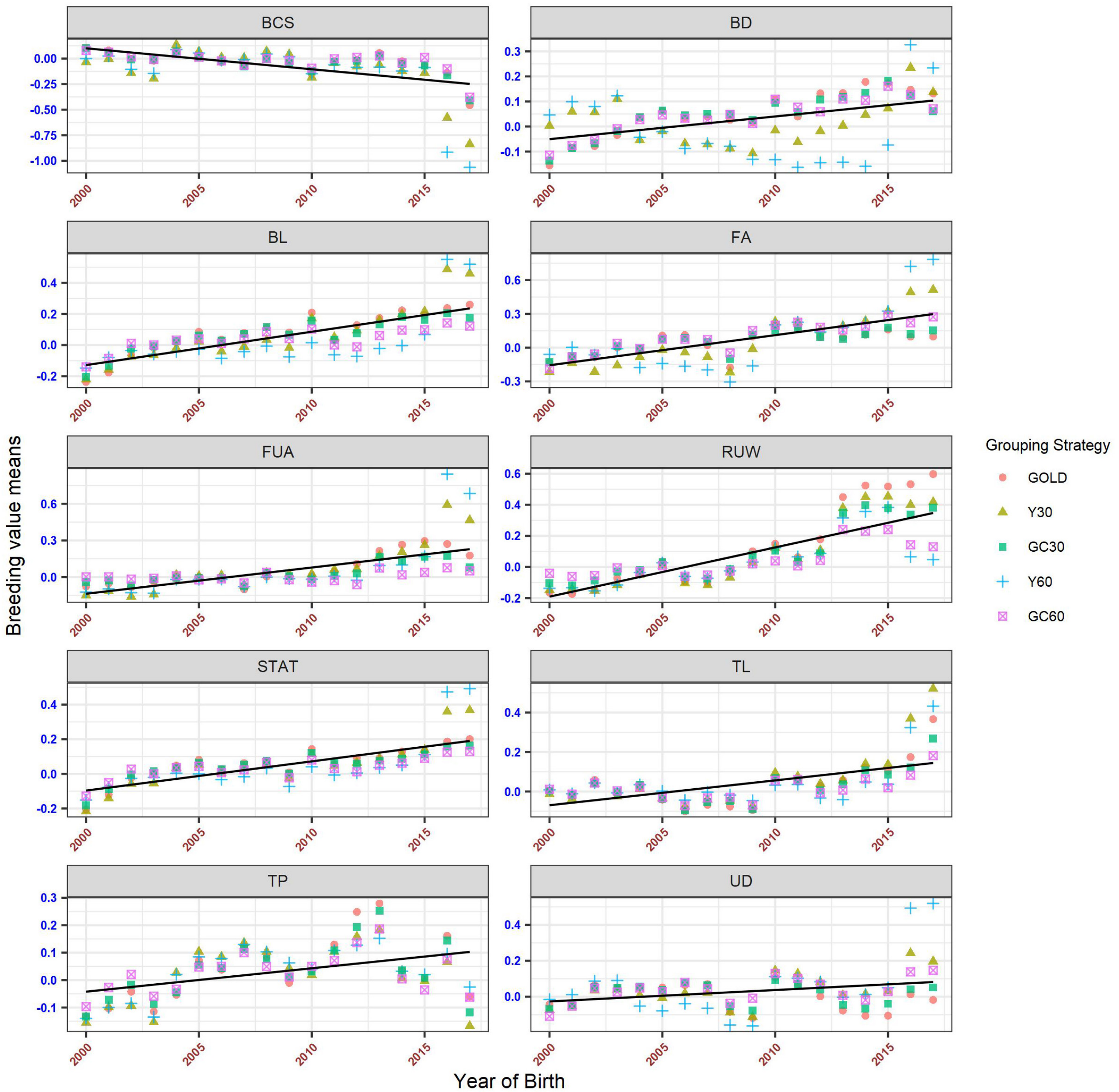

FIGURE 7 | Genetic trend by year of birth for the linear trait, according to the different genetic group.

Considering SEf, several studies suggest that it can be used as a measure of the correlation between the ranking of the best animals obtained in the different analyzes and that would in turn provide information on the degree of efficacy of the genetic grouping strategy (Theron et al., 2002; Peškovičová et al., 2004; Petrini et al., 2015). A value above 70% would indicate that the ranking observed in the different scenarios is stable and does not undergo a significant re-ranking. In relation to what we observed in this study, when the selection intensity is 10%, practically all traits exceeded this threshold (14/15 traits in Y30 and 15/15 in GC30). Meanwhile, in the scenario where the proportion of missing of pedigrees was 60% only 5/15 traits showed a value of SEf higher than 70%. These results suggest that bulls that are above the 90th percentile would experience virtually no important changes in their ranking. Another aspect worth noticing is the standard deviation of SEf. If a large standard deviation is observed, the response to selection will be more unstable and less accurate (Peškovičová et al., 2004). In this regard, the genetic group GC30 showed the lowest standard deviation while results obtained from GC60 and Y60 were more unstable. Consequently, when considering a high correlation and SEf, in addition to a low SD, we retain that the ranking of the bulls will be consistent.

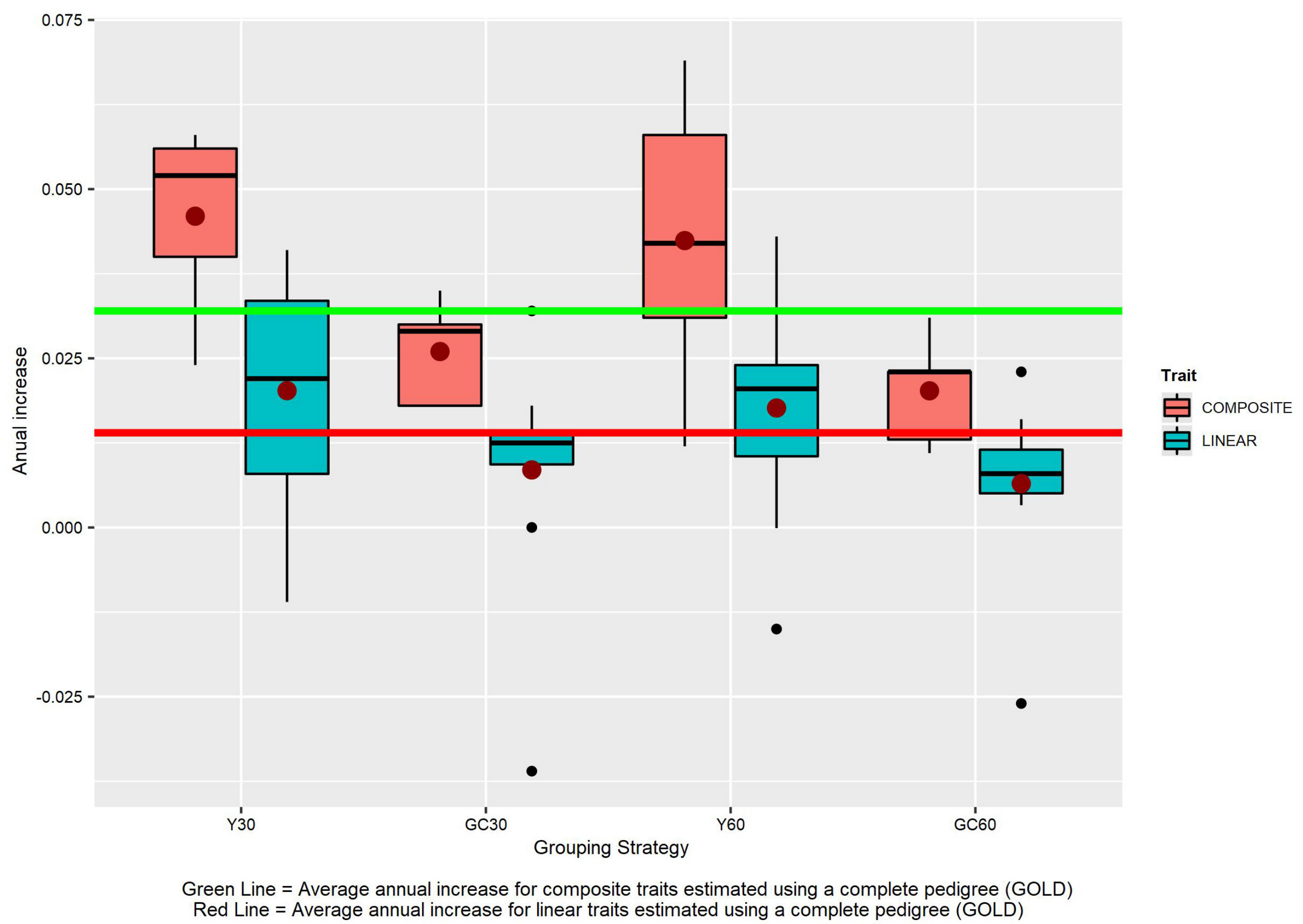

FIGURE 8 | The effect of different grouping strategy on EBVs scale. Average annual increase for the composite and linear trait EBVs, according to the different genetic group.

The inclusion of GG in the genetic evaluation could have unpredictable but substantial effects on the estimated genetic trend (Saavedra, 2019). Furthermore, the exclusion of genetic groups or having paternities with "phantom" parents could lead to biased estimates of selection response (Theron et al., 2002). In our study, these expectations are met, observing how the cumulative genetic trends without genetic groups were slightly lower than those estimated with the Y30/Y60 genetic group. Upward trends may indicate that the grouping type "year of birth" may be comparable to those obtained in GOLD. Other study, obtained some indication that the best strategy was grouping phantom sires according to the year of birth and the phantom dams in a single group due to the slow genetic change in females over the generations (Casellas et al., 2007). Theron et al. (2002) and Shiotsuki et al. (2013) observed higher genetic trends when they included GG in the analyses. Those results did not agree with (Petrini et al., 2015) where the inclusion of GG in genetic analyses showed a lower genetic trend.

The effectiveness of including GG on genetic evaluation depends on the genetic structure of the population, the nature of the observed trait (Petrini et al., 2015) and the criterion adopted to define GG. Several authors recommended that the definition of the GG should be a balance between complexity of the method and the adequate representation of genetic differences (Rodriguez et al., 1996; Peškovičová et al., 2004; Petrini et al., 2015; Carneiro et al., 2017; Shiotsuki et al., 2018). The adoption of an inappropriate method may not only have consequences on genetic progress (at the population level), but also on the choice of the best animals that will be used at the herd level. On the other hand, a change in the pedigree structure tends to have a higher impact on traits with medium-low heritability. In our study, this fact occurred with the FA trait, where GC30/GC60 scenarios had the largest correlation with GOLD. On the other hand, for traits with high heritability, the weight of the phenotypic information is high, therefore, the use of GG would have a lesser effect on the estimates. According to Cardoso and Tempelman (Cardoso and Tempelman, 2003), differences between the models that take into account uncertain paternity do not necessarily increase with increasing heritability, but these differences will be greater for the traits of medium-low heritability. In addition, individuals that have a greater number of ancestors or progeny with an incomplete pedigree will be more affected, in particular young animals with no own phenotypic information.

The lack of pedigree information is a common problem among domestic species, being more pronounced in less represented breeds that are mainly managed by small farmers with scarce economic resources. Resolving the uncertainty of paternity has always been a topic of interest to the scientific community and for decades various methodologies have been developed that allow managing the presence of gaps in a relationship matrix. Nowadays, there are different tools to improve the knowledge of genealogical information, such as DNA-based methods, but these are still expensive for breeders. Likewise, in those species that have recently implemented the genetic evaluation system they may face this problem, as they may be in the situation where they possess historical phenotypic data from which it is almost impossible to obtain biological samples due to the absence of a DNA banks.

The prediction of the genetic value with models that consider the uncertainty in paternity have been shown to have better precision (Cardoso and Tempelman, 2003; Sapp et al., 2007; Shiotsuki et al., 2012; Shiotsuki et al., 2013; Carneiro et al., 2017; Shiotsuki et al., 2018). Its effectiveness depends on the definition of the grouping strategy (Petrini et al., 2015), which requires prior knowledge of: (a) the selection process of the breed, (b) the sources of genetic variation present in the population, (c) the intensity of selection or the generational interval. It is clear that GG should be included in the model to improve the accuracy of the EBV of animals with some degree of unknown paternity (Saavedra, 2019). Therefore, the use of genetic groups can be considered an effective alternative in the absence of relationship data for VC and EBV.

CONCLUSION

Pedigree completeness is a fundamental requirement of any genetic evaluation. In species other than dairy cattle, the presence of individuals with phenotypic records but with an incomplete pedigree is not a trivial matter. Buffalo breeding is an example of such a situation. We do expect a more extended use of DNA testing which will eventually increase the implementation of genomic selection approaches in Buffalo species as well. However, missing information in the pedigree will still be present and even genomic selection will be faced with the same problem. When a variable proportion of missing pedigree information is present in a population under selection, including genetic groups in the mixed model equations for both VC and EBV estimation is a worth-while and low-demanding approach to mitigate the loss in accuracy. Different strategies can be used to create genetic grouping depending on data distribution across years and on population structure. In the IMB population the best results were obtained when grouping was based on the year of birth. These findings confirmed the possibility of developing a genetic evaluation in populations with uncertain paternities without the need to exclude data or to use only a select of the available population.

DATA AVAILABILITY STATEMENT

The data analyzed in this study was obtained from Italian National Association of Buffalo Breeders (ANASB). Requests to access these datasets should be directed to d.rossi@anasb.it

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because Animal welfare and use committee approval was not needed for this study as datasets were obtained from pre-existing databases based on routine animal recording procedures.

AUTHOR CONTRIBUTIONS

SB and MGC conceived and designed the work. DR, RC, GZ, YG, and DA were responsible for updating and editing the data. SB, RDP, and MGC contributed to analyzing the data and interpreting the results. MGC and SB wrote the manuscript with input from all the authors. All authors revised the manuscript, contributed to the article, and approved the submitted version.

ACKNOWLEDGMENTS

We would like to acknowledge Dr. Francesco Tiezzi of the Department of Animal Science, North Carolina State University, Raleigh 27695 for his comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.625335/full#supplementary-material

REFERENCES

Abdel-Shafy, H., Awad, M. A. A., El-Regalaty, H., Ismael, A., El-Assal, S. E.-D., and Abou-Bakr, S. (2020). A single-step genomic evaluation for milk production in Egyptian buffalo. *Livest. Sci.* 234:103977. doi: 10.1016/j.livsci.2020.103977

Agudelo-Gómez, D., Pineda-Sierra, S., and Cerón-Muñoz, M. F. (2015). Genetic evaluation of dual-purpose buffaloes (*Bubalus bubalis*) in colombia using principal component analysis. *PLoS One* 10:e0132811. doi: 10.1371/journal.pone.0132811

Aguilar, I., Fernandez, E. N., Blasco, A., Ravagnolo, O., and Legarra, A. (2020). Effects of ignoring inbreeding in model-based accuracy for BLUP and SSGBLUP. *J. Anim. Breed Genet.* 137, 356–364. doi: 10.1111/jbg.12470

Associazione Nazionale Allevatori Specie Bufalina (ANASB) (2020). *Dati ANASB*. Available online at: http://www.anasb.it/statistiche/ (Accessed May 11, 2020)

Biscarini, F., Schwarzenbacher, H., Pausch, H., Nicolazzi, E. L., Pirola, Y., and Biffani, S. (2016). Use of SNP genotypes to identify carriers of harmful recessive mutations in cattle populations. *BMC Genomics* 17:857. doi: 10.1186/s12864-016-3218-9

Boselli, C., De Marchi, M., Costa, A., and Borghese, A. (2020). Study of milkability and its relation with milk yield and somatic cell in mediterranean italian water buffalo. *Front. Vet. Sci.* 7:432. doi: 10.3389/fvets.2020.00432

Cardoso, F. F., and Tempelman, R. J. (2003). Bayesian inference on genetic merit under uncertain paternity. *Genet. Sel. Evol.* 35, 469–487. doi: 10.1186/1297-9686-35-6-469

Carneiro, F. F. D., Lôbo, A. M. B. O., Silva, L. P. D., Silva, K. D. M., Landim, A. V., and Lôbo, R. N. B. (2017). Genetic evaluation is possible on community pastoral small ruminant flocks in the presence of multiple sires and uncertain of paternity. *Small Ruminant Res.* 151, 72–81. doi: 10.1016/j.smallrumres.2017.04.017

Casellas, J., Piedrafita, J., and Varona, L. (2007). Bayes factor for testing between different structures of random genetic groups: a case study using weaning weight in Bruna dels Pirineus beef cattle. *Genet. Sel. Evol.* 39, 39–53. doi: 10.1186/1297-9686-39-1-39

Costa, A., Neglia, G., Campanile, G., and De Marchi, M. (2020a). Milk somatic cell count and its relationship with milk yield and quality traits in Italian water buffaloes. *J. Dairy Sci.* 103, 5485–5494. doi: 10.3168/jds.2019-18009

Costa, A., Negrini, R., De Marchi, M., Campanile, G., and Neglia, G. (2020b). Phenotypic characterization of milk yield and quality traits in a large population of water buffaloes. *Animals* 10:327. doi: 10.3390/ani10020327

Crews, D. H. Jr., and Franke, D. E. (1998). Heterogeneity of variances for carcass traits by percentage Brahman inheritance2. *J. Anim. Sci.* 76, 1803–1809. doi: 10.2527/1998.7671803x

Food and Agriculture Organization (FAO) (2020). *Live Animals*. Available online at: http://www.fao.org/faostat/en/#data/QA (Accessed September 28, 2020)

Gómez, M. M., Gama, L. T., León, J. M., Fernández, J., Attalla, S. A., and Delgado, J. V. (2016). Genetic parameters for harmony and gaits in hispano-arabe horses estimated by bayesian methods and Restricted Maximum Likelihood. *Livest. Sci.* 188, 159–165. doi: 10.1016/j.livsci.2016.04.016

Henderson, C. R. (1988). Use of an average numerator relationship matrix for multiple-sire joining. *J. Anim. Sci.* 66, 1614–1621. doi: 10.2527/jas1988.6671614x

ISMEA (2020). *Istituto Di Servizi per Il Mercato Agricolo Alimentare è un Ente Pubblico Economico Nazionale Sottoposto Alla Vigilanza Del Ministero Delle Politiche Agricole Alimentari e Forestali*. Available online at: http://www.ismeamercati.it/flex/cm/pages/ServeBLOB.php/L/IT/IDPagina/5186 (Accessed September 15, 2020)

Kaufman, L., and Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ: John Wiley & Sons.

Macedo, F. L., Christensen, O. F., Astruc, J.-M., Aguilar, I., Masuda, Y., and Legarra, A. (2020). Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. *Genet. Sel. Evol.* 52:47. doi: 10.1186/s12711-020-00567-1

Misztal, I., Lourenco, D., and Legarra, A. (2020). Current status of genomic evaluation. *J. Anim. Sci.* 98, skaa101. doi: 10.1093/jas/skaa101

Misztal, I., Tsuruta, S., Lourenco, D. A. L., Masuda, Y., Aguilar, I., Legarra, A., et al. (2018). *Manual for BLUPF90 Family Programs*. Georgia: University of Georgia.

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., and Lee, D. (eds) (2002). "BLUPF90 and related programs (BGF90)," in *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, Montpellier.

Neglia, G., de Nicola, D., Esposito, L., Salzano, A., D'Occhio, M. J., and Fatone, G. (2020). Reproductive management in buffalo by artificial insemination. *Theriogenology* 150, 166–172. doi: 10.1016/j.theriogenology.2020.01.016

Nwogwugwu, C. P., Kim, Y., Chung, Y. J., Jang, S. B., Roh, S. H., Kim, S., et al. (2020). Effect of errors in pedigree on the accuracy of estimated breeding value for carcass traits in Korean Hanwoo cattle. *Asian-Australas. J. Anim. Sci.* 33, 1057–1067. doi: 10.5713/ajas.19.0021

Parlato, E., and Van Vleck, L. D. (2012). Effect of parentage misidentification on estimates of genetic parameters for milk yield in the Mediterranean Italian buffalo population. *J. Dairy Sci.* 95, 4059–4064. doi: 10.3168/jds.2011-4855

Perez-Enciso, M., and Fernando, R. L. (1992). Genetic evaluation with uncertain parentage: a comparison of methods. *Theor. Appl. Genet.* 84, 173–179. doi: 10.1007/bf00223997

Peškovičová, D., Groeneveld, E., and Wolf, J. (2004). Effect of genetic groups on the efficiency of selection in pigs. *Livest. Prod. Sci.* 88, 213–222. doi: 10.1016/j.livprodsci.2003.12.003

Petrini, J., Pertile, S. F. N., Eler, J. P., Ferraz, J. B. S., Mattos, E. C., Figueiredo, L. G. G., et al. (2015). Genetic grouping strategies in selection efficiency of composite beef cattle (*Bos taurus* × *Bos indicus*)1. *J. Anim. Sci.* 93, 541–552. doi: 10.2527/jas.2014-8088

Phocas, F., and Laloë, D. (2004). Should genetic groups be fitted in BLUP evaluation? Practical answer for the French AI beef sire evaluation. *Genet. Sel. Evol.* 36, 325–345. doi: 10.1186/1297-9686-36-3-325

Pieramati, C., and Van Vleck, L. D. (1993). Effect of genetic groups on estimates of additive genetic variance1. *J. Anim. Sci.* 71, 66–70. doi: 10.2527/1993.71166x

Postma, E. (2006). Implications of the difference between true and predicted breeding values for the study of natural selection and micro-evolution. *J. Evol. Biol.* 19, 309–320. doi: 10.1111/j.1420-9101.2005.01007.x

Purohit, G., Thanvi, P., Pushp, M., Gaur, M., Shekher, C., Saraswat, A. S. A., et al. (2019). Estrus synchronization in buffaloes: prospects, approaches and limitations. *Pharma Innov. J.* 8, 54–62.

Quaas, R. L. (1988). Additive genetic model with groups and relationships. *J. Dairy Sci.* 71, 1338–1345. doi: 10.3168/jds.S0022-0302(88)79691-5

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team.

Raoul, J., Palhière, I., Astruc, J. M., and Elsen, J. M. (2016). Genetic and economic effects of the increase in female paternal filiations by parentage assignment in sheep and goat breeding programs1. *J. Anim. Sci.* 94, 3663–3683. doi: 10.2527/jas.2015-0165

Rodriguez, M. C., Toro, M., and Silió, L. (1996). Selection on lean growth in a nucleus of Landrace pigs: an analysis using Gibbs sampling. *Anim. Sci.* 63, 243–253. doi: 10.1017/S1357729800014806

Rodríguez-Martínez, H., and Peña Vega, F. (2013). Semen technologies in domestic animal species. *Anim. Front.* 3, 26–33. doi: 10.2527/af.2013-0030

Saavedra, L. (2019). *Variabilidad Genética de Ácidos Grasos y Fracciones Nitrogenadas en Leche de Bovinos*. Ph. D. thesis, Chapingo, MX: Universidad Autónoma Chapingo.

Safari, A., Ghavi Hossein-Zadeh, N., Shadparvar, A. A., and Abdollahi Arpanahi, R. (2018). A review on breeding and genetic strategies in Iranian buffaloes (*Bubalus bubalis*). *Trop. Anim. Health Prod.* 50, 707–714. doi: 10.1007/s11250-018-1563-1

Santana, M. L., Eler, J. P., and Ferraz, J. B. S. (2013). Alternative contemporary group structure to maximize the use of field records: application to growth traits of composite beef cattle. *Livest. Sci.* 157, 20–27. doi: 10.1016/j.livsci.2013.06.034

Sapp, R. L., Zhang, W., Bertrand, J. K., and Rekaya, R. (2007). Genetic evaluation in the presence of uncertain additive relationships. I. Use of phenotypic information to ascertain paternity. *J. Anim. Sci.* 85, 2391–2400. doi: 10.2527/jas.2006-667

Shiotsuki, L., Cardoso, F. F., and Albuquerque, L. G. (2018). Method for the estimation of genetic merit of animals with uncertain paternity under Bayesian inference. *J. Anim. Breed Genet.* 135, 116–123. doi: 10.1111/jbg.12322

Shiotsuki, L., Cardoso, F. F., and Silva, J. A. II (2013). Albuquerque LG. Comparison of a genetic group and unknown paternity models for growth traits in Nellore cattle. *J. Anim. Sci.* 91, 5135–5143. doi: 10.2527/jas.2011-4989

Shiotsuki, L., Cardoso, F. F., Silva, J. A. I. V., Rosa, G. J. M., and Albuquerque, L. G. (2012). Evaluation of an average numerator relationship matrix model and a Bayesian hierarchical model for growth traits in Nellore cattle with uncertain paternity. *Livest. Sci.* 144, 89–95. doi: 10.1016/j.livsci.2011.11.002

Singh, I., and Balhara, A. K. (2016). New approaches in buffalo artificial insemination programs with special reference to India. *Theriogenology* 86, 194–199. doi: 10.1016/j.theriogenology.2016.04.031

Sullivan, P. G. (1995). Alternatives for genetic evaluation with uncertain parentage. *Can. J. Anim. Sci.* 75, 31–36. doi: 10.4141/cjas95-004

Theron, H., Kanfer, F., and Rautenbach, L. (2002). The effect of phantom parent groups on genetic trend estimation. *S. Afr. J. Anim. Sci.* 32, 130–135. doi: 10.4314/sajas.v32i2.3755

Tonussi, R. L., Silva, R. M. D. O., Magalhães, A. F. B., Espigolan, R., Peripolli, E., Olivieri, B. F., et al. (2017). Application of single step genomic BLUP under different uncertain paternity scenarios using simulated data. *PLoS One* 12:e0181752. doi: 10.1371/journal.pone.0181752

Ugur, M. R., Saber Abdelrahman, A., Evans, H. C., Gilmore, A. A., Hitit, M., Arifiantini, R. I., et al. (2019). Advances in cryopreservation of bull sperm. *Front. Vet. Sci.* 6:268. doi: 10.3389/fvets.2019.00268

Wellmann, R. (2019). Optimum contribution selection for animal breeding and conservation: the R package optiSel. *BMC Bioinformatics* 20:25. doi: 10.1186/s12859-018-2450-5

Westell, R. A., Quaas, R. L., and Van Vleck, L. D. (1988). Genetic Groups in an Animal Model. *J. Dairy Sci.* 71, 1310–1318.

Wolak, M. E., and Reid, J. M. (2017). Accounting for genetic differences among unknown parents in microevolutionary studies: how to include genetic groups in quantitative genetic animal models. *J. Anim. Ecol.* 86, 7–20.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 04 February 2021
doi: 10.3389/fgene.2020.622494

Complete *CSN1S2* Characterization, Novel Allele Identification and Association With Milk Fatty Acid Composition in River Buffalo

Gianfranco Cosenza[1], Daniela Gallo[1], Barbara Auzino[1], Giustino Gaspa[2] and Alfredo Pauciullo[2]**

[1] Department of Agriculture, University of Napoli Federico II, Portici, Italy, [2] Department of Agricultural, Forest and Food Sciences, University of Torino, Grugliasco, Italy

The αs2-casein is one of the phosphoproteins secreted in all ruminants' milk, and it is the most hydrophilic of all caseins. However, this important gene (*CSN1S2*) has not been characterized in detail in buffaloes with only two alleles detected (reported as alleles A and B), and no association studies with milk traits have been carried out unlike what has been achieved for other species of ruminants. In this study, we sequenced the whole gene of two Mediterranean river buffalo homozygotes for the presence/absence of the nucleotide C (g.7539G>C) realized at the donor splice site of exon 7 and, therefore, responsible for the skipping of the same exon at mRNA level (allele B). A high genetic variability was found all over the two sequenced *CSN1S2* alleles. In particular, 74 polymorphic sites were found in introns, six in the promoter, and three SNPs in the coding region (g.11072C>T, g.12803A>T, and g.14067A>G) with two of them responsible for amino acid replacements. Considering this genetic diversity, those found in the database and the SNP at the donor splice site of exon 7, it is possible to deduce at least eight different alleles (*CSN1S2* A, B, B1, B2, C, D, E, and F) responsible for seven different possible translations of the buffalo αs2-casein. Haplotype data analysis suggests an evolutionary pathway of buffalo *CSN1S2* gene consistent with our proposal that the published allele *CSN1S2* A is the ancestral αs2-CN form, and the B2 probably arises from interallelic recombination (single crossing) between the alleles D and B (or B1). The allele *CSN1S2* C is of new identification, while *CSN1S2* B, B1, and B2 are deleted alleles because all are characterized by the mutation g.7539G>C. Two SNPs (g.7539G>C and g.14067A>G) were genotyped in 747 Italian buffaloes, and major alleles had a relative frequency of 0.83 and 0.51, respectively. An association study between these SNPs and milk traits including fatty acid composition was carried out. The SNP g.14067A>G showed a significant association ($P < 0.05$) on the content of palmitic acid in buffalo milk, thus suggesting its use in marker-assisted selection programs aiming for the improvement of buffalo milk fatty acid composition.

Keywords: *CSN1S2*, alleles, candidate gene, mediterranean river buffalo, milk, palmitic acid

OPEN ACCESS

Edited by:
Wai Yee Low,
University of Adelaide, Australia

Reviewed by:
Salvatore Mastrangelo,
University of Palermo, Italy
Giovanni Chillemi,
University of Tuscia, Italy

****Correspondence:***
Alfredo Pauciullo
alfredo.pauciullo@unito.it
Gianfranco Cosenza
giacosen@unina.it

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *28 October 2020*
Accepted: *24 December 2020*
Published: *04 February 2021*

Citation:
Cosenza G, Gallo D, Auzino B, Gaspa G and Pauciullo A (2021) Complete CSN1S2 Characterization, Novel Allele Identification and Association With Milk Fatty Acid Composition in River Buffalo. Front. Genet. 11:622494. doi: 10.3389/fgene.2020.622494

INTRODUCTION

The αs2-casein (207 aa) is one of the phosphoproteins (αs1, β, αs2, and k) secreted in ruminants' milk in the form of stable calcium–phosphate micelles, and it is the most hydrophilic of all caseins. The αs2-casein (αs2-CN) appears to be readily susceptible to proteolysis as assessed by the activities of chymosin and plasmin toward the protein. The molecular weight of this protein was assessed to be 22,741 Da in buffalo vs. 25,226 in cattle (Feligini et al., 2009).

The proportion of αs2-CN in milk changes considerably between species and is absent from human and marsupial milk (Kim et al., 2015). In buffalo milk, the αs2-CN is the third most abundant casein fraction (4.99 g/L), and the corresponding coding gene (*CSN1S2*) showed a lower translation efficiency (0.25) compared to the other casein genes as *CSN3* (k-CN, 2.69), *CSN2* (β-CN, 2.39), and *CSN1S1* (αs1-CN, 1.31) (Cosenza et al., 2011).

Among ruminants, goat and sheep showed a higher level of genetic diversity at *CSN1S2*, and nowadays, at least seven alleles associated with three different αs2-CN levels have been characterized in both species (Boisnard et al., 1991; Ramunno et al., 2001a,b; Giambra and Erhardt, 2011). In cattle, only four variants A, B, C, and D have been found (Farrell et al., 2004). The alleles B and C are specific for the zebu and yak cattle, respectively (Ibeagha-Awemu et al., 2007).

Conversely, this *locus* is less polymorphic in buffalo, probably as a result of the little studies realized in this species. Chianese et al. (1996) have reported three variants that differ for the content of phosphates, and D'Ambrosio et al. (2008) have indicated different αs2-CN isoforms with 13, 12, 11, and 10 phosphate groups realized at the same positions as those observed in cattle. At the DNA level, the only example of biallelic polymorphism (alleles A and B) observed, so far, at the buffalo *CSN1S2* has been identified and characterized by Cosenza et al. (2009a). The mutation that characterizes the allele B is an SNP (FM865620:g.773G>C) realized at the donor splice site of exon 7 and, therefore, responsible for the skipping of the same exon at mRNA level.

Contrary to what has been studied in other ruminants, until now, this important gene has not been characterized in detail in buffaloes. In 2006, Sukla et al. characterized the cDNA sequence in the Murrah breed (GeneBank no. DQ173244.1), and only very recently, the complete and annotated sequence of *CSN1S2* gene has been published for the Mediterranean breed (*Bubalus bubalis* breed Mediterranean chromosome 7, ASM312139v1, whole genome shotgun sequence; GenBank no. NC_037551.1, 32020000-32040337, complement).

Although a new reference genome assembly (UOA_WB_1) has been published (Low et al., 2019), and the first SNP array designed specifically for buffaloes has become available (Iamartino et al., 2017), its use is still very limited. Therefore, the candidate gene approach is still today a valid method for the identification of genetic variability and its relationship with milk production traits. Several studies have been carried out in river buffalo aiming the discovery of polymorphisms in *loci* coding for milk proteins that, in other ruminants, have well-known effects on milk characteristics (Masina et al., 2007; Cosenza et al., 2009a,b; Balteanu et al., 2013; Vinesh et al., 2013; Cosenza et al., 2015). For instance, these studies allowed the identification of positive associations between markers at *CSN1S1* and *CSN3* and traits of economic interest, like the protein yield (Cosenza et al., 2015) and milk coagulation properties (Bonfatti et al., 2012a,b). Conversely, in this respect, no association studies have been carried out in the buffalo for the *CSN1S2* so far, unlike what has been achieved for other species of ruminants. In fact, significant differences were found between genotypes of the goat, sheep, and cattle *CSN1S2 locus* in relation to milk protein and casein content (Ramunno et al., 2001b; Noce et al., 2016; Ardicli et al., 2018). Besides, *CSN1S2* genotypes were significantly associated with milk and/or fat yield in goat and sheep (Wessels et al., 2004; Lan et al., 2005; Yue et al., 2013; Vacca et al., 2018). For years, the interest of several research groups also focused on the study of connection between milk fat and fatty acid composition and the different milk protein polymorphisms and/or genetic polymorphisms of casein-encoding genes (Bobe et al., 1999, 2004; Chilliard et al., 2006; Cebo et al., 2012). In particular, it has been shown that fat globule size, the incidence of each globule size class on total measured bovine milk fat globules, and fatty acid composition are strongly influenced by single casein *loci* or casein haplotype (Perna et al., 2016).

The aim of this study was to sequence the whole *CSN1S2* for the samples reported as alleles A and B by Cosenza et al. (2009a), to characterize and annotate extensively the gene, to compare the alleles in their complex genetic diversity, and to investigate possible association with traits that might affect the nutritional and technological quality of buffalo milk.

MATERIALS AND METHODS

DNA Samples and Phenotypes Collection

Samples used in this study belong to DNA collections of the University of Napoli Federico II and University of Turin.

The original biological tissue used for DNA isolation was blood, collected during routine treatments according to Italian national rules on animal welfare and achieved by official veterinarians in collaboration with the Italian National Association of Buffalo Breeders (A.N.A.S.B.).

DNA was isolated from leukocytes using the procedure described by Goossens and Kan (1981). DNA concentration and the $OD_{260/280}$ ratio of the samples were measured by a Nanodrop ND-2000 Spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA).

DNA from two Mediterranean river buffaloes, homozygotes for the alleles A (FM865620:g.773G) and B (FM865621:g.773C) as determined by Cosenza et al. (2009a), have been used for the complete sequencing of the *CSN1S2*. In addition, individual DNA samples randomly chosen from 747 female Mediterranean river buffaloes belonging to 14 farms with intensive breeding system, located in Salerno, Caserta, and Potenza provinces (Southern Italy) were used for population analysis.

For assessing possible associations between polymorphisms identified at the *CSN1S2 locus* and milk traits, such as milk yield, fat percentage, single fatty acid percentage, and fatty acid classes, we used single milk samples collected from a sub-group of 310 lactating buffaloes. These subjects were at third calving, had similar days in milking (DIM: 110–120), feeding management and diet, with a reduced occurrence of unsaturated fatty acids, compared to graze-based systems.

Fatty acid (FA) composition, FA classes, and fat percentage of the 310 individual milk samples have been assessed and previously reported by Cosenza et al. (2017a, 2018a). The same phenotypes were also used in the present work to assess possible associations with the genetic diversity found at the *CSN1S2 locus* by using the mixed linear model as reported by Cosenza et al. (2017a).

PCR Amplification Conditions and Genotyping

Using primers designed on bubaline genome (GenBank accession no. NC_037551.1, from 32020000 to 32040337 complement) and bubaline mRNA sequence (GenBank accession nos. FM865618.1, FM865619.1) (**Supplementary Table 1**), the DNA regions of the *CSN1S2* gene spanning from the 5′- to the 3′-UTR of two Mediterranean river buffalo homozygotes for the alleles A and B were amplified by iCycler (BioRad, CA, USA). A typical 50-μl PCR reaction mix including 100 ng of genomic DNA, 50 mM KCl, 10 mM Tris-HCl (pH 9.0), 0.1% Triton X-100, 3 mM $MgCl_2$, 200 nmol of each primer, dNTPs each at 400 μM, 2.5 U of Taq DNA Polymerase (Promega, Madison, WI), and 0.04% BSA. The thermal condition for the amplification consisted of an initial denaturation at 95°C for 4 min, followed by 35 cycles at 94°C for 45 s, 54.0–57.4°C for 45 s (according to the amplicon) and 72°C for 2 min. A final extension of 10 min was accomplished to end the reaction. All PCR products were analyzed directly by electrophoresis in 1.5% TBE agarose gel (Bio-Rad, CA, USA) in 0.5X TBE buffer and stained with SYBR® green nucleic acid stain (Lonza Rockland, Inc., USA). PCR products were sequenced on both strands at CEINGE–Biotecnologie Avanzate (Naples, Italy) using Sanger DNA sequencing technology.

The entire panel of 747 Mediterranean river buffalo DNA samples was genotyped in outsourcing (KBiosciences, Herts, UK, http://www.kbioscience.co.uk) for the SNPs g.7539G>C (FM865620:g.773G>C) and g.14067A>G.

Bioinformatics and Statistical Analysis

Allelic frequencies and Hardy–Weinberg equilibrium (chi square test) were calculated. Homology searches, comparisons among nucleotide and amino acid sequences, and multiple alignments for polymorphism discovery were accomplished using Dnasis Pro (Hitachi Software Engineering Co.). Measures of linkage disequilibrium (D' and r^2) were estimated using Haploview software ver. 4.2 (http://www.broadinstitute.org/haploview/haploview). The haplotype structure was defined according to Gabriel et al. (2002). The regulatory regions were analyzed for potential transcription factors (TFs) by Transfac® 7.0. (http://gene-regulation.com/index2.html). Associations between *CSN1S2* genotypes and fat traits were tested using a mixed linear model by SAS (*ver* 9.2) as reported by Cosenza et al. (2017a).

RESULTS

CSN1S2 Gene Structure in Mediterranean River Buffalo

By using genomic DNA as template, we sequenced the whole gene encoding the αs2-casein (*CSN1S2*) of two Mediterranean river buffalo homozygotes for the presence/absence of the nucleotide C (FM865620:g.773G>C) that caused inactivation of the intron 7 splice donor site and, consequently, the allele-specific exon skipping characteristic of the *CSN1S2* B allele (GenBank accession nos. MW159135 and MW159136).

Using as reference the sample homozygote for the allele FM865620:g.773G (previously misidentified as *CSN1S2* A and from now named allele *CSN1S2* D), the sequenced DNA region including the *CSN1S2* gene is about 20,300-bp long, and it includes 1,025 bp of exonic regions, 17,578 bp of intronic regions, 937 nucleotides at the 5′ flanking region, and 707 nucleotides at the 3′ flanking region. The level of sequence similarity with the allele *CSN1S2* B is about 98% as a consequence of an elevated polymorphism.

The main feature of the buffalo *CSN1S2* gene is the extremely split architecture. It contains 18 exons ranging in size from 21 (exon 4) to 267 bp (exon 18). The first exon (44 bp) is not coding at all. The whole highly conserved signal peptide (15 amino acids, MKFFIFTCLLAVALA) of the mature protein (207 amino acids) is encoded by the nucleotides 13–57 of exon 2 (63 bp), and the translation stop codon TAA is created by nucleotides 10–12 of exon 17. The deduced CDS length of bubaline *CSN1S2* gene is 669-bp long. These results are in agreement with what was reported by Sukla et al. (2006). All splice junctions follow the 5′ GT/3′ AG splice rule, similarly as it was described in different ruminant species. The only peculiarity is represented by the polymorphism at the splice donor site of exon 7 of the allele *CSN1S2* B (g.7539G>C, corresponding to FM865620:g.773G>C).

Consequently, the *CSN1S2* B allele (GenBank MW159136) compared to the *CSN1S2* D (GenBank MW159135) allele is characterized by 17 exons.

Finally, different microsatellite sequences are present in the buffalo *CSN1S2* gene, many of which flanking retroposonic sequences (**Supplementary Figure 1**).

Polymorphism Detection

The analysis and the alignment of the *CSN1S2* intronic sequences of the two subjects used in this study have highlighted a remarkable genetic diversity.

In detail, 74 polymorphic sites (24 transversions, 37 transitions, 13 deletions/insertions) and several variable microsatellites were found between the two sequenced subjects

(**Supplementary Figure 1**, **Supplementary Table 2**). Except for the g.7539G>C at the splicing donor site of exon 7 and causative event of the *CSN1S2* B allele, none of the remaining polymorphisms are apparently located in the regulatory regions (splicing donor/acceptor site, enhancer/silencer, etc.) and as a consequence, we hypothesize that they do not affect the *CSN1S2* expression.

Then, the comparison between our sequences and the reference sequence recorded in GenBank (NC_037551.1) highlighted further 15 new intronic mutations. In particular, two polymorphisms are responsible for the differences in the number of mononucleotide thymine (T) repeats, while one is a multiple substitution: NC_037551.1:g.32034006A>G>T (**Supplementary Table 2**). This genomic sequence is particularly interesting because it is also characterized by a cytosine at the splice donor site of exon 7 (NC_037551.1:g.32033131C), and consequently, it can be considered an allele B derived.

As expected, the comparison of the exonic regions showed a reduced level of polymorphism. We identified three SNPs in total. The first, g.11072C>T, is located at the 18th nucleotide of exon 13; it is a conservative SNP, and it is not generating any amino acid change. The further SNPs are located at the 16th nucleotide of exon 14 and at the 31st nucleotide of exon 16. They are the transversions g.12803A>T and g.14067A>G responsible for the amino acid substitutions p.I162>F and p.190T>A, respectively (**Supplementary Figure 1**). The g.14067A>G has been observed in heterozygosis in the sample homozygote for the SNP g.7539G, giving two new alleles named *CSN1S2* C (g.14067A) and D (g.14067G).

Furthermore, by comparing the sequences analyzed in this work and those available in the database (https://blast.ncbi.nlm.nih.gov/Blast.cgi) for the buffalo *CSN1S2* gene, it is possible to identify other four exonic polymorphisms (**Table 1**) and consequently several haplotypes.

Two SNPs were conservative, the transition g.3165T>C (27th nt of the exon 2, GenBank acc. no. DQ173244.1) and the transversion g.9220T>A (90th nt of the exon 11, GenBank acc. no. DQ133467.1). The other SNPs were not conservative: the transition g.9221A>G (91st nt of the exon 11, GenBank acc. no. DQ133467.1), responsible for the amino acid change p.128K>E and the transversion g.14141C>G (105th nt of the exon 16, GenBank acc. no. DQ173244.1), which generates the amino acid replacement p.214N>K.

Rearrangement of Allele Nomenclature and Phylogenetic Relationship Among the Markers

Considering all the SNPs (both from the database and newly determined in the present study), it is possible to deduce at least eight different alleles (*CSN1S2* A, B, B1, B2, C, D, E, and F; **Table 1**) responsible for seven different possible translations of the buffalo αs2-casein.

The allele that we named *CSN1S2* A (GenBank NM_001290865) is stated as ancestral αs2-CN form according to nucleotide and amino acid sequence of cattle and goat. By several mutational events often responsible for either amino acid substitution or deletions, starting from *CSN1S2* A, we propose two different phylogenetic road maps. The first map generates four alleles that are different for a single amino acid substitution: p.162F>I (*CSN1S2* C, present work), p.162F>I and p.190A>T (*CSN1S2* D, XM_006071123.2, KY399458.2, FM865618.1, JQ292811.1, AJ005431.2, present work), p.162F>I, p.190A>T, and p.214N>K (*CSN1S2* E, DQ173244.1). Similar to *CSN1S2* E, also the allele named *CSN1S2* F (DQ133467.1) originated from the allele *CSN1S2* D because they differ from each other only for the amino acid substitution p.128K>E (**Figure 1**).

The second phylogenetic road map is generated by the point mutation g.7539G>C, which brings to the inactivation of the intron 7 splice donor site. Thus, as a consequence, the alleles named *CSN1S2* B1 (KX896650) and B (FM865619.1; present work) are characterized by the complete skipping of exon 7 (nine amino acids, EVIRNANEE from 58 to 66). Moreover, *CSN1S2* B1 and B differ from each other for the single polymorphism g.11072C>T in the exon 13 (**Table 1**, **Figure 1**).

Finally, the comparison of specific haplotypes defined for each of the *CSN1S2* alleles (**Table 1**) indicates that the B2 probably arises from interallelic recombination (single crossing) between alleles D and B or B1 (**Figure 1**).

Regulatory Elements and Polymorphism Detection at the Gene Promoter

Variations in regulatory regions are known to affect the composition, structure, and expression of milk caseins (Martin et al., 2002; Szymanowska et al., 2003; Cosenza et al., 2007, 2016). Therefore, the proximal promoter regions of both *CSN1S2* D and B alleles were sequenced and characterized.

Using the database Transfact® 7.0, we identified the potential transcription factors (TFs) that could affect the gene expression. Together with the TATA box, we identified the following TFs: C/EBP (CCAAT/enhancer-binding protein), Oct-1 (octamer-binding factor-1), HNF-3beta (hepatocyte nuclear factor-3 beta), AP (activator protein), YY1 (Yiang Yang factor-1), POU1F1a (Pit1, growth hormone factor 1), PR (progesterone receptor), GR (glucocorticoid receptor), and MGF (mammary gland factor) (**Supplementary Figure 1**). This gene structure is similar to the homologous gene identified in *Bos taurus* and *Bos indicus* (Kishore et al., 2013) as demonstrated by the conserved position of the TATA box (between nucleotide −25 and −30, where +1 is the first nucleotide of the first exon).

The sequence comparison of the gene promoters for the alleles *CSN1S2* B and D showed six SNPs in total: four transitions, one transversion, and the deletion of one adenine (**Supplementary Figure 1**, **Supplementary Table 2**). None of the polymorphisms identified generates or deletes known TFs, and consequently, no influence on gene expression was expected.

However, in the comparison with the only *Bubalus bubalis* promoter sequence available at GenBank (accession number EF066480), three additional sites of variation were detected: g.595A>G, g.620_622delG, and g.996T>A. The latter polymorphism fell within the putative transcriptional factor binding site for Oct-1.

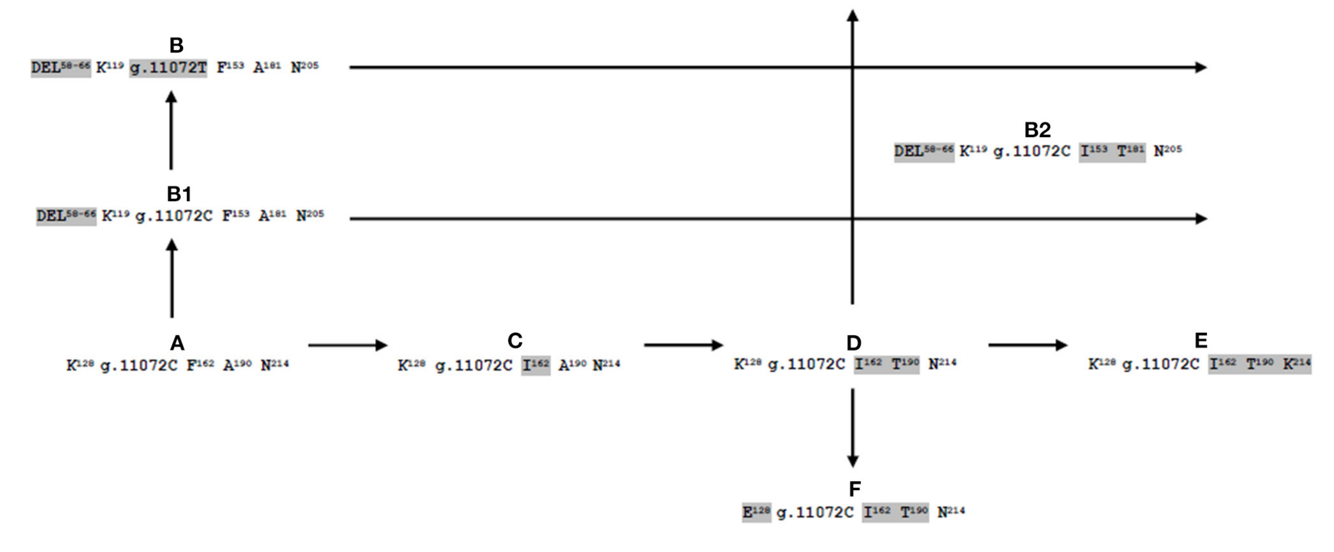

FIGURE 1 | Possible evolutionary pathway of buffalo αs2-casein-encoding gene (*CSN1S2*).

At the 3′-end of the gene, the polyadenylation site AATAAAA is located between nucleotides 247–253, with reference to the first nucleotide of the 18th exon. In addition, a G/T cluster was found downstream of the poly-A site. This sequence motif also contributes to the information for the polyadenylation. Both the AATAAA sequence and the G/T cluster are underlined in **Supplementary Figure 1**. With the exception of a polymorphic stretch of T, we do not report any further mutation in this region.

Repeated Sequences Within the Mediterranean River Buffalo *CSN1S2* Gene

The buffalo *CSN1S2* gene sequence is characterized by at least 13 retrotransposons (**Supplementary Figure 1**). In particular, the first (A) is located in the promoter region (GenBank MW159135 from 136 to 305) and appears to be a retroposon of Bov-tA2 type.

Further, two elements are located in the first intron (B, from 1,122 to 1,290, and C, from 1,952 to 2,137, respectively) and showed a strong similarity with an L1_Art sequence. Then, we found two Bov-tA2 located in intron 2 that we named element D (from the nucleotide 3,696 to 3,906) and element E (from nucleotide 4,155 to 4,313). At intron 8, we found a Bov-B (element F, from nucleotide 8,320 to 8,574), whereas in intron 12, we identified a Bov-A2 (retroposon G, from nucleotide 10,338 to nucleotide 10,621). Furthermore, five retroposons (H, I, L, M, and N) are located in intron 13 (Bov-tA1, from 12,138 to 12,356), in intron 15 (Bov-A2, from 13,313 to 13,587), and in the intron 17 (Bov-tA1, from 15,051 to 15,236, Bov-B from 15,931 to 17,485, and Bov-A2, from 17,763 to 18,046). Finally, a further element Bov-tA2 (O) is located in the 3′-UT region between the nucleotide 19,682 and 19,885, closely to the last exon.

The sequence similarity between these elements and those used as reference (Lenstra et al., 1993; AC150707.3; GenBank: AC150561.6) ranges from 75 to 90%.

Genotyping and Association of *CSN1S2* Polymorphisms With Milk Fatty Acid Composition Traits

To estimate the frequencies at the two polymorphic sites g.7539G>C and g.14067A>G, and to determine the possible haplotypes, specific genotyping protocols have been developed by the company Kbioscience (http://www.kbioscience.co.uk/genotyping/genotyping_intro.html).

The genotype distributions and the allelic frequencies of the two SNPs, determined in 747 buffaloes reared in Salerno, Caserta, and Potenza provinces (Italy) are reported in **Table 2**. The major alleles had a relative frequency of 0.83 and 0.51 for g.7539G and g.14067G, respectively, and the χ^2 value showed that there was no evidence of departure from the Hardy–Weinberg equilibrium ($P \leq 0.05$). Using Haploview software ver. 4.2 (http://www.broadinstitute.org/haploview/haploview), three different allelic combinations (out of the four expected) were observed: haplotypes 1 (7539G/14067A), 2 (7539G/14067G), and 3 (7539C/14067G). The first haplotype was the most represented with a frequency of 0.491, followed by the haplotypes 2 (0.336) and 3 (0.173). Although not observed, the fourth expected haplotype (7539C/14067A) was recorded on database (GenBank acc.no NC_037551.1).

The majority of mutations identified at this *locus* were either conservative (g.3165T>C, g.9220T>A, and g.11072C>T) or specific for an allele (g.9221A>G and g.14141C>G), and for these reasons, only the SNPs g.7539G>C and g.14067A>G were genotyped and used for running the model according to Cosenza et al. (2017a) (1).

Genotype distributions and allelic frequencies of both total- and sub-population genotyped are reported in **Table 2** and **Supplementary Table 3**, respectively.

The analysis of the relationships between the *CSN1S2* polymorphisms and the FA profile showed a significant effect

TABLE 1 | Discovery and diffusion of the genetic variants of buffalo αs2-casein-encoding gene (*CSN1S2*).

***CSN1S2* alleles**	**Exon, nucleotide, and amino acid position**																**Breed**
	Exon 2		**Exon 7**		**Exon 11**		**Exon 11**		**Exon 13**		**Exon 14**		**Exon 16**		**Exon 16**		
	nt	**aa**	**nt**	**aa**	**nt**	**aa**	**nt**	**aa**	**nt**	**aa**	**nt**	**aa**	**nt**	**aa**	**nt**	**aa**	
	3165	**5**	**7539**	**58-66**	**9220**	**127**	**9221**	**128**	**11072**	**153**	**12803**	**162**	**14067**	**190**	**14141**	**214**	
A[1]	T	I	G	EVIRNANEE	T	V	A	K	C	T	T	F	G	A	C	N	Murrah
C[2]	T	I	G	EVIRNANEE	T	V	A	K	C	T	A	I	G	A	C	N	Mediterranean
D[3]	T	I	G	EVIRNANEE	T	V	A	K	C	T	A	I	A	T	C	N	Mediterranean/ Egyptian/ Murrah
E[4]	C	I	G	EVIRNANEE	T	V	A	K	C	T	A	I	A	T	G	K	Murrah
F[5]	T	I	G	EVIRNANEE	A	V	G	E	C	T	A	I	A	T	C	N	Murrah
B[6]	T	I	C	–	T	V^{117}	A	K^{118}	T	T^{144}	T	F^{153}	G	A^{181}	C	N^{205}	Mediterranean
B1[7]	T	I	C	–	T	V^{117}	A	K^{118}	C	T^{144}	T	F^{153}	G	A^{181}	C	N^{205}	Carabao
B2[8]	T	I	C	–	T	V^{117}	A	K^{118}	C	T^{144}	A	I^{153}	A	T^{181}	C	N^{205}	Mediterranean

The CSN1S2 A allele is the putative original one from which the different alleles originated.
Numbering refers to CSN1S2 allele D (GenBank MW159135) both for nucleotides (nt) and the corresponding predicted protein (aa).
*References—**1**: NM_001290865.1; **2**: present work; **3**: XM_006071123.2, KY399458.2, FM865618.1, JQ292811.1, AJ005431.2, present work (MW159135); **4**: DQ173244.1; **5**: DQ133467.1; **6**: FM865619.1; present work (MW159136); **7**: KX896650.1; **8**: NC_037551.1.*

TABLE 2 | Genotyping data, allele frequency, relative frequencies of the SNP g.14067A>G at exons 16 and g.7539G>C in the splice donor site of intron 7 of the *CSN1S2* gene in the Mediterranean river buffalo population.

			Genotype distribution						**Allelic frequency**			
			g.14067A>G			**Obs.**	**Exp.**	**χ^2**	**g.7539**		**g.14067**	
			A/A	**G/A**	**G/G**				**G**	**C**	**A**	**G**
Genotype distribution	g.7539G>C	G/G	192	229	94	515	512.1		0.83	0.17	0.49	0.51
		G/C	–	123	84	207	212.79	0.55				
		C/C	–	–	25	25	192.29					
	Obs.		192	352	203	747						
	Exp.		181.29	373.42	192.29							
	χ^2			2.45								

($P < 0.05$) only for the SNP g.14067A>G on the content of palmitic acid in buffalo milk. In particular, the homozygous GG and heterozygous buffaloes showed a lower amount with 34.13% and 34.71% palmitic acid compared with the AA genotype (35.23%), respectively (**Table 3**).

DISCUSSION

Caseins αs1, β, αs2, and k have an important role for the production of milk-derived products in terms of quality and quantity. For this reason, in the last decades, many studies have been published in the main ruminant species (cow, sheep, and goat) about the identification of possible association between genetic markers and protein structure with milk traits of economic interest (Caroli et al., 2009; Selvaggi et al., 2014a,b; Ozdemir et al., 2018). Different from the abovementioned species, water buffalo has not been deeply investigated, and to our knowledge, the complete genomic sequence of the bubaline αs2-casein gene has not been reported yet. Therefore, this study focused first on the structure of the buffalo *CSN1S2* gene, exploring the genetic diversity within the Italian Mediterranean breed and testing possible associations between the detected polymorphisms and milk traits.

Structure and Analysis of Mediterranean River Buffalo *CSN1S2* Gene

On the whole, the buffalo *CSN1S2* gene shares a similar organization with the bovine counterpart (Groenen et al., 1993), with some differences in intronic size, mainly as a consequence of the presence/absence of artiodactyla retroposons.

Transposable elements (TE) are the largest class of sequences in mammalian genomes, elements that replicate and jump throughout the genome in a manner similar to retroviruses. The TEs are distributed primarily as retrotransposons (98.62%) rather than transposons (1.38%). DNA transposons have been extensively studied beyond mammals (Berg and Howe, 1989; Capy et al., 1998; Craig, 2002), whereas they are not well-documented in mammalian genomes. Based on their size and mode of propagation, retrotransposons can be divided into two

TABLE 3 | Least squares means of the SNP g.14067A>G genotypes for palmitic acid, estimation of average substitution effects (α) for the adenine to guanine replacement, and contribution of the polymorphism to the phenotypic variance (r^2).

SNP	Trait	P	Genotype			α	r^2
			AA (97)	AG (142)	GG (71)		
g.14067A>G	C16:0	0.05	35.23[a]	34.71[ab]	34.13[b]	0.55	0.15

[a,b]*Means within columns without a common superscript differ (P < 0.05).*

separate classes, the long terminal repeat (LTR) and non-LTR (Han, 2010). The non-LTR LINEs (long interspersed repeat elements, L1_Art, BovB) and SINEs (short interspersed repeat elements, BOV-A2, Bov-tA, tRNA, MIR, and others) are widely distributed and represent a major component of ruminant genomes. For example, the BovB LINEs and related SINEs occupy about 22% of the cow genome. In particular, two retroposon families, Bov-A2 and Bov-tA, are the most distributed in the genomes of ruminants (Lenstra et al., 1993). The Bov-A2 and Bov-tA retroposons share a common Bov-B LINE-derived region, called the Bov-A unit, suggesting a common origin for these two retroposons (Okada and Hamada, 1997; Shimamura et al., 1999).

Although many retroposons are common for the ruminants and non-ruminant species and, thus, are likely of ancestral origin, every species has a definite number of short interspersed nuclear elements, which contributes to make each genome specific for each species (Ramunno et al., 2004; Cosenza et al., 2005; Pauciullo et al., 2013, 2019).

The 13 repetitive elements observed at the buffalo *CSN1S2* gene and its promoter represent the 19.45% of the sequence deposited in the EMBL database. This figure decreases considerably in the bovine (GenBank no. M94327.1), caprine (GenBank no. NC_030813.1), and ovine (GenBank no. NC_040257.1) counterpart because of the presence/absence of other repetitive elements observed in these species. In particular, the bovine *CSN1S2* (similarity of 75.4 %) is characterized for the absence of the elements B and C and, at the same time, an expanded Bov-A2 (G element in buffalo in intron 12), which consisted of three Bov-A monomers (Bov-A3) in agreement with Onami et al. (2007). In sheep and goat, the number of retroposonic elements is lower. Both species have a similar gene structure (homology of 96%), and when compared to water buffalo, we noted the absence of elements C, G, I, and N. However, in the promoter region, there is an extra Bov-tA3, and in the intron 1, there is an expanded Bov-A2-derived sequences, which consisted of four Bov-A monomers: Bov-A4. Overall, it appears that the elements B, C, G, I, and N are rather young insertions. These ruminant-specific retrotransposon insertions are often polymorphic (present or absent) at orthologous *loci*, and they are highly informative genetic markers that can be considered a powerful phylogenetic tool for clustering studies, animal evolutionary history, population structure and demography, rather than the set-up of methods for the species discrimination in meat and dairy products (Cosenza et al., 2019).

The accumulation of interspersed repeats within or near genes has been studied in ruminants as well as in camelid casein genes (Groenen et al., 1993, Ramunno et al., 2004, Cosenza et al., 2009b, Pauciullo et al., 2013, 2014, 2019). It has not been observed that insertions within or near promoters or 3′ UTR can alter gene expression. Conversely, insertions into exons are often incorporated into existing protein-coding genes and modulate gene expression. For instance, the alleles E and G of the *CSN1S1* in goat and cattle, respectively, are characterized by the insertion of a truncated LINE in the last exon, which is, in both species, responsible for a reduction in transcriptional rate of the corresponding protein (Jansa Pérez et al., 1994; Rando et al., 1998). The interaction between the LINE sequence and the poly(A) sequence of the mature transcript, reduced the mRNA stability causing a rapid degradation of the transcript and a low protein synthesis efficiency (Rando et al., 1998). However, none of the elements observed at the *CSN1S2 locus* in the buffalo species would appear to be potentially responsible for differences in the gene expression.

Furthermore, these transposable elements are known to affect the genome in many other different ways: contributing to genome size increase, genomic instability, exonization, epigenetic regulation, RNA editing, and have the ability to generate microsatellites because they contain homopolymeric tracts and, in particular, mutations at many *loci* in the genome by Cordaux and Batzer (2009). In the buffalo, retroposons at the *CSN1S2 locus* are responsible for the majority of genetic variability. In fact, the comparison between the retroposonic sequences (4,157 and 4,139 bp for the alleles D and B, respectively) showed a homology level lower (98.92%) than that of the remaining part of the gene (99.49%) assessed on 16,164 and 16,179 bp, respectively, for alleles D and B. The increase in the genetic diversity of the retroposons is over eight-fold higher (8.23). Considering the number of mutational events (SNP, insertion/deletion) within each region, 24 mutations found in retroposons vs. 58 polymorphic sites found in the rest of the gene represent almost a double incidence of genetic variation (on average, one mutation every 160 vs. 279 bp, respectively). This finding confirms that interspersed repeats are major drivers of *CSN1S2* gene evolution.

The buffalo *CSN1S2* proximal promoter region showed, as expected, stronger similarities with sequences of other ruminants (about 96% with yak, cattle, zebu, and about 91–93% with goat, sheep, and common red deer) than those observed with non-ruminants (about 76% with lama, dromedary, pig, horse, donkey, and about 67% with rabbit).

Detection of Genetic Variability and Allele Discovery

In the last decades, several studies have highlighted the importance of the genetic variability in non-coding regions, which regulates the expression of genes involved in milk quali-quantitative properties. Such polymorphisms are often located in the promoter region of milk protein genes that regulate their transcriptional rate and thus determine the amount of transcripts in milk (Malewski, 1998; Szymanowska et al., 2004a,b).

Also polymorphisms located in the 3′ untranslated region (UTR) are important because they could modify the target sequence of microRNA (miRNA), an important class of non-coding RNA responsible for the regulation of many physiological processes (including lactation) by influencing mRNA stability (Chen et al., 2010). So far, many SNPs located in non-coding regions of genes involved in the milk production traits have been identified. For instance, SNPs responsible of splicing mechanism modification (Cosenza et al., 2009a; Giambra et al., 2010; Balteanu et al., 2013) and mutations affecting transcription factor binding sites are associated with the regulation of gene expression (Kuss et al., 2003; Liefers et al., 2005; Ordovás et al., 2009; Pauciullo et al., 2012a,b; Yang et al., 2015; Cosenza et al., 2016, 2018b; Gu et al., 2019).

The comparison between the promoter sequences of alleles B and D at the *locus CSN1S2* of the water buffalo and the sequences recorded in GenBank has highlighted nine SNPs. Among them, only the mutation g.996T>A is located within the putative binding site for Oct-1, and consequently, it could affect the *CSN1S2* gene expression. The transcription factors Oct-1 belongs to a family of structurally related POU domain factors found throughout the eukaryotes. Oct-1 is the most studied member of the POU factors. It is expressed in all eukaryotic cells and regulates, either positively or negatively, the expression of a variety of genes (Dong and Zhao, 2007). In fact, mutations in the consensus sequences of the ubiquitous Oct-1 transcription factor are reported to reduce hormonal induction in different gene promoters, like the β-casein-encoding gene (*CSN2*) promoter in mice (Dong and Zhao, 2007) or the oxytocin gene (*OXT*) promoter in sheep (Cosenza et al., 2017b).

One of the main finding of this study was the discovery of a great genetic diversity at this *locus* and the understanding of phylogenetic relationship among the markers. Therefore, the clarification and rearrangement of allele nomenclature were considered a priority.

Regarding the high genetic variability found all over the *CSN1S2* gene, the most interesting polymorphisms identified are the transversion g.7539G>C at the donor splicing site of exon 7 (responsible for the *CSN1S2* B allele) and three SNPs in the coding region (g.11072C>T, g.12803A>T, and g.14067A>G) with two of them responsible for amino acid replacements.

Besides these SNPs, the comparative analysis with the bubaline *CSN1S2* sequences in GenBank identified further mutations. In total, eight observed markers allow to identify eight different alleles: *CSN1S2* A, B, B1, B2, C, D, E, and F (**Table 1**).

As a consequence, for the first time, it was possible also to propose an evolutionary pathway of the buffalo *CSN1S2* gene (**Figure 1**), as it was already published for different casein-encoding genes in ruminants (Formaggioni et al., 1999; Cosenza et al., 2008; Giambra and Erhardt, 2011).

Among the eight alleles, the *CSN1S2* C is of novel identification because it was never observed or reported earlier in databases. Furthermore, the identification of three B-derived alleles is interesting because they are characterized by the mutation g.7539G>C, which brings to the inactivation of the intron 7 splice donor site. In particular, *CSN1S2* B and B1 differ only for the conservative mutation g.11072C>T at the 18th nucleotide of exon 13, i.e., coding for the same protein 198-aa long vs. the 207 aa of the normal αs2-CN. Conversely, the haplotype of the allele *CSN1S2* B2 (DEL$^{58-66}$ K^{119} g.11072C I^{153} T^{181} N^{205}) likely suggests an interallelic recombination between the alleles D (K^{128} g.11072C I^{162} T^{190} N^{214}) and B (DEL$^{58-66}$ K^{119} g.11072T F^{153} A^{181} N^{205}) or B1 (DEL$^{58-66}$ K^{119} g.11072C F^{153} A^{181} N^{205}) (**Figure 1**).

This hypothesis was strengthened by genomic sequencing data, the sequence of the *CSN1S2* B2 allele being available. Although a mutation-driven convergence cannot be excluded, an interallelic recombination/gene conversion event seems to be the most plausible. Indeed, a detailed comparative analysis at 94 polymorphic sites (15 belonging exclusively to allele B2) spanning a large part of the gene sequence (**Supplementary Table 1**) provides a haplotype formula allowing each allele to be precisely characterized. Thus, the B2 allele unequivocally appears to be a hybrid structure made of B-type allele sequences in its 5′ part (from the beginning of exon 12) followed by D allele sequences in its 3′ part (from exon 12 to 3′ flanking region). Following such a scheme, a recombination event would have occurred around exons 11 and 13. This is, to our knowledge, the first hypothesis of a genomic recombination event that happened for genetic polymorphism and generating a new allelic diversity at a *locus* encoding a milk protein in the buffalo. Similar examples were observed in the goat and llama for the *CSN1S1 locus* (Bevilacqua et al., 2002; Ramunno et al., 2005; Pauciullo et al., 2017). The resulting phylogenetic trees of the bubaline αs2-CN-encoding gene can certainly help to understand the history of buffalo breeds and their genetic distances, as recently illustrated also by Luo et al. (2020).

Genetic Association With the Milk Palmitic Fatty Acid

The study of the correlations between the identified genetic variability and the phenotypic variability of animals is important especially for economic traits such as milk production and composition that are controlled by a cluster of genes (polygenes) where each gene has a small effect on the trait.

Different molecular genetic methods are used to identify the candidate genes involved in these quali-quantitative traits. Recently, a commercial buffalo SNP chip array, Axiom_Buffalo Genotyping Array 90K (Affymetrix), has been created to investigate the structure of buffalo populations (Iamartino et al., 2017) and performing genome-wide association studies (GWAS). However, the use of the array is very limited, and the few studies available still refer to bovine genome for the SNP positions and gene annotations. This represents a great restriction despite the recent efforts in the new annotation release of the buffalo genome (Low et al., 2019). For this reason, the genome annotation is still necessary in this species, as well as the understanding of the candidate gene functions and their mechanisms in the regulation of milk production traits. In this respect, the approach of candidate gene association study is still a powerful method in river buffalo, especially for markers falling within genes or regulatory sequences and with putative causative effects. Thus, an additional aim of the present study was the identification of possible associations between two genetic markers (g.7539G>C and g.14067A>G) found at this *locus* and the water buffalo milk traits.

In our study, the SNP g.14067A>G showed a significant association ($P < 0.05$) with the content of palmitic acid in buffalo milk.

Palmitic acid is the main SFA in milk fat in all investigate species (Markiewicz-Keszycka et al., 2013; Gantner et al., 2015). Palmitic acid, also known as palmitate and belonging to the class of organic compounds known as long-chain fatty acids (C16:0), exists in all living species, ranging from bacteria to humans, and it is found naturally in palm oil and palm kernel oil, as well as in meat, milk, butter, and cheese. Palmitic acid is an essential component of cell membranes, secretory and transport lipids, with crucial roles in protein palmitoylation and palmitoylated signal molecules (German, 2011). In milk, the C16:0 originates both from diet and endogenous synthesis by the mammary gland (Chilliard et al., 2007).

In buffalo raw milk, the percentage of palmitic acid is about 34.8% of the total SFA (Cosenza et al., 2017a). This percentage is the highest among those observed in the milk of the majority of ruminants such as cattle (31.6%), goat (23.1%), and sheep (19.8%), and non-ruminants such as donkey (20.9%) (Blasi et al., 2008) and camel (18.4%) (Gorban and Izzeldin, 1999). On the contrary, the contribution of the short-chain FA (C8:0, C10:0, and C12:0) is rather low compared with what was observed in other ruminant species (Correddu et al., 2017).

High concentrations of palmitic acid are also present in buffalo dairy products, such as mozzarella di Bufala Campana PDO (24.7%, Romano et al., 2008), yogurt (31.7 %, Naydenova et al., 2013), and ghee butter (28.7%, Peña-Serna et al., 2019).

Nutrition and supplementation of feed rations constitute a natural and economical way for farmers to increase the content of unsaturated fatty acids in milk (Chilliard et al., 2007), but some authors reported their negative effect on milk flavor (Stoop et al., 2008). Moreover, it can cause milk fat depression and decrease in milk yield (Markiewicz-Keszycka et al., 2013).

An alternative way of acting on the concentration of milk fatty acids is the application of genetic selection. Indeed, Stoop et al. (2008) found that there is a considerable genetic variation for fatty acid composition, with genetic variation being high for C16:0.

However, few studies have shown possible associations between genetic variability and the variability of palmitic acid concentration in milk. Schennink et al. (2007) found that the acyl CoA:diacylglycerol acyltransferase1 (*DGAT*1) 232A variant is associated with less C16:0 in cow milk. This association has also been observed by Bouwman et al. (2011), which suggests that the gene 1-acylglycerol-3-phosphate O-acyltransferase 6 (AGPAT6) might be a candidate for this association. Similarly, Zidi et al. (2010) detected a suggestive association between *PRLR* genotype and palmitic acid in goat.

Recently, many studies have been performed to identified possible associations with fatty acid composition in water buffalo milk (Misra et al., 2008; Pauciullo et al., 2010; Cosenza et al., 2017a, 2018a; Gu et al., 2017, 2019). In particular, Cosenza et al. (2017a) reported that the genotype CC at the oxytocin receptor (*OXTR*) was significantly associated to a lower level of palmitic acid in milk of Mediterranean river buffalo.

It is well-documented that palmitic acid is associated with obesity, with decreased insulin sensitivity that could increase risk of type 2 diabetes and higher cardiovascular disease risk through increased level of blood cholesterol much more of other SFAs (Mensink et al., 2003; Bermudez et al., 2014; Praagman et al., 2016; Imamura et al., 2018). Therefore, its presence at high concentrations in human diets has a negative impact on health, and it should be avoided preferring foods with higher concentrations of MUFA and/or PUFA. In this respect, increasing the frequency of *CSN1S2* GG and *OXTR* CC genotypes in river buffalo might guarantee a lower content of C16:0 in milk and dairy to be desirable for the consumer of buffalo products.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

No animals were used in the present study. The samples used herein belonged to DNA collections available from past studies (Cosenza et al., 2017a, 2018a) already approved by different ethic committees. Therefore, according to the Committee on the Ethics of Animal Experiments of the University of Torino (D.R. n. 2128 released on 06/11/2015) further ethics approval was not required.

AUTHOR CONTRIBUTIONS

GC and AP conceived and designed the experiments. BA and DG performed the experiments. GC, AP, and GG analyzed the data. GC contributed reagents, materials, and analysis tools. GC and AP wrote the paper. GC, AP, GG, BA, and DG revised the article critically for important intellectual content. GC, AP, GG, BA, and DG gave final approval of the version to be published. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by the Italian Ministry of Research (Project number PON01_00486 GENOBU) and by the University of Turin (ORACLE Project number PAUA_RILO_20_01).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.622494/full#supplementary-material

REFERENCES

Ardicli, S., Soyudal, B., Samli, H., Dincel, D., and Balci, F. (2018). Effect of *STAT1*, *OLR1*, *CSN1S1*, *CSN1S2*, and *DGAT1* genes on milk yield and composition traits of Holstein breed. *R. Bras. Zootec.* 47:e20170247. doi: 10.1590/rbz4720170247

Balteanu, V. A., Carsai, T. C., and Vlaic, A. (2013). Identification of an intronic regulatory mutation at the buffalo αS1-casein gene that triggers the skipping of exon 6. *Mol. Biol. Rep.* 40, 4311–4316. doi: 10.1007/s11033-013-2518-2

Berg, D. E., and Howe, M. M. (1989). *Mobile DNA*. Washington, DC: American Society for Microbiology, xii, 972.

Bermudez, B., Ortega-Gomez, A., Varela, L. M., Villar, J., Abia, R., Muriana, F. J. G., et al. (2014). Clustering effects on postprandial insulin secretion and sensitivity in response to meals with different fatty acid compositions. *Food Funct.* 5, 1374–1380. doi: 10.1039/c4fo00067f

Bevilacqua, C., Ferranti, P., Garro, G., Veltri, C., Lagonigro, R., Leroux, C., et al. (2002). Interallelic recombination is probably responsible for the occurrence of a new alpha(s1)-casein variant found in the goat species. *Eur. J. Biochem.* 269, 1293–1303. doi: 10.1046/j.1432-1033.2002.02777.x

Blasi, F., Montesano, D., De Angelis, M., Maurizi, A., Ventura, F., Cossignani, L., et al. (2008). Results of stereospecific analysis of triacylglycerol fraction from donkey, cow, ewe, goat and buffalo milk. *J. Food Compos. Anal.* 21, 1–7. doi: 10.1016/j.jfca.2007.06.005

Bobe, G., Beitz, D. C., Freeman, A. E., and Lindberg, G. L. (1999). Associations among individual proteins and fatty acids in bovine milk as determined by correlations and factor analyses. *J. Dairy Res.* 66, 523–536. doi: 10.1017/s0022029999003799

Bobe, G., Freeman, A. E., Lindberg, G. L., and Beitz, D. C. (2004). The influence of milk protein phenotypes on fatty acid composition of milk from Holstein cows. *Milchwissenschaft* 59, 3–6.

Boisnard, M., Hue, D., Bouniol, C., Mercier, J. C., and Gaye, P. (1991). Multiple mRNA species code for two non-allelic forms of ovine alpha-s2-casein. *Eur. J. Biochem.* 201, 633–641. doi: 10.1111/j.1432-1033.1991.tb16324.x

Bonfatti, V., Giantin, M., Gervaso, M., Coletta, A., Dacasto, M., and Carnier, P. (2012a). Effect of *CSN1S1-CSN3* (α(S1)-κ-casein) composite genotype on milk production traits and milk coagulation properties in Mediterranean water buffalo. *J. Dairy Sci.* 95, 3435–3443. doi: 10.3168/jds.2011-4901

Bonfatti, V., Giantin, M., Gervaso, M., Rostellato, R., Coletta, A., Dacasto, M., et al. (2012b). *CSN1S1-CSN3* (αs1-k-casein) composite genotypes affect detailed milk protein composition of Mediterranean water buffalo. *J. Dairy Sci.* 95, 6801–6805. doi: 10.3168/jds.2012-5601

Bouwman, A. C., Bovenhuis, H., Visker, M. H. P. W., and van Arendonk, J. A. M. (2011). Genome-wide association of milk fatty acids in Dutch dairy cattle. *BMC Genet.* 12, 43. doi: 10.1186/1471-2156-12-43

Capy, P., Bazin, C., Higuet, D., and Langin, T. (1998). *Dynamics and Evolution of Transposable Elements*. 1° France: Landes Bioscience, Springer, Heidelberg, 197.

Caroli, A. M., Chessa, S., and Erhardt, G. J. (2009). Invited review: Milk protein polymorphisms in cattle: effect on animal breeding and human nutrition. *J. Dairy Sci.* 92, 5335–5352. doi: 10.3168/jds.2009-2461

Cebo, C., Lopez, C., Henry, C., Beauvallet, C., Ménard, O., Bevilacqua, C., et al. (2012). Goat αs1-casein genotype affects milk fat globule physicochemical properties and the composition of the milk fat globule membrane. *J. Dairy Sci.* 95, 6215–6229. doi: 10.3168/jds.2011-5233

Chen, X., Gao, C., Li, H., Huang, L., Sun, Q., Dong, Y., et al. (2010). Identification and characterization of microRNAs in raw milk during different periods of lactation, commercial fluid and powdered milk products. *Cell Res.* 20, 1128–1137. doi: 10.1038/cr.2010.80

Chianese, L., Garro, G., Nicolai, M. A., Mauriello, R., Laezza, P., Ferranti, P., et al. (1996). Ocurrence of five αs1 casein variants in ovine milk. *J. Dairy Res.* 63, 49–59. doi: 10.1017/s0022029900031538

Chilliard, Y., Glasser, F., Ferlay, A., Bernard, L., Rouel, J., and Doreau, M. (2007). Diet, rumen biohydrogenation and nutritional quality of cow and goat milk fat. *Eur. J. Lipid Sci. Technol.* 109, 828–855. doi: 10.1002/ejlt.200700080

Chilliard, Y., Rouel, J., and Leroux, C. (2006). Goat's alpha-s1 casein genotype influences its milk fatty acid composition and delta-9 desaturation ratios. *Anim. Feed Sci. Technol.* 131, 474–487. doi: 10.1016/j.anifeedsci.2006.05.025

Cordaux, R., and Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. *Nat Rev Genet.* 10, 691–703. doi: 10.1038/nrg2640

Correddu, F., Serdino, J., Manca, M. G., Cosenza, G., Pauciullo, A., Ramunno, L., et al. (2017). Use of multivariate factor analysis to characterize the fatty acid profile of buffalo milk. *J. Food Compost. Anal.* 60, 25–31. doi: 10.1016/j.jfca.2017.03.008

Cosenza, G., Ciampolini, R., Iannaccone, M., Gallo, D., Auzino, B., and Pauciullo, A. (2018b). Sequence variation and detection of a functional promoter polymorphism in the lysozyme c-type gene from Ragusano and Grigio Siciliano donkeys. *Anim. Genet.* 49, 270–271. doi: 10.1111/age.12647

Cosenza, G., Feligini, M., Mancusi, A., D'Avino, A., Coletta, A., Di Berardino, D., et al. (2009b). Italian Mediterranean river buffalo *CSN2* gene structure and promoter analysis. *Ital. J. Anim. Sci.* 8, 57–59. doi: 10.4081/ijas.2009.s2.57

Cosenza, G., Iannaccone, M., Auzino, B., Macciotta, N. P. P., Kovitvadhi, A., Nicolae, I., et al. (2018a). Remarkable genetic diversity detected at river buffalo Prolactin Receptor (*PRLR*) gene and association studies with milk fatty acid composition. *Anim. Genet.* 49, 159–168. doi: 10.1111/age.12645

Cosenza, G., Iannaccone, M., Gallo, D., and Pauciullo, A. (2019). A fast and reliable PCR method based on SINEs detection for the discrimination of buffalo, cattle, goat and sheep species in dairy products. *Asian-Australas J. Anim. Sci.* 32, 891–895. doi: 10.5713/ajas.18.0459

Cosenza, G., Iannaccone, M., Pico, B. A., Gallo, D., Capparelli, R., and Pauciullo, A. (2017b). Molecular characterization, genetic variability and detection of a functional polymorphism influencing the promoter activity of OXT gene in goat and sheep. *J. Dairy Res.* 84, 165–169. doi: 10.1017/S00220299170 00097

Cosenza, G., Iannaccone, M., Pico, B. A., Ramunno, L., and Capparelli, R. (2016). The SNP g1311T>C associated with the absence of β-casein in goat milk influences *CSN2* promoter activity. *Anim. Genet.* 47, 615–617. doi: 10.1111/age.12443

Cosenza, G., Macciotta, N. P. P., Nudda, A., Coletta, A., Ramunno, L., and Pauciullo, A. (2017a). A novel polymorphism in the Oxytocin receptor encoding gene (*OXTR*) affects milk fatty acid composition in Italian Mediterranean river buffalo. *J. Dairy Res.* 84, 170–180. doi: 10.1017/S0022029917000127

Cosenza, G., Pauciullo, A., Coletta, A., Di Francia, A., Feligini, M., Gallo, D., et al. (2011). Translational efficiency of casein transcripts in Mediterranean river buffalo. *J. Dairy Sci.* 94, 5691–5694. doi: 10.3168/jds.2010-4086

Cosenza, G., Pauciullo, A., Colimoro, L., Mancusi, A., Rando, A., Di Berardino, D., et al. (2007). A SNP in the goat *CSN2* promoter region is associated with the absence of β-casein in the milk. *Anim. Genet.* 38, 655–658. doi: 10.1111/j.1365-2052.2007.01649.x

Cosenza, G., Pauciullo, A., Feligini, M., Coletta, A., Colimoro, L., Di Berardino, D., et al. (2009a). A point mutation in the splice donor site of intron 7 in the alphas2-casein encoding gene of the Mediterranean River buffalo results in an allele-specific exon skipping. *Anim. Genet.* 40, 791. doi: 10.1111/j.1365-2052.2009.01897.x

Cosenza, G., Pauciullo, A., Gallo, D., Colimoro, L., D'Avino, A., Mancusi, A., et al. (2008). Genotyping at the *CSN1S1 locus* by PCR-RFLP and AS-PCR in a Neapolitan Goat Population. *Small Rumin. Res.* 74, 84–90. doi: 10.1016/j.smallrumres.2007.03.010

Cosenza, G., Pauciullo, A., Illario, R., Gallo, D., Di Berardino, D., and Ramunno, L. (2005). A preliminary analysis of the goat lactoferrin encoding gene. *Ital. J. Anim. Sci.* 4, 49–51. doi: 10.4081/ijas.2005.2s.49

Cosenza, G., Pauciullo, A., Macciotta, N. P. P., Apicella, E., Steri, R., La Battaglia, A., et al. (2015). Mediterranean River Buffalo *CSN1S1* gene: search for polymorphisms and association studies. *Anim. Prod. Sci.* 55, 654–660. doi: 10.1071/AN13438

Craig, N. L. (2002). Mobile DNA II. *ASM Press*,Washington DC. *pp.xviii*, 1204.

D'Ambrosio, C., Arena, S., Salzano, A. M., Renzone, G., Ledda, L., and Scaloni, A. (2008). Proteomic characterization of water buffalo milk fractions describing PTM of major species and the identification of minor components involved in nutrient delivery and defense against pathogens. *Proteomics* 8, 3657–3666. doi: 10.1002/pmic.200701148

Dong, B., and Zhao, F. Q. (2007). Involvement of the ubiquitous Oct-1 transcription factor in hormonal induction of beta-casein gene expression. *Biochem J.* 401, 57–64. doi: 10.1042/BJ20060570

Farrell, H. M. J., Jimenez-Flores, R., Bleck, G. T., Brown, E. M., Butler, J. E., Creamer, L. K., et al. (2004). Nomenclature of the proteins of cows' milk–sixth revision. *J. Dairy Sci.* 87, 1641–1674. doi: 10.3168/jds.S0022-0302(04)73319-6

Feligini, M., Bonizzi, I., Buffoni, J. N., Cosenza, G., and Ramunno, L. (2009). Identification and Quantification of alpha(S1), alpha(S2), beta, and kappa-Caseins in Water Buffalo Milk by Reverse Phase-High Performance Liquid Chromatography and Mass Spectrometry. *J. Agric. Food Chem.* 57, 2988–2992. doi: 10.1021/jf803653v

Formaggioni, P., Summer, A., Malacarne, M., and Mariani, P. (1999). *Milk Protein Polymorphism: Detention and Diffusion of the Genetic Variants in Bos Genus*. Annali della Facoltà di Medicina Veterinaria, Università di Parma 19, 127–165.

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., et al. (2002). The structure of haplotype blocks in the human genome. *Science* 296, 2225–2229. doi: 10.1126/science.1069424

Gantner, V., Mijić, P., Baban, M., Škrti,ć, Z., and Turalija, A. (2015). The overall and fat composition of milk of various species. *Mljekarstvo: Journal for dairy production and processing improvement* 65, 223–231. doi: 10.15567/mljekarstvo.2015.0401

German, J. B. (2011). Dietary lipids from an evolutionary perspective: sources, structures and functions. *Matern. Child. Nutr.* 7, 2–16. doi: 10.1111/j.1740-8709.2011.00300.x

Giambra, I. J., Chianese, L., Ferranti, P., and Erhardt, G. (2010). Short communication: molecular genetic characterization of ovine alpha(S1)-casein allele H caused by alternative splicing. *J. Dairy Sci.* 93, 792–795. doi: 10.3168/jds.2009-2615

Giambra, I. J., and Erhardt, G. (2011). Molecular genetic characterization of ovine CSN1S2 variants C and D reveal further important variability within CSN1S2. *Anim. Genet.* 43, 642–645. doi: 10.1111/j.1365-2052.2011.02299.x

Goossens, M., and Kan, Y. W. (1981). DNA analysis in the diagnosis of hemoglobin disorders. *Methods Enzymol.* 76, 805–817. doi: 10.1016/0076-6879(81)76159-7

Gorban, A. M. S., and Izzeldin, O. M. (1999). Study on cholesteryl ester fatty acids in camel and cow milk lipid. *Int. J. Food Sci.* 34, 229–234. doi: 10.1046/j.1365-2621.1999.00254.x

Groenen, M. A., Dijkhof, R. J., Verstege, A. J., and van der Poel, J. J. (1993). The complete sequence of the gene encoding bovine alpha s2-casein. *Gene* 123,187–193. doi: 10.1016/0378-1119(93)90123-k

Gu, M., Cosenza, G., Iannaccone, M., Macciotta, N. P. P., Guo, Y., Di Stasio, L., et al. (2019). The single nucleotide polymorphism g.133A>C in the stearoyl CoA desaturase gene (SCD) promoter affects gene expression and quali-quantitative properties of river buffalo milk. *J. Dairy Sci.* 102, 442–451. doi: 10.3168/jds.2018-15059

Gu, M., Cosenza, G., Nicolae, I., Bota, A., Guo, Y., Di Stasio, L., et al. (2017). Transcript analysis at DGAT1 reveals different mRNA profiles for river buffaloes with extreme phenotypes for milk fat. *J. Dairy Sci.* 100, 8265–8276. doi: 10.3168/jds.2017-12771

Han, J. S. (2010). Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. *Mob. DNA* 1:15. doi: 10.1186/1759-8753-1-15

Iamartino, D., Nicolazzi, E. L., Van Tassell, C. P., Reecy, J. M., Fritz-Waters, E. R., Koltes, J. E., et al. (2017). Design and validation of a 90K SNP genotyping 1 assay for the Water Buffalo (Bubalus bubalis). *Plos One*, 12, e0185220. doi: 10.1371/journal.pone.0185220

Ibeagha-Awemu, E. M., Prinzenberg, E. M., Jann, O. C., Lühken, G., Ibeagha, A. E., Zhao, X., et al. (2007). Molecular characterization of bovine CSN1S2*B and extensive distribution of zebu-specific milk protein alleles in European cattle. *J. Dairy Sci.* 90, 3522–3529. doi: 10.3168/jds.2006-679

Imamura, F., Fretts, A., Marklund, M., Ardisson Korat, A. V., Yang, W. S., Lankinen, M., et al. (2018). Fatty acid biomarkers of dairy fat consumption and incidence of type 2 diabetes: A pooled analysis of prospective cohort studies. *PLoS Med.* 15:e1002670. doi: 10.1371/journal.pmed.1002670

Jansa Pérez, M. J., Leroux, C., Sanchez Bonastre, A., and Martin, P. (1994). Occurrence of a LINE sequence in the 3'UTR of the goat αs1-casein E encoding allele associated with a reduced protein synthesis level. *Gene* 147, 177–179. doi: 10.1016/0378-1119(94)90063-9

Kim, J. J., Yu, J., Bag, J., Bakovic, M., and Cant, J. P. (2015). Translation attenuation via 3′ terminal codon usage in bovine csn1s2 is responsible for the difference in αs2- and β-casein profile in milk. *RNA Biol.* 12, 354–367. doi: 10.1080/15476286.2015.1017231

Kishore, A., Sodhi, M., Mukesh, M., Mishra, B. P., and Sobti, R. C. (2013). Sequence analysis and identification of new variations in the 5′-flanking region of αS2-casein gene in Indian zebu cattle. *Mol. Biol. Rep.* 40, 4473–4481. doi: 10.1007/s11033-013-2539-x

Kuss, A. W., Gogol, J., and Geldermann, H. (2003). Associations of a polymorphic AP-2 binding site in the 5′-flanking region of the bovine beta-lactoglobulin gene with milk proteins. *J. Dairy Sci.* 86, 2213–2218. doi: 10.3168/jds.s0022-0302(03)73811-9

Lan, X. Y., Chen, H., Zhang, R. F., Tian, Y., Zhang, Y. D., Fang, X. T., et al. (2005). Association of polymorphisms of *CSN1S2* gene with average milk yield and body sizes indexes in Xinong Saanen dairy goat. *Acta Vet. Zootech. Sin.* 36, 318–322.

Lenstra, J. A., Van Boxtel, J. A. F., Zwaagstra, K. A., and Schwerin, M. (1993). Short interspersed nuclear elements (SINE) sequences of the Bovidae. *Anim. Genet.* 24, 33–39. doi: 10.1111/j.1365-2052.1993.tb00916.x

Liefers, S. C., Veerkamp, R. F., Te Pas, M. F. W., Delavaud, C., Chilliard, Y., Platje, M., et al. (2005). Leptin promoter mutations affect leptin levels and performance traits in dairy cows. *Anim. Genet.* 36, 111–118. doi: 10.1111/j.1365-2052.2005.01246.x

Low, W. Y., Tearle, R., Bickhart, D. M., Rosen, B. D., Kingan, S. B., Swale, T., et al. (2019). Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity. *Nat Commun.* 10:260. doi: 10.1038/s41467-018-08260-0

Luo, X., Zhou, Y., Zhang, B., Zhang, Y., Wang, X., Feng, T., et al. (2020). Understanding divergent domestication traits from the whole-genome sequencing of swamp and river buffalo populations. *Natl. Sci. Rev.* 7, 686–701. doi: 10.1093/nsr/nwaa024

Malewski, T. (1998). Computer analysis of distribution of putative cis- and trans-regulatory elements in milk protein gene promoters. *Biosystems* 45, 29–44. doi: 10.1016/s0303-2647(97)00059-2

Markiewicz-Keszycka, M., Czyzak-Runowska, G., Lipińska, P., and Wójtowski, J. (2013). Fatty acid profile of milk - a review. *Bull. Vet. Inst. Pulawy* 57, 135–139. doi: 10.2478/bvip-2013-0026

Martin, P., Szymanowska, M., Zwierzchowski, L., and Leroux, C. (2002). The impact of genetic polymorphisms on the protein composition of ruminant milks. *Reprod Nutr Dev.* 42, 433–459. doi: 10.1051/rnd:2002036

Masina, P., Rando, A., Di Gregorio, P., Cosenza, G., and Mancusi, A. (2007). Water buffalo kappa-casein gene sequence. *Ital. J. Anim. Sci.* 6, 353–355. doi: 10.4081/ijas.2007.s2.353

Mensink, R. P., Zock, P. L., Kester, A. D., and Katan, M. B. (2003). Effects of dietary fatty acids and carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and apolipoproteins: a meta-analysis of 60 controlled trials. *Am. J. Clin. Nutr.* 77, 1146–1155. doi: 10.1093/ajcn/77.5.1146

Misra, S. S., Sharma, A. A., Bhattacharya, T. K., Kumar, P., and Roy, S. S. (2008). Association of breed and polymorphism of αs1- and αs2-casein genes with milk quality and daily milk and constituent yield traits of buffaloes (*Bubalus bubalis*). *Buffalo Bulletin* 27, 294–301.

Naydenova, N., Iliev, T., and Mihaylova, G. (2013). Fatty acids and lipid indices of buffalo milk yogurt. *Agricultural Science and Technology* 5, 331–334.

Noce, A., Pazzola, M., Dettori, M. L., Amills, M., Castell,ó, A., Cecchinato, A., et al. (2016). Variations at regulatory regions of the milk protein genes are associated with milk traits and coagulation properties in the Sarda sheep. *Anim. Genet.* 47, 717–726. doi: 10.1111/age.12474

Okada, N., and Hamada, M. (1997). The 3′ ends of tRNA derived SINEs originated from the 3′ ends of LINEs: a new example from the bovine genome. *J. Mol. Evol.*. 44(1), S52–S56. doi: 10.1007/pl00000058

Onami, J., Nikaido, M., Mannen, H., and Okada, N. (2007). Genomic expansion of the Bov-A2 retroposon relating to phylogeny and breed management. *Mamm. Genome* 18, 187–196. doi: 10.1007/s00335-007-9000-1

Ordovás, L., Roy, R., Pampín, S., Zaragoza, P., Osta, R., Rodríguez-Rey, J. C., et al. (2009). The g.763G>C SNP of the bovine FASN gene affects its promoter activity via Sp-mediated regulation: implications for the bovine lactating mammary gland. *Physiol. Genomics* 34,144–148. doi: 10.1152/physiolgenomics.00043.2008

Ozdemir, M., Kopuzlu, S., Topal, M., and Bilgin, O. C. (2018). Relationships between milk protein polymorphisms and production traits in cattle: a systematic review and meta-analysis. *Arch. Anim. Breed.* 61, 197–206. doi: 10.5194/aab-61-197-2018

Pauciullo, A., Cosenza, G., D'avino, A., Colimoro, L., Nicodemo, D., Coletta, A., et al. (2010). Sequence analysis and genetic variability of stearoyl CoA desaturase (SCD) gene in the Italian Mediterranean river buffalo. *Mol. Cell. Probes* 24, 407–410. doi: 10.1016/j.mcp.2010.07.009

Pauciullo, A., Cosenza, G., Steri, R., Coletta, A., Jemma, L., Feligini, M., et al. (2012a). An association analysis between *OXT* genotype and milk yield and flow in Italian Mediterranean river buffalo. *J. Dairy Res.* 79, 150–156. doi: 10.1017/S0022029911000914

Pauciullo, A., Cosenza, G., Steri, R., Coletta, A., La Battaglia, A., Di Berardino, D., et al. (2012b). A single nucleotide polymorphism in the promoter region of River buffalo Stearoyl CoA Desaturase gene (SCD) is associated with milk yield. *J. Dairy Res.* 79, 429–435. doi: 10.1017/S0022029912000507

Pauciullo, A., Gauly, M., Cosenza, G., Wagner, H., and Erhardt, G. (2017). Lama glama αs1-casein: identification of new polymorphisms at the *CSN1S1* gene. *J. Dairy Sci.* 100, 1282–1289. doi: 10.3168/jds.2016-11918

Pauciullo, A., Giambra, I. J., Iannuzzi, L., and Erhardt, G. (2014). The β-casein in camels: molecular characterization of the *CSN2* gene, promoter analysis and genetic variability. *Gene* 547, 159–168. doi: 10.1016/j.gene.2014.06.055

Pauciullo, A., Shuiep, E. S., Cosenza, G., Ramunno, L., and Erhardt, G. (2013). Molecular characterization and genetic variability at k-casein gene (*CSN3*) in Sudanese camels. *Gene* 513, 22–30. doi: 10.1016/j.gene.2012.10.083

Pauciullo, A., Shuiep, E. T., Ogah, M. D., Cosenza, G., Di Stasio, L., and Erhardt, G. (2019). Casein gene cluster in camelids: a comparative analysis, haplotype structure and physical mapping. *Front. Genet.* 10:748, 1–18. doi: 10.3389/fgene.2019.00748

Peña-Serna, C., Gómez-Ramirez, B., and Zapata-López, N. (2019). Nutritional Aspects of Ghee Based on Lipid Composition. *Pak. J. Nutr.* 18, 1107–1114. doi: 10.3923/pjn.2019.1107.1114

Perna, A., Intaglietta, I., Simonetti, A., and Gambacorta, E. (2016). The influence of casein haplotype on morphometric characteristics of fat globules and fatty acid composition of milk in Italian Holstein cows. *J. Dairy Sci.* 99, 2512–2519. doi: 10.3168/jds.2015-10397

Praagman, J., de Jonge, E. A., Kiefte-de Jong, J. C., Beulens, J. W., Sluijs, I., Schoufour, J. D., et al. (2016). Dietary Saturated Fatty Acids and Coronary Heart Disease Risk in a Dutch Middle-Aged and Elderly Population. *Arterioscler. Thromb. Vasc. Biol.* 36, 2011–2018. doi: 10.1161/ATVBAHA.116.307578

Ramunno, L., Cosenza, G., Pappalardo, M., Longobardi, E., Gallo, D., Pastore, N., et al. (2001b). Characterization of two new alleles at the goat *CSN1S2 locus*. *Anim. Genet.* 32, 264–268. doi: 10.1046/j.1365-2052.2001.00786.x

Ramunno, L., Cosenza, G., Rando, A., Illario, R., Gallo, D., Di Berardino, D., et al. (2004). The goat αs1-casein gene: gene structure and promoter analysis. *Gene* 334, 105–111. doi: 10.1016/j.gene.2004.03.006

Ramunno, L., Cosenza, G., Rando, A., Pauciullo, A., Illario, R., Gallo, D., et al. (2005). Comparative analysis of gene sequence of goat *CSN1S1* F and N alleles and characterization of *CSN1S1* transcript variants in mammary gland. *Gene,* 345, 289–299. doi: 10.1016/j.gene.2004.12.003

Ramunno, L., Longobardi, E., Pappalardo, M., Rando, A., Di Gregorio, P., Cosenza, G., et al. (2001a). An allele associated with a non dectable amount of αs2 casein in goat milk. *Anim. Genet.* 32, 19–26. doi: 10.1046/j.1365-2052.2001.00710.x

Rando, A., Di Gregorio, P., Ramunno, L., Mariani, P., Fiorella, A., Senese, C., et al. (1998). Characterization of the CSN1AG allele of the bovine αs1-casein *locus* by the insertion of a relict of a long interspersed element. *J. Dairy Sci.* 81, 1735–1742. doi: 10.3168/jds.S0022-0302(98)75741-8

Romano, R., Borriello, I., Chianese, L., and Addeo, F. (2008). Quali-quantitative determination of triglicerydic, fatty acids and CLA in "Mozzarella di Bufala Campana" by high resolution gascromatography (HRGC). *Progress in Nutrition* 10, 22–29.

Schennink, A., Stoop, W. M., Visker, M. H. P. W., Heck, J. M. L., Bovenhuis, H., van der Poel, J. J., et al. (2007). DGAT1 underlies large genetic variation in milk-fat composition of dairy cows. *Anim. Genet.* 38, 467–473. doi: 10.1111/j.1365-2052.2007.01635.x

Selvaggi, M., Laudadio, V., Dario, C., and Tufarelli, V. (2014a). Investigating the genetic polymorphism of sheep milk proteins: a useful tool for dairy production. *J. Sci. Food Agric.* 94, 3090–3099. doi: 10.1002/jsfa.6750

Selvaggi, M., Laudadio, V., Dario, C., and Tufarelli, V. (2014b). Major proteins in goat milk: an updated overview on genetic variability. *Mol. Biol. Rep.* 41, 1035–1048. doi: 10.1007/s11033-013-2949-9

Shimamura, M., Abe, H., Nikaido, M., Ohshima, K., and Okada, N. (1999). Genealogy of families of SINEs in cetaceans and artiodactyls: The presence of a huge superfamily of tRNA(Glu)-derived families of SINEs. *Mol. Biol. Evol.* 16, 1046–1060. doi: 10.1093/oxfordjournals.molbev.a026194

Stoop, W. M., van Arendonk, J. A. M., Heck, J. M. L., van Valenberg, H. J. F., and Bovenhuis, H. (2008). Genetic Parameters for Major Milk Fatty Acids and Milk Production Traits of Dutch Holstein-Friesians. *J. Dairy Sci.* 91, 385–394. doi: 10.3168/jds.2007-0181

Sukla, S., Bhattacharya, T. K., Venkatachalapathy, R. T., Kumar, P., and Sharma, A. (2006). Cloning and characterization of alpha(s2)-casein gene of Riverine buffalo. *DNA Seq.* 17, 458–464. doi: 10.1080/10425170600886474

Szymanowska, M., Malewski, T., and Zwierzchowski, L. (2004a). Transcription factor binding to variable nucleotide sequences in 5′ flanking regions of bovine casein genes. *Int. Dairy J.* 14,103–115. doi: 10.1016/S0958-6946(03)00153-5

Szymanowska, M., Siadkowska, E., Lukaszewicz, M., and Zwierzchowski, L. (2004b). Association of nucleotide-sequence polymorphism in the 5′-flanking regions of bovine casein genes with casein content in cow's milk. *Lait* 84, 579–590. doi: 10.1051/lait:2004030

Szymanowska, M., Strzalkowska, N., Siadkowska, E., Krzyzewski, J., Ryniewicz, Z., and Zwierzchowski, L. (2003). Effects of polymorphism at 5'-noncoding S2-caseingenes on selected milk S1- and regions (promoters) of production traits in Polish Black-and White cows. *Anim. Sci. Pap. Rep.* 21, 97–108.

Vacca, G. M., Stocco, G., Dettori, M. L., Pira, E., Bittante, G., and Pazzola, M. (2018). Milk yield, quality, and coagulation properties of 6 breeds of goats: Environmental and individual variability. *J. Dairy Sci.* 101, 7236–7247. doi: 10.3168/jds.2017-14111

Vinesh, P. V., Brahma, B., Kaur, R., Datta, T. K., Goswami, S. L., and De, S. (2013). Characterization of β-casein gene in Indian riverine buffalo. *Gene* 527, 683–688. doi: 10.1016/j.gene.2013.06.029

Wessels, G., Hamann, H., Erhardt, G., and Distl, O. (2004). Genotype effects of milk protein polymorphisms on milk production in East Friesian dairy sheep. *Berliner Munchener Tierarztl Wochenschr* 117, 414–419

Yang, S., Li, C., Xie, Y., Cui, X., Li, X., Wei, J., et al. (2015). Detection of functional polymorphisms influencing the promoter activity of the SAA2 gene and their association with milk production traits in Chinese Holstein cows. *Anim. Genet.* 46, 591–598. doi: 10.1111/age.12332

Yue, X. P., Fang, Q., Zhang, X., Mao, C. C., Lan, X. Y., Chen, H., et al. (2013). Effects of *CSN1S2* genotypes on economic traits in chinese dairy goats. *Asian-Australas. J. Anim. Sci.* 26, 911–915. doi: 10.5713/ajas.2013.13018

Zidi, A., Serradilla, J. M., Jordana, J., Carrizosa, J., Urrutia, B., Polvillo, O., et al. (2010). Pleiotropic effects of the goat prolactin receptor genotype on milk fatty acid composition. *Domest. Anim. Endocrinol.* 39, 85–89. doi: 10.1016/j.domaniend.2010.02.005

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 08 March 2021
doi: 10.3389/fgene.2021.633017

Genetic Analysis of Persistency for Milk Fat Yield in Iranian Buffaloes (*Bubalus bubalis*)

Mohammad Ali Nazari[1], Navid Ghavi Hossein-Zadeh[1]*, Abdol Ahad Shadparvar[1] and Davood Kianzad[2]

[1] Department of Animal Science, Faculty of Agricultural Sciences, University of Guilan, Rasht, Iran, [2] Animal Breeding Center and Promotion of Animal Products, Karaj, Iran

OPEN ACCESS

Edited by:
Tingxian Deng,
Institute of Buffalo (CAAS), China

Reviewed by:
Madhu Tantia,
National Bureau of Animal Genetic Resources (NBAGR), India
Hossam Eldin Rushdi Ahmed Ali Osman,
Cairo University, Egypt

***Correspondence:**
Navid Ghavi Hossein-Zadeh
nhosseinzadeh@guilan.ac.ir;
navid.hosseinzadeh@gmail.com
orcid.org/0000-0001-9458-5860

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 24 November 2020
Accepted: 18 February 2021
Published: 08 March 2021

Citation:
Nazari MA, Ghavi Hossein-Zadeh N, Shadparvar AA and Kianzad D (2021) Genetic Analysis of Persistency for Milk Fat Yield in Iranian Buffaloes (Bubalus bubalis). Front. Genet. 12:633017. doi: 10.3389/fgene.2021.633017

This study aimed to estimate heritabilities and genetic trends for different persistency measures for milk fat yield and their genetic correlations with 270-day milk yield in Iranian buffaloes. The records of test-day milk fat yield belonging to the first three lactations of buffaloes within 523 herds consisting of 43,818 records were got from the Animal Breeding Center and Promotion of Animal Products of Iran from 1996 to 2012. To fit the lactation curves based on a random regression test-day model, different orders of Legendre polynomial (LP) functions were selected. Three persistency measures were altered according to the specific condition of the lactation curve in buffaloes: (1) The average of estimated breeding values (EBVs) for test day fat yield from day 226 to day 270 as a deviation from the average of EBVs from day 44 to day 62 (PM_1), (2) A summation of contribution for each day from day 53 to day 247 as a deviation from day 248 (PM_2), and (3) The difference between EBVs for day 257 and day 80 (PM_3). The estimates of heritability for PM_1, PM_2, and PM_3 ranged from 0.20 to 0.48, from 0.36 to 0.47, and from 0.19 to 0.35 over the first three lactations, respectively. The estimate of genetic trends for different persistency measures of milk fat yield was not significant over the lactations ($P > 0.05$). Genetic correlation estimates between various measures of persistency were generally high over the first three lactations. Also, genetic correlations estimates between persistency measures and 270-day milk yield were mostly low and varied from 0.00 to 0.24 (between PM_1 and 270-day milk yield), from −0.19 to 0.13 (between PM_2 and 270-day milk yield), and from −0.02 to 0.00 (between PM_1 and 270-day milk yield) over the first three lactations, respectively. Persistency measures that showed low genetic correlations with milk fat yield were considered the most suitable measures in selection schemes. Besides, medium to high heritability estimates for different persistency measures for milk fat yield indicated that relevant genetic variations detected for these characters could be regarded in outlining later genetic improvement programs of Iranian buffaloes.

Keywords: dairy buffalo, genetic parameter, genetic trend, lactation persistency, random regression model

INTRODUCTION

One important step for reaching self-sufficiency in any country is to identify the productive potential of native breeds of animals. The great adaptability of native animals to harsh conditions such as high environmental humidity and temperature, irregular rainfall, the incidence of different diseases, weak management practices, and low quality of feeds causes native buffaloes of Iran to play an important role in supplying milk and meat as major protein sources. Currently, many Asian countries depend mainly on buffalo as a source of milk and dairy products, especially in rural areas (Safari et al., 2018).

One of the main factors in determining the total milk production over a lactation period is persistency (Muir, 2004). Persistency is defined as the potential of an animal to maintain milk yield at a high extent after reaching the peak of production. The other definition of persistency is the gradual decrease of daily milk production after reaching the peak of the lactation curve (Togashi and Lin, 2004). The major cause for the worth of buffaloes with more persistent curves is that they can relatively satisfy most parts of their nutrient requirements from roughages (Sölkner and Fuchs, 1987). Therefore, not only metabolic problems, reproduction disorders, and diseases are lower in cows with more persistent lactations, but also production costs would be lower (Dekkers et al., 1998). Determining the method of measuring persistency is a critical point in estimating genetic progress for this trait. However, no general agreement is existent on the most appropriate method to describe the persistency of lactation (Cole and VanRaden, 2006). Various measures were suggested for calculating persistency (Gengler, 1996): measures based on the functions describing persistency; measures based on a fraction of total yield, peak yield, or parts of lactation; and those based on the breeding value of animals derived from analyzing random regression models.

The method used for defining persistency measures would determine the genetic parameter estimates for these measures and their genetic relationship with milk production (Swalve and Gengler, 1999; Jakobsen et al., 2002; Khorshidie et al., 2012). A measure of persistency must have two characteristics: association with lactation curve flatness, and independent explanation from production level. The latter item implies that the genetic correlation between milk yield and persistency measures to be decreased because milk production explains some genetic variance of persistency measures under study (Muir, 2004; Cole and VanRaden, 2006; Khorshidie et al., 2012). The independence of these two traits causes genetic selection for persistency of lactation and total yield to avoid unfavorable consequences of peak yield stress in high-yielding cows. Also, the incidence of metabolic diseases and reproductive disorders would be minimized while high milk production is maintained (Elmaghraby, 2012; Ghavi Hossein-Zadeh et al., 2017).

Previous studies carried out on dairy cattle indicated that lactation persistency positively correlated with favorable reproductive performance and health status (Jakobsen et al., 2002; Muir, 2004). Such favorable correlations along with the positive economic value for persistency would support including lactation persistency in the genetic improvement programs of cattle and buffalo (Dekkers et al., 1998; Khorshidie et al., 2012; Ghavi Hossein-Zadeh et al., 2017).

The random regression models enable fitting random genetic and environmental effects at different stages of lactation, which results in higher accuracy of estimated breeding values (EBVs) compared with other statistical models (Li et al., 2020). These models provide insights about the temporal variation of biological processes and physiological implications underlying the studied traits. Therefore, random regression models generate relevant information to be exploited in breeding programs (Oliveira et al., 2019). The functions generally used to model the lactation curve include Wood's model (Wood, 1967), Wilmink's function (Wilmink, 1987), spline function (White et al., 1999), and Legendre polynomial (LP) function (Kirkpatrick et al., 1990). Because of variations in production environments and management systems, optimal functions for test-day models in various countries may be different (Mrode et al., 2003). But several studies have indicated that LPs performed well in random regression test-day models (Li et al., 2020).

Milk constituents can be used as a simple indicator of the nutritional status of the lactating animals. Because of the dilution effect, milk fat percentage shows the opposite direction of the lactation curve for milk yield (Eicher, 2004; Ghavi Hossein-Zadeh, 2016), but fat yield follows a variation trend similar to milk yield over the lactation. When trying to apply milk composition as a nutritional evaluation tool, these fluctuations should be noticed. Although several researchers have studied the genetic analysis of the persistency for milk yield and components in dairy cattle (Cole and Null, 2009; Khorshidie et al., 2012; Canaza-Cayo et al., 2015), limited studies have been performed to estimate genetic parameters of persistency for milk production traits in buffaloes (Ghavi Hossein-Zadeh et al., 2017). Therefore, the objective of the present study was to estimate the heritability and genetic trend of distinct persistency measures for milk fat yield and their genetic correlations with 270-day milk yield in Iranian buffaloes using random regression test-day models.

MATERIALS AND METHODS

Data

Records of test-day milk fat yield belonging to the first three lactations of Iranian native buffaloes in 523 herds consisting of 43,818 records were provided by the Animal Breeding Center and Promotion of Animal Products of Iran during 1996–2012. According to climatic conditions, Iranian native buffaloes can be grouped into three main classes: Azari ecotype, Kuhzestani ecotype, and Mazandarani or North ecotype (Ghavi Hossein-Zadeh et al., 2012). Borghese (2005) and Ghavi Hossein-Zadeh (2015a,b) described the overall management practices and population structure of buffaloes in Iran. Outliers that appeared to deviate markedly from other observations in the original data set were discarded. Therefore, the subsequent analyses included only production records corresponding to the first three lactations in which days in milk (DIM) were between 5 and 270. Calving ages ranged between 24–60, 39–76, and 54–100 months for the first, second, and third lactations, respectively. The total number of test-day records per animal was from 4 to 9. Summary statistics of the edited data set are presented in **Table 1**. The number of animals, sires, and dams in the pedigree of Iranian buffaloes was 42,285, 549, and 6,376, respectively.

Statistical and Genetic Analysis

Legendre polynomial functions were chosen to fit the lactation curves in the framework of a random regression test-day model for estimating (co)variance components. Model specification and the choice of fixed effects to be included in the model were based on the backward elimination method and variables which were significant at $P < 0.05$ were considered in the model. To obtain the appropriate random regression test-day model for the genetic analysis of test day fat yield, with the minimum number of parameters, different orders of fit for random regression coefficients of additive genetic and permanent environmental effects were evaluated. Also, the optimum set of polynomials was selected according to the logarithm of the likelihood function at the point of conversion and the total number of parameters to be estimated. The difference of these models was based on the LPs applied to fit the covariance functions for additive genetic and permanent environmental effects. The maximum logarithm likelihood of the models was compared and models with the lowest values of this criterion were selected for further analysis. Test day records were analyzed using the following random regression model:

$$\begin{aligned} &\mathrm{Y_{ijmnptv}} \\ &= \mathrm{G_i} + \mathrm{YS_j} + \mathrm{HTD_m} + \sum_{f=0}^{2} \mathrm{c_f}(\mathrm{age_n})^{f} + \sum_{r=0}^{k} \beta_r \emptyset_r(\mathrm{dim_t}) \\ &+ \sum_{r=0}^{k_{a-1}} \alpha_{pr} \emptyset_r\left(\mathrm{dim_t}\right) + \sum_{r=0}^{k_{p-1}} \gamma_{pr} \emptyset_r\left(\mathrm{dim_t}\right) + \mathrm{e_{ijmnptv}} \end{aligned}$$

Where,

$\mathrm{Y_{ijmnptv}}$: test day record i obtained at DIM t of cow p calved at the n^{th} age in herd-test date m,

G_i : fixed effect of i^{th} breed or ecotype,

$\mathrm{YS_j}$: fixed effect of j^{th} calving year-season,

HTD_m: fixed effect of m^{th} herd-test date,

c_f: the f^{th} fixed regression coefficient for calving age,

age_n: the n^{th} calving age,

k: the order of fit for fixed regression coefficients ($k = 2$),

β_r: the r^{th} fixed regression coefficient,

k_a: the order of fit for additive genetic random regression coefficients,

k_p: the order of fit for permanent environmental random regression coefficients,

α_{pr}: the r^{th} random regression coefficient of additive genetic value for p^{th} cow,

γ_{pr}: the r^{th} random regression coefficient of permanent environmental effect for p^{th} cow,

$\emptyset_r\left(\mathrm{dim_t}\right)$: the r^{th} coefficient of LPs evaluated at days in milk t,

$\mathrm{e_{ijmnptv}}$: the random residual error.

All random regression analyses were conducted using the Average Information Restricted Maximum Likelihood (AIREML) algorithm of the WOMBAT program (Meyer, 2006).

Lactation Persistency Measures

The following measures were used to describe lactation persistency in this study. These measures were modified based on the lactation curve conditions of buffaloes and adapted for 270 days lactation period:

1. The average of EBVs for test day fat yield from day 226 to day 270 as a deviation from the average of EBVs from day 44 to day 62 [adapted from Kistemaker (2003)]:

$$PM_1 = \frac{1}{44}\sum_{i=226}^{270} \mathrm{EBV_i} - \frac{1}{21}\sum_{i=44}^{62} \mathrm{EBV_i}$$

2. A summation of contribution for each day from day 53 to day 247 as a deviation from day 248 [adapted from Cobuci et al. (2007) and Jakobsen et al. (2002)]:

$$PM_2 = \sum_{i=53}^{247} (\mathrm{EBV_i} - \mathrm{EBV}_{\ 248})$$

3. The difference between EBVs for day 257 and day 80 [adapted from Cobuci et al. (2004, 2007)]:

$$PM_3 = (\mathrm{EBV}_{257} - \mathrm{EBV}_{\ 80})$$

Small absolute values of the abovementioned measures indicate a high lactation persistency. If $\hat{\alpha}_i$ was a ($k_a \times 1$) vector of the estimates of additive genetic random regression coefficients specific to the animal i, and Z_t was

TABLE 1 | Summary statistics of edited milk fat yield data used in this study.

Days in milk classes	Lactation 1			Lactation 2			Lactation 3		
	N	Mean (kg)	SD (kg)	N	Mean (kg)	SD (kg)	N	Mean (kg)	SD (kg)
5–30	756	0.432	0.225	686	0.461	0.251	654	0.487	0.252
31–60	943	0.426	0.225	956	0.464	0.247	859	0.487	0.252
61–90	1,095	0.488	0.243	985	0.473	0.249	989	0.499	0.257
91–120	1,252	0.477	0.251	1,071	0.492	0.257	1,033	0.508	0.256
121–150	1,176	0.487	0.252	1,013	0.497	0.254	945	0.500	0.263
151–180	1,156	0.474	0.252	1,028	0.489	0.261	906	0.481	0.256
181–210	1,014	0.466	0.252	783	0.480	0.253	711	0.450	0.245
211–240	806	0.444	0.245	611	0.463	0.244	592	0.462	0.255
241–270	569	0.469	0.246	455	0.459	0.244	420	0.433	0.235

a $\left(k_a \times 1\right)$ vector of LP coefficients evaluated at day t, the EBV of animal i for day t was calculated as follows:

$$EBV_{\mathrm{a}} = \sum_{\mathrm{i}=0}^{\mathrm{ka}-1} \mathrm{a_{ij}}\ \emptyset_{\mathrm{j}}\ \left(\mathrm{dim_t}\right) = \hat{a}_{0i}\emptyset_{0t} + \hat{a}_{1i}\emptyset_{1t} + \hat{a}_{2i}\emptyset_{2t} + \ \hat{a}_{3i}\emptyset_{3t}$$

Therefore, the EBV of animal i for 270-day production was obtained by summing the EBVs from day 5 to day 270:

$$EBVT_i \ = \ \sum_{5}^{270} \left(\hat{a}_{0i}\emptyset_{0i}\ \hat{a}_{1i}\emptyset_{1i}\ \hat{a}_{2i}\emptyset_{2i}\ \hat{a}_{3i}\emptyset_{3i}\right)$$

$$= \left(\sum_{5}^{270} \emptyset_{0t}\ \sum_{5}^{270} \emptyset_{0t}\ \sum_{5}^{270} \emptyset_{0t}\ \sum_{5}^{270} \emptyset_{0t}\right)\hat{a}_i \ = \ Z_{c270}\widehat{\mathrm{a}}_{\mathrm{i}}$$

Where, Z_{c270} is a vector of the summations of LPs corresponding to total lactation yield. In addition to the 270-day yield, we could estimate a Z_c corresponding to each persistency measures used in the current study as follows:

For the first lactation fat yield:

$$\begin{aligned}
\mathrm{Zc}_{270} &= (0.7071 \quad 1.42\mathrm{E}{-17} \quad 0.0059)\\
\mathrm{ZcP}_{1\mathrm{g}} &= (0 \quad 0.7839 \quad 0.8491)\\
\mathrm{ZcP}_{2\mathrm{g}} &= (0 \quad 1.6361 \quad 1.4825)\\
\mathrm{ZcP}_{3\mathrm{g}} &= (0 \quad -0.9058 \quad -1.2003)
\end{aligned}$$

For second lactation fat yield:

$$\begin{aligned}
\mathrm{Zc}_{270} &= (0.7071 \quad 1.42\mathrm{E}{-17} \quad 0.0059 \quad -6.7\mathrm{E}{-18} \quad 0.0081)\\
\mathrm{ZcP}_{1\mathrm{g}} &= (0 \quad 0.7839 \quad 0.8491 \quad -0.0664 \quad 0.9943)\\
\mathrm{ZcP}_{2\mathrm{g}} &= (0 \quad 1.6361 \quad 1.4825 \quad 0.0645 \quad 0.8387)\\
\mathrm{ZcP}_{3\mathrm{g}} &= (0 \quad -0.9058 \quad -1.2003 \quad -0.3943 \quad 0.1711)
\end{aligned}$$

For third lactation fat yield:

$$\begin{aligned}
\mathrm{Zc}_{270} &= (0.7071 \quad 1.42\mathrm{E}{-17} \quad 0.0059 \quad -6.7\mathrm{E}{-18} \quad 0.0081)\\
\mathrm{ZcP}_{1\mathrm{g}} &= (0 \quad 0.7839 \quad 0.8491 \quad -0.0664 \quad 0.9943)\\
\mathrm{ZcP}_{2\mathrm{g}} &= (0 \quad 1.6361 \quad 1.4825 \quad 0.0645 \quad 0.8387)\\
\mathrm{ZcP}_{3\mathrm{g}} &= (0 \quad -0.9058 \quad -1.2003 \quad -0.3943 \quad 0.1711)
\end{aligned}$$

Estimation of Genetic Parameters and Genetic Trends

The following formulas were applied to estimate additive genetic, permanent environmental and residual variances and heritabilities for different measures of persistency for fat yield and 270-day milk yield:

$$\begin{aligned}
&\sigma_{a_{(p_i,EBV_{270MY})}} = Z_{c_{p_{ig}}} K_a Z_{c_{270MYg}}{}'\\
&\sigma^2_{pe_{p_i}} = Z_{c_{p_{ipe}}} K_{pe} Z_{c_{p_{ipe}}}{}'\\
&h^2_{p_i} = \frac{\sigma^2_{a_{p_i}}}{\sigma^2_{ph_{p_i}}}\\
&\sigma^2_{a_{270MY}} = Z_{c_{270MYg}} K_a Z_{c_{270MYg}}{}'\\
&\sigma^2_{pe_{270MY}} = Z_{c_{270MYpe}} K_a Z_{c_{270MYpe}}{}'
\end{aligned}$$

$$\begin{aligned}
&\sigma^2_e = 8.85Kg^2\\
&\sigma^2_{e_{p_1}} = \left(\tfrac{1}{44} + \tfrac{1}{18}\right)\sigma^2_e\\
&\sigma^2_{e_{p_2}} = 48620\sigma^2_e\\
&\sigma^2_{e_{p_3}} = 2\sigma^2_e\\
&\sigma^2_{e_{270MY}} = 266\sigma^2_e
\end{aligned}$$

Where, K_{a} and K_{pe} are matrices of direct additive genetic and permanent environmental (co)variances of random regression coefficients, $\sigma^2_{a_{p_i}}$, $\sigma^2_{pe_{p_i}}$, $\sigma^2_{ph_{p_i}}$, and $h^2_{p_i}$ are the additive genetic, permanent environmental, phenotypic variances, and heritability estimate for ith persistency measure and $\sigma^2_{a_{270MY}}$, $\sigma^2_{pe_{270MY}}$, $\sigma^2_{ph_{270MY}}$, and h^2_{270MY} are the additive genetic, permanent environmental, phenotypic variances, and heritability estimate for 270-day milk yield, respectively. σ^2_e is a constant residual variance estimated for each day of lactation and $\sigma^2_{e_{p_1}}$, $\sigma^2_{e_{p_2}}$, $\sigma^2_{e_{p_3}}$, and $\sigma^2_{e_{270MY}}$ are residual variances for persistency measures PM_1, PM_2, PM_3, and 270-day milk yield, respectively. Also, phenotypic variances were obtained by summing the genetic, permanent environmental, and residual variances for different persistency measures and milk yield. Estimates of genetic correlations among persistency measures and with 270-day milk yield were obtained as follows:

$$\sigma_{a_{(p_i,p_j)}} = Z_{c_{p_{ig}}} K_a Z_{c_{p_{jg}}}{}'$$

$$\sigma_{a_{(p_i,EBV_{270MY})}} = Z_{c_{p_{ig}}} K_a Z_{c_{270MYg}}{}'$$

$$R_{a_{(p_i,p_j)}} = \frac{\sigma_{a_{(p_i,p_j)}}}{\sqrt{\left(\sigma^2_{a_{p_i}}\right)\left(\sigma^2_{a_{p_j}}\right)}}$$

$$R_{a_{(p_i,EBV_{270MY})}} = \frac{\sigma_{a_{(p_i,EBV_{270MY})}}}{\sqrt{\left(\sigma^2_{a_{p_i}}\right)\left(\sigma^2_{a_{270MY}}\right)}}$$

Where, $\sigma_{a_{(p_i,p_j)}}$, $\sigma_{a_{(p_i,EBV_{270MY})}}$, $R_{a_{(p_i,p_j)}}$, and $R_{a_{(p_i,EBV_{270MY})}}$ are genetic covariances and correlations between persistency measures and 270-day milk yield, respectively. Estimates of genetic trends for persistency measures were obtained by regressing the average EBVs on the calving year of animals.

RESULTS

The orders of fit for different random regression test-day models of milk fat production are given in **Table 2**. The maximum log-likelihood values of test-day models 1, 10, and 10 differed significantly ($P < 0.05$) from the other models for fat yield in the first three lactations, respectively. Thus, models 1, 10, and 10 were chosen to fit the additive genetic and permanent environmental effects for the analysis of fat production in the first three lactations of buffaloes, respectively.

Heritability estimates of persistency measures for fat production and estimates of genetic correlation among distinct fat yield persistency measures with each other and with 270-day milk production in Iranian buffaloes are presented in **Table 3**. Heritability estimates for PM_1, PM_2, and PM_3 ranged between 0.20–0.48, 0.36–0.47, and 0.19–0.35 for the first, second, and third lactations, respectively. In general, heritability estimates fluctuated largely among lactations and persistency measures. The highest estimate of heritability was observed for PM_1 in the third lactation (0.48), while the lowest one was recorded for PM_3 also in the third lactation.

Genetic correlation estimates among various measures of persistency were generally high and ranged from 0.98 to 0.99 (between PM_1 and PM_2), from −0.98 to −0.87 (between PM_1 and PM_3), and from −0.99 to −0.95 (between PM_2 and PM_3) over the first three lactations, respectively. Also, genetic correlation estimates between persistency measures and milk yield were mostly low and varied from 0.00 to 0.24 (between PM_1 and 270-day milk yield), from −0.19 to 0.13 (between PM_2 and 270-day milk yield), and from −0.02 to 0.00 (between PM_3 and 270-day milk yield) across the first three lactations, respectively (**Table 3**).

Variation of milk yield and milk fat yield across the first three lactations of Iranian buffaloes are depicted in **Figures 1**, **2**. The trend of observed milk yield and milk fat yield for all lactations

TABLE 2 | Orders of fit for different random regression test-day models of milk fat yield evaluated in this study.

Model	Order of fit		NP[3]	Maximum log-likelihood		
	k_a[1]	k_{pe}[2]		Lactation 1	Lactation 2	Lactation 3
1	3	3	21	−5,593.84*	−7,188.50	−6,996.65
2	3	4	25	−5,478.18	−7,187.89	−7,021.78
3	3	5	30	−5,486.11	−7,190.54	−7,030.21
4	3	6	36	−5,487.47	−7,194.83	−7,046.79
5	4	3	25	−5,577.40	−7,190.17	−7,026.02
6	4	4	29	−5,483.49	−7,191.63	−7,025.94
7	4	5	34	−5,482.32	−7,192.05	−7,026.09
8	4	6	40	−5,492.64	−7,198.73	−7,049.79
9	5	5	39	−5,495.48	−7,200.18	−7,041.73
10	5	6	45	−5,495.71	−7,204.1*	−7,050.93*

[1]k_a = orders of fit for additive genetic effects.
[2]k_{pe} = orders of fit for permanent environmental effects.
[3]NP: number of the parameter for estimated variance function.
**Significant at $P < 0.05$.*

TABLE 3 | Heritability estimates of different persistency measures for milk fat yield and genetic correlations among distinct fat yield persistency measures with each other and with 270-day milk production in Iranian buffaloes.

Trait	Lactation 1	Lactation 2	Lactation 3
		Heritability	
PM_1	0.20	0.39	0.48
PM_2	0.47	0.36	0.46
PM_3	0.31	0.35	0.19
		Genetic correlation	
PM_1-PM_2	0.99	0.98	0.98
PM_1-PM_3	−0.98	−0.90	−0.87
PM_2-PM_3	−0.99	−0.95	−0.95
PM_1-270 d MY	0.05	0.00	0.24
PM_2-270 d MY	0.01	−0.19	0.13
PM_3-270 d MY	0.00	0.00	−0.02

PM_1, the average EBVs for test day milk fat yield from day 226 to day 270 as a deviation from the average of EBVs from day 44 to day 62; PM_2, a summation of contribution for each day from day 53 to day 247 as a deviation from day 248; PM_3, The difference between EBVs for day 257 and day 80; 270 d MY, 270-day milk production.

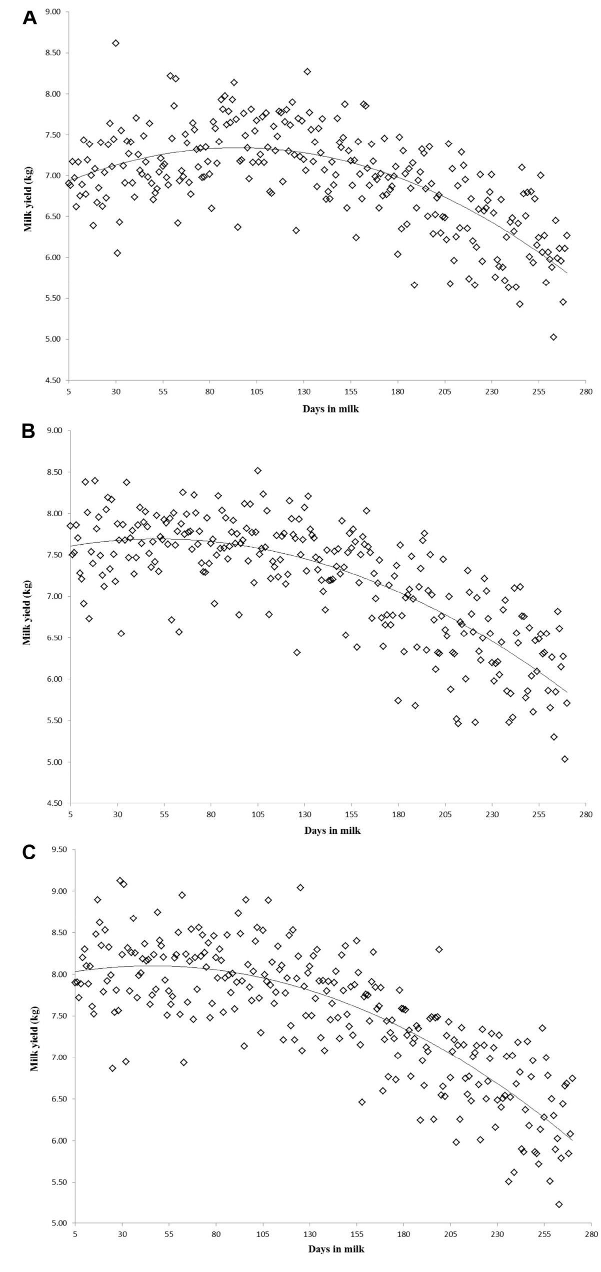

FIGURE 1 | Variation of milk yield across the first **(A)**, second **(B)**, and third **(C)** lactations in Iranian buffaloes.

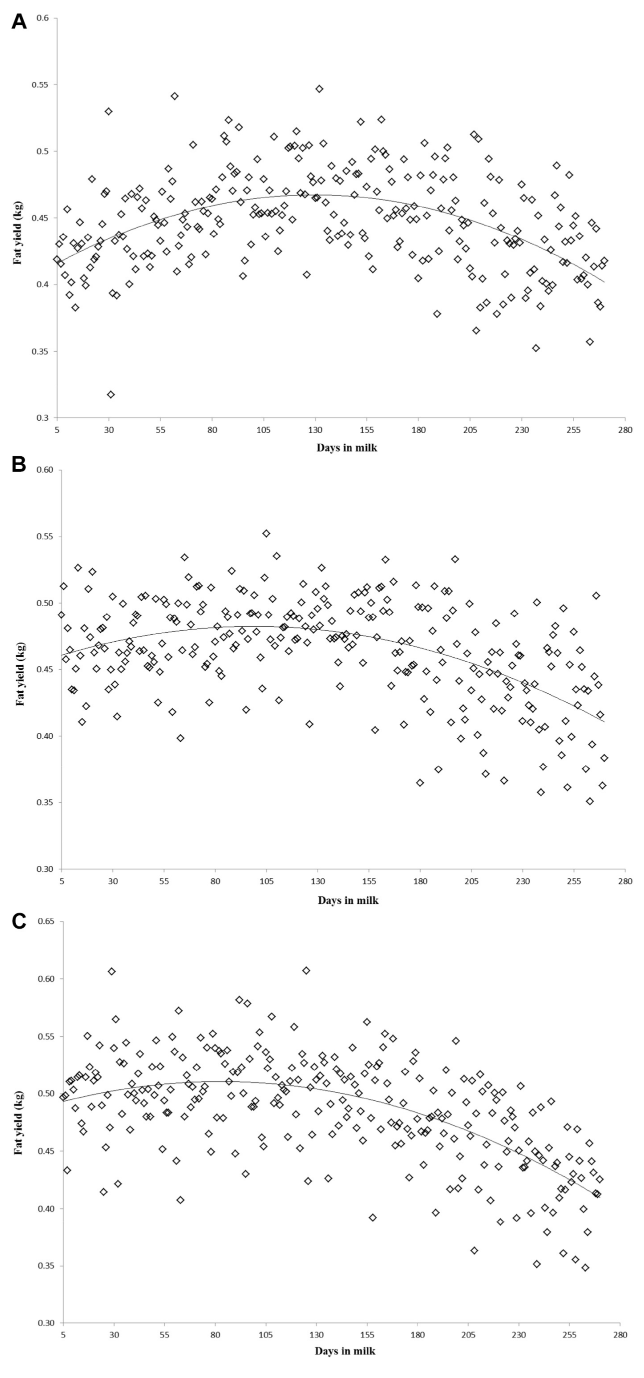

FIGURE 2 | Variation of milk fat yield across the first **(A)**, second **(B)**, and third **(C)** lactations in Iranian buffaloes.

increased from day 5 of lactation to a peak several weeks later, declining thereafter until day 270. Genetic trends of persistency measures for milk fat yield are illustrated in **Table 4**. In general, all estimates are very low and not significant ($P > 0.05$). Therefore, they would not be considered different from zero. Changes in EBVs of buffaloes for three persistency measures of milk fat yield according to calving year and lactations are illustrated in

TABLE 4 | Estimates of genetic trends for various persistency measures of fat production in buffaloes.

Trait	Lactation 1	Lactation 2	Lactation 3
PM_1	−0.00004 ± 0.000035	−0.00007 ± 0.00091	−0.0006 ± 0.0008
PM_2	−0.000013 ± 0.00018	−0.00024 ± 0.00074	−0.0007 ± 0.0008
PM_3	0.000013 ± 0.00019	−0.000085 ± 0.00026	−0.00052 ± 0.00051

PM_1, the average EBVs for test day milk fat production from day 226 to day 270 as a deviation from the average of EBVs from day 44 to day 62; PM_2, a summation of contribution for each day from day 53 to day 247 as a deviation from day 248; PM_3, the difference between EBVs for day 257 and day 80.

Figures 3–5. In general, irregular fluctuations were observed in the annual mean predicted breeding values of animals for different persistency measures across the first three lactations.

DISCUSSION

For many years, the breeding objectives of dairy animals emphasized increased milk yield. But negative genetic associations were observed between numerous functional characters with production traits, and decreases in genetic excellence for fitness and health have been detected in dairy farms (Egger-Danner et al., 2015). The management practices must be directed toward the compensation of these effects and to equalize reproduction performance, metabolic diseases, and udder health vs. enhanced production to maximize profit without any negative impact on animal welfare. Because concerns on animal welfare and consumers' appeal for natural and health products are increasing, the functional traits have received greater importance in animal breeding programs (Egger-Danner et al., 2015). In this regard, it is required to have valid genetic parameter estimates for outstanding traits related to the farm profit, including functional traits, in the animal breeding programs (Fleming et al., 2018). Interest to include new traits in the current animal breeding programs is extending to improve simultaneously the production and reproduction performance along with animal health and well-being in dairy farms. Although, for the inclusion of a specific trait into a genetic selection program, it would be inheritable, profitable, quantifiable, and changeable (Wood et al., 2003). Although, there were some reports on the genetic analysis of persistency measures for milk components in dairy cows (Cole and Null, 2009; de Oliveira Biassus et al., 2010), to the knowledge of authors, this is the first report on the genetic analysis of fat production persistency measures in buffaloes.

In general, medium to high heritability estimates for three milk fat persistency measures in this study could be due to the reasonable additive genetic variations for these traits indicating that improvement in these traits could be attained by genetic selection. Regardless of the simpler estimation of PM_3 in contrast to other measures of persistency, the estimate of heritability for this measure was between the estimates of heritability for PM_1 and PM_2 measures for fat yield in the first lactation and had the smallest estimate in second and third parities. If a measure of persistency had higher heritability compared with

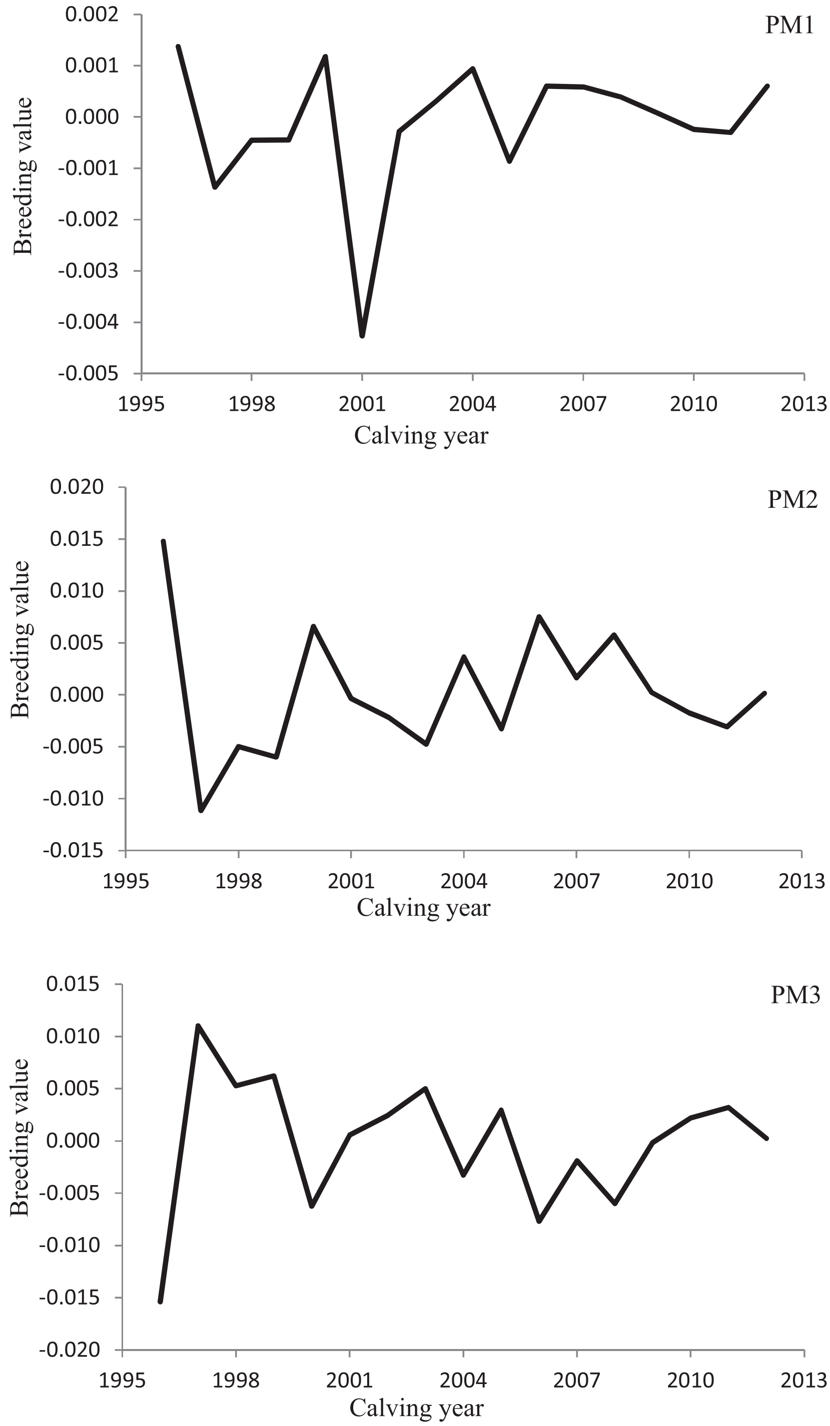

FIGURE 3 | Variation in estimated breeding values of animals for persistency measures of milk fat yield according to calving year in the first lactation.

other measures, this measure would be an appropriate measure to be considered in the selection objective (Ghavi Hossein-Zadeh et al., 2017). Respecting this explanation, the PM_2 measure would be regarded as the selection criterion in the first lactation, but the measure of PM_1 would be included as a selection objective in the second-, and third parities. Although there is no report of genetic parameters for persistency measures of fat yield in buffaloes, Cole and Null (2009) reported the estimates of heritability for fat production persistency measure varied from 0.07 to 0.12 in five breeds of dairy cows. Also, de Oliveira Biassus et al. (2010) reported the heritability estimates for fat yield persistency measures ranged from 0.00 to 0.23 in

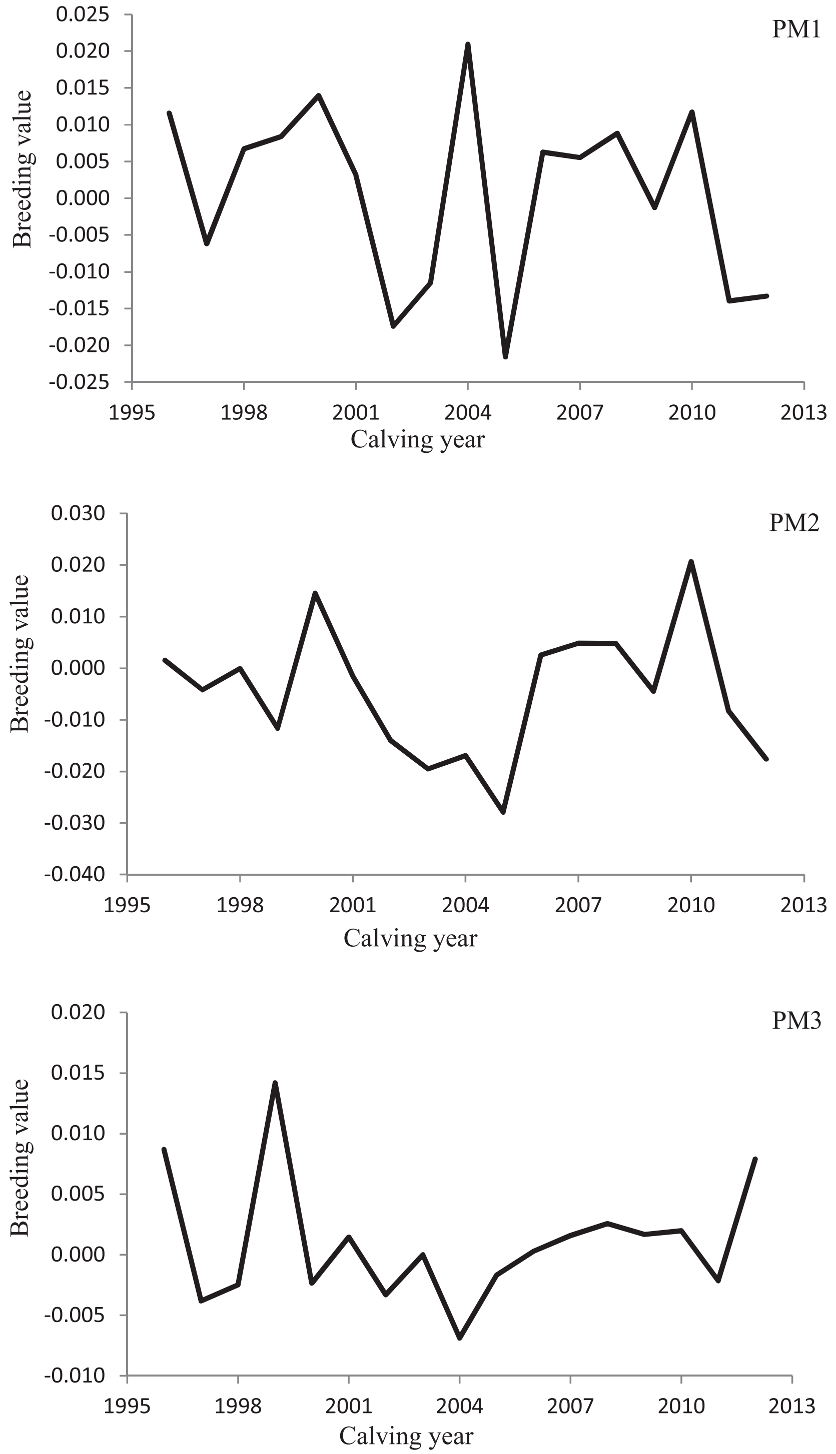

FIGURE 4 | Variation in estimated breeding values of animals for persistency measures of milk fat yield according to calving year in the second lactation.

primiparous Holstein cows. Besides, Gengler (1995) estimated the heritability for fat yield persistency measure would be equal to 0.06 in dairy cows. In general, several factors could influence the variation in heritability estimates for milk fat yield persistency obtained in different studies, including the breed of the animal, within-population genetic diversity, management procedures, environmental conditions, and methods used for estimating genetic parameters. According to de Oliveira Biassus et al. (2010), different factors would influence the variations of heritability estimates for persistency measures among studies: the

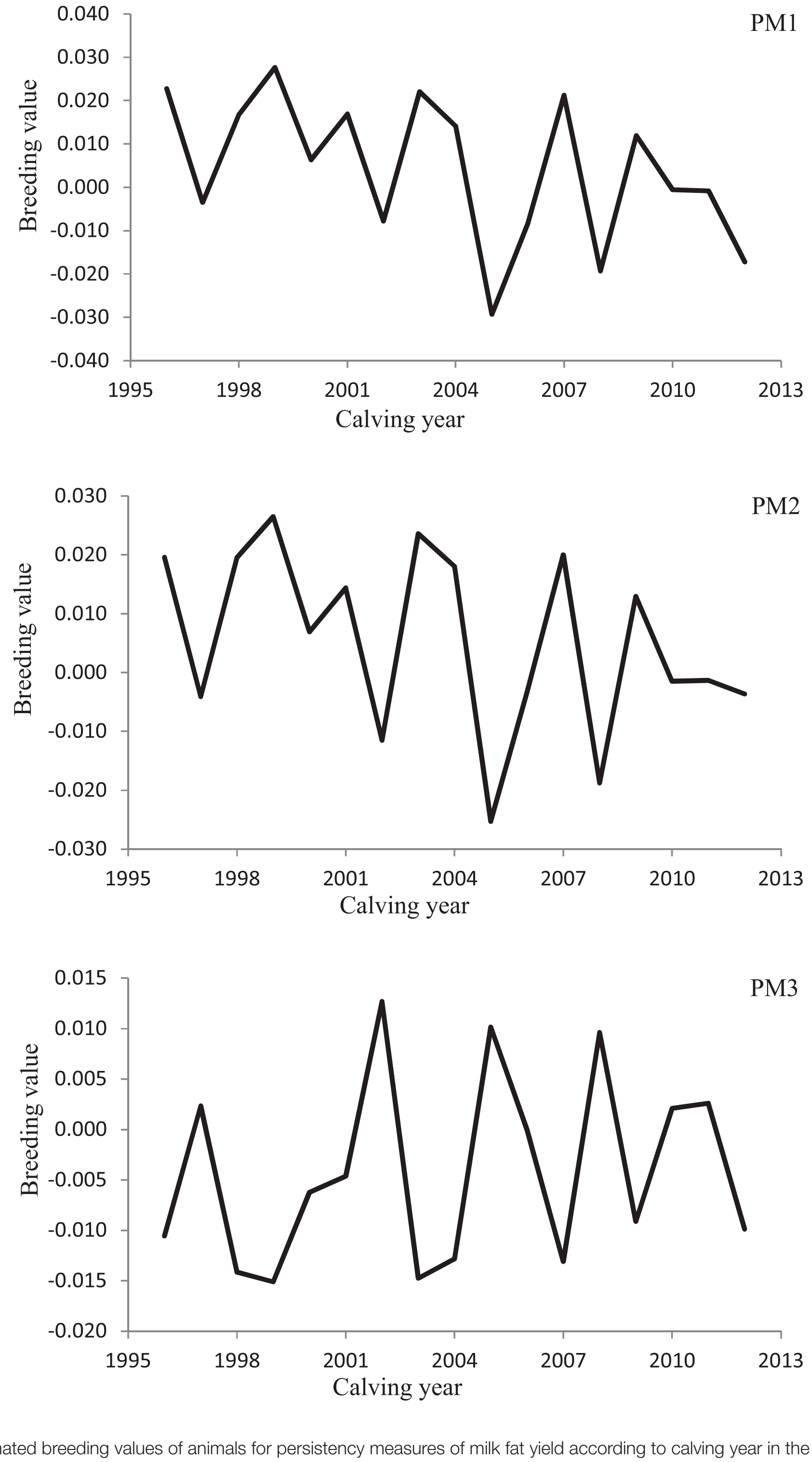

FIGURE 5 | Variation in estimated breeding values of animals for persistency measures of milk fat yield according to calving year in the third lactation.

definition of persistency measure as absolute or relative terms, the statistical adequacy of the specific measure of persistency for under study population, the lactation period used to calculate the measure of persistency, and the method or model used to calculate a specific persistency measure. Compared with the first and third lactations, less variation in heritability estimates between the three persistency measures in the second lactation would be due to the differences in lactation curves, yield persistencies, and variation of records across the first three lactations.

The production difference in two different parts of the lactation would be evaluated by the PM_1 and PM_3 persistency measures (Ghavi Hossein-Zadeh et al., 2017). Compared with PM_1 and PM_3 measures, the PM_2 measure displayed a domain below the lactation curve at a definite time that has been adjusted for yield at the end section of that period (Khorshidie et al., 2012). The procedure for defining the PM_2 and PM_3 persistency measures resulted in a high and negative genetic association between them. The positive genetic correlations between PM_1 and PM_2 are proof for the same genetic and physiological systems managing these persistency measures and would cause the same ranking of buffaloes according to these criteria in breeding and genetic schemes (Ghavi Hossein-Zadeh et al., 2017). Contrarily, high negative genetic correlations between PM_3 with two other persistency measures implied the existence of various mechanisms to govern them. In general, low genetic correlations between different persistency measures for fat yield with milk production point out that selection for a persistency measure for milk fat yield would slightly affect milk yield. In a selection program, it would be favorable to have persistency measures that had low genetic correlations with milk yield (Dekkers et al., 1998; Ghavi Hossein-Zadeh et al., 2017). According to this explanation and regarding the low genetic correlations of persistency measures for fat yield with milk production in the present study, all three measures would be considered as selection criteria that were relatively independent of production level in buffaloes. This finding indicates that a buffalo cow with the highest EBV for 270-day milk yield does not necessarily has the highest EBV for fat yield persistency and vice versa. In the other words, low estimates of genetic association between fat yield persistency measures with milk production signified that buffaloes with the identical quantity of 270-day milk yield could have a distinct extent of persistency across the lactation period (Jamrozik et al., 1998; Cobuci et al., 2007; Ghavi Hossein-Zadeh et al., 2017). The appropriateness of genetic correlation between a specific persistency measure for milk fat yield and milk production depends on the positive or negative mean of the persistency measure in the population under study (Khorshidie et al., 2012). Generally similar to the results of the present study, Cole and Null (2009) observed the estimates of genetic associations between persistency measure of fat yield with 305-day milk production varied from 0.07 to 0.29 in five breeds of dairy cows.

Predicting accurately the animals' breeding value is an appropriate way to increase the genetic gain in a specific breeding scheme (Ghavi Hossein-Zadeh, 2012). The successfulness of a selection scheme would be assessed by testing the actual alteration in breeding value indicated as a fraction of the expected theoretical modification in the average breeding value of the character under study (Jurado et al., 1994; Ghavi Hossein-Zadeh, 2012). Non-significant genetic progress estimated for all fat production persistency measures in the present study and irregular changes in average EBVs of animals over the years demonstrated the non-presence of a clear breeding design for making better the lactation persistency for fat yield in Iranian buffaloes until now. A possible reason for the non-significant genetic trends of milk fat persistency measures would be the low and close to zero estimates of genetic correlation between fat yield persistency measures and 270-day milk yield in the population under study.

CONCLUSION

The persistency measures of fat yield proposed in the present study had favorable low genetic correlations with 270-day milk production. These low correlations would be a benefit in designing a selection program to enhance the milk yield in Iranian buffaloes because buffaloes with the identical quantity of 270-day milk yield could have a distinct extent of persistency across the lactation period. The PM_2 measure had the highest heritability estimate for the first lactation buffaloes, but the PM_1 measure had the highest estimate in the second- and -third lactations. Therefore, the PM_2 measure would be regarded as the selection criterion in the first lactation, but the measure of PM_1 could be suggested as a selection objective in the second- and third parities. Based on the results of this study, it would be necessary to consider the persistency of fat yield in the selection objective of buffaloes in Iran together with main characters such as production and reproduction traits, and persistency for milk production.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The data analyzed in this study was obtained from the Animal Breeding Center and Promotion of Animal Products of Iran. Requests to access these datasets should be directed to http://abc.org.ir.

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because datasets used in this study were obtained from pre-existing databases based on routine animal recording procedures.

AUTHOR CONTRIBUTIONS

MN participated in the acquisition of data, statistical, and genetic analyses of data. NH-Z designed and conceived this study and contributed to the statistical and genetic analyses of data, and prepared the manuscript. AS contributed to the conception of the study and assisted with the interpretation of the outputs. DK assisted with the interpretation of data. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors acknowledge the Animal Breeding Center and Promotion of Animal Products of Iran for providing the data used in this study.

REFERENCES

Borghese, A. (2005). *Buffalo Production and Research.' (FAO Regional Office for Europe Inter-Regional Cooperative Research Network on Buffalo (ESCORENA: Rome, Italy)*. Available online at http://www.fao.org/3/ah847e/ah847e.pdf (accessed November 23, 2020).

Canaza-Cayo, A. W., Sávio Lopes, P., da Silva, M. V. G. B., de Almeida Torres, R., Fonseca Martins, M., Arbex, W. A., et al. (2015). Genetic parameters for milk yield and lactation persistency using random regression models in Girolando cattle. *Asian Australas. J. Anim. Sci.* 28, 1407–1418. doi: 10.5713/ajas.14.0620

Cobuci, J. A., Euclydes, R. F., Costa, C. N., Lopes, P. S., Torres, R. A., and Perreira, C. S. (2004). Analysis of persistency in the lactation of Holstein cows using test-day yield and random regression model. *Revis. Brasil Zootec.* 33, 546–554.

Cobuci, J. A., Euclydes, R. F., Costa, C. N., Torres, R. A., Lopes, P. S., and Perreira, C. S. (2007). Genetic evaluation for persistency of lactation in Holstein cows using a random regression model. *Genet. Mol. Biol.* 30, 349–355.

Cole, J. B., and Null, D. J. (2009). Genetic evaluation of lactation persistency for five breeds of dairy cattle. *J. Dairy Sci.* 92, 2248–2258. doi: 10.3168/jds.2008-1825

Cole, J. B., and VanRaden, P. M. (2006). Genetic evaluation and best prediction of lactation persistency. *J. Dairy Sci.* 89, 2722–2728. doi: 10.3168/jds.S0022-0302(06)72348-7

de Oliveira Biassus, I., Cobuci, J. A., Costa, C. N., Rorato, P. R. N., Neto, J. B., and Cardoso, L. L. (2010). Persistence in milk, fat and protein production of primiparous Holstein cows by random regression models. *R. Bras. Zootec.* 39, 2617–2624. doi: 10.1590/S1516-35982010001200009

Dekkers, J. C. M., ten Hag, J. H., and Weersink, A. (1998). Economic aspects of persistency of lactation in dairy cattle. *Livest. Prod. Sci.* 53, 237–252. doi: 10.1016/S0301-6226(97)00124-3

Egger-Danner, C., Cole, J. B., Pryce, J. E., Gengler, N., Heringstad, B., Bradley, A., et al. (2015). Invited review: overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. *Animal* 9, 191–207. doi: 10.1017/S1751731114002614

Eicher, R. (2004). Evaluation of the metabolic and nutritional situation in dairy herds: diagnostic use of milk components. *Med. Vet. du Quebec.* 34, 36–38.

Elmaghraby, M. (2012). Lactation persistency and prediction of total milk yield from monthly yields in Egyptian buffaloes. *Lucrări ?tiin?ifice* 53, 130–137.

Fleming, A., Schenkel, F. S., Malchiodi, F., Ali, R. A., Mallard, B., Sargolzaei, M., et al. (2018). Genetic correlations of mid-infrared-predicted milk fatty acid groups with milk production traits. *J. Dairy Sci.* 101, 4295–4306. doi: 10.3168/jds.2017-14089

Gengler, N. (1995). Multiple-trait genetic evaluations for milk, fat, and protein yields and persistency. *Interbull Bull.* 11, 1–6.

Gengler, N. (1996). Persistency of lactation yields: a review. *Interbull Bull.* 12, 87–96.

Ghavi Hossein-Zadeh, N. (2012). Bayesian estimates of genetic changes for body weight traits of Moghani sheep using Gibbs sampling. *Trop. Anim. Health Prod.* 44, 531–536. doi: 10.1007/s11250-011-9930-1

Ghavi Hossein-Zadeh, N. (2015b). Bayesian analysis of direct and maternal effects for birthweight in Iranian buffaloes using Gibbs sampling. *Anim. Prod. Sci.* 56, 859–865. doi: 10.1071/AN14564

Ghavi Hossein-Zadeh, N. (2015a). Analysis of population structure and genetic variability in Iranian buffaloes (*Bubalus bubalis*) using pedigree information. *Anim. Prod. Sci.* 56, 1130–1135. doi: 10.1071/AN14738

Ghavi Hossein-Zadeh, N. (2016). Modelling lactation curve for milk fat to protein ratio in Iranian buffaloes (*Bubalus bubalis*) using non-linear mixed models. *J. Dairy Res.* 83, 334–340. doi: 10.1017/S0022029916000340

Ghavi Hossein-Zadeh, N., Madad, M., Shadparvar, A. A., and Kianzad, D. (2012). An observational analysis of secondary sex ratio, stillbirth and birth weight in Iranian buffaloes (*Bubalus bubalis*). *J. Agric. Sci. Technol.* 14, 1477–1484.

Ghavi Hossein-Zadeh, N., Nazari, M. A., and Shadparvar, A. A. (2017). Genetic perspective of milk yield persistency in the first three lactations of Iranian buffaloes (*Bubalus bubalis*). *J. Dairy Res.* 84, 434–439. doi: 10.1017/S0022029917000498

Jakobsen, J. H., Madsen, P., Jensen, J., Pedersen, J., Christensen, L. G., and Sorensen, D. A. (2002). Genetic parameters for milk production and persistency for Danish Holsteins estimated in random regression models using REML. *J. Dairy Sci.* 85, 1607–1616. doi: 10.3168/jds.S0022-0302(02)74231-8

Jamrozik, J., Jansen, G., Schaeffer, L. R., and Liu, Z. (1998). Analysis of persistency of lactation calculated from a random regression test day model. *Interbull Bull.* 17, 64–69.

Jurado, J. J., Alonso, A., and Alenda, R. (1994). Selection response for growth in Spanish Merino flock. *J. Anim. Sci.* 72, 1433–1440. doi: 10.2527/1994.7261433x

Khorshidie, R., Shadparvar, A. A., Ghavi Hossein-Zadeh, N., and Joezy Shakalgurabi, S. (2012). Genetic trends for 305-day milk yield and persistency in Iranian Holsteins. *Livest. Sci.* 144, 211–217. doi: 10.1016/j.livsci.2011.11.016

Kirkpatrick, M., Lofsvold, D., and Bulmer, M. (1990). Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* 124, 979–993. doi: 10.1093/genetics/124.4.979

Kistemaker, G. J. (2003). Comparison of persistency definitions in random regression test day models. *Interbull Bull.* 30, 96–98.

Li, J., Gao, H., Madsen, P., Li, R., Liu, W., Bao, P., et al. (2020). Impact of the order of Legendre polynomials in random regression model on genetic evaluation for milk yield in dairy cattle population. *Front. Genet.* 11:586155. doi: 10.3389/fgene.2020.586155

Meyer, K. (2006). *WOMBAT – A Program for Mixed Model Analyses by Restricted Maximum Likelihood. User Notes*. Armidale: Animal Genetics and Breeding Unit, University of New England.

Mrode, R. A., Swanson, G. J. T., and Paget, M. F. (2003). "Implementation of a test day model for production traits in the UK," in *Proceedings of the Interbull Meeting*, (Cambridge: Cambridge University Press), 193–196.

Muir, B. L. (2004). *Genetics of Lactation Persistency and Relationships with Reproductive Performance in Holsteins*. Ph.D. Dissertation, University of Guelph, Guelph, ON.

Oliveira, H. R., Brito, L. F., Lourenco, D. A. L., Silva, F. F., Jamrozik, J., Schaeffer, L. R., et al. (2019). Invited review: advances and applications of random regression models: from quantitative genetics to genomics. *J. Dairy Sci.* 102, 7664–7683. doi: 10.3168/jds.2019-16265

Safari, A., Ghavi Hossein-Zadeh, N., Shadparvar, A. A., and Abdollahi Arpanahi, R. (2018). A review on breeding and genetic strategies in Iranian buffaloes (*Bubalus bubalis*). *Trop. Anim. Health Prod.* 50, 707–714. doi: 10.1007/s11250-018-1563-1

Sölkner, J., and Fuchs, W. (1987). A comparison of different measures of persistency with special respect to variation of test-day milk yields. *Livest. Prod. Sci.* 16, 305–319. doi: 10.1016/0301-6226(87)90001-7

Swalve, H. H., and Gengler, N. (1999). Genetics of lactation persistency. *Occ. Publ. Br. Soc. Anim. Sci.* 24, 75–82. doi: 10.1017/S1463981500043090

Togashi, K., and Lin, C. Y. (2004). Efficiency of different selection criteria for persistency and lactation milk yield. *J. Dairy Sci.* 87, 1528–1535. doi: 10.3168/jds.S0022-0302(04)73304-4

White, I. M. S., Thompson, R., and Brotherstone, S. (1999). Genetic and environmental smoothing of lactation curves with cubic splines. *J. Dairy Sci.* 82, 632–638. doi: 10.3168/jds.S0022-0302(99)75277-X

Wilmink, J. B. M. (1987). Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. *Livest. Prod. Sci.* 16, 335–348. doi: 10.1016/0301-6226(87)90003-0

Wood, G. M., Boettcher, P. J., Jamrozik, J., Jansen, G. B., and Kelton, D. F. (2003). Estimation of genetic parameters for concentrations of milk urea nitrogen. *J. Dairy Sci.* 86, 2462–2469. doi: 10.3168/jds.S0022-0302(03)73840-5

Wood, P. (1967). Algebraic model of the lactation curve in cattle. *Nature* 216:164. doi: 10.1038/216164a0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

MINI REVIEW
published: 22 March 2021
doi: 10.3389/fgene.2021.629861

Opportunities and Challenges for Improving the Productivity of Swamp Buffaloes in Southeastern Asia

Paulene S. Pineda[1], Ester B. Flores[1], Jesus Rommel V. Herrera[2] and Wai Yee Low[3]*

[1]Philippine Carabao Center National Headquarters and Genepool, Science City of Muñoz, Philippines, [2]Philippine Carabao Center at University of the Philippines – Los Baños, Laguna, Philippines, [3]The Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Adelaide, SA, Australia

The swamp buffalo is a domesticated animal commonly found in Southeast Asia. It is a highly valued agricultural animal for smallholders, but the production of this species has unfortunately declined in recent decades due to rising farm mechanization. While swamp buffalo still plays a role in farmland cultivation, this species' purposes has shifted from draft power to meat, milk, and hide production. The current status of swamp buffaloes in Southeast Asia is still understudied compared to its counterparts such as the riverine buffaloes and cattle. This review discusses the background of swamp buffalo, with an emphasis on recent work on this species in Southeast Asia, and associated genetics and genomics work such as cytogenetic studies, phylogeny, domestication and migration, genetic sequences and resources. Recent challenges to realize the potential of this species in the agriculture industry are also discussed. Limited genetic resource for swamp buffalo has called for more genomics work to be done on this species including decoding its genome. As the economy progresses and farm mechanization increases, research and development for swamp buffaloes are focused on enhancing its productivity through understanding the genetics of agriculturally important traits. The use of genomic markers is a powerful tool to efficiently utilize the potential of this animal for food security and animal conservation. Understanding its genetics and retaining and maximizing its adaptability to harsher environments are a strategic move for food security in poorer nations in Southeast Asia in the face of climate change.

Keywords: swamp buffalo, genomics, genetic improvement, genetic diversity, Southeast Asia agriculture

OPEN ACCESS

Edited by:
Jiuzhou Song,
University of Maryland, College Park, United States

Reviewed by:
Filippo Biscarini,
National Research Council (CNR), Italy
Lingyang Xu,
Chinese Academy of Agricultural Sciences, China

Correspondence:
Wai Yee Low
wai.low@adelaide.edu.au

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *16 November 2020*
Accepted: *26 February 2021*
Published: *22 March 2021*

Citation:
Pineda PS, Flores EB, Herrera JRV and Low WY (2021) Opportunities and Challenges for Improving the Productivity of Swamp Buffaloes in Southeastern Asia. Front. Genet. 12:629861. doi: 10.3389/fgene.2021.629861

INTRODUCTION

The majority (~97%) of the 207 million buffalo population in the world is found in Asia, wherein about 20.51% are swamp buffaloes (FAOSTAT, 2018). There are two types of water buffaloes: swamp buffaloes and river buffaloes. Swamp buffaloes are mainly found in China and Southeast Asian countries. River buffaloes' populations are larger than swamp buffaloes' populations. They differ in chromosome number, phenotypic characteristics, and geographical locations, where they are usually found (Degrandi et al., 2014; Colli et al., 2018; Zhang et al., 2020).

Swamp buffaloes in Southeast Asia are raised by smallhold farmers because of their powerful draft capacity (OECD, 2017). This animal is utilized mostly for land cultivation; though it also provides milk, meat, hide, and horn, which are additional income sources to the farmers. However, due to increased farm mechanization, swamp buffalo have declined in value and its production has decreased by 4.92% for the last two decades (FAOSTAT, 2018). While swamp buffalo still holds a significant role in farmland cultivation, the purpose of this animal has shifted from draft power to meat and milk production.

One way to address the decline in production of swamp buffalo is to use genomic markers to selectively breed this animal for food security and conservation. Many countries in Southeast Asia have only started their breeding programs for swamp buffaloes in recent decades. Genetic improvement for buffalo in Thailand started in 1979 through their Department of Livestock Development. Genetic evaluation procedures, such as using estimated breeding values (EBVs), were conducted as part of their selection criteria for superior swamp buffaloes (Sanghuayphrai et al., 2013). Although genetic evaluation procedures are used in Thailand, breeding improvement and disease prevention are still lacking in some buffalo herds, leading to its low productivity, and hence highlight the need for upgraded buffalo management (Koobkaew et al., 2013; Sapapanan et al., 2013; Suphachavalit et al., 2013).

In the Philippines, a centralized research agency – Philippine Carabao Center (PCC) was established in 1992 to strengthen research and development on the Philippine carabaos. The PCC has several programs, such as the nationwide dispersal of semen for artificial insemination and bull loan programs, to upgrade buffaloes (Cruz, 2015). Cross breeding of the two types of water buffalo was carried out to improve the efficiency of the animal as their progeny showed increased body weight and milk production when compared to local swamp buffaloes. However, the crossbred progeny showed a decline in reproductivity, and hence backcrossing with a purebred swamp- or river-type was done to produce a ¾ Philippine swamp-type for draft power or ¾ river-type for dairy, respectively (Salas et al., 2000; Cruz, 2015). Genetic evaluation has also been done to select elite animals to improve milk traits in the Philippine dairy buffaloes (Herrera et al., 2018).

While there is no centralized agency exclusively for the development of water buffaloes in Malaysia, Indonesia, and Vietnam, regional efforts have been carried out to increase the performance of buffaloes in terms of reproductive performance, weight gain, and meat and milk production (Suryanto et al., 2002; Othman, 2014; Ariff et al., 2015). Buffalo management in Indonesia still follows the traditional approach leading to low productivity of the animal due to poor breeding plans, which has led to inbreeding within the population (Komariah et al., 2020). Despite breeding inefficiency, buffalo rearing by smallhold farmers is expected to contribute to the development of dairy industry in Indonesia. Vietnam produced and consumed more buffalo meat than beef; however, limited resources for research have stumped its intensified breeding program and buffalo development (Nguyen, 2000).

CYTOGENETICS, PHYLOGENY, DOMESTICATION, AND MIGRATION

River and swamp buffaloes have 50 and 48 chromosomes, respectively. Although their chromosome numbers are dissimilar, these two sub-species can produce fertile offspring when crossed, which inherits 49 chromosomes due to the preserved characteristics of its chromosome arms (Degrandi et al., 2014). However, reproductivity is decreased in the hybrid progeny (Harisah et al., 1989; Borghese, 2011). This difference in chromosome number between the swamp and river buffalo is due to a tandem fusion translocation between river buffalo chromosomes 4 and 9 and swamp buffalo chromosome 1 (Di Berardino and Iannuzzi, 1981; Harisah et al., 1989), which was later confirmed when swamp buffalo genome assembly was made available (Luo et al., 2020). Studies on the karyotypes of swamp buffaloes that originated from the Philippines, Thailand, Malaysia, and Brazil showed conflicting results on the centromeres' positions but they all agreed that the species has 48 chromosomes (Bondoc et al., 2002; Supanuam et al., 2012; Degrandi et al., 2014; Shaari et al., 2019). There are at least two possible reasons that account for differences in the centromeres' positions: (1) different methods were used in the cytogenetic study (e.g., an addition of alcohol might have affected the arrangement of the chromosomes) and (2) subjective determination of each chromosomes' centromere locations. Further investigation using a standardized method is needed to confirm the typical karyotype of swamp buffaloes.

Both river- and swamp-type have the same ancestral origin from wild Asiatic buffalo, *Bubalus arnee* (Cockrill, 1981). There is genetic separation for the two types of water buffaloes (**Figure 1**) and divergence between them is higher than the divergence observed between cattle subspecies (Yindee et al., 2010). Interestingly, comparison between river- and swamp-type buffaloes showed higher genetic variation within swamp populations despite the homogenous characteristics of their phenotypes and small number of breeds (Zhang et al., 2016; Paraguas et al., 2018; Sun et al., 2020b).

Divergence of the water buffalo to river- and swamp-type is estimated to have happened from 10 Kya to 1.7 Mya with the most probable period being from around 230 Kya or 900–860 Kya based on overlapping events such as geographical changes and concurrences from multiple studies (Tanaka et al., 1996; Wang et al., 2017; Sun et al., 2020a).

Swamp buffalo during post-domestication period followed two separate migration events from about 3,000 to 6,000 years ago in Asia (Wang et al., 2017). One was from Indochina border spreading around mainland China to the Philippines and the other was from mainland Southeast Asia and Southwest China border disseminating down to Indonesia (Zhang et al., 2016; Wang et al., 2017; Colli et al., 2018; Sun et al., 2020b). There is a genetically distinct population of swamp buffaloes in Southeast Asia that is thought to have arisen from the founder effect (Zhang et al., 2016; Colli et al., 2018; Sun et al., 2020b). A rare haplogroup was found in Thailand by Sun et al., 2020b using mtDNA D-loop sequences, which supported the hypothesis that Thai buffalo population may have come from an ancestral lineage (Colli et al., 2018). Considering that the wild Asiatic buffalo still exists in some parts of Thailand (Sarataphan et al., 2017), the ancestor of water buffalo may have also originated in mainland Southeast Asia (Lau et al., 1998).

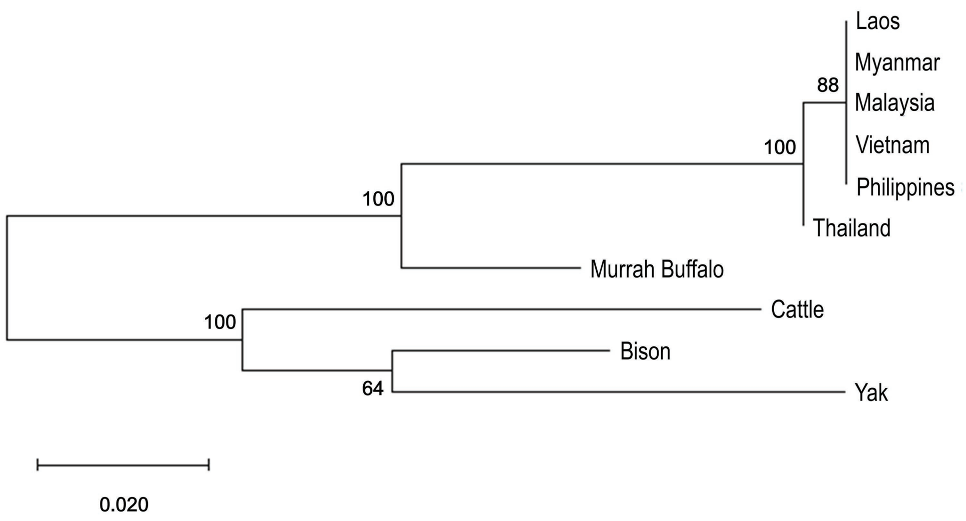

FIGURE 1 | Phylogenetic analysis of mtDNA partial D-loop of swamp buffalo, Murrah buffalo, and three outgroup species was inferred by using a Maximum Likelihood method and a Tamura 3-parameter model in MEGA-X (Tamura, 1992; Kumar et al., 2018). Sequences were downloaded from the GenBank with the following accession numbers: Laos swamp buffalo (PopSet: 1174238592, KR008969-KR009068), Myanmar swamp buffalo (PopSet: 1174238592), Malaysia swamp buffalo (PopSet: 1605320276), Vietnam swamp buffalo (PopSet: 1174238592, 966874160), Philippines swamp buffalo (FJ873676-FJ873683), Thailand swamp buffalo (PopSet: 1174238592, KR008886-KR008939), Murrah buffalo – river-type buffalo (NC_049568), Cattle (NC_006853), American bison (NC_012346), and Yak (NC_006380). Initial trees were obtained by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. One thousand bootstraps were done and their percentage values are displayed in the nodes.

GENETIC SEQUENCE AND RESOURCE AVAILABILITY

The whole genome sequence of a Mediterranean breed (UMD_CASPUR_WB_2.0) river buffalo was released in the NCBI in 2013 and published 4 years later (Williams et al., 2017; **Table 1**). A 90K SNP Buffalo Genotyping Array (Iamartino et al., 2013, 2017) has been available for use by researchers in the past few years; however, the SNP panel was created using a cattle reference genome (UMD3.1). The disadvantage of using the SNP panel for water buffalo is that it only represents 75% and 24.5% of the high quality, known polymorphic SNPs of river- and swamp-type buffaloes, respectively. The majority of the samples used in the SNP validation belonged to river buffalo, and hence a specific SNP panel for the swamp buffaloes is recommended since it is underrepresented in the 90K SNP Panel (Iamartino et al., 2013, 2017; Colli et al., 2018). Despite the limitation of missing some water buffalo specific SNPs, the genotyping array is still useful for genomic studies in river buffaloes but its usefulness remains limited in swamp buffalo (Herrera et al., 2016).

The river buffalo assembly based on the same animal used to create UMD_CASPUR_WB_2.0 was recently upgraded using long read sequencing for contig assembly and chromatin conformation capture technologies for scaffolding. The final assembly is called as UOA_WB_1 (Low et al., 2019) and is the best representative assembly of the river buffalo based on contiguity metric such as contig N50 (**Table 1**). The next assembly upgrade for the river buffalo will be a completely haplotype-resolved genome as demonstrated in cattle (Low et al., 2020). There are eight river buffalo assemblies but only one swamp genome assembly (Luo et al., 2020) in the literature and public databases. Besides genome assemblies and SNP panel, there are transcriptome resources that were used to create a large-scale gene expression atlas for the river buffalo and 3 million intestinal microbial gene catalogs from both buffalo and cattle (Williams et al., 2017; Zhang et al., 2017; Young et al., 2019).

COMPARISONS BETWEEN RIVER AND SWAMP BUFFALOES

The latest river buffalo reference assembly (UOA_WB_1) is approximately 2.5 times more contiguous than the best swamp buffalo assembly (GWHAAJZ00000000) based on contig N50. Both of these assemblies benefited from long read PacBio sequencing to preserve assembly continuity and scaffolding with Hi-C reads has helped to produce chromosome-scale scaffolds. However, despite the availability of an impressive genome assembly, only about 0.76% of the submitted water buffalo nucleotide sequences were from swamp buffaloes in the GenBank as of January 2021. The river buffalo sequences represented the majority of water buffalo sequences in the public database. Additionally, there were only 17 genes for swamp-type, if one excluded the annotation from the recent swamp genome (Luo et al., 2020), which was a few magnitudes lower than the ~35,000 genes submitted for river-type buffaloes.[1]

[1]https://www.ncbi.nlm.nih.gov

TABLE 1 | Genome assemblies and resources available for water buffalo.

Assembly name	Genome size (Gb)	Contig N50 (kb)	Scaffold N50 (Mb)	Breed/origin	Type	Resources available for river and/or swamp buffaloes	References
UOA_WB_1	2.66	22441.5	117.2	Mediterranean	River	90K SNP Panel for buffaloes (Iamartino et al., 2013) – river and swamp buffalo	https://www.nature.com/articles/s41467-018-08260-0#citeas
Murrah_sire	2.62	9500.0	82.0	Murrah	River	Gene expression atlas (Young et al., 2019) – river buffalo	https://www.biorxiv.org/content/10.1101/618785v2.full
Murrah_dam	2.62	5230.0	83.2	Murrah	River	Transcriptome resource (Williams et al., 2017) – river buffalo	https://www.biorxiv.org/content/10.1101/618785v2.full
GWHAAKA00000000	2.65	3100.0	116.1	Murrah	River	Intestinal microbial gene catalog (Zhang et al., 2017) – cattle and river buffalo	https://academic.oup.com/nsr/article/7/3/686/5737567
Bubbub1.0	2.77	25.0	7.0	Bangladesh	River	Breeding programs Italy: Mediterranean breed – genetic improvement with genetic evaluation (http://www.anasb.it/) – river buffalo	https://onlinelibrary.wiley.com/doi/full/10.1002/ece3.4965
UMD_CASPUR_ WB_2.0	2.84	21.9	1.4	Mediteranean	River	Brazil: Genetic improvement program (Bernardes, 2007) – river buffalo	https://academic.oup.com/gigascience/article/6/10/gix088/4101552
ASM299383v1	3.00	14.6	3.6	Egypt	River	India: Genetic improvement and dispersal of semen from different breeds (Sahu et al., 2019) – river buffalo	Unpublished
Bubalus_bubalis_ Jaffrabadi_v3.0	3.80	14.0	0.1	Jafarabadi	River	Pakistan: Genetic improvement of buffalo in Pakistan (GIBP; http://parc.gov.pk/index.php/en/faqy/131-narc/animal-sciences-institute/610-asi-breeding-genetics) – river buffalo Bangladesh: Buffalo Development Project (Hamid et al., 2017) – river buffalo	Unpublished
GWHAAJZ00000000	2.63	8800.0	117.3	Fuzhong	Swamp	Philippines: Genetic Improvement Program – upgrading and crossbreeding of river and swamp buffaloes (national dispersion of semen; https://www.pcc.gov.ph/genetic-improvement/) – river and swamp buffalo China: Upgrading and crossbreeding of river and swamp buffaloes (regional dispersion of semen; Yang et al., 2013) – river and swamp buffalo Thailand: BREEDPLAN program (analysis system developed in Australia; https://breedplan.une.edu.au/) – swamp buffalo	https://academic.oup.com/nsr/article/7/3/686/5737567

The scientific name for river buffalo is Bubalus bubalis and swamp buffalo is Bubalus carabanesis.

Genomic regions that may be under selection have been analyzed in both swamp and river buffaloes. Interestingly, swamp buffaloes showed the signs of selection in docile behavior, muscle development, and fatigue resistance (Luo et al., 2020; Sun et al., 2020a). Among the genes under selection, *HDAC9* was found to be associated with muscle development in other species (Mei et al., 2019; Sun et al., 2020a). Luo et al. (2020) study on swamp buffalo genome also showed the expansion of *AMD1* gene that promotes muscle growth. This suggests the possibility of prospecting swamp buffaloes as a meat resource. Two critical starch digestion-enzyme genes, *AMY2B* and *SI*, were also identified that makes this species unique from other ruminants, which may suggest a new mechanism for adapting to rumen acidosis (Luo et al., 2020).

Signature of selection in river buffaloes showed over-representation in genes associated with immune-response, milk production, growth, and feed efficiency, which can be due to selection for milk production (Luo et al., 2020; Sun et al., 2020a). From the genes identified, *thyroglobulin* gene was associated with milk and meat quality traits, and was found to be a good candidate gene marker for meat marbling and milk fat percentage (Gan et al., 2008; Dubey et al., 2015).

Genetic variations in *DGAT1*, *MUC1*, *INSIG2*, and *GHR* in both river and swamp buffaloes were also associated with milk components, milk yield, and mastitis resistance, which are potential candidates for genetic selection (Deng et al., 2016; Li et al., 2018; da Rosa et al., 2020; El-Komy et al., 2020).

CHALLENGES AND OPPORTUNITIES

While Southeast Asian countries are experiencing improvements in agricultural productivity, it still remains relatively small (OECD, 2017). Considering the limited number of available genetic sequences and studies of swamp buffalo, it can be said that research funding allocation for this animal is low when compared to other bovine species. Countries from Southeast Asia should take a more progressive approach in studying the animal through genome science. Given the limited budget for research and development, this may be challenging as the costs for genomic research is high. Nevertheless, the trend of smaller farm sizes, increases in population and the effect of climate change, as well as agricultural innovations and developments, will likely push swamp buffalo farming toward intensified, profitable, and efficient farming (OECD, 2017).

Incorporation of genomic selection in genetic improvement programs has proven its success in dairy cattle and other livestock species, but which usually carried out in large-scale breeding programs and with intensive breeding selection (Sonstegard et al., 2001; Miller, 2010; Dekkers, 2012; Xu et al., 2020). On the contrary, local breeds are usually farmed in smaller population size and remain inferior in terms of productivity. Although the incorporation of genome science will maximize genetic gains of the animals, and hence an increase in productivity and income, the costs are relatively higher on a per animal basis (Iamartino et al., 2013; Biscarini et al., 2015). Despite the opportunities in breeding swamp buffaloes, economic constraints in smallhold farming remain a challenge for large scale and cost-effective genetic improvement programs (Biscarini et al., 2015; El Debaky et al., 2019). Nonetheless, the improvement of breeding stock through EBVs and proper management has shown significant increase in milk production in the Philippines, which demonstrated the value of systematic breeding programs for dairy buffalo (Flores et al., 2007). Rural farmers have seen buffalo rearing as a less risky source of income when compared to recurrent crop failures due to calamities such as typhoons and droughts (Escarcha et al., 2020). For example, through the support from government and organized groups, buffalo rearing holds the promise to enable sustainable living in smallhold farmers in the Philippines (Del Rosario and Vargas, 2013).

Genome editing (GE) technologies use zinc-finger nucleases, transcription activator-like effector nucleases and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 to reproduce animals with economically important traits (Lee et al., 2020). It has been used in livestock species to produce polled (i.e., hornless) cattle (Young et al., 2020), mastitis resistant cows through insertion of *lysozyme* gene (Liu et al., 2014) and enhanced wool quality in goats and sheep by altering their *FGF5* gene (Hu et al., 2017; Li et al., 2017, 2019). The GE system has also been used to edit the swamp buffalo *GDF8* gene in cell line, which is a regulatory gene for myostatin that inhibits muscle development and differentiation (Su et al., 2018; Lee et al., 2020). Gene knockout of *GDF8* can increase the production of meat in cattle, goat, and sheep as double muscling was observed in experimental animals (Proudfoot et al., 2015; He et al., 2018; Wu et al., 2018; Ding et al., 2020). Examples of GE in water buffalo are limited but the opportunity to use this technology to enhance their economic traits remains to be explored. The applications of GE in livestock need to adhere to ethical standards and regulatory policies (McFarlane et al., 2019) that vary between countries. For example, the hornless cattle created using GE tools by the company Recombinetics was meant to proceed further in Brazil, but the plan was abandoned when unintended integration of plasmid was found in edited animals (Molteni, 2019; Norris et al., 2020). AquAdvantage salmon and GalSafe pigs are the only approved genetically modified animals for food specifically in United States and Canada (FDA, 2020).[2] In Asia-Pacific region, it is unclear if livestock made using GE technologies will be acceptable in the near future (FAO, 2019).

Precision livestock farming (PLF) incorporates artificial intelligence technology to automatically monitor and manage animal production, predicts solutions for problems that may arise in the farm, and uses deep learning for genomic prediction (Banhazi et al., 2012; Pérez-Enciso and Zingaretti, 2019; Tullo et al., 2019). PLF assists large farms to be economically and environmentally sustainable; however, the cost of PLF still outweighs its efficiency for smallhold farmers (Hostiou et al., 2017; Carillo and Abeni, 2020). Genomic prediction using deep learning requires large datasets that are currently unavailable for the swamp buffalo. While PLF should be embraced in Southeast Asia, the limitation of high cost

[2]https://aquabounty.com

means its application to swamp buffalo farming remains infeasible in the near future.

Microbiome analysis for swamp buffaloes showed intrinsic difference to cattle microbiota that might explain buffalo's efficiency in digesting fibers (Zhang et al., 2017; Iqbal et al., 2018). Rumen manipulation to reduce methane emission is also of interest in livestock management as it decreases the environmental impact of livestock production (Ungerfeld, 2018). In large-scale farmed populations, besides rumen related measurements, there are other low-cost proxies such as body weights and high-throughput milk mid-infrared that are also suitable to monitor methane emission (Negussie et al., 2017). Management and genetic improvement of swamp buffalo based on combination of these proxies may lead to production animals with less negative environmental footprint (Negussie et al., 2017; Ungerfeld, 2018).

With the increasing demand for food and mechanization in farming, swamp buffalo should be bred for meat and milk production through wide-scale or institutionalized development programs (Palacpac, 2010; Cruz, 2013). Buffaloes are well suited for tropical climate of Southeast Asia, and thus there is potential in upgrading local buffaloes to maximize milk production, which cannot be easily done with species maladapted to hotter and humid climates. Although swamp buffaloes are still susceptible to heat stress (Upadhyay et al., 2007; Rojas-Downing et al., 2017), their wallowing behavior and adaptability to warm conditions give them an advantage for hotter climate (Nardone et al., 2010).

CONCLUSION

The potential of swamp buffaloes in food production is still untapped and genome research to increase its production is still limited. Understanding the capabilities of this species through a genomic approach can increase its productivity and benefit the farmers in the long run. The availability of high-quality swamp buffalo assembly is a leap forward in swamp buffalo genome science, because it opens up opportunities for technological advancement such as the creation of SNP panels specific to swamp buffalo for genetic improvement, diagnosis of diseases, and the study of genetic diversity. Although the cost of genomics is expensive and remains a challenge for developing countries in Southeast Asia, the opportunities to improve this animal for milk and meat production and animal conservation remain to be explored. With the rapid progress of technology and changing climates, rearing swamp buffaloes is a strategic option to increase smallhold farmers' income. Breeding the animals through genomic selection is a good strategy to select meat and milk type swamp buffaloes while retaining its adaption to hotter, humid climates.

AUTHOR CONTRIBUTIONS

All authors contributed to the conception of the study, manuscript revision, read, and approved the submitted version. PP wrote the first draft of the manuscript.

FUNDING

Publishing fee for this article review is funded by the Research and Development Division, Philippine Carabao Center.

ACKNOWLEDGMENTS

We would like to thank the Research and Development Division, Philippine Carabao Center (PCC) for their support to this article review, the PCC-Publication and Presentation Review Committee, the PCC-Animal Genetic Resource Section, and the PCC-Animal Breeding and Genomics Section for contributing their idea to the paper.

REFERENCES

Ariff, O. M., Sharifah, N. Y., and Hafidz, A. W. (2015). Status of beef industry of Malaysia. *Mal. J. Anim. Sci.* 18, 1–21.

Banhazi, T. M., Lehr, H., Black, J. L., Crabtree, H., Schofield, P., Tscharke, M., et al. (2012). Precision livestock farming: an international review of scientific and commercial aspects. *Int. J. Agric. Biol. Eng.* 5:1. doi: 10.3965/j.ijabe.20120503.001

Bernardes, O. (2007). Buffaloes breeding in Brasil. *Ital. J. Anim. Sci.* 6(Suppl. 2), 162–167. doi: 10.4081/ijas.2007.s2.162

Biscarini, F., Nicolazzi, E., Alessandra, S., Boettcher, P., and Gandini, G. (2015). Challenges and opportunities in genetic improvement of local livestock breeds. *Front. Genet.* 6:33. doi: 10.3389/fgene.2015.00033

Bondoc, O. L., Flor, M. C. G. T., Rebollos, S. D. N., and Albarace, A. G. (2002). Variations in karyotypic characteristics of different breed groups of water buffaloes (*Bubalus bubalis*). *Asian Australas. J. Anim. Sci.* 15, 321–325. doi: 10.5713/ajas.2002.321

Borghese, A. (2011). Situation and perspectives of buffalo in the World, Europe and Macedonia. *Maced. J. Anim. Sci.* 1, 281–296.

Carillo, F., and Abeni, F. (2020). An estimate of the effects from precision livestock farming on a productivity index at farm level. Some evidences from a dairy farms' sample of lombardy. *Animals* 10, 1–11. doi: 10.3390/ani10101781

Cockrill, W. R. (1981). The water buffalo: a review. *Br. Vet. J.* 137, 8–16. doi: 10.1016/S0007-1935(17)31782-7

Colli, L., Milanesi, M., Vajana, E., Iamartino, D., Bomba, L., Puglisi, F., et al. (2018). New insights on water buffalo genomic diversity and post-domestication migration routes from medium density SNP chip data. *Front. Genet.* 9:53. doi: 10.3389/fgene.2018.00053

Cruz, L. (2013). Changing faces of swamp buffaloes in an industrializing Asia. *Buffalo Bull.* 32, 32–49.

Cruz, L. (2015). "Institutionalization of Swamp Buffalo Development in The Philippines" in *Proceeding of International Seminar "Improving Tropical Animal Production for Food Security"*; November 3-5, 2015; 15–37.

da Rosa, F. T., Moreira, C. G. A., Barbero, M. M. D., Hurtado Lugo, N. A., de Camargo, G. M. F., Aspicueta Borquis, R. R., et al. (2020). Associations between MUC1 gene polymorphism and resistance to mastitis, milk production and fertility traits in Murrah water buffaloes. *J. Appl. Anim. Res.* 48, 151–155. doi: 10.1080/09712119.2020.1749641

Degrandi, T., Marques, J., Gunski, R., Costa, M., Marques, L., Figueiró, M., et al. (2014). Cytogenetic identification of four generations of crossbred buffaloes maintained in a conservation program in the Marajó island/Brazil. *J. Biotech. Biodivers.* 5, 162–171. doi: 10.20873/jbb.uft.cemaf.v5n2.degrandi

Dekkers, J. C. M. (2012). Application of genomics tools to animal breeding. *Curr. Genomics* 13, 207–212. doi: 10.2174/138920212800543057

Del Rosario, W., and Vargas, D. (2013). Sustainability of Philippine Carabao Center and Primary Cooperative Partnership in Carabao-Based Enterprise. *Buffalo Bull.* 32, 1226–1229.

Deng, T., Pang, C., Ma, X., Lu, X., Duan, A., Zhu, P., et al. (2016). Four novel polymorphisms of buffalo INSIG2 gene are associated with milk production traits in Chinese buffaloes. *Mol. Cell. Probes* 30, 294–299. doi: 10.1016/j.mcp.2016.09.003

Di Berardino, D., and Iannuzzi, L. (1981). Chromosome banding homologies in swamp and murrah buffalo. *J. Hered.* 72, 183–188. doi: 10.1093/oxfordjournals.jhered.a109469

Ding, Y., Zhou, S. W., Ding, Q., Cai, B., Zhao, X. -e., Zhong, S., et al. (2020). The CRISPR/Cas9 induces large genomic fragment deletions of MSTN and phenotypic changes in sheep. *J. Integr. Agric.* 19, 1065–1073. doi: 10.1016/S2095-3119(19)62853-4

Dubey, P. K., Goyal, S., Mishra, S. K., Yadav, A. K., Kathiravan, P., Arora, R., et al. (2015). Association analysis of polymorphism in thyroglobulin gene promoter with milk production traits in riverine buffalo (*Bubalus bubalis*). *Meta Gene* 5, 157–161. doi: 10.1016/j.mgene.2015.07.005

El Debaky, H. A., Kutchy, N. A., Ul-Husna, A., Indriastuti, R., Akhter, S., Purwantara, B., et al. (2019). Review: Potential of water buffalo in world agriculture: challenges and opportunities. *Appl. Anim. Sci.* 35, 255–268. doi: 10.15232/aas.2018-01810

El-Komy, S. M., Saleh, A. A., Abdel-Hamid, T. M., and El-Magd, M. A. (2020). Association of GHR polymorphisms with milk production in buffaloes. *Animals* 10, 1–16. doi: 10.3390/ani10071203

Escarcha, J. F., Lassa, J. A., Palacpac, E. P., and Zander, K. K. (2020). Livelihoods transformation and climate change adaptation: the case of smallholder water buffalo farmers in the Philippines. *Environ. Dev.* 33:100468. doi: 10.1016/j.envdev.2019.100468

FAO (2019). The status of application, capacities and the enabling environment for agricultural biotechnologies in the Asia-Pacific region. Regional background study (Licence: CC BY-NC-SA 3.0 IGO).

FAOSTAT (2018). About live animals, data on buffaloes. Available at: http://www.fao.org/faostat/en/#data/TA (Accessed July 14, 2020).

FDA (2020). FDA approves first-of-its-kind intentional genomic alteration in line of domestic pigs for both human food, potential therapeutic uses. Available at: https://www.fda.gov/news-events/press-announcements/fda-approves-first-its-kind-intentional-genomic-alteration-line-domestic-pigs-both-human-food (Accessed January 27, 2021).

Flores, E. B., Maramba, J. F., Aquino, D. L., Abesamis, A. F., Cruz, A. F., and Cruz, L. C. (2007). Evaluation of milk production performance of dairy buffaloes raised in various herds of the Philippine Carabao Center. *Ital. J. Anim. Sci.* 6(Suppl. 2), 295–298. doi: 10.4081/ijas.2007.s2.295

Gan, Q.-F., Zhang, L.-P., Li, J.-Y., Hou, G.-Y., Li, H.-D., Gao, X., et al. (2008). Association analysis of thyroglobulin gene variants with carcass and meat quality traits in beef cattle. *J. Appl. Genet.* 49, 251–255. doi: 10.1007/BF03195621

Hamid, M., Zaman, M. A., Rahman, A., and Hossain, K. M. (2017). Buffalo genetic resources and their conservation in Bangladesh. *Res. J. Vet. Sci.* 10, 1–13. doi: 10.3923/rjvs.2017.1.13

Harisah, M., Azmi, T. I., Hilmi, M., Vidyadaran, M. K., Bongso, T. A., Nava, Z. M., et al. (1989). Identification of crossbred buffalo genotypes and their chromosome segregation patterns. *Genome* 32, 999–1002. doi: 10.1139/g89-544

He, Z., Zhang, T., Jiang, L., Zhou, M., Wu, D., Mei, J., et al. (2018). Use of CRISPR/Cas9 technology efficiently targetted goat myostatin through zygotes microinjection resulting in double-muscled phenotype in goats. *Biosci. Rep.* 38, 1–8. doi: 10.1042/BSR20180742

Herrera, J. R., Flores, E. B., Duijvesteijn, N., Gondro, C., and Werf, J. H. J. (2018). "Genome-wide association study for milk traits in Philippine dairy buffaloes" in *Proceedings of the World Congress on Genetics Applied to Livestock Production*. 825.

Herrera, J. R., Flores, E., Gondro, C., and Van Der Werf, J. (2016). Performance of the Axiom 90k Buffalo Genotyping Array in four Philippine water buffalo populations. *Revista CES Medicina Veterinaria y Zootecnia* 11:210.

Hostiou, N., Fagon, J., Chauvat, S., Turlot, A., Kling-Eveillard, F., Boivin, X., et al. (2017). Impact of precision livestock farming on work and human-animal interactions on dairy farms. A review. *Biotechnol. Agron. Soc. Environ.* 21, 268–275. doi: 10.25518/1780-4507.13706

Hu, R., Fan, Z. Y., Wang, B. Y., Deng, S. L., Zhang, X. S., Zhang, J. L., et al. (2017). Rapid communication: generation of FGF5 knockout sheep via the CRISPR/Cas9 system. *J. Anim. Sci.* 95, 2019–2024. doi: 10.2527/jas.2017.1503

Iamartino, D., Nicolazzi, E. L., Van Tassell, C. P., Reecy, J. M., Fritz-waters, E. R., Koltes, J. E., et al. (2017). Design and validation of a 90K SNP genotyping assay for the water buffalo (*Bubalus bubalis*). *PLoS One* 12:e0185220. doi: 10.1371/journal.pone.0185220

Iamartino, D., Williams, J. L., Sonstegard, T., Reecy, J., Van Tassell, C., Nicolazzi, E. L., et al. (2013). The buffalo genome and the application of genomics in animal management and improvement. *Buffalo Bull.* 32, 151–158. doi: 10.13140/RG.2.2.30951.85924

Iqbal, M. W., Zhang, Q., Yang, Y., Li, L., Zou, C., Huang, C., et al. (2018). Comparative study of rumen fermentation and microbial community differences between water buffalo and Jersey cows under similar feeding conditions. *J. Appl. Anim. Res.* 46, 740–748. doi: 10.1080/09712119.2017.1394859

Komariah, B., Dzaki, M., Aditia, E. L., and Mendrofa, V. A. (2020). Performance and development strategy for Swamp Buffalo (*Bubalus bubalis*) in Serang District Indonesia. *Jurnal Ilmu Produksi Dan Teknologi Hasil Peternakan* 8, 54–60. doi: 10.29244/jipthp.8.2.54-60

Koobkaew, K., Nakavisut, S., and Kiyothong, K. (2013). "Thailand Buffalo Strategic Plan 2012–2016." in *Conference: 10th World Buffalo Congress/7th Asian Buffalo Congress*; May 6-8, 2013; 32, 83–89.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lau, C. H., Drinkwater, R. D., Yusoff, K., Tan, S. G., Hetzel, D. J. S., and Barker, J. S. F. (1998). Genetic diversity of Asian water buffalo (*Bubalus bubalis*): mitochondrial DNA D-loop and cytochrome b sequence variation. *Anim. Genet.* 29, 253–264. doi: 10.1046/j.1365-2052.1998.00309.x

Lee, K., Uh, K., and Farrell, K. (2020). Current progress of genome editing in livestock. *Theriogenology* 150, 229–235. doi: 10.1016/j.theriogenology.2020.01.036

Li, J., Liu, S., Li, Z., Zhang, S., Hua, G., Salzano, A., et al. (2018). DGAT1 polymorphism in Riverine buffalo, Swamp buffalo and crossbred buffalo. *J. Dairy Res.* 85, 412–415. doi: 10.1017/S0022029918000468

Li, W. R., Liu, C. X., Zhang, X. M., Chen, L., Peng, X. R., He, S. G., et al. (2017). CRISPR/Cas9-mediated loss of FGF5 function increases wool staple length in sheep. *FEBS J.* 284, 2764–2773. doi: 10.1111/febs.14144

Li, G., Zhou, S., Li, C., Cai, B., Yu, H., Ma, B., et al. (2019). Base pair editing in goat: nonsense codon introgression into FGF5 results in longer hair. *FEBS J.* 286, 4675–4692. doi: 10.1111/febs.14983

Liu, X., Wang, Y., Tian, Y., Yu, Y., Gao, M., Hu, G., et al. (2014). "Generation of mastitis resistance in cows by targeting human lysozyme gene to β-casein locus using zinc-finger nucleases" in *Proceedings of the Royal Society B: Biological Sciences*. 281.

Low, W. Y., Tearle, R., Bickhart, D. M., Rosen, B. D., Kingan, S. B., Swale, T., et al. (2019). Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity. *Nat. Commun.* 10:260. doi: 10.1038/s41467-018-08260-0

Low, W. Y., Tearle, R., Liu, R., Koren, S., Rhie, A., Bickhart, D. M., et al. (2020). Haplotype-resolved genomes provide insights into structural variation and gene content in Angus and Brahman cattle. *Nat. Commun.* 11:2071. doi: 10.1038/s41467-020-15848-y

Luo, X., Zhou, Y., Zhang, B., Zhang, Y., Wang, X., Feng, T., et al. (2020). Understanding divergent domestication traits from the whole-genome sequencing of swamp- and river-buffalo populations. *Natl. Sci. Rev.* 7, 686–701. doi: 10.1093/nsr/nwaa024

McFarlane, G. R., Salvesen, H. A., Sternberg, A., and Lillico, S. G. (2019). On-farm livestock genome editing using cutting edge reproductive technologies. *Front. Sustain. Food Syst.* 3:106. doi: 10.3389/fsufs.2019.00106

Mei, C., Wang, H., Liao, Q., Khan, R., Raza, S. H. A., Zhao, C., et al. (2019). Genome-wide analysis reveals the effects of artificial selection on production and meat quality traits in Qinchuan cattle. *Genomics* 111, 1201–1208. doi: 10.1016/j.ygeno.2018.09.021

Miller, S. (2010). Genetic improvement of beef cattle through opportunities in genomics. *Rev. Bras. Zootec.* 39, 247–255. doi: 10.1590/S1516-35982010001300027

Molteni, M. (2019). Brazil's plans for gene-edited cows got scrapped—here's why. Wired – Science. Available at: https://www.wired.com/story/brazils-plans-for-gene-edited-cows-got-scrappedheres-why/ (Accessed January 27, 2021).

Nardone, A., Ronchi, B., Lacetera, N., Ranieri, M. S., and Bernabucci, U. (2010). Effects of climate changes on animal production and sustainability of livestock systems. *Livest. Sci.* 130, 57–69. doi: 10.1016/j.livsci.2010.02.011

Negussie, E., de Haas, Y., Dehareng, F., Dewhurst, R. J., Dijkstra, J., Gengler, N., et al. (2017). Invited review: large-scale indirect measurements for enteric methane emissions in dairy cattle: A review of proxies and their potential for use in management and breeding decisions. *J. Dairy Sci.* 100, 2433–2453. doi: 10.3168/jds.2016-12030

Nguyen, V. T. (2000). "Buffalo production and performance in Vietnam." in *Third Asian Buffalo Congress*; March 27–31, 2000; Kandy, Sri Lanka, 375–383.

Norris, A. L., Lee, S. S., Greenlees, K. J., Tadesse, D. A., Miller, M. F., and Lombardi, H. A. (2020). Template plasmid integration in germline genome-edited cattle. *Nat. Biotechnol.* 38, 163–164. doi: 10.1038/s41587-019-0394-6

OECD (2017). "Southeast Asia: prospects and challenges" in *OECD & Food and Agriculture Organization of the United Nations, OECD-FAO Agricultural Outlook 2017–2026*. 59–99.

Othman, R. (2014). Improving the reproductive performance of buffaloes in Sabah, Malaysia. *J Anim. Health Prod.* 2, 1–4. doi: 10.14737/journal.jahp/2014/2.1.1.4

Palacpac, E. P. (2010). Spurring dairy buffalo development in the Philippines through cooperatives, negotiations, and networks. *J. Rural Dev.* 38, 70–86.

Paraguas, A. M., Cailipan, T. P. C., Flores, E. B., and Villamor, L. P. (2018). Morphology and phylogeny of swamp buffaloes (*Bubalus bubalis*) in Calayan Island, Cagayan. *Philipp. J. Vet. Anim. Sci.* 44, 59–67.

Pérez-Enciso, M., and Zingaretti, L. M. (2019). A guide for using deep learning for complex trait genomic prediction. *Genes* 10:553. doi: 10.3390/genes10070553

Proudfoot, C., Carlson, D. F., Huddart, R., Long, C. R., Pryor, J. H., King, T. J., et al. (2015). Genome edited sheep and cattle. *Transgenic Res.* 24, 147–153. doi: 10.1007/s11248-014-9832-x

Rojas-Downing, M. M., Nejadhashemi, A. P., Harrigan, T., and Woznicki, S. A. (2017). Climate change and livestock: impacts, adaptation, and mitigation. *Clim. Risk Manag.* 16, 145–163. doi: 10.1016/j.crm.2017.02.001

Sahu, J., Yadav, A., Pal, P., Kumar, R., Chaudhary, S., Kumar, S., et al. (2019). Recent breed improvement projects on buffaloes in India. *Indian Farmer* 6, 322–328.

Salas, R. C. D., Van Der Lende, T., Udo, H. M. J., Mamuad, F. V., Garillo, E. P., and Cruz, L. C. (2000). Comparison of growth, milk yield and draughtability of Murrah-Philippine crossbred and Philippine Native Buffaloes. *Asian Australas. J. Anim. Sci.* 13, 580–586. doi: 10.5713/ajas.2000.580

Sanghuayphrai, N., Nakavisut, S., Dongpaletum, C., Phothikanit, G., and Supanun, S. (2013). Genetic parameters and trends for weaning weight and calving interval of department of Livestock Development swamp buffalo. *Buffalo Bull.* 32, 717–720.

Sapapanan, S., Rimkeeree, K., Chusen, P., Dongpaleetan, C., Photikanit, G., Na, A., et al. (2013). "The problems and obstacles on raising buffaloes of local farmers in Central Thailand: a case study of Saraburi Province" in *10th World Buffalo Congress and 7th Asian Buffalo Congress*; May 6–8, 2013; Phuket, Thailand.

Sarataphan, N., Narongwanichgarn, W., and Maneerat, S. (2017). Phylogenetic analysis of a thai wild water buffalo (*Bubalus arnee*) through mitochondrial control region. *Int. J. Conserv. Sci.* 8, 105–112.

Shaari, N. A. L., Jaoi-Edward, M., Loo, S. S., Salisi, M. S., Yusoff, R., Ab Ghani, N. I., et al. (2019). Karyotypic and mtDNA based characterization of Malaysian water buffalo. *BMC Genet.* 20:37. doi: 10.1186/s12863-019-0741-0

Sonstegard, T. S., Van Tassell, C. P., and Ashwell, M. S. (2001). Dairy cattle genomics: tools to accelerate genetic improvement? *J. Anim. Sci.* 79, E307–E315. doi: 10.2527/jas2001.79E-SupplE307x

Su, X., Cui, K., Du, S., Li, H., Lu, F., Shi, D., et al. (2018). Efficient genome editing in cultured cells and embryos of Debao pig and swamp buffalo using the CRISPR/Cas9 system. *In Vitro Cell. Dev. Biol. Anim.* 54, 375–383. doi: 10.1007/s11626-018-0236-8

Sun, T., Shen, J., Achilli, A., Chen, N., Chen, Q., Dang, R., et al. (2020a). Genomic analyses reveal distinct genetic architectures and selective pressures in buffaloes. *Gigascience* 9:giz166. doi: 10.1093/gigascience/giz166

Sun, T., Wang, S., Chanthakhoun, V., Dang, R., Huang, Y., Chen, H., et al. (2020b). Multiple domestication of swamp buffalo in China and South East Asia. *J. Anim. Breed. Genet.* 137, 331–340. doi: 10.1111/jbg.12445

Supanuam, P., Tanomtong, A., Jantarat, S., Kakampuy, W., Kaewsri, S., and Kenthao, A. (2012). Standardized karyotype and idiogram of Thai native swamp buffalo, *Bubalus bubalis* (Artiodactyla, Bovidae) by convention staining, G-banding, C-banding and NOR-banding techniques. *Thai J. Genet.* 3, 63–93. doi: 10.14456/tjg.2010.8

Suphachavalit, S., Sricharoen, P., Luesopha, T., Srisakdi, T., Chiangmai, A., and Boonprong, S. (2013). "Swamp buffalo production system and needs for extension on local scale farmers in the lower northeast of Thailand." in *10th World Buffalo Congress and 7th Asian Buffalo Congress*; May 6–8, 2013; Phuket, Thailand.

Suryanto, B., Arifin, M., and Rianto, E. (2002). Potential of swamp buffalo development in Central Java, Indonesia. *Buffalo Bull.* 21, 3–9.

Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol. Biol. Evol.* 9, 678–687. doi: 10.1093/oxfordjournals.molbev.a040752

Tanaka, K., Solis, C. D., Masangkay, J. S., Maeda, K. I., Kawamoto, Y., and Namikawa, T. (1996). Phylogenetic relationship among all living species of the genus Bubalus based on DNA sequences of the cytochrome b gene. *Biochem. Genet.* 34, 443–452. doi: 10.1007/BF00570125

Tullo, E., Finzi, A., and Guarino, M. (2019). Review: Environmental impact of livestock farming and Precision Livestock Farming as a mitigation strategy. *Sci. Total Environ.* 650, 2751–2760. doi: 10.1016/j.scitotenv.2018.10.018

Ungerfeld, E. M. (2018). Inhibition of rumen methanogenesis and ruminant productivity: a meta-analysis. *Front. Vet. Sci.* 5:113. doi: 10.3389/fvets.2018.00113

Upadhyay, R. C., Singh, S. V., Kumar, A., Gupta, S. K., and Ashutosh, A. (2007). Impact of climate change on milk production of Murrah buffaloes. *Ital. J. Anim. Sci.* 6(Suppl. 2), 1329–1332. doi: 10.4081/ijas.2007.s2.1329

Wang, S., Chen, N., Capodiferro, M. R., Zhang, T., Lancioni, H., Zhang, H., et al. (2017). Whole mitogenomes reveal the history of Swamp Buffalo: initially shaped by glacial periods and eventually modelled by domestication. *Sci. Rep.* 7:4708. doi: 10.1038/s41598-017-18302-0

Williams, J. L., Iamartino, D., Pruitt, K. D., Sonstegard, T., Smith, T. P. L., Low, W. Y., et al. (2017). Genome assembly and transcriptome resource for river buffalo, *Bubalus bubalis* (2n = 50). *Gigascience* 6, 1–6. doi: 10.1093/gigascience/gix088

Wu, M., Du, L., Liu, R., Wei, C., Wang, Y., Yang, L., et al. (2018). Double-muscled phenotype in mutant sheep directed by the CRISPRCas9 system. *Cloning Transgenes.* 7, 3–7. doi: 10.4172/2168-9849.1000161

Xu, Y., Liu, X., Fu, J., Wang, H., Wang, J., Huang, C., et al. (2020). Enhancing genetic gain through genomic selection: from livestock to plants. *Plant Commun.* 1:100005. doi: 10.1016/j.xplc.2019.100005

Yang, B. Z., Liang, X. W., Qin, J., Yang, C. J., and Shang, J. H. (2013). Brief introduction to the development of Chinese dairy buffalo industry. *Buffalo Bull.* 32, 111–120.

Yindee, M., Vlamings, B. H., Wajjwalku, W., Techakumphu, M., Lohachit, C., Sirivaidyapong, S., et al. (2010). Y-chromosomal variation confirms independent domestications of swamp and river buffalo. *Anim. Genet.* 41, 433–435. doi: 10.1111/j.1365-2052.2010.02020.x

Young, R., Lefevre, L., Bush, S. J., Joshi, A., Young, R., and Hume, D. A. (2019). A gene expression atlas of the domestic water buffalo (*Bubalus bubalis*). *Front. Genet.* 10, 1–14. doi: 10.3389/fgene.2019.00668

Young, A. E., Mansour, T. A., McNabb, B. R., Owen, J. R., Trott, J. F., Brown, C. T., et al. (2020). Genomic and phenotypic analyses of six offspring of a genome-edited hornless bull. *Nat. Biotechnol.* 38, 225–232. doi: 10.1038/s41587-019-0266-0

Zhang, Y., Colli, L., and Barker, J. S. F. (2020). Asian water buffalo: domestication, history and genetics. *Anim. Genet.* 51, 177–191. doi: 10.1111/age.12911

Zhang, Yi, Lu, Y., Yindee, M., Li, K. Y., Kuo, H. Y., Ju, Y.-T., et al. (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. doi: 10.1111/mec.13518

Zhang, J., Xu, C., Huo, D., Hu, Q., and Peng, Q. (2017). Comparative study of the gut microbiome potentially related to milk protein in Murrah buffaloes (*Bubalus bubalis*) and Chinese Holstein cattle. *Sci. Rep.* 7, 1–11. doi: 10.1038/srep42189

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 23 March 2021
doi: 10.3389/fgene.2021.662609

Comparative Genomics, Evolutionary and Gene Regulatory Regions Analysis of Casein Gene Family in *Bubalus bubalis*

Saif ur Rehman[1†], Tong Feng[1†], Siwen Wu[1], Xier Luo[1], An Lei[2], Basang Luobu[3], Faiz-ul Hassan[4*] and Qingyou Liu[1*]

[1] State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangxi University, Nanning, China, [2] National Engineering Laboratory for Animal Breeding, Key Laboratory of Animal Genetics, Breeding and Reproduction of the Ministry of Agriculture, College of Animal Science and Technology, China Agricultural University, Beijing, China, [3] Shannan Animal Husbandry and Veterinary Terminus, Xizang, China, [4] Faculty of Animal Husbandry, Institute of Animal and Dairy Sciences, University of Agriculture, Faisalabad, Pakistan

OPEN ACCESS

Edited by:
Guohua Hua,
Huazhong Agricultural University, China

Reviewed by:
Yongwang Miao,
Yunnan Agricultural University, China
Gregorio Miguel Ferreira De Camargo,
Federal University of Bahia, Brazil

***Correspondence:**
Faiz-ul Hassan
f.hassan@uaf.edu.pk
Qingyou Liu
qyliu-gene@gxu.edu.cn;
qyliu-gene@qq.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

***Received:** 01 February 2021*
***Accepted:** 01 March 2021*
***Published:** 23 March 2021*

Citation:
Rehman S, Feng T, Wu S, Luo X, Lei A, Luobu B, Hassan F and Liu Q (2021) Comparative Genomics, Evolutionary and Gene Regulatory Regions Analysis of Casein Gene Family in Bubalus bubalis. Front. Genet. 12:662609. doi: 10.3389/fgene.2021.662609

Buffalo is a luxurious genetic resource with multiple utilities (as a dairy, draft, and meat animal) and economic significance in the tropical and subtropical regions of the globe. The excellent potential to survive and perform on marginal resources makes buffalo an important source for nutritious products, particularly milk and meat. This study was aimed to investigate the evolutionary relationship, physiochemical properties, and comparative genomic analysis of the casein gene family (*CSN1S1, CSN2, CSN1S2,* and *CSN3*) in river and swamp buffalo. Phylogenetic, gene structure, motif, and conserved domain analysis revealed the evolutionarily conserved nature of the casein genes in buffalo and other closely related species. Results indicated that casein proteins were unstable, hydrophilic, and thermostable, although αs1-CN, β-CN, and κ-CN exhibited acidic properties except for αs2-CN, which behaved slightly basic. Comparative analysis of amino acid sequences revealed greater variation in the river buffalo breeds than the swamp buffalo indicating the possible role of these variations in the regulation of milk traits in buffalo. Furthermore, we identified lower transcription activators STATs and higher repressor site YY1 distribution in swamp buffalo, revealing its association with lower expression of casein genes that might subsequently affect milk production. The role of the main motifs in controlling the expression of casein genes necessitates the need for functional studies to evaluate the effect of these elements on the regulation of casein gene function in buffalo.

Keywords: buffalo breeds, caseins, evolution, regulatory regions, milk yield

INTRODUCTION

Buffalo is a luxurious genetic resource with multiple utilities (as a dairy, draft, and meat animal) and economic significance in the tropical and subtropical regions of the globe (Rehman et al., 2019, 2020; Luo et al., 2020). The domesticated buffalo is grouped into river buffalo with karyotype 2n = 50 primarily present in southwestern Asia, India, south Mediterranean Europe, and Egypt and swamp buffalo with 2n = 48 distributed across Southeast Asia, southern and southeast China, where the swamp buffalo is used as draft power in the rice paddy fields while the river buffalo is mainly reared for milk production (Moioli et al., 2001; Fan et al., 2020; Luo et al., 2020). The excellent potential to survive and perform on marginal resources under harsh environmental conditions makes buffalo an important source for nutritious products, particularly milk and meat. Buffalo contributes about 13% of global milk production where the river buffalo produces 2,000 kg milk per year and swamp buffalo annual production is 500–600 kg (Basilicata et al., 2018; Fan et al., 2020; Lu et al., 2020). Moreover, the physio-chemical characteristics of buffalo milk are different from cow milk, and buffalo milk is relished due to its peculiar taste and higher butterfat content (Li et al., 2020).

Buffalo milk contains higher protein, fat, and total solid contents relative to dairy cow milk (Ahmad et al., 2013). The milk proteins are broadly categorized into whey (serum) protein and casein protein families based on their physio-chemical properties. Casein (CN) is the major milk protein, contributing 80% of the whole milk proteins including α-s1-CN, α-s2-CN, β-CN, and k-CN. Each CN protein has its unique amino acid configuration, genetic and functional properties (Fan et al., 2020). Milk CNs are physiologically important as they provide food to the newborn and are associated with milk processing properties and lactation behaviors of dairy animals (Nilsen et al., 2009).

Notably, the CN protein is characterized into calcium-sensitive αS1, αS2, and β caseins, in young one sustenance bone growth through providing calcium, and phosphorus enriched stable micelles, and the Ca-insensitive κ-casein (Pauciullo and Erhardt, 2015). So far, in mammals, caseins are the main constituent of milk proteins. The casein proteins coding genes CSN1S1 (αs1-casein), CSN1S2 (αs2-casein), CSN2 (β-casein), and CSN3 (κ-casein), have been mapped in the 250-350kb genomic DNA cluster on chromosome 6 in sheep, goat, and cattle (Rijnkels, 2002).

Casein is considered a powerful molecular model for evolutionary research (Kawasaki et al., 2011). It is also a useful tool to better understand the genetic architecture of less-studied species, phylogenetic relationships among mammalian species, and domestic animals, particularly the buffalo breeds (rive and swamp). From a physiological standpoint, there is a difference in milk yield and composition traits, including protein, fat, and solid contents among different species or breeds, suggesting the potential role of gene regulatory regions in these breeds. Exploring the genetic architecture and evolutionary processes is imperative to understand the regulatory mechanisms of the casein gene family in the buffalo. This study aims to investigate the evolutionary relationship, physiochemical properties, comparative genomics, and gene regulatory regions analysis of the casein gene family in river and swamp buffalo.

MATERIALS AND METHODS

The sequences of different casein genes (*CSN1S1, CSN2, CSN1S2,* and *CSN3*) of *Bos taurus* were retrieved from NCBI[1] and used as queries for the identification of casein genes from the buffalo genome. The buffalo (river and swamp) whole-genome sequences were downloaded from the Bigdata center and NCBI[1,2]. The *Bos taurus* casein protein sequences (XP_005208084.1, XP_024848786.1, XP_010804480.2, and XP_024848756.1) were used in BLAST search with an E value less or equal to $1.0 \times e^{-5}$ with all default parameters, to retrieve non-redundant protein sequences of the buffalo. To avoid ambiguity, the redundancy of the sequences was checked. The chromosomal locations of casein genes were obtained from buffalo genome resources through the GFF file of annotated buffalo genome with corresponding gene positions in the MCScanX program as reported earlier (Wang et al., 2012).

The Maximum Likelihood method based on the JTT matrix model was used to infer the evolutionary history of representative species (Jones et al., 1992). The accessions number of amino acid sequences used to construct the phylogenetic tree and holology of the representative speices sequence are given in **Supplementary Table S1**. The likelihood phylogram of 44 amino acid sequences with the highest log (−1641.52) was downloaded and the percentage of trees in which the associated taxa clustered together presented next to the branches. A bootstrap value of 3,000 replicates was used and the percentage of resampling was visualized on the node of the phylogram. All the missing and gaped positions were eliminated and MEGA7 was used to conduct the evolutionary analyses (Kumar et al., 2016).

Moreover, the genomic and coding sequence data of casein genes from buffalo and cattle were submitted to Gene Structure Display Server 2.0[3], for gene structure analysis and visualization of untranslated regions and exon-intron structure (Hu et al., 2015). Additionally, 10 MEME (Multiple EM for Motif Elicitation) conserved motifs of caseins were explored using the MEME Suite[4] (Bailey et al., 2006). The NCBI conserved domain (CDD) database was used to confirm the conserved domains[5].

ProtParam tool was used to illustrate the physio-chemical properties of buffalo casein proteins including the isoelectric point (pI), grand average of hydropathicity (GRAVY), molecular weight (MW), number of amino acids, instability index (II), and aliphatic index (AI) (Gasteiger et al., 2003). Multiple sequence alignment of casein protein sequences was performed in Multiple Align Show to visualize the sequence variations and indels[6].

The genomic sequences of casein genes of Mediterranean and swamp buffalo were subjected to the Promoter 2.0 Prediction Server[7] to detect potential signals for putative transcription binding factor. The site with a score > 1.0 was presumed as a high likelihood predicted site and the putative transcription binding factor site sequence was searched in the 100bp upstream regions from the high likelihood predicted site (Knudsen, 1999). Further, the genomic sequences were analyzed in TFBIND software[8] by using the transcription factor database TRANSFAC R.3.4 weight

[1] https://www.ncbi.nlm.nih.gov/

[2] https://bigd.big.ac.cn

[3] http://gsds.gao-lab.org/

[4] http://meme-suite.org/tools/meme

[5] https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi

[6] https://www.bioinformatics.org/sms/multi_align.html

[7] http://www.cbs.dtu.dk/services/Promoter/

[8] http://tfbind.hgc.jp/

matrix to find the transcription factor binding sites (Tsunoda and Takagi, 1999). According described previously, four potential transcription factor binding sites (GATA, TATA, STAT, and OCT1) (Hennighausen and Robinson, 1998; Robinson et al., 1998; Rosen et al., 1999; Wheeler et al., 2001; Wyszomierski and Rosen, 2001; Yamashita et al., 2001; Chughtai et al., 2002; Pauciullo et al., 2019) and one repressor site (YY1) (Helman et al., 1998; Tomic et al., 1999) in casein genes of Mediterranean and swamp buffalo in 100bp upstream regions of the potential signal site were calculated (Wyszomierski and Rosen, 2001). The significant difference for the distribution of putative transcription factor binding and repressor sites in Mediterranean and swamp buffalo was statistically evaluated by using a *t*-test with a *P*-value of < 0.05 as statistical significance. Moreover, the potential nuclear hormone receptor sites in the genome of Mediterranean buffalo were detected by using the NHR scan[9].

RESULTS

The molecular phylogenetic analysis of representative bovine species revealed that all the casein gene sequences were clustered into four groups; *CSN1S1*, *CSN2*, *CSN1S2*, and *CSN3* (**Figure 1**). Additionally, overall phylogenetic relationships revealed that *Bubalus bubalis CSN* gene family is more closely related to *Bos mutus, Bos taurus,* and *Bos indicus* sharing higher sequence homology about 93, 91, and 90%, respectively, as compared to the *Capra hircus, Ovis aries* and hybrid cattle with 86, 84, and 74% similarity respectively. Moreover, distantly related species included *Camelus ferus* and *Equus caballus* with 55 and 50% resemblance, respectively (**Supplementary Table S2**).

Furthermore, to perform the structural characterization of the *CSN* gene family in different species, analysis of gene organization, motifs pattern, and the conserved domains were carried out considering their phylogenetic relationships (**Figure 2**). In casein genes, 10 MEME conserved motifs were identified (**Figure 2C**). Motif 3 corresponding to 21 amino acid was annotated as kappa casein (K-CN) domain while motif 4, 5, and 6 were annotated as casein domain after the Pfams search (**Table 1**). The CDD BLAST was used to confirm the identified conserved domains (**Figure 2D**). Additionally, the ODAM and PHA03247 superfamily domain has also been dredged up in *CSN* genes (**Figure 2D**). Besides, the upstream and downstream untranslated regions (UTRs) and intron structure considerably varied, structural analysis of the gene indicated that buffalo *CSN* genes in the same group possess a corresponding number of introns and exons (**Figure 2B**). However, different *CSN* gene groups exhibited a variable pattern of introns and exons (**Figure 2B**).

Physiochemical properties of the *CSN* gene family in *Bubalus bubalis* was determined in terms of their distribution on the chromosome, exon count, molecular weight (Da), number of the amino acids (A.A) in each peptide, aliphatic index (AI), isoelectric point (pI), instability index (II) and Grand Average of hydropathicity Index (GRAVY) (**Table 2**). All the *CSN* genes were found on chromosome 7 in the region between ∼250 kb that harbors a variable number of exons and inconsistent length of the gene with amino acid residues (**Table 2**). The molecular weight of CN proteins ranged from 21 to 29 kDa. The CN peptides in buffalo were observed as unstable but thermostable proteins as the aliphatic index for all caseins had values > 65. Further, the pI values revealed that all CN proteins αs1-CN, β-CN, and κ-CN were acidic peptides except αs2-CN which behaved slightly basic in nature (**Table 2**). Lower values of GRAVY indicate the hydrophilic nature of buffalo CN proteins (**Table 2**).

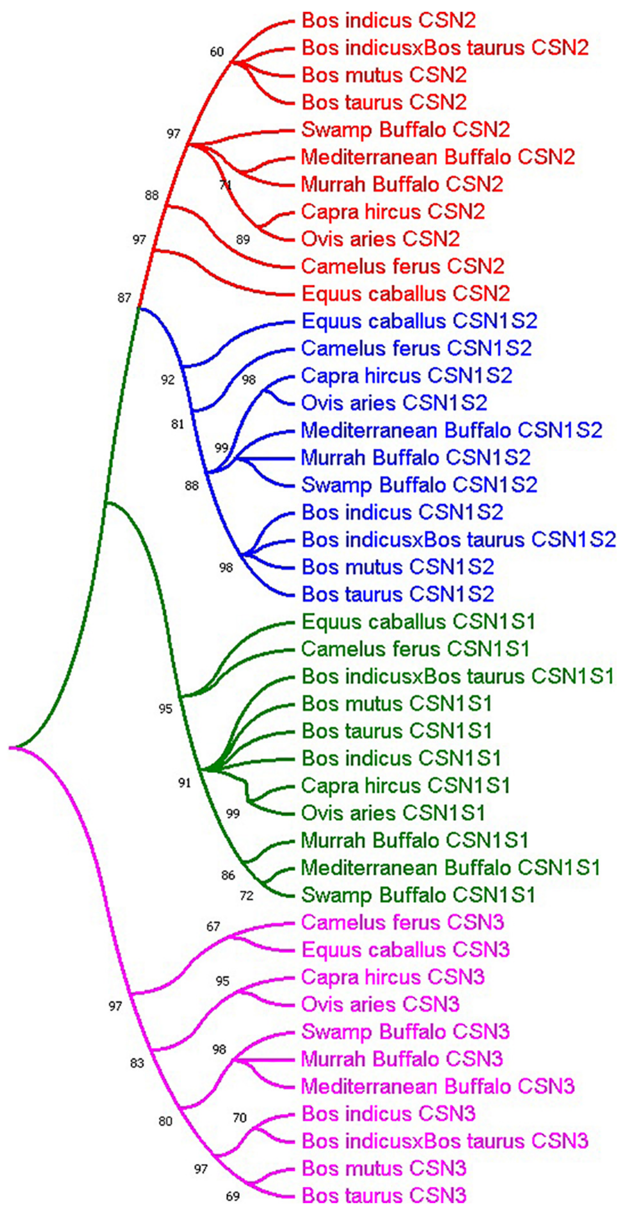

FIGURE 1 | Molecular phylogenetic analysis of casein gene family (green; *CSN1S1*, blue; *CSN1S2*, red; *CSN2* and fuchsia; *CSN3*) in representative species.

[9]http://www.cisreg.ca/cgi-bin/NHR-scan/nhr_scan.cgi

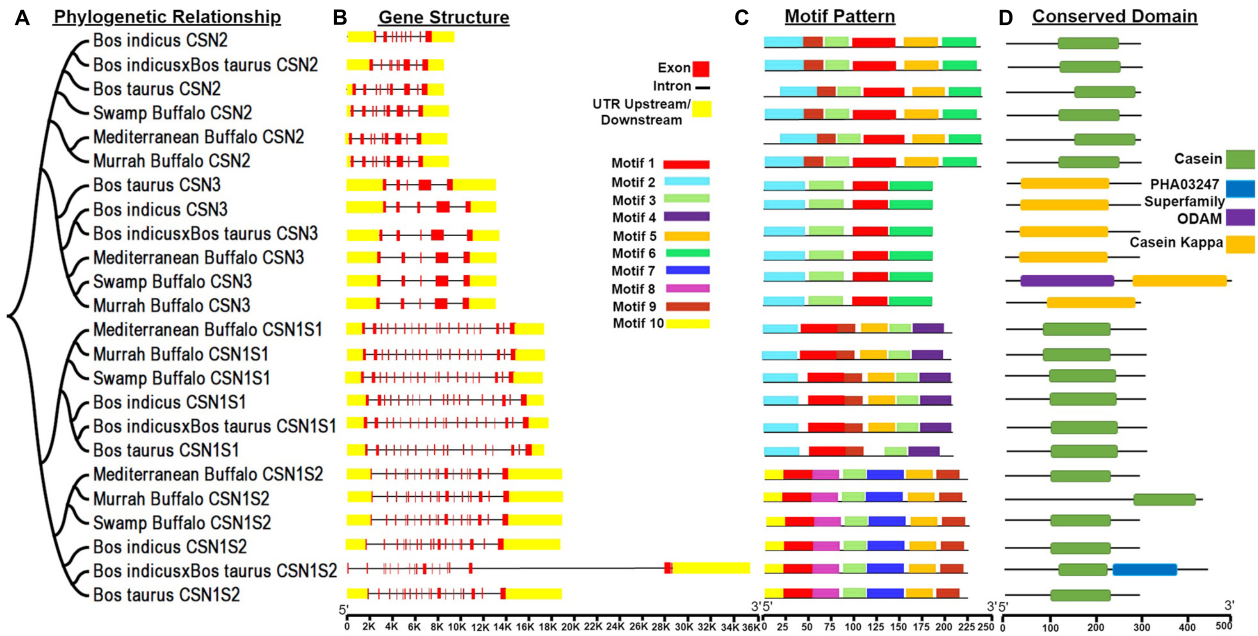

FIGURE 2 | Phylogenetic relationships, gene structure, motif patterns, and conserved domain regions of the casein proteins gene family. **(A)** Phylogenetic relationship of 24 amino acid sequences of casein proteins. **(B)** Gene structure of casein. **(C)** Motif pattern. **(D)** Conserved domain regions of the casein proteins gene family. Ten putative motifs are indicated in different colored boxes. For details of motifs refer to **Table 1**.

TABLE 1 | Ten differentially conserved motifs detected in casein protein (*CSN1S1*, *CSN1S2*, *CSN2*, and *CSN3*) gene family.

Motif	Protein sequence	Length	Pfam domain
MEME-1	NTLPENISSAEETDVAREPYKQLEAMAISPSKEALAT	37	–
MEME-2	MKLLILTCLVALALARPLEELKVQGEPQEVLNENEERFFVA	41	–
MEME-3	BKYQQKELALINNQYLAYPYY	21	K-CN
MEME-4	FRQFYQLDAYPSGAWYYVPLGTQYTDAPSFSDIPNPIGSENSGKTTMPLW	50	CN
MEME-5	VEVFTEKTKLTEEDVERLNLLKKJSQSYMHFPK	33	CN
MEME-6	IPSINKILPVEPKAVPYPZADEPIVAFLEYSEEVJGPVPEP	41	CN
MEME-7	QYLYQGPIVLNPWDQVKRNAVPITPTLNR	29	–
MEME-8	TFCKEVVRNANEEEYSIGSSSEESAEVAT	29	–
MEME-9	NKEVEKFQKEEKPST	15	–
MEME-10	MKFFIFTCLLAVALA	15	–

K-CN; kappa casein, CN; Casein.

TABLE 2 | Physiochemical properties of the casein gene family in *Bubalus bubalis* (Mediterranean breed).

Buffalo breed	Gene	Chromosome	Exon count	MW (Da)	A.A	pI	AI	II	GRAVY
Italian	CSN1S1	7	19	23451.87	206	4.89	90.87	59.32	−0.332
Italian	CSN1S2	7	18	25081.53	213	7.66	73.66	45.54	−0.699
Italian	CSN2	7	9	29110.29	259	6.31	100.04	92.21	−0.124
Italian	CSN3	7	5	21409.62	190	6.83	86.21	49.60	−0.232

MW, Molecular Weight in Daltons; A.A, number of amino acids; pI, Isoelectric point; AI, Aliphatic Index; II, Instability Index; and GRAVY, Grand Average of hydropathicity Index.

Comparative amino acid analysis of buffalo breeds revealed 7 indels in *CSN* genes including a single indel in both *CSN1S1* and *CSN3* while two indels in *CSN1S2* and 3 in *CSN2*. The *CSN1S1* gene has an indel of 8 amino acids at position 50 > 57 whereas single amino acid change V46 > M in Murrah and S193 > L in Mediterranean buffalo was also observed (**Figure 3A**). Two indels of variable length were in CSN1S2, where 9 amino acid indel is positioned at 149 > 157, presumably is due to an alternative splicing of exon 13, and the second indel toward the terminal end of the peptide with a length of three amino acids 220 > 222. In

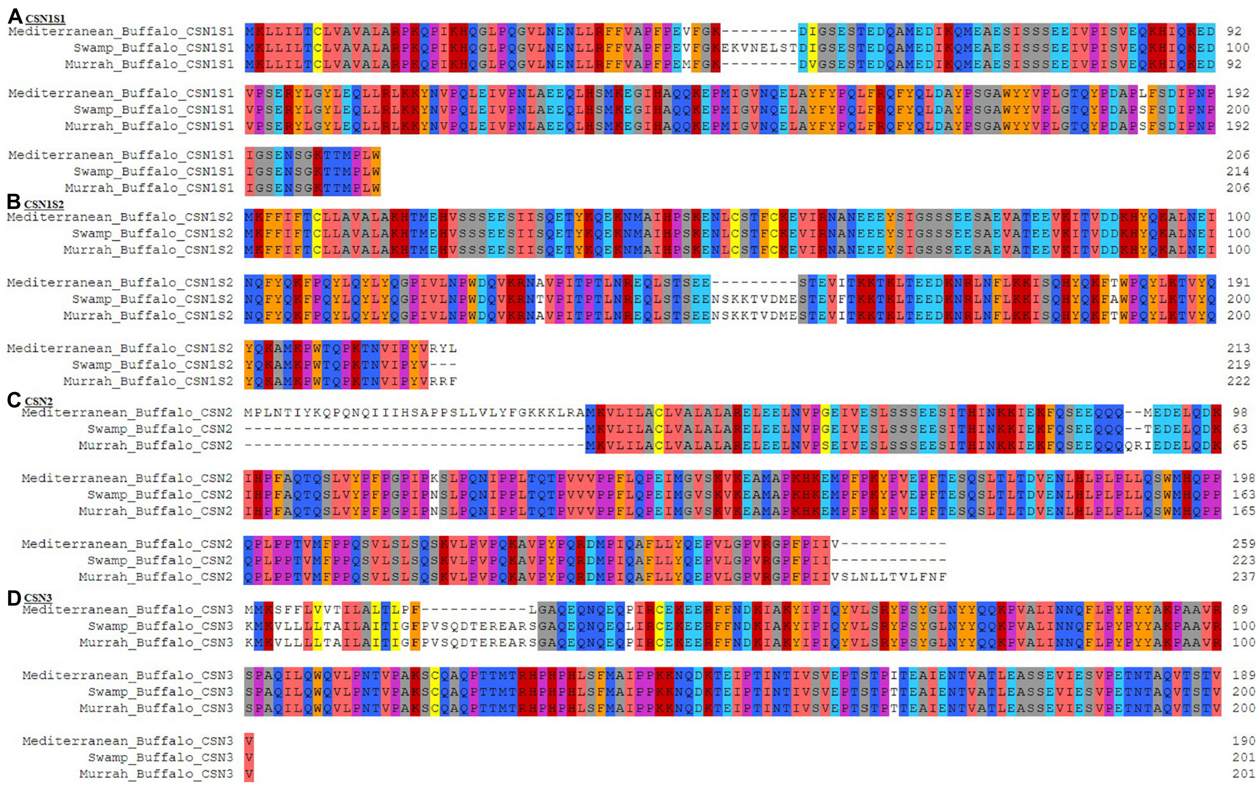

FIGURE 3 | Comparative amino acid analysis of casein gene family in Mediterranean, Swamp, and Murrah buffalo breeds. **(A)** *CSN1S1* gene, **(B)** *CSN1S2* gene, **(C)** *CSN2* gene, **(D)** *CSN3* gene.

swamp buffalo, three amino acid variations A131 > T, I162 > F, and T190 > A were also detected in *CSN1S2* (**Figure 3B**). Furthermore, in *CSN2* two prominent indels toward terminal ends with a length of 35 amino acids (5′end) at 1 > 35 and 12 amino acids (3′end) at 261 > 272, and a short indel of 2 amino acids 91 > 92 was observed. A single amino acid modification was observed in the Mediterranean buffalo (N120 > K) but much variable amino acid in three buffalo breeds was observed at 93 M > T > I (**Figure 3C**). Moreover, a highly variable region toward the 5′ end in *CSN3* was perceived with an indel of 11 amino acids 19 > 29. All single amino acid differences were marked in Mediterranean buffalo except P40 > L which is observed in swamp buffalo (**Figure 3D**).

The genome sequences of Mediterranean and swamp buffalo *CSN* gene family was scanned to find out putative transcription factors binding sites by selecting previously reported four potential transcription sites (GATA, TATA, STAT, and OCT1), and one repressor binding site (YY1) (**Supplementary Tables S3**, **S4**). Both Mediterranean and swamp buffalo shared approximately an equal number of respective transcription sites except the repressor site YY1 that was highly distributed ($P < 0.05$) in the swamp buffalo as compared to the Mediterranean buffalo (**Figure 4** and **Supplementary Table S5**). The distribution of GATA in the Mediterranean was 35, 6, 44, and 15 correspondings to *CSN1S1, CSN2, CSN1S2,* and *CSN3*, respectively, while swamp buffalo had 41, 29, 43, and 8, respectively (**Figure 4** and **Supplementary Table S5**). Furthermore, TATA site distribution in Mediterranean buffalo was 3, 1, 7, and 1 in *CSN1S1, CSN2, CSN1S2,* and *CSN3*, respectively but in swamp buffalo, it was 1, 2, 12, and 3, respectively (**Figure 4** and **Supplementary Table S5**). A considerable difference ($P > 0.05$) was observed in the STAT site's distribution in *CSN1S1* (9 vs. 3), *CSN2* (4 vs. 2), *CSN1S2* (7 vs. 7), and *CSN3* (4 vs. 5) of Mediterranean and swamp buffalo (**Figure 4** and **Supplementary Table S5**). The distribution of OCT1 transcription sites varied across the *CSN1S1* (21 vs. 13), *CSN2* (9 vs. 6), *CSN1S2* (32 vs. 45), and *CSN3* (4 vs. 9) of Mediterranean and swamp buffalo (**Figure 4** and **Supplementary Table S5**).

The pattern of nuclear hormone receptors (NHRs) sites in the *CSN* gene family of *Bubalus bubalis* was explored using genome sequence data of Mediterranean buffalo. A total of 58 NHRs sites were observed in the buffalo *CSN* gene family that was mostly distributed toward 5′end (**Figure 5** and **Supplementary Table S6**). Moreover, the number of NHRs identified in *CSN1S1, CSN1S2, CSN2,* and *CSN3* were 17, 22, 4, and 15, respectively (**Figure 5**). A total of 7 inverted repeats (IR) were observed in different *CSN* genes that are primarily used as the hormonal response element (HRE) important for steroid receptors. Single IR in each of *CSN1S1* and *CSN3*, while 5 IR were observed in *CSN1S2* whereas, *CSN2* harbored no IR (**Figure 5** and **Supplementary Table S6**). In total 22 direct repeats (DR) and

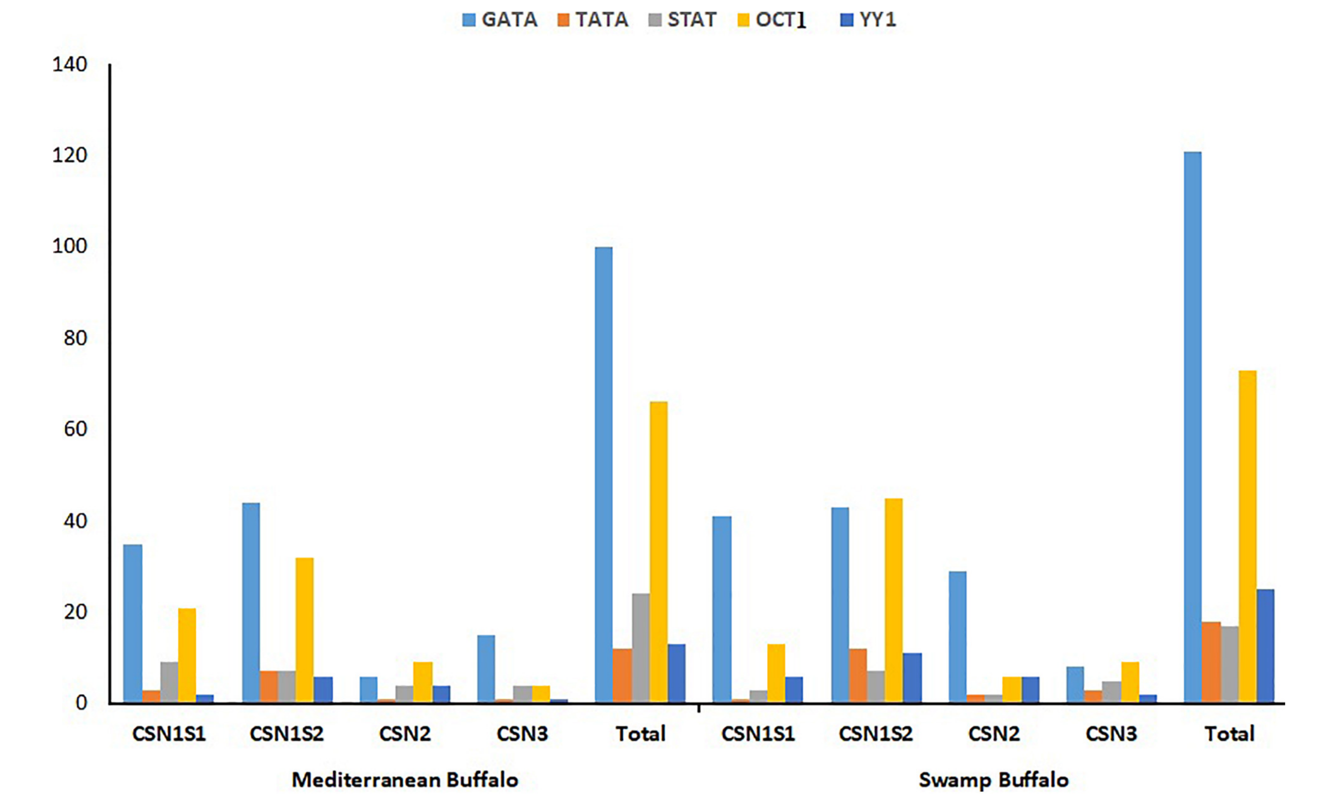

FIGURE 4 | Comparative distribution of putative transcription binding site (GATA, TATA, STAT, and OCT1) and repressor site (YY1) in the genomic sequences of Mediterranean and Swamp buffalo casein gene family.

29 everted repeats (ER) were found in the buffalo *CSN* genes which are prominently used by type II receptors (RXR) and some type III receptors (orphan receptors) can also able to use DR. The number of DR distributed in *CSN1S1, CSN1S2, CSN2,* and *CSN3* was 6, 4, 3, and 9, and ER was 10, 13, 1, and 5, respectively (**Figure 5** and **Supplementary Table S6**). All these HRE were detected close to the putative transcription binding sites (**Supplementary Tables S3**, **S4**, **S6**).

DISCUSSION

The advances in genome sequencing technology particularly next-generation sequencing has led to the availability of sequenced genomes for different animal species that opens up new ways to understand genomic architecture at the molecular level (Luo et al., 2020). Comparative genomics provides an opportunity for discovering novel genes and their functional components (Wei et al., 2002; Rijnkels et al., 2003). Exploring the genetics and evolutionary processes is required to understand the regulatory mechanisms of different physiological important genes like the *CSN* gene family in mammals. Buffalo possesses significant economic attributes owing to its high milk protein contents which are imperative for the production of commercial dairy products like cheese. The milk proteins and related coding genes have been ubiquitously studied due to their extensive distribution in all mammalian species, as an enriched nutrient source for neonates. Caseins (αs1, αs2, β, and κ) are the primary components of milk protein content in dairy animals. All the mammalian *CSN* genes are rapidly evolving genes and are mainly classified into four types including *CSN1S1, CSN2*, *CSN1S2,* and *CSN3* (Madende and Osthoff, 2019). The results of our molecular phylogenetic analysis of the *CSN* gene family are in consensus as all the representative species were clustered into four taxa. The buffalo species were grouped with cattle, *Capra hircus,* and *Ovis aries* sharing higher sequence homology with cattle breeds (**Figure 1**).

The amino acid sequence of protein data can impersonate a better prototype of biologically substantial conserved evolutionary motifs. For protein structural and functional analyses, these conserved regions are vital and can be traced by Multiple sequence alignment (Neuwald, 2016). In reference to the aligned sequence of the *CSN* gene, high variation has been reported in all the *CSN* genes. Even though closely related species represent increased sequence similarity with conserved and non-conserved genomic regions (Madende and Osthoff, 2019). In the present study, sequence analysis of CN protein revealed 10 conserved motifs in buffalo, and cattle using the MEME tool. Apart from the sequence variations in the *CSN* gene, further differences and divergence were observed because of different incidents including exon skipping (Martin and Orgogozo, 2013). Besides, the upstream and downstream UTRs

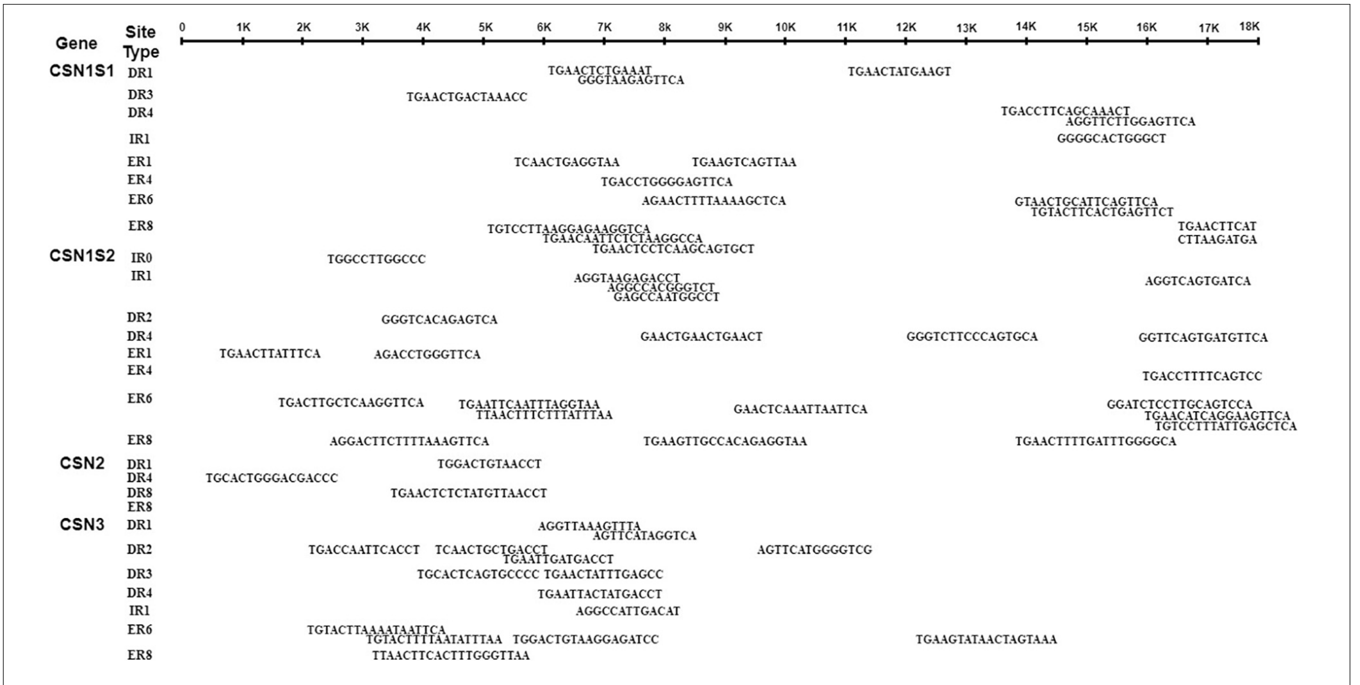

FIGURE 5 | Nuclear hormone receptor sites patterns in the casein gene family of *Bubalus bubalis*.

and introns structure considerably varied, structural analysis of the gene indicated that buffalo *CSN* genes in the same group have a consistent number of exons and introns but variable patterns of exons and introns have also been observed. The variability of UTRs and intronic regions is mostly because of the absence or presence of retroposonic elements. In fact, these ruminants-specific retrotransposons insertions are often polymorphic (absent or present) at orthologous loci and they are highly informative genetic markers that can be considered a powerful phylogenetic tool for clustering studies, animal evolutionary history, population structure, and demography. In general, these elements are known to affect the genome in many other different ways: contributing to the genome size increase, genomic instability, exonization, epigenetic regulation, RNA editing, and so on (Cosenza et al., 2009; Giambra et al., 2010).

All these caseins are encoded by autosomal genes *CSN1S1, CSN2*, *CSN1S2,* and *CSN3*, respectively in closely linked DNA clusters (Pauciullo et al., 2019). The genomic cluster of the casein gene spans between 250 and 350 kb in different mammalian species (Ryskaliyeva, 2018), and in buffalo entire *CSN* gene covers a region of 250kb. This was hypothesized that the exon duplications events in the ancestral gene result in casein gene evolution (Jones et al., 1985). For instance, donkeys, horses, rabbits, and rodents possess an extra copy of αs2-casein indicating the event of recent paralogous gene duplication (Stewart et al., 1987; Ginger et al., 1999). While no evidence for the paralogous gene duplication in buffalo was practically observed that confirms the previous findings of phylogenetic data, which demonstrated Artiodactyla gene loss, whereas gain of an extra copy of the gene in other species was somewhat attained by differential exon usage (Rijnkels, 2002). Caseins are intrinsically disordered proteins (IDPs) related groups of proteins, manifested in milk as roughly spherical, amorphous, polydisperse particles, classically encompassing protein chains, and calcium phosphate nanoclusters. These particles are termed as casein micelles (Cosenza et al., 2010). Caseins have flexible open conformation with an abundance of poly-L-proline II secondary structures and cannot be considered as hydrophobic proteins (Carver and Holt, 2019). Similarly, lower values of GRAVY represent the hydrophilic nature of buffalo CN proteins.

Moreover, short phosphorylated sequences and flexible conformation remarkably increase the casein's ability to keep calcium phosphate nanoclusters and develop a dense shell of peptide around the calcium phosphate to form a thermodynamically stable core-shell complex, even at quite higher phosphate and calcium concentrations (Carver and Holt, 2019). In the present study, the aliphatic index showed that all CN proteins have values >65 so perused as thermostable. The casein micelles formation is essential for the effective transportation of phosphate and calcium via milk from the mother to the neonate (Holt et al., 2013). Thus, a readily digestible calcium-enriched diet in the form of casein micelle is available for the neonate. Caseins as IDPs play an important role in mammary gland protection against pathological calcification, amyloid formation, and other dysfunctional processes that can minimize the reproductive success of the mother (Carver and Holt, 2019). Our findings illustrated all the CN peptides in buffalo were determined as unstable protein and the pI revealed all casein proteins α s1-, β-, and κ-CN were determined as acidic peptides except αs2 which behaved slightly basic nature.

In recent years, the polymorphisms of milk proteins have aroused great research interest because of the genotypes of milk

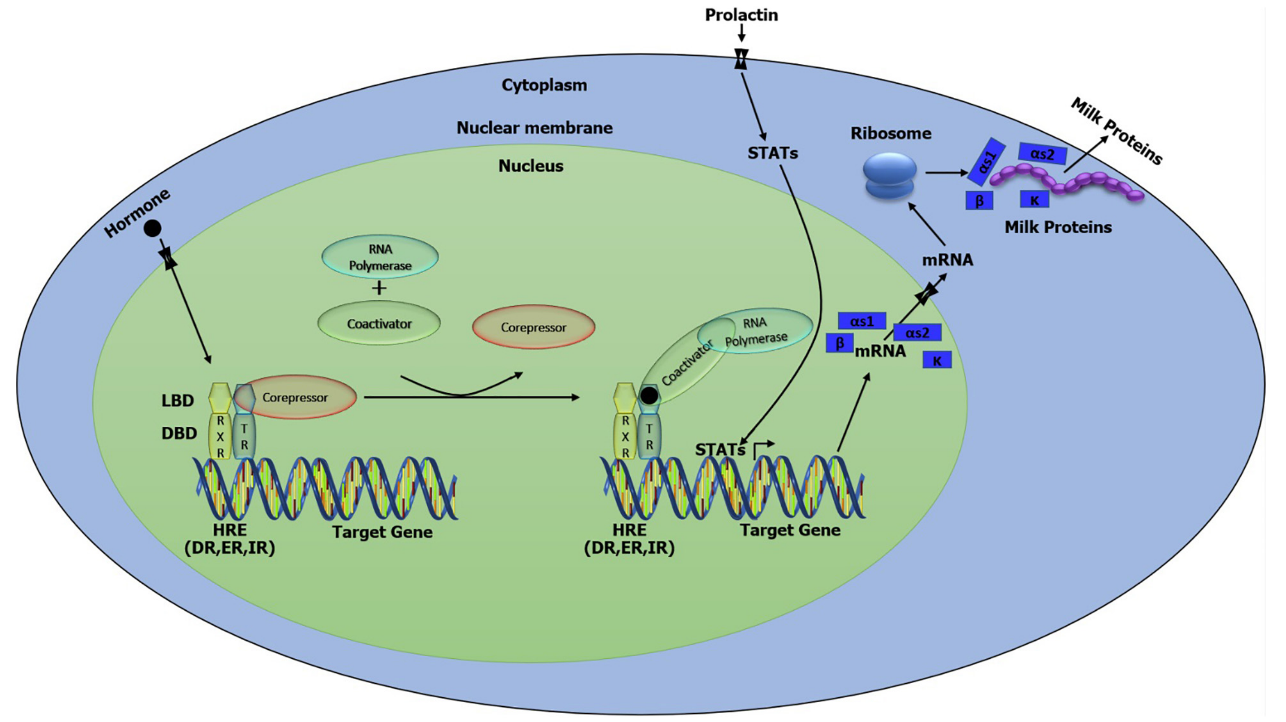

FIGURE 6 | Putative model of the network involved in the regulation and biosynthesis of casein protein in buffalo mammary tissue.

proteins may be related to milk composition and milk yield of dairy mammals (Nilsen et al., 2009). The amino acid changes possibly have a functional effect on the buffalo caseins (Fan et al., 2020). Comparative amino acid sequence analysis revealed that CN protein harbor higher amino acid variations in river buffalo (Mediterranean and Murrah) as compared to the swamp buffalo. The results of the present study are in line with previous studies (Masina et al., 2007; Azevedo et al., 2008; Massella et al., 2017; Rangel et al., 2017; Miluchova et al., 2018; Fan et al., 2020; de Oliveira et al., 2021) which reported the potential association of genetic variants in *CSN* genes with lactation performance, milk composition, and attributes of milk products. Thus, casein gene-based markers are important candidates for the selective breeding of buffalo to improve the quantity and quality of milk (de Oliveira et al., 2021). Moreover, further insights are required to ubiquitously apply these candidate markers to other mammals due to genetic variability and locus distribution (Sulimova et al., 2007; Cosenza et al., 2015).

Nevertheless, understanding the molecular basis for the regulation of *CSN* gene expression is very crucial for improving milk production (Debeljak et al., 2005). Sequence analysis of promotor region of *CSN* genes has shown various transcription factor binding sites including transcription initiation sites such as STAT5, NF1 and GR, C/EBP (Hennighausen and Robinson, 1998; Robinson et al., 1998; Rosen et al., 1999; Wheeler et al., 2001; Wyszomierski and Rosen, 2001; Yamashita et al., 2001; Chughtai et al., 2002) and potential repressors sites such as YY1, CIS3, SOCS-1, and SOCS-3 (Helman et al., 1998; Tomic et al., 1999). The identification of critical regulatory regions

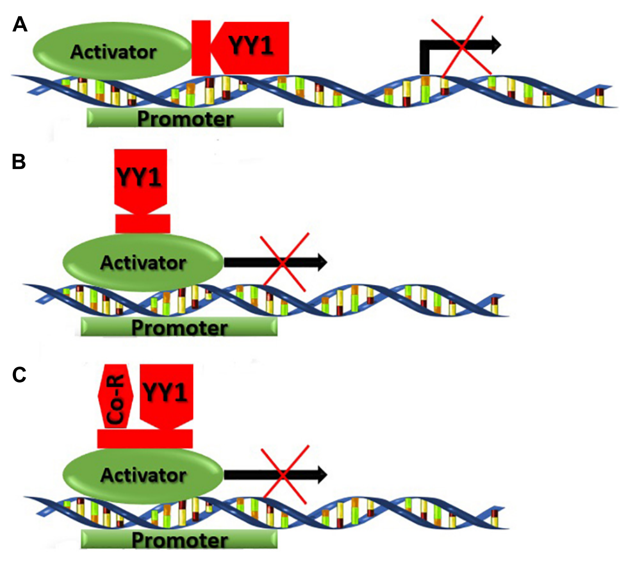

FIGURE 7 | Different transcriptional suppression mechanisms used by YY1.

responsible for the expression of the *CSN* genes provides valuable information for the selection of markers in dairy mammals especially the buffaloes. So, in both Mediterranean and swamp buffalo, we selected four potential transcription sites (GATA, TATA, STAT, and OCT1) and one repressor binding site (YY1), for comparative genomic analysis (**Figure 4**). OCT1 affects the factors of acute myeloid leukemia (AML), forming a complex that reduces its inhibitory role in DNA binding and promotes the expression of the casein gene (Inman et al., 2005).

Various lactogenic hormones like prolactin, insulin, hydrocortisone, and some growth factors such as insulin-like growth factor 1 (*IGF-1*) and epidermal growth factor (*EGF*) are crucial for mammary gland activation and eventually the milk proteins gene expression regulation (Hennighausen and Robinson, 1998; Tsunoda and Takagi, 1999). Therefore, we further analyzed the distribution of HRE including DR, ER, and IR in the genome of the buffalo. All these HRE were detected close to the putative transcription binding sites. Therefore, the combined action of the transcription factor and HRE can mediate the activation of caseins (**Figure 6**). STAT5 is the principal transcription factor in milk protein gene expression that could be activated by the action of growth hormone (*GH*) and prolactin (*PR*) via the STAT/JAK2 signaling pathway or Src-kinase/STAT signaling pathway through the *EGF* action (Gallego et al., 2001). Dimerization and phosphorylation activate the STAT5 and translocate it to the nucleus where STAT5 dimers bind with the DNA and induce transcription (**Figure 6**; Gallego et al., 2001).

Multiple mechanisms are being used by YY1 for transcriptional suppression. Mostly YY1 competes with activator factors and overlaps the binding site ultimately repressing the gene transcription. In mammary epithelial cells, YY1 competes with a β-CN activating promoter also known as mammary gland factor (MGF), fallouts in transcription repression (**Figure 7A**). Moreover, the c-fos promoter possesses two extra YY1 sites between the TATA box and calcium or cyclic AMP response element (CRE) in addition to YY1 overlapping sites (Gordon et al., 2006). The YY1 binding remotely caused direct suppression of the upstream CRE promoter. YY1 can repress the c-fos promoter in a site-dependent or independent manner, including the interaction of zinc finger patterns or binding with cAMP response element-binding (CREB) at the basic leucine zipper region (bZIP) in YY1 (**Figure 7B**). Most likely, the YY1 and CREB interact in the nucleus and inhibit transcription (Gordon et al., 2006). The YY1 can recruit corepressors that directly induce transcriptional repression or facilitate chromatin condensing to assist further YY1 mediated repression. The repression activity of YY1 is generally because of its glycine-rich and zinc-finger regions. Simultaneous deletions in each individual or both regions reduce the GAL4-YY1 fusion proteins deficient for transcriptional repression (**Figure 7C**). Thus, cofactors interactions are often required with repression domains of YY1 to facilitate repression like mRPD3 (Yang et al., 1996) or Smad family members (Kurisaki et al., 2003). A considerably higher ratio of STATs distribution and lower number of repressor binding site YY1 was observed in Mediterranean buffalo as compared to swamp buffalo. This envisages that lower STATs and higher YY1 site distribution in swamp buffalo might lead to a lower expression of *CSN* gene subsequently leading to poor milk yield in swamp buffalo.

Our study provides inclusive insights into the regulation of the casein gene family revealing a plausible association of STATs and YYI distribution with a poor milk production potential of swamp buffalo. Moreover, we report striking findings regarding genetic variations in transcription activators and repressor elements from evolutionary standpoint. Further investigations are required to confirm these findings to elucidate the putative role of STATs and repressor sites in the regulation of *CSN* gene expression and their potential utility for the genomic selection of buffaloes for effective utilization and enhanced production.

CONCLUSION

The present study provides a comprehensive insight into the molecular structure and function of the casein gene family in buffalo. Phylogenetic, gene structure, motif, and conserved domain analysis elucidated the evolutionary conserved nature of the casein gene in buffalo and closely related species. Buffalo casein proteins were observed as unstable, hydrophilic, and thermostable. The α s1-, β-, and κ-CN behaved as acidic peptides except for αs2, which was slightly basic. Comparative genomic analysis revealed higher amino acid variations in the river buffalo (Mediterranean and Murrah breeds) than swamp buffalo, revealing that these variations may influence milk production traits in buffalo. Moreover, for the first time, our findings indicate lower STATs and higher YY1 site distribution in swamp buffalo as a plausible reason for the comparatively lower expression of casein genes that ultimately affect milk production.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

FH and QL: conceptualization. SR, TF, and QL: resources. FH, BL, TF, and XL: data curation. SR, TF, and XL: methodology and software. QL and AL: supervision. SR: writing—original draft preparation. FH, XL, SW, TF, AL, BL, and QL: writing—review and editing. All authors have read and agreed to the published version of the manuscript.

FUNDING

This study was granted and supported by the National Natural Science Fund (U20A2051, 31760648, and 31860638), Guangxi Natural Science Foundation (AB18221120), and Guangxi Distinguished Scholars Program (201835).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.662609/full#supplementary-material

REFERENCES

Ahmad, S., Anjum, F. M., Huma, N., Sameen, A., and Zahoor, T. (2013). Composition and physicochemical characteristics of buffalo milk with particular emphasis on lipids, proteins, minerals, enzymes and vitamins. *J. Anim. Plant Sci.* 23, 62–74.

Azevedo, A. L., Nascimento, C. S., Steinberg, R. S., Carvalho, M. R., Peixoto, M. G., Teodoro, R. L., et al. (2008). Genetic polymorphism of the kappa-casein gene in Brazilian cattle. *Genet. Mol. Res.* 7, 623–630. doi: 10.4238/vol7-3gmr428

Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34, W369–W373.

Basilicata, M. G., Pepe, G., Sommella, E., Ostacolo, C., Manfra, M., Sosto, G., et al. (2018). Peptidome profiles and bioactivity elucidation of buffalo-milk dairy products after gastrointestinal digestion. *Food Res. Int.* 105, 1003–1010.

Carver, J. A., and Holt, C. (2019). "Functional and dysfunctional folding, association and aggregation of caseins," in *Advances in Protein Chemistry and Structural Biology*, Vol. 118, ed. R. Donev (Amsterdam: Academic Press), 163–216. doi: 10.1016/bs.apcsb.2019.09.002

Chughtai, N., Schimchowitsch, S., Lebrun, J. J., and Ali, S. (2002). Prolactin induces SHP-2 association with Stat5, nuclear translocation, and binding to the β-casein gene promoter in mammary cells. *J. Biol. Chem.* 277, 31107–31114. doi: 10.1074/jbc.m200156200

Cosenza, G., Pauciullo, A., Annunziata, A. L., Rando, A., Chianese, L., Marletta, D., et al. (2010). Identification and characterization of the donkey CSN1S2 I and II cDNAs. *Ital. J. Anim. Sci.* 9:e40.

Cosenza, G., Pauciullo, A., Feligini, M., Coletta, A., Colimoro, L., Di Berardino, D., et al. (2009). A point mutation in the splice donor site of intron 7 in the as2-casein encoding gene of the Mediterranean River buffalo results in an allele-specific exon skipping. *Anim. Genet.* 40:791. doi: 10.1111/j.1365-2052.2009.01897.x

Cosenza, G., Pauciullo, A., Macciotta, N. P. P., Apicella, E., Steri, R., La Battaglia, A., et al. (2015). Mediterranean river buffalo CSN1S1 gene: search for polymorphisms and association studies. *Anim. Prod. Sci.* 55, 654–660. doi: 10.1071/an13438

de Oliveira, L. S. M., Alves, J. S., Bastos, M. S., da Cruz, V. A. R., Pinto, L. F. B., Tonhati, H., et al. (2021). Water buffaloes (*Bubalus bubalis*) only have A2A2 genotype for beta-casein. *Trop. Anim. Health Prod.* 53:145.

Debeljak, M. A., Frajman, P. O., Lenasi, T. I., Narat, M. O., Baldi, A. N., and Dovc, P. E. (2005). Functional analysis of the bovine beta-and kappa casein gene promoters using homologous mammary gland derived cell line. *Arch. Anim. Breed.* 48, 334–345. doi: 10.5194/aab-48-334-2005

Fan, X., Gao, S., Fu, L., Qiu, L., and Miao, Y. (2020). Polymorphism and molecular characteristics of the CSN1S2 gene in river and swamp buffalo. *Arch. Anim. Breed.* 63, 345–354. doi: 10.5194/aab-63-345-2020

Gallego, M. I., Binart, N., Robinson, G. W., Okagaki, R., Coschigano, K. T., Perry, J., et al. (2001). Prolactin, growth hormone, and epidermal growth factor activate Stat5 in different compartments of mammary tissue and exert different and overlapping developmental effects. *Dev. Biol.* 229, 163–175. doi: 10.1006/dbio.2000.9961

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31, 3784–3788. doi: 10.1093/nar/gkg563

Giambra, I. J., Chianese, L., Ferranti, P., and Erhardt, G. (2010). Short communication:molecular genetic characterization of ovine αS1-casein allele H caused by alternative splicing. *J. Dairy Sci.* 93, 792–795. doi: 10.3168/jds.2009-2615

Ginger, M. R., Piotte, C. P., Otter, D. E., and Grigor, M. R. (1999). Identification, characterisation and cDNA cloning of two caseins from the common brushtail possum (*Trichosurus vulpecula*). *Biochim. Biophys. Acta Gen. Subj.* 1427, 92–104. doi: 10.1016/s0304-4165(99)00008-2

Gordon, S., Akopyan, G., Garban, H., and Bonavida, B. (2006). Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene* 25, 1125–1142. doi: 10.1038/sj.onc.1209080

Helman, D., Sandowski, Y., Cohen, Y., Matsumoto, A., Yoshimura, A., Merchav, S., et al. (1998). Cytokine-inducible SH2 protein (CIS3) and JAK2 binding protein (JAB) abolish prolactin receptor-mediated STAT5 signaling. *FEBS Lett.* 441, 287–291. doi: 10.1016/s0014-5793(98)01555-5

Hennighausen, L., and Robinson, G. W. (1998). Think globally, act locally: the making of a mouse mammary gland. *Genes Dev.* 12, 449–455. doi: 10.1101/gad.12.4.449

Holt, C., Carver, J. A., Ecroyd, H., and Thorn, D. C. (2013). Invited review: caseins and the casein micelle: their biological functions, structures, and behavior in foods. *J. Dairy Sci.* 96, 6127–6146. doi: 10.3168/jds.2013-6831

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene features visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

Inman, C. K., Li, N., and Shore, P. (2005). Oct-1 counteracts autoinhibition of Runx2 DNA binding to form a novel Runx2/Oct-1 complex on the promoter of the mammary gland-specific gene β-casein. *Mol. Cell. Biol.* 25, 3182–3193. doi: 10.1128/mcb.25.8.3182-3193.2005

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 8, 275–282. doi: 10.1093/bioinformatics/8.3.275

Jones, W. K., Yu-Lee, L. Y., Clift, S. M., Brown, T. L., and Rosen, J. M. (1985). The rat casein multigene family. fine structure and evolution of the beta-casein gene. *J. Biol. Chem.* 260, 7042–7050. doi: 10.1016/s0021-9258(18)88885-8

Kawasaki, K., Lafont, A. G., and Sire, J. Y. (2011). The evolution of milk casein genes from tooth genes before the origin of mammals. *Mol. Biol. Evol.* 28, 2053–2061. doi: 10.1093/molbev/msr020

Knudsen, S. (1999). Promoter2. 0: for the recognition of PolII promoter sequences. *Bioinformatics* 15, 356–361. doi: 10.1093/bioinformatics/15.5.356

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Kurisaki, K., Kurisaki, A., Valcourt, U., Terentiev, A. A., Pardali, K., Ten Dijke, P., et al. (2003). Nuclear factor YY1 inhibits transforming growth factor β-and bone morphogenetic protein-induced cell differentiation. *Mol. Cell. Biology.* 23, 4494–4510. doi: 10.1128/mcb.23.13.4494-4510.2003

Li, Z., Lu, S., Cui, K., Shafique, L., Rehman, S., Luo, C., et al. (2020). Fatty acid biosynthesis and transcriptional regulation of Stearoyl-CoA Desaturase 1 (SCD1) in buffalo milk. *BMC Genet.* 21:23. doi: 10.1186/s12863-020-0829-6

Lu, X. R., Duan, A. Q., Li, W. Q., Abdel-Shafy, H., Rushdi, H. E., Liang, S. S., et al. (2020). Genome-wide analysis reveals genetic diversity, linkage disequilibrium, and selection for milk production traits in Chinese buffalo breeds. *J. Dairy Sci.* 103, 4545–4556. doi: 10.3168/jds.2019-17364

Luo, X., Zhou, Y., Zhang, B., Zhang, Y., Wang, X., Feng, T., et al. (2020). Understanding divergent domestication traits from the whole-genome sequencing of swamp-and river-buffalo populations. *Natl. Sci. Rev.* 7, 686–701. doi: 10.1093/nsr/nwaa024

Madende, M., and Osthoff, G. (2019). Comparative genomics of casein genes. *J. Dairy Res.* 86, 323–330. doi: 10.1017/s0022029919000414

Martin, A., and Orgogozo, V. (2013). The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* 67, 1235–1250.

Masina, P., Rando, A., Di Gregorio, P., Cosenza, G., and Mancusi, A. (2007). Water buffalo kappa-casein gene sequence. *Ital. J. Anim. Sci.* 6(Suppl. 2), 353–355.

Massella, E., Piva, S., Giacometti, F., Liuzzo, G., Zambrini, A. V., and Serraino, A. (2017). Evaluation of bovine beta casein polymorphism in two dairy farms located in northern Italy. *Ital. J. Food Saf.* 6:6904.

Miluchova, M., Gábor, M., Candrák, J., Trakovická, A., and Candráková, K. (2018). Association of HindIII-polymorphism in kappa-casein gene with milk, fat and protein yield in holstein cattle. *Acta Biochim. Pol.* 65, 403–407.

Moioli, B., Georgoudis, A., Napolitano, F., Catillo, G., Giubilei, E., Ligda, C., et al. (2001). Genetic diversity between Italian, Greek and Egyptian buffalo populations. *Livest. Prod. Sci.* 70, 203–211. doi: 10.1016/s0301-6226(01)00175-0

Neuwald, A. F. (2016). Gleaning structural and functional information from correlations in protein multiple sequence alignments. *Curr. Opin. Struct. Biol.* 38, 1–8. doi: 10.1016/j.sbi.2016.04.006

Nilsen, H., Olsen, H. G., Hayes, B., Sehested, E., Svendsen, M., Nome, T., et al. (2009). Casein haplotypes and their association with milk production traits in Norwegian Red cattle. *Genet. Sel. Evol.* 41:24. doi: 10.1186/1297-9686-41-24

Pauciullo, A., and Erhardt, G. (2015). Molecular characterization of the llamas (*Lama glama*) casein cluster genes transcripts (CSN1S1, CSN2, CSN1S2, CSN3) and regulatory regions. *PLoS One* 10:e0124963. doi: 10.1371/journal.pone.0124963

Pauciullo, A., Shuiep, E. T., Ogah, D. M., Cosenza, G., Di Stasio, L., and Erhardt, G. (2019). Casein gene cluster in camelids: comparative genome analysis and new findings on haplotype variability and physical mapping. *Front. Genet.* 10:748. doi: 10.3389/fgene.2019.00748

Rangel, A. H., Zaros, L. G., Lima, T. C., Borba, L. H., Novaes, L. P., Mota, L. F., et al. (2017). Polymorphism in the beta casein gene and analysis of milk characteristicsin Gir and GuzerÃ¡ dairy cattle. *Genet. Mol. Res.* 16:gmr16029592. doi: 10.4238/gmr16029592

Rehman, S., Nadeem, A., Javed, M., Hassan, F., Luo, X., Khalid, R. B., et al. (2020). Genomic identification, evolution and sequence analysis of the heat-shock protein gene family in buffalo. *Genes* 11:1388. doi: 10.3390/genes11111388

Rehman, S., Shafique, L., Yousuf, M. R., Liu, Q., Ahmed, J. Z., and Riaz, H. (2019). Spectrophotometric calibration and comparison of different semen evaluation methods in Nili-Ravi buffalo bulls. *Pak. Vet. J.* 39, 568–572. doi: 10.29261/pakvetj/2019.073

Rijnkels, M. (2002). Multispecies comparison of the casein gene loci and evolution of casein gene family. *J. Mammary Gland Biol. Neoplasia* 7, 327–345.

Rijnkels, M., Elnitski, L., Miller, W., and Rosen, J. M. (2003). Multispecies comparative analysis of a mammalian-specific genomic domain encoding secretory proteins. *Genomics* 82, 417–432. doi: 10.1016/s0888-7543(03)00114-9

Robinson, G. W., Johnson, P. F., Hennighausen, L., and Sterneck, E. (1998). The C/EBPβ transcription factor regulates epithelial cell proliferation and differentiation in the mammary gland. *Genes Dev.* 12, 1907–1916. doi: 10.1101/gad.12.12.1907

Rosen, J. M., Wyszomierski, S. L., and Hadsell, D. (1999). Regulation of milk protein gene expression. *Annu. Rev. Nutr.* 19, 407–436.

Ryskaliyeva, A. (2018). *Exploring the Fine Composition of Camelus Milk from Kazakhstan with Emphasis on Protein Components*. Doctoral dissertation, Université Paris Saclay, Paris.

Stewart, A. F., Bonsing, J., Beattie, C. W., Shah, F., Willis, I. M., and Mackinlay, A. G. (1987). Complete nucleotide sequences of bovine alpha S2-and beta-casein cDNAs: comparisons with related sequences in other species. *Mol. Biol. Evol.* 4, 231–241.

Sulimova, G. E., Azari, M. A., Rostamzadeh, J., Abadi, M. M., and Lazebny, O. E. (2007). κ-casein gene (CSN3) allelic polymorphism in Russian cattle breeds and its information value as a genetic marker. *Russ. J. Genet.* 43, 73–79. doi: 10.1134/s1022795407010115

Tomic, S., Chughtai, N., and Ali, S. (1999). SOCS-1,-2,-3: selective targets and functions downstream of the prolactin receptor. *Mol. Cell. Endocrinol.* 158, 45–54. doi: 10.1016/s0303-7207(99)00180-x

Tsunoda, T., and Takagi, T. (1999). Estimating transcription factor bindability on DNA. *Bioinformatics* 15, 622–630. doi: 10.1093/bioinformatics/15.7.622

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293

Wei, L., Liu, Y., Dubchak, I., Shon, J., and Park, J. (2002). Comparative genomics approaches to study organism similarities and differences. *J. Biomed. Inform.* 35, 142–150. doi: 10.1016/s1532-0464(02)00506-3

Wheeler, T. T., Broadhurst, M. K., Sadowski, H. B., Farr, V. C., and Prosser, C. G. (2001). Stat5 phosphorylation status and DNA-binding activity in the bovine and murine mammary glands. *Mol. Cell. Endocrinol.* 176, 39–48. doi: 10.1016/s0303-7207(01)00481-6

Wyszomierski, S. L., and Rosen, J. M. (2001). Cooperative effects of STAT5 (signal transducer and activator of transcription 5) and C/EBP β (CCAAT/enhancer-binding protein-β) onβ-casein gene transcription are mediated by the glucocorticoid receptor. *Mol. Endocrinol.* 15, 228–240. doi: 10.1210/me.15.2.228

Yamashita, H., Nevalainen, M. T., Xu, J., LeBaron, M. J., Wagner, K. U., Erwin, R. A., et al. (2001). Role of serine phosphorylation of Stat5a in prolactin-stimulated β-casein gene expression. *Mol. Cell. Endocrinol.* 183, 151–163. doi: 10.1016/s0303-7207(01)00546-9

Yang, W. M., Inouye, C., Zeng, Y., Bearss, D., and Seto, E. (1996). Transcriptional repression by YY1 is mediated by interaction with a mammalian homolog of the yeast global regulator RPD3. *Proc. Natl Acad. Sci. U.S.A.* 93, 12845–12850. doi: 10.1073/pnas.93.23.12845

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

REVIEW
published: 23 March 2021
doi: 10.3389/fgene.2021.617128

Genetic Features of Reproductive Traits in Bovine and Buffalo: Lessons From Bovine to Buffalo

Baoshun Shao[1†], Hui Sun[1†], Muhammad Jamil Ahmad[1†], Nasser Ghanem[2], Hamdy Abdel-Shafy[2], Chao Du[1], Tingxian Deng[1,3], Shahid Mansoor[4], Yang Zhou[1,5,6], Yifen Yang[1,2], Shujun Zhang[1,5,6], Liguo Yang[1,5,6] and Guohua Hua[1,5,6]*

[1] Key Lab of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, College of Animal Science and Technology, Huazhong Agricultural University, Wuhan, China, [2] Department of Animal Production, Faculty of Agriculture, Cairo University, Giza, Egypt, [3] Guangxi Provincial Key Laboratory of Buffalo Genetics, Breeding and Reproduction Technology, Buffalo Research Institute, Chinese Academy of Agricultural Sciences, Nanning, China, [4] National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad, Pakistan, [5] International Joint Research Centre for Animal Genetics, Breeding and Reproduction, Wuhan, China, [6] Hubei Province's Engineering Research Center in Buffalo Breeding and Products, Wuhan, China

OPEN ACCESS

Edited by:
Fabyano Fonseca Silva,
Universidade Federal de Viçosa, Brazil

Reviewed by:
Sirlene Fernandes Lazaro,
Purdue University, United States
Mahdi Mokhber,
Urmia University, Iran

****Correspondence:***
Guohua Hua
huaguohua09@gmail.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *14 October 2020*
Accepted: *25 February 2021*
Published: *23 March 2021*

Citation:
Shao B, Sun H, Ahmad MJ, Ghanem N, Abdel-Shafy H, Du C, Deng T, Mansoor S, Zhou Y, Yang Y, Zhang S, Yang L and Hua G (2021) Genetic Features of Reproductive Traits in Bovine and Buffalo: Lessons From Bovine to Buffalo. Front. Genet. 12:617128. doi: 10.3389/fgene.2021.617128

Bovine and buffalo are important livestock species that have contributed to human lives for more than 1000 years. Improving fertility is very important to reduce the cost of production. In the current review, we classified reproductive traits into three categories: ovulation, breeding, and calving related traits. We systematically summarized the heritability estimates, molecular markers, and genomic selection (GS) for reproductive traits of bovine and buffalo. This review aimed to compile the heritability and genome-wide association studies (GWASs) related to reproductive traits in both bovine and buffalos and tried to highlight the possible disciplines which should benefit buffalo breeding. The estimates of heritability of reproductive traits ranged were from 0 to 0.57 and there were wide differences between the populations. For some specific traits, such as age of puberty (AOP) and calving difficulty (CD), the majority beef population presents relatively higher heritability than dairy cattle. Compared to bovine, genetic studies for buffalo reproductive traits are limited for age at first calving and calving interval traits. Several quantitative trait loci (QTLs), candidate genes, and SNPs associated with bovine reproductive traits were screened and identified by candidate gene methods and/or GWASs. The IGF1 and LEP pathways in addition to non-coding RNAs are highlighted due to their crucial relevance with reproductive traits. The distribution of QTLs related to various traits showed a great differences. Few GWAS have been performed so far on buffalo age at first calving, calving interval, and days open traits. In addition, we summarized the GS studies on bovine and buffalo reproductive traits and compared the accuracy between different reports. Taken together, GWAS and candidate gene approaches can help to understand the molecular genetic mechanisms of complex traits. Recently, GS has been used extensively and can be performed on multiple traits to improve the accuracy of prediction even for traits with low heritability, and can be combined with multi-omics for further analysis.

Keywords: reproduction, breeding, genetic improvement, heritability, GWAS

INTRODUCTION

Reproductive traits are economically important for sustainable food production, especially for monotocous livestock, such as cattle and buffalo. Low reproductive capacity or infertility can be described as an extended duration between two calvings. This problem requires additional inseminations, more veterinary attention, and hormonal treatments, which consequently alters the current and subsequent lactations (Boichard, 1990). Also, additional costs are raised due to culling and replacing animals with fertility problems (Roxström and Strandberg, 2002). Enhancing fertility is the best choice not only to reduce the culling cost but also to save important genetic materials and increase farm profit (Dekkers, 1991). Several countries have included female reproductive traits in the breeding goals to emphasize the genetic aspects of reducing fertility costs (FCOST) in dairy cattle (Kadarmideen and Simm, 2002). Herein, we emphasize the recent literature about genetic parameters, genome-wide association study (GWAS), and genomic selection (GS) for reproductive traits in cattle and buffalo over the past 20 years for researchers, who can integrate these traits in cattle and buffalo breeding programs and achieve optimum fertility.

In the previous study, reproductive traits were divided into binary, interval, and continuous traits with respect to statistical distribution (Berry and Evans, 2014). To better understand and utilize reproductive traits in livestock and breeding programs, they are reclassified as ovulation, mating, and calving-related traits from the physiological viewpoint (Cammack et al., 2009; **Table 1**).

HERITABILITY ESTIMATES OF REPRODUCTIVE TRAITS

Genetic variation, which is a variability in breeding values within a population for a trait under selection, significantly affects the accuracy of genetic selection. Heritability measures how much of the phenotypic variation is attributed to genetic variation, and affects the rate of genetic improvement for a trait over generations. Over the past 20 years, several studies were conducted to estimate the heritability of different reproductive traits in dairy cattle (**Table 2**), beef cattle (**Table 3**), and buffalo cows (**Table 4**).

In dairy cattle, all ovulation-related traits range from low to moderate heritabilities (**Table 2**). The heritability estimate of the superovulation response was about 0.15 in Holstein cows (Jaton et al., 2020). Regarding mating-related traits, heritability estimates for age of puberty (AOP) and age at first calving (AFC) were moderate in most cattle populations, except for AFC in the Chile population ($h^2 = 0.01$) (Montaldo et al., 2017). Likewise, the heritabilities of non-return rate (NRR) and pregnancy rate (PR) of Holstein dairy cows and Brown Swiss cattle were low (Gaddis et al., 2016; Tiezzi et al., 2018; Ansari-Mahyari et al., 2019; Zhang et al., 2019). Regarding the superovulation response and twinning rate, heritability was higher for superovulation, indicating a response to hormone treatment is more heritable than natural ovulation in dairy cows. Non-return and PR are directly related to reproductive outcomes. Unfortunately, the heritability estimates for these two traits were remarkably low. Besides, dairy cows' calving-related traits, including calving interval, days open, calving difficulty (CD), and the length of the productive life, were all of low heritability. Therefore, management practices (reproductive management, balanced nutrition, etc.) and/or environmental factors are of significant importance for improving reproductive efficiency and preventing reproductive disorders in dairy cows. Thus, selection on dairy cows' AOP, first calving, and superovulation response may gain more progression than other traits.

In beef cattle, the superovulation response had higher heritability than those of ovulation rate, and twinning rate was similar to those reported in dairy cattle (**Table 3**). Regarding mating-related traits, AOP had moderate to high heritability estimates in most beef populations; for example, the estimate reached 0.78 in the Brahman bull population (Fortes et al., 2012). The h^2 for scrotal circumference was also reported to have moderate to high estimates. Excluding the Angus population (0.2) (Doyle et al., 2000) in beef cattle, the NRR and PR of heritability were low, as reported in dairy cattle. The heritabilities for calving difficulties in beef cattle had moderate to high estimates, unlike those reported in dairy cattle with low heritabilities. In comparison, other mating-related reproductive traits, such as DO, NRR, CI, and length of productive life had low heritabilities similar to dairy cattle. Taken together, selections on beef cow's AOP, calving difficulties, DO, NRR, and CI traits may gain more progression due to the moderate to high estimates of heritabilities compared with other traits (Cassell, 2009).

The excellent milk quality and the limitation of buffalo milk yield contribute to the breeding selection focusing more on milk production traits in buffalo compared with reproductive traits. Currently, there are limited studies for estimating genetic parameters for reproductive traits in buffalo species, mainly for AFC and CI (**Table 4**). The heritability estimates of AFC in the buffalo population is close to Holstein cattle (Gupta et al., 2015; Kumar et al., 2015; Barros et al., 2016; Rathod et al., 2018). Most studies showed that the heritability of CI is low, mostly below 0.1 (Morammazi et al., 2007; Thiruvenkadan et al., 2010; El-Bramony and Reclamation, 2014; Camargo et al., 2015). However, the highest record for CI was 0.55 in Surti buffalo, which may be due to the limited numbers of lactation records and/or number of parities per sire monitored (Rathod et al., 2018). The heritabilities of DO (Camargo et al., 2015) and CD (Al-Khuzai et al., 2019) were similar to those reported in dairy cattle.

Comparing heritabilities between different traits in dairy and beef cattle along with buffalo, we found that:

(1) Most of the reproductive traits had low habitabilities, but not all. In the dairy and beef cattle, AOP showed high heritabilities. The heritability estimates for scrotal circumference of the beef bull were medium to high. Also, the superovulation response in dairy and beef cattle was worthy of notice. These moderate to high heritability traits could be applied to the selection and breeding system.

TABLE 1 | Physiological classification and description of reproductive traits.

Trait category	Parameter	Description
Ovulation	Ovulation rate	Corpus luteum (CL) number during mid-luteal phase of the estrous cycle
	Superovulation response	The biological potentiality of the cow in terms of total number of ova (TNO), transferable embryos (NTE), unfertilized ova (NUO) and degenerated embryos (NDE); total number of embryos (NE) and number of viable embryos (VE)
	Twinning rate	The proportion of cows giving birth to two or more calves in one pregnancy
Mating	Age of puberty (AOP)	Male: the age when a bull scrotal circumference reaches 26–29 cm (AGESC)*, or the age at which a bull first produces an ejaculate containing at least 50 million sperm with a minimum of 10% motility Female: the appearance of the first corpus luteum (AGECL), age at first behavioral estrus (AFO) or standardized age at first behavioral estrus (SFO) and plasma progesterone concentration
	Age at first calving (AFC)	The interval between the date of first calving and the date of birth of the cow
	Non-return rate (NRR)	The proportion of cows that are not subsequently rebred
	pregnancy rate (PR)	The percentage of cows to become pregnant
Calving	Calving interval (CI)	The period of time (days or months) between the birth of a calf and the birth of a subsequent calf, both from the same cow
	Days open (DO)	The period between calving and conception
	Calving difficulty (CD)	Dystocia, which is categorized into three degrees, including easy calving, slight problems, and difficult calving
	Length of production life (LPL)	Mainly focused on dairy cattle, length of service, tenure, etc. Such as fertility-/mastitis-/production-/determined PL (FPL/MPL/PPL)

**Most of the heritability studies for bulls' puberty employed the AGESC 26–29 cm.*

(2) The heritability estimates for calving intervals, NRR, days open, and length of reproductive life in most populations were very low, which indicated that these traits would be influenced and improved by proper management practices. The application of synchronization-timed AI protocol (Goodling et al., 2005), body composition control, reproductive disorder treatment, and culling on time would benefit the related performance.

(3) The heritability of the same trait varies greatly among different breeds. For instance, the heritability of age at first calving was as high as 0.4 in a crossbreed of dairy cows (Effa et al., 2011), while the Dairy Overo Colorado breed was as low as 0.01 (Montaldo et al., 2017). The heritability of CI reported in Surti buffalo is 0.55 (Rathod et al., 2018) compared to the Murrah buffalo cows near to 0.1 (Thiruvenkadan et al., 2010). Although heritability was estimated using paternal half-sib correlation methods in both studies, lactation records, number of buffaloes, and sired by bulls were higher for Murrah buffaloes. Even in the same breed, the different populations showed varied values, which may related to different management and performance.

(4) For most of the reproductive traits, beef cattle had higher heritability estimates compared to those estimated in dairy cattle for the AOP and CD (**Tables 2**, **3**). Either the genetic makeup or the fact that dairy cows are more susceptible to reproductive diseases, such as endometritis, vaginitis, ovarian cyst, and mastitis, due to high energy consumption for milk production may be the reason for this difference.

(5) The breeding progress of buffalo is slow compared to dairy and beef cattle, as a few studies have reported during the last decade. Further large-scale studies are required to accurately estimate the genetic parameters for different reproductive traits in buffalo populations.

MARKER-ASSOCIATED STUDIES FOR BOVINE AND BUFFALO REPRODUCTIVE TRAITS

Concerning the disadvantages of the long cycle and not up-to-mark efficiency of traditional breeding, several association analyses were performed to identify genomic loci associated with the trait variation to find the possible candidate genes or to detect causative mutations. This section summarized the GWAS and candidate gene studies for bovine and buffalo reproductive traits published in the past 20 years (2000–2020) (**Supplementary Tables 1–3**).

At present, there are few marker-assisted selection (MAS) studies on the reproductive traits of buffalo. In this regard, *FSHR*, *INHA*, *LHCGR*, and *OPN* were reported to have significant effects on the buffalo superovulation responses. So far, few GWAS have been performed on buffalo reproductive traits (Camargo et al., 2015; Li et al., 2018a,b; de Araujo Neto et al., 2020). Previous GWASs for reproductive traits (Camargo et al., 2015; Li et al., 2018a) were conducted using the bovine reference genome assembly, and the results are expressed for bovine autosomes (BTA). Camargo et al. (2015) reported some candidate genes (*TPCN1*, *SCG5*, and *Fig 4*) associated with reproductive traits such as AFC, CI, and DO in buffalo. Also, Li et al. (2018a; 2018b) found 25 SNPs in 13 genes related to reproductive traits by integrating RNA-seq and GWAS methods. They also described significant SNPs on BBU 6, 9, and 15 [corresponding to bovine chromosomes 3, 7, 14, and 8: equivalence presented by Cribiu et al. (2001)]. Recently, ssGBLUP was employed to identify genomic regions affecting AFC and first calving interval (FCI) in buffalo cows and select candidate loci and gene expression (de Araujo Neto et al., 2020). They reported that the observed candidate regions for both traits (CI, AFC; explaining a large proportion of variance for both traits) were

TABLE 2 | Heritability estimates of reproduction traits in dairy cattle.

Category	Trait	Heritability	Breeds (Numbers/records)	References
Ovulation	Superovulation responses	0.231 ± 0.091	Holstein (2,489)	König et al., 2007
		0.27 ± 0.08	Holstein (926)	Gaddis et al., 2017
		0.234 ± 0.046(CL) 0.159 ± 0.087(EM)	Holstein–Friesian (56)	Bényei et al., 2004
		0.15 ± 0.01	Holstein (150,971)	Jaton et al., 2020
		0.15 ± 0.01/0.17 ± 0.01(NE) 0.14 ± 0.01/0.14 ± 0.01(VE) (Log/Ans)	Holstein (137,446)	Jaton et al., 2016a
		0.145 ± 0.007/0.188 ± 0.033(NE) 0.136 ± 0.007/0.187 ± 0.034(VE) (*in vivo/vitro*)	Holstein (145661/5310 records) (*in vivo/vitro*)	Jaton et al., 2016b
	Twinning rate	0.11 ± 0.01(parity1) 0.16 ± 0.01(parity2) 0.14 ± 0.01(parity3)	Japanese Holsteins (1,323,946) (1053469) (750600)	Yutaka et al., 2015
		0.0192 ± 0.0009/0.142 ± 0.007 (LM/TLM)	Holsteins (658436 cows/1440540 records)	Lett and Kirkpatrick, 2018
		0.1	12 multiple breeds (9272 females)	Allan et al., 2007
		0.013(parity1) 0.022(parity2) 0.024(parity3) 0.026(parity4) 0.031(parity5)	Israeli Holstein (671,361) (460940) (304213) (188077) (104434)	Weller et al., 2008
Mating	Age of puberty	0.38	Friesian × Ethiopian Boran (399) Jersey × Ethiopian Boran (151)	Effa et al., 2011
	Age at first calving	0.4	Friesian × Ethiopian Boran (399) Jersey × Ethiopian Boran (151)	Effa et al., 2011
		0.26 ± 0.02	South African Holstein (20419)	Makgahlela et al., 2008
		0.20 ± 0.03/0.21 ± 0.03(uni-trait/bi-trait analysis)	Brazilian Girolando (10,900)	Canaza-Cayo et al., 2018
		0.219 ± 0.162	multiple dairy cows (224)	Ali et al., 2019
		0.17 ± 0.01 0.093 ± 0.037	Holstein–Friesian Other dairy breeds	Berry and Evans, 2014
		0.15 ± 0.04 (PM)/0.16 ± 0.04 (GPM)	7 breeds (9,106)	Konkruea et al., 2019
		0.111	Holstein (276,573)	Changhee et al., 2013
		0.103 ± 0.025	German Holstein heifers (721919)	Heise et al., 2017
		0.01 ± 0.07	Dairy Overo Colorado breed (2,043)	Montaldo et al., 2017
	Non-return rate	0.1292 (NRR45) 0.1460 (NRR90)	Holstein (21,405)	Ansari-Mahyari et al., 2019
		0.02 (Paternal NRR90) 0.02 (Maternal NRR90)	German Holstein (1193) (1283)	Kaupe et al., 2007
		0.012 (heifer NRR56) 0.015 (cow NRR56)	Holstein (2,527)	Müller et al., 2017
		0.011 ± 0.001(NRR56)	Holstein (386869)	Zhang et al., 2019
		0.027 ± 0.0004 0.020 ± 0.001	Holstein–Friesian Other dairy breeds	Berry et al., 2014
	Pregnancy rate	0.04/0.02/0.01 (DPR/CCR/HCR)	Holstein (2,107)	Gaddis et al., 2016
		0.04	Spanish Holstein (113375 records)	Gonzálezrecio and Alenda, 2005
Calving	Calving interval	0.17	Friesian × Ethiopian Boran (847) Jersey × Ethiopian Boran (559)	Effa et al., 2011
		0.16 ± 0.12 0.00 ± 0.09	Holstein (624) Swedish Red (460)	Tarekegn et al., 2019
		0.14 ± 0.211	multiple dairy cow (224)	Ali et al., 2019
		0.106 ± 0.015 (linear sire model) 0.103 ± 0.013 (linear animal model) 0.059 ± 0.006 (repeatability animal model)	Iranian Holstein (22,269)	Chegini et al., 2019a
		0.07 ± 0.013	Holstein (11674 records)	Toghiani, 2012

(Continued)

TABLE 2 | Continued.

Category	Trait	Heritability	Breeds (Numbers/records)	References
		0.044 ± 0.01	Holstein (20544)	Chegini et al., 2019b
		0.04 ± 0.003	Iranian Holstein (129199)	Hossein Salimi et al., 2017
		0.04	Spanish Holstein (96346 records)	Gonzálezrecio and Alenda, 2005
		0.034 ± 0.001 0.029 ± 0.004	Holstein–Friesian Other dairy breeds	Berry et al., 2014
		0.002 ± 0.02	Dairy Overo Colorado breed (3,488)	Montaldo et al., 2017
		0.01 ± 0.02 (CI_1) 0.00 ± 0.04 (CI_2) 0.08 ± 0.07 (CI_3)	Brazilian Girolando (5327) (3444) (2229)	Canaza-Cayo et al., 2018
		0.03 ± 0.01(CI_1) 0.04 ± 0.01(CI_2) 0.04 ± 0.01(CI_3) 0.03 ± 0.01(CI_4)	South African Holstein (20419) (18589) (10681) (15529)	Makgahlela et al., 2008
		0.088 (CI_1) 0.142(CI_2)	Holstein (167996 records) (128080 records)	Changhee et al., 2013
	Days open/calving to conception interval	0.102	Canadian Holstein (3,729)	Nayeri et al., 2016
		0.09 ± 0.121	multiple dairy cows (224)	Ali et al., 2019
		0.06 ± 0.03	Holstein (3,682)	Saowaphak et al., 2017
		0.06 ± 0.008	Holstein (15895)	Toghiani, 2012
		0.04	Spanish Holstein (113375 records)	Gonzálezrecio and Alenda, 2005
		0.04 ± 0.003	Iranian Holstein (129199)	Hossein Salimi et al., 2017
		0.033/0.024 (Model1/2)	Korean Holstein (14,188)	Lee and Han, 2004
		0.026	Holstein (2,527)	Müller et al., 2017
		0.038 ± 0.002 0.030 ± 0.001	Holstein–Friesian Other dairy breeds	Berry et al., 2014
	Calving difficulty	0.132 ± 0.003	Holstein (734)	Maryam et al., 2016
		0.121 ± 0.024 (LM) 0.074 ± 0.012 (TM)	Walloon Holstein	Vanderick et al., 2015
		0.05 (paternal CE) 0.05 (maternal CE)	German Holstein (1267) (1287)	Kaupe et al., 2007
		0.048 (paternal CE) 0.039 (maternal CE)	Holstein (2,527)	Müller et al., 2017
		0.043 ± 0.0031/0.010 ± 0.0016 (LM1) 0.041 ± 0.0030/0.010 ± 0.0015 (LM2) 0.046 ± 0.0032/0.011 ± 0.0016 (LM3) 0.086 ± 0.0091/0.023 ± 0.0037 (TM) (direct/maternal CE)	Portuguese dairy cattle (320,953 records)	Silvestre et al., 2019
		0.02 ± 0.002	Iranian Holstein (132831)	Hossein Salimi et al., 2017
		0.015/0.030 (Model1/2)	Korean Holstein (14,188)	Lee and Han, 2004
	Length of productive life	0.16	German Holstein (1,286)	Kaupe et al., 2007
		0.12	Pinzgau Cattle	Egger-Danner et al., 2005
		0.102	Holstein (276,573)	Changhee et al., 2013
		0.10 ± 0.03	Holstein (4,739)	Saowaphak et al., 2017
		0.06/0.10/0.18/0.25 (LPL/FPL/MPL/PPL)	Swedish Red and White dairy cattle (538783)	Roxström and Strandberg, 2002
		0.04	Hungarian Holstein (1403747)	van der Linde et al., 2006

located on BBU 3, 12, 21, and 22. Also, candidate regions were found on BBU 6, 7, 8, 9, and 15 for age at first calving and on BBU 4, 14, and 19 for FCI. The *ROCK2*, *PMVK*, *ADCY2*, *MAP2K6*, *BMP10*, and *GFPT1* genes are the main candidates for reproductive traits in water dairy buffaloes, which may be used in the future for animal breeding programs or for gene expression studies of the species (de Araujo Neto et al., 2020). The *GFPT1* and *BMP10* genes are interesting because they have been detected for both traits, which may be related to a possible pleiotropic effect.

The candidate gene studies for bovine reproductive traits mostly used genes of hormones and/or growth factors and their receptors as candidates (Tang et al., 2011; Yang et al., 2013; Arslan et al., 2017). For example, polymorphisms in the *GnRH*, *GnRHR*, *LEP*, and *LHCGR* were studied for association with reproductive traits of buffalo bulls. Notably, genes involved in IGF1 and LEP

TABLE 3 | Heritability estimates of reproduction traits in beef cattle.

Category	Trait	Heritability	Breeds (Numbers/Records)	References
Ovulation	Ovulation rate	0.12	MARC twinning herd (16,035)	Allan et al., 2014
		0.08	MARC 12 breeds of cattle (29485 records)	Allan et al., 2007
		0.02	multiple breeds	Piper et al., 2017
	Superovulation responses (VE)	0.56–0.65 (1 flush) 0.20–0.26 (3 flushes)	Nellore (405) (858)	Peixoto et al., 2004
	Twinning	0.1	MARC twinning herd (16,035)	Allan et al., 2014
		0.1	MARC 12 breeds of cattle (9272 records)	Allan et al., 2007
		0.062 ± 0.093 (RThM) 0.014 ± 0.018 (RLM)	Maremmana cattle (1,260)	Moioli et al., 2017
Mating	Age of puberty	0.31 ± 0.05 (AFO) 0.27 ± 0.04 (SFO) 0.56 ± 0.11 (AGECL) 0.78 ± 0.10 (AGE26)	Angus cattle (1513 records) (1588 records) Brahman heifers (1007) Brahman bulls (1118)	Morris et al., 2000 Fortes et al., 2012
		0.57 ± 0.12 0.52 ± 0.12 (AGECL)	Brahman heifers (1007) Tropical Composite heifers (1108)	Johnston et al., 2009
		0.35/0.22/0.11 0.22/0.33 0.24/0.32 (AGECL)	Brahman (397/371/206) Santa Gertrudis (1022/776) Droughtmaster (222/688)	Engle et al., 2019
		0.42–0.44	Nelore cattle (12964)	Forni and Albuquerque, 2005
		0.26 ± 0.03	Heifer Angus (629)	Morris et al., 2011
		0.221 ± 0.08 (univariate) 0.198 ± 0.06 (multivariate)	50% Red Angus, 25%Charolais and 25%Tarentaise (890)	Toghiani et al., 2017
		0.310 ± 0.050 (AFO)	Beef cattle	Berry and Evans, 2014
		0.16–0.20	1828 Beef CRC (868 Brahman and 960 Tropical Composite) 3695 SMF (979 Brahman,1802 Santa Gertrudis and 914 Droughtmaster)	Warburton et al., 2020
	Scrotal circumference	0.37 ± 0.06(SC-8 month) 0.44 ± 0.06 (SC-10 month) 0.42 ± 0.06 (SC-12 month)	Angus cattle (1702 records) (1691 records) (1671 records)	Morris et al., 2000
		0.48 ± 0.02 (AGE365) 0.52 ± 0.02 (AGE450)	Brazilian Nellore (27567 records)	Kluska et al., 2018
		0.397 ± 0.011 (AGE365)	Nelore (135862 records)	Schmidt et al., 2019
		0.33 ± 0.07 (AGE365) 0.41 ± 0.07 (AGE450)	Guzera beef cattle (1773) Guzera beef cattle (2091)	Tramonte et al., 2019
		0.29 (AGE365)	Nelore cattle (66986 records)	Costa et al., 2020
		0.18 ± 0.02 (AGE365)	Charolais, Charbray, and Charolais-Zebu crosses (18,972)	Martínez-Velázquez et al., 2020
	Age at first calving	0.31 ± 0.016	Crossbred *Bos taurus* (64380 records)	Berry et al., 2014
		0.27 ± 0.12	Asturiana de los Valles (1226 records)	Goyache and Gutiérrez, 2001
		0.24 ± 0.04	Brazilian Nelore cattle (762)	Mota et al., 2017
		0.235 ± 0.018	Asturiana de los Valles (2533 records)	Gutiérrez et al., 2002
		0.220 ± 0.11	Jersey × Red Sindhi (313)	Vinothraj et al., 2016
		0.215 ± 0.026	Japanese Black Cows (24595 records)	Oyama et al., 2002
		0.20	Nelore cattle (1853)	Costa et al., 2019
		0.20–0.22	Simmental (3,063)	Amaya-Martínez et al., 2020
		0.17 ± 0.04	Brahman-Angus (909)	Elzo et al., 2018
		0.158 ± 0.039	Japanese Black cows (2,078)	Setiaji and Oikawa, 2019
		0.137 ± 0.008	beef cattle	Berry et al., 2014
		0.13 ± 0.130	Crossbred heifers (538 records)	Akanno et al., 2015
		0.11 ± 0.01	Brazilian Nellore (18526 records)	Kluska et al., 2018
		0.10 ± 0.01 (multi-trait model) 0.08 ± 0.01 (single-trait model)	Hanwoo cows (15,355)	Lopez et al., 2019

(Continued)

TABLE 3 | Continued.

Category	Trait	Heritability	Breeds (Numbers/Records)	References
		0.10 ± 0.01	Nelore beef cattle (25,594)	Boligon and Albuquerque, 2011
		0.20/0.19/0.18/0.09 (LM/SM/PM/TLcens)	Brazilian Brahman cattle (53703 records)	Lázaro et al., 2019
		0.08	Nelore cattle (374665 records)	Costa et al., 2020
		0.06/0–0.15 0.13/0.06–0.13 (AMI/MHNRHOP1)	Limousine (18,500) Charolais (4,330)	de Rezende et al., 2020
		0.06–0.08	Nelore cattle (18615)	Forni and Albuquerque, 2005
		0.039 ± 0.039 (univariate) 0.031 ± 0.01 (multivariate)	50% Red Angus, 25%Charolais and 25%Tarentaise (1117)	Toghiani et al., 2017
	Non-return rate	0.020 ± 0.029 (1st parity) 0.014 ± 0.022 (2nd parity) 0.013 ± 0.034 (3rd parity) 0.013 ± 0.017 (repeatability model)	Japanese Black cows (2,078)	Setiaji and Oikawa, 2019
	Pregnancy rate	0.21 ± 0.009	Angus (1,299)	Doyle et al., 2000
		0.14 ± 0.099	Crossbred heifers (734 records)	Akanno et al., 2015
		0.12 ± 0.05 (yearlings) 0.08 ± 0.064 (2-year-olds)	Angus cattle (1190 records) (711 records)	Morris et al., 2000
		0.027 ± 0.38 (1st parity) 0.023 ± 0.034 (2nd parity) 0.021 ± 0.036 (3rd parity) 0.022 ± 0.007 (repeatability model)	Japanese Black cows (2,078)	Setiaji and Oikawa, 2019
		0.025/0.014/0.023/0.014 (model 1/2/3/4)	Sistani beef cattle (1489 records)	Faraji-Arough and Rokouei, 2016
Calving	Calving interval	0.222 ± 0.101	Jersey × Red Sindhi (522)	Vinothraj et al., 2016
		0.125 ± 0.020	Asturiana de los Valles (2007 records)	Gutiérrez et al., 2002
		0.12 ± 0.03	Asturiana de los Valles (1851 records)	Goyache and Gutiérrez, 2001
		0.105 ± 0.008	Nelore (33735 records)	Schmidt et al., 2019
		0.09 ± 0.02 (CI_1)	Brahman-Angus (447)	Elzo et al., 2018
		0.02 ± 0.02 (CI_1) 0.02 ± 0.04 (CI_2) 0.06 ± 0.03 (mean CI)	Nelore (2642) (1437) (2888)	do Amaral Grossi et al., 2016
		0.049 ± 0.048 (CI_1) 0.043 ± 0.045 (CI_2) 0.048 ± 0.042 (CI_3) 0.047 ± 0.009 (repeatability model)	Japanese Black cows (2,078)	Setiaji and Oikawa, 2019
		0.047 ± 0.009	Japanese Black Cows (72740 records)	Oyama et al., 2002
		0.032 ± 0.004	beef cattle	Berry et al., 2014
		0.056/0.040/0.033/0.032 (model 1/2/3/4)	Sistani beef cattle (1489 records)	Faraji-Arough and Rokouei, 2016
		0.01 ± 0.05 (CI_1) 0.04 ± 0.02 (CI_2) 0.07 ± 0.03 (CI_3) 0.03 ± 0.01 (multi-trait model)	Hanwoo cows (1936) (11144) (8201) (32599)	Lopez et al., 2019
		0.02 ± 0.004	Crossbred *Bos taurus* (101864 records)	Berry and Evans, 2014
	Days open/calving to conception interval	0.192 (model 1) 0.091 (model 2) 0.168/0.197/0.170/0.091 (model3) 0.154/0.132 (model4) 0.135/0.090/0.086 (model5)	Asturiana de los Valles (21349 records) (3250/3416/13783/900 records) (6666/14683 records) (21349 records)	Goyache et al., 2005
		0.110 ± 0.04	beef cattle	Berry et al., 2014
		0.110 ± 0.04	Angus (1680 records)	Morris et al., 2000
		0.09/0.045/0.096/0.049 (model 1/2/3/4)	Sistani beef cattle (1489 records)	Faraji-Arough and Rokouei, 2016
		0.047 ± 0.009	Japanese Black cows (72740 records)	Oyama et al., 2002

(Continued)

TABLE 3 | Continued.

Category	Trait	Heritability	Breeds (Numbers/Records)	References
		0.042 ± 0.044 (1st parity) 0.034 ± 0.052 (2nd parity) 0.034 ± 0.033 (3rd parity) 0.036 ± 0.021 (repeatability model)	Japanese Black cows (2,078)	Setiaji and Oikawa, 2019
		0.02 ± 0.05 (1st parity) 0.09 ± 0.02 (2nd parity) 0.08 ± 0.03 (3rd parity) 0.03 ± 0.01(multi-trait model)	Hanwoo cows (1726) (7308) (5888) (32465)	Lopez et al., 2019
	Calving difficulty	0.42	Asturiana de los Valles (7298 records)	Goyache and Gutiérrez, 2001
		0.325 ± 0.022	Asturiana de los Valles (35,395 records)	Cervantes et al., 2010
		0.32 ± 0.174	Crossbred heifers (543 records)	Akanno et al., 2015
		0.29 ± 0.10	multi breeds (5,795)	Ahlberg, 2014
		0.250 ± 0.018	Crossbred *Bos taurus* (100445 records)	Berry and Evans, 2014
	Length of productive life	0.096 ± 0.001	Multiple breeds (21,895)	Brzáková et al., 2019

pathways were reported to affect multiple reproductive traits. For example, *IGF1* could affect a variety of ovulation- and mating-related traits. *LEP* and *LEPR* showed significant effects on both breeding- and calving-related traits. Moreover, long non-coding RNA and ribosomal RNA could be future research directions since non-coding RNAs (U6 spliceosomal RNA) were reported to affect reproductive traits (Fortes et al., 2013; Nascimento et al., 2016; Buzanskas et al., 2017). The combination of GWAS and other omics studies are becoming more useful, as they provide a broad space for exploring candidate gene functions and related mechanisms.

Further, we visualized the chromosomal distribution of quantitative trait loci (QTL) in cattle related to each reproductive trait using the Cattle Quantitative Trait Locus Database (Cattle QTLdb) (Hu et al., 2019) (**Supplementary Figures 1–3**). Only 11 QTL related to ovulation-related traits were identified, and four of these were located on chromosome 5, where the *IGF1* gene is placed (Miller et al., 1992) (**Supplementary Figure 1**). The QTL for mating-related traits were spread throughout different chromosomes (**Supplementary Figure 1A**). The most abundant chromosome is BTX with 10237 QTL (96.4%) related to puberty. BTA2 (21QTLs, 19.6%) and BTA14 (15 QTLs, 14.0%) had the most associated loci for AFC (**Supplementary Figure 1B**). Most of the QTL for NRR were located on BTA17 (233421 QTLs, 94.7%). However, QTL for PR-related were scattered (**Supplementary Figure 2**). About 37.1% of QTL related to calving interval were enriched in BTA25 (17.5%) and BTA29 (19.6%). Whereas, BTA 21 enriched the most QTLs (44.8%) related to CD, and BTA18 had 30.7% of QTL related to the length of productive life.

Undoubtedly, these significantly enriched chromosomes (BTX related to puberty, BTA related to NRR, and BTA related to CD) could be directions for future research. Moreover, certain areas that affect multiple traits of different species also deserve further attention. For example, McClure et al. (2010) found one SNP related to CD at 49.1 Mb of BTA 20 in Angus cattle (McClure et al., 2010), while Ke et al. (2014) reported SNP in a similar region in dairy cattle affecting age at first calving. The relationship between these highly enriched chromosomal regions and various traits is worthy of further investigation.

Based on morphological and behavioral criteria, the domestic Asian water buffalo has two types (Macgregor, 1941). The two types have different chromosome numbers: river buffalo (*Bubalus bubalis*, 2n = 50) and swamp buffalo (*Bubalus bubalis carabanesis*, 2n = 48) (Ulbrich and Fischer, 1966). In addition, the chromosomal karyotype of hybrid buffalo is more complicated. Although presenting different species, buffalo and bovine share highly homologous chromosomes banding, as well as gene mapping (Amaral et al., 2008; Michelizzi et al., 2010; Kale et al., 2014). It is also reported that river buffalo and bovine chromosomes can be matched arm for arm at the cytogenetic level (Williams et al., 2017; Du et al., 2019). Despite the complicated genomic background of buffalo, candidate genes or their chromosome locations identified for the bovine reproductive traits could be considered as a valuable reference for buffalo.

GENOMIC SELECTION FOR REPRODUCTIVE TRAITS IN BOVINE AND BUFFALO

Phenotypic records for a trait of individuals and their relatives are used to estimate breeding values by employing the best linear unbiased prediction (BLUP) to facilitate animal selection for economically important traits (Henderson, 1984). It is believed for genetic selection that information at the DNA level can quicken the genetic progression compared to phenotypic data alone. The sparse map of genetic markers can be used to detect QTL (Georges et al., 1995). Combining genetic marker information with BLUP (Fernando and Grossman, 1989) showed an increase in the genetic gain by 8–38% (Fernando and Grossman, 1989; Goddard, 1996). The effectiveness of sparse markers in outbreeding species was limited, as an establishment of linkage phase between a marker and QTL is necessary for

TABLE 4 | Heritability estimates of reproduction traits in buffalo.

Trait	Heritability	Breeds (Numbers/records)	References
Age at first calving	0.28 ± 0.03	Murrah buffalo (827)	Kumar et al., 2015
	0.226 ± 0.154 0.16	Surti buffalo (48) Murrah water buffalo (2290 records)	Rathod et al., 2018 de Araujo Neto et al., 2020
	0.16 ± 0.04	Murrah buffalo (2389 records)	Barros et al., 2016
	0.16 ± 0.12	Murrah buffalo (167)	Thiruvenkadan et al., 2010
	0.17 ± 0.02	Murrah buffaloes (3,431 records)	Camargo et al., 2015
	0.135 ± 0.035	Indian Murrah buffalo (1,456 records)	Gupta et al., 2015
	0.11 ± 0.06	Egyptian buffalo (1911 records)	El-Bramony, 2011
	0.07 ± 0.05	Murrah buffalo (1,578)	Seno et al., 2010
calving interval	0.55 ± 0.131	Surti buffalo (158)	Rathod et al., 2018
	0.234 ± 0.175	Indian Murrah buffalo (1,456 records)	Gupta et al., 2015
	0.14 ± 0.07 (CI_1)	Murrah buffalo (1,578)	Seno et al., 2010
	0.09 ± 0.13	Murrah buffalo (506)	Thiruvenkadan et al., 2010
	0.085 ± 0.134	Iranian Khuzestan buffalo (146 records)	Morammazi et al., 2007
	0.07 ± 0.05	Egyptian buffalo (1911 records)	El-Bramony, 2011
	0.06 ± 0.01	Egyptian buffalo (2,066)	El-Bramony and Reclamation, 2014
	0.06 ± 0.01	Murrah buffaloes (4729 records)	Camargo et al., 2015
	0.05 ± 0.08	Mehsana buffalo (812 records)	Galsar et al., 2016
	0.05 ± 0.01	Murrah buffalo (5672 records)	Barros et al., 2016
	0.03(CI_1)	Murrah water buffalo (765 records)	de Araujo Neto et al., 2020
Days open	0.14 ± 0.03	Murrah buffaloes (6894 records)	Camargo et al., 2015
Calving difficulty	0.16/0.19/0.06/0.08/ 0.09/0.04/0.11 (parity1–7)	Iraqi Buffalo (360)	Al-Khuzai et al., 2019

every family in which the marker is to be used for selection (Meuwissen et al., 2001).

The total number of SNP estimated at millions and the advent of DNA Chip technology made genotyping of many animals for many of these markers feasible and cost-effective. However, a dense marker map improved precision for QTL mapping by traditional linkage analysis (Darvasi et al., 1993). Therefore, a search for a different approach to efficiently use all this marker information remained necessary.

Considering a denser marker map, not only could some markers be close to QTL but also, in linkage disequilibrium with it, it was anticipated that some markers could have a positive effect on the quantitative traits across all families and be used for selection without the need to establish a Linkage phase in each family. Close markers can also be combined into a haplotype. Chromosome bearing the rare marker haplotype is likely to be identical by descent and hence carry the same QTL allele. Meuwissen et al. (2001), estimated the effect of the quantitative trait of the small chromosome segment defined by the haplotype of the allele that they carry. They concluded that it's possible to accurately estimate the breeding value of animals that have no phenotypic records by estimating a large number of haplotype effects. Using least squares, all haplotype effects could not be estimated simultaneously. Even when only the largest effects were included, they were overestimated and the accuracy of predicting breeding value was low. Methods that assumed prior distribution for the variance associated with each chromosome segment gave a more accurate prediction of breeding values even when the prior was not correct. Selection based on breeding values predicted from markers could substantially increase the rate of genetic gain in animals and plants, especially if combined with reproductive techniques to shorten the generation interval. Selection based on pedigree has played an important role in the selective breeding improvement in domestic animals.

Quantitative traits are usually affected by many genes and, consequently, the benefits from the MAS are limited by the proportion of the genetic variance explained by the QTL. Hence, it is warranted to utilize all the QTL affecting the traits in MAS. Nevertheless, a dense marker map defines a very large number of chromosome segments and so there will be many effects to be estimated, probably more than there are phenotypic data points from which to estimate them (Meuwissen et al., 2001).

With the emergence of high-density SNP chips, such as Illumina chips [BovineHD BeadChip SNP, BovineSNP50 chip, High-Density Bovine SNP chip (777K)] and Axiom® Buffalo Genotyping Array (90K), GS methods are improving livestock genetic evaluation systems. They have the advantages of high accuracy, short interval between generations, and rapid genetic progress.

At present, GS has been applied in cattle on a large scale, but mainly focus on milk production and carcass traits (Silva et al., 2014; Weller et al., 2017). The GS studies on reproductive traits in dairy and beef cattle, including AFC, puberty, NRR, PR, days open, and CD, are listed on **Table 5**.

For AFC, the accuracy of genomic prediction was varied among different populations and methods. In the Nellore breed, the accuracy of prediction for AFC was 0.64 (Boddhireddy et al., 2014); however, another scholarly journal reported that the accuracy ranged between 0.38 and 0.42 by three different models (Costa et al., 2019). The prediction accuracy is around 0.23–0.33 in another Nellore cow population (Mota et al., 2018). Using the ssGBLUP model, the accuracy of prediction for AFC was 0.299 in the Thai native breed (Laodim et al., 2019), and was 0.56 in the Gyr dairy cattle breed (Boison et al., 2017).

TABLE 5 | A summary of genomic selection studies for reproductive traits.

Traits studied	Breed (country)	Chip size	Validation population size	Models	Response variable	Accuracy of prediction	Regression coefficients	References
Age at first calving	Nelore (Brazil)	Illumina BovineHD	1,853	GBLUP BAYESCπ IBLASSO	dEBV	0.38(GBLUP), 0.39(IBLASSO) 0.42(BAYESCπ)	0.88(GBLUP), 1.14(IBLASSO) 0.81(BAYESC)	Costa et al., 2019
	Nelore (Brazil)	Illumina Bovine 70 K	714	BayesA BayesB BayesCπ BLASSO BRR	dEBV	0.24(BayesA) 0.23(BayesB) 0.33(BayesCπ) 0.24(BLASSO) 0.38	0.62 0.63 0.65 0.83 0.65	Mota et al., 2018
	Nelore (Brazil)	Illumina BovineHD	2,241	BayesC	EBVs	0.64	0.9	Boddhireddy et al., 2014
	crossbred animals (Thai)	GeneSeek 80k chip	8,361	ss GBLUP ssGBLUPS1 ssGBLUPS2	EBV	0.297 0.298 0.264		Laodim et al., 2019
	Gyr dairy cattle (Brazil)	GeneSeek SGGP-20Ki Illumina BovineSNP50 GeneSeek GGP-75Ki Illumina BovineHD	422 bulls and 1582 cows	GBLUP	dEBVs	0.380	0.968/0.960 0.966/0.958 0.967/0.959 0.968/0.970 (bulls/bulls and cows)	Boison et al., 2017
	CGC: 50%Red Angus 25%Charolais 25%Tarentaise	BovineSNP50 chip	1117 records	BayesA BayesB BayesCπ	EBVs	0.148 0.143/0.154/0.146 (π = 0.99/0.95/0.90) 0.150		Toghiani et al., 2017
Scrotal circumference	Braford and Hereford (Brazil)	Illumina BovineSNP50K Illumina BovineHD	3680 (2997 Braford and 683 Hereford)	tsGBLUP/ ssGBLUP	EBVs/ dEBVs	0.28–0.33 0.15–0.17	0.50–1.10 0.55–1.13	Piccoli et al., 2020
	Brangus	GGP−LDV3 chip (1074) GGP−LDV4 chip (1535) Illumina BovineSNP50 (261) GGP−HDT (295) GGP−UHD (628) Illumina Bovine HD (4)	3,797	tsGBLUP ssGBLUP	EBVs/ dEBVs	0.717 0.634		Lopes et al., 2018
	Nelore cattle (Brazil)	Illumina BovineHD (763) Illumina BovineSNP50 (1478)	2,241	BayesC	EBVs	0.59/0.59 (AGE365/450) 0.57/0.56 (AGE365/450)	0.95/0.93 (AGE365/450) 0.89/0.86 (AGE365/450)	Boddhireddy et al., 2014
	Nelore bulls (Brazil)	Illumina BovineHD	691	GBLUP Bayes C BLASSO	dEBV	0.68(GBLUP0) 0.71(GBLUP20) 0.72(Bayes C) 0.72(BLASSO)	1.27 (GBLUP0) 1.44(GBLUP2) 1.68(BAYESC) 1.65(BLASSO)	Neves et al., 2014
	Angus' sires (America)	Illumina BovineSNP50	439	BayesC	dEBVs	0.487 (K-means)/0.600 (Random)	0.916 (K-means)/ 0.983 (Random)	Saatchi et al., 2011
Puberty (age at first corpus luteum)	Beef CRC: (882 Brahman and 990 Tropical Composite) Smart Futures: (974 Brahman, 1798 Santa Gertrudis, and 910 Droughtmaster)	Illumina BovineSNP50 chip GeneSeek GGP-LD array	1,872 3682	GBLUP	EBVs	0.49 ± 0.06 (Tropical Composite) 0.52 ± 0.07 (Brahman) (80% CRC + SF)		Engle et al., 2019

(Continued)

TABLE 5 | Continued.

Traits studied	Breed (country)	Chip size	Validation population size	Models	Response variable	Accuracy of prediction	Regression coefficients	References
	50%Red Angus 25%Charolais 25%Tarentaise	BovineSNP50 chip	890	BayesA BayesB BayesC	EBVs	0.237 0.188/0.235/0.242 (π = 0.99/0.95/0.90) 0.226		Toghiani et al., 2017
	CRC(2174) and Validation cows (4286)	Illumina BovineHD Illumina 7K Illumina BovineSNP50K	6,460	GBLUP	EBVs	0.33 (Brahman) 0.15 (Tropical Composite)		Zhang et al., 2014
Non-return rate	Holstein (Canada)	Illumina Bovine SNP50	317 (first) and 489 (later)	ssGBLUP msGBLUP	GEBV DGV	0.39/0.33 (first/later)	0.63–0.97 (first) 0.81–1.35 (later)	Guarini et al., 2018
Heifer pregnancy rate	Angus sires (America)	Illumina BovineSNP50	133	BayesC	dEBVs	0.269 (K-means)/0.378 (Random)	1.337 (K-means)/1.580 (Random)	Saatchi et al., 2011
	Nelore (Brazil)	Illumina BovineHD (763) Illumina BovineSNP50 (1478)	2,241	BayesC	EBVs	0.64 0.64	0.89 0.87	Boddhireddy et al., 2014
Days open	Holstein (North America)	Illumina Bovine SNP 50 TM Chip	6,515	GBLUP	dEBV	0.50	0.9	Forutan et al., 2018
Calving ease direct/maternal (CED/CEM)	Brangus (CED/CEM)	GGP−LDV3 chip (1074) GGP−LDV4 chip (1535) Illumina BovineSNP50 (261) GGP−HDT (295) GGP−UHD (628) Illumina Bovine HD (4)	3,797	tsGBLUP ssGBLUP	EBVs dEBVs	0.451/0.512 0.337/0.266 (CED/CEM)		Lopes et al., 2018
	Holstein (Canada) (calving ease)	Illumina Bovine SNP50	438 (first) and 363 (later)	ssGBLUP msGBLUP	GEBV DGV	0.76/0.69 (first/later)	0.71–1.09 (first) 0.56–0.82 (later)	Guarini et al., 2018
	Angus bulls (America) (CED/CEM)	Illumina BovineSNP50 BeadChip	3180	BayesC	dEBVs	CED:0.488/0.617 CEM:0.416/0.571 (K-means/Random)	CED:0.942/1.007 CEM:1.181/1.277 (K-means/ Random)	Saatchi et al., 2011
	Norwegian Red bulls (calving ease)	Affymetrix 25K MIP-SNP chip	500	GBLUP BayesB MIXTURE	GW-EBV	0.406/0.382 0.411/0.392 0.429/0.401 (Cohort//Random masking)	1.192/1.104 0.932/0.953 0.998/0.862 (Cohort//Random masking)	Luan et al., 2009

Genomic selection studies on puberty (scrotal circumference and age at first corpus luteum) showed that the accuracy performance of different models is above 0.6 (Boddhireddy et al., 2014; Neves et al., 2014; Toghiani et al., 2017; Lopes et al., 2018; Engle et al., 2019). However, the accuracy was decreased dramatically in crossbred populations (Zhang et al., 2014; Piccoli et al., 2020). The limited reference population in the hybrid population and the general traits of the reference population have no direct counterpart in the validation population, which may be the reason for this decrease.

In the PR studies, the accuracy of prediction was 0.269 in the Angus population (Saatchi et al., 2011) and 0.64 in Nelore cattle (Boddhireddy et al., 2014). For CD, the highest accuracy was 0.516 in Brangus using GBLUP models (Lopes et al., 2018), and the prediction accuracy of different beef cattle breeds is around 0.45 among different models (Luan et al., 2009; Saatchi et al., 2011), while the accuracy in dairy cows was lower by 0.24–0.34 (Guarini et al., 2018).

Regarding buffalo studies, genomic evaluation reports are very limited either for productive or reproductive traits. There is only one published study for AFC and CI in buffalo (de Araujo Neto et al., 2020). Genomic evaluation studies in buffalo are still in the developing stage. The main limitation of applying genomic evaluation in buffalo is the lack of a well-structured reference population. Since the number of individuals with both genotypic and phenotypic information in each country is still limited, a multi-breed genomic evaluation would be the best alternative (Liu et al., 2018; Abdel-Shafy et al., 2020a,b).

CONCLUSION AND PERSPECTIVES

Reproductive traits were depreciated during selection indexes to improve the genetic potential of livestock. Hence, the recently desired gains are being practiced to ensure that the all TMI (total merit index) traits show a positive response or, at the very least, no negative response. However, the statistical data from the Council on Dairy Cattle Breeding (CDCB)[1] indicated that, without severely slowing genetic gain for milk production, the daughter PR has stabilized and the declining trend has been reversing since 2003. A similar trend has also been demonstrated by García-Ruiz et al. (2016). Moreover, several pregnancy-related SNPs with neutral associations with milk production in Holstein bulls were identified (Cochran et al., 2013). It elicits the possibility of increasing fertility without reducing productive performance during selection.

Unlike dairy and beef cattle, few studies have been performed so far for reproductive traits in buffalo. Methods such as GWAS and GS require a large group size, well-structured pedigree, and accurate phenotypic records, which are big challenges for buffalo populations. The first reference for buffalo genome sequencing was released in 2017 (Williams et al., 2017), lacking the sequence in the chromosome and genes annotation, which was completed and updated in 2019 (Low et al., 2019; Mintoo et al., 2019). It will quicken the GS research and be significantly helpful in promoting buffalo breeding.

Dissimilar to dairy production traits, GWAS for reproductive traits seems to be underpowered and has difficulty in finding major QTL. It still provides genetic variability across many genome-wide genes and intragenic regions for complex trait studies, which greatly increases the understanding of complex traits' molecular genetic mechanisms.

[1] https://queries.uscdcb.com/eval/summary/trend.cfm

For reproductive traits with low heritability, the genetic gain using GS is improved three to four times per year compared to traditional methods (García-Ruiz et al., 2016). However, GS is also facing some difficulties, especially for buffalo, such as lacking an optimum population structure with record and some species having no dense marker maps yet. Its accuracy is limited by the reference population's size and SNP marker density, which is obvious in some hybrid populations. In developing countries, there is a lack of complete historical records, and the number of genotyped animals has limited the development of GS. Also, for those traits with low to high heritability (such as puberty, age at first calving, and CD), multivariate GS can performed on multiple traits to improve prediction accuracy. In addition, multi-breed genomic evaluation can be used for populations with limited size. Besides, multi-omics data integration and analysis are gaining more attention from fields such as genomics, transcriptomics, and epigenomics.

AUTHOR CONTRIBUTIONS

GH contributed to the conception and design of the study. BS wrote the first draft of the manuscript and collected the data. CD, HS, MA, and YY wrote sections of the manuscript. NG, HA, SM, YZ, TD, LY, and SZ revised the manuscript and made profound suggestions. All authors contributed to manuscript revision and read and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (31872352), Fundamental Research Funds for the Central Universities (2662018PY037), and the Earmarked Fund for Modern Agro-Industry Technology Research System (CARS-36).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.617128/full#supplementary-material

REFERENCES

Abdel-Shafy, H., Awad, M. A. A., El-Regalaty, H., El-Assal, S. E.-D., and Abou-Bakr, S. (2020a). Prospecting genomic regions associated with milk production traits in Egyptian buffalo. *J. Dairy Res.* 87, 389–396. doi: 10.1017/s0022029920000953

Abdel-Shafy, H., Awad, M. A. A., El-Regalaty, H., Ismael, A., El-Assal, S. E.-D., and Abou-Bakr, S. (2020b). A single-step genomic evaluation for milk production in Egyptian buffalo. *Livestock Sci.* 234:103977. doi: 10.1016/j.livsci.2020.103977

Ahlberg, C. (2014). *Genetic Parameter Estimates and Breed Effects for Calving Difficulty and Birth Weight in a Multi-Breed Population*. Lincoln, NE: University of Nebraska.

Akanno, E. C., Plastow, G., Fitzsimmons, C., Miller, S. P., Baron, V., Ominski, K., et al. (2015). Genome-wide association for heifer reproduction and calf performance traits in beef cattle. *Genome* 58, 549–557. doi: 10.1139/gen-2015-0031

Ali, I., Muhammad Suhail, S., and Shafiq, M. (2019). Heritability estimates and genetic correlations of various production and reproductive traits of different grades of dairy cattle reared under subtropical condition. *Reprod. Domest. Anim.* 54, 1026–1033. doi: 10.1111/rda.13458

Al-Khuzai, H. M., Al-Dulaimi, M. K., and Zaini, Z. E. (2019). Estimation of heritability for reproductive traits and newborn mortality in iraqi buffaloes through the relationship of dams with daughters. *Indian J. Public Health Res. Dev.* 10, 660–663. doi: 10.5958/0976-5506.2019.02508.7

Allan, M., Kuehn, L., Cushman, R., Snelling, W., Echternkamp, S., and Thallman, R. (2014). Confirmation of quantitative trait loci using a low-density single nucleotide polymorphism map for twinning and ovulation rate on bovine chromosome 5. *J. Anim. Sci.* 87, 46–56. doi: 10.2527/jas.2008-0959

Allan, M., Thallman, R., Cushman, R., Echternkamp, S., White, S., Kuehn, L., et al. (2007). Association of a single nucleotide polymorphism in SPP1 with growth traits and twinning in a cattle population selected for twinning rate. *J. Anim. Sci.* 85, 341–347. doi: 10.2527/jas.2006-460

Amaral, M. E. J., Grant, J. R., Riggs, P. K., Stafuzza, N. B., Rodrigues Filho, E. A., Goldammer, T., et al. (2008). A first generation whole genome RH map of the river buffalo with comparison to domestic cattle. *BMC Genomics* 9:631. doi: 10.1186/1471-2164-9-631

Amaya-Martínez, A. A., Martínez, R., and Cerón-Muñoz, M. F. (2020). Genetic parameters for growth and reproduction in Simmental cattle from pedigree and genomic relationship. *Rev. MVZ Córdoba* 25:1520. doi: 10.21897/rmvz.1520

Ansari-Mahyari, S., Ojali, M. R., Forutan, M., Riasi, A., and Brito, L. F. (2019). Investigating the genetic architecture of conception and non-return rates in Holstein cattle under heat stress conditions. *Trop. Anim. Health Prod.* 51, 1847–1853. doi: 10.1007/s11250-019-01875-5

Arslan, K., Akyüz, B., Akçay, A., İlgar, E. G., Macun, H. C., and Çınar, M. U. (2017). Association of number of artificial inseminations per pregnancy in holstein dairy cows with polymorphism in luteinizing hormone receptor and follicle stimulating hormone receptor genes. *Slovenian Vet. Res.* 54, 91–98.

Barros, C. D. C., Aspilcueta-Borquis, R. R., Fraga, A. B., and Tonhati, H. (2016). Genetic parameter estimates for production and reproduction traits in dairy buffaloes. *Rev. Caat.* 29, 216–221. doi: 10.1590/1983-21252016v29n125rc

Bényei, B., Gáspárdy, A., Komlósi, I., and Pécsi, A. (2004). Repeatability and heritability of ovulation number and embryos in dam-daughters pairs in superovulated holstein–friesian cows. *Reprod. Domest. Anim.* 39, 99–102. doi: 10.1111/j.1439-0531.2004.00487.x

Berry, D. P., and Evans, R. (2014). Genetics of reproductive performance in seasonal calving beef cows and its association with performance traits. *J. Anim. Sci.* 92, 1412–1422. doi: 10.2527/jas.2013-6723

Berry, D. P., Wall, E., and Pryce, J. (2014). Genetics and genomics of reproductive performance in dairy and beef cattle. *Animal* 8, 105–121. doi: 10.1017/s1751731114000743

Boddhireddy, P., Prayaga, K., Barros, P., Lôbo, R., and Denise, S. (2014). "Genomic predictions of economically important traits in Nelore cattle of Brazil," in *Proceedings of the 10th World Congr. Genet. Appl. Livest. Prod*, Vancouver, BC.

Boichard, D. (1990). Estimation of the economic value of conception rate in dairy cattle. *Livestock Prod. Sci.* 24, 187–204. doi: 10.1016/0301-6226(90)90001-m

Boison, S., Utsunomiya, A., Santos, D., Neves, H., Carvalheiro, R., Mészáros, G., et al. (2017). Accuracy of genomic predictions in Gyr (Bos indicus) dairy cattle. *J. Dairy Sci.* 100, 5479–5490. doi: 10.3168/jds.2016-11811

Boligon, A., and Albuquerque, L. G. D. (2011). Genetic parameters and relationships of heifer pregnancy and age at first calving with weight gain, yearling and mature weight in Nelore cattle. *Livestock Sci.* 141, 12–16. doi: 10.1016/j.livsci.2011.04.009

Brzáková, M., Svitáková, A., Čítek, J., Veselá, Z., and Vostrý, L. (2019). Genetic parameters of longevity for improving profitability of beef cattle. *J. Anim. Sci.* 97, 19–28. doi: 10.1093/jas/sky390

Buzanskas, M. E., do Amaral Grossi, D., Ventura, R. V., Schenkel, F. S., Chud, T. C. S., Stafuzza, N. B., et al. (2017). Candidate genes for male and female reproductive traits in Canchim beef cattle. *J. Anim. Sci. Biotechnol.* 8:67.

Camargo, G. D., Aspilcueta-Borquis, R. R., Fortes, M., Porto-Neto, R., Cardoso, D. F., Santos, D., et al. (2015). Prospecting major genes in dairy buffaloes. *BMC Genomics* 16:872. doi: 10.1186/s12864-015-1986-2

Cammack, K., Thomas, M., and Enns, R. (2009). Reproductive traits and their heritabilities in beef cattle. *Prof. Anim. Sci.* 25, 517–528. doi: 10.15232/s1080-7446(15)30753-1

Canaza-Cayo, A. W., Lopes, P. S., Cobuci, J. A., Martins, M. F., and Silva, M. V. G. B. D. (2018). Genetic parameters of milk production and reproduction traits of Girolando cattle in Brazil. *Ital. J. Anim. Sci.* 17, 22–30. doi: 10.1080/1828051x.2017.1335180

Cassell, B. G. (2009). *Using Heritability for Genetic Improvement*. Blacksburg: Virginia Cooperative Extension.

Cervantes, I., Gutiérrez, J. P., Fernández, I., and Goyache, F. (2010). Genetic relationships among calving ease, gestation length, and calf survival to weaning in the Asturiana de los Valles beef cattle breed. *J. Anim. Sci.* 88, 96–101. doi: 10.2527/jas.2009-2066

Changhee, D., Nidarshani, W., Kwanghyun, C., Yunho, C., Taejeong, C., Byungho, P., et al. (2013). The effect of age at first calving and calving interval on productive life and lifetime profit in korean holsteins. *Asian Austral. J. Anim. Sci.* 26, 1511–1517. doi: 10.5713/ajas.2013.13105

Chegini, A., Hossein-Zadeh, N. G., Moghaddam, S. H. H., and Shadparvar, A. A. (2019a). Genetic aspects of some reproductive, udder health and energy status traits in Holstein cows. *Theriogenology* 130, 1–7. doi: 10.1016/j.theriogenology.2019.02.027

Chegini, A., Shadparvar, A. A., Hossein-Zadeh, N. G., and Mohammad-Nazari, B. (2019b). Genetic and environmental relationships among milk yield, persistency of milk yield, somatic cell count and calving interval in Holstein cows. *Rev. Colomb. Ciencias Pecuarias* 32, 81–89. doi: 10.17533/udea.rccp.v32n2a01

Cochran, S. D., Cole, J. B., Null, D. J., and Hansen, P. J. (2013). Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. *BMC Genet.* 14:49. doi: 10.1186/1471-2156-14-49

Costa, E. V., Ventura, H. T., Veroneze, R., Silva, F. F., Pereira, M. A., and Lopes, P. S. (2020). Estimated genetic associations among reproductive traits in Nellore cattle using Bayesian analysis. *Anim. Reprod. Sci.* 214:106305. doi: 10.1016/j.anireprosci.2020.106305

Costa, R. B., Irano, N., Diaz, I. D. P. S., Takada, L., Da Costa Hermisdorff, I., Carvalheiro, R., et al. (2019). Prediction of genomic breeding values for reproductive traits in nellore heifers. *Theriogenology* 125, 12–17. doi: 10.1016/j.theriogenology.2018.10.014

Cribiu, E. P., Di Berardino, D., Di Meo, G. P., Eggen, A., Gallagher, D. S., Gustavsson, I., et al. (2001). International system for chromosome nomenclature of domestic Bovids (ISCNDB 2000). *Cytogenet. Cell Genet.* 92, 283–299. doi: 10.1159/000056917

Darvasi, A., Weinreb, A., Minke, V., Weller, J., and Soller, M. (1993). Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* 134, 943–951. doi: 10.1093/genetics/134.3.943

de Araujo Neto, F. R., Takada, L., Dos Santos, D. J. A., Aspilcueta-Borquis, R. R., Cardoso, D. F., Do Nascimento, A. V., et al. (2020). Identification of genomic regions related to age at first calving and first calving interval in water buffalo using single-step GBLUP. *Reprod. Domest. Anim.* 55, 1565–1572. doi: 10.1111/rda.13811

de Rezende, M. P. G., Malhado, C. H. M., Biffani, S., Carneiro, P. L. S., Carrillo, J. A., and Bozzi, R. (2020). Genotype-environment interaction for age at first calving in Limousine and Charolais cattle raised in Italy, employing reaction norm model. *Livestock Sci.* 232:103912. doi: 10.1016/j.livsci.2019.103912

Dekkers, J. (1991). Estimation of economic values for dairy cattle breeding goals: bias due to sub-optimal management policies. *Livestock Prod. Sci.* 29, 131–149. doi: 10.1016/0301-6226(91)90062-u

do Amaral Grossi, D., Berton, M. P., Buzanskas, M. E., Chud, T. C. S., Grupioni, N. V., De Paz, C. C. P., et al. (2016). Genetic analysis on accumulated productivity and calving intervals in Nelore cattle. *Trop. Anim. Health Prod.* 48, 207–210. doi: 10.1007/s11250-015-0915-3

Doyle, S., Golden, B., Green, R., and Brinks, J. (2000). Additive genetic parameter estimates for heifer pregnancy and subsequent reproduction in Angus females. *J. Anim. Sci.* 78, 2091–2098. doi: 10.2527/2000.7882091x

Du, C., Deng, T., Zhou, Y., Ye, T., Zhou, Z., Zhang, S., et al. (2019). Systematic analyses for candidate genes of milk production traits in water buffalo (Bubalus Bubalis). *Anim. Genet.* 50, 207–216. doi: 10.1111/age.12739

Effa, K., Wondatir, Z., Dessie, T., and Haile, A. (2011). Genetic and environmental trends in the long-term dairy cattle genetic improvement programmes in the central tropical highlands of Ethiopia. *J. Cell Anim. Biol.* 5, 96–104.

Egger-Danner, C., Kadlečík, O., Fuerst, C., and Kasarda, R. (2005). "Joint genetic evaluation for functional longevity for Pinzgau cattle," in *Proceedings of the 56th Annual Meeting of the European Association for Animal Production*, Uppsala.

El-Bramony, M. M. (2011). Genetic and phenotypic parameters of milk yield and reproductive performance in the first three lactations of Egyptian buffalo. *Egypt. J. Anim. Prod.* 48, 1–10. doi: 10.21608/ejap.2011.94365

El-Bramony, M. M., and Reclamation, D. (2014). Estimation of genetic and phenotypic parameters for milk yield, lactation length, calving interval and body weight in the first lactation of Egyptian buffalo. *Life Sci. J.* 11, 1012–1019.

Elzo, M., Mateescu, R., Rae, D., Carr, C., Scheffler, T., Scheffler, J., et al. (2018). "Genomic-polygenic EBV for reproduction, ultrasound-carcass, and tenderness traits in the Florida multibreed Brahman-Angus population," in *Proceedings of the World Congress on Genetics Applied to Livestock Production*, Rome, 3–7.

Engle, B. N., Corbet, N. J., Allen, J. M., Laing, A. R., Fordyce, G., Mcgowan, M. R., et al. (2019). Multivariate genomic predictions for age at puberty in tropically adapted beef heifers. *J. Anim. Sci.* 97, 90–100. doi: 10.1093/jas/sky428

Faraji-Arough, H., and Rokouei, M. (2016). Bayesian inference of genetic parameters for reproductive traits in Sistani native cows using Gibbs sampling. *J. Livestock Sci. Technol.* 4, 39–49.

Fernando, R., and Grossman, M. (1989). Marker assisted selection using best linear unbiased prediction. *Genet. Select. Evol.* 21, 1–11.

Forni, S., and Albuquerque, L. (2005). Estimates of genetic correlations between days to calving and reproductive and weight traits in Nelore cattle. *J. Anim. Sci.* 83, 1511–1515. doi: 10.2527/2005.8371511x

Fortes, M., Lehnert, S., Bolormaa, S., Reich, C., Fordyce, G., Corbet, N., et al. (2012). Finding genes for economically important traits: Brahman cattle puberty. *Anim. Product. Sci.* 52, 143–150. doi: 10.1071/an11165

Fortes, M. R., Reverter, A., Kelly, M., Mcculloch, R., and Lehnert, S. A. (2013). Genome-wide association study for inhibin, luteinizing hormone, insulin-like growth factor 1, testicular size and semen traits in bovine species. *Andrology* 1, 644–650. doi: 10.1111/j.2047-2927.2013.00101.x

Forutan, M., Mahyari, S., Schenkel, F., and Sargolzaei, M. (2018). Improving genomic evaluation of Holstein cattle using a haplotype-based relationship matrix. *Iran. J. Anim. Sci. Res.* 10, 393–402.

Gaddis, K. L. P., Null, D. J., and Cole, J. B. (2016). Explorations in genome-wide association studies and network analyses with dairy cattle fertility traits. *J. Dairy Sci.* 99, 6420–6435. doi: 10.3168/jds.2015-10444

Gaddis, K. P., Dikmen, S., Null, D., Cole, J., and Hansen, P. (2017). Evaluation of genetic components in traits related to superovulation, in vitro fertilization, and embryo transfer in Holstein cattle. *J. Dairy Sci.* 100, 2877–2891. doi: 10.3168/jds.2016-11907

Galsar, N. S., Shah, R., Pandey, J. P. D., and Prajapati, K. (2016). 7. ANALYSIS OF FIRST PRODUCTION AND REPRODUCTION TRAITS OF MEHSANA BUFFALOES MAINTAINED AT TROPICAL AND SEMI-ARID REGION OF GUJARAT, INDIA by NIRALI S. GALSAR 1, RR SHAH 2, JAY PRAKASH GUPTA 3, DP PANDEY 4, K. B. *Life Sci. Leaflets* 77, 65–75.

García-Ruiz, A., Cole, J. B., Vanraden, P. M., Wiggans, G. R., Ruiz-López, F. J., and Van Tassell, C. P. (2016). Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc. Natl. Acad. Sci. U.S.A.* 113, E3995–E4004.

Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A. T., et al. (1995). Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* 139, 907–920. doi: 10.1093/genetics/139.2.907

Goddard, M. (1996). The use of marker haplotypes in animal breeding schemes. *Genet. Select. Evol.* 28, 161–176. doi: 10.1186/1297-9686-28-2-161

Gonzálezrecio, O., and Alenda, R. (2005). Genetic parameters for female fertility traits and a fertility index in Spanish dairy cattle. *J. Dairy Sci.* 88, 3282–3289. doi: 10.3168/jds.s0022-0302(05)73011-3

Goodling, R. Jr., Shook, G., Weigel, K., and Zwald, N. (2005). The effect of synchronization on genetic parameters of reproductive traits in dairy cattle. *J. Dairy Sci.* 88, 2217–2225. doi: 10.3168/jds.s0022-0302(05)72897-6

Goyache, F., and Gutiérrez, J. P. (2001). Heritability of reproduetive traits in Asturiana de los Valles beef cattle breed. *Arch. Anim. Breed.* 44, 489–496. doi: 10.5194/aab-44-489-2001

Goyache, F., Gutiérrez, J. P., Fernández, I., Royo, L., and Álvarez, I. (2005). Genetic analysis of days open in beef cattle. *Livestock Prod. Sci.* 93, 283–289. doi: 10.1016/j.livprodsci.2004.10.002

Guarini, A., Lourenco, D., Brito, L., Sargolzaei, M., Baes, C. F., Miglior, F., et al. (2018). Comparison of genomic predictions for lowly heritable traits using multi-step and single-step genomic best linear unbiased predictor in Holstein cattle. *J. Dairy Sci.* 101, 8076–8086. doi: 10.3168/jds.2017-14193

Gupta, J. P., Sachdeva, G. K., Gandhi, R., and Chakaravarty, A. (2015). Developing multiple-trait prediction models using growth and production traits in Murrah buffalo. *Buffalo Bull.* 34, 347–355.

Gutiérrez, J. P., Álvarez, I., Fernández, I., Royo, L., Dıez, J., and Goyache, F. (2002). Genetic relationships between calving date, calving interval, age at first calving and type traits in beef cattle. *Livestock Prod. Sci.* 78, 215–222. doi: 10.1016/s0301-6226(02)00100-8

Heise, J., Stock, K. F., Reinhardt, F., Ha, N. T., and Simianer, H. (2017). Phenotypic and genetic relationships between age at first calving, its component traits, and survival of heifers up to second calving. *J. Dairy Sci.* 101:S0022030217309967.

Henderson, C. R. (1984). *Applications of Linear Models in Animal Breeding*. University of Guelph.

Hossein Salimi, M., Hossein-Zadeh, N. G., Shadparvar, A. A., and Eghbal, A. R. (2017). Genetic evaluation of dystocia and its relationship with productive and reproductive traits in Holstein cows. *Rev. Colomb. Ciencias Pecuarias* 30, 126–137. doi: 10.17533/udea.rccp.v30n2a04

Hu, Z.-L., Park, C. A., and Reecy, J. M. (2019). Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Res.* 47, D701–D710.

Jaton, C., Koeck, A., Sargolzaei, M., Malchiodi, F., Price, C., Schenkel, F., et al. (2016a). Genetic analysis of superovulatory response of Holstein cows in Canada. *J. Dairy Sci.* 99, 3612–3623. doi: 10.3168/jds.2015-10349

Jaton, C., Koeck, A., Sargolzaei, M., Price, C., Baes, C., Schenkel, F., et al. (2016b). Genetic correlations between number of embryos produced using in vivo and in vitro techniques in heifer and cow donors. *J. Dairy Sci.* 99, 8222–8226. doi: 10.3168/jds.2016-11356

Jaton, C., Schenkel, F., Chud, T., Malchiodi, F., Sargolzaei, M., Price, C., et al. (2020). Genetic and genomic analyses of embryo production in dairy cattle. *Reprod. Fertil. Dev.* 32, 50–55. doi: 10.1071/rd19275

Johnston, D., Barwick, S., Corbet, N., Fordyce, G., Holroyd, R., Williams, P., et al. (2009). Genetics of heifer puberty in two tropical beef genotypes in northern Australia and associations with heifer-and steer-production traits. *Anim. Prod. Sci.* 49, 399–412. doi: 10.1071/ea08276

Kadarmideen, H., and Simm, G. (2002). "Selection responses expected from index selection including disease resistance, fertility and longevity in dairy cattle," in *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, Montpellier, 01–19.

Kale, D., Yadav, B., and Prasad, J. (2014). DNA polymorphisms at candidate gene loci and their relation with milk production traits in Murrah Buffalo (Bubalus bubalis) *Iran. J. Appl. Anim. Sci.* 4, 39–43.

Kaupe, B., Brandt, H., Prinzenberg, E., and Erhardt, G. (2007). Joint analysis of the influence of CYP11B1 and DGAT1 genetic variation on milk production, somatic cell score, conformation, reproduction, and productive lifespan in German Holstein cattle. *J. Anim. Sci.* 85, 11–21. doi: 10.2527/jas.2005-753

Ke, H., Iqbal, A., and Jj, K. (2014). A genome wide association study on age at first calving using high density single nucleotide polymorphism chips in Hanwoo (*Bos taurus* coreanae). *Asian Austral. J. Anim. Sci.* 27, 1406–1410. doi: 10.5713/ajas.2014.14273

Kluska, S., Olivieri, B. F., Bonamy, M., Chiaia, H. L. J., Feitosa, F. L. B., Berton, M. P., et al. (2018). Estimates of genetic parameters for growth, reproductive, and carcass traits in Nelore cattle using the single step genomic BLUP procedure. *Livestock Sci.* 216, 203–209. doi: 10.1016/j.livsci.2018.08.015

König, S., Bosselmann, F., and Von Borstel, U. H. S. (2007). Genetic analysis of traits affecting the success of embryo transfer in dairy cattle. *J. Dairy Sci.* 90, 3945–3954. doi: 10.3168/jds.2007-0089

Konkruea, T., Koonawootrittriron, S., Elzo, M. A., and Suwanasopee, T. (2019). Accuracy of genomic-polygenic and polygenic breeding values for age at first calving and milk yield in thai multibreed dairy cattle. *Ann. Anim. Sci.* 19, 633–645. doi: 10.2478/aoas-2019-0032

Kumar, V., Chakravarty, A., Patil, C., Valsalan, J., and Mahajan, A. (2015). Estimate of genetic and non-genetic parameters for age at first calving in Murrah buffalo. *Indian J. Anim. Sci.* 85, 84–85.

Laodim, T., Elzo, M. A., Koonawootrittriron, S., Suwanasopee, T., and Jattawa, D. (2019). Genomic-polygenic and polygenic predictions for milk yield, fat yield, and age at first calving in Thai multibreed dairy population using genic and functional sets of genotypes. *Livestock Sci.* 219, 17–24. doi: 10.1016/j.livsci.2018.11.008

Lázaro, S. F., Varona, L., Silva, F. F., Ventura, H. T., Veroneze, R., Brito, L. C., et al. (2019). Censored Bayesian models for genetic evaluation of age at first calving in Brazilian Brahman cattle. *Livestock Sci.* 221, 177–180. doi: 10.1016/j.livsci.2018.11.014

Lee, D., and Han, K. (2004). Genetic relationship between milk production, calving ease and days open at first parity in Holstein cows. *Asian Austral. J. Anim. Sci.* 17, 153–158. doi: 10.5713/ajas.2004.153

Lett, B. M., and Kirkpatrick, B. W. (2018). Short communication: heritability of twinning rate in Holstein cattle. *J. Dairy Sci.* 101, 4307-4311.

Li, J., Liu, J., Campanile, G., Plastow, G., Zhang, C., Wang, Z., et al. (2018a). Novel insights into the genetic basis of buffalo reproductive performance. *BMC Genomics* 19:814. doi: 10.1186/s12864-018-5208-6

Li, J., Liu, J., Liu, S., Plastow, G., Zhang, C., Wang, Z., et al. (2018b). Integrating RNA-seq and GWAS reveals novel genetic mutations for buffalo reproductive traits. *Anim. Reprod. Sci.* 197, 290–295. doi: 10.1016/j.anireprosci.2018.08.041

Liu, J. J., Liang, A. X., Campanile, G., Plastow, G., Zhang, C., Wang, Z., et al. (2018). Genome-wide association studies to identify quantitative trait loci affecting milk production traits in water buffalo. *J. Dairy Sci.* 101, 433–444. doi: 10.3168/jds.2017-13246

Lopes, F., Wu, X. L., Li, H., Xu, J., Perkins, T., Genho, J., et al. (2018). Improving accuracy of genomic prediction in Brangus cattle by adding animals with imputed low-density SNP genotypes. *J. Anim. Breed. Genet.* 135, 14–27. doi: 10.1111/jbg.12312

Lopez, B. I., Son, J.-H., Seo, K., and Lim, D. (2019). Estimation of genetic parameters for reproductive traits in Hanwoo (Korean Cattle). *Animals* 9:715. doi: 10.3390/ani9100715

Low, W. Y., Tearle, R., Bickhart, D. M., Rosen, B. D., Kingan, S. B., Swale, T., et al. (2019). Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity. *Nat. Commun.* 10, 1–11.

Luan, T., Woolliams, J. A., Lien, S., Kent, M., Svendsen, M., and Meuwissen, T. H. (2009). The accuracy of genomic selection in Norwegian red cattle assessed by cross-validation. *Genetics* 183, 1119–1126. doi: 10.1534/genetics.109.107391

Macgregor, R. (1941). The domestic buffalo. *Vet. Rec.* 53, 443–450. doi: 10.7313/upo9781907284991.034

Makgahlela, M., Banga, C., Norris, D., Dzama, K., and Ngambi, J. (2008). Genetic analysis of age at first calving and calving interval in South African Holstein cattle. *Asian J. Anim. Vet. Adv.* 3, 197–205. doi: 10.3923/ajava.2008.197.205

Martínez-Velázquez, G., Ríos-Utrera, A., Román-Ponce, S., Baeza-Rodríguez, J., Arechavaleta-Velasco, M., Montaño-Bermúdez, M., et al. (2020). Genetic correlations between scrotal circumference, heifer fertility and stayability in Charolais–Charbray cattle. *Livestock Sci.* 232:103914. doi: 10.1016/j.livsci.2019.103914

Maryam, K., Beiginassiri, M., Nejad, A. N., Chaji, M., Roshanfekr, H., and Nazari, B. M. (2016). Genetic parameter estimation of dystocia variable in iranian holstein dairy cattle. *Basrah J. Vet. Res.* 15:4.

McClure, M., Morsci, N., Schnabel, R., Kim, J., Yao, P., Rolf, M., et al. (2010). A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. *Anim. Genet.* 41, 597–607. doi: 10.1111/j.1365-2052.2010.02063.x

Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

Michelizzi, V. N., Dodson, M. V., Pan, Z., Amaral, M. E. J., Michal, J. J., Mclean, D. J., et al. (2010). Water buffalo genome science comes of age. *Int. J. Biol. Sci.* 6:333. doi: 10.7150/ijbs.6.333

Miller, J., Thomsen, P., Dixon, S., Tucker, E., Konfortov, B., and Harbitz, I. (1992). Synteny mapping of the bovine IGHG2, CRC and IGF1 genes. *Anim. Genet.* 23, 51–58. doi: 10.1111/j.1365-2052.1992.tb00017.x

Mintoo, A. A., Zhang, H., Chen, C., Moniruzzaman, M., Deng, T., Anam, M., et al. (2019). Draft genome of the river water buffalo. *Ecol. Evol.* 9, 3378–3388. doi: 10.1002/ece3.4965

Moioli, B., Steri, R., Marchitelli, C., Catillo, G., and Buttazzoni, L. (2017). Genetic parameters and genome-wide associations of twinning rate in a local breed, the Maremmana cattle. *Animal* 11, 1660–1666. doi: 10.1017/s1751731117000283

Montaldo, H., Trejo, C., and Lizana, C. (2017). Genetic parameters for milk yield and reproduction traits in the Chilean Dairy Overo Colorado cattle breed. *Ciencia Investig. Agraria* 44, 24–34.

Morammazi, S., Torshizi, R., Rouzbehan, Y., and Sayyadnejad, M. (2007). PosterEstimates of genetic parameters for production and reproduction traits in Khuzestan buffalos of Iran. *Italian J. Anim. Sci.* 6, 421–424. doi: 10.4081/ijas.2007.s2.421

Morris, C., Amyes, N., and Hickey, S. (2011). Responses of prolactin and hair growth to selection for age at puberty in Angus cattle. *Animal* 5, 198–201. doi: 10.1017/s1751731110001825

Morris, C., Wilson, J., Bennett, G., Cullen, N., Hickey, S., and Hunter, J. (2000). Genetic parameters for growth, puberty, and beef cow reproductive traits in a puberty selection experiment. *N. Zeal. J. Agric. Res.* 43, 83–91. doi: 10.1080/00288233.2000.9513411

Mota, R., Guimarães, S., Fortes, M., Hayes, B., Silva, F., Verardo, L., et al. (2017). Genome-wide association study and annotating candidate gene networks affecting age at first calving in Nellore cattle. *J. Anim. Breed. Genet.* 134, 484–492. doi: 10.1111/jbg.12299

Mota, R. R., E Silva, F. F., Guimarães, S. E. F., Hayes, B., Fortes, M. R. S., Kelly, M. J., et al. (2018). Benchmarking Bayesian genome enabled-prediction models for age at first calving in Nellore cows. *Livestock Sci.* 211, 75–79. doi: 10.1016/j.livsci.2018.03.009

Müller, M. P., Rothammer, S., Seichter, D., Russ, I., Hinrichs, D., Tetens, J., et al. (2017). Genome-wide mapping of 10 calving and fertility traits in Holstein dairy cattle with special regard to chromosome 18. *J. Dairy Sci.* 100, 1987–2006. doi: 10.3168/jds.2016-11506

Nascimento, A. V., Matos, M. C., Seno, L. O., Romero, A. R., Garcia, J. F., and Grisolia, A. B. (2016). Genome wide association study on early puberty in *Bos indicus*. *Genet. Mol. Res.* 15, 1–6.

Nayeri, S., Sargolzaei, M., Abo-Ismail, M. K., May, N., Miller, S. P., Schenkel, F., et al. (2016). Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. *BMC Genet.* 17:75. doi: 10.1186/s12863-016-0386-1

Neves, H. H., Carvalheiro, R., O'brien, A. M. P., Utsunomiya, Y. T., Do Carmo, A. S., Schenkel, F. S., et al. (2014). Accuracy of genomic predictions in *Bos indicus* (Nellore) cattle. *Genet. Select. Evol.* 46:17. doi: 10.1186/1297-9686-46-17

Oyama, K., Katsuta, T., Anada, K., and Mukai, F. (2002). Heritability and repeatability estimates for reproductive traits of Japanese Black cows. *Asian Austral. J. Anim. Sci.* 15, 1680–1685. doi: 10.5713/ajas.2002.1680

Peixoto, M., Pereira, C., Bergmann, J., Penna, V., and Fonseca, C. (2004). Genetic parameters of multiple ovulation traits in Nellore females. *Theriogenology* 62, 1459–1464. doi: 10.1016/j.theriogenology.2004.02.019

Piccoli, M. L., Brito, L. F., Braccini, J., Oliveira, H. R., Cardoso, F. F., Roso, V. M., et al. (2020). Comparison of genomic prediction methods for evaluation of adaptation and productive efficiency traits in Braford and Hereford cattle. *Livestock Sci.* 231:103864. doi: 10.1016/j.livsci.2019.103864

Piper, L., Bindon, B., Swan, A., and Brewer, H. (2017). Genetic selection for litter size in cattle. *Proc. Assoc. Advmt. Breed. Genet.* 21, 101–105.

Rathod, A., Vaidya, M., and Ali, S. S. (2018). Genetic studies of productive and reproductive attributes of surti buffalo in Maharashtra. *Int. J. Livestock Res.* 8, 309–314. doi: 10.5455/ijlr.20171016061752

Roxström, A., and Strandberg, E. (2002). Genetic analysis of functional, fertility-, mastitis-, and production-determined length of productive life in Swedish dairy cattle. *Livestock Prod. Sci.* 74, 125–135. doi: 10.1016/s0301-6226(01)00300-1

Saatchi, M., McClure, M. C., Mckay, S. D., Rolf, M. M., Kim, J., Decker, J. E., et al. (2011). Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. *Genet. Select. Evol.* 43:40. doi: 10.1186/1297-9686-43-40

Saowaphak, P., Duangjinda, M., Plaengkaeo, S., Suwannasing, R., and Boonkum, W. (2017). Genetic correlation and genome-wide association study (GWAS) of the length of productive life, days open, and 305-days milk yield in crossbred Holstein dairy cattle. *Genet. Mol. Res.* 16:16029091.

Schmidt, P., Campos, G., Roso, V., Souza, F., and Boligon, A. (2019). Genetic analysis of female reproductive efficiency, scrotal circumference and growth traits in Nelore cattle. *Theriogenology* 128, 47–53. doi: 10.1016/j.theriogenology.2019.01.032

Seno, L. O., Cardoso, V. L., El-Faro, L., Sesana, R. C., Aspilcueta-Borquis, R. R., Camargo, G. M. F. D., et al. (2010). Genetic parameters for milk yield, age at first calving and interval between first and second calving in milk buffaloes. *Ital. J. Anim. Sci.* 6, 397–400. doi: 10.4081/ijas.2007.s2.397

Setiaji, A., and Oikawa, T. (2019). Genetic parameters of reproductive traits from artificial insemination records of Japanese Black cows. *Livestock Sci.* 229, 85–89. doi: 10.1016/j.livsci.2019.09.018

Silva, M. V., Dos Santos, D. J., Boison, S. A., Utsunomiya, A. T., Carmo, A. S., Sonstegard, T. S., et al. (2014). The development of genomics applied to dairy breeding. *Livestock Sci.* 166, 66–75.

Silvestre, A., Martins, Â, Santos, V., and Colaço, J. (2019). Genetic parameters of calving ease in dairy cattle using threshold and linear models. *Ital. J. Anim. Sci.* 18, 80–87. doi: 10.1080/1828051x.2018.1482801

Tang, K.-Q., Li, S.-J., Yang, W.-C., Yu, J.-N., Han, L., Li, X., et al. (2011). An MspI polymorphism in the inhibin alpha gene and its associations with superovulation traits in Chinese Holstein cows. *Mol. Biol. Rep.* 38, 17–21. doi: 10.1007/s11033-010-0072-8

Tarekegn, G., Gullstrand, P., Strandberg, E., Båge, R., Rius-Vilarrasa, E., Christensen, J., et al. (2019). Genetic parameters of endocrine fertility traits based on in-line milk progesterone profiles in Swedish Red and Holstein dairy cows. *J. Dairy Sci.* 102, 11207–11216. doi: 10.3168/jds.2019-16691

Thiruvenkadan, A., Panneerselvam, S., Rajendran, R., and Murali, N. (2010). Analysis on the productive and reproductive traits of Murrah buffalo cows maintained in the coastal region of India. *Appl. Anim. Husbandry Rural Dev.* 3, 1–5. doi: 10.21608/jpd.2006.45174

Tiezzi, F., Arceo, M. E., Cole, J. B., and Maltecca, C. (2018). Including gene networks to predict calving difficulty in holstein, brown swiss and jersey cattle. *BMC Genet.* 19:20. doi: 10.1186/s12863-018-0606-y

Toghiani, S. (2012). Genetic relationships between production traits and reproductive performance in Holstein dairy cows. *Arch. Anim. Breed.* 55, 458–468. doi: 10.5194/aab-55-458-2012

Toghiani, S., Hay, E., Sumreddee, P., Geary, T., Rekaya, R., and Roberts, A. (2017). Genomic prediction of continuous and binary fertility traits of females in a composite beef cattle breed. *J. Anim. Sci.* 95, 4787–4795. doi: 10.2527/jas2017.1944

Tramonte, N. C., Grupioni, N. V., Stafuzza, N. B., Guidolin, D. G. F., Savegnago, R. P., Bezerra, L. A. F., et al. (2019). Genetic parameters, genetic trends, and principal component analysis for productive and reproductive traits of Guzera beef cattle. *Rev. Bras. Zootecnia* 48:34.

Ulbrich, F., and Fischer, H. (1966). *The Chromosomes of the Asiatic Buffalo (Bubalus bubalis) and the African Buffalo (Cyncerus caffer)*. Hoboken, NJ: Wiley Online Library.

van der Linde, C., De Jong, G., Simai, S., Gombacsi, P., and Wellisch, P. (2006). Genetic evaluation for longevity in Hungary. *Interbull. Bull.* 3:35.

Vanderick, S., Troch, T., Gillon, A., Glorieux, G., and Gengler, N. (2015). Genetic parameters for direct and maternal calving ease in Walloon dairy cattle based on linear and threshold models. *J. Anim. Breed. Genet.* 131, 513–521. doi: 10.1111/jbg.12105

Vinothraj, S., Subramaniyan, A., Venkataramanan, R., Joseph, C., and Sivaselvam, S. (2016). Genetic evaluation of reproduction performance of Jersey× Red Sindhi crossbred cows. *Vet. World* 9:1012. doi: 10.14202/vetworld.2016.1012-1017

Warburton, C. L., Engle, B. N., Ross, E. M., Costilla, R., Moore, S. S., Corbet, N. J., et al. (2020). Use of whole-genome sequence data and novel genomic selection strategies to improve selection for age at puberty in tropically-adapted beef heifers. *Genet. Select. Evol.* 52, 1–13.

Weller, J., Ezra, E., and Ron, M. (2017). Invited review: A perspective on the future of genomic selection in dairy cattle. *J. Dairy Sci.* 100, 8633–8644. doi: 10.3168/jds.2017-12879

Weller, J. I., Golik, M., Seroussi, E., Ron, M., and Ezra, E. (2008). Detection of quantitative trait loci affecting twinning rate in Israeli Holsteins by the daughter design. *J. Dairy Sci.* 91, 2469–2474. doi: 10.3168/jds.2007-0915

Williams, J. L., Iamartino, D., Pruitt, K. D., Sonstegard, T., Smith, T. P., Low, W. Y., et al. (2017). Genome assembly and transcriptome resource for river buffalo, Bubalus bubalis (2 n= 50). *Gigascience* 6:gix088.

Yang, W. C., Yang, L. G., Riaz, H., Tang, K. Q., Chen, L., and Li, S. J. (2013). Effects in cattle of genetic variation within the IGF1R gene on the superovulation performance and pregnancy rates after embryo transfer. *Anim. Reprod. Sci.* 143, 24–29. doi: 10.1016/j.anireprosci.2013.10.008

Yutaka, M., Toshimi, B., and Mitsuyoshi, S. (2015). Genetic analysis of twinning rate and milk yield using a threshold-linear model in Japanese Holsteins. *Anim. Sci. J.* 86, 31–36. doi: 10.1111/asj.12236

Zhang, Y., Johnston, D., Bolormaa, S., Hawken, R., and Tier, B. (2014). Genomic selection for female reproduction in Australian tropically adapted beef cattle. *Anim. Prod. Sci.* 54, 16–24. doi: 10.1071/an13016

Zhang, Z., Kargo, M., Liu, A., Thomasen, J. R., Pan, Y., and Su, G. (2019). Genotype-by-environment interaction of fertility traits in Danish Holstein cattle using a single-step genomic reaction norm model. *Heredity* 123, 202–214. doi: 10.1038/s41437-019-0192-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 26 March 2021
doi: 10.3389/fgene.2021.626158

lncSAMM50 Enhances Adipogenic Differentiation of Buffalo Adipocytes With No Effect on Its Host Gene

Ruirui Zhu[1†], Xue Feng[2†], Yutong Wei[2], Duo Guo[1], Jiaojiao Li[1], Qingyou Liu[1], Jianrong Jiang[1], Deshun Shi[1] and Jieping Huang[1*]*

[1] State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangxi University, Nanning, China, [2] College of Life Sciences, Xinyang Normal University, Xinyang, China

OPEN ACCESS

Edited by:
Guohua Hua,
Huazhong Agricultural University,
China

Reviewed by:
Andres Contreras,
Michigan State University,
United States
Alfredo Pauciullo,
University of Turin, Italy
Faizul Hassan,
University of Agriculture, Faisalabad,
Pakistan

***Correspondence:**
Jieping Huang
huangjieping1989@126.com
Deshun Shi
ardsshi@gxu.edu.cn;
ardsshi@163.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: 05 November 2020
Accepted: 22 February 2021
Published: 26 March 2021

Citation:
Zhu R, Feng X, Wei Y, Guo D, Li J, Liu Q, Jiang J, Shi D and Huang J (2021) lncSAMM50 Enhances Adipogenic Differentiation of Buffalo Adipocytes With No Effect on Its Host Gene. Front. Genet. 12:626158. doi: 10.3389/fgene.2021.626158

Fat deposition is one of the most important traits that are mediated by a set of complex regulatory factors in meat animals. Several researches have revealed the significant role of long non-coding RNAs (lncRNAs) in fat deposition while the precise regulatory mechanism is still largely elusive. In this study, we investigated the lncRNA profiles of adipose and muscle tissues in buffalo by using the Illumina HiSeq 3000 platform. In total, 43,809 lncRNAs were finally identified based on the computer algorithm. A comparison analysis revealed 241 lncRNAs that are differentially expressed (DE) in adipose and muscle tissues. We focused on lncSAMM50, a DE lncRNA that has a high expression in adipose tissue. Sequence alignment showed that lncSAMM50 is transcribed from the antisense strand of the upstream region of sorting and assembly machinery component 50 homolog (*SAMM50*), a gene involved in the function of mitochondrion and is subsequently demonstrated to inhibit the adipogenic differentiation of 3T3-L1 adipocyte cells in this study. lncSAMM50 is highly expressed in adipose tissue and upregulated in the mature adipocytes and mainly exists in the nucleus. Gain-of-function experiments demonstrated that lncSAMM50 promotes the adipogenic differentiation by upregulating adipogenic markers but with no effect on its host gene *SAMM50* in buffalo adipocytes. These results indicate that lncSAMM50 enhances fat deposition in buffalo and provide a new factor for the regulatory network of adipogenesis.

Keywords: ***Bubalus bubalis*, adipose, RNA sequencing, long non-coding RNA, adipogenesis**

INTRODUCTION

The buffalo (*Bubalus bubalis*) is a globally important domestic animal providing economic value from meat, milk, and draft power. In China, the number of buffaloes is more than 27 million, second only to India and Pakistan (FAO, http://www.fao.org/, 2019). Traditionally, buffaloes are raised for draught power in China. Recently, with the increasing agricultural mechanization, the utility of buffaloes in draught power has gradually decreased, indicating that the role of buffaloes can be changed into a meat source (Kiran and Naveena, 2014). The fat deposition level in Chinese buffalo is very low due to the long-term breeding for draught power. However, both backfat thickness and intramuscular fat (IMF) content, which are associated with fat deposition, are vital traits for meat animals as buffalo. Especially, IMF content is highly correlated with tenderness, juiciness, and flavor of buffalo meat. A lower backfat thickness and a higher IMF content are of benefit to beef production. However, it is nearly impossible to decrease the backfat thickness and to increase the IMF deposition at the same time, indicating that the regulatory mechanism of fat deposition is far from complete to be understood, as new regulatory factors need to be discovered.

In animals, excess energy is stored as triglycerides within the lipid droplets of adipocytes and then expressed as fat deposition. Adipogenesis is the process of cell differentiation from preadipocytes to mature adipocytes, with lipid accumulation in cells. This process has been widely studied for decades. Researches *in vitro* and *in vivo* show that adipogenesis is a highly complex process that can be regulated by a large number of factors (Lowe et al., 2011; Mota de Sá et al., 2017). Peroxisome proliferator-activated receptor gamma (PPARγ or PPARG) is the most well-studied one and is undoubtedly the most significant modulator in adipogenesis of animals (Lowe et al., 2011; Mota de Sá et al., 2017). Many other factors, such as the CCAAT/enhancer-binding protein family (C/EBPs; Cao et al., 1991; Yeh et al., 1995; Hamm et al., 2001), Kruppel-like transcription factors (KLFs; Mori et al., 2005; Oishi et al., 2005; Birsoy et al., 2008), and GATA transcription factors (Tong et al., 2000, 2005; Jack and Crossley, 2010), have also been identified as important modulators in adipogenesis. However, most evidences are based on studies in humans and model animals as rodents. In buffaloes, researches on adipogenesis are still very limited. The genetic diversities of adipogenesis relative genes have been suggested to be with the adipogenesis of milk fat (Gu et al., 2017, 2019, 2020). Phosphoenolpyruvate carboxykinase 1 has been identified as a significant candidate gene that is involved in IMF deposition by transcriptome sequencing analysis and functional validation in buffalo adipocytes (Huang et al., 2020).

Although the major regulatory activity of adipogenesis has been revealed, the precisely orchestrated process is far from complete, as new modulators in this process are gradually identified. In recent years, increasing long non-coding RNAs (lncRNAs) have been demonstrated to have profound effects on adipogenesis (Li et al., 2016; Huang et al., 2019). lncRNAs are a kind of well-known non-coding RNAs that have more than 200 nucleotides and have become a research hotspot in recent years. With the development of high-throughput sequencing technology, increasing lncRNAs have been demonstrated to modulate fat deposition (Nuermaimaiti et al., 2018; Huang et al., 2019; Zhang and Fu, 2020; Zhang S. et al., 2020). The majority of studies that reveal a significant role of lncRNAs in adipogenesis are performed in humans (Nuermaimaiti et al., 2018; Zhang T. et al., 2020) or murine (Cai et al., 2018; Chen et al., 2020). In livestock animals, several lncRNAs also have been identified to modulate adipogenesis. In pigs, knockdown lncIMF4 promotes adipogenesis by attenuating autophagy to repress the lipolysis in intramuscular adipocytes (Sun et al., 2020). In cattle, lncRNA ADNCR suppresses adipogenic differentiation by targeting miR-204 (Li et al., 2016). Recently, a new lncRNA lncFAM200B is found to have an important role in the development of adipocytes in cattle (Zhang S. et al., 2020). In buffaloes, the *NDUFC2-AS* lncRNA promotes adipogenic differentiation by upregulating adipogenesis relative genes (Huang et al., 2019). Compared to the larger number of lncRNAs identified in adipose tissue (Huang et al., 2019), the number of present identified lncRNAs with effects on adipogenesis is very limited, suggesting that the modulatory role of lncRNAs is still poorly understood.

To uncover novel lncRNAs involved in the regulatory network of adipogenesis, lncRNA profiles of adipose and muscle tissues were characterized by high-throughput RNA sequencing using the Illumina HiSeq 3000 platform in this study. Differential expression analysis was performed, and the host gene was revealed to yield candidate lncRNAs with putative effects on adipogenesis. Further gain-of-function experiments demonstrated that an lncRNA, which transcribed from the upstream region of sorting and assembly machinery component 50 homolog (*SAMM50*), promotes the adipogenic differentiation of buffalo adipocytes by upregulating the adipogenesis relative gene. This study further supplies the buffalo lncRNA data and proposes a novel lncRNA that has a significant role in fat deposition of buffalo.

MATERIALS AND METHODS

Animals and Sample Preparation

Chinese swamp buffaloes (bull, n = 3) were raised under equivalent forage and feeding management condition in Xinyang Buffalo Breeding Farm (Guangshan, Henan province, China) as previously described (Huang et al., 2019). Animals were weaned at 6 months of age and slaughtered at 30 months of age. Tissues (the longissimus dorsi muscle, back subcutaneous fat, heart, liver, spleen, lung, and kidney) were sampled immediately after slaughter and were frozen in liquid nitrogen for RNA sequencing and qRT-PCR experiments. Meanwhile, the fresh back subcutaneous fat was kept at ∼30°C in phosphate-buffered saline (PBS) with 1% streptomycin and penicillin and taken back to the lab for primary adipocyte isolation.

RNA Isolation and Sequencing

Total RNA was isolated by TRIzol (Invitrogen, Carlsbad, CA, United States) according to the manufacturer's instructions. RNA quantity was measured with NanoDrop 2000 (NanoDrop, Wilmington, DE, United States) and 1.5% agarose gels. RNA with $1.8 < 260/280$ value <2.0 and concentration >500 ng/μL was used for further analysis. Isolation of nuclear and cytoplasmic RNA was performed by PARIS kit (Life Technologies, Carlsbad, CA, United States) according to the manufacturer's instructions. Details of RNA isolation and high-throughput RNA sequencing were described previously (Huang et al., 2019). The longissimus dorsi muscle (n = 3) and the back subcutaneous fat (n = 3) were used for RNA sequencing.

Quality Control, Transcriptome Assembly, lncRNA Prediction, and Differential Expression Analysis

Quality control, transcriptome assembly, and lncRNA prediction were performed as previously described (Huang et al., 2019). Briefly, the low-quality reads and those containing adapters were removed to obtain clean reads. Then, clean reads that are high-quality were used for the subsequent analysis. The cattle genome (UMD3.1) was used as the reference, for the annotation information of buffalo genome is not available. Clean reads were mapped to the reference genome to obtain complete transcripts. Transcripts with more than 200 bp and without coding capability were identified as lncRNAs. The expression level of lncRNA was indicated as log2(FPKM+1). lncRNA with the absolute value of log2(fold change) $\geq$ 2 and the FDR value $\leq$ 0.05 was considered to be differentially expressed (DE).

qRT-PCR Analysis

Details of primer design, reverse transcription reaction, and quantitative PCR were described in our previous study (Huang et al., 2019). The ubiquitously expressed prefoldin-like chaperone (*UXT*) gene and the glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) gene were used to normalize the expression level of the candidate gene in tissues and adipocytes of buffalo, respectively (Huang et al., 2019; Feng et al., 2020). For the 3T3-L1 cells, β-*actin* was used as the internal reference gene. The cycle threshold ($2^{-\Delta\Delta Ct}$) method was used to calculate the relative expression level of the candidate gene. In particular, for cell localization, *U6* and β-*actin* were respectively used as nuclear and cytoplasmic markers, and the $2^{-\Delta Ct}$ method was used to calculate the gene expression level. Three replicates were run per sample, and the qRT-PCR experiment was performed three times. Details of the primers used for qRT-PCR are shown in **Supplementary Table 1**.

Vector Construction

The CDS region of mouse *SAMM50* (NCBI Reference Sequence: NM_178614.5) was amplified from the cDNA of mouse muscle tissue, which was kindly provided by Dr. Yongjie Xu of Xinyang Normal University (Xinyang, China) and cloned into the *Hin*dIII and *Xho*I restriction sites of pcDNA3.1(+) vector. Primers used to amplify the CDS region were as follows: F-5′-CCC*aagctt*GCCGAGCCTCTTGTGTTTG-3′; R-5′-CCGctcgagCCAGAAGCACTCAACCGTGT-3′. The lowercase indicates the restriction enzyme site.

Cell Culture

The 3T3-L1 preadipocytes were purchased from ATCC (Shanghai, China). Buffalo primary adipocytes were isolated from adipose tissues of male buffaloes (n = 3) using the tissue block method as described in our previous study (Huang et al., 2019). Buffaloes used here were different than those used for RNA sequencing, but all the animals were raised under equivalent forage and feeding management conditions in the same farm and slaughtered at similar months of age. Adipocytes were cultured with a complete culture medium [Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum and 1% penicillin–streptomycin] in 5% CO_2 at 37°C. All the reagents used for cell culture were purchased from Gibco (Grand Island, NY, United States). Before transfection and transduction, cells were plated in a 6-well plate in triplicate.

Transfection, Adipogenic Differentiation, Oil Red O Staining, and Quantification

For the 3T3-L1 preadipocytes, transfection was conducted when the cells reached 80% confluence by using Lipofectamine 3000 (Invitrogen, Carlsbad, CA, United States) following the manufacturer's protocol. Two days after transfection, cells were induced to adipogenic differentiation treatment with an inducing medium (containing 10 μg/mL insulin, 1 μM rosiglitazone, 1 μM dexamethasone, and 0.5 mM IBMX). Two days later, cells were treated with a maintenance medium which contains 10 μg/mL insulin and 1 μM rosiglitazone. Meanwhile, transfection was performed again. After inducing with adipogenic agents for 8 days, Oil Red O staining and quantification were performed as previous described (Huang et al., 2019).

Adenovirus Packaging and Transduction

Adenovirus packaging was performed at Hanbio Biotechnology Co., Ltd. (Shanghai, China). Briefly, the full length of lncSAMM50 was synthesized and ligated to the AdMax system to obtain Ad-lncSAMM50. EGFP was used as an internal indicator. Ad-EGFP was used as a negative control.

Similar to transfection, adenoviral transduction was conducted when the buffalo adipocytes reached 80% confluence. Twenty-four hours later, cells were treated with an inducing medium for 2 days and then treated with a maintenance medium for 4 days. The maintenance medium was changed every 2 days. After inducing with adipogenic agents for 6 days, Oil Red O staining and quantification of lipid content were performed as previously described (Huang et al., 2019).

Statistical Analysis

Comparison was analyzed by using the SPSS 19.0 software. Student's t-test was used when the data had a normal distribution; otherwise, a non-parametric test was performed. A value of $p < 0.05$ was considered to indicate statistically significant differences. Data were presented as mean ± SD by using the OriginPro 8.5 program.

RESULTS

Differential Expression Analysis and Validation

In total, 43,809 lncRNAs were identified by a computer algorithm in buffalo adipose and muscle tissues in this study (**Supplementary Table 2**). Differential expression analysis revealed that 241 lncRNAs were DE between adipose and muscle tissues in buffalo (**Supplementary Table 3**). Among them, 125 were upregulated in adipose tissue compared with muscle tissue while others were downregulated (**Supplementary Table 3**).

To evaluate the quality of differential expression analysis, 13 lncRNAs (5 lncRNAs were upregulated and 8 were downregulated in adipose tissue) were randomly selected for validation by qRT-PCR. As shown in **Figure 1**, the expression patterns of 5/5 upregulated and 6/8 downregulated lncRNAs

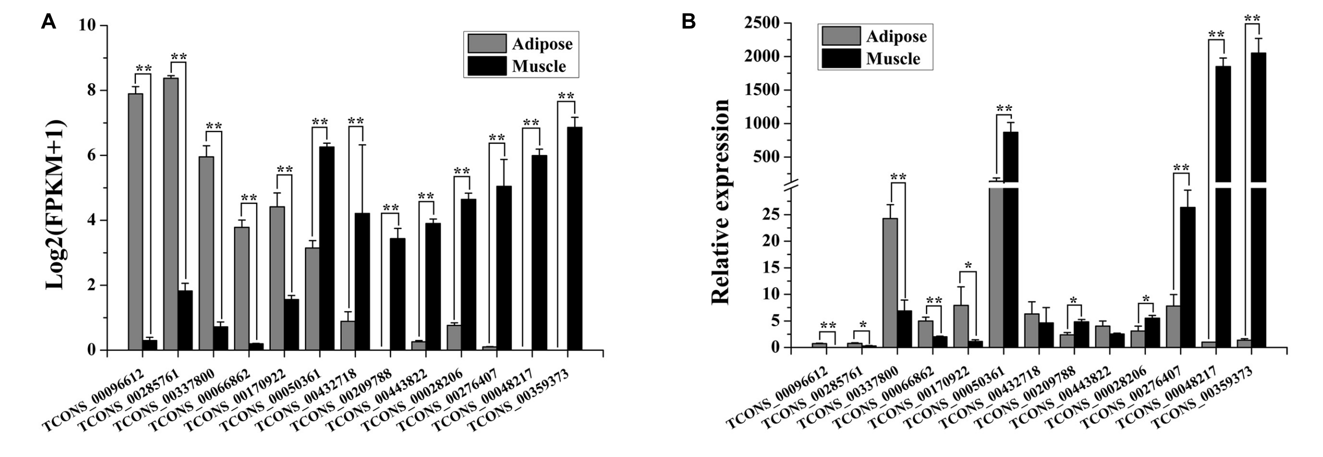

FIGURE 1 | Validation of differentially expressed (DE) lncRNAs by qRT-PCR. **(A)** Expression patterns of the 13 DE lncRNAs in the adipose and muscle tissues of buffalo based on RNA sequencing. The expression level of lncRNA is indicated as log2(FPKM+1). Data are presented as mean ± SD ($n = 3$; $^{*}p < 0.05$; $^{**}p < 0.01$). **(B)** Expression patterns of the 13 DE lncRNAs in the adipose and muscle tissues of buffalo analyzed by qRT-PCR. The RNA expression levels are normalized to those of *UXT*. Data are presented as mean ± SD ($n = 6$; $^{*}p < 0.05$; $^{**}p < 0.01$).

TABLE 1 | Candidate lncRNAs associated with fat deposition in buffalo.

Transcript _id	Host gene	Strand	Symbol	Adipose _1	Adipose _2	Adipose _3	Muscle _1	Muscle _2	Muscle _3	Mean _adipose	Mean _muscle	Log2 (Fold Change)	*p*-value	FDR
TCONS _00096612	FABP4	Antisense strand	FABP4-AS lncRNA	7.49	8.25	7.94	0.12	0.47	0.30	7.89	0.29	−7.60	0.0034	0.0188
TCONS _00285761	NDUFC2	Antisense strand	NDUFC2-AS lncRNA	8.52	8.25	8.36	1.35	2.07	2.05	8.38	1.82	−6.56	0.0001	0.0004
TCONS _00285845	Intergenic region	–	–	6.99	7.19	6.29	0.78	0.90	1.54	6.82	1.07	−5.75	0.0014	0.0088
TCONS _00337800	SAMM50	Antisense strand	lncSAMM50	6.59	5.85	5.43	0.46	0.99	0.69	5.96	0.71	−5.24	0.0001	0.0004

in qRT-PCR analysis were consistent with that in RNA sequencing analysis.

Candidate lncRNAs Associated With Fat Deposition in Buffalo

The aim of this study was to identify lncRNAs with significant effect on fat deposition in buffalo. We noticed that four DE lncRNAs have log2(fold change) ≥ -5 and showed a high expression level in adipose tissue (**Table 1**). Among them, TCONS_00096612, TCONS_00285761, and TCONS_00337800 are transcribed from the antisense strand of fatty acid-binding protein 4 (*FABP4*), ubiquinone oxidoreductase subunit C2 (*NDUFC2*), and *SAMM50* gene, respectively. Interestingly, these genes have been confirmed to be associated with fat deposition. In addition, the *p* value for lncSAMM50 and *NDUFC2*-AS lncRNA was the lowest. Thus, we further focused on the effect of lncSAMM50 on the fat deposition in buffalo.

Characterization of lncSAMM50

The full length of lncSAMM50 is 3,169 nt (**Supplementary Table 4**), and the sequence is reverse complementary with the upstream region, exon 1, and part of intron 1 of *SAMM50* (**Figure 2A**). Both Coding Potential Calculator (CPC) and Coding Potential Assessment Tool (CPAT) indicated that lncSAMM50 is a non-coding RNA (**Figures 2B,C**). Results of semiquantitative PCR for nuclear and cytoplasmic fractions showed that lncSAMM50 was mainly expressed in the nucleus (**Figure 2E**). The qRT-PCR detection confirmed that the expression pattern of lncSAMM50 was the same as a nuclear marker U6 (**Figure 2D**).

Expression Pattern of lncSAMM50 and *SAMM50*

Based on RNA sequencing, the expression level of lncSAMM50 in adipose tissue is higher than that in muscle tissue (**Figure 3A**, $p < 0.01$), which was further conformed by qRT-PCR analysis (**Figure 3B**, $p < 0.05$). By contrast, *SAMM50*, the host gene of lncSAMM50, showed a similar expression level in adipose and muscle tissues (**Figures 3A,B**). Analysis of the tissue expression profile revealed that lncSAMM50 is mainly expressed in adipose and muscle tissues while *SAMM50* is widely expressed in variable tissues (**Figures 3C,D**). During adipogenic differentiation, lncSAMM50 was upregulated in the mature adipocytes of buffalo (**Figure 3E**) while *SAMM50* was widely expressed in different stages (**Figure 3F**).

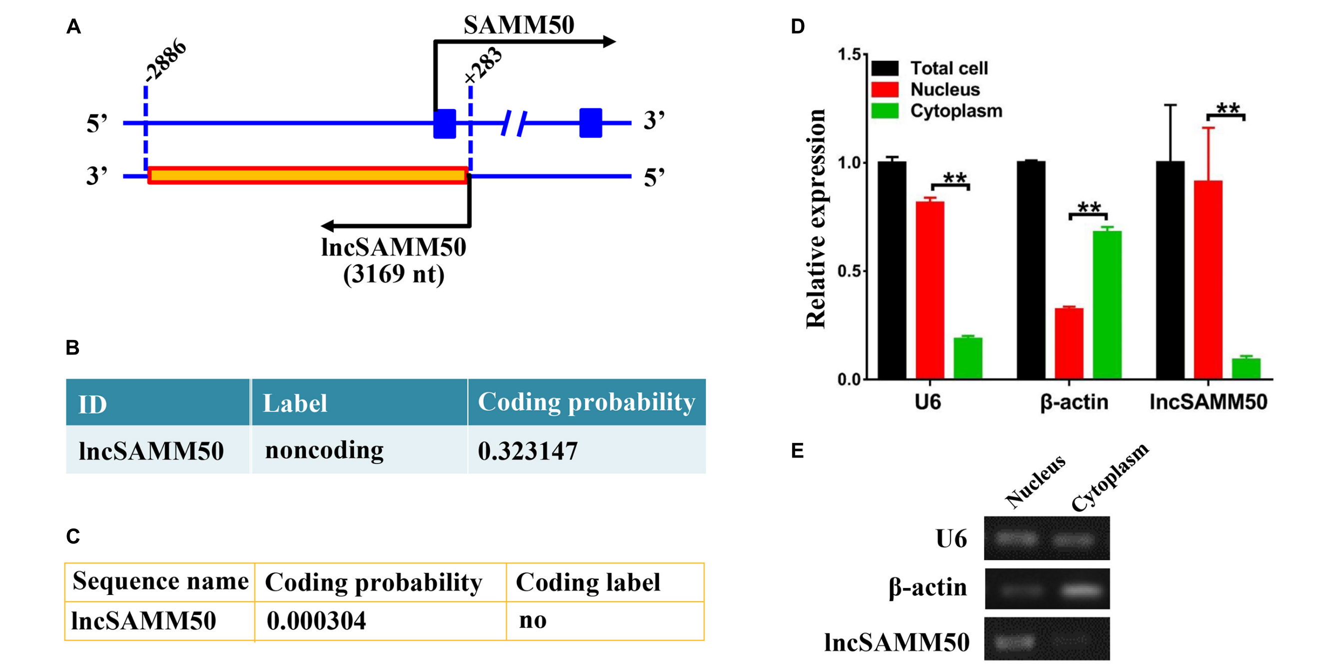

FIGURE 2 | Characterization of buffalo lncSAMM50. **(A)** Positional relationship between SAMM50 and lncSAMM50. **(B)** The Coding Potential Calculator (CPC) program suggests that lncSAMM50 is a non-coding RNA. **(C)** The Coding Potential Assessment Tool (CPAT) indicates that lncSAMM50 is a non-coding RNA. **(D)** Cell localization of lncSAMM50 by qRT-PCR. Adipocytes induced to differentiation for 6 days were used for separation of nucleus and cytoplasm RNA. *U6* and β-*actin* were respectively used as nuclear and cytoplasmic markers. The $2^{-\Delta Ct}$ method was used to calculate the gene expression level. Data are presented as the mean ± SD (n = 3; $^{**}p < 0.01$). **(E)** Cell localization of lncSAMM50 by semiquantitative PCR.

SAMM50 Inhibits the Adipogenic Differentiation of 3T3-L1 Cells

To access the function of *SAMM50* in fat deposition, gain-of-function experiments for *SAMM50* were performed in 3T3-L1 adipocytes. The strategy of transfection, adipogenic differentiation, and Oil Red O staining is shown in **Figure 4A**. As expected, the mRNA expression of *SAMM50* was highly significantly upregulated in pcDNA3.1_SAMM50 group (**Figure 4B**, $p < 0.01$). Meanwhile, *C/EBPα* was significantly downregulated in the pcDNA3.1_SAMM50 group (**Figure 4D**, $p < 0.05$). Accordingly, lipid accumulation in the pcDNA3.1_SAMM50 group was less than that in the pcDNA3.1 group (**Figures 4E,F**). No effect was detected on the expression of *PPARG* (**Figure 4C**).

lncSAMM50 Promotes the Adipogenic Differentiation of Buffalo Adipocytes

To evaluate the effect of lncSAMM50 on fat deposition in buffalo, the full length of lncSAMM50 (**Supplementary Table 4**) was packaged into an adenovirus system for overexpression (ad_lncSAMM50). The time axis of overexpression of LncSAMM50, induction, quantification is shown in **Figure 5A**. Indicator EGFP was highly expressed 1 day after adenoviral transduction and continued until the 6th day of adipogenic induction (**Figure 5B**). The expression of lncSAMM50 in the ad_lncSAMM50 group was significantly higher than that in the ad_EGFP group, and the overexpression was continued until the 6th day of adipogenic induction (**Figure 5E**, $p < 0.01$). Meanwhile, lipid accumulation in the ad_lncSAMM50 group was significantly enhanced (**Figures 5C,D**, $p < 0.01$). As to the adipogenic markers, the mRNA expressions of *PPARG* and *C/EBPα* were slightly upregulated on day_0 and day_6 of adipogenic induction, respectively (**Figure 5F**). Lipoprotein lipase (*LPL*), a lipolysis gene, was upregulated on day_0 of adipogenesis induction (24 h after lncSAMM50 overexpression) in the ad_lncSAMM50 group (**Figure 5I**). Confusingly, the fatty acid transporter (*FAT/CD36*), a fatty acid uptake marker, was downregulated in the ad_lncSAMM50 group (**Figure 5G**). For the expression of the host gene *SAMM50*, no significant difference was observed between the ad_lncSAMM50 group and the ad_EGFP group (**Figure 5E**).

DISCUSSION

This study characterizes the lncRNA expression profiles of buffalo adipose and muscle tissues based on RNA sequencing analysis and evaluates the effects of lncSAMM50 on the adipogenesis of buffalo adipocytes. This study demonstrates that (1) the expression profiles of lncRNAs in buffalo adipose and muscle are significantly different with each other; (2) lncSAMM50 is a nuclear-location non-coding RNA; (3) *SAMM50* inhibits adipogenic differentiation in 3T3-L1 cells; and (4) lncSAMM50 promotes adipogenic differentiation by slightly upregulating

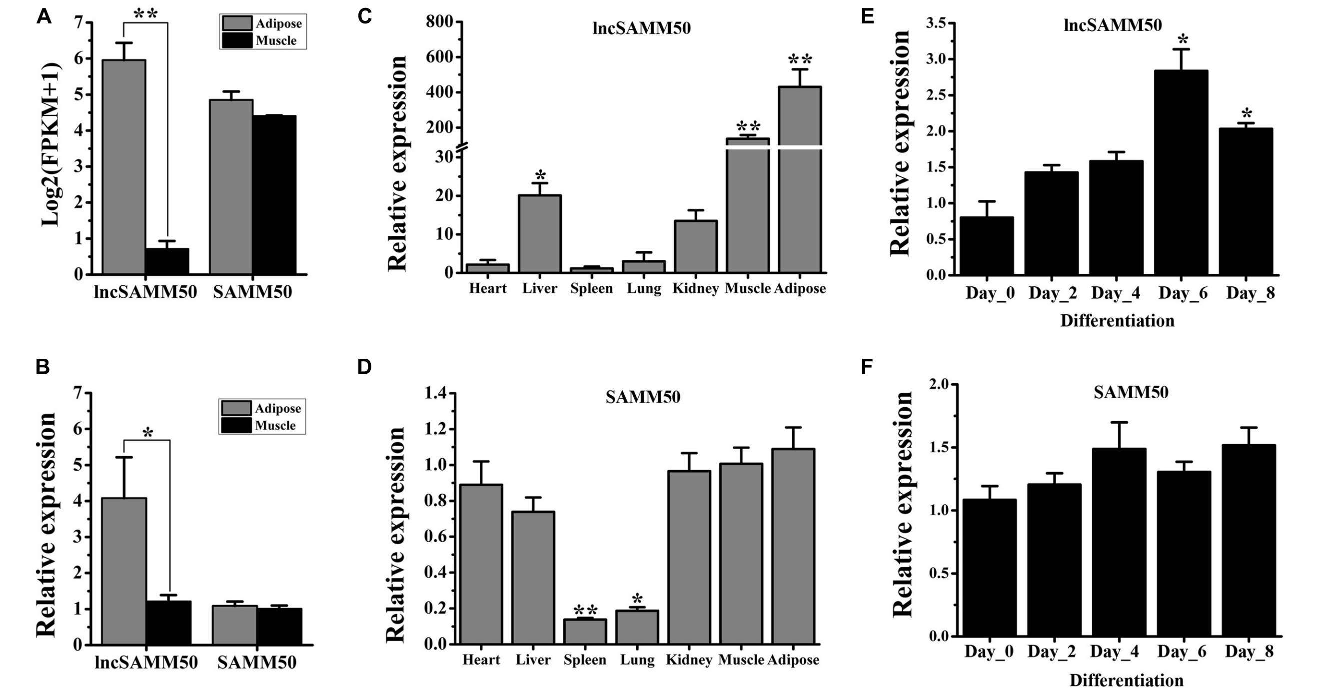

FIGURE 3 | Expression pattern of lncSAMM50 and SAMM50 in buffalo. **(A)** The expression pattern of lncSAMM50 and *SAMM50* in adipose and muscle tissues by RNA sequencing. **(B)** The expression pattern of lncSAMM50 and *SAMM50* in adipose and muscle tissues by qRT-PCR. **(C)** The expression profile of lncSAMM50 in heart, liver, spleen, lung, kidney, muscle, and adipose tissues. (D) The expression profile of SAMM50 in heart, liver, spleen, lung, kidney, muscle, and adipose tissues. For panels **(A–D)**, Xinyang buffalo (30 months of age, $n = 3$) was used; the *UXT* gene was used to normalize the expression level of the candidate gene. **(E)** The expression pattern of lncSAMM50 during adipocyte differentiation. **(F)** The expression pattern of *SAMM50* during adipocyte differentiation. *GAPDH* was used to normalize the expression level of the candidate gene in adipocytes. The cycle threshold ($2^{-\Delta\Delta Ct}$) method was used to calculate the relative expression level of the candidate gene. Data are presented as the mean ± SD ($n = 3$; $^{*}p < 0.05$; $^{**}p < 0.01$).

PPARG, *C/EBP*α, and *LPL* in buffalo adipocytes, but with no effect on its host gene *SAMM50*.

Each of the activities in living organisms is precisely mediated by a genome. Generally, the gene is expressed in a time- and stage-specific manner and is regulated by multiple factors. With the development of RNA sequencing, the lncRNA expression profile has been characterized in multiple tissues in livestock animals (Huang et al., 2019; Li et al., 2020; Song et al., 2020). In the present study, the comparison of the lncRNA expression profiles of adipose and muscle tissues identified 241 DE lncRNAs (**Supplementary Table 3**). The quality of the differential expression analysis was further identified by qRT-PCR. These results indicated a significant difference in the biological function between adipose and muscle tissues in buffalo. Among the DE lncRNAs, four with high expression are significantly upregulated in adipose tissue (**Table 1**). The *NDUFC2*-AS lncRNA (TCONS_00096612) has been demonstrated to promote the adipogenesis in buffalo adipocytes (Huang et al., 2019). FABP4 is a significant protein in fatty acid transportation (Boord et al., 2002) and adipocyte differentiation (Garin-Shkolnik et al., 2014). SAMM50 is a mitochondrial membrane protein and is associated with energy metabolism in mammals (Liu et al., 2016). Considering the function of host genes and the lowest p value (**Table 1**), we focused on a lncRNA transcribed from the antisense strand of *SAMM50*, lncSAMM50. Interestingly, lncSAMM50 is mainly expressed in adipose tissue (**Figure 3C**) and is upregulated during the adipogenic differentiation of buffalo adipocytes (**Figure 3E**). These results indicated a vital role of lncSAMM50 in fat deposition of buffalo (Li et al., 2016; Huang et al., 2019).

The existing data suggest that lncRNA can play a role by regulating the expression of a host gene (Guo et al., 2019; Song et al., 2020), meaning that the function of a lncRNA is associated with its host gene. SAMM50 is the core component of the sorting and assembly machinery and plays a critical role in regulating mitochondrial dynamics and mitophagy (Liu et al., 2016; Jian et al., 2018), indicating a significant role of SAMM50 in energy metabolism. In the present study, we found that SAMM50 is widely expressed across different tissues in buffalo, especially in tissues with high level in energy metabolism such as the heart, liver, muscle, and adipose (**Figure 3D**). These results are consistent with its vital role in mitochondria (Liu et al., 2016; Jian et al., 2018). However, the effect of SAMM50 on adipogenic differentiation of adipocytes had not been revealed. By gain-of-function experiments, we demonstrated that SAMM50 inhibits the adipogenic differentiation of 3T3-L1 adipocytes (**Figures 4D–F**). These results further indicate that lncSAMM50 may affect the fat deposition by regulating the expression of its host gene SAMM50.

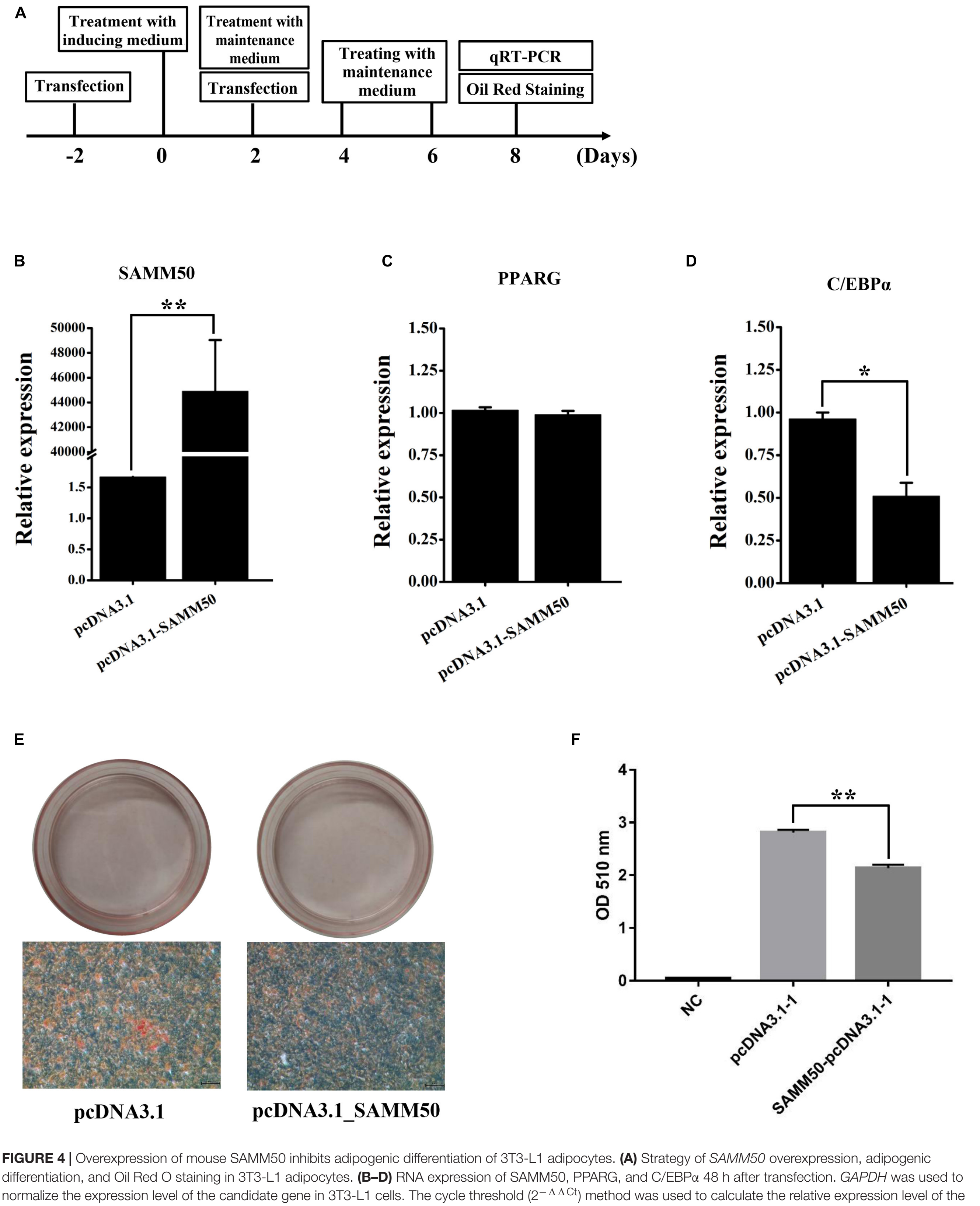

FIGURE 4 | Overexpression of mouse SAMM50 inhibits adipogenic differentiation of 3T3-L1 adipocytes. **(A)** Strategy of *SAMM50* overexpression, adipogenic differentiation, and Oil Red O staining in 3T3-L1 adipocytes. **(B–D)** RNA expression of SAMM50, PPARG, and C/EBPα 48 h after transfection. *GAPDH* was used to normalize the expression level of the candidate gene in 3T3-L1 cells. The cycle threshold ($2^{-\Delta\Delta Ct}$) method was used to calculate the relative expression level of the candidate gene. **(E)** Images of Oil Red O staining in 3T3-L1 cells transfected with pcDNA3.1 and pcDNA3.1-SAMM50 on day 8 of adipogenic differentiation. Scale bar, 200 μm. **(F)** Histogram showing the quantitation of Oil Red O staining by spectrophotometry. NC, negative control. Data are presented as the mean ± SD ($n = 3$; $^{*}p < 0.05$; $^{**}p < 0.01$).

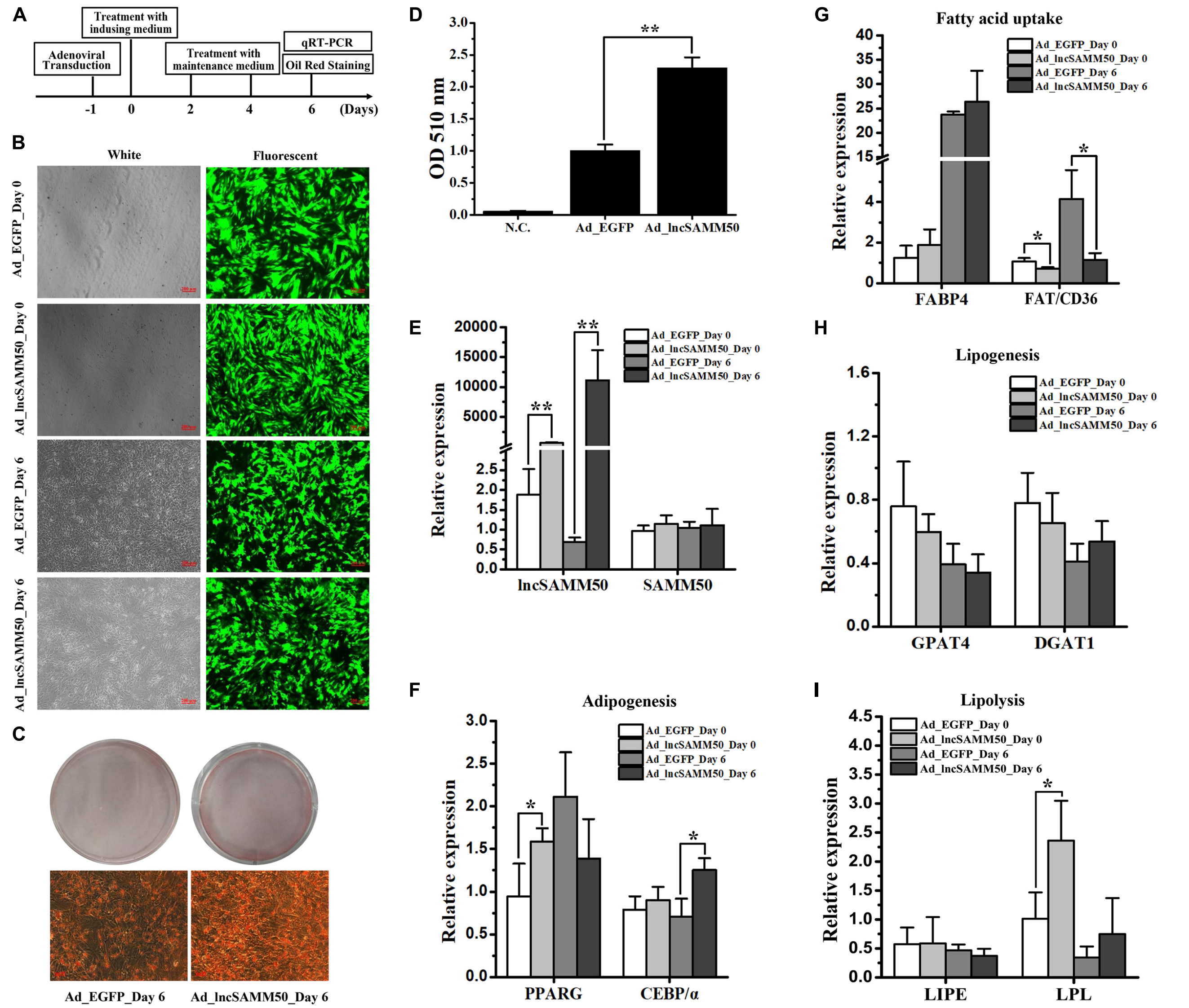

FIGURE 5 | Overexpression of lncSAMM50 enhances adipogenic differentiation of buffalo adipocytes. **(A)** Strategy of lncSAMM50 overexpression, adipogenic differentiation, and Oil Red O staining in buffalo adipocytes. **(B)** Micrographs of EGFP-positive cells in the ad_EGFP (control) and ad_lncSAMM50 groups on day 0 and day 6 of adipogenic differentiation. Scale bar, 200 μm. **(C)** Images of Oil Red O staining in buffalo adipocytes transduced with ad_EGFP and ad_lncSAMM50 on day 6 of adipogenic differentiation. Scale bar, 50 μm. **(D)** Histogram showing the quantitation of Oil Red O staining by spectrophotometry. NC, negative control. **(E–I)** The RNA expression levels of lncSAMM50, SAMM50, PPARG, C/EBPα, FABP4, FAT/CD36, GPAT4, DGAT1, LIPE, and LPL on day 0 and day 6 of adipogenic differentiation in buffalo adipocytes transduced with ad_EGFP and ad_lncSAMM50. *GAPDH* was used to normalize the expression level of the candidate gene in buffalo adipocytes. The cycle threshold ($2^{-\Delta\Delta Ct}$) method was used to calculate the relative expression level of the candidate gene. Data are presented as the mean ± SD ($n = 3$; *$p < 0.05$; **$p < 0.01$).

To confirm the effect of lncSAMM50 on fat deposition, an overexpression of lncSAMM50 in buffalo adipocytes was performed by an efficient adenovirus system. As expected, lncSAMM50 significantly enhances the lipid accumulation in buffalo adipocytes (**Figures 5C–E**). Meanwhile, eight lipid metabolism-associated genes, including two adipogenesis markers *PPARG* and *C/EBPα*, two fatty acid uptake markers *FAT/CD36* and fatty acid-binding protein 4 (FABP4), two lipogenesis markers glycerol-3-phosphate acyltransferase 4 (*GPAT4*) and diacylglycerol *O*-acyltransferase 1 (*DGAT1*), and two lipolysis markers lipase E (*LIPE*) and *LPL*, were used to predict the potential regulatory mechanisms of lncSAMM50 in buffalo adipocytes. PPARG and C/EBPα are well known as the crucial determinants of adipogenesis in adipocytes (Lowe et al., 2011; Mota de Sá et al., 2017). With the significant increase of lncSAMM50, both *PPARG* and *C/EBPα* were slightly upregulated (**Figures 5E,F**). These results indicate that lncSAMM50 may not have a direct impact on the expression of *PPARG* and *C/EBPα* but promote the adipogenic differentiation of buffalo adipocytes. FABP4 is a member of the fatty acid-binding protein family which is responsible for the intracellular transport of fatty acids (Lappas, 2014). FAT/CD36 is a membrane protein expressed in adipose tissue and plays an important role in the transport of fatty acid into adipocytes (Bonen et al., 2007). LPL can be produced by adipocytes and transferred to the surface of adipocytes to hydrolyze triglycerides and liberate free fatty acids (Merkel et al., 2002; Yagyu et al., 2003). The fatty acid produced by LPL lipase can be transported into adipocytes, synthesized again, and stored in adipose tissue (Merkel et al., 2002). In the present study, though the expression of FABP4 was not stimulated and the expression of FAT/CD36 was slightly inhibited by lncSAMM50 (**Figure 5G**), the expression of *LPL* was slightly upregulated (**Figure 5I**) in the ad_lncSAMM50 group, indicating that lncSAMM50 may enhance the fatty acid transport into buffalo adipocytes. GPAT4 and DGAT1 are key markers for triglyceride synthesis (Lappas, 2014; Yan and Ajuwon, 2015). Regretfully, both GPAT4 and DGAT1 were not stimulated by the overexpression of lncSAMM50 (**Figure 5H**), indicating that lncSAMM50 has no effect on the expression of these two genes.

Existing evidence suggests that lncRNAs can repress or activate the hose gene in the *cis* method (Fatica and Bozzoni, 2014; Wang et al., 2016; Song et al., 2020). Sirt1 antisense (AS) lncRNA is transcribed from the AS strand of the Sirt1 gene. Sirt1 AS lncRNA promotes myoblast proliferation and inhibits differentiation by interacting with Sirt1 3′UTR to rescue Sirt1 transcriptional suppression by competing with miR-34a (Wang et al., 2016). Similarly, another lncRNA IGF2 AS promotes the proliferation and differentiation of bovine myoblasts by complementing the IGF2 intron and affecting the expression of IGF2 mRNA (Song et al., 2020). In the present study, the sequence of lncSAMM50 is reverse complementary to the upstream region, exon 1, and part of intron 1 of *SAMM50* (**Figure 2A**). Additionally, lncSAMM50 is a nuclear localization transcript (**Figures 2D,E**). Thus, the physical proximity of lncSAMM50 and *SAMM50* inspired us to investigate a relationship in regulation between them. Unfortunately, overexpression of lncSAMM50 does not affect the expression of *SAMM50* in buffalo adipocytes (**Figure 5E**). Previously, we also identified a similar lncRNA, *NDUFC2*-AS lncRNA, which promotes the adipogenic differentiation by upregulating adipogenesis relative genes but with no obvious effect on the host gene as well (Huang et al., 2019). Thus, the precise regulatory mechanism of lncSAMM50 promoting the adipogenesis of buffalo adipocytes still needs further investigation.

Meanwhile, limitations still exist in this study. Firstly, the sample size (n = 3) and the gender (male only) for RNA sequencing seem to be limited. A higher sample size and use of both male and female animals will harvest a more accurate expression profile of lncRNAs. Secondly, identification of the effect of SAMM50 activity in buffalo adipocytes will contribute to a clearer relationship between SAMM50 and lncSAMM50. However, the effect of SAMM50 on lipid accumulation in adipocytes was only evaluated in the 3T3-L1 cell line but not in buffalo adipocytes. This is because the transfection by a simple liposome method is practicable in 3T3-L1 cells but not in buffalo adipocytes. Moreover, overexpression must be performed through the more complex and time-consuming virus system in buffalo adipocytes.

In conclusion, the present study provides a valuable genomic resource for identification of functional lncRNAs in buffalo and reveals the important role of lncSAMM50 in lipid accumulation of buffalo adipocytes. These data further perfects the molecular theory on buffalo fat deposition, which will instruct the buffalo breeding by genetic engineering or genome editing.

DATA AVAILABILITY STATEMENT

The RNA sequencing data were deposited in the GEO profiles of NCBI. The accession number of three adipose tissues is GSE112744 and that of three muscle tissues is GSE139102.

ETHICS STATEMENT

The animal study was reviewed and approved by Institutional Animal Care and Use Committee (IACUC) of Xinyang Normal University.

AUTHOR CONTRIBUTIONS

JH, QL, and DS designed the experiment. XF, YW, and RZ collected the samples. XF, YW, RZ, DG, and JL performed the experiments. JH and XF analyzed the data and wrote the manuscript. All authors have read and approved the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant Nos. 31702094 and 32060747).

ACKNOWLEDGMENTS

We would like to acknowledge the Xinyang Buffalo Breeding Farm (Guangshan, Henan province, China) for providing the buffalo.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.626158/full#supplementary-material

Supplementary Table 1 | Details of primers used for qRT-PCR detection.

Supplementary Table 2 | Total lncRNAs identified in the adipose and muscle tissues of buffalo.

Supplementary Table 3 | The lncRNAs showed differential expression between adipose and muscle tissues in buffalo.

Supplementary Table 4 | Full length of lncSAMM50.

REFERENCES

Birsoy, K., Chen, Z., and Friedman, J. (2008). Transcriptional regulation of adipogenesis by KLF4. *Cell Metab*. 7, 339–347. doi: 10.1016/j.cmet.2008.02.001

Bonen, A., Chabowski, A., Luiken, J. J., and Glatz, J. F. (2007). Is membrane transport of FFA mediated by lipid, protein, or both? mechanisms and regulation of protein-mediated cellular fatty acid uptake: molecular, biochemical, and physiological evidence. *Physiology* 22, 15–29. doi: 10.1152/physiologyonline.2007.22.1.15

Boord, J. B., Fazio, S., and Linton, M. F. (2002). Cytoplasmic fatty acid-binding proteins: emerging roles in metabolism and atherosclerosis. *Curr. Opin. Lipidol*. 13, 141–147. doi: 10.1097/00041433-200204000-00005

Cai, R., Sun, Y., Qimuge, N., Wang, G., Wang, Y., Chu, G., et al., (2018). Adiponectin as lncRNA inhibits adipogenesis by transferring from nucleus to cytoplasm and attenuating adiponectin mRNA translation. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* 1863, 420–432. doi: 10.1016/j.bbalip.2018.01.005

Cao, Z., Umek, R. M., and McKnight, S. L. (1991). Regulated expression of three C/EBP isoforms during adipose conversion of 3T3-L1 cells. *Genes Dev*. 5, 1538–1552. doi: 10.1101/gad.5.9.1538

Chen, Y., Li, K., Zhang, X., Chen, J., Li, M., and Liu, L. (2020). The novel long noncoding RNA lncRNA-Adi regulates adipogenesis. *Stem Cells Transl. Med*. 9, 1053–1067. doi: 10.1002/sctm.19-0438

Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet*. 15, 7–21. doi: 10.1038/nrg3606

Feng, X., Cao, X., Zhu, R., and Huang, J. (2020). Selection and validation of reference genes for RT-qPCR in adipose and longissimus dorsi muscle tissues of buffalo. *Anim. Biotechnol*. Epub ahead of print.

Garin-Shkolnik, T., Rudich, A., Hotamisligil, G. S., and Rubinstein, M. (2014). FABP4 attenuates PPARγ and adipogenesis and is inversely correlated with PPARγ in adipose tissues. *Diabetes* 63, 900–911. doi: 10.2337/db13-0436

Gu, M., Cosenza, G., Gaspa, G., Iannaccone, M., Macciotta, N. P. P., Chemello, G., et al., (2020). Sequencing of lipoprotein lipase gene in the Mediterranean river buffalo identified novel variants affecting gene expression. *J. Dairy Sci*. 103, 6374–6382. doi: 10.3168/jds.2019-17968

Gu, M., Cosenza, G., Iannaccone, M., Macciotta, N. P. P., Guo, Y., Di Stasio, L., et al., (2019). The single nucleotide polymorphism g.133A>C in the stearoyl CoA desaturase gene (SCD). promoter affects gene expression and quali-quantitative properties of river buffalo milk. *J. Dairy Sci*. 102, 442–451. doi: 10.3168/jds.2018-15059

Gu, M., Cosenza, G., Nicolae, I., Bota, A., Guo, Y., Di Stasio, L., et al., (2017). Transcript analysis at DGAT1 reveals different mRNA profiles in river buffaloes with extreme phenotypes for milk fat. *J. Dairy Sci*. 100, 8265–8276. doi: 10.3168/jds.2017-12771

Guo, W., Liang, X., Liu, L., Guo, Y., Shen, S., Liang, J., et al., (2019). MiR-6872 host gene SEMA3B and its antisense lncRNA SEMA3B-AS1 function synergistically to suppress gastric cardia adenocarcinoma progression. *Gastric Cancer* 22, 705–722. doi: 10.1007/s10120-019-00924-0

Hamm, J. K., Park, B. H., and Farmer, S. R. (2001). A role for C/EBPbeta in regulating peroxisome proliferator-activated receptor gamma activity during adipogenesis in 3T3-L1 preadipocytes. *J. Biol. Chem*. 276, 18464–18471. doi: 10.1074/jbc.m100797200

Huang, J., Feng, X., Zhu, R., Guo, D., Wei, Y., Cao, X., et al., (2020). Comparative transcriptome analysis reveals that PCK1 is a potential gene affecting IMF deposition in buffalo. *BMC Genomics* 21:710. doi: 10.1186/s12864-020-07120-w

Huang, J., Zheng, Q., Wang, S., Wei, X., Li, F., and Ma, Y. (2019). High-throughput RNA sequencing reveals NDUFC2-AS lncRNA promotes adipogenic differentiation in Chinese buffalo (*Bubalus bubalis* L). *Genes (Basel)* 10:689. doi: 10.3390/genes10090689

Jack, B. H., and Crossley, M. (2010). GATA proteins work together with friend of GATA (FOG). and C-terminal binding protein (CTBP). co-regulators to control adipogenesis. *J. Biol. Chem*. 285, 32405–32414. doi: 10.1074/jbc.m110.141317

Jian, F., Chen, D., Chen, L., Yan, C., Lu, B., Zhu, Y., et al., (2018). Sam50 regulates PINK1-Parkin-mediated mitophagy by controlling PINK1 stability and mitochondrial morphology. *Cell Rep*. 23, 2989–3005. doi: 10.1016/j.celrep.2018.05.015

Kiran, M., and Naveena, B. M. (2014). Buffalo meat quality, composition, and processing characteristics: contribution to the global economy and nutritional security. *Anim. Front*. 4, 18–24. doi: 10.2527/af.2014-0029

Lappas, M. (2014). Effect of pre-existing maternal obesity, gestational diabetes and adipokines on the expression of genes involved in lipid metabolism in adipose tissue. *Metabolism* 63, 250–262. doi: 10.1016/j.metabol.2013.10.001

Li, H., Huang, K., Wang, P., Feng, T., Shi, D., Cui, K., et al., (2020). Comparison of long non-coding RNA expression profiles of cattle and buffalo differing in muscle characteristics. *Front. Genet*. 11:98. doi: 10.3389/fgene.2020.00098

Li, M., Sun, X., Cai, H., Sun, Y., Plath, M., Li, C., et al., (2016). Long non-coding RNA ADNCR suppresses adipogenic differentiation by targeting miR-204. *Biochim. Biophys. Acta* 1859, 871–882. doi: 10.1016/j.bbagrm.2016.05.003

Liu, S., Gao, Y., Zhang, C., Li, H., Pan, S., Wang, X., et al., (2016). SAMM50 affects mitochondrial morphology through the association of Drp1 in mammalian cells. *FEBS Lett*. 590, 1313–1323. doi: 10.1002/1873-3468.12170

Lowe, C. E., O'Rahilly, S., and Rochford, J. J. (2011). Adipogenesis at a glance. *J. Cell Sci*. 124, 2681–2686. doi: 10.1242/jcs.079699

Merkel, M., Eckel, R. H., and Goldberg, I. J. (2002). Lipoprotein lipase: genetics, lipid uptake, and regulation. *J. Lipid Res*. 43, 1997–2006.

Mori, T., Sakaue, H., Iguchi, H., Gomi, H., Okada, Y., Takashima, Y., et al., (2005). Role of Krüppel-like factor 15 (KLF15). in transcriptional regulation of adipogenesis. *J. Biol. Chem*. 280, 12867–12875. doi: 10.1074/jbc.m410515200

Mota de Sá, P., Richard, A. J., Hang, H., and Stephens, J. M. (2017). Transcriptional regulation of adipogenesis. *Compr. Physiol*. 7, 635–674. doi: 10.1002/cphy.c160022

Nuermaimaiti, N., Liu, J., Liang, X., Jiao, Y., Zhang, D., Liu, L., et al., (2018). Effect of lncRNA HOXA11-AS1 on adipocyte differentiation in human adipose-derived stem cells. *Biochem. Biophys. Res. Commun*. 495, 1878–1884. doi: 10.1016/j.bbrc.2017.12.006

Oishi, Y., Manabe, I., Tobe, K., Tsushima, K., Shindo, T., Fujiu, K., et al., (2005). Krüppel-like transcription factor KLF5 is a key regulator of adipocyte differentiation. *Cell Metab*. 1, 27–39. doi: 10.1016/j.cmet.2004.11.005

Song, C., Yang, Z., Jiang, R., Cheng, J., Yue, B., Wang, J., et al., (2020). lncRNA IGF2 AS regulates bovine myogenesis through different pathways. *Mol. Ther. Nucleic Acids* 21, 874–884. doi: 10.1016/j.omtn.2020.07.002

Sun, Y., Cai, R., Wang, Y., Zhao, R., Qin, J., and Pang, W. (2020). A newly identified LncRNA LncIMF4 controls adipogenesis of porcine intramuscular preadipocyte through attenuating autophagy to inhibit lipolysis. *Animals* 10:926. doi: 10.3390/ani10060926

Tong, Q., Dalgin, G., Xu, H., Ting, C. N., Leiden, J. M., and Hotamisligil, G. S. (2000). Function of GATA transcription factors in preadipocyte-adipocyte transition. *Science* 290, 134–138. doi: 10.1126/science.290.5489.134

Tong, Q., Tsai, J., Tan, G., Dalgin, G., and Hotamisligil, G. S. (2005). Interaction between GATA and the C/EBP family of transcription factors is critical in GATA-mediated suppression of adipocyte differentiation. *Mol. Cell Biol*. 25, 706–715. doi: 10.1128/mcb.25.2.706-715.2005

Wang, G. Q., Wang, Y., Xiong, Y., Chen, X. C., Ma, M. L., Cai, R., et al., (2016). Sirt1 as lncRNA interacts with its mRNA to inhibit muscle formation by attenuating function of miR-34a. *Sci. Rep*. 6:21865.

Yagyu, H., Chen, G., Yokoyama, M., Hirata, K., Augustus, A., Kako, Y., et al., (2003). Lipoprotein lipase (LpL). on the surface of cardiomyocytes increases lipid uptake and produces a cardiomyopathy. *J. Clin. Invest*. 111, 419–426. doi: 10.1172/jci16751

Yan, H., and Ajuwon, K. M. (2015). Mechanism of butyrate stimulation of triglyceride storage and adipokine expression during adipogenic differentiation of porcine stromovascular cells. *PLoS One* 10:e0145940. doi: 10.1371/journal.pone.0145940

Yeh, W. C., Cao, Z., Classon, M., and McKnight, S. L. (1995). Cascade regulation of terminal adipocyte differentiation by three members of the C/EBP family of leucine zipper proteins. *Genes Dev*. 9, 168–181. doi: 10.1101/gad.9.2.168

Zhang, S., Kang, Z., Cai, H., Jiang, E., and Pan, C. (2020b). Identification of novel alternative splicing of bovine lncRNA lncFAM200B and its effects on preadipocyte proliferation. *BMC Bioinformatics* 21:541.

Zhang, T., Liu, H., Mao, R., Yang, H., Zhang, Y., Zhang, Y., et al., (2020a). The lncRNA RP11-142A22.4 promotes adipogenesis by sponging miR-587 to modulate Wnt5β expression. *Cell Death Dis.* 11:475.

Zhang, Y., and Fu, Y. (2020). Identification of differentially expressed mRNA and the hub mRNAs modulated by lncRNA Meg3 as a competing endogenous RNA in brown adipose tissue of mice on a high-fat diet. *Adipocyte* 9, 346–358.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ORIGINAL RESEARCH
published: 07 July 2021
doi: 10.3389/fgene.2021.643497

The Expression Profiles of mRNAs and lncRNAs in Buffalo Muscle Stem Cells Driving Myogenic Differentiation

Ruimen Zhang[1†], Jinling Wang[1†], Zhengzhong Xiao[2†], Chaoxia Zou[1], Qiang An[1], Hui Li[1], Xiaoqing Zhou[2], Zhuyue Wu[2], Deshun Shi[1], Yanfei Deng[1], Sufang Yang[1,3*] and Yingming Wei[1*]*

[1] *State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Animal Reproduction Institute, Guangxi University, Nanning, China,* [2] *The Animal Husbandry Research Institute of Guangxi Autonomous, Nanning, China,* [3] *International Zhuang Medical Hospital Affiliated to Guangxi University Chinese Medicine, Nanning, China*

OPEN ACCESS

Edited by:
Guohua Hua, Huazhong Agricultural University, China

Reviewed by:
Ikhide G. Imumorin, Georgia Institute of Technology, United States
Bo Wang, China Agricultural University, China

***Correspondence:**
Yanfei Deng
yanfei-dun@163.com
Sufang Yang
ysfang3511@163.com
Yingming Wei
dkywym@163.com

[†] *These authors have contributed equally to this work*

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *18 December 2020*
Accepted: *25 May 2021*
Published: *07 July 2021*

Citation:
Zhang R, Wang J, Xiao Z, Zou C, An Q, Li H, Zhou X, Wu Z, Shi D, Deng Y, Yang S and Wei Y (2021) The Expression Profiles of mRNAs and lncRNAs in Buffalo Muscle Stem Cells Driving Myogenic Differentiation. Front. Genet. 12:643497. doi: 10.3389/fgene.2021.643497

Buffalo breeding has become an important branch of the beef cattle industry. Hence, it is of great significance to study buffalo meat production and meat quality. However, the expression profiles of mRNA and long non-coding RNAs (lncRNA) molecules in muscle stem cells (MuSCs) development in buffalo have not been explored fully. We, therefore, performed mRNA and lncRNA expression profiling analysis during the proliferation and differentiation phases of MuSCs in buffalo. The results showed that there were 4,820 differentially expressed genes as well as 12,227 mRNAs and 1,352 lncRNAs. These genes were shown to be enriched in essential biological processes such as cell cycle, p53 signaling pathway, RNA transport and calcium signaling pathway. We also identified a number of functionally important genes, such as *MCMC4*, *SERDINE1*, *ISLR*, LOC102394806, and LOC102403551, and found that interference with *MYLPF* expression significantly inhibited the differentiation of MuSCs. In conclusion, our research revealed the characteristics of mRNA and lncRNA expression during the differentiation of buffalo MuSCs. This study can be used as an important reference for the study of RNA regulation during muscle development in buffalo.

Keywords: buffalo, muscle stem cells, mRNAs, non-coding RNAs, myogenesis

INTRODUCTION

There is an annual increase in the global consumption of beef and it is an indispensable food in our modern society, and therefore the beef cattle industry occupies an increasingly important position in modern agricultural practices (Bonny et al., 2015). According to statistics, in 2019, China's beef production was 6.85 million tons and beef imports were 1.66 million tons with a year-on-year increase of approximately 57%. It is anticipated that China's future beef demand will continue to rise. Therefore, China urgently needs a viable and thriving beef cattle industry in order to provide its society with larger amounts of high-quality beef (Mwangi et al., 2019; Ornaghi et al., 2020). There is a need for us to conduct research on the growth and meat quality of locally produced beef as well as to explore the potential molecular information of breeding stocks so as to provide reference values for future breeding protocols (Grigoletto et al., 2020).

In ruminants, skeletal muscle tissue accounts for about 40–60% of the adult animal body weight, which not only determines the level of meat production performance, but also has an important impact on meat quality. There is a group of myoblasts-muscle stem cells (MuSCs), which are the source of skeletal muscle formation and regeneration, and these have the potential for differentiation and proliferation of muscle-derived stem cells (Feige and Rudnicki, 2018; Feige et al., 2018). This is also the current cell model for studying skeletal muscle development. Under certain conditions, these cells can be activated causing the MuSCs proliferate and differentiate.

One of the main challenges in the field of muscle research is to understand how the genes that are involved in specialized muscle functions at the transcriptional and post-transcriptional levels are regulated. Undoubtedly, myogenic regulatory factors (MRFs) (Hernandez-Hernandez et al., 2017), myocyte enhancer factor-2 (*MEF2*) (Taylor and Hughes, 2017), and *PAX3*/*PAX7* genes are the main genes involved in the growth and development of skeletal muscle. Initially, long non-coding RNAs (lncRNAs) were considered to be transcriptional noise but later studies showed these RNAs play an important function in many biological processes (Jae and Dimmeler, 2020). Epigenetic control and transcriptional regulation, translation, RNA metabolism, stem cell maintenance and differentiation, autophagy and apoptosis, embryonic development, and other aspects have also been shown to play important roles (Chen et al., 2020). With the discovery of a large number of important muscle regulators such as lncRNA H19 (Xu et al., 2017), Neat1 (Wang et al., 2019), lnc-133b (Jin et al., 2017), circLOM7 (Wei et al., 2017), more and more ncRNAs related to muscle development have also been widely characterized (Martone et al., 2019). At the same time, the important role of related coding RNAs, lncRNAs, and other molecules in the development of skeletal muscle in agricultural animals are gradually being explored.

So far, with the emergence of RNA structure detection technologies such as Frag-seq (Underwood et al., 2010), (ss/dsRNA)-seq, and SHAPE-seq, have allowed scientists to characterize the structure of RNAs obtained from different tissues and cell components. When these data were combined with knowledge of RNA transformation events, such as miRNA targeting, RNA modification, and the function of RNA binding proteins (RPBs), they have emphasized the importance of RNA structure during gene regulation (Li et al., 2012). Moreover, most of these studies are focused on mRNAs and ncRNAs in order to explore the biological functions of RNA structure.

As a characteristic species of southern China, the potential use of the buffalo as a meat source has gradually attracted attention. The buffalo breeding industry has become a food basket project for urban residents, but the meat production and meat quality of buffalo needs to be improved for it to be an acceptable alternative to cattle (Li et al., 2020). Previously, several breakthroughs have been made in studies of buffalo embryos, stem cells, and somatic cells, covering traits such as milk production, reproduction, and meat production. This culminated in the successful construction of the buffalo genomic DNA sequence map (Low et al., 2019). Recently our laboratory analyzed the regulatory networks of lncRNA-mRNA interactions in the muscle tissue of cattle and buffalo (Li et al., 2020).

However, when compared to cattle, buffalo muscle has the characteristics of possessing greater shear force and consisting of thicker muscle fibers. At present, the molecular mechanisms that regulate buffalo muscle fibers formation are still unclear (Huang et al., 2021). We hypothesized that there are key signaling pathway(s) which control the myogenic differentiation of MuSCs. We, therefore, analyzed the mRNA expression of MuSCs before and after myogenic differentiation through transcriptome sequencing strategies in an attempt to screen the signal pathways that may regulate muscle fiber development. Other recent studies have also shown that differential expression lncRNAs also play an important physiological function during cellular differentiation of MuSCs (Zhu et al., 2017). This study further expands the understanding of skeletal muscle biology, and provides a reference target for the genetic improvement of buffalo and the production and cultivation of meat *in vitro* and *in vivo*.

Abbreviations: MuSCs, muscle stem cells; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; lncRNAs, long non-coding RNAs; RT-qPCR, quantitative real time PCR.

MATERIALS AND METHODS

MuSCs Culture and Differentiation

All experiments regarding animals were performed in the State Key Laboratory for Conservation and Utilization of Subtropical Agro-bio-resources, and were conducted in accordance with its guidelines for the care and use of laboratory animals. Primary water buffalo MuSCs were isolated and cultured from fetal-derived longissimus muscle as described in Supplementary File 1, using a combination digestion method of type I collagenase and trypsin. MuSCs were cultured in high-glucose DMEM supplemented with fetal bovine serum (Hyclone, USA; 10% FBS and 20% FBS, respectively) and antibiotics [1% penicillin and streptomycin; growth medium (GM)] at 5% CO_2, 37°C. To induce MuSCs myogenic differentiation, MuSCs were switched to a differentiation medium (DMEM, 2% horse serum; DM) when cells were almost 90% confluent for up to 4 days.

Sample Preparation

The tissues from Chinese buffalo at embryonic stage (90 days) were collected at a local slaughterhouse in Nanning, Guangxi province. Tissue samples, including muscle, liver, heart, lung, skin, kidney, brain, stomach, and intestine, were collected and immediately frozen in liquid nitrogen. Proliferation of MuSCs was labeled as the GM samples ($n = 3$) and differentiation of these was then called the DM samples ($n = 3$). The samples were kept at −80°C until RNA was isolated.

Total RNA Extraction

Total RNA from cells and tissues samples were extracted with TRizol reagent (Invitrogen, Carlsbad, CA, USA) in accordance with manufacturer's instructions.

RNA-Seq and Transcriptome Data Analysis

About 3 μg RNA per sample was used as the initial material for RNA sample preparation. PolyA-Seq libraries were prepared following the described protocol at RiboBio (Guangzhou, China) in accordance with the manufacturer's instructions. The identification of mRNAs and lncRNAs was carried out with reference to RiboBio's technical methods. We have provided a detailed description of the methods and analysis in **Supplementary Table 9**. All data were uploaded to the GEO database.

Analysis of Differentially Expressed Genes, mRNAs, and lncRNAs

The RPKM (expected number of Reads Per Kilobase of transcript sequence per Millions base pairs sequenced) value was used to estimate the expression levels of mRNAs and lncRNAs. Genes with a RPKM value of <1 in no <50% of samples were defined as unreliably expressed genes, while those with a RPKM value of ≥1 in more than 50% of samples were considered as reliably expressed genes. Differentially Expressed Genes DE mRNAs, and DE lncRNAs were analyzed using DESeq2, which defined them as reliably expressed genes with |log2 (Fold Change)|>1 and *Q*-values <0.05 between any two groups.

Gene Ontology and KEGG Analysis

Gene ontology (GO; http://www.geneontology.org) and KEGG pathway (http://www.kegg.jp) were analyzed as described previously.

Quantitative Real-Time PCR

Total RNA was extracted using TRizol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. Reverse transcription was performed by using the HiScript R II One Step RT-PCR kit (Vazyme, Nanjing, China). RT-qPCR was performed with ChamQ SYBR qPCR Master Mix (Vazyme, Nanjing, China) using the $2^{-\Delta\Delta Ct}$ method. Beta-actin was used as the internal control. All primer sequences used are listed in Supplementary File 2.

Western Blotting

Cells were collected from different treatment groups, pelleted by centrifugation, and then lysed in RIPA buffer. Total protein was prepared and protein concentrations were determined using the Bradford method. Proteins were then separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and subsequently transferred to nitrocellulose membranes. These were then blocked with 5% skimmed milk powder solution for 1.5–2 h at room temperature. The membranes were then incubated overnight with primary antibodies. Anti-PCNA, anti-CDK2, and anti-β-actin were purchased from Abcam (Cambridge, MA, USA). After that, the membranes were washed with PBS-tween and incubated for 1.5 h with horseradish peroxidase-conjugated secondary antibodies (Abcam, Cambridge, MA, USA). Protein bands were detected after treatment with Super Signal West Femto reagent purchased from Thermo (Thermo Scientific, Karlsruhe, Germany).

Vector Construction

Construction and sequencing identification of *MYLPF* interference vectors were completed by a Biological Company (GeneCopoeia, Guangzhou, China). The interference MYLPF expression vector plasmids, which were named sh-MYLPF-A, sh-MYLPF-B, sh-MYLPF-C, and the control plasmids were named NC. All primers sequences used are shown in Supplementary File 2.

Treatment of Cells

Muscle stem cells were grown to 70% confluence and then trypsinized and plated at 5×10^5 cells/well into six-well plates (Thermo Fisher Scientific, USA). They were then transfected with vectors using X-treme GENE HP DNA Transfection Reagent (Roche, Basel, Switzerland). After incubation, the MuSCs were used for the different assays outlined below. In order to induce differentiation of myoblasts, the culture medium was changed to high-glucose DM medium.

Immunofluorescence and Microscopy

Myoblasts of MuSCs were washed three times with PBS buffer (pH 7.4), and permeabilized for 15 min in PBS containing 0.5% Triton X-100 before fixation in PBS containing 4% paraformaldehyde for 20–30 min. Immunostaining was carried out as follows: cells were incubated overnight at 4°C with primary anti-MyoD1 (1:200; Abcam) diluted in 5% bovine serum albumin. After that, cells were washed with PBS and incubated at room temperature for 3–4 h with the corresponding secondary antibody, goat anti-mouse IgG (H+L; 1:1,000; Invitrogen) diluted in PBS. DNA was visualized using 5 mg/ml DAPI staining. Finally, the prepared cells were washed four times with PBS and observed under a fluorescence microscope (Nikon).

Statistical Analysis

The quantitative results are presented as means ± SEMs based on at least three independent experiments. Significant variance by treatments in comparison to the untreated samples was determined by one-way ANOVA performed with GraphPad Prism version 8.0 (GraphPad Software, La Jolla, CA, USA). Differences were considered significant when *P*-values were ≤ 0.05.

RESULTS

Variation of Phenotypic Characteristics During Differentiation of Buffalo MuSCs

A combination digestion method of type I collagenase and trypsin was used to obtain buffalo fetal-derived MuSCs. This cell type is similar to fibroblasts and spindle-shaped in appearance. These cells have good proliferation capacity (**Figure 1A**), which is referred to as the proliferation phase (GM samples) of MuSCs. In addition, when the medium was

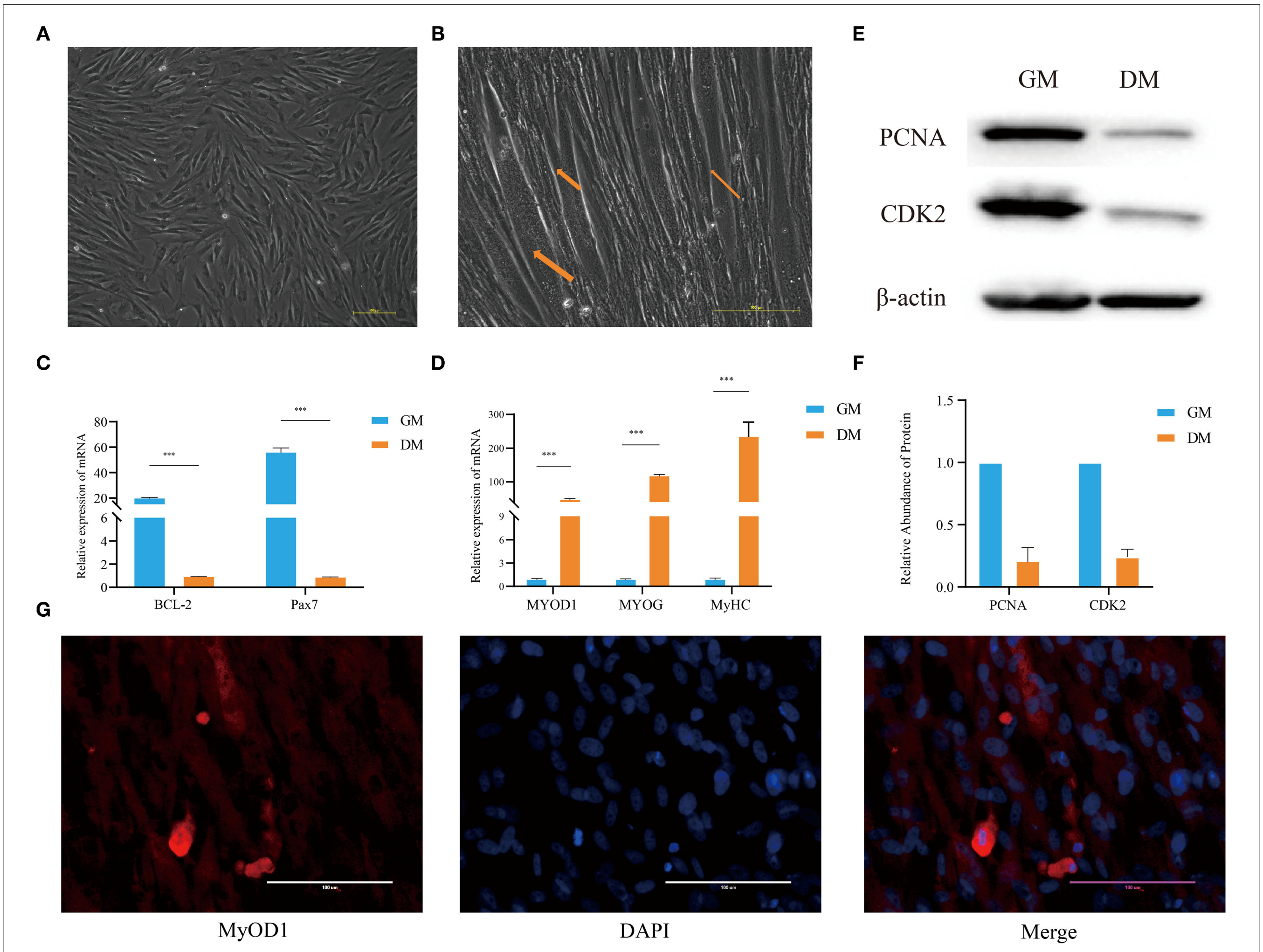

FIGURE 1 | Analysis of the characteristics of buffalo MuSCs during proliferation and differentiation. **(A)** Proliferation phenotype of muscle stem cells (GM samples). **(B)** Differentiation phenotype of muscle stem cells (DM samples). **(C)** MuSCs grown as GM and DM samples were subjected to real-time PCR for *BCL-2* and *Pax7*. **(D)** The differentiation phenotype (DM) samples of MuSCs were subjected to real-time PCR for *MyoD1*, *MyoG,* and *MyHC*. **(E,F)** MuSCs grown as GM and DM samples were subjected to western blotting for determination of *PCNA* and *CDK2* proteins. **(G)** The differentiation phenotype (DM) samples of MuSCs were subjected to immunofluorescence for *MyOD1*. The data are presented as means ± *SDs*, $n = 3$ per group. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. Scale bars = 100/200 μm.

replaced with DM, after 2 days, the cells began to show myotube fusion. On the fifth day, the number of myotubes increased and the myotube fusion became more obvious, which is referred to as the differentiation phase (DM samples) of MuSCs (**Figure 1B**).

Western blotting showed that the expression of *PCNA* and *CDK2* in MuSCs GM samples were significantly higher than that in DM samples (**Figures 1E,F**). The expression levels of *BCL-2* and *Pax7* (paired Homeobox transcription factors) in GM samples were significantly higher than those in MuSCs DM samples (**Figure 1C**). Immunofluorescence experiments showed that the muscle marker molecule, *MyOD1,* was enriched in DM samples of MuSCs (**Figure 1G**). The expression levels of muscle-derived marker molecules, *MYOD1*, *MYOG*, and *MyHC,* increased significantly in DM samples of MuSCs (**Figure 1D**). These results suggest that the cells obtained were MuSCs with the capability of myogenic differentiation.

PolyA-Seq Characteristics of Buffalo MuSCs

In order to identify the mRNAs and lncRNAs involved in proliferation and differentiation, we compared the polyA-seq status of GM and DM samples of MuSCs (**Supplementary Figure S1**). Analysis of sequencing data revealed that a very large number of clean reads, total maps, and uniquely mapped areas were involved in these processes (**Figure 2A**). The analysis of uniquely map profiles of MuSCs, showed the distribution for the reads in different chromosomes (**Figures 2B,C**). Among them, most of the reads from GM and DM samples were found to be targeted to exonic areas.

A

RNA-seq information statistics of samples

Sample	Effective Reads	Total mapped	Uniquely mapped	Reads mapped in proper pairs
DM1	57,596,044(100%)	51,872,397(90.06%)	50,100,867(86.99%)	48,606,484(84.39%)
DM2	61,516,490(100%)	54,935,455(89.30%)	53,113,346(86.34%)	50,910,210(82.76%)
DM3	63,795,084(100%)	55,847,360(87.54%)	54,035,324(84.70%)	52,169,222(81.78%)
GM1	64,813,710(100%)	59,227,087(91.38%)	57,442,917(88.63%)	55,846,712(86.16%)
GM2	65,025,170(100%)	58,897,366(90.58%)	57,076,410(87.78%)	55,085,782(84.71%)
GM3	60,247,382(100%)	54,869,417(91.07%)	53,185,590(88.28%)	51,545,542(85.56%)

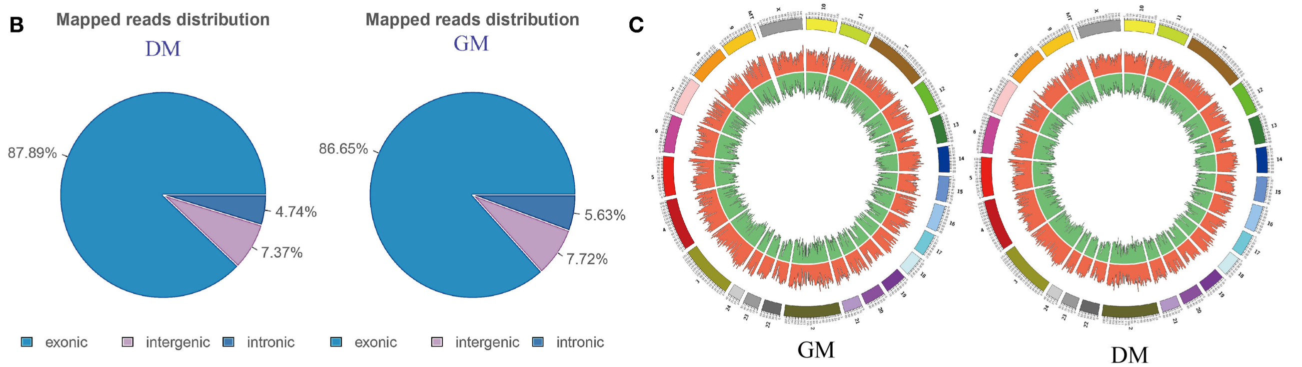

FIGURE 2 | RNA-seq analysis of buffalo MuSCs during proliferation and differentiation. **(A)** RNA-seq information statistics of samples. **(B)** Mapped reads distribution. **(C)** The distribution for the reads in different chromosomes.

TABLE 1 | PolyA-seq statistics of the different results obtained.

Differential type	Upregulated	Downregulated	Total
Genes	2,979	1,841	4,820
mRNAs	7,505	4,722	12,227
lncRNAs	831	521	1,352

Profiles of DE Genes, mRNAs, and lncRNAs During Differentiation of Buffalo MuSCs

Analysis of sequencing data revealed that a total of 31,819 genes, 57,640 mRNAs, and 11,357 lncRNAs were involved in the proliferation and differentiation of MuSCs (**Supplementary Table 1**). We also performed heatmaps and volcano plots for the genes, mRNAs, and lncRNAs in MuSCs ($|\log 2$ (FoldChange)$| > 1$, Q-value < 0.05) (**Supplementary Figure S2**). There were 4,820 DEGs, 12,227 DE mRNAs, and 1,352 DE lncRNAs (**Supplementary Tables 2–4**). Compared with the GM samples of MuSCs, 2,979 genes (61.80%), 7,505 mRNAs (61.38%), and 831 lncRNAs (61.46%) were upregulated, while 1,841 genes (38.20%), 4,722 mRNAs (38.62%), and 521 lncRNAs (38.54%) were downregulated in DM samples of MuSCs (**Table 1**).

Signal Pathway Enrichment Analysis of DEGs Between Proliferation and Differentiation Phases of Buffalo MuSCs

Since we mainly analyzed mRNAs transcripts, and also involved a small number of known lncRNAs, we did not perform functional correlation analysis on these lncRNAs. We performed GO and KEGG enrichment analysis on the related regulatory DEGs in the processes of proliferation, differentiation, transformation, and maturation. This produced signal pathway information which was then used to predict the functions and mechanisms of the mRNAs (**Supplementary Tables 5**, **6**). The results of pathway analysis of DEGs showed that GO analyzed and annotated these into three main categories: biological processes, cellular components, and molecular functions, including ATP binding and nucleus RNA-directed DNA polymerase activity (**Figure 3**). In addition, we employed KEGG pathway enrichment analysis to further understand the biological functions and molecular interactions of most DEGs with the assumption that the identified pathways may be involved in the development and growth of buffalo skeletal muscle. We found more than 300 pathways to be enriched, and the top 30 most significant terms were uncovered, including biological processes such as cell cycle, p53 signaling pathway, RNA transport, and calcium signaling pathway (**Figure 4**). In short, these signal pathways related to DEGs play an important role in the regulation of MuSC proliferation and differentiation, which provides an important basis for subsequent research on buffalo myogenesis.

The Verification of DEGs and DE lncRNAs in MuSCs

Based on the expression levels of DEGs and DE lncRNAs, including 12 genes (*MCM4*, *MCM7*, *SERDINE1*, *SEMA7A*, *C1QTNF6*, *CPZ*, *VDR*, *PLAC9*, *ISLR*, *MyOG*, *PCNA,* and

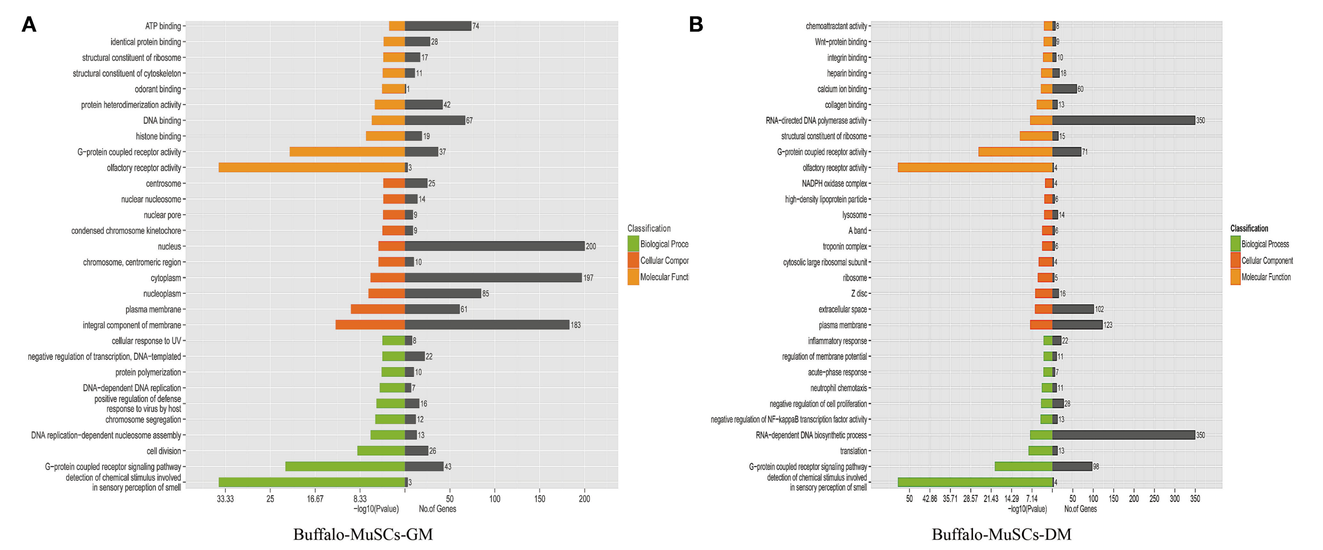

FIGURE 3 | Gene ontology (GO) analysis. **(A)** GO analysis of DEGs in the GM samples of MuSCs. **(B)** GO analysis of DEGs in the DM samples of MuSCs.

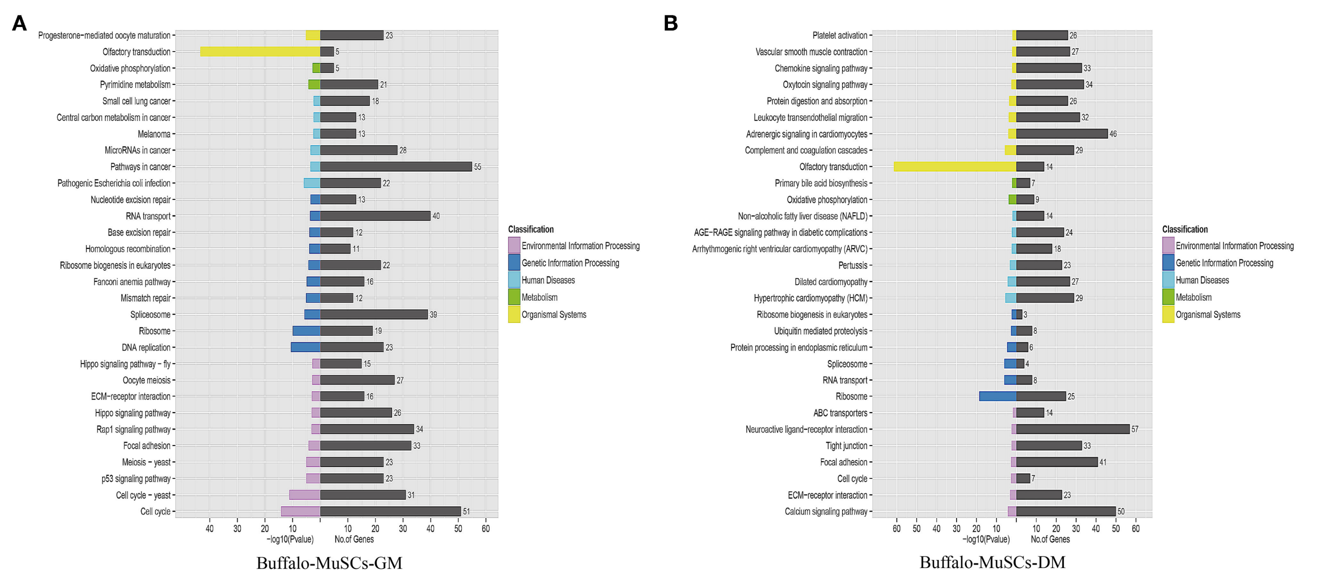

FIGURE 4 | Kyoto Encyclopedia of Genes and Genomes pathway analysis. **(A)** Top 30 KEGG terms of DEGs in the GM samples of MuSCs. **(B)** Top 30 KEGG terms of DEGs in the DM samples of MuSCs.

cyclin D1) related to the cell cycle, actin cytoskeleton, cell differentiation, and lipid metabolism (**Figures 5A,B**), and seven random lncRNAs (LOC102403551, LOC112586870, LOC112584513, LOC102399397, LOC112581569, LOC102395296, and LOC102394806; **Figure 5C**), were selected for RT-qPCR verification. After comparisons with the RNA-seq data, similar expression trends for RT-qPCR were discovered, showing the strong consistency between RT-qPCR and RNA-seq data.

The Role of *MYLPF* in Buffalo MuSCs

We identified a dysregulated gene, *MYLPF*, which was shown to be upregulated significantly (by almost 60-fold) in DM compared with GM samples when measured by RT-qPCR (**Figure 6A**). In addition, *MYLPF* was expressed in various tissues, such as heart and liver and the highest expression levels seen in muscle (**Figure 6B**). A previous report also showed that the relationship between *MYLPF* and meat quality can be used as an important genetic consideration when dealing with gene-assisted selection

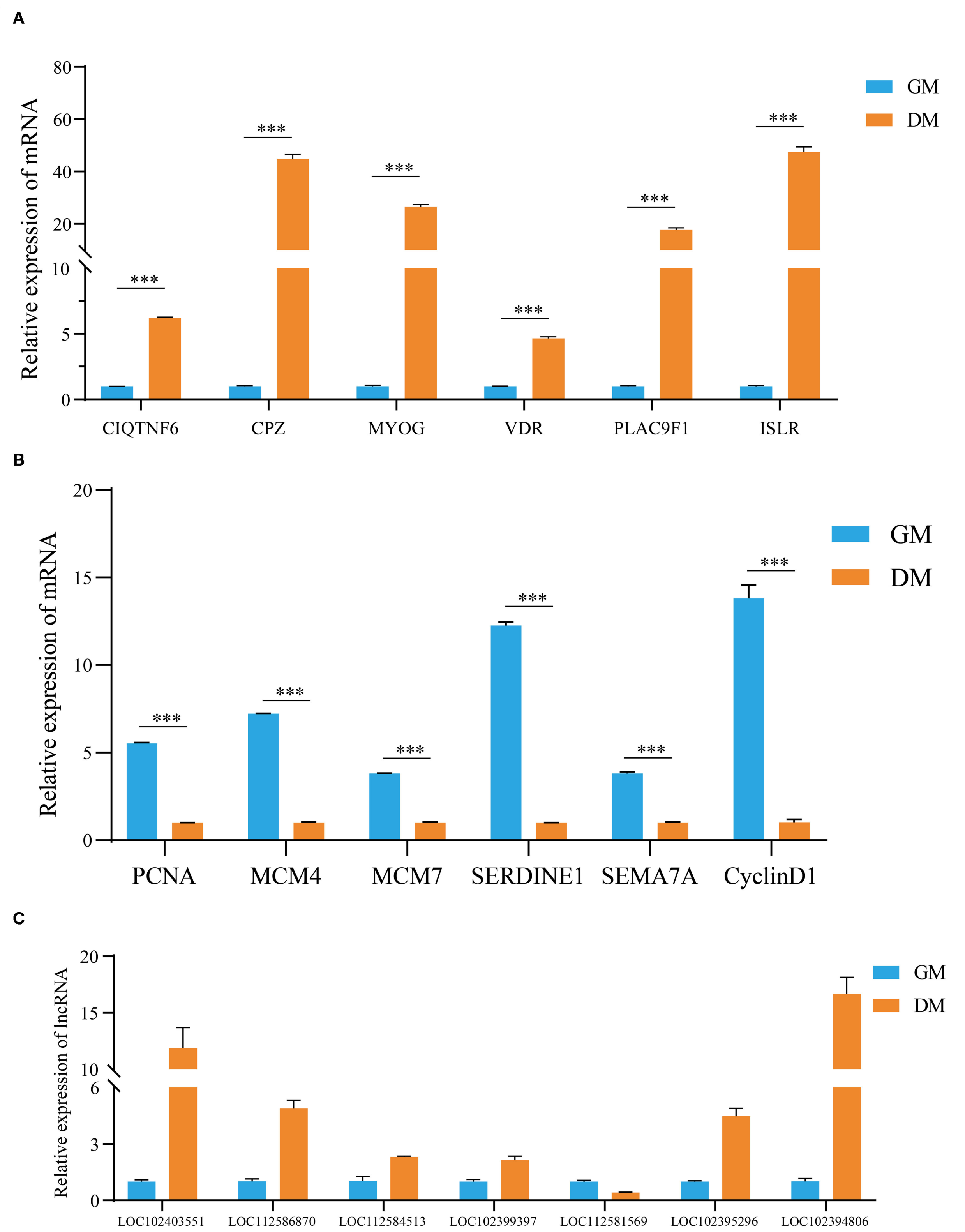

FIGURE 5 | Validation of the expression levels of DEGs, DE lncRNAs by RT-qPCR. **(A,B)** The validation results of DEGs and **(C)** DE lncRNAs. The data are present as means ± SDs, $n = 3$ per group. $^{***}P < 0.001$.

programs. This suggests that *MYLPF* may play an important role in muscle development. Therefore, *MYLPF* was selected for more in-depth study in order to further explore its potential functions in MuSCs.

Subsequently, interference vector (sh-MYLPF-A/B/C/NC) plasmids were introduced into the 293T cell line which is derived from human embryonic kidney 293 cells and contains the SV40 T-antigen. After 24 h, the reporter gene, green fluorescence protein (GFP), was found to be expressed in the transfected cells, with a strong fluorescence observed under the fluorescence microscope (**Supplementary Figure S3**). Further qPCR showed that the expression levels of *MYLPF* declined by 70% in the transfected cells with the sh-MYLPF-A plasmid (**Figure 6C**).

Sh-MYLPF-A vectors were then transferred into P2 MuSCs (**Supplementary Figure S4**). The cells were cultured for a further 24 h, followed by replacement of the medium with myogenic differentiation medium. On the fourth day, the knockdown of *MYLPF* was found to have inhibited the formation of myotubes (**Figure 6D**). The marker gene of myoblast differentiation, *MyoD1*, was then measured by qPCR. The results showed that there were significantly lower levels of *MYLPF* in the knockdown group compared to the controls (**Figure 6E**). These findings suggested that *MYLPF* knockdown inhibited differentiation of MuSCs.

DISCUSSION

Currently, the global population is 7.7 billion, and it is expected to exceed 9 billion by 2050 (Bonny et al., 2015). By then, mankind will face a bigger challenge of food provision for the growing population, and this will have a major impact on global meat consumption which will increase accordingly (Ornaghi et al., 2020). Muscle development is an important factor that affects the growth rate, meat yield, meat quality, and other important economic traits of livestock, and this process is dependent on the proliferation and differentiation of MuSCs (Feige and Rudnicki, 2018; Boscolo Sesillo et al., 2020).

Initially, we established a successful protocol for *in vitro* culture of buffalo MuSCs, which provided a good working foundation for subsequent research on candidate factors that regulate buffalo muscle development. With the rise of *in vitro* cultured meat such as laboratory meat, the role played by MuSCs is becoming more important. The *in vitro* cultured meat production technology is still in its infancy, and it is necessary to strengthen and improve the technical systems involved in MuSCs production of beef (Bhat et al., 2017). Therefore, buffalo MuSCs can play an important role in the emerging research fields of animal husbandry, such as providing improvement of buffalo meat quality and production as well as increasing our biochemical knowledge of MuSCs *in vitro*.

In this study, we constructed the expression profiles of mRNAs and lncRNAs in the process of myogenic differentiation of buffalo MuSCs. During this process, a total of 4,820 genes, 12,227 mRNAs, and 1,352 lncRNAs were differentially expressed. Among them, 2,979 genes, 7,505 mRNAs, and 831 lncRNAs were significantly related to the myogenic differentiation of these cells, and they affected the formation of myoblasts and the fusion of myotubes. In addition, we performed target gene analysis on differentially expressed lncRNAs, and obtained many lncRNAs-target gene relationship networks. We can indirectly predict the function of the candidate lncRNA through the target gene (**Supplementary Tables 7**, **8**). Previous studies had confirmed that compared with cattle, buffalo muscles have larger muscle fiber diameters and poorer meat texture. Of course, myotube fusion is an important factor affecting the formation of muscle fibers (Picard and Gagaoua, 2020). The mRNAs and lncRNAs which are related to myogenic differentiation of MuSCs may regulate the diameter of muscle fibers through myotube fusion, which further affected the quality of meat. We also found some key signal transduction pathways, such as p53 signal transduction pathway, TGF-β signal pathway, calcium signal transduction pathway that were related to these RNAs. These signal pathways are involved in cell development and maintenance of muscle structure and function, suggesting that they were also likely to be important regulatory signals for regulating buffalo muscle fiber hypertrophy (Liu et al., 2018; Valle-Tenney et al., 2020).

We also randomly selected a batch of candidate molecules for verification, and their expression trends were found to be consistent with the RNA-seq results, indicating that the sequencing data was reliable. We found that genes such as *VDR, PLAC9, ISLR* were involved in the myogenic differentiation process of MuSCs, but their molecular mechanisms needed to be further explored (Bass et al., 2020; Cui et al., 2020). It had been confirmed that the immunoglobulin superfamily containing leucine-rich repeats (*ISLR*) promoted skeletal muscle regeneration by activating canonical Wnt signaling. Loss of function of *ISLR* resulted in defective differentiation of myoblasts leading to a block in myotube formation (Zhang et al., 2018). Therefore, *ISLR* may be an important biological regulator to control buffalo muscle development. It had also been reported that *MYLPF* was one of the muscle-derived marker genes involved in the process of muscle metabolism and related to meat quality traits (Rosa et al., 2018; Chong et al., 2020). As one of the muscle markers, *MYLPF* is expected to become a target for regulating the quality traits of buffalo meat (Silva et al., 2019). We also found that decreased *MYLPF* was linked to an inhibition of myogenic differentiation of buffalo MuSCs, but the molecular mechanism of this phenomenon is not yet fully understood. Therefore, how *MYLPF* regulates buffalo muscle regeneration is worthy of further investigation.

At present, lncRNA is also one of the research hotspots in the field of ncRNA (Martone et al., 2019). However, we only discovered the number of known lncRNAs and their expression levels involved in the myogenic differentiation of MuSCs. Then, we screened out a batch of potential candidate lncRNAs, such as LOC102403551, LOC112586870, and LOC102394806. These potential candidate lncRNAs may affect the myoblast differentiation of MuSCs by regulating gene expression through miRNAs, RPBs, and other ways (Chi et al., 2019; Xu et al., 2019; Guo et al., 2020; Liu et al., 2020). In future studies, we and others will also investigate the interactions between lncRNAs and enhancers in order to regulate fate of MuSCs (Lin et al., 2019; Nikonova et al., 2019; Williams et al., 2020). The biological effects

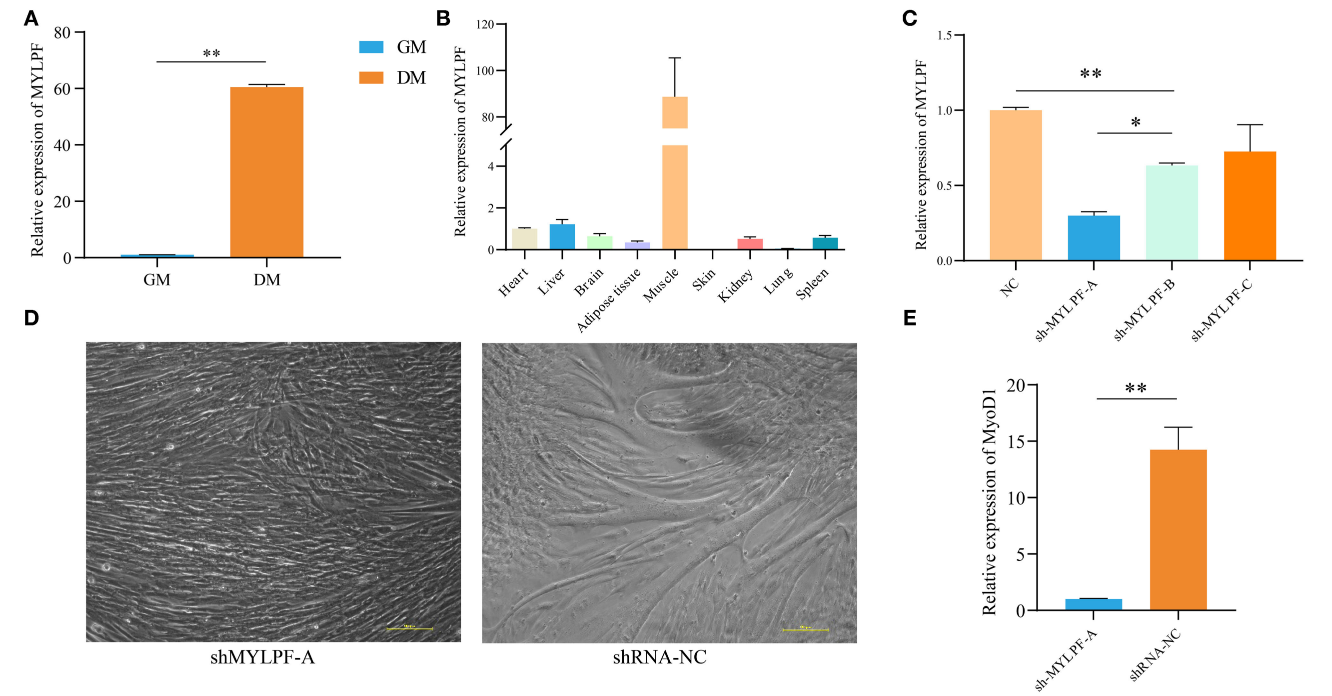

FIGURE 6 | Expression and characterization of *MYLPF* in buffalo skeletal muscle. **(A)** Validation of the expression of *MYLPF* by RT-qPCR. **(B)** The expression levels of *MYLPF* in different tissues of buffalo. **(C)** The interference efficiency of *MYLPF* was measured by RT-qPCR. **(D)** Inhibition of *MYLPF* expression on myotubule formation. **(E)** The mRNA expression of myogenesis marker gene, *MyOD1*, was measured by RT-qPCR. The data are presented as means ± SDs, $n = 3$ per group. $^*P < 0.05$; $^{**}P < 0.01$. Scale bars = 100 µm.

of these lncRNAs related to buffalo MuSCs, lncRNA evolution, lncRNA SNP issues, etc. are also worth pursuing (Qian et al., 2019). At the same time, how these lncRNAs and coding genes regulate the molecular mechanisms of farmed beef production and their contributions to the *in vitro* meat production process also need to be further explored.

In summary, we have established the mRNA and lncRNA expression profiles that regulate the myogenic differentiation of buffalo MuSCs, and further predicted and verified the signaling pathways and candidate regulators involved in cell proliferation and differentiation. These results enrich the expression information of factors that regulate the development of MuSCs in Chinese local fine beef cattle breeds, and provide effective genetic information for future programs of breeding high-yield beef cattle.

CONCLUSIONS

In this study, the proliferation and myogenic differentiation phenotypic characteristics of buffalo MuSCs were compared for the first time, and the expression of mRNAs and lncRNAs in these cells were reviewed. Many coding RNAs and lncRNAs were found to be differentially expressed during the proliferation and myogenic differentiation phases of MuSCs. We further identified and verified a number of differentially expressed molecules such as *SERDINE1*, *ISLR*, *MYLPF*, LOC102403551, LOC112586870, and LOC112584513. This study lays the foundation for further research on the role of lncRNAs in the muscle development of buffalo with a view to improving its share as a desirable beef alternative in the marketplace.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: NCBI GEO, accession no: GSE164808.

AUTHOR CONTRIBUTIONS

YD, SY, YW, JW, and RZ conceived and designed the experiments. RZ, JW, and CZ performed the experiments and analyzed the data. RZ, QA, CZ, XZ, ZW, HL, and DS contributed reagents, materials, and helped to analyze the data. RZ and ZX wrote the manuscript and SY, YD, DS, and YW revised it. All authors contributed to the article and approved the submitted version.

FUNDING

The research was funded by the National Natural Science Foundation of China (31860644 and 31760334), Natural Science Foundation of Guangxi Province (AA17204051, AA17204052, 2017GXNSFBA198170, and 2019GXNSFDA185002), Nanning Scientific Research Technological Development Foundation (20194147 and

20192087), and Guangxi Innovation Team Construction Project of National Modern Agricultural Industry Technology System (nycytxgxcxtd-09-01).

ACKNOWLEDGMENTS

We would like to thank RiboBio (Guangzhou, China) for conducting the whole-genome RNA sequencing. We are also grateful to Dr. Dev Sooranna of Imperial College London for English language edits of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.643497/full#supplementary-material

REFERENCES

Bass, J. J., Nakhuda, A., Deane, C. S., Brook, M. S., Wilkinson, D. J., Phillips, B. E., et al. (2020). Overexpression of the vitamin D receptor (VDR) induces skeletal muscle hypertrophy. *Mol. Metab.* 42:101059. doi: 10.1016/j.molmet.2020.101059

Bhat, Z. F., Kumar, S., and Bhat, H. F. (2017). *In vitro* meat: a future animal-free harvest. *Crit. Rev. Food Sci. Nutr.* 57, 782–789. doi: 10.1080/10408398.2014.924899

Bonny, S. P. F., Gardner, G. E., Pethick, D. W., and Hocquette, J.-F. (2015). What is artificial meat and what does it mean for the future of the meat industry? *J. Integr. Agric.* 14, 255–263. doi: 10.1016/S2095-3119(14)60888-1

Boscolo Sesillo, F., Wong, M., Cortez, A., and Alperin, M. (2020). Isolation of muscle stem cells from rat skeletal muscles. *Stem. Cell Res.* 43:101684. doi: 10.1016/j.scr.2019.101684

Chen, R., Lei, S., Jiang, T., Zeng, J., Zhou, S., and She, Y. (2020). Roles of lncRNAs and circRNAs in regulating skeletal muscle development. *Acta Physiol.* 228:e13356. doi: 10.1111/apha.13356

Chi, Y., Wang, D., Wang, J., Yu, W., and Yang, J. (2019). Long non-coding RNA in the pathogenesis of cancers. *Cells* 8:1015. doi: 10.3390/cells8091015

Chong, J. X., Talbot, J. C., Teets, E. M., Previs, S., Martin, B. L., Shively, K. M., et al. (2020). Mutations in MYLPF cause a novel segmental amyoplasia that manifests as distal arthrogryposis. *Am. J. Hum. Genet.* 107, 293–310. doi: 10.1016/j.ajhg.2020.06.014

Cui, C., Han, S., Shen, X., He, H., Chen, Y., Zhao, J., et al. (2020). ISLR regulates skeletal muscle atrophy via IGF1-PI3K/Akt-Foxo signaling pathway. *Cell Tissue Res.* 381, 479–492. doi: 10.1007/s00441-020-03251-4

Feige, P., Brun, C. E., Ritso, M., and Rudnicki, M. A. (2018). Orienting muscle stem cells for regeneration in homeostasis, aging, and disease. *Cell Stem Cell* 23, 653–664. doi: 10.1016/j.stem.2018.10.006

Feige, P., and Rudnicki, M. A. (2018). Muscle stem cells. *Curr. Biol.* 28, R589–R590. doi: 10.1016/j.cub.2018.02.064

Grigoletto, L., Ferraz, J. B. S., Oliveira, H. R., Eler, J. P., Bussiman, F. O., Abreu Silva, B. C., et al. (2020). Genetic architecture of carcass and meat quality traits in montana tropical[(R)] composite beef cattle. *Front. Genet.* 11:123. doi: 10.3389/fgene.2020.00123

Guo, C. J., Ma, X. K., Xing, Y. H., Zheng, C. C., Xu, Y. F., Shan, L., et al. (2020). Distinct processing of lncRNAs contributes to non-conserved functions in stem cells. 181, 621.e22–636.e22. doi: 10.1016/j.cell.2020.03.006

Hernandez-Hernandez, J. M., Garcia-Gonzalez, E. G., Brun, C. E., and Rudnicki, M. A. (2017). The myogenic regulatory factors, determinants of muscle development, cell identity and regeneration. *Semin. Cell Dev. Biol.* 72, 10–18. doi: 10.1016/j.semcdb.2017.11.010

Huang, K., Chen, M., Zhong, D., Luo, X., Feng, T., Song, M., et al. (2021). Circular RNA profiling reveals an abundant circEch1 that promotes myogenesis and differentiation of bovine skeletal muscle. *J. Agric. Food Chem.* 69, 592–601. doi: 10.1021/acs.jafc.0c06400

Jae, N., and Dimmeler, S. (2020). Noncoding RNAs in vascular diseases. *Circ. Res.* 126, 1127–1145. doi: 10.1161/CIRCRESAHA.119.315938

Jin, C. F., Li, Y., Ding, X. B., Li, X., Zhang, L. L., Liu, X. F., et al. (2017). lnc133b, a novel, long non-coding RNA, regulates bovine skeletal muscle satellite cell proliferation and differentiation by mediating miR-133b. *Gene* 630, 35–43. doi: 10.1016/j.gene.2017.07.066

Li, F., Zheng, Q., Ryvkin, P., Dragomir, I., Desai, Y., Aiyer, S., et al. (2012). Global analysis of RNA secondary structure in two metazoans. *Cell. Rep.* 1, 69–82. doi: 10.1016/j.celrep.2011.10.002

Li, H., Huang, K., Wang, P., Feng, T., Shi, D., Cui, K., et al. (2020). Comparison of long non-coding RNA expression profiles of cattle and buffalo differing in muscle characteristics. *Front. Genet.* 11:98. doi: 10.3389/fgene.2020.00098

Lin, X., Spindler, T. J., de Souza Fonseca, M. A., Corona, R. I., Seo, J. H., Dezem, F. S., et al. (2019). Super-enhancer-associated LncRNA UCA1 interacts directly with AMOT to activate YAP target genes in epithelial ovarian cancer. *iScience* 17, 242–255. doi: 10.1016/j.isci.2019.06.025

Liu, L., Charville, G. W., Cheung, T. H., Yoo, B., Santos, P. J., Schroeder, M., et al. (2018). Impaired notch signaling leads to a decrease in p53 activity and mitotic catastrophe in aged muscle stem cells. *Cell Stem Cell* 23, 544.e4–556.e4. doi: 10.1016/j.stem.2018.08.019

Liu, W., Wang, Z., Liu, L., Yang, Z., Liu, S., Ma, Z., et al. (2020). LncRNA Malat1 inhibition of TDP43 cleavage suppresses IRF3-initiated antiviral innate immunity. *Proc. Natl. Acad. Sci. U.S.A.* 117, 23695–23706. doi: 10.1073/pnas.2003932117

Low, W. Y., Tearle, R., Bickhart, D. M., Rosen, B. D., Kingan, S. B., Swale, T., et al. (2019). Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity. *Nat. Commun.* 10:260. doi: 10.1038/s41467-018-08260-0

Martone, J., Mariani, D., Desideri, F., and Ballarino, M. (2019). Non-coding RNAs shaping muscle. *Front. Cell Dev. Biol.* 7:394. doi: 10.3389/fcell.2019.00394

Mwangi, F. W., Charmley, E., Gardiner, C. P., Malau-Aduli, B. S., Kinobe, R. T., and Malau-Aduli, A. E. O. (2019). Diet and genetics influence beef cattle performance and meat quality characteristics. *Foods* 8:648. doi: 10.3390/foods8120648

Nikonova, E., Kao, S. Y., Ravichandran, K., Wittner, A., and Spletter, M. L. (2019). Conserved functions of RNA-binding proteins in muscle. *Int. J. Biochem. Cell Biol.* 110, 29–49. doi: 10.1016/j.biocel.2019.02.008

Ornaghi, M. G., Guerrero, A., Vital, A. C. P., de Souza, K. A., Passetti, R. A. C., Mottin, C., et al. (2020). Improvements in the quality of meat from beef cattle fed natural additives. *Meat. Sci.* 163:108059. doi: 10.1016/j.meatsci.2020.108059

Picard, B., and Gagaoua, M. (2020). Muscle fiber properties in cattle and their relationships with meat qualities: an overview. *J. Agric. Food Chem.* 68, 6021–6039. doi: 10.1021/acs.jafc.0c02086

Qian, X., Zhao, J., Yeung, P. Y., Zhang, Q. C., and Kwok, C. K. (2019). Revealing lncRNA structures and interactions by sequencing-based approaches. *Trends Biochem. Sci.* 44, 33–52. doi: 10.1016/j.tibs.2018.09.012

Rosa, A. F., Moncau, C. T., Poleti, M. D., Fonseca, L. D., Balieiro, J. C. C., Silva, S. L. E., et al. (2018). Proteome changes of beef in Nellore cattle with different genotypes for tenderness. *Meat Sci.* 138, 1–9. doi: 10.1016/j.meatsci.2017.12.006

Silva, L. H. P., Rodrigues, R. T. S., Assis, D. E. F., Benedeti, P. D. B., Duarte, M. S., and Chizzotti, M. L. (2019). Explaining meat quality of bulls and steers by differential proteome and phosphoproteome analysis of skeletal muscle. *J. Proteomics* 199, 51–66. doi: 10.1016/j.jprot.2019.03.004

Taylor, M. V., and Hughes, S. M. (2017). Mef2 and the skeletal muscle differentiation program. *Semin. Cell Dev. Biol.* 72, 33–44. doi: 10.1016/j.semcdb.2017.11.020

Underwood, J. G., Uzilov, A. V., Katzman, S., Onodera, C. S., Mainzer, J. E., Mathews, D. H., et al. (2010). FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* 7, 995–1001. doi: 10.1038/nmeth.1529

Valle-Tenney, R., Rebolledo, D. L., Lipson, K. E., and Brandan, E. (2020). Role of hypoxia in skeletal muscle fibrosis: synergism between hypoxia and TGF-beta signaling upregulates CCN2/CTGF expression specifically in muscle fibers. *Matrix Biol.* 87, 48–65. doi: 10.1016/j.matbio.2019.09.003

Wang, S., Zuo, H., Jin, J., Lv, W., Xu, Z., Fan, Y., et al. (2019). Long noncoding RNA Neat1 modulates myogenesis by recruiting Ezh2. *Cell Death Dis.* 10:505. doi: 10.1038/s41419-019-1742-7

Wei, X., Li, H., Yang, J., Hao, D., Dong, D., Huang, Y., et al. (2017). Circular RNA profiling reveals an abundant circLMO7 that regulates myoblasts differentiation and survival by sponging miR-378a-3p. *Cell Death Dis.* 8:e3153. doi: 10.1038/cddis.2017.541

Williams, K., Ingerslev, L. R., Bork-Jensen, J., Wohlwend, M., Hansen, A. N., Small, L., et al. (2020). Skeletal muscle enhancer interactions identify genes controlling whole-body metabolism. *Nat. Commun.* 11:2695. doi: 10.1038/s41467-020-16537-6

Xu, X., Ji, S., Li, W., Yi, B., Li, H., Zhang, H., et al. (2017). LncRNA H19 promotes the differentiation of bovine skeletal muscle satellite cells by suppressing Sirt1/FoxO1. *Cell Mol. Biol. Lett.* 22:10. doi: 10.1186/s11658-017-0040-6

Xu, Y., Wu, W., Han, Q., Wang, Y., Li, C., Zhang, P., et al. (2019). New insights into the interplay between non-coding RNAs and RNA-binding protein HnRNPK in regulating cellular functions. *Cells* 8:63. doi: 10.3390/cells8010062

Zhang, K., Zhang, Y., Gu, L., Lan, M., Liu, C., Wang, M., et al. (2018). Islr regulates canonical Wnt signaling-mediated skeletal muscle regeneration by stabilizing Dishevelled-2 and preventing autophagy. *Nat. Commun.* 9:5129. doi: 10.1038/s41467-018-07638-4

Zhu, M., Liu, J., Xiao, J., Yang, L., Cai, M., Shen, H., et al. (2017). Lnc-mg is a long non-coding RNA that promotes myogenesis. *Nat. Commun.* 8:14718. doi: 10.1038/ncomms14718

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

BRIEF RESEARCH REPORT
published: 19 October 2021
doi: 10.3389/fgene.2021.673697

Comparative Signatures of Selection Analyses Identify Loci Under Positive Selection in the Murrah Buffalo of India

*Shiv K. Tyagi[1], Arnav Mehrotra[1], Akansha Singh[1], Amit Kumar[1], Triveni Dutt[2], Bishnu P. Mishra[3] and Ashwni K. Pandey[1]**

[1]Animal Genetics Division, ICAR-Indian Veterinary Research Institute, Izatnangar, Bareilly, India, [2]Livestock Production and Management, Indian Council of Agricultural Research (ICAR)-Indian Veterinary Research Institute, Bareilly, India, [3]Animal Biotechnology, Indian Council of Agricultural Research (ICAR)-Indian Veterinary Research Institute, Bareilly, India

India is home to a large and diverse buffalo population. The Murrah breed of North India is known for its milk production, and it has been used in breeding programs in several countries. Selection signature analysis yield valuable information about how the natural and artificial selective pressures have shaped the genomic landscape of modern-day livestock species. Genotype information was generated on six buffalo breeds of India, namely, Murrah, Bhadawari, Mehsana, Pandharpuri, Surti, and Toda using ddRAD sequencing protocol. Initially, the genotypes were used to carry out population diversity and structure analysis among the six breeds, followed by pair-wise comparisons of Murrah with the other five breeds through XP-EHH and F_{ST} methodologies to identify regions under selection in Murrah. Admixture results showed significant levels of Murrah inheritance in all the breeds except Pandharpuri. The selection signature analysis revealed six regions in Murrah, which were identified in more than one pair-wise comparison through both XP-EHH and F_{ST} analyses. The significant regions overlapped with QTLs for milk production, immunity, and body development traits. Genes present in these regions included *SLC37A1*, *PDE9A*, *PPBP*, *CXCL6*, *RASSF6*, *AFM*, *AFP*, *ALB*, *ANKRD17*, *CNTNAP2*, *GPC5*, *MYLK3*, and *GPT2*. These genes emerged as candidates for future polymorphism studies of adaptability and performance traits in buffaloes. The results also suggested ddRAD sequencing as a useful cost-effective alternative for whole-genome sequencing to carry out diversity analysis and discover selection signatures in Indian buffalo breeds.

Keywords: ddRAD, genotypes, bubalus, Fst, XP-EHH

OPEN ACCESS

Edited by:
*Yang Zhou,
Huazhong Agricultural University, China*

Reviewed by:
*Pablo Fonseca,
University of Guelph, Canada
Monika Sodhi,
National Bureau of Animal Genetic Resources (NBAGR), India*

****Correspondence:***
*Ashwni K. Pandey
ashwni.pandey@gmail.com*

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *28 February 2021*
Accepted: *17 September 2021*
Published: *19 October 2021*

Citation:
Tyagi SK, Mehrotra A, Singh A, Kumar A, Dutt T, Mishra BP and Pandey AK (2021) Comparative Signatures of Selection Analyses Identify Loci Under Positive Selection in the Murrah Buffalo of India. Front. Genet. 12:673697. doi: 10.3389/fgene.2021.673697

INTRODUCTION

Water buffalo is considered as an important livestock resource in tropical and sub-tropical countries due to its high milk production ability along with adaptability to hot and humid environment, and high feed conversion efficiency (Kumar et al., 2019). Buffaloes are the major contributors of milk production in India accounting for 49.2% of 187.7 million tons of total milk production (DAHD&F, 2018). India possesses a remarkably large and diverse buffalo population with 109.85 million buffaloes and 17 registered breeds (DAHD&F, 2018; NBAGR Karnal, 2021).

Murrah is the most important buffalo breed of India, constituting about 44.3% of the total buffalo population of the country. The main breeding area of this breed is the northern states of India, namely Punjab, Haryana, and Western Uttar Pradesh. Due to its high milk potential in varied environmental conditions, the germplasm of the breed has been extensively used throughout the country. It has also been imported in several countries like China, Brazil, Vietnam, Egypt, Bangladesh, etc., due to its higher milk production potential (Zhang et al., 2020). As part of the breed improvement schemes, Murrah buffalo has been selected for improved milk production for the past 30 years, and the process is going on. By investigation of selection sweeps in the Murrah genome, we may gain insights into the genes and genomic regions related to important economic traits in buffaloes. Recently, Dutta et al. (2020) identified selection sweeps in seven Indian riverine buffaloes and compared patterns of between-species selective sweeps with different cattle breeds using whole-genome sequencing (WGS) data. Since WGS is a costly process, several workers have proposed reduced representation genotyping techniques such as the double digest restriction site-associated DNA sequencing (ddRAD-seq) as a useful alternative to WGS for genotyping Indian buffaloes (Surya et al., 2019; Mishra et al., 2020). For the present study, the genotype data of six Indian buffalo breeds (Murrah, Surti, Mehsana, Bhadawari, Pandharpuri, and Toda) was generated using ddRAD sequencing.

This study aimed to assess the genetic diversity and population structure among the six Indian buffalo breeds using ddRAD data. Furthermore, we attempted to unravel signatures of positive selection in Murrah by comparing it with other reference Indian breeds (Surti, Mehsana, Bhadawari, Pandharpuri, and Toda) through cross-population extended haplotype homozygosity (XP-EHH) and cross-population fixation index (F_{ST}) approaches.

MATERIAL AND METHODS

Sample Collection and Generation of Double Digest Restriction Site-Associated DNA Data

Ninety-six samples were collected from six breeds of riverine buffalo from different parts of India. These breeds are diverse in terms of physical features, milk production, and adaptation. Selection of the animals was done in a way to cover the genepool of the respective breeds. So the animals of all the breeds in the present study were chosen randomly from their respective institutional farms (except animals of the Toda breed of buffalo for which random samples were collected from its breeding tracts in the Nilgiri Hills area of Tamilnadu state of India). As the Murrah breed is mainly found in the northern part of India, the random samples were collected from three institutional farms of the area, i.e., the Livestock Research Station (LRS) ICAR-IVRI situated in Izatnagar, Bareilly (Uttar Pradesh), the Buffalo Farm at livestock research station of GBPUA and T, Pantnagar (Uttarakhand), and the Livestock Farm, GADVASU Ludhiana. The samples of Bhadawari buffalo were collected from the Buffalo Farm, ICAR-IGFRI, Jhansi (Uttar Pradesh), Mehsana buffalo samples were collected from the Livestock Research Station, SDAU, SK Nagar (Gujarat), Surti buffalo samples were collected from the Livestock Research Station, CVAS, Udaipur (Rajasthan), and Pandharpuri buffalo samples were collected from the Buffalo Farm, Zonal Agriculture Research Station, Kolhapur (Maharashtra). All these farms are situated in their respective breeding tract, and animals were randomly selected from these institution farms as to cover substantially the genepool of the population. The breed-wise details of sample numbers and location are also provided in **Supplementary Table S1**. Whole-blood samples were collected from the jugular vein of the animals in 10-ml vacutainers under aseptic condition, and genomic DNA was extracted using the standard phenol–chloroform method (Sambrook and Russell, 2006). The concentration and purity of the DNA were measured using agarose gel electrophoresis and NanoDrop spectrophotometer. Following the ddRAD protocol (Peterson et al., 2012), the double digestion of genomic DNA was carried out using Sph I and MluC I enzymes as mentioned in Kumar et al. (2020), and the samples were sequenced on Illumina Hi-seq 2000 platform to generate 150-bp reads.

Quality Control and Variant Calling

The reads were quality checked using FastQC (Andrews, 2010). Trimming of Illumina universal adapters and quality filtering was performed by the *process_radtags* function of the STACKS v2 software (Rochette et al., 2019). Reads were examined using a sliding window spanning 15% of the read length, and the reads having average phred score of <15 were discarded. The barcode of the reads was removed using Cutadapt 2.10 (Martin, 2011).

The paired reads were aligned to the *Bubalus bubalis* assembly UOA_WB_1 downloaded from NCBI (Low et al., 2019; https://www.ncbi.nlm.nih.gov/assembly/GCF_003121395.1/) using BWA-MEM 0.7.17 (Li, 2013) with default settings. The percentage of reads aligning to the reference genome was determined by Samtools (v1.7) flagstats (Li et al., 2009) function. Variant calling was performed through the bcftools mpileup utility of the Samtoolsv1.7 suite in a multi-sample mode as recommended by Wright et al. (2019). SNPs with quality score greater than 30 and a read depth of 10 were retained for further analysis.

The structural and functional annotation of the retained SNPs was performed using SnpEff v4.3 (Cingolani et al., 2012). Quality filtering of the annotated variants was performed by removing unmapped and non-autosomal SNPs. SNPs missing in more than 25% of the individuals and below the minor allele frequency (MAF) threshold of 0.01 were also filtered out using PLINK 1.9 (Purcell et al., 2007). Genotype imputation of sporadically missing genotypes was done using Beagle 4.1 (Browning and Browning, 2016).

Genetic Diversity and Population Structure Analysis

Linkage disequilibrium (LD) pruning of the SNPs was carried out using the *indep-pairwise* command parameters (indep-pairwise 50 5 0.2) of the PLINK software. The observed (Ho) and expected (He) heterozygosities for different buffalo breeds were estimated using PLINK 1.9. Furthermore, admixture analysis was performed on the LD pruned data for K values ranging from K = 2 to K = 6 using ADMIXTURE 1.3 software (Alexander et al., 2009). The results of the admixture analysis were visualized using PONG (Behr et al., 2016). A genomic relationship matrix was prepared in GCTA (Yang et al., 2011), and the first 10 principal components were extracted. The top principal components were plotted in R (R Core team, 2018) to visualize population clustering. A maximum-likelihood phylogram was constructed using TREEMIX (Pickrell and Pritchard, 2012) to infer the ancestral relationships and migration patterns between the breeds.

Analysis of Selection Signatures

Cross-population selection signatures between Murrah buffalo and five other Indian water buffalo breeds (Bhadawari, Surti, Mehsana, Pandharpuri, and Toda) were derived using XP-EHH (Sabeti et al., 2007) and F_{ST} (Weir and Cockerham, 1984) methodologies. The genotypic data of all the breeds were phased using BEAGLE v5.1 (Browning et al., 2018) using default settings (burnin = 6; iterations = 12; and phase-states = 280). The XP-EHH scores of the Murrah buffalo were calculated for each breed comparison using the R package *rehh* (Gautier et al., 2017), taking the other water buffalo breeds in the study as the reference populations. To detect positive selection, average XP-EHH scores were computed for 100-kb regions with a 50-kb overlap. Regions with absolute XP-EHH scores of four or above were considered as putative candidate regions in Murrah.

The pairwise F_{ST} estimates between the Murrah and other buffalo breeds were calculated with VCFTOOLS (Danecek et al., 2011), with a sliding window of 100 kb and a 50-kb step size. Windows belonging to the top 0.1% of the F_{ST} values were considered as potential regions under selection (Singh et al., 2020).

The candidate genes in the selected regions were annotated using the GTF (gene transfer format) file supplied with the UOA_WB_1 assembly, using BEDTools (Quinlan and Hall, 2010) *intersect* function. Each putative selected region was cross-referenced with the literature to find previously detected regions of functional importance.

RESULTS

In the present study, total 397.8 million paired-end reads of 150-bp length were obtained for the 96 buffalo breeds, averaging 4.14 million reads per sample. After initial quality control, a total of 367.2 million reads (92.3% of the total reads) of average 135-bp length were retained. The average alignment rate of the reads was 99.82% with the reference genome. Sample-wise alignment percentages are given in **Supplementary Table S2**. A total of 569,535 variants were identified, out of which 502,476 were SNPs and 67,059 were indels. A total of 551,458 variants were present on autosomes, 15,315 on the X chromosome and 12 on the mtDNA, and 2,750 variants were located on unmapped contigs (**Supplementary Table S3**). A variant was discovered for every 4,637 bp of the genome length. The total number of SNPs and indels of each buffalo breed at read depth 10 is mentioned in **Supplementary Table S4**. The highest number of SNPs was found for the Mehsana (489,738) buffalo followed by the Murrah buffalo (484,449), and lowest for the Toda buffalo (448,714). After quality control and imputation of sporadically missing genotypes, a total of 237,762 SNPs, which were common across all the breeds, were used for downstream analysis.

TABLE 1 | Number of animals, means of observed (HO) and expected heterozygosity (HE), and differentiation (FST) between each breed and the Murrah.

S.No	Breeds	Number of animals	Ho	He	F_{ST}
1	Murrah	30	0.2372	0.2462	-
2	Bhadawari	15	0.2343	0.2366	0.11
3	Mehsana	15	0.2314	0.2239	0.17
4	Surti	15	0.2361	0.2255	0.09
5	Pandharpuri	15	0.2366	0.2390	0.15
6	Toda	6	0.2150	0.2111	0.13

Genome-wide Annotation of SNPs in Water Buffalo Breeds

Based on the sequence ontology terms, a greater number of identified SNPs were located within the intronic regions (66.57%), followed by the intergenic regions (22.13%), and 0.34% of SNPs were found to be located in the transcript region (**Supplementary Figure S1**). The impact-wise and region-wise distribution of variant effects, as generated by SNPeff, are given in **Supplementary Table S5**.

About 71.89% of the annotated SNPs were identified as transitions (Ts) while 28.10% as transversions (Tv) with a T_S/T_V ratio of 2.5578. The Ts/Tv ratio serves as a quality control indicator of high-throughput sequencing data. Our values are consistent with previous reports of targeted sequencing methods in buffalo (Surya et al., 2019; Kumar et al., 2020).

Genetic Diversity

For the genetic diversity and population structure analyses, we used a subset of 67,798 SNPs after pruning the SNPs in LD. The average observed heterozygosity (Ho) and expected heterozygosity (He) of all breeds in the study are presented in **Table 1**. The Ho and He was found highest for the Murrah (0.237 and 0.246) and lowest for the Toda (0.215 and 0.211). The genetic distances (F_{ST}) of the Murrah with the Bhadawari, Mehsana, Surti, Pandharpuri, and Toda were 0.11, 0.17, 0.09, 0.15, and 0.13, respectively.

Population Structure

The population structure of the Indian water buffalo breeds was identified using PCA. The first and second principal component (PC) explained 3.4 and 2.86% of the total variance. PC1 separated the crossbred Mehsana individuals from the rest of the breeds, while PC2 separated the Pandharpuri, Surti, and Toda from the Murrah and Bhadawari (**Figure 1A**). PC3 explained 2.71% of the

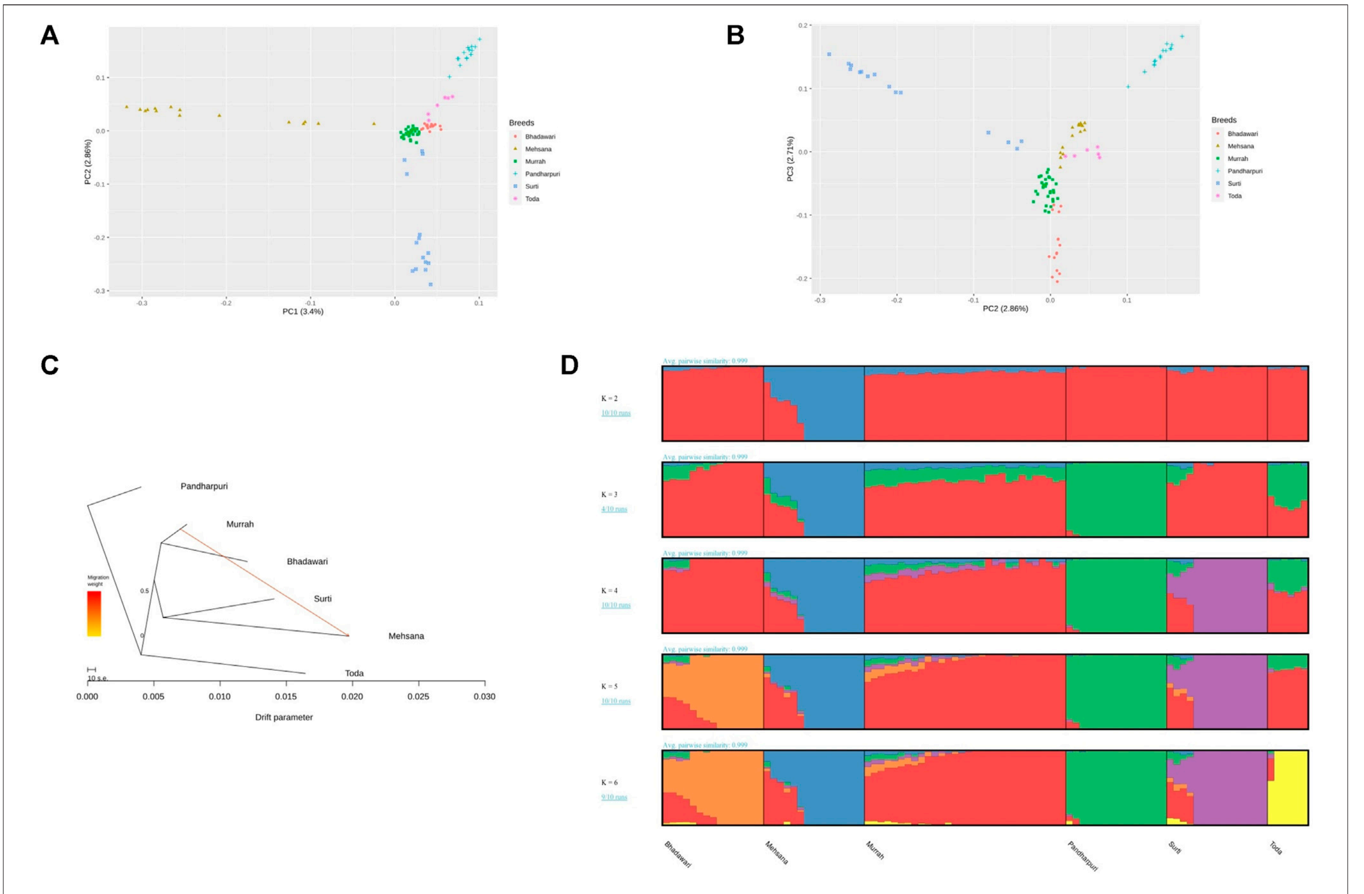

FIGURE 1 | (**A**) Plot of the first two principal components (PC1 and PC2) of the genomic relationship matrix of the 96 animals under study. (**B**) Plot of PC2 and PC3. (**C**) Treemix phylogram showing one migration path. (**D**) Admixture results from K = 2 to K = 6.

total variation and showed clear separation between the Murrah and Bhadawari (**Figure 1B**).

The maximum-likelihood phylogram constructed with Treemix also displayed a similar tree (**Figure 1C**). The addition of one migration path in Treemix revealed the introgression of the Murrah inheritance in the Mehsana buffaloes. This tree explained 99.6% of the covariance observed between populations, whereas the tree without any migration events included explained only 98.3% of the covariance.

As seen with PC1, the Mehsana was separated from the rest of the breeds at K = 2 in the admixture analysis. K = 3 separated the Pandharpuri as a distinct population from the rest of the breeds, which gives credence to the results of the phylogenetic analyses. The Toda samples in our study showed a mixture of Pandharpuri and Murrah inheritance. At K = 6, all the breeds were assigned to their own clusters, with varying levels of Murrah ancestry appearing in other breeds (Bhadawari, Mehsana, Surti, and Toda) (**Figure 1D**).

Cross-Population Signatures of Selection (XP-EHH and F_{ST})

The distribution of XP-EHH scores for the Murrah buffalo (positive values) against other water buffalo breeds in the study is visualized in **Figure 2**. A total of 164 putative selection regions for the Murrah buffalo were identified in comparison with the reference breeds (**Supplementary Table S6**). Ten selection sweeps were detected in comparisons of the Murrah with more than one breed (**Table 2**).

The Manhattan plot for pairwise F_{ST} across all comparisons are shown in **Figure 3**. A total of 58 positive regions were identified from all comparisons. The selection sweeps were located on all autosomes except for chromosome 5, 14, and 21. The highest number of selected regions were identified on chromosome 8 (seven regions), followed by chromosomes 1, 9, and 10 from all pairwise comparisons (**Supplementary Table S7**).

A total of six fully or partially overlapping selection sweeps were identified from both the approaches XP-EHH and F_{ST} (**Table 3**). These regions were distributed on chromosomes 1, 7, 8, 13, 15, and 18.

DISCUSSION

In the present study, ddRAD sequencing was used to identify genetic variants in six water buffalo breeds of India. The average heterozygosity levels ranged from 0.215 to 0.237, which were lower compared with a previous study (Kumar et al., 2006).

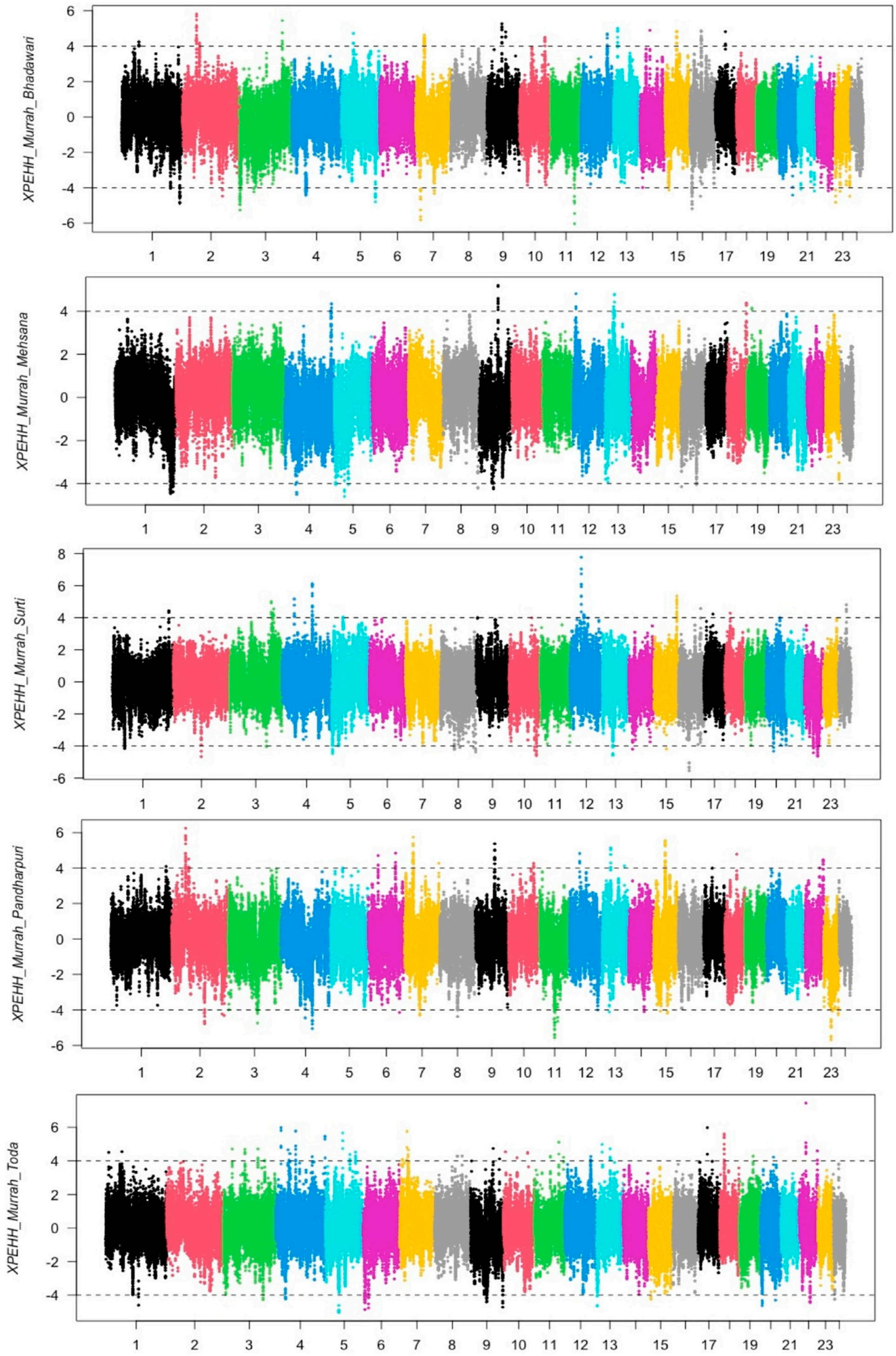

FIGURE 2 | Cross-population extended haplotype heterozygote (XP-EHH) plot of the Murrah in comparison with the Bhadawari, Mehsana, Surti, Pandharpuri, and Toda.

However, they used microsatellite data, which suffers from ascertainment bias due to the most polymorphic microsatellite markers being studied, resulting in inflated heterozygosity estimates (Fischer et al., 2017). The population structure analysis separated the six breeds under study. Our findings confirmed two existing notions about the Indian buffaloes. First, it has been traditionally believed that the Mehsana breed is of the Murrah and Surti lineage (Patel et al., 2017; Sathwara

TABLE 2 | Common selection sweeps identified by cross-population extended haplotype homozygosity (XP-EHH) in two or more pairwise comparisons involving the Murrah.

S.No	References breeds	Chr	Start	End	Annotated gene
1	Bhadawari Mehsana	1	192,319,897	192,322,098	*LOC112580862*
2	Bhadawari Pandharpuri	2	56,674,658	56,740,551	*HS6ST1*
3	Bhadawari Surti	3	143,278,931	143,620,455	*DAPK1, CTSL, FBP2*
4	Surti Toda	4	41,323,382	41,449,515	*IP O 8, CAPRIN2*
5	Bhadawari Toda Pandharpuri	7	28,640,078	30,146,985	*AFM, AFP, ALB*
6	Mehsana Pandharpuri	9	64,216,990	64,326,407	*NEUROG1, TIFAB*
7	Bhadawari Pandharpuri	10	84,290,283	84,562,847	*BCKDHB*
8	Toda Bhadawari	12	86,340,919	86,501,726	*KCNF1*
9	Toda Bhadawari	20	49,776,417	49,968,750	*LOC112580801*
10	Pandharpuri Bhadawari	23	48,880,371	49,056,564	*LOC112580801*

et al., 2020). The maximum-likelihood phylogram constructed using Treemix in our study showed the Mehsana and Surti emerging from the same node in the phylogenetic tree, with introgression of the Murrah germplasm into the Mehsana, which supports the anecdotal knowledge about this breed. The admixture analyses also showed varying levels of Murrah inheritance into the Mehsana breed. Second, the western Indian buffalo, the Pandharpuri, formed a separate lineage from the rest of the breeds and appeared free of any Murrah inheritance, which was in agreement with previous studies (Kumar et al., 2006). However, in our study, the geographically distinct semi-wild Toda breed clustered with the Murrah. Admixture analysis showed all the Toda samples to contain significant levels of Murrah inheritance, which is a cause for concern. The samples were collected from the hamlets of the Toda tribes, situated in the jungles in and around Nilgiris district. In the 1990s, some of the Murrah bulls were introduced in Toda hamlets near small towns. This may be one of the reasons for the inheritance of the Murrah in Toda, which is reflected predominantly due to only six samples taken in the study.

The second objective of this study was to identify positive signatures of selection in the Murrah buffaloes. Humans have exerted strong artificial selection on different breeds of buffalo for similar traits since domestication (Dutta et al., 2020). Probably milk production formed the basis of selection and breeding, which resulted in the evolution of the dairy breeds of the farmers of riverine buffalo like the Murrah, Bhadawari, Mehsana, Surti, Pandharpuri, etc. (CIRB, Hisar, 2017). The Toda, on the other end is a semi wild breed purposely used for religious values from the past in the Nilgiri hills. These breeds may share mutations in the same gene(s) or regulatory region and, consequently, may have selective sweeps in the same area of the genome. However, the scope of selective sweeps may differ among breeds sharing mutations in the same genes because of differences in breed history, effective population size, and mutation rate (Pollinger et al., 2005), and also, differences may be caused by large environmental variations and different managemental practices throughout the country.

The positive signatures of selection in the Murrah buffaloes were identified using XP-EHH and F_{ST} approaches. Several fully or partially overlapping candidate regions in Murrah were identified through XP-EHH comparisons against more than one breed, which indicated recent artificial selection in the Murrah, given the characteristics of the XP-EHH test (Cheruiyot et al., 2018). Many of these regions overlap with previous reports in the Murrah.

On chromosome 1, a region was identified around the 192.2 Mb position against the Bhadawari, Mehsana, and Toda, which was in agreement with Dutta et al. (2020). This region includes *UPK1B* (Uroplakin 1 B), *B4GALT4* (Beta-1,4-Galactosyltransferase), and *ARHGAP31* (Rho GTPase-activating protein 31) genes, which could be putative candidate genes undergoing selection in the Murrah. The *UPK1B* and *ARHGAP31*genes have previously been linked with growth and carcass traits in cattle breeds (Kim et al., 2012; Medeiros de Oliveira Silva et al., 2017). Another partly overlapping region (17.4–17.5 Mb) in agreement with Dutta et al. (2020) was identified on chromosome two against the Pandharpuri. The region includes *FABP3* (fatty acid-binding protein 3) gene, which is involved in the synthesis of long-chain fatty acid and, thus, regulates milk fat composition (Li et al., 2014).

A selection sweep (28.5–29.1 Mb) on chromosome seven in comparisons of the Murrah with the Pandharpuri, Toda, and Bhadawari also confirms a previously reported selection sweep

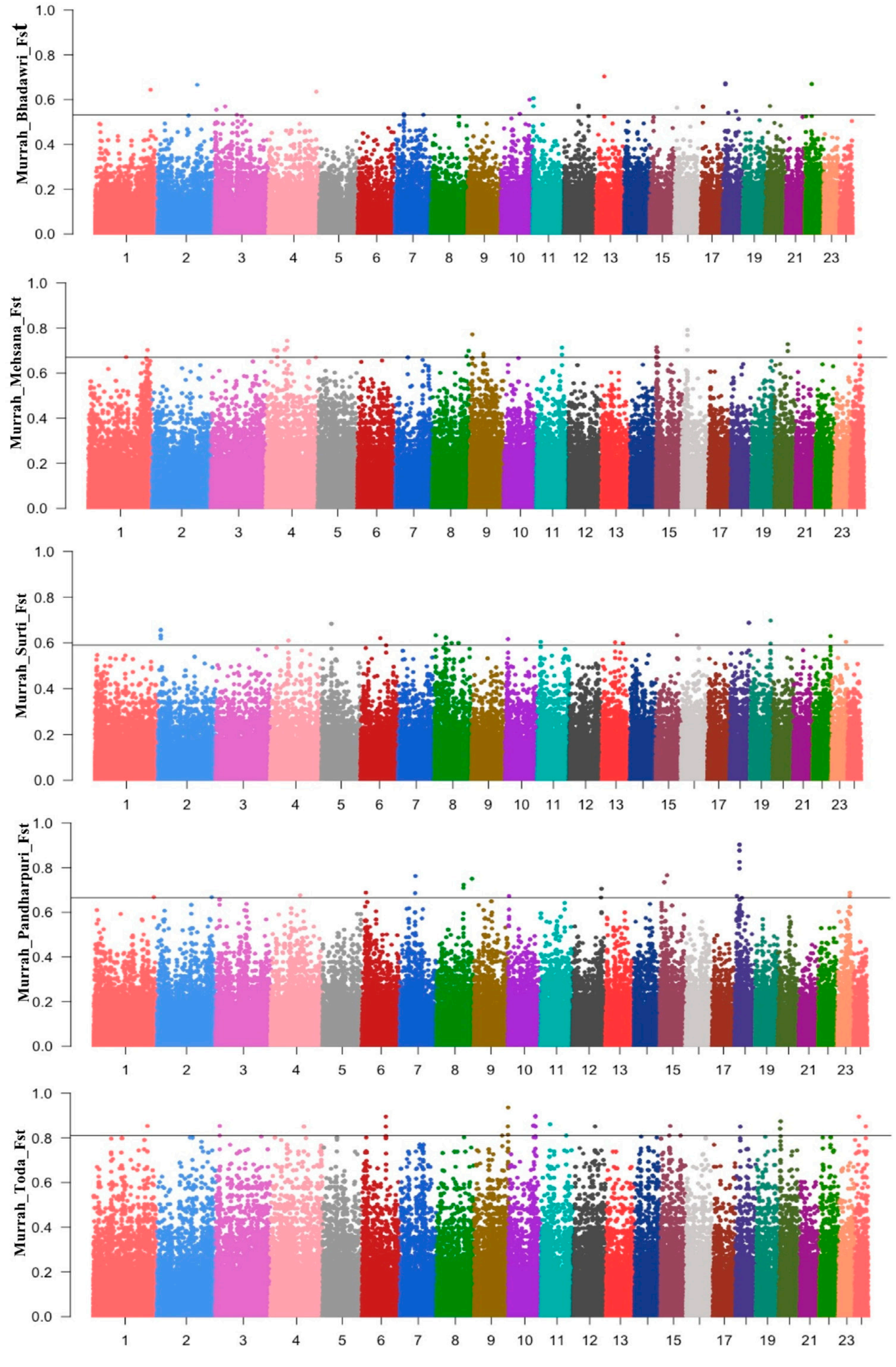

FIGURE 3 | Manhattan plot for FST between the Murrah in comparison with the Bhadawari, Mehsana, Surti, Pandharpuri, and Toda.

(chromosome 7, 26.5–30.5 Mb) in the Murrah genome by Dutta et al. (2020). This region contains *ALB*, *AFP*, and *AFM* belonging to the family of albumin genes. The *ALB* (albumin) gene encodes albumin protein, which is involved in the transportation of varied endogenous molecules. *ALB* was reported to be significantly associated with total milk yield, milk fat, and protein percentage in the Holstein cattle (Seo et al., 2016) and obesity in humans (Kunej et al., 2013).

In agreement with Dutta et al. (2020), two regions on chromosome 13 (23.4–24.9 Mb) and chromosome 18

TABLE 3 | Selection signatures in the Murrah identified by both XP-EHH and F_{ST} approaches.

S. No	Test	Chr	Start	End	Genes
1	XP-EHH (Surti); F_{ST} (Mehsana)	1	187,322,925	187,600,000	*SLC37A1, PDE9A*
2	XP-EHH (Pandharpuri, Bhadawri, Toda); F_{ST} (Bhadawari)	7	28,553,887	29,108,103	*PPBP, CXCL6, RASSF6, AFM, AFP, ALB, ANKRD17*
3	XP-EHH (Surti); F_{ST} (Mehsana)	8	109,432,200	1,117,495,711	*CNTNAP2*
4	XP-EHH(Pandharpuri); F_{ST} (Bhadawari)	13	23,401,830	24,977,050	*GPC5*
5	XP-EHH (Pandharpuri); F_{ST} (Pandharpuri)	15	22,545,641	22,557,701	*LOC112579137*
6	XP-EHH (Toda); F_{ST} (Pandharpuri)	18	14,622,913	14,929,335	*C18H16orf87, MYLK3, GPT2*

(14.6–14.9 Mb) were identified in our study. The region on chromosome 13 included *GPC5* (glypican 5) gene, which is linked with fatty acid composition (Li et al., 2014), fertility traits (Purfield et al., 2019), and feed efficiency (Serão et al., 2013) in cattle. The *MYLK3* (myosin light chain kinase 3) and *GPT2* (glutamic pyruvic transaminase 2) genes on chromosome 18 are involved in muscle cell development (Silva-Vignato et al., 2019; Cheng et al., 2020) and Ca^{+2} signaling pathway in contraction of striated muscles (Zhang et al., 2009).

In addition, several novel regions of positive selection were also identified. These regions contain candidate genes, which are associated with the phenotypes that are under selection in the Murrah buffalo, including milk production and fat metabolism (*HS6ST1*, *FBP2*, and *PDE9A*), immunity-related pathways (*DAPK1*), stature (*CTSL*), and fertility traits (*KCNF1* and *CNTNAP2*) (Jiang et al., 2011; Abo-Ismail et al., 2017; Guan et al., 2020). The regions included *HS6ST1* (heparin sulfate 6-O sulfotransferase 1) gene located on chromosome 2, which plays a pivotal role in heparin metabolism pathway and regulates the fatty acid composition (Jiang et al., 2011). Another region on chromosome 3 contains *DAPK1* (death-associated protein kinase 1), *CTSL* (cathepsin L), and *FBP2* (fructose bisphophatase 2) genes, which are involved in various metabolic processes such as immunity and milk production (Vineeth et al., 2019; Guan et al., 2020). The *KCNF1* (potassium voltage-gated channel modifier subfamily F member 1) gene on chromosome 12 has been previously reported to be associated with fertility traits in buffaloes (de Araujo Neto et al., 2020). Another candidate region spanning 280 kb on chromosome 1, which was detected by both approaches, contains *PDE9A* gene (phosphodiesterase 9A). This gene is involved in the signaling pathway, which regulates the level of cGMP inside the cell. Yang et al. (2015) has reported the strong association of *PDE9A* gene with milk production in Chinese Holstein cattle. On chromosome 8, *CNTNAP2* (contactin-associated protein 2) gene was present in a significant region. This gene has been reported to be associated with immunity and growth traits in cattle (Abo-Ismail et al., 2017). *CNTNAP2* gene is also reported to play an important role in milk synthesis pathway in water buffalo (Mishra et al., 2020). These positively selected genes may create the observed differences in the Murrah buffaloes from the rest of the buffalo breeds included in the study and makes the Murrah as one of the high milk-producing buffalo breed with high fertility and immunity.

CONCLUSION

The genetic diversity and population structure analysis revealed varying levels of the Murrah inheritance in the Bhadawari, Mehsana, Surti, and Toda buffalo breeds. The selection signature analysis provides several genomic regions as selection signature in the Murrah, which is the prominent milch breed in India. Using reduced representation ddRAD data, our results confirm many regions, which have been previously identified as selection sweeps in the Murrah genome using WGS data. In addition, novel regions were also identified, which are involved in several biological pathways. The candidate genes, found to be positively selected, are involved in milk production (*ALB*, *FBP2*, *PDE9A*, and *GPC5*), immunity-related traits (*DAPK1*), muscle cell development (*MYLK3* and *GPT2*), and fertility traits (*KCNF1* and *CNTNAP2*). These genes are suitable candidates for future polymorphism studies to detect causative variants associated with these phenotypes in buffaloes.

DATA AVAILABILITY STATEMENT

The genotypes of the 96 individuals under study have been uploaded to Figshare under the DOI https://doi.org/10.6084/m9.figshare.14130389.v1. The data will be made public upon acceptance of the article.

ETHICS STATEMENT

The study was carried out in accordance with recommendation of Institute Animal Ethics Committee of ICAR-IVRI, Bareilly India.

AUTHOR CONTRIBUTIONS

AP and AK conceived and designed the experiments. ST performed the experiments. AM and AS analyzed the data and wrote the manuscript. TD and BM contributed reagents/materials/analysis tools. ST, AK and AP edited the manuscript.

FUNDING

This project work was supported by the CAAST-ACLH project of NAHEP, and institute funding of ICAR-IVRI.

ACKNOWLEDGMENTS

The authors acknowledge the logistic and laboratory support rendered by the Director ICAR-IVRI. We also thank Dr. P.P Dubey and Dr. Puneet Malhotra (GADVASU), Dr. B.N. Sahi (GBPUA&T), Dr. B.P. Kushwaha (ICAR-IGFRI, Jhansi), Dr. Mitesh Gaur (LRS, Vallabhnagar), Director of Research (SDAU, SK Nagar, Gujarat), Dr. R.S. Kataria (ICAR-NBAGR), Dr. Jayakumar S. (ICAR-Directorate of Poultry Research), and Dr. A. P. Fernandes (MPKV Rahuri) for their contribution in providing the samples used in this study. We also wish to acknowledge Prof. Hubert Pausch (Animal Genomics, ETH Zurich), for his inputs regarding the SNP calling pipeline.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.673697/full#supplementary-material

REFERENCES

Abo-Ismail, M. K., Brito, L. F., Miller, S. P., Sargolzaei, M., Grossi, D. A., Moore, S. S., et al. (2017). Genome-Wide Association Studies and Genomic Prediction of Breeding Values for Calving Performance and Body Conformation Traits in Holstein Cattle. *Genet. Sel. Evol.* 49, 1–29. doi:10.1186/s12711-017-0356-8

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome Res.* 19, 1655–1664. doi:10.1101/gr.094052.109

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available at: at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Behr, A. A., Liu, K. Z., Liu-Fang, G., Nakka, P., and Ramachandran, S. (2016). Pong: Fast Analysis and Visualization of Latent Clusters in Population Genetic Data. *Bioinformatics* 32, 2817–2823. doi:10.1093/bioinformatics/btw327

Browning, B. L., and Browning, S. R. (2016). Genotype Imputation with Millions of Reference Samples. *Am. J. Hum. Genet.* 98, 116–126. doi:10.1016/j.ajhg.2015.11.020

Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am. J. Hum. Genet.* 103, 338–348. doi:10.1016/j.ajhg.2018.07.015

Cheng, J., Cao, X., Hanif, Q., Pi, L., Hu, L., Huang, Y., et al. (2020). Integrating Genome-Wide CNVs into QTLs and High Confidence GWAScore Regions Identified Positional Candidates for Sheep Economic Traits. *Front. Genet.* 11, 569. doi:10.3389/fgene.2020.00569

Cheruiyot, E. K., Bett, R. C., Amimo, J. O., Zhang, Y., Mrode, R., and Mujibi, F. D. N. (2018). Signatures of Selection in Admixed Dairy Cattle in Tanzania. *Front. Genet.* 9, 607. doi:10.3389/fgene.2018.00607

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff. *Fly* 6, 80–92. doi:10.4161/fly.19695

CIRB, Hisar (2017). Available at: https://cirb.icar.gov.in/history/. (Accessed February 25, 2021).

DAHD&F (2018). *Department of Animal Husbandry Dairying & Fisheries*. New Delhi: Ministry of Fisheries, Animal Husbandry & Dairying.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The Variant Call Format and VCFtools. *Bioinformatics* 27, 2156–2158. doi:10.1093/bioinformatics/btr330

de Araujo Neto, F. R., Takada, L., Santos, D. J. A., Aspilcueta-Borquis, R. R., Cardoso, D. F., Nascimento, A. V., et al. (2020). Identification of Genomic Regions Related to Age at First Calving and First Calving Interval in Water buffalo Using Single-Step GBLUP. *Reprod. Dom Anim.* 55, 1565–1572. doi:10.1111/rda.13811

Dutta, P., Talenti, A., Young, R., Jayaraman, S., Callaby, R., Jadhav, S. K., et al. (2020). Whole Genome Analysis of Water Buffalo and Global Cattle Breeds Highlights Convergent Signatures of Domestication. *Nat. Commun.* 11, 1–13. doi:10.1038/s41467-020-18550-1

Fischer, M. C., Rellstab, C., Leuzinger, M., Roumet, M., Gugerli, F., Shimizu, K. K., et al. (2017). Estimating Genomic Diversity and Population Differentiation - an Empirical Comparison of Microsatellite and SNP Variation in Arabidopsis Halleri. *BMC Genomics* 18, 1–15. doi:10.1186/s12864-016-3459-7

Gautier, M., Klassmann, A., and Vitalis, R. (2017). rehh2.0: A Reimplementation of the R Packagerehhto Detect Positive Selection from Haplotype Structure. *Mol. Ecol. Resour.* 17, 78–90. doi:10.1111/1755-0998.12634

Guan, D., Landi, V., Luigi-Sierra, M. G., Delgado, J. V., Such, X., Castelló, A., et al. (2020). Analyzing the Genomic and Transcriptomic Architecture of Milk Traits in Murciano-Granadina Goats. *J. Anim. Sci Biotechnol* 11, 1–19. doi:10.1186/s40104-020-00435-4

Jiang, Z., Michal, J. J., Wu, X.-L., Pan, Z., and MacNeil, M. D. (2011). The Heparan and Heparin Metabolism Pathway Is Involved in Regulation of Fatty Acid Composition. *Int. J. Biol. Sci.* 7, 659–663. doi:10.7150/ijbs.7.659

Kim, K. S., Kim, S. W., Raney, N. E., and Ernst, C. W. (2012). Evaluation of BTA1 and BTA5 QTL Regions for Growth and Carcass Traits in American and Korean Cattle. *Asian Australas. J. Anim. Sci.* 25, 1521–1528. doi:10.5713/ajas.2012.12218

Kumar, M., Dahiya, S. P., Ratwan, P., Kumar, S., and Chitra, A. (2019). Status, Constraints and Future Prospects of Murrah Buffaloes in India. *Indian J. Anim. Sci.* 89, 1291–1302.

Kumar, S., Gupta, J., Kumar, N., Dikshit, K., Navani, N., Jain, P., et al. (2006). Genetic Variation and Relationships Among Eight Indian Riverine Buffalo Breeds. *Mol. Ecol.* 15, 593–600. doi:10.1111/j.1365-294X.2006.02837.x

Kunej, T., Skok, D. J., Zorc, M., Ogrinc, A., Michal, J. J., Kovac, M., et al. (2013). Obesity Gene Atlas in Mammals. *J. Genomics* 1, 45–55. doi:10.7150/jgen.3996

Li, C., Sun, D., Zhang, S., Wang, S., Wu, X., Zhang, Q., et al. (2014). Genome Wide Association Study Identifies 20 Novel Promising Genes Associated with Milk Fatty Acid Traits in Chinese Holstein. *PLoS One* 9, e96186. doi:10.1371/journal.pone.0096186

Li, H. (2013). Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. arXiv preprint arXiv:1303.3997

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Low, W. Y., Tearle, R., Bickhart, D. M., Rosen, B. D., Kingan, S. B., Swale, T., et al. (2019). Chromosome-Level Assembly of the Water Buffalo Genome Surpasses Human and Goat Genomes in Sequence Contiguity. *Nat. Commun.* 10, 1–11. doi:10.1038/s41467-018-08260-0

Martin, M. (2011). Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet j.* 17, 10–12. doi:10.14806/ej.17.1.200

Medeiros de Oliveira Silva, R., Bonvino Stafuzza, N., de Oliveira Fragomeni, B., Miguel Ferreira de Camargo, G., Matos Ceacero, T., Noely dos Santos Gonçalves Cyrillo, J., et al. (2017). Genome-wide Association Study for Carcass Traits in an Experimental Nelore Cattle Population. *PLoS One* 12, e0169860. doi:10.1371/journal.pone.0169860

Mishra, D. C., Sikka, P., Yadav, S., Bhati, J., Paul, S. S., Jerome, A., et al. (2020). Identification and Characterization of Trait-Specific SNPs Using ddRAD Sequencing in Water Buffalo. *Genomics* 112, 3571–3578. doi:10.1016/j.ygeno.2020.04.012

NBAGR Karnal (2021). Breed Profiles. Available at: https://nbagr.icar.gov.in/en/registered-buffalo/(Accessed July 04, 2021).

Patel, S., Thakkar, J., Koringa, P., Jakhesara, S., Patel, A., De, S., et al. (2017). Evolution and Diversity Studies of Innate Immune Genes in Indian Buffalo (Bubalus Bubalis) Breeds Using Next Generation Sequencing. *Genes Genom* 39, 1237–1247. doi:10.1007/s13258-017-0585-9

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS One* 7, e37135. doi:10.1371/journal.pone.0037135

Pickrell, J., and Pritchard, J. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genet.* 8 (11), e1002967. doi:10.1371/journal.pgen.1002967

Pollinger, J. P., Bustamante, C. D., Fledel-Alon, A., Schmutz, S., Gray, M. M., and Wayne, R. K. (2005). Selective Sweep Mapping of Genes with Large Phenotypic Effects. *Genome Res.* 15, 1809–1819. doi:10.1101/gr.4374505

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* 81, 559–575. doi:10.1086/519795

Purfield, D. C., Evans, R. D., Carthy, T. R., and Berry, D. P. (2019). Genomic Regions Associated with Gestation Length Detected Using Whole-Genome Sequence Data Differ between Dairy and Beef Cattle. *Front. Genet.* 10, 1068. doi:10.3389/fgene.2019.01068

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* 26, 841–842. doi:10.1093/bioinformatics/btq033

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ravi Kumar, D., Joel Devadasan, M., Surya, T., Vineeth, M. R., Choudhary, A., Sivalingam, J., et al. (2020). Genomic Diversity and Selection Sweeps Identified in Indian Swamp Buffaloes Reveals It's Uniqueness with Riverine Buffaloes. *Genomics* 112, 2385–2392. doi:10.1016/j.ygeno.2020.01.010

Rochette, N. C., Rivera-Colón, A. G., and Catchen, J. M. (2019). Stacks 2: Analytical Methods for Paired-end Sequencing Improve RADseq-based Population Genomics. *Mol. Ecol.* 28, 4737–4754. doi:10.1111/mec.15253

Sabeti, P. C., Varilly, P., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., et al. (2007). Genome-Wide Detection and Characterization of Positive Selection in Human Populations. *Nature* 449, 913–918. doi:10.1038/nature06250

Sambrook, J., and Russell, D. W. (2006). Purification of Nucleic Acids by Extraction with Phenol: Chloroform. *Cold Spring Harbor Protoc.* 2006, pdb-prot4455. doi:10.1101/pdb.prot4455

Sathwara, R. N., Gupta, J. P., Chaudhari, J. D., Parmar, G. A., Prajapati, B. M., Srivastava, A. K., et al. (2020). Analysis of Association between Various Fertility Indicators and Production Traits in Mehsana Buffaloes. *Trop. Anim. Health Prod.* 52, 2585–2592. doi:10.1007/s11250-020-02288-5

Seo, M., Kim, K., Yoon, J., Jeong, J. Y., Lee, H.-J., Cho, S., et al. (2016). RNA-seq Analysis for Detecting Quantitative Trait-Associated Genes. *Sci. Rep.* 6, 24375. doi:10.1038/srep24375

Serão, N. V., González-Peña, D., Beever, J. E., Faulkner, D. B., Southey, B. R., and Rodriguez-Zas, S. L. (2013). Single Nucleotide Polymorphisms and Haplotypes Associated with Feed Efficiency in Beef Cattle. *BMC Genet.* 14, 1–20. doi:10.1186/1471-2156-14-94

Silva-Vignato, B., Coutinho, L. L., Poleti, M. D., Cesar, A. S. M., Moncau, C. T., Regitano, L. C. A., et al. (2019). Gene Co-Expression Networks Associated with Carcass Traits Reveal New Pathways for Muscle and Fat Deposition in Nelore Cattle. *BMC Genomics* 20, 32. doi:10.1186/s12864-018-5345-y

Singh, A., Mehrotra, A., Gondro, C., Romero, A. R. d. S., Pandey, A. K., Karthikeyan, A., et al. (2020). Signatures of Selection in Composite Vrindavani Cattle of India. *Front. Genet.* 11, 589496. doi:10.3389/fgene.2020.589496

Surya, T., Vineeth, M. R., Sivalingam, J., Tantia, M. S., Dixit, S. P., Niranjan, S. K., et al. (2019). Genomewide Identification and Annotation of SNPs in Bubalus Bubalis. *Genomics* 111, 1695–1698. doi:10.1016/j.ygeno.2018.11.021

Vineeth, M. R., Surya, T., Sivalingam, J., Kumar, A., Niranjan, S. K., Dixit, S. P., et al. (2019). Genome-Wide Discovery of SNPs in Candidate Genes Related to Production and Fertility Traits in Sahiwal Cattle. *Trop. Anim. Health Pro* 52 (4), 1707–1715. doi:10.1007/s11250-019-02180-x

Weir, B. S., and Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 38, 1358–1370. doi:10.2307/2408641

Wright, B., Farquharson, K. A., McLennan, E. A., Belov, K., Hogg, C. J., and Grueber, C. E. (2019). From Reference Genomes to Population Genomics: Comparing Three Reference-Aligned Reduced-Representation Sequencing Pipelines in Two Wildlife Species. *BMC Genomics* 20, 1–10. doi:10.1186/s12864-019-5806-y

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a Tool for Genome-Wide Complex Trait Analysis. *Am. J. Hum. Genet.* 88, 76–82. doi:10.1016/j.ajhg.2010.11.011

Yang, S.-H., Bi, X.-J., Xie, Y., Li, C., Zhang, S.-L., Zhang, Q., et al. (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. doi:10.3390/ijms161125976

Zhang, S.-Z., Xu, Y., Xie, H.-Q., Li, X.-Q., Wei, Y.-Q., and Yang, Z.-M. (2009). The Possible Role of Myosin Light Chain in Myoblast Proliferation. *Biol. Res.* 42, 121–132. doi:10.4067/S0716-97602009000100013

Zhang, Y., Colli, L., and Barker, J. S. F. (2020). Asian Water buffalo: Domestication, History and Genetics. *Anim. Genet.* 51, 177–191. doi:10.1111/age.12911

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 October 2021
doi: 10.3389/fgene.2021.682576

Accuracy of Genomic Prediction for Milk Production Traits in Philippine Dairy Buffaloes

Jesus Rommel V. Herrera[1,2]*, *Ester B. Flores*[3], *Naomi Duijvesteijn*[1]*, *Nasir Moghaddar*[1] and *Julius H. van der Werf*[1]

[1]School of Environmental and Rural Science, University of New England, Armidale, NSW, Australia, [2]Philippine Carabao Center-University of the Philippines Los Banos, Laguna, Philippines, [3]Philippine Carabao Center National Headquarters, Muñoz, Philippines

OPEN ACCESS

Edited by:
Yang Zhou,
Huazhong Agricultural University,
China

Reviewed by:
Lubos Vostry,
Czech University of Life Sciences
Prague, Czechia
Grum Gebreyesus Teklewold,
Aarhus University, Denmark

***Correspondence:**
Jesus Rommel V. Herrera
jrvh4171@yahoo.com
Naomi Duijvesteijn
nduijves@une.edu.au

Specialty section:
This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 18 March 2021
Accepted: 28 September 2021
Published: 28 October 2021

Citation:
Herrera JRV, Flores EB, Duijvesteijn N, Moghaddar N and van der Werf JH (2021) Accuracy of Genomic Prediction for Milk Production Traits in Philippine Dairy Buffaloes. Front. Genet. 12:682576. doi: 10.3389/fgene.2021.682576

The objective of this study was to compare the accuracies of genomic prediction for milk yield, fat yield, and protein yield from Philippine dairy buffaloes using genomic best linear unbiased prediction (GBLUP) and single-step GBLUP (ssGBLUP) with the accuracies based on pedigree BLUP (pBLUP). To also assess the bias of the prediction, the regression coefficient (slope) of the adjusted phenotypes on the predicted breeding values (BVs) was also calculated. Two data sets were analyzed. The GENO data consisting of all female buffaloes that have both phenotypes and genotypes (n = 904 with 1,773,305-days lactation records) were analyzed using pBLUP and GBLUP. The ALL data, consisting of the GENO data plus females with phenotypes but not genotyped (n = 1,975 with 3,821,305-days lactation records), were analyzed using pBLUP and ssGBLUP. Animals were genotyped with the Affymetrix 90k buffalo genotyping array. After quality control, 60,827 single-nucleotide polymorphisms were used for downward analysis. A pedigree file containing 2,642 animals was used for pBLUP and ssGBLUP. Accuracy of prediction was calculated as the correlation between the predicted BVs of the test set and adjusted phenotypes, which were corrected for fixed effects, divided by the square root of the heritability of the trait, corrected for the number of lactations used in the test set. To assess the bias of the prediction, the regression coefficient (slope) of the adjusted phenotypes on the predicted BVs was also calculated. Results showed that genomic methods (GBLUP and ssGBLUP) provide more accurate predictions compared to pBLUP. Average GBLUP and ssGBLUP accuracies were 0.24 and 0.29, respectively, whereas average pBLUP accuracies (for GENO and ALL data) were 0.21 and 0.22, respectively. Slopes of the two genomic methods were also closer to one, indicating lesser bias, compared to pBLUP. Average GBLUP and ssGBLUP slopes were 0.89 and 0.84, respectively, whereas the average pBLUP (for GENO and ALL data) slopes were 0.80 and 0.54, respectively.

Keywords: dairy buffalo, ssGBLUP, bias, accuracy of genomic prediction, pBLUP, GBLUP

INTRODUCTION

The Philippine Carabao Center (PCC) has put in place a genetic improvement program that includes a system of evaluating genetically superior individual animals for milk and milk component traits and maintenance of nucleus herds of dairy buffaloes as source of breeding animals and provision of frozen semen from the best riverine buffalo germplasm (identified through progeny testing) for artificial insemination (AI). PCC maintains 12 institutional herds of dairy buffaloes [mostly Bulgarian Murrahs (BUL)] dispersed throughout the archipelago as source of breeding animals and frozen semen from the best riverine buffalo germplasm for AI to riverine, crossbred, and swamp buffaloes. Recording and evaluation of performance are presently limited to animals in these herds, numbering ~1,200 females, of which ~400 can be considered as elite dams (open-nucleus scheme). However, present constraints of the breeding program are as follows: the number of recorded cows is not expected to increase substantially in the immediate future; currently progeny is testing only eight bulls per year; accuracies of progeny test bulls are low due to small number of daughters with lactation records; and generation interval is long for AI sires, ~8 years (Flores, 2014).

The availability of the Affymetrix 90K Buffalo Genotyping Array (Affymetrix, Inc., Santa Clara, CA) in 2013 made it possible to do genomic studies in the bubaline species (Iamartino et al., 2017). When the trait of interest cannot be recorded on the selection candidate, genomic selection schemes are very attractive even when the number of phenotypic records is limited, because traditional breeding requires progeny testing schemes with long generation intervals (Schaeffer, 2006). Having similarities with dairy cattle breeding, for example, long generation interval, traits that are sex-limited, and measured late in life, it is probable that the advantages of genomic selection seen in dairy cattle will also be observed in dairy buffalo.

Genomic prediction studies in dairy buffaloes are very limited and were based on small data sets. Tonhati et al. (2016) used single-step genomic best linear unbiased prediction (ssGBLUP) to estimate the predicted transmitting ability accuracies for seven milk traits on 452 Brazilian buffaloes. Using a fivefold cross-validation, Liu et al. (2017) evaluated the reliability of genomic estimated BVs and their correlation with EBVs for six milk production traits from 412 Italian Mediterranean (ITA) buffaloes.

The objective of this study was to determine the accuracy of genomic prediction and bias for milk yield (MY), fat yield (FY), and protein yield (PY) from Philippine dairy buffaloes using GBLUP and ssGBLUP compared to prediction accuracy and bias based on pedigree BLUP (pBLUP).

MATERIALS AND METHODS

Phenotype data and blood samples used in this study were obtained from the PCC. All animals are housed in institutional farms and cooperatives managed by PCC. Data collection and storage are managed by the center's Animal Breeding and Genomics Section (ABGS).

Phenotype Data

Traits investigated in this study are 305-days MY, FY, and PY. Descriptive statistics of the phenotypic data are presented in **Tables 1** and **2**. The numbers of animals with one, two, and three lactation records are shown in **Tables 3** and **4**.

Two data sets were analyzed. One contains only female buffaloes that have both phenotypes and genotypes (hereby referred to as GENO) (**Table 1**). Analyses done on these data were pBLUP and GBLUP. The other data set (hereby referred to as ALL) (**Table 2**) contains all the above animals, plus females with phenotypes but are not genotyped. Analyses done on these data were pBLUP and ssGBLUP. A pedigree file containing 2,642 animals spanning six generations was used for pBLUP and ssGBLUP.

Genotype Data

Genomic DNA was extracted using the Promega ReliaPrep Blood gDNA Miniprep System according to the manufacturer's protocol. DNA quantification was done using the Promega Quantus Fluorometer. Samples were first subjected to RNA purification prior to shipment to Affymetrix, Inc. Submitted samples were genotyped using the Axiom 90k Buffalo Genotyping Array. Generated ".CEL" files were analyzed using the Axiom Analysis Suite using default settings, wherein polymorphic markers were identified. Additional quality control measures applied include a single-nucleotide polymorphism (SNP) removed if its minor allele frequency is less than 0.05, is out of Hardy-Weinberg equilibrium ($p < 1 \times 10^{-15}$), has no genome location, and is not found in the autosomes. After applying the quality control measures, only 60,827 SNPs in 29 autosomes were used for the determination of accuracy of genomic prediction and bias.

Statistical Methods

BVs were estimated using three methods: pBLUP, GBLUP, and ssBLUP. The three methods used the following model:

$$305DTrait_{ijkp} = \mu + breed_i + lactation\,number_j + HYS_k + animal_p + permanent\,env_p + e_{ijkp}$$

where *305dTrait* is a 305-days record for the desired trait (MY, FY, PY); μ is the general mean; *breed* is the fixed breed effect; *lactation number* is the fixed effect for lactation number; *HYS* is the fixed effect for herd-year-season; and *animal* and *permanent env* are the individual effect and permanent environmental effect on animal p; and e is random residual with $e \sim N(0,e^2)$.

The difference among the three methods is the type of relationships that was used. pBLUP uses a numerator relationship matrix (also known as an A-matrix) based on the pedigree (family relationships). The creation of the genomic relationship matrix (GRM), also known as the G-matrix, was used in GBLUP, and ssGBLUP is based on VanRaden (2008). The

TABLE 1 | Descriptive statistics of GENO data to be used for pBLUP and GBLUP analyses.

Trait	No. of animals	No. of records	No. genotyped	Mean (kg)	Min (kg)	Max (kg)	SD (kg)
MY	904	1,773	904	1,573.2	103.1	3,054.5	505.9
FY	856	1,384	856	119.0	30.2	206.9	27.7
PY	856	1,384	856	70.7	22.5	127.9	16.0

MY, milk yield; FY, fat yield; PY, protein yield.

TABLE 2 | Descriptive statistics of ALL data to be used for pBLUP and ssGBLUP analyses.

Trait	No. of animals	No. of records	No. genotyped	Mean (kg)	Min (kg)	Max (kg)	SD (kg)
MY	1,975	3,821	904	1,466.3	103.1	3,150.9	518.0
FY	1,918	3,405	856	111.9	29.3	210.1	29.1
PY	1,918	3,405	856	66.3	19.9	128.8	17.3

MY, milk yield; FY, fat yield; PY, protein yield.

TABLE 3 | Number of animals (number of records) for test and training sets for MY.

Test set	Training set	
	GENO	ALL
329[a] (329)	575 (1,444)	1,646 (3,492)
281[b] (562)	623 (1,211)	1,694 (3,259)
294[c] (882)	610 (891)	1,681 (2,939)

[a,b,c]Number of animals with 1, 2, and 3 lactation records, respectively.

TABLE 4 | Number of animals (number of records) for test and training sets for FY and PY.

Test set	Training set	
	GENO	ALL
441[a] (441)	415 (943)	1,477 (2,964)
302[b] (604)	554 (780)	1,616 (2,801)
113[c] (339)	743 (1,045)	1,805 (3,066)

[a,b,c]Number of animals with 1, 2, and 3 lactation records, respectively.

ssGLUP (Misztal et al., 2009; Legarra et al., 2014) uses an H-matrix (combination of family and genomic relationships), where the G-matrix replaces the A_{22} matrix (A-matrix containing only females that were genotyped).

Validation Scheme

A threefold cross-validation scheme was used to compare accuracy of prediction and bias using GBLUP and ssGBLUP with those of pBLUP. Animals were assigned to one of three test sets: one lactation record, two lactation records, and three lactation records (**Tables 3** and **4**). One lactation record could mean that the animal has a record for the first lactation, second lactation, or third lactation. An animal with two lactation records could mean that it has the first two lactations, the first and the third lactations, or the second and third lactations. In each case, the training set is composed of animals in the data set that are not part of the test set. Phenotypes of animals in the test sets were masked, and BVs were then estimated for each set either by pBLUP and GBLUP for the GENO data or pBLUP and ssGBLUP for ALL data using ASReml 4.1 (Gilmour et al., 2015).

Accuracy of Genomic Prediction

Accuracy of prediction was calculated as the correlation between the predicted BVs of the test set and its corresponding adjusted phenotypes, which were corrected for fixed effects, divided by the square root of the heritability of the trait, corrected for the number of lactations used in the test set:

$$r = \frac{\mathrm{corr}\left(\mathrm{BV}, \mathrm{adj.pheno}\right)}{\sqrt{\frac{h^2}{\mathrm{rep}+\left(1-\frac{\mathrm{rep}}{n}\right)}}}$$

where r is the accuracy of prediction; corr is the correlation; BV is the predicted BV; adj. pheno is the adjusted phenotype corrected for fixed effects; h^2 is the heritability of the trait; rep is the repeatability of test set; and n is the number of lactations records used in test set. Note that if $n = 1$, denominator is equal to h.

The average of the accuracies of the three test sets is the accuracy of prediction of a trait.

Prediction Bias

To assess the bias of prediction, the regression coefficient (slope) of the adjusted phenotypes on the predicted BVs was also calculated, with slopes of approximately 1 showing zero bias. Slopes greater than or less than 1 indicate underestimation and overestimation, respectively, of BVs. The average of the slopes of the three test sets is the slope of a trait.

RESULTS

Accuracy of Genomic Prediction

Accuracies of genomic prediction of the three traits through cross-validation are shown in **Table 5**. Heritabilities used are 0.19, 0.17, and 0.19 for MY, FY and PY, respectively, which were derived using pBLUP. Results showed that genomic methods (GBLUP and ssGBLUP) provide more accurate predictions compared to pBLUP. For the GENO data, GLUP accuracies

TABLE 5 | Accuracy of prediction for pBLUP, GBLUP, and ssGBLUP estimated from threefold cross-validation scheme.

Trait	GENO			ALL		
	pBLUP	GBLUP	Increase in accuracy	pBLUP	ssGBLUP	Increase in accuracy
MY	0.20 ± 0.04	0.28 ± 0.06	0.08	0.17 ± 0.02	0.30 ± 0.04	0.13
FY	0.23 ± 0.04	0.24 ± 0.05	0.01	0.26 ± 0.14	0.30 ± 0.01	0.04
PY	0.20 ± 0.05	0.20 ± 0.05	0	0.23 ± 0.14	0.26 ± 0.02	0.03
Average	0.21	0.24	0.03	0.22	0.29	0.07

MY, milk yield; FY, fat yield; PY, protein yield.

TABLE 6 | Estimated slopes calculated from breeding values from pBLUP, GBLUP, and ssGBLUP.

Trait	GENO		ALL	
	pBLUP	GBLUP	pBLUP	ssGBLUP
MY	0.69 ± 0.39	0.85 ± 0.28	0.42 ± 0.07	0.85 ± 0.16
FY	0.94 ± 0.17	0.99 ± 0.22	0.62 ± 0.36	0.88 ± 0.04
PY	0.76 ± 0.11	0.83 ± 0.34	0.57 ± 0.38	0.79 ± 0.10
Average	0.80	0.89	0.54	0.84

MY, milk yield; FY, fat yield; PY, protein yield.

increased for MY and FY by 0.08 and 0.01, respectively, whereas there was no increase for PY if compared to pBLUP accuracies. In the case of ALL data, ssGBLUP accuracies are higher by 0.13, 0.04, and 0.07 for MY, FY, and PY, respectively, if compared to pBLUP accuracies. Average pBLUP (for GENO and ALL data) accuracies for the three traits were 0.21 and 0.22, respectively, whereas the average GBLUP and ssGBLUP (for GENO and ALL data) accuracies were 0.24 and 0.29, respectively. GBLUP and ssGBLUP accuracies were, on average, 0.03 and 0.07 higher, respectively, compared to pBLUP accuracies.

Prediction Bias

In the case of bias of prediction, slopes for all methods were less than 1, indicating overestimation of BVs (**Table 6**). However, slopes of the two genomic methods are closer to 1, indicating lesser bias, compared to pBLUP slopes. Average pBLUP (for GENO and ALL data) slopes for the three traits were 0.80 and 0.54, respectively, whereas GBLUP and ssGBLUP slopes were 0.89 and 0.84, respectively.

DISCUSSION

With a limited number of progeny-tested bulls, a reference population of females with at most three lactations per animal was used in this study to determine the accuracy of genomic prediction and bias for MY, FY, and PY using GBLUP and ssGBLUP and compared to prediction accuracy and bias based on pedigree pBLUP. The accuracy of prediction was based on threefold cross-validation scheme (test sets are the number of lactations per animal), and bias was calculated as the regression coefficient (slope) of the adjusted phenotypes on the predicted BVs.

Several genomic prediction studies in dairy cattle have been done wherein the reference populations are cows. Brown et al. (2016) used crossbred cows from Kenya as no bulls were available that can be ranked because there is very little phenotypic and pedigree data available. In the case of Nayee et al. (2018), Holstein crossbred cows in India were used as the reference population because the annual numbers of progeny tested bulls are limited to 20 to 40 per year. With limited number of progeny-tested bulls with highly reliable EBV (reliability >0.8), Ding et al. (2013) established a reference population of Chinese Holstein females. In the case of dairy buffalo, two genomic prediction studies (Tonhati et al., 2016; Liu et al., 2017) were done based on small data sets of genotyped female buffaloes as the reference population.

Combining different breeds is another option to increase the reference population (Hayes et al., 2009; Cole and Silva, 2016). In this study, three breeds were included BUL, Brazilian Murrah (BRA), and American Murrah (AME). Based on their breed histories, these three breeds all have the riverine buffalo blood from India as ancestors. The BUL was created by crossing the Indian Murrah imported into Bulgaria in 1962 and 1975 with the native Bulgarian Mediterranean buffaloes (Alexiev, 1998; Borghese, 2013). Buffaloes imported by PCC from Brazil in 2013 were all Indian Murrah and their crosses. The AME came from one buffalo herd from Florida; the most probable source of the foundation stock came from the University of Florida, wherein in 1979, 14 cows and 2 bulls of the Bufalypso breed from Trinidad were delivered, which were created during 1949–1960 from 7 imported Indian buffalo breeds [(Alexiev, 1998). A principal component analysis (PCA) (**Figure 1**] was done in a previous study wherein these three breeds were grouped together. PCC also has an ITA buffalo population but was not included in this study as it formed a separate group in the PCA plot (**Figure 1**). Included also in the reference population are crosses of BUL bulls with BRA (BUL × BRA) and AME (BUL × BRA) females. Moreover, all the institutional herds, dispersed throughout the archipelago, are linked using BUL sires.

The increase in accuracy in GBLUP could be due to the realized relationships of animals in GBLUP compared to just expected relationships of animals in pBLUP. For example, full sibs would have an expected relationship of 0.5 in pBLUP, but this could be 0.3 to 0.6 in GBLUP. The increase in accuracy in ssGLUP could also be due to the above plus the linking of unrelated families, which is not possible with pBLUP. As an example, two families in pBLUP are not related because they do not share a common ancestor. In ssGBLUP, genotyping only one animal in each family would serve as a link between these two families; this relationship between these two

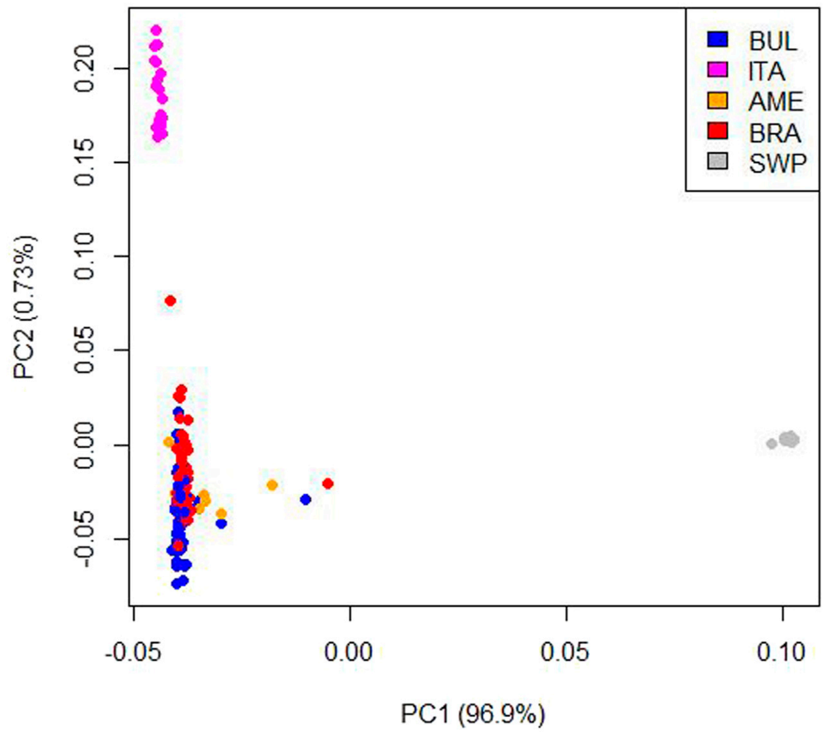

FIGURE 1 | PCA plot generated based on the genomic relationship matrix of the five buffalo populations (*n* = 250). BUL, Bulgarian Murrah; BRA, Brazilian Murrah; ITA, Italian Mediterranean; AME, American Murrah; SWP, Philippine swamp.

genotyped animals will now create relationships among all animals in both families.

The accuracy of prediction for MY in this study using GBLUP and ssGBLUP was 0.28 and 0.30, respectively. These were lower than reported studies using dairy cows as the reference population. Brown et al. (2016) had an accuracy of prediction of 0.32–041 for MY using GBLUP with a reference population of 1,013 crossbred Kenyan cows. The creation of the GRM (G-matrix) here made it possible to estimate the genetic relationships among the animals, all of which do not have pedigree information. The accuracy of prediction of (Nayee et al., 2018) using ssGBLUP for MY was 0.387–0.405 with a larger reference population of 10,797 Holstein crossbred cows. In the case of Ding et al. (2018), accuracies of prediction for MY, FY, and PY were 0.37, 0.32, and 0.40, respectively, using 3,087 Chinese Holstein cows. In the case of dairy buffaloes, accuracies of prediction in Liu et al. (2017) are similar for MY (0.28), but higher for FY (0.35 vs. 0.24) and PY (0.24 vs. 0.20). The study by Liu et al. reported reliabilities, whereas accuracy is the square root of reliability.

A limitation of this study is the small data set. Female animals with production and genotype data will be added yearly to

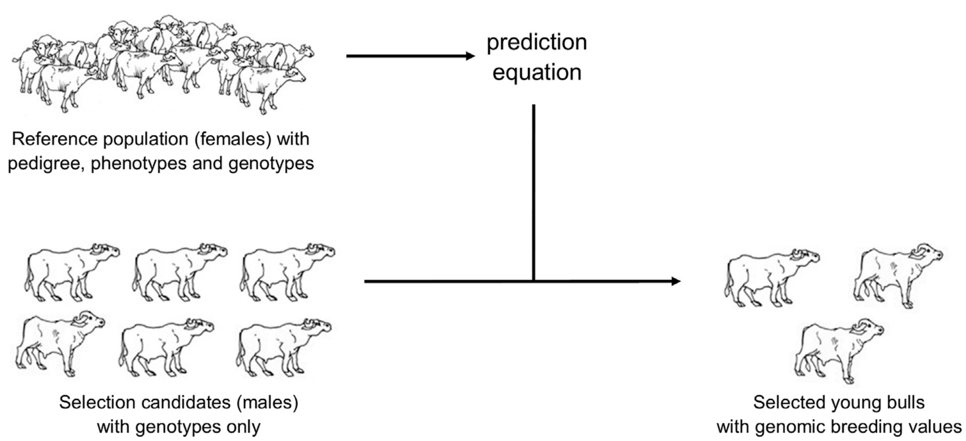

FIGURE 2 | Genomic selection in Philippine dairy buffaloes.

increase the reference population. Potential semen donor bulls will be genotyped to determine their BVs using the population of cows as the reference population.

Implications

At present, the generation interval of AI buffalo sires is ~8 years. With GS, young genotyped candidate bulls can be given BVs using females in the institutional herds as the reference population (**Figure 2**). ssGBLUP method can be used to generate BVs as some females with performance data cannot be genotyped anymore (ie, dead). Moreover, limited funds allocated per year may not allow genotyping of all cows with at least one lactation record. Selected candidate bulls coming from the institutional herds (and cooperatives) that will be genotyped are closely related to the reference population as their female relatives (dams, granddams, siblings) are in that population. Young bulls can now be selected at a younger age; generation interval can be lowered to ~3.5 years old. A future study will be done to compare the present progeny testing breeding scheme and a genomic breeding scheme, that is, GBLUP in terms of genetic gain and cost savings from the point of view of PCC as the breeding entity.

CONCLUSIONS

This study determined the accuracy of genomic prediction and bias for MY, FY, and PY in Philippine dairy buffaloes wherein the reference population is composed solely of cows. GBLUP and ssGBLUP accuracies were, on average, 0.03 and 0.07 higher, respectively, compared to pBLUP accuracies. Moreover, prediction bias of the two genomic methods is lesser (closer to 1) compared to pBLUP. With the higher accuracy of prediction and lesser bias, it is suggested that PCC adopts the genomic method, that is, GLUP or ssGBLUP, in its genetic evaluation.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The datasets for this article are not publicly available because these are the exclusive property of the Philippine Carabao Center. Requests to access these datasets should be directed to Ronnie D. Domingo, OIC-Executive Director, pcc-oed@mozcom.com.

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because the implementation of this study was monitored and supervised by the Livestock Research Division of DOST-PCAARRD to ensure acceptable guidelines and regulations were followed. All animals used in the study are directly managed by PCC. All data and samples were collected under the supervision of PCC licensed veterinarians.

AUTHOR CONTRIBUTIONS

All authors agreed on the concept of this work. JH performed the analysis and wrote the manuscript. EBF, ND, NM and JW provided feedback and reviewed the manuscript. EBF provided and consolidated phenotype data.

FUNDING

Research was supported by the Philippine Council for Agriculture, Aquatic and Natural Resources Research and Development—Department of Science and Technology (PCAARRD-DOST) and PCC. Scholarship of JRVH was provided by the Philippine Carabao Center- Southeast Asian Regional Center for Graduate Study and Research in Agriculture (PCC-SEARCA).

ACKNOWLEDGMENTS

The authors would like to acknowledge: 1) PCAARRD-DOST for funding the genotyping project, 2) various coordinators, herd managers and directors of various PCC centers for making data and samples available, 3) staff of PCC's ABGS for preparing the blood samples for genotyping. JRVH would like to acknowledge PCC-SEARCA for his PhD scholarship.

REFERENCES

Alexiev, A. (1998). *The Water buffalo*. St: Kilment Ohridski University Press.

Borghese, A. (2013). Buffalo Livestock and Products in Europe. *Buffalo Bull.* 32 (1), 50–74.

Brown, A., Ojango, J., Gibson, J., Okeyo, M., and Mrode, R. (2016). Short Communication: Genomic Selection in a Crossbred Cattle Population Using Data from the Dairy Genetics East Africa Project. *J. Dairy Sci.* 99, 7308–7312. doi:10.3168/jds.2016-11083

Cole, J. B., and Silva, M. V. G. B. D. (2016). Genomic Selection in Multi-Breed Dairy Cattle Populations. *R. Bras. Zootec.* 45 (4), 195–202. doi:10.1590/s1806-92902016000400008

Ding, X., Zhang, Z., Li, X., Liu, X., Wang, S., Wu, X., et al. (2013). Accuracy of Genomic Prediction for Milk Production Traits in the Chinese Holstein Population Using a Reference Population Consisting of Cows. *J. Dairy Sci.* 96, 5315–5323. doi:10.3168/jds.2012-6194

Flores, E. B. (2014). PhD Thesis. University of New England.

Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., and Thompson, R. (2015). *ASReml User Guide Release 4.0*. Hemel Hempstead, UK: VSN International Ltd.

Hayes, B. J., Bowman, P. J., Chamberlain, A. C., Verbyla, K., and Goddard, M. E. (2009). Accuracy of Genomic Breeding Values in Multi-Breed Dairy Cattle Populations. *Genet. Sel. Evol.* 41, 51. doi:10.1186/1297-9686-41-51

Iamartino, D., Nicolazzi, E. L., Van Tassell, C. P., Reecy, J. M., Fritz-Waters, E. R., Koltes, J. E., et al. (2017). Design and Validation of a 90K SNP Genotyping Assay for the Water buffalo (Bubalus Bubalis). *PLoS ONE* 12 (10), e0185220. doi:10.1371/journal.pone.0185220

Legarra, A., Christensen, O. F., Aguilar, I., and Misztal, I. (2014). Single Step, a General Approach for Genomic Selection. *Livestock Sci.* 166, 54–65. doi:10.1016/j.livsci.2014.04.029

Liu, J., Liang, A. X., Campanile, G., Plastow, G., Zhang, C., Wang, Z., et al. (2017). Genome-Wide Association Studies to Identify Quantitative Trait Loci Affecting Milk Production Traits in Water Buffalo. *J. Dairy Sci.* 101, 433–444. doi:10.3168/jds.2017-13246

Misztal, I., Legarra, A., and Aguilar, I. (2009). Computing Procedures for Genetic Evaluation Including Phenotypic, Full Pedigree, and Genomic Information. *J. Dairy Sci.* 92, 4648–4655. doi:10.3168/jds.2009-2064

Nayee, N. K., Su, G., Gaijar, S., Sahana, G., Saha, S., Trivedi, K., Guldbrandtsen, B., and Lund, M. (2018). "Genomic Prediction by Single-step Genomic BLUP Using Cow Reference Population in Holstein Crossbred Cattle in India," in Proceedings of the World Congress on Genetics applied to Livestock Production, 11, 411.

Schaeffer, L. R. (2006). Strategy for Applying Genome-Wide Selection in Dairy Cattle. *J. Anim. Breed. Genet.* 123, 218–223. doi:10.1111/j.1439-0388.2006.00595.x

Tonhati, H., Cardoso, D. F., Jordan, D., and Santos, A. (2016). "Genomic Tools Applied to Dairy Buffaloes," in 11th World Buffalo Congress paper presentation.

VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* 91, 4414–4423. doi:10.3168/jds.2007-0980

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 07 December 2021
doi: 10.3389/fgene.2021.608186

Check for updates

Linkage Disequilibrium and Effective Population Size of Buffalo Populations of Iran, Turkey, Pakistan, and Egypt Using a Medium Density SNP Array

Shirin Rahimmadar[1], Mokhtar Ghaffari[1]*, Mahdi Mokhber[1] and John L. Williams[2,3]

[1]Department of Animal Science, Faculty of Agricultural Science, Urmia University, Urmia, Iran, [2]Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, SA, Australia, [3]Department of Animal Science, Food and Nutrition, Università Cattolica Del Sacro Cuore, Piacenza, Italy

OPEN ACCESS

Edited by:
Guohua Hua,
Huazhong Agricultural University,
China

Reviewed by:
Romdhane Rekaya,
University of Georgia, United States
Rostam Abdollahi,
University of Georgia, United States
Mohammad Bagher Zandi,
University of Zanjan, Iran

***Correspondence:**
Mokhtar Ghaffari
araz.11306@gmail.com

Specialty section:
This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 19 September 2020
Accepted: 03 November 2021
Published: 07 December 2021

Citation:
Rahimmadar S, Ghaffari M,
Mokhber M and Williams JL (2021)
Linkage Disequilibrium and Effective
Population Size of Buffalo Populations
of Iran, Turkey, Pakistan, and Egypt
Using a Medium Density SNP Array.
Front. Genet. 12:608186.
doi: 10.3389/fgene.2021.608186

Linkage disequilibrium (LD) across the genome provides information to identify the genes and variations related to quantitative traits in genome-wide association studies (GWAS) and for the implementation of genomic selection (GS). LD can also be used to evaluate genetic diversity and population structure and reveal genomic regions affected by selection. LD structure and Ne were assessed in a set of 83 water buffaloes, comprising Azeri (AZI), Khuzestani (KHU), and Mazandarani (MAZ) breeds from Iran, Kundi (KUN) and Nili-Ravi (NIL) from Pakistan, Anatolian (ANA) buffalo from Turkey, and buffalo from Egypt (EGY). The values of corrected r^2 (defined as the correlation between two loci) of adjacent SNPs for three pooled Iranian breeds (IRI), ANA, EGY, and two pooled Pakistani breeds (PAK) populations were 0.24, 0.28, 0.27, and 0.22, respectively. The corrected r^2 between SNPs decreased with increasing physical distance from 100 Kb to 1 Mb. The LD values for IRI, ANA, EGY, and PAK populations were 0.16, 0.23, 0.24, and 0.21 for less than 100Kb, respectively, which reduced rapidly to 0.018, 0.042, 0.059, and 0.024, for a distance of 1 Mb. In all the populations, the decay rate was low for distances greater than 2Mb, up to the longest studied distance (15 Mb). The r^2 values for adjacent SNPs in unrelated samples indicated that the Affymetrix Axiom 90 K SNP genomic array was suitable for GWAS and GS in these populations. The persistency of LD phase (PLDP) between populations was assessed, and results showed that PLPD values between the populations were more than 0.9 for distances of less than 100 Kb. The Ne in the recent generations has declined to the extent that breeding plans are urgently required to ensure that these buffalo populations are not at risk of being lost. We found that results are affected by sample size, which could be partially corrected for; however, additional data should be obtained to be confident of the results.

Keywords: water buffalo, linkage disequilibrium, LD phase persistency, NE, linkage disequilibrium, LD phase persistency

INTRODUCTION

Recognizing and protecting the genetic diversity of domestic species is important in the development of breeding strategies (Al-Mamun et al., 2015; Wultsch et al., 2016). Recent progress in the field of genome sequencing has increased the availability of genomic data, which has facilitated the assessment of the genetic diversity and population structure (Vonholdt et al., 2010; Decker et al., 2014) using parameters such as population admixture, linkage disequilibrium (LD), and effective population size (Ne) (Al-Mamun et al., 2015).

The non-random association between alleles at different loci is referred to as LD or gametic phase disequilibrium. Knowledge of the pattern of LD in a population is an important prerequisite for GWAS, exploring population structures, and implementing genomic selection (GS) (Niu et al., 2016). The pattern of LD can be used to estimate the rate of genetic drift, level of inbreeding, and the effects of evolutionary forces such as mutation, selection, and migration (Shin et al., 2013). There have been studies of LD in several livestock species, including cattle (McKay et al., 2007; Karimi et al., 2015; Biegelmeyer et al., 2016; Jemaa et al., 2019), buffalo (Mokhber et al., 2019a; Deng et al., 2019; Lu et al., 2020), pig (Badke et al., 2012; Wang et al., 2013), sheep (Meadows et al., 2008), goat (Brito et al., 2015), chicken (Qanbari et al., 2010; Fu et al., 2015), horse (Corbin et al., 2010), dog (Pfahler & Distl., 2015), and cat (Alhaddad et al., 2013).

Several statistics have been suggested to measure LD (Hill and Weir, 1994; Terwilliger, 1995; Zhao et al., 2005; Gianola et al., 2013). Evaluation of these methods has shown that r^2 is less affected by allelic frequency and sample size than D' (Pritchard & Przeworski, 2001; Sved, 2009; Bohmanova et al., 2010). Even when the level of LD of populations is similar, this may still be the result of different evolutionary histories. In this regard, determining patterns of the persistency of LD phase (PLDP) is useful for genetic studies (Pritchard et al., 2000). A SNP in LD with quantitative trait loci may have one marker allele in phase with the beneficial allele for the trait in one breed, while in another breed, the phase may be different. Therefore, GS based on marker information in one population may not lead to genetic progress in another (De Roos et al., 2008). PLDP represents the amount of LD that is maintained between populations and is dependent on the divergence time of the breeds (Badke et al., 2012; Wang et al., 2013). Higher values of PLDP between populations indicate more ancestral LD in common, such that the genomic information can be more reliably inferred between them (Mokry et al., 2014). PLDP can also be used to evaluate the relationships among populations, with those having a common history showing higher PLDP (Wang et al., 2013).

LD provides information to identify the genes and variations affecting quantitative traits in genome-wide association studies (GWAS) by inferring the distribution of recombination events. LD can also be used to evaluate diversity and population structure and to identify genomic regions affected by selection (Mokry et al., 2014). The pattern of LD can reveal the genetic history and the previous demography of a population and can be used to infer the effective population size (Ne) (Qanbari, 2020). Effective population size, Ne, is considered to be one of the most important parameters in population genetics and reflects the amount of genetic diversity, inbreeding, and genetic drift in the population (Frankham, 2005; Tenesa et al., 2007). A low value of Ne indicates limited genetic diversity in a population and affects the amount of genetic progress that can be made in breeding programs (Hayes et al., 2003). Ne can be determined by assessing the amount of LD at various distances along the genome (Sved, 1971; Hayes et al., 2003). High LD at long recombination distances reflects low Ne in recent generations (Hayes et al., 2003).

Buffaloes were introduced into Egypt from India, Iran, and Iraq during the seventh B.C. (Minervino et al., 2020). The three breeds from Iran are reared in three different geographical areas with completely different climatic conditions. The Azeri breed is mainly reared in the north-west and north of Iran (West Azerbaijan, East Azerbaijan province, Ardebil, and eastern parts of Gilan provinces), which have cold, sub-zero winters with heavy snowfall and hot, dry summers with temperatures reaching 35 C, the Khuzestani breed is found in the southwest (mainly in Khuzestan province), which has very hot and occasionally humid summers, with temperatures routinely exceeding 45°C degrees, while in the winter, it can drop below freezing, and the Mazandarani breed is reared along the coast of the Caspian Sea in the Mazandaran and Golestan provinces, which have a moderate climate with occasional humidity all around the year (Mokhber et al., 2019a). The Anatolian water buffalo is widespread in Northwestern Turkey, especially along the coast of the Black Sea, the middle of Anatolia, and also in Eastern Anatolia (Soysal et al., 2007). The Egyptian buffaloes are spread along the River Nile, in the Delta Region, and at the Fayum Oasis. With more than three million head, buffalo is the most important livestock species for milk production in Egypt. The Nili-Ravi breed is the most important livestock breed in Pakistan with more than 10 million head in Punjab, while the Kundi, with more than five million head, is the second most important breed in Pakistan (Minervino et al., 2020).

The present study describes genetic diversity, LD between adjacent SNPs, the trend of LD with increasing distance, and the patterns of PLDP and Ne using genomic data from buffalo breeds of Turkey, Egypt, Pakistan, and Iran, which are genetically closer together than other water buffaloes across the world (Colli et al., 2018).

MATERIALS AND METHOD

Genotype Determination and Data Edition

The present study used data for 83 water buffaloes, including 14 Azeri (AZI), 11 Khuzestani (KHU), and eight Mazandarani (MAZ) from Iran, 12 Anatolian buffalo (ANA) from Turkey, nine Kundi (KUN), and 14 Nili-Ravi (NIL) from Pakistan, and 15 Egyptian buffalo (EGY) to assess LD structure and calculate Ne (**Table 1**).

The samples were genotyped using the Axiom® Buffalo Genotyping 90 K array (Affymetrix, Santa Clara, CA,

TABLE 1 | Descriptive statistics for the studied buffalo populations.

Row	Population name	Population label	Country	Region	N before QC	Number after QC	SNP number after separating QC	SNP number after mergence
1	Azeri	AZI	Iran	Urmia, West Azerbaijan Province	14	14	66,989	57,455
2	Khuzestani	KHU	Iran	Ahvaz, Khuzestan Province	11	11	66,145	57,455
3	Mazandarani	MAZ	Iran	Miankaleh peninsula, Mazandaran Province	8	8	67,900	57,455
4	Anatolian	ANA	Turkey	Istanbul, Afyonkarahisar (western Anatolia) and Tokat (central Anatolia) Provinces	15	12	66,692	57,455
5	Egyptian	EGY	Egypt	-	16	15	66,145	57,455
6	Kundhi	KUN	Pakistan	-	10	9	69,451	57,455
7	Nili-Ravi	NIR	Pakistan	-	15	14	69,820	57,455
Total					89	83	82,043	57,455

United States) that were mapped to the bovine sequence (UMD3.1 Bos Taurus) (Iamartino et al., 2017). Details on the animals and the genomic data are presented in **Table 1**. The genotype data were edited with Plink software (Purcell et al., 2007), and animals and loci with more than 5% missing genotypes (CR_{IND} and CR_{SNP}), monomorphic genotypes, and genotypes with minor allele frequency (MAF) less than 5% were eliminated. MAF and missing genotypes of individuals and SNPs were filtered separately for each genotypic group. Then, the genomic data of all genetic groups were integrated, and the common genetic markers were identified. Finally, the SNPs that were not in the Hardy-Weinberg equilibrium were excluded, and the missing genotypes were imputed using BEAGLE software (Browning & Browning, 2007).

Assessment of Population Structure

Discriminant analysis principal component (DAPC), principal component analysis (PCA), Weir and Cockerham unbiased fixation index (F_{ST}), and population admixture were used to obtain a general overview of the structure of each population and identify animals falling outside their breed group. DAPC, PCA, and F_{ST} were performed using the adegenet package (Jombart and Ahmed, 2011), GeneABEL software (Price et al., 2006), and R scripts using R software (http://www.rproject.org/), respectively. Additionally, the genetic structure of the populations was evaluated using ADMIXTURE software (Alexander et al., 2009).

LD Analysis

After determining the population structure of each genetic group, the patterns of LD were estimated. The values of LD between adjacent SNP as well as paired bases at distances of 0–15 Mb were obtained in each population and evaluated using the statistics r^2 (Hill and Robertson, 1968) and D′, which were calculated as follows:

$$r^2 = \frac{(D)^2}{(freq\,A * freq\,a * freq\,B * frq\,b)},$$

where

$$D = freq\,AB - freq\,A * freq\,B$$

and

$$D' = \begin{cases} \dfrac{D}{\min(freq\,A * freq\,b,\ freq\,B * freq\,a)}\ if\ D > 0 \\ \dfrac{D}{\min(freq\,A * freq\,B * freq\,a * freq\,b)}\ if\ D < 0 \end{cases},$$

where SNP pairs had alleles A and a at the first locus and B and b at the second locus, *freq A*, *freq a*, *freq B*, and *freq b* denote frequencies of alleles *A, a, B*, and *b*, respectively, and *freq AB* denotes frequency of the haplotype *AB* in the population.

The r^2 statistic represents the correlation between alleles at two loci and is less dependent on allele frequencies in finite population sizes compared with other LD measures (Lewontin, 1988; Abecasis et al., 2001; Mueller, 2004) and is the preferred measure for biallelic markers (Zhao et al., 2007). Therefore, r^2 was used in the Ne, LD decay, and PLDP analyses. The r^2 statistic is biased by sample size, and this bias is higher for a smaller sample size. Correction methods discussed by Hui and Burt (2020), Waples et al. (2016), Villa-Angulo et al. (2009), Weir and Hill (1980), and Sved (1971) were applied to the estimate of r^2 in this study. Due to the small sample size for each population, the information was corrected for the sample number and uncertainty of the gametic phase using the following equation (Weir and Hill, 1980; Corbin et al., 2012), which was implemented in SNeP software (Barbato et al., 2015).

$$r^2_{adj} = r^2 - (\beta n)^{-1},$$

where n is the number of individuals sampled, $\beta = 2$ when the gametic phase is known, and $\beta = 1$ if instead the phase is not known (Weir and Hill, 1980).

To determine LD decay, paired markers that were common to all populations were grouped at distances between 0 and 15 Mb at 100 Kb intervals, and the mean r^2 was calculated for each group. The PLDP between populations was expressed as the correlation between the roots of the r^2 calculated for adjacent markers using the formula provided by Badke et al. (2012).

$$r_{ij} = \frac{\sum_{(i,j)} \left(r_{ij(A)} - \bar{r}_A\right)\left(r_{ij(B)} - \bar{r}_B\right)}{S_A S_B},$$

where r_{ij} is the correlation of phase between $r_{ij(A)}$ in population A and $r_{ij(B)}$ in population B, S_A and S_B are the standard deviation of

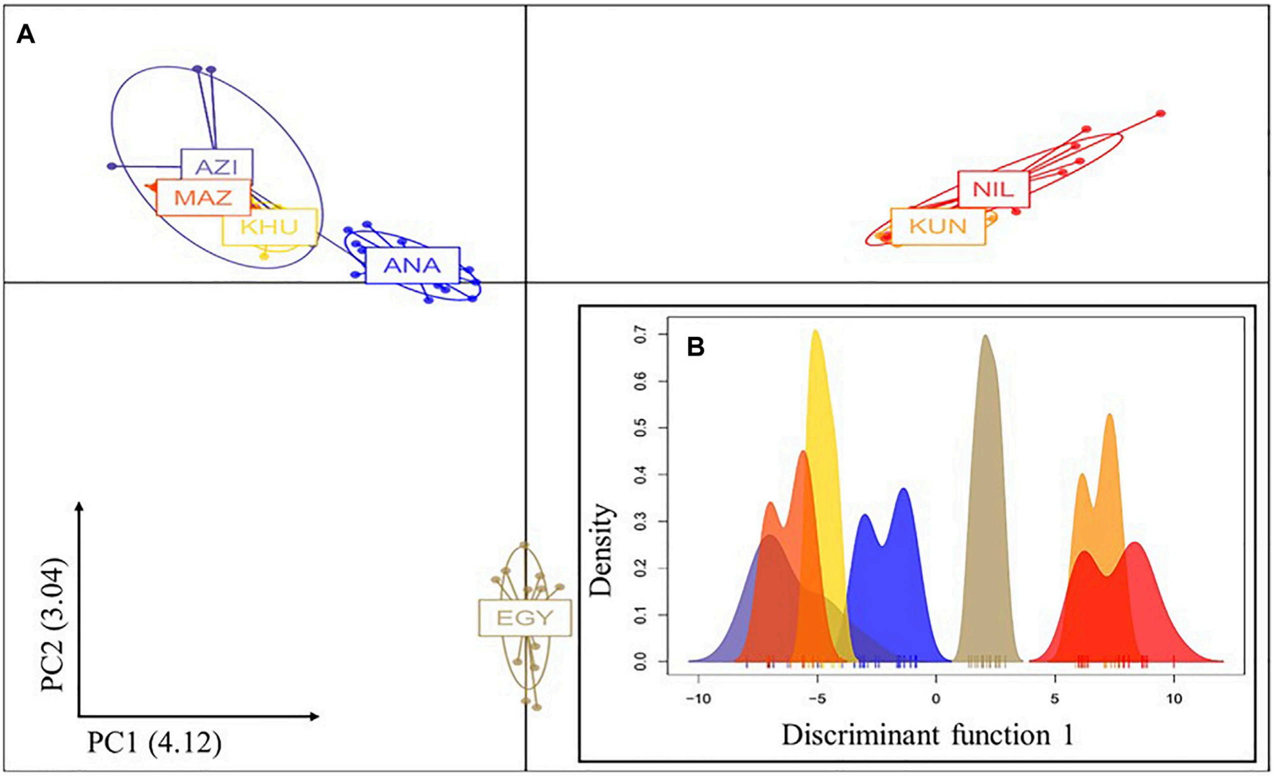

FIGURE 1 | A two first PCs and B first PC only of the DAPC analysis of the water buffalo populations studied. Clusters are indicated by different colors (blue, gray, and green), and populations are identified by their abbreviations AZI, KHU, MAZ, ANA, EGY, KUN, and NIL.

$r_{ij(A)}$ and $r_{ij(B)}$, respectively, and r_A and r_B are the average r_{ij} across all SNP i and j within the common set of markers.

Effective Population Size (Ne)

The corrected LD for each population was used to calculate Ne by applying the formula of $Ne = (\frac{1}{4c})(\frac{1}{r^2} - 1)$ (Sved, 1971), where Ne represents the effective population size of generation T, r^2 indicates the mean of LD for a given distance, and c is the distance between markers in Morgan (1 centimorgan was considered to be approximately equal to one megabase pair, Tenesa et al., 2007; Villa-Angulo et al., 2009). Generation was calculated to determine Ne (T) based genomic distance using the formula of T = 1/2c (Hayes et al., 2003).

RESULTS AND DISCUSSION

Quality of Data

Before frequency and genotyping pruning, there were 89,988 SNPs and 89 individuals. In the first step, six individuals were removed for low genotyping success (MIND >0.05), 637 markers were excluded based on HWE (p≤5.7e-007), and 7,618 SNPs for missing information (GENO >0.05). A total of 83 individuals with 82,043 SNPs passed the first step of QC; the total genotyping rate of these remaining individuals was 0.985. In the second step, MAF was assessed in each population separately, and SNPs with MAF>0.05 were removed (**Table 1**). Then, the populations were merged to create a common dataset of 57,426 SNP markers with MAF higher than 0.05 for each population that passed all the filters. These were used in subsequent analyses in snppLD software (Sargolzaei M, University of Guelph, Canada). These markers covered 2,646.07 Mb of the bovine genome. The mean distance between these markers was 46.07 Kb, and minimum and maximum distances were 42.4 Kb on chromosome BTA 24 and 68.2 Kb on the BTA X, respectively.

Assessment of Population Structure

Understanding of population genetic structure is important to assess population stratification for GWAS, breeding program design, and developing strategies for genetic resources preservation. DAPC, PCA, and admixture analysis results were used to assess population structure. Both PCA and DAPC methods gave similar results. In both methods, genotype data formed three distinct clusters in the first two PCs. The ANA population from Turkey was partially separated from the Iranian cluster, which includes AZI, KHU, and MAZ (**Figure 1** and **Supplementary Figure S1**). The first two PCs in the DAPC accounted for 7.16% of the total variance, 4.12% in the first, and 3.04% in the second dimension (**Figure 1**). The first 10 PCs of DAPC only accounted for about 24% of the total variance (**Supplementary Figure S2**). In the PCA analysis, the first and second PCs explained 4% and 2% of the variance, respectively (**Supplementary Figure S1**). The ANA along with AZI, KHU,

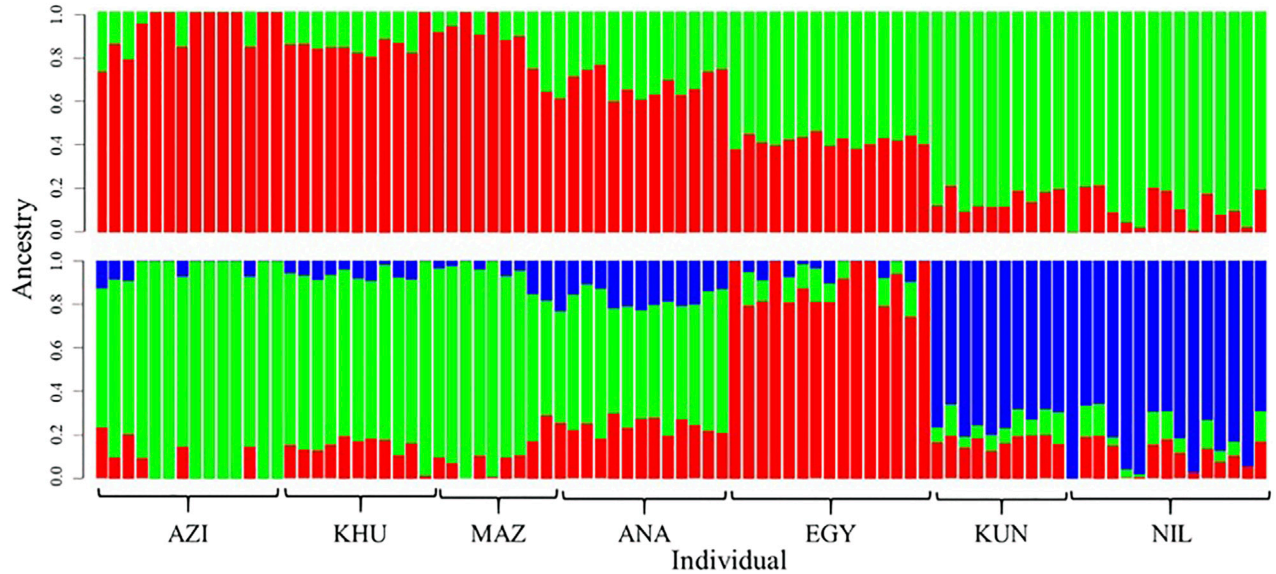

FIGURE 2 | Genetic composition of buffalo breeds revealed with ADMIXTURE software at K = 2 **(top)** and K = 3 **(bottom)**. Individuals are represented with vertical colored bars. Genomic components are assigned different colors.

and MAZ formed overlapping groups with the AZI buffalo being interspersed among the KHU, MAZ, and ANA populations (**Figure 1A,B**). The EGY and populations from Pakistan (KUN and NIL) formed two additional distinct clusters (**Figure 1**). The geographic proximity of Iranian populations with the ANA in Turkey makes gene flow between these two populations likely, which would reduce the differentiation between them. In the analysis of Colli et al. (2018), the populations assessed in the present study belonged to one cluster, which is because these populations are genetically similar when compared with other more genetically distinct breeds worldwide. The results presented here are consistent with other studies focused on Iranian buffaloes where no differences (Strillacci et al., 2021) or very small genetic differentiation was observed (Rahmaninia et al., 2015; Azizi et al., 2016; Mokhber et al., 2018; Ghoreishifar et al., 2020).

There were small differences in F_{ST} among the studied populations (Supplementary Table 1); in most cases, the difference between pairs of populations was less than 0.05, indicating low genetic differentiation according to Wright's classification. The reason for this is because there was high within, compared with between-population variance. However, the F_{ST} results confirmed the DPCA and PCA analyses by separating the populations into three genetic groups. The mean F_{ST} value across populations was 0.045 and varied from 0.011 for AZI from Iran and ANA from Turkey to 0.077 for MAZ from Iran and KUN from Pakistani. The smallest genetic distance was between the Iranian buffaloes and ANA from Turkey, while the largest distance was between the Iranian buffalo and KUN and NIL from Pakistani.

Population structures were investigated using ADMIXTURE software, assuming K as ancestral populations ranging from one to seven. Based on cross-validation error criteria, K = 2 and three had suitable resolution (**Figure 2**). The first subdivision at K = 2 distinguished between Pakistani (KUN and NIL) and the others populations (AZI, KHU, MAZ, ANA, and EGY) (**Figure 2**). At K = 3, the EGY population becomes genetically distinct, giving three groups that coincide with DAPC and PCA clusters. The ADMIXTURE analysis shows that there are genetic components shared among all the populations explaining the overlap between clusters.

LD Analysis

We calculated both r^2 and D′ for adjacent SNPs in the populations for each chromosome (see S1 **Supplementary Table S1**). Because of the small sample size, uncorrected LD values were similar among breeds within clusters, in particular the Iranian breeds, AZI, KHU, and MAZ and Pakistani breeds, KUN and NIL. Results were also corrected for sample size. The values of corrected r^2 for the pooled Iranian breeds (IRI), ANA, EGY, and PAK populations were 0.24, 0.28, 0.27, and 0.22, respectively (**Table 2**). At the chromosome level, chromosomes 25 of the PAK population and chromosomes X of the ANA had the maximum corrected r^2 values, respectively (**Table 2** and **Supplementary Table S2**). Previous studies reported that a small sample size (less than 25) leads to an overestimate of r^2 (Khatkar et al., 2008; Deng et al., 2019), while Bohmanova et al. (2010) reported that at least 55 and 444 individuals were required for accurate estimation of r^2 and D′, respectively. Other studies have found that D′ statistics are more affected by population size than r^2 (Ardlie et al., 2002; Jemaa et al., 2019). Therefore, estimated r^2 values in the present study are more reliable than the D' statistics. Comparing uncorrected and corrected r^2 for sample size revealed that the differences in smaller populations are greater. The corrected vs. uncorrected r^2 values changed from 0.27 to 0.24 (around 0.02 units) in the pooled IRI, which has 33 individuals, but from 0.35 to 0.28 (around 0.07 units) in ANA with 12 individuals, 0.34 to 0.27 (around 0.07 units) in EGY with

TABLE 2 | Distance and linkage disequilibrium (corrected r^2) between adjacent polymorphic SNPs for IRI, ANA, EGY, and PAK water buffalo populations.

Chromosome	SNP number	Distance (Kb)	IRI	ANA	EGY	PAK
1	3,583	44.1	0.24 ± 0.25	0.27 ± 0.26	0.28 ± 0.27	0.23 ± 0.24
2	3,024	45.1	0.24 ± 0.26	0.3 ± 0.28	0.27 ± 0.27	0.24 ± 0.24
3	2,708	44.8	0.23 ± 0.25	0.27 ± 0.27	0.28 ± 0.27	0.22 ± 0.23
4	2,731	44.1	0.23 ± 0.24	0.28 ± 0.27	0.28 ± 0.27	0.22 ± 0.22
5	2,601	46.3	0.24 ± 0.26	0.29 ± 0.27	0.29 ± 0.27	0.24 ± 0.24
6	2,649	45	0.23 ± 0.25	0.29 ± 0.28	0.26 ± 0.26	0.21 ± 0.22
7	2,505	44.9	0.22 ± 0.24	0.26 ± 0.27	0.25 ± 0.26	0.22 ± 0.23
8	2,416	46.8	0.24 ± 0.26	0.28 ± 0.27	0.28 ± 0.27	0.23 ± 0.23
9	2,268	46.4	0.22 ± 0.24	0.3 ± 0.28	0.26 ± 0.26	0.23 ± 0.23
10	2,307	45	0.22 ± 0.24	0.28 ± 0.26	0.28 ± 0.27	0.22 ± 0.23
11	2,368	45.2	0.23 ± 0.25	0.31 ± 0.28	0.27 ± 0.26	0.23 ± 0.23
12	1,933	47.1	0.22 ± 0.25	0.27 ± 0.26	0.26 ± 0.26	0.22 ± 0.23
13	1,872	44.7	0.21 ± 0.23	0.23 ± 0.24	0.26 ± 0.26	0.2 ± 0.22
14	1,945	42.7	0.22 ± 0.24	0.25 ± 0.25	0.26 ± 0.26	0.21 ± 0.22
15	1,798	47.2	0.19 ± 0.24	0.29 ± 0.28	0.27 ± 0.27	0.2 ± 0.21
16	1,742	46.6	0.23 ± 0.26	0.27 ± 0.26	0.26 ± 0.27	0.24 ± 0.24
17	1,658	45.1	0.23 ± 0.26	0.29 ± 0.28	0.25 ± 0.25	0.22 ± 0.23
18	1,397	47	0.2 ± 0.23	0.25 ± 0.26	0.24 ± 0.26	0.19 ± 0.21
19	1,384	45.9	0.22 ± 0.25	0.27 ± 0.26	0.25 ± 0.25	0.22 ± 0.23
20	1,584	45.3	0.2 ± 0.25	0.28 ± 0.27	0.28 ± 0.28	0.22 ± 0.23
21	1,510	45.7	0.22 ± 0.24	0.29 ± 0.28	0.22 ± 0.24	0.21 ± 0.22
22	1,379	44.4	0.22 ± 0.24	0.23 ± 0.24	0.25 ± 0.26	0.21 ± 0.23
23	1,115	46.7	0.22 ± 0.25	0.28 ± 0.28	0.27 ± 0.27	0.21 ± 0.22
24	1,462	42.4	0.21 ± 0.22	0.28 ± 0.27	0.26 ± 0.26	0.21 ± 0.22
25	991	43.1	0.2 ± 0.24	0.26 ± 0.26	0.24 ± 0.26	0.18 ± 0.21
26	1,178	43.5	0.2 ± 0.23	0.26 ± 0.26	0.24 ± 0.26	0.19 ± 0.21
27	1,017	44.6	0.2 ± 0.22	0.25 ± 0.25	0.24 ± 0.26	0.21 ± 0.22
28	1,043	44.1	0.22 ± 0.24	0.24 ± 0.25	0.27 ± 0.27	0.22 ± 0.23
29	1,076	47.2	0.2 ± 0.23	0.25 ± 0.26	0.24 ± 0.25	0.19 ± 0.21
30	2,181	68.2	0.29 ± 0.3	0.59 ± 0.32	0.39 ± 0.31	0.3 ± 0.3
Average	1914	46.7	0.24 ± 0.24	0.28 ± 0.27	0.27 ± 0.26	0.22 ± 0.29

15 individuals, and 0.27 to 0.22 (around 0.05 units) in PAK with 23 individuals (**Supplementary Table S2**). If Iranian and Pakistani populations were considered individually, the bias in r^2 estimates increased because of the smaller sample size in the individual populations. These results show that correction of r^2 for sample size is necessary.

The corrected average r^2 values for individual populations from Iran, including AZI, KHU, and MAZ, were consistent and slightly lower than the values reported by Mokhber et al. (2019a) for AZI and KHU but not for MAZ. They found an r^2 of 0.27, 0.29, and 0.32 for AZI, KHU, and MAZ, respectively, using a larger dataset for AZI and KHU, but not MAZ. The difference in r^2 for MAZ was due to the correction method for average r^2 values.

Much lower values that obtaining in the present study were obtained r^2 values were obtained using the 90 K Buffalo SNP genotyping array in a study of 430 pure Mediterranean buffaloes and 65 Chinese crossbred buffalo, which gave an r^2 of 0.13 and 0.09, respectively (Deng et al., 2019). The mean value r^2 for adjacent SNPs in a study of 384 Brazilian Murrah buffaloes using the Bovine HD array in buffalo (Borquis et al., 2014), which provided 16,580 polymorphic loci from the 688,593 markers on the array, obtained and r^2 of 0.29. When the 90 K Buffalo Axiom array was used with a sample of 452 Brazilian Murrah buffaloes, 58,585 SNPs were polymorphic, and the same genome-wide r^2 of 0.29 was obtained, while the r^2 and |D|' for each chromosome were between 0.17 and 0.33 and 0.41 and 0.80, respectively (Cardoso et al., 2015). Using genomic information for 70 Iranian native cattle belonging to seven breeds (10 samples for each breed), Karimi et al. (2015) obtained average r^2 for the adjacent SNP markers of between 0.321 and 0.393.

The percentages of adjacent markers in IRI, ANA, EGY, and PAK populations with corrected r^2 greater than 0.2 (0.12) were 46, 52, 51, and 47% (**Supplementary Table S3**). The mean r^2 for adjacent markers can be used to assess their suitability for GWAS and the estimation of breeding values. An r^2 higher than 0.3 is recommended for GWAS (Ardlie et al., 2002), while an LD of more than 0.2 is considered essential for estimating genomic breeding values (Meuwissen et al., 2001).

The mean and standard deviation of D′, which represents the frequency of recombination events between adjacent SNPs, was 0.74, 0.67, 0.64, and 0.72 for IRI, ANA, EGY, and PAK, respectively (see **Supplementary Table S2**). A D′ value close to one implies that ancestral haplotypes have not been separated by recombination over time. In general, D′ is more affected by sample size than r^2 but less influenced by allele frequency. The pooled Iranian (IRI) population had the highest D' (0.74), while the EGY had the lowest (0.64).

Population history, including mutation, selection, recombination, and migration, affects the genome structure and will be reflected in the value of r^2. Factors such as sample size, the threshold for the frequency of rare alleles, the density of SNP, and the distances between markers will also affect the results. Further, the way that samples are selected may distort the diversity estimated for a

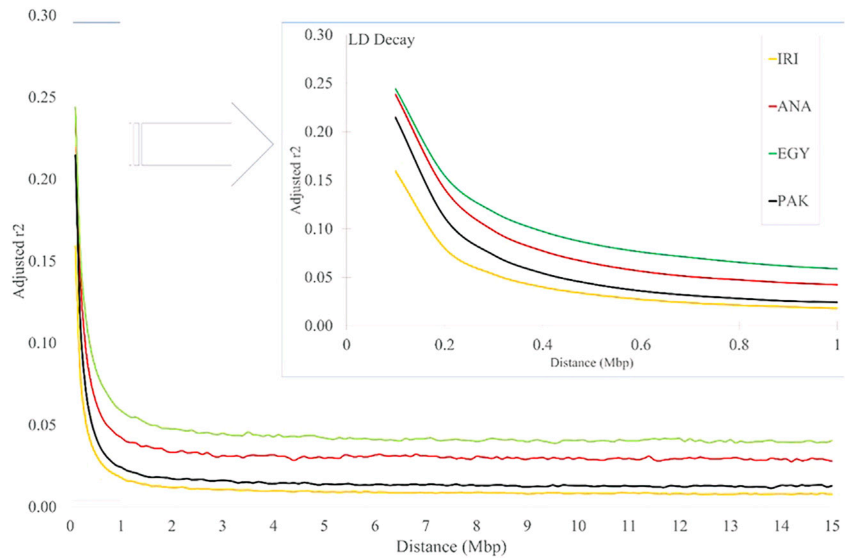

FIGURE 3 | LD decay for increasing distance (Mb) for IRI, ANA, EGY, and PAK water buffalo populations.

population. A study on pig breeds using a 50 K SNP array and a large number of samples in each genetic group identified high selection pressure and low diversity in populations as the reasons for the high LD found (Badke et al., 2012). In the present study, we pooled some populations because of the small sample size; in addition, we corrected LD estimates for sample size, and only SNPs with reasonable MAF (>0.05) were included. Because D' is more sensitive to sample size, we used the corrected r^2 values for subsequent analysis of LD decay, PLDP, and Ne.

LD Decay

As expected, the average r^2 values decreased with increasing distance between pairwise SNPs for all the studied populations (**Figure 3** and **Supplementary Table S4**). The values for IRI, ANA, EGY, and PAK were 0.367, 0.441, 0.411, and 0.432, respectively, for distances less than 10 Kb and 0.16, 0.24, 0.24, and 0.21, respectively, for distances less than 100Kb, which reduced rapidly to 0.018, 0.042, 0.059 and 0.024 (respectively) for a distance between markers of 1 Mb (**Figure 3** and **Supplementary Table S4**). In all the populations, the LD then remained constant for distances greater than 2 Mb to the longest distance considered (15 Mb) (**Supplementary Table S4**). The LD decayed slowly in EGY and ANA and in individual Iranian and Pakistani breeds. The highest LD, especially at longer distances, was seen MAZ and KUN. This may be due to the rapid decline of these populations in more recent generations. The effect of correcting r^2 was smaller (6–20 percent) for distances <10 kb and increased to more than 50 percent for distances >1 Mb and to 70–80 percent for distances >10 Mb. This suggests that r^2 values are more affected at longer distances by population size (**Supplementary Table S4**). Comparing the LD for individual Iran populations (AZI, KHU, and MAZ) obtained here with Mokhber et al. (2019a), which used a larger sample size (more than 200), LD estimates at >100 Kb were similar, whereas at greater distances, the results were significantly different.

Lu et al. (2020) calculated the rate of LD decay in Chinese river and swamp buffaloes and found that the LD of river buffaloes was higher than that of a swamp and that the rate of LD decay in swamp buffaloes was higher than for river buffaloes for all marker distances. These data reflect the stronger genetic selection in the river buffalo breeds compared with the swamp breeds. The rate of LD decay in Chinese crossbred buffaloes has been reported to be higher than in pure Mediterranean buffalo at a distance of 600 Kb (Deng et al., 2019), possibly as a result of recent cross-breeding.

A similar situation is seen for cattle where the LD is higher in dairy cattle, which are under stronger selection than beef breeds (Qanbari et al., 2010). The pattern of LD in German Holstein cattle gave an r^2 of about 0.3 for a distance less than 25Kb, which decreased to 0.24 for distances of 50–75 Kb (Qanbari et al., 2010), whereas in Australian Holstein bulls, r^2 varied from 0.402 to 0.073 as the distance increased from 20 to 500 Kb (Khatkar et al., 2008). For beef cattle, where selection is less intense, the r^2 for Angus, Charolais, and crossbred beef breeds (Angus × Charolais) decreased from 0.23 to 0.19, 0.16 to 0.12, and 0.15 to 0.11, respectively, for distances 30 to 100Kb, respectively (Lu et al., 2012).

Persistency of LD Phase

PLDP was calculated from the correlation between paired SNPs at distances of 0–15 Mb. An increase in the distance led to a decrease in PLDP between breeds (see **Table 3** and **Supplementary Table S4**). At distances less than 100Kb, PLPD in all the populations was higher than 0.95 for buffalo populations from Iran, Turkey, Egypt, and Pakistan, which decreased to between 0.7 and 0.97 at 200Kb and then reduced rapidly. However, from 500 Kb to 1 Mb, the reduction in PLPD was less than seen between 200 and 500 Kb (**Table 3** and **Supplementary Table S5**). The PLDP within breeds from the same geographical area that formed pools was higher than the other comparisons (**Supplementary Table S5**).

PLPD among individual populations from Iran was above 0.95 for a distance less than 100Kb, which is similar results of Mokhber

TABLE 3 | Consistency of gametic phase at given distances between IRI, ANA, EGY, and PAK water buffalo populations.

Populations	Distances between paired SNPs (kbp)									
	>100	100–200	200–300	300–400	400–500	500–600	600–700	700–800	800–900	900–1,000
IRI-ANA	0.956	0.876	0.773	0.463	0.407	0.229	0.296	0.011	0.140	0.216
IRI_EGY	0.969	0.905	0.715	0.287	0.271	0.203	0.099	0.391	0.173	0.208
IRI_PAK	0.958	0.848	0.620	0.292	0.324	0.180	0.195	0.176	0.060	0.141
ANA-EGY	0.956	0.876	0.773	0.463	0.407	0.229	0.296	0.011	0.140	0.216
ANA_PAK	0.969	0.905	0.715	0.287	0.271	0.203	0.099	0.391	0.173	0.208
EGY_PAK	0.966	0.923	0.713	0.423	0.282	0.149	0.204	0.184	-0.059	0.145

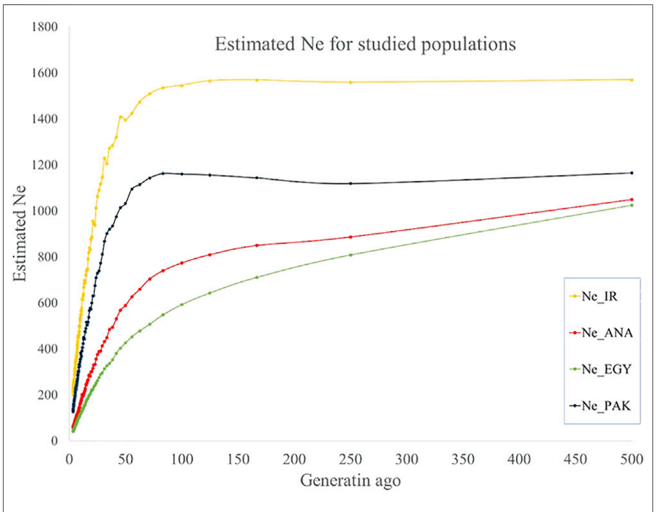

FIGURE 4 | Estimated Ne for IRI (pooled Iranian breeds including AZI, KHU, and MAZ), ANA, EGY, and PAK (pooled Pakistani breeds including KHU and NIL) water buffalo populations for past generations.

et al. (2019a) who reported values of 0.99, 0.96, and 0.95 at distances less than 100Kb, which reduced to 0.74, 0.25, and 0.12 at distances below than 1 Mb for AZI-KHU, AZI-MAZ, and KHU-MAZ populations, respectively.

These high PLPD values suggest that there may have been genetic exchange among these populations. The highest correlations previously reported among other pure and crossbred buffalo populations were 0.47 at the distance of 100 Kb (Deng et al., 2019), showing that the LD phase between independent populations tends not to be maintained. The value of PLDP among European, African, and African-European cattle breeds has been reported as 0.77, 0.71, and 0.65, respectively, at distances less than 10Kb and below 0.5 at distances greater than 50 Kb (Gautier et al., 2007). In Australian Holstein and New Zealand Jersey breeds, the PLDP correlation was 0.97 (De Roos et al., 2008), which is surprisingly high for breeds with different genetic histories. For beef breeds, PLDP between Charolais and Angus, Charolais and crossbred cattle, and Angus and Crosses was 0.84, 0.81, and 0.77, respectively, at distances less than 70 Kb (Lu et al., 2012), so that exchange of information among these populations should be treated with caution.

Ne

Ne was estimated from the last 500 to recent generations in the present study. A trend of decreasing Ne was observed from more distant to recent generations: from 1,570 to 212, 1,049 to 59, 1,025 to 43, and 1,165 to 131 for IRI, ANA, EGY, and PAK breeds, respectively, from 500 generations ago to three last generations (**Figure 4** and **Supplementary Table S6**). Similar trends for a decline in Ne from past to recent generations have been reported for buffalo (Mokhber et al., 2019b) other species (Sargolzaei et al., 2008; Moradi et al., 2012). The Ne of Canadian and American Holstein cattle decreased from 1,400 to less than 100 from 500 generations ago to recent generations (Sargolzaei et al., 2008). For sheep, the Ne of Zel and Lori-Bakhtiari breeds reduced from 4,900 to 840 and 4,900 to 532 animals from 2000 generations ago to the 20 last generations, respectively (Moradi et al., 2012). Ne for Sunite, German Mutton Merino, and Dorper sheep breeds has decreased from 1,506 to 207, 1,678 to 74, and 1,506 to 67, respectively, from 2000 generations ago to the seven last generations (Zhao et al., 2014).

The conservation of genetic and biological diversity is dependent on Ne (Wang, 2005). According to the FAO (1992), when Ne is equal to 25, 50, 125, 250, and 500, genetic diversity will shrink 18, 10, 4, 1.6, and 0.8 percent over 10 next generations, respectively. Evidence accumulated since 1980 shows that a Ne of more than 100 is necessary to maintain fitness over the subsequent 10 generations. Meuwissen (2009) showed that, with Ne greater than 100 individuals, the population would be sufficiently genetic diverse to survive in the long term, while to conserve the evolutionary potential of the population, it is better than Ne is more than 1,000 individuals (Frankham et al., 2014).

The present study showed that Ne of Iranian and Pakistani populations are greater than the population size threshold necessary to be genetically viable (Meuwissen, 2009). The main concern for all the studied populations is the rapid reduction in Ne in recent generations. Therefore, controlling the decline in Ne and increase in efficiency of economic production, e.g., by well-designed breeding programs, is necessary to prevent increasing inbreeding and eventually genetic extinction.

CONCLUSION

In the present study, the LD structure, PLDP, and Ne were determined for seven buffalo populations and two populations pooled based on country or origin. The level of LD found indicated that it is appropriate to use the Affymetrix Axiom 90 K SNP genomic array for GWAS and GS in these populations. The correlation between the LD information and PLDP between geographically close populations was high, meaning that genomic information from one population can be used efficiently to predict genetic effects in another. We found that results are affected by sample size, which could be partially corrected for; however, additional data should be obtained to be confident of the results. The Ne in recent generations has declined to the extent that breeding plans are urgently required to ensure that these buffalo populations are not at risk of being lost.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Ethical review and approval was not required as no animal work was undertaken and the data were obtained from research published by Colli et al. (2018).

AUTHOR CONTRIBUTIONS

SR, MG, MM, and JW participated in the conception and design of the study. SR and MM analyzed the data. MG and MM drafted the manuscript. MM and JW revised themanuscript. All authors read and approved thefinal manuscript.

ACKNOWLEDGMENTS

MG and MM are supported by the Urmia University and JW was supported by the project LEO: Livestock Environment Open Data, 16.2—PSRN 2014-2020, funded through Fondo Europeo Agricolo per lo Sviluppo Rurale (FEASR).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.608186/full#supplementary-material

Supplementary Figure S1 | Principal components analysis based on the genomic kinship coefficients between all studied individuals.

Supplementary Figure S2 | Principal components analysis based on the genomic kinship coefficients between all studied individuals.

Supplementary Figure S3 to 11 | LD decay over physical distance for AZI, KHU, MAZ, IRI, ANA, EGY, KUN, NIL, and PAK, separated for each population and chromosomes.

Supplementary Table S1 | Genetic diversity between all studied populations by unbiased F_{ST} statistics.

Supplementary Table S2 | Table of mean| D'| and uncorrected and corrected r^2 values for AZI, KHU, MAZ, IRI, ANA, EGY, KUN, NIL, and PAK buffalo populations.

Supplementary Table S3 | Frequency of r^2 and |D'| values for AZI, KHU, MAZ, IRI, ANA, EGY, KUN, NIL, and PAK buffalo populations.

Supplementary Table S4 | Average LD decay over physical distance for AZI, KHU, MAZ, IRI, ANA, EGY, KUN, NIL, and PAK buffalo populations.

Supplementary Table S5 | Consistency of gametic phase at given distances for AZI, KHU, MAZ, IRI, ANA, EGY, KUN, NIL, and PAK buffalo breed pairs.

Supplementary Table S6 | Effective population size for AZI, KHU, MAZ, IRI, ANA, EGY, KUN, NIL, and PAK buffalo breed in a given number of generations ago.

REFERENCES

Abecasis, G. R., Noguchi, E., Heinzmann, A., Traherne, J. A., Bhattacharyya, S., and Cookson, W. O. (2001). Extent and Distribution of Linkage Disequilibrium in Three Genomic Regions. *Am. J. Hum. Genet.* 68, 191–197. doi:10.1086/316944

Al-Mamun, H. A., A Clark, S., Kwan, P., and Gondro, C. (2015). Genome-wide Linkage Disequilibrium and Genetic Diversity in Five Populations of Australian Domestic Sheep. *Genet. Sel Evol.* 47, 90. doi:10.1186/s12711-015-0169-6

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome Res.* 19 (9), 1655–1664. doi:10.1101/gr.094052.109

Alhaddad, H., Khan, R., Grahn, R. A., Gandolfi, B., Mullikin, J. C., Cole, S. A., et al. (2013). Extent of Linkage Disequilibrium in the Domestic Cat, *Felis silvestris* Catus, and its Breeds. *PLoS One* 8 (1), e53537. doi:10.1371/journal.pone.0053537

Ardlie, K. G., Kruglyak, L., and Seielstad, M. (2002). Patterns of Linkage Disequilibrium in the Human Genome. *Nat. Rev. Genet.* 3, 299–309. doi:10.1038/nrg777

Azizi, Z., Moradi Shahrbabak, H., Moradi Shahrbabak, M., Rafat, S. A., and Shodja, J. (2016). Genetic Classification of Azari and North Ecotype Buffalo Population Using SVM Method. *Iranian J. Anim. Sci.* 47 (2), 279–290. doi:10.22059/ijas.2016.59033

Badke, Y. M., Bates, R. O., Ernst, C. W., Schwab, C., and Steibel, J. P. (2012). Estimation of Linkage Disequilibrium in Four US Pig Breeds. *BMC Genomics* 13, 24. doi:10.1186/1471-2164-13-24

Barbato, M., Orozco-terWengel, P., Tapio, M., and Bruford, M. W. (2015). SNeP: a Tool to Estimate Trends in Recent Effective Population Size Trajectories Using Genome-wide SNP Data. *Front. Genet.*, 6, 109.þdoi:10.3389/fgene.2015.00109

Biegelmeyer, P., Gulias-Gomes, C. C., Caetano, A. R., Steibel, J. P., and Cardoso, F. F. (2016). Linkage Disequilibrium, Persistence of Phase and Effective Population Size Estimates in Hereford and Braford Cattle. *BMC Genet.* 17 (1), 32. doi:10.1186/s12863-016-0339-8

Bohmanova, J., Sargolzaei, M., and Schenkel, F. S. (2010). Characteristics of Linkage Disequilibrium in North American Holsteins. *BMC Genomics* 11, 421–435. doi:10.1186/1471-2164-11-421

Borquis, R. R. A., Baldi, F., de Camargo, G. M. F., Cardoso, D. F., Santos, D. J. A., Lugo, N. H., et al. (2014). Water buffalo Genome Characterization by the Illumina BovineHD BeadChip. *Genet. Mol. Res.* 13, 4202–4215. doi:10.4238/2014.june.9.6

Brito, L. F., Jafarikia, M., Grossi, D. A., Kijas, J. W., Porto-Neto, L. R., Ventura, R. V., et al. (2015). Characterization of Linkage Disequilibrium, Consistency of Gametic Phase and Admixture in Australian and Canadian Goats. *BMC Genet.* 16 (1), 67. doi:10.1186/s12863-015-0220-1

Browning, S. R., and Browning, B. L. (2007). Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies by Use of Localized Haplotype Clustering. *Am. J. Hum. Genet.* 81 (5), 1084–1097. doi:10.1086/521987

Cardoso, D., Aspilcueta-Borquis, A., Santos, D., Hurtado-Lugo, H. N., De Camargo, G., Scalez, D., et al. (2015). "Study of Linkage Disequilibrium in Brazilian Dairy Buffaloes," in Proceedings, 10th World Congress on Genetics Applied to Livestock Production, Vancouver, August 17–22, 2014.

Colli, L., Milanesi, M., Vajana, E., Iamartino, D., Bomba, L., Puglisi, F., et al. (2018). New Insights on Water Buffalo Genomic Diversity and Post-Domestication Migration Routes from Medium Density SNP Chip Data. *Front. Genet.* 9, 53. doi:10.3389/fgene.2018.00053

Corbin, L. J., Blott, S. C., Swinburne, J. E., Vaudin, M., Bishop, S. C., and Woolliams, J. A. (2010). Linkage Disequilibrium and Historical Effective Population Size in the Thoroughbred Horse. *Anim. Genet.* 41 (Suppl. 2), 8–15. doi:10.1111/j.1365-2052.2010.02092.x

Corbin, L. J., Liu, A. Y. H., Bishop, S. C., and Woolliams, J. A. (2012). Estimation of Historical Effective Population Size Using Linkage Disequilibria with Marker Data. *J. Anim. Breed. Genet.* 129 (4), 257–270. þ. doi:10.1111/j.1439-0388.2012.01003.x

De Roos, A. P. W., Hayes, B. J., Spelman, R. J., and Goddard, M. E. (2008). Linkage Disequilibrium and Persistence of Phase in Holstein-Friesian, Jersey and Angus Cattle. *Genetics* 179, 1503–1512. doi:10.1534/genetics.107.084301

Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J., Molina Alcalá, A., Sonstegard, T. S., et al. (2014). Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. *Plos Genet.* 10, e1004254. doi:10.1371/journal.pgen.1004254

Deng, T., Liang, A., Liu, J., Hua, G., Ye, T., Liu, S., et al. (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. doi:10.3389/fgene.2018.00688

Frankham, R., Bradshaw, C. J. A., and Brook, B. W. (2014). Genetics in Conservation Management: Revised Recommendations for the 50/500 Rules, Red List Criteria and Population Viability Analyses. *Biol. Conservation* 170, 56–63. doi:10.1016/j.biocon.2013.12.036

Frankham, R. (2005). Stress and Adaptation in Conservation Genetics. *J. Evol. Biol* 18 (4), 750–755. doi:10.1111/j.1420-9101.2005.00885.x

Fu, W., Dekkers, J. C., Lee, W. R., and Abasht, B. (2015). Linkage Disequilibrium in Crossbred and Pure Line Chickens. *Genet. Sel Evol.* 47 (1), 11. doi:10.1186/s12711-015-0098-4

Gautier, M., Faraut, T., Moazami-Goudarzi, K., Navratil, V., Foglio, M., Grohs, C., et al. (2007). Genetic and Haplotypic Structure in 14 European and African Cattle Breeds. *Genetics* 177, 1059–1070. doi:10.1534/genetics.107.075804

Ghoreishifar, S. M., Moradi-Shahrbabak, H., Fallahi, M. H., Jalil Sarghale, A., Moradi-Shahrbabak, M., Abdollahi-Arpanahi, R., et al. (2020). Genomic Measures of Inbreeding Coefficients and Genome-wide Scan for Runs of Homozygosity Islands in Iranian River buffalo, Bubalus Bubalis. *BMC Genet.* 21 (1), 16–12. doi:10.1186/s12863-020-0824-y

Gianola, D., Qanbari, S., and Simianer, H. (2013). An Evaluation of a Novel Estimator of Linkage Disequilibrium. *Heredity* 111, 275–285. doi:10.1038/hdy.2013.46

Hayes, B. J., Visscher, P. M., McPartlan, H. C., and Goddard, M. E. (2003). Novel Multilocus Measure of Linkage Disequilibrium to Estimate Past Effective Population Size. *Genome Res.* 13 (4), 635–643. doi:10.1101/gr.387103

Hill, W. G., and Weir, B. S. (1994). Maximum-likelihood Estimation of Gene Location by Linkage Disequilibrium. *Am. J. Hum. Genet.* 54, 705–714.

Hill, W. G., and Robertson, A. (1968). Linkage Disequilibrium in Finite Populations. *Theoret. Appl. Genet.* 38 (6), 226–231. doi:10.1007/bf01245622

Hui, T. J., and Burt, A. (2020). Estimating Linkage Disequilibrium from Genotypes under Hardy-Weinberg Equilibrium. *BMC Genet.* 21 (1), 21–11. þ. doi:10.1186/s12863-020-0818-9

Iamartino, D., Nicolazzi, E. L., Van Tassell, C. P., Reecy, J. M., Fritz-Waters, E. R., Koltes, J. E., et al. (2017). Design and Validation of a 90K SNP Genotyping Assay for the Water Buffalo (Bubalus Bubalis). *PLoS ONE* 12, e0185220. doi:10.1371/journal.pone.0185220

Jemaa, S. B., Thamri, N., Mnara, S., Rebours, E., Rocha, D., and Boussaha, M. (2019). Linkage Disequilibrium and Past Effective Population Size in Native Tunisian Cattle. *Genet. Mol. Biol.* 42 (1), 52–61. doi:10.1590/1678-4685-gmb-2017-0342

Jombart, T., and Ahmed, I. (2011). Adegenet 1.3-1: New Tools for the Analysis of Genome-wide SNP Data. *Bioinformatics* 27 (21), 3070–3071. þ. doi:10.1093/bioinformatics/btr521

Karimi, K., Esmailizadeh Koshkoiyeh, A., and Gondro, C. (2015). Comparison of Linkage Disequilibrium Levels in Iranian Indigenous Cattle Using Whole Genome SNPs Data. *J. Anim. Sci. Technol.* 57 (1), 47. doi:10.1186/s40781-015-0080-2

Khatkar, M. S., Nicholas, F. W., Collins, A. R., Zenger, K. R., Cavanagh, J. A., Barris, W., et al. (2008). Extent of Genome-wide Linkage Disequilibrium in Australian Holstein-Friesian Cattle Based on a High-Density SNP Panel. *BMC genomics* 9, 187. doi:10.1186/1471-2164-9-187

Lewontin, R. C. (1988). On Measures of Gametic Disequilibrium. *Genetics* 120, 849–852. doi:10.1093/genetics/120.3.849

Lu, D., Sargolzaei, M., Kelly, M., Li, C., VanderVoort, G., Wang, Z., et al. (2012). Linkage Disequilibrium in Angus, Charolais, and Crossbred Beef Cattle. *Front. Gene* 3, 152. doi:10.3389/fgene.2012.00152

Lu, X. R., Duan, A. Q., Li, W. Q., Abdel-Shafy, H., Rushdi, H. E., Liang, S. S., et al. (2020). Genome-wide Analysis Reveals Genetic Diversity, Linkage Disequilibrium, and Selection for Milk Production Traits in Chinese buffalo Breeds. *J. Dairy Sci.* 103, 4545. doi:10.3168/jds.2019-17364

McKay, S. D., Schnabel, R. D., Murdoch, B. M., Matukumalli, L. K., Aerts, J., Coppieters, W., et al. (2007). Whole Genome Linkage Disequilibrium Maps in Cattle. *BMC Genet.* 8, 74. doi:10.1186/1471-2156-8-74

Meadows, J. R., Chan, E. K., and Kijas, J. W. (2008). Linkage Disequilibrium Compared Between Five Populations of Domestic Sheep. *BMC Genet.* 9 (1), 1–10. doi:10.1186/1471-2156-9-61

Meuwissen, T. (2009). Genetic Management of Small Populations: A Review. *Acta Agriculturae Scand. Section A - Anim. Sci.* 59 (2), 71–79. þ. doi:10.1080/09064700903118148

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-wide Dense Marker Maps. *Genetics* 157, 1819–1829. doi:10.1093/genetics/157.4.1819

Minervino, A. H. H., Zava, M., Vecchio, D., and Borghese, A. (2020). Bubalus Bubalis: A Short story. *Front. Vet. Sci.*, 7, 971.þdoi:10.3389/fvets.2020.570413

Mokhber, M., Moradi Shahre Babak, M., Sadeghi, M., Moradi Shahrbabak, H., and Rahmani-Nia, J. (2019b). Estimation of Effective Population Size of Iranian Water buffalo by Genomic Data. *Iranian J. Anim. Sci.* 50 (3), 197–205. þ.

Mokhber, M., Moradi-Shahrbabak, M., Sadeghi, M., Moradi-Shahrbabak, H., Stella, A., Nicolzzi, E., et al. (2018). A Genome-wide Scan for Signatures of Selection in Azeri and Khuzestani buffalo Breeds. *BMC genomics* 19, 449. doi:10.1186/s12864-018-4759-x

Mokhber, M., Shahrbabak, M. M., Sadeghi, M., Shahrbabak, H. M., Stella, A., Nicolzzi, E., et al. (2019a). Study of Whole Genome Linkage Disequilibrium Patterns of Iranian Water buffalo Breeds Using the Axiom Buffalo Genotyping 90K Array. *PLoS One* 14, e0217687. doi:10.1371/journal.pone.0217687

Mokry, F. B., Buzanskas, M. E., de Alvarenga Mudadu, M., do Amaral Grossi, D., Higa, R. H., Ventura, R. V., et al. (2014). Linkage Disequilibrium and Haplotype Block Structure in a Composite Beef Cattle Breed. *BMC Genomics* 15 (Suppl. 7), S6. doi:10.1186/1471-2164-15-S7-S6

Moradi, M. H., Nejati-Javaremi, A., Moradi-Shahrbabak, M., Dodds, K. G., and McEwan, J. C. (2012). Genomic Scan of Selective Sweeps in Thin and Fat Tail Sheep Breeds for Identifying of Candidate Regions Associated with Fat Deposition. *BMC Genet.* 13, 10. doi:10.1186/1471-2156-13-10

Mueller, J. C. (2004). Linkage Disequilibrium for Different Scales and Applications. *Brief. Bioinform.* 5, 355–364. doi:10.1093/bib/5.4.355

Niu, H., Zhu, B., Guo, P., Zhang, W., Xue, J., Chen, Y., et al. (2016). Estimation of Linkage Disequilibrium Levels and Haplotype Block Structure in Chinese Simmental and Wagyu Beef Cattle Using High-Density Genotypes. *Livestock Sci.* 190, 1–9. doi:10.1016/j.livsci.2016.05.012

Pfahler, S., and Distl, O. (2015). Effective Population Size, Extended Linkage Disequilibrium and Signatures of Selection in the Rare Dog Breed Lundehund. *PloS one* 10 (4), e0122680. doi:10.1371/journal.pone.0122680

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal Components Analysis Corrects for Stratification in Genome-wide Association Studies. *Nat. Genet.* 38, 904–909. doi:10.1038/ng1847

Pritchard, J. K., and Przeworski, M. (2001). Linkage Disequilibrium in Humans: Models and Data. *Am. J. Hum. Genet.* 69, 1–14. doi:10.1086/321275

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155, 945–959. doi:10.1093/genetics/155.2.945

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* 81, 559–575. doi:10.1086/519795

Qanbari, S., Pimentel, E. C., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A. R., et al. (2010). The Pattern of Linkage Disequilibrium in German Holstein Cattle. *Anim. Genet.* 41, 346–356. doi:10.1111/j.1365-2052.2009.02011.x

Qanbari, S. (2020). On the Extent of Linkage Disequilibrium in the Genome of Farm Animals. *Front. Genet.* 10, 1304. doi:10.3389/fgene.2019.01304

Rahmaninia, J., Miraei-Ashtiani, S. R., and Moradi Shahrbabak, H. (2015). Unsupervised Clustering Analysis of Population and Subpopulation Structure Using Dense SNP Markers. *Iranian J. Anim. Sci.* 46 (3), 277–287.

Sargolzaei, M., Schenkel, F. S., Jansen, G. B., and Schaeffer, L. R. (2008). Extent of Linkage Disequilibrium in Holstein Cattle in North America. *J. Dairy Sci.* 91, 2106–2117. doi:10.3168/jds.2007-0553

Shin, J.-B., Krey, J. F., Hassan, A., Metlagel, Z., Tauscher, A. N., Pagana, J. M., Sherman, N. E., Jeffery, E. D., Spinelli, K. J., Zhao, H., Wilmarth, P. A., Choi, D., David, L. L., Auer, M., and Barr-Gillespie, P. G. (2013). Molecular Architecture of the Chick Vestibular Hair Bundle. *Nat. Neurosci.* 16 (3), 365–374. doi:10.1038/nn.3312

Soysal, M. I., Tuna, Y. T., Gurcan, E. K., Ozkan, E., Kok, S., Castellano, N., et al. (2007). Anatolian Water Buffaloes Husbandry in Turkey: Preliminary Results on Somatic Characterization. *Ital. J. Anim. Sci.* 6 (Suppl. 2), 1302–1307. þ. doi:10.4081/ijas.2007.s2.1302

Strillacci, M. G., Moradi-Shahrbabak, H., Davoudi, P., Ghoreishifar, S. M., Mokhber, M., Masroure, A. J., et al. (2021). A Genome-wide Scan of Copy Number Variants in Three Iranian Indigenous River Buffaloes. *BMC genomics* 22 (1), 305–314. þ. doi:10.1186/s12864-021-07604-3

Sved, J. A. (1971). Linkage Disequilibrium and Homozygosity of Chromosome Segments in Finite Populations. *Theor. Popul. Biol.* 2, 125–141. doi:10.1016/0040-5809(71)90011-6

Sved, J. A. (2009). Linkage Disequilibrium and its Expectation in Human Populations. *Twin Res. Hum. Genet.* 12, 35–43. doi:10.1375/twin.12.1.35

Tenesa, A., Navarro, P., Hayes, B. J., Duffy, D. L., Clarke, G. M., Goddard, M. E., et al. (2007). Recent Human Effective Population Size Estimated from Linkage Disequilibrium. *Genome Res.* 17, 520–526. doi:10.1101/gr.6023607

Terwilliger, J. D. (1995). A Powerful Likelihood Method for the Analysis of Linkage Disequilibrium between Trait Loci and One or More Polymorphic Marker Loci. *Am. J. Hum. Genet.* 56, 777–787.

Villa-Angulo, R., Matukumalli, L. K., Gill, C. A., Choi, J., Van Tassell, C. P., and Grefenstette, J. J. (2009). High-resolution Haplotype Block Structure in the Cattle Genome. *BMC Genet.* 10, 19. doi:10.1186/1471-2156-10-19

Vonholdt, B. M., Stahler, D. R., Bangs, E. E., Smith, D. W., Jimenez, M. D., Mack, C. M., Niemeyer, C. C., Pollinger, J. P., and Wayne, R. K. (2010). A Novel Assessment of Population Structure and Gene Flow in Grey Wolf Populations of the Northern Rocky Mountains of the United States. *Mol. Ecol.* 19 (20), 4412–4427. doi:10.1111/j.1365-294x.2010.04769.x

Wang, J. (2005). Estimation of Effective Population Sizes From Data on Genetic Markers. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360 (1459), 1395–1409.

Wang, L., Sørensen, P., Janss, L., Ostersen, T., and Edwards, D. (2013). Genome-wide and Local Pattern of Linkage Disequilibrium and Persistence of Phase for 3 Danish Pig Breeds. *BMC Genet.* 14 (1), 115. doi:10.1186/1471-2156-14-115

Waples, R. K., Larson, W. A., and Waples, R. S. (2016). Estimating Contemporary Effective Population Size in Non-model Species Using Linkage Disequilibrium across Thousands of Loci. *Heredity* 117, 233–240. doi:10.1038/hdy.2016.60

Weir, B. S., and Hill, W. G. (1980). Effect of Mating Structure on Variation in Linkage Disequilibrium. *Genetics* 95, 477–488. doi:10.1093/genetics/95.2.477

Wultsch, C., Waits, L. P., and Kelly, M. J. (2016). A Comparative Analysis of Genetic Diversity and Structure in Jaguars (*Panthera onca*), Pumas (Puma Concolor), and Ocelots (Leopardus Pardalis) in Fragmented Landscapes of a Critical Mesoamerican Linkage Zone. *PloS one* 11 (3), 0151043. doi:10.1371/journal.pone.0151043

Zhao, H., Nettleton, D., and Dekkers, J. C. M. (2007). Evaluation of Linkage Disequilibrium Measures between Multi-Allelic Markers as Predictors of Linkage Disequilibrium between Single Nucleotide Polymorphisms. *Genet. Res.* 89, 1–6. doi:10.1017/s0016672307008634

Zhao, H., Nettleton, D., Soller, M., and Dekkers, J. C. M. (2005). Evaluation of Linkage Disequilibrium Measures between Multi-Allelic Markers as Predictors of Linkage Disequilibrium between Markers and QTL. *Genet. Res.* 86, 77–87. doi:10.1017/S001667230500769X

Zhao, Y., Wang, H., Chen, W., and Li, Y. (2014). Genetic Structure, Linkage Disequilibrium and Association Mapping of Verticillium Wilt Resistance in Elite Cotton (Gossypium Hirsutum L.) Germplasm Population. *PloS one* 9 (1), e86308. doi:10.1371/journal.pone.0086308

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership