

Complete Genomic Landscape Reveals Hidden Evolutionary History and Selection Signature in Asian Water Buffaloes (*Bubalus bubalis*)

Jingfang Si, Dongmei Dai, Neena Amatya Gorkhali, Mingshan Wang, Sheng Wang, Saroj Sapkota, Ram Chandra Kadel, Amir Sadaula, Aashish Dhakal, Md Omar Faruque, Abdullah Ibne Omar, Eka Meutia Sari, Hidayat Ashari, Muhammad Ihsan Andi Dagong, Marnoch Yindee, Hossam E. Rushdi, Hussein Elregalaty, Ahmed Amin, Mohamed A. Radwan, Lan Doan Pham, W. M. M. P. Hulugalla, G. L. L. Pradeepa Silva, Wei Zheng, Shahid Mansoor, Muhammad Basil Ali, Farhad Vahidi, Sahar Ahmed Al-Bayatti, Alfredo Pauciullo, Johannes A. Lenstra, J. Stuart F. Barker, Lingzhao Fang, Dong-Dong Wu,* Jianlin Han,* and Yi Zhang*

To identify the genetic determinants of domestication and productivity of Asian water buffaloes (*Bubalus bubalis*), 470 genomes of domesticated river and swamp buffaloes along with their putative ancestors, the wild water buffaloes (*Bubalus arnee*) are sequenced and integrated. The swamp buffaloes inherit the morphology of the wild buffaloes. In contrast, most river buffaloes are unique in their morphology, but their genomes cluster with the wild buffaloes. The levels of genomic diversity in Italian river and Indonesian swamp buffaloes decrease at opposite extremes of their distribution range. Purifying selection prevented the accumulation of harmful loss-of-function variants in the Indonesian buffaloes. Genes that evolved rapidly (e.g., *GKAP1*) following differential selections in the river and swamp buffaloes are involved in their reproduction. Genes related to milk production (e.g., *CSN2*) and coat color (e.g., *MC1R*) underwent strong selections in the dairy river buffaloes via soft and hard selective sweeps, respectively. The selective sweeps and single-cell RNA-seq data revealed the luminal cells as the key cell type in response to artificial selection for milk production of the dairy buffaloes. These findings show how artificial selection has been driving the evolutionary divergence and genetic differentiation in morphology and productivity of Asian water buffaloes.

1. Introduction

The evolutionary emergence, genetic differentiation, and breed development of domesticated animals have primarily been influenced by two biological processes, the adaptation to different natural environments and human-controlled managements and the artificial selection for specialized utilities.^[1,2] These processes also promote phenotypic and genetic diversity due to variations among species and breeding objectives. Genomic studies on chickens,^[3] dogs,^[4] pigs,^[5] bovine species,^[6–9] goats,^[10] and sheep^[11] have identified candidate genes and causative mutations. Interesting examples of convergent evolution have been reported, most notably for domestication,^[12] environmental adaptation,^[13–15] coat color,^[16] and genetic disorders.^[17] The identification of causative genes/mutations within species collectively contributes to our expanding understanding of the functional diversity of mammalian and avian gene repertoires.

In this context, the genomic landscape of domestication and breed improvement of



The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/advs.202407615>

© 2024 The Author(s). Advanced Science published by Wiley-VCH GmbH. This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/advs.202407615

J. Si, D. Dai, A. Dhakal, A. I. Omar, Y. Zhang
State Key Laboratory of Animal Biotech Breeding, National Engineering Laboratory for Animal Breeding
Key Laboratory of Animal Genetics
Breeding and Reproduction of Ministry of Agriculture and Rural Affairs
College of Animal Science and Technology
China Agricultural University
Beijing 100193, China
E-mail: yizhang@cau.edu.cn

Asian water buffaloes have been explored less extensively compared to other livestock species.^[18–21] The domesticated buffaloes represent an important farm animal genetic resource in the tropical and subtropical zones, in contrast to most other livestock species that are more productive in the temperate zones. There

N. A. Gorkhali, S. Sapkota
National Animal Breeding and Genetics Research Centre
National Animal Science Research Institute
Nepal Agriculture Research Council
Khumaltar, Lalitpur, Nepal

M. Wang, S. Wang, D.-D. Wu
State Key Laboratory of Genetic Resources and Evolution
Yunnan Laboratory of Molecular Biology of Domestic Animals
Kunming Institute of Zoology
Chinese Academy of Sciences
Kunming 650023, China
E-mail: wudongdong@mail.kiz.ac.cn

R. C. Kadel
Ministry of Industry, Tourism, Forests and Environment
Government of Gandaki Province
Pokhara, Nepal

A. Sadaula
National Trust for Nature Conservation- Biodiversity Conservation Center
Sauraha, Chitwan, Nepal

M. O. Faruque
Department of Animal Breeding and Genetics
Bangladesh Agricultural University
Mymensingh 2202, Bangladesh

E. M. Sari, H. Ashari
Department of Animal Science
Faculty of Agriculture
Syiah Kuala University
Darussalam-Banda Aceh 23111, Indonesia

M. I. A. Dagong
Animal Science Faculty
Hasanuddin University
Makassar, Indonesia

M. Yindee
Akkhratchakumari Veterinary College
Walailak University
Thaiburi 222, Thailand

H. E. Rushdi, A. Amin, M. A. Radwan
Department of Animal Production
Faculty of Agriculture
Cairo University
Algammaa Street, Giza 12613, Egypt

H. Elregalaty
Animal Production Research Institute
Agricultural Research Center
Dokki, Giza, Egypt

L. D. Pham
Key Laboratory of Animal Cell Technology
National Institute of Animal Sciences
Tu Liem, Hanoi 100000, Vietnam

W. M. M. P. Hulugalla, G. L. L. P. Silva
Department of Animal Science
Faculty of Agriculture
University of Peradeniya
, Sri Lanka

W. Zheng
Guangxi Key Laboratory of Buffalo Genetics, Reproduction and Breeding
Guangxi Buffalo Research Institute
Chinese Academy of Agricultural Sciences
Nanning 530001, China

S. Mansoor, M. B. Ali
National Institute for Biotechnology and Genetic Engineering (NIBGE)
Faisalabad, Pakistan

are two distinct types of domesticated buffaloes^[22] with different chromosome numbers of $2n = 50$ for river buffalo and $2n = 48$ for swamp buffalo.^[23–25] Genetic studies have revealed considerable differences between the river and swamp buffaloes, which are justified to be classified as two (sub)species of *Bubalus bubalis bubalis* and *Bubalus bubalis kerabau*, respectively.^[26] After their independent domestications,^[27–29] the swamp buffaloes have been bred into excellent draught animals playing an essential role in the development of rice culture. In contrast, the river buffaloes have been developed into proficient dairy animals producing milk with high fat but low cholesterol. In this study, we sequenced and integrated 470 genomes from 51 Asian water buffalo populations for an improved understanding of their complete genomic landscape driven by natural and artificial selections underlying their evolutionary divergence and domestication and for providing solid evidence of the genes controlling the coat color and milk production of the river buffaloes and the genes underwent a rapid evolution of reproductive traits in all buffaloes.

2. Results

2.1. Genomic Variation in Asian Water Buffaloes

We analyzed whole genome sequencing data of 470 Asian water buffaloes across their entire range from East Asia, South-east Asia, South Asia, the Middle East, to the Mediterranean

S. Mansoor
Jamil ur Rehman Center for Genome Research
International Center for Chemical and Biological Sciences
University of Karachi
Karachi, Pakistan

F. Vahidi
Department of Genomics
Agricultural Biotechnology Research Institute of Iran-North Branch (ABRII)
Rasht, Iran

S. A. Al-Bayatti
Medical Laboratory Techniques Department
Al-Farabi University College
Ministry of Higher Education and Scientific Research
Baghdad, Iraq

A. Paucillo
Department of Agricultural, Forest and Food Sciences
University of Turin
Grugliasco (TO) 10095, Italy

J. A. Lenstra
Faculty of Veterinary Medicine
Utrecht University
Yalelaan 104, Utrecht CM 3584, The Netherlands

J. S. F. Barker
School of Environmental and Rural Science
University of New England
Armidale, NSW 2351, Australia

L. Fang
The Center for Quantitative Genetics and Genomics (QGG)
Aarhus University
Aarhus 8000, Denmark

J. Han
Yazhouwan National Laboratory
Sanya 572024, China
E-mail: hanjianlin@yzwlab.cn

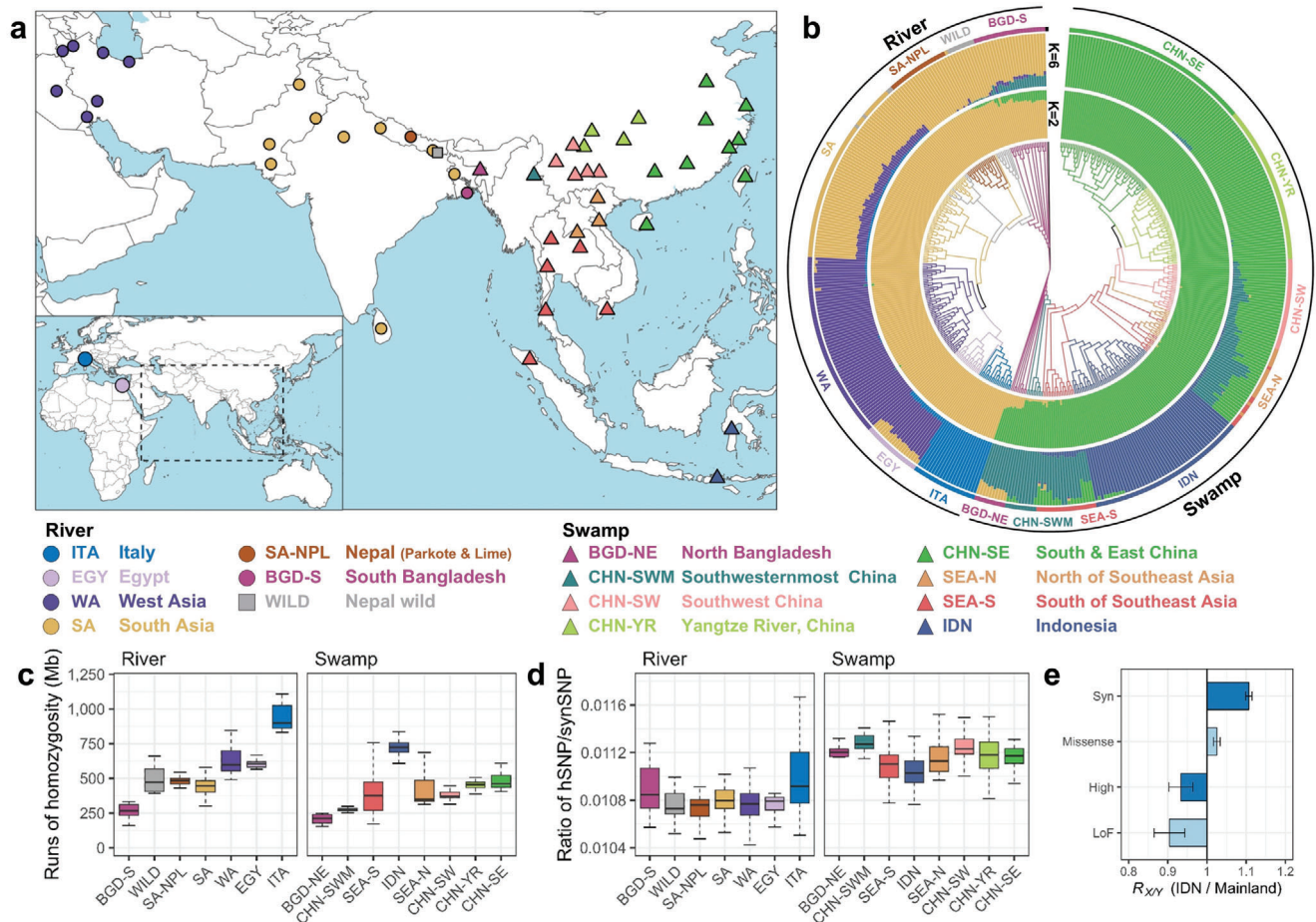


Figure 1. Geographic distribution and population genetic structure of buffaloes used in this study. a) The geographic locations of wild (square), river (circles), and swamp (triangle) buffalo populations. Different colors show the classifications of different geographic groups. The dash-lined box is enlarged to indicate all the populations except the Italian and Egyptian buffaloes. b) Maximum-likelihood phylogenetic tree and admixture patterns at $K = 2$ and 6 . c) The total lengths of homozygosity (ROH, > 100 kb) in different geographic groups. d) The ratios of high-impact SNPs (hSNP)/synonymous (synSNP) in different geographic groups. e) The R_{xy} ratios of derived alleles for synonymous (Syn), missense, high-impact (High), and loss of function (LoF) variants. $R_{xy} < 1$ indicates a relative frequency deficit of the corresponding category in the Indonesian (IND) compared to the mainland buffaloes. Whiskers represent 95% confidence interval.

region, including 301 new genomes generated in this study (Table S1, Supporting Information). The dataset comprised 200 domestic river, 257 domestic swamp, and 13 wild water buffaloes (Figure 1a). The depth coverages range from 5.22 to 47.80 × across the 470 genomes (Table S1, Supporting Information). In addition, we incorporated one publicly available genome of African buffalo (*Syncerus caffer*)^[30] as an outgroup. By mapping these data to the river buffalo reference genome,^[31] we found that the average mapping rates of river and swamp buffaloes reached to as high as 99.068% and 99.512%, respectively, an clear indication of no mapping bias against the reference genome.^[31] We identified 58.2 million single-nucleotide polymorphisms (SNPs), of which 11.7 million and 14.8 million are private (polymorphic in one type but fixed in another type) to the river and swamp buffaloes, respectively (Table S2, Supporting Information). The ratio of transitions to transversions (Ti/Tv) across all SNPs is 2.15, suggesting their high quality.^[32] After a strict quality control based on missing genotype rate and minor allele frequency, 34977918 SNPs are retained for subsequent analyses.

2.2. Population Genetic Structure of the River and Swamp Buffaloes

Population genomic analyses, including the maximum likelihood (ML) phylogeny, principal components analysis (PCA), and model-based clustering using the Admixture program (Figure 1b; Figures S1 and S2, Supporting Information), consistently unveil a deep divergence between the river and swamp buffaloes.^[20,21,28,29,33] The distribution patterns of genomic diversity in individual river and swamp buffalo populations agree with the geographic distances to their independent centers of domestications (Figure 1c,d; Figure S3, Supporting Information).

A NeighborNet visualization of the Fixation index (F_{ST}) genetic distances (Figure S4, Supporting Information) shows that the differentiations between the river or swamp buffalo populations were relatively low compared to the divergence between the river and swamp buffaloes. It also shows the phylogenetic position of wild water buffaloes sampled in Nepal (WILD) as a sister taxon of the river buffaloes. Additionally, it highlights the intermediate

positions of the river and swamp buffaloes from Bangladesh, a spatial pattern of hybrid zone based on their historically overlapped distributions and supported by haplotype inference of karyotype (Figures S5 and S6, Supporting Information).

The Indonesian buffaloes (IDN), distributed in the islands of Sulawesi and Sumbawa located in the southeasternmost range of the swamp buffaloes, comprise a distinct cluster in the PCA (Figure S1, Supporting Information) and carry higher levels of runs of homozygosity (ROH) than those of other swamp buffalo populations (Figure 1c and Table S3, Supporting Information). The MSMC2 analysis shows that the IDN is an isolated population with a long history of strong inbreeding, driving its divergence from other swamp buffalo populations 4779–16243 years ago (Figure S7a, Supporting Information). However, the IDN has relatively low mutational load that is measured by calculating the ratio of high-impact SNPs (hSNPs) to synonymous SNPs (synSNPs) (Figure 1d). The ratios of derived alleles (R_{xy}) of the loss of function SNPs (LoFSNPs) and hSNPs are also lower in the IDN than in other swamp buffalo populations (Figure 1e). This is not expected if genetic isolation leads to the fixation of deleterious alleles (the “cost of isolation”). Furthermore, the IDN has relatively low ratios of LoFSNPs to synSNPs and hSNPs to synSNPs in heterozygous states (Figure S8, Supporting Information), suggesting that the LoFSNPs and hSNPs are primarily deleterious in homozygous states. These results suggest that most deleterious mutations have been purged in the IND, which supports previous observations that inbreeding may enhance purifying selection, leading to the purging of deleterious mutations in small populations.^[34–36]

The WILD clusters with the river buffaloes and is positioned at a basal node in the river buffalo branch (Figure 1b; Figure S4, Supporting Information). The divergence time estimated by the MSMC2 between the WILD and domesticated river buffaloes (4045–8306 years ago; Figure S7b, Supporting Information) is in line with the time frame of domestication of the river buffaloes.^[18] These results support the proposal that the wild water buffalo is the ancestor of domesticated river buffaloes. It is worth noting that two domestic buffalo populations indigenous to Nepal (Lime and Parkote) are close to the WILD in the phylogenetic tree and ancestral composition (Figure 1b). Considering that both the WILD and these specific Nepalese domestic buffaloes (e.g., Lime and Parkote) are more similar to the swamp buffaloes in their physiognomy, we believe that the ancestral morphology of the WILD (e.g., the skin and hair colors as well as horn shape) are inherited by the swamp buffaloes.^[18,29,37]

2.3. Genomic Divergence Between the River and Swamp Buffaloes

To investigate genomic divergence between the two types of domesticated buffaloes, we selected 146 river and 205 swamp buffaloes that contain relatively pure genomic ancestry based on the admixture inference ($K = 2$, ancestral compositions $\geq 99.999\%$; Table S4, Supporting Information). The average F_{ST} value between the two types is 0.456 across 19506 protein-coding genes. The Hudson-Kreitman-Aguadé (HKA) tests, based on the ratio of polymorphic to differentiated sites identified at a significance threshold based on the top 5% of P -values, identify 838 genes un-

der selection in the river buffaloes, 831 genes in the swamp buffaloes, and 496 genes across the two types (Table S5, Supporting Information). Genes under the selection in the river or swamp buffaloes exhibit reduced ratios of polymorphic to fixed differentiated sites. Similarly, genes under the selection across the two types show markedly low ratios of polymorphic to fixed differentiated sites (Figure 2a).

Genes under the selection across the two types have an average F_{ST} value of 0.740, which is higher than the average F_{ST} values for genes under the selection in the river (0.685, one-tailed t -test: $P = 1.21 \times 10^{-13}$) or swamp (0.655, one-tailed t -test: $P = 2.20 \times 10^{-16}$) buffaloes (Figure 2a) or for genes that are not under the selection (0.441) (Figure 2a). For example, *CNOT4* is under the selection across the two types and shows a pronounced F_{ST} peak but exceptionally low level of polymorphism in the river or swamp buffaloes (Figure 2b). In contrast, *TEX9* and *TBX19* are under the selection with elevated F_{ST} peaks but reduced levels of polymorphisms in the river and swamp buffaloes, respectively (Figure 2b).

From the 496 genes under selection in the two types, 124 are significantly enriched in the gene set of testis-specific expression (Figure S9, Supporting Information). Of these genes, G Kinase Anchoring Protein 1 (*GKAP1*) has the highest level of divergence ($F_{ST} = 0.911$) and is under the selection in the two types as shown by a low level of polymorphism and nearly fixed distinct haplotypes (Figure 2c,d). *GKAP1* is specifically and highly expressed in the testis of buffaloes (Figure S10, Supporting Information). A recent study shows that *GKAP1* regulates the apoptosis, number, and motility of germ cells.^[38] These findings indicate that selection pressures in the two types have resulted in a significant divergence in *GKAP1*, supporting the rapid evolution of testis-related genes based on their accelerated fixation rates (Figure S11, Supporting Information).^[39]

2.4. Directional Selection in the River Buffaloes

The river buffaloes have a greater phenotypic diversity among populations than the swamp buffaloes. Specifically, long-term artificial selection has resulted in several well-defined dairy breeds, including three most famous breeds of Murrah, Nili-Ravi, and the Italian Mediterranean buffaloes with distinct morphological features and outstanding milk productivity. This offers an opportunity for us to localize artificial selection signals underlying milk traits by contrasting these three dairy breeds with non-dairy wild and indigenous buffaloes in Nepal based on F_{ST} and θ_{π} values across their genomes. These analyses reveal 69 statistically significant genomic regions (adjusted $P < 0.05$, $F_{ST} > 0.157$, and $\theta_{\pi} > 0.731$), which harbor 122 genes (Figure S12; Tables S6 and S7, Supporting Information). These genes are functionally enriched in mammary gland development, lipid transport, and multiple-compound catabolic processes (Table S8, Supporting Information).

2.5. Selection Signatures in the Casein Genes Associated with Milk Production

The BBU7:32020001-32430000 region is diverged the most between the dairy and non-dairy buffaloes (Figure 3a; Figure S12,

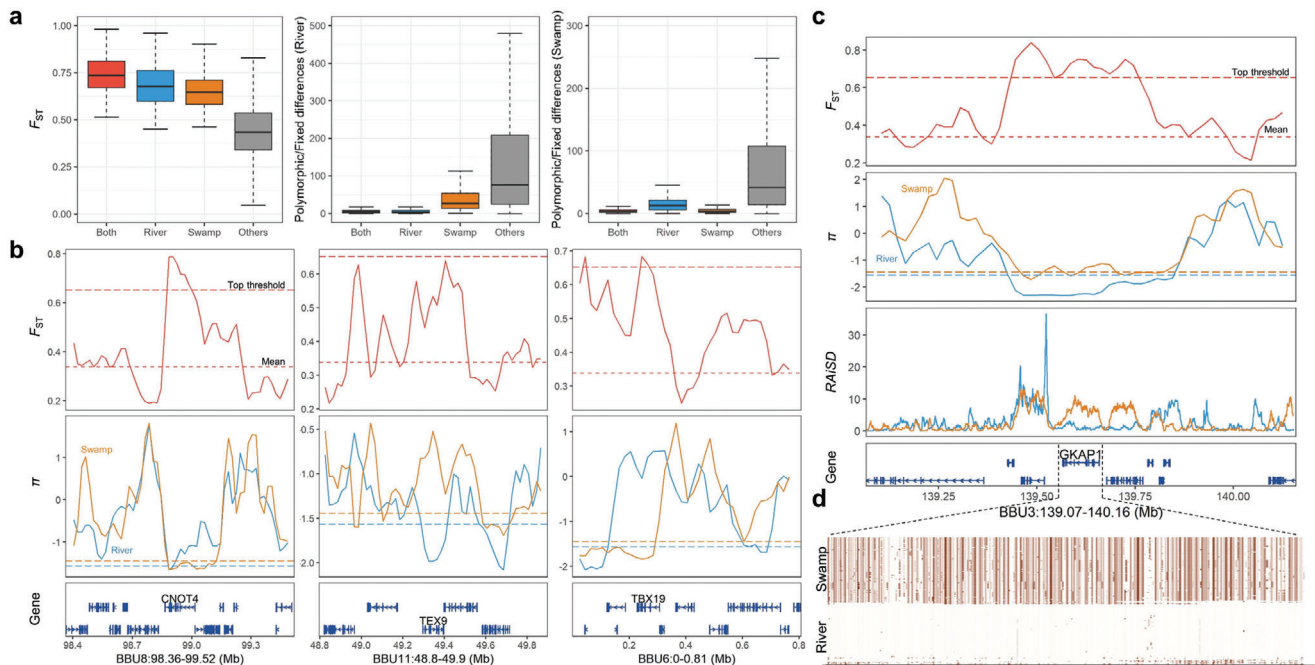


Figure 2. Genomic differentiation between the river and swamp buffaloes. a) Left panel: The F_{ST} estimates of four types of genes: genes under the selection in both the river and swamp (red), the river (blue), and the swamp (orange) buffaloes, and genes free of selection in all buffaloes (gray). The middle and right panels show the ratios of polymorphic and differentiated loci for the four types of genes by comparing the river to swamp and African buffaloes (middle) or the swamp to river and African buffaloes (right). In the middle panel, the polymorphic loci are those exhibiting polymorphisms in the river buffaloes, while the differentiated loci are those fixed in the swamp and African buffaloes. In the right panel, the polymorphic loci are those exhibiting polymorphisms in the swamp buffaloes, while the differentiated loci are those fixed in the swamp and African buffaloes. b) Three candidate genes under the selection in both the river and swamp (left panel, *CNOT4*), the river (middle panel, *TEX9*), and the swamp (right panel, *TBX19*) buffaloes. The F_{ST} , π_{river} , and π_{swamp} approaches are used to scan the differentiation between and the polymorphism in the genomic regions of the river and swamp buffaloes. The top threshold represents the F_{ST} threshold for windows with an FDR-corrected P -value < 0.05 . c) The differentiation and polymorphism measured by the F_{ST} , π_{river} , π_{swamp} , $RAiSD_{river}$, and $RAiSD_{swamp}$ estimates in the *GKAP1* gene region (BBU3:138.94–140.17 Mb), which is under the selection and highly expressed in the testis of the river buffaloes. d) The haplotype patterns of *GKAP1* in the swamp and river buffaloes. Brown and light yellow indicate alternative and reference alleles, respectively.

Supporting Information), in which the dairy buffaloes have a specific haploblock with few polymorphisms, while the non-dairy buffaloes have extensive variations (Figure 3b). There are seven genes in this region, including *CSN1S1*, *CSN2*, *CSN1S2*, *LOC102397372*, *SULT1B1*, *STATH*, and *SULT1E1* (Table S6, Supporting Information). *CSN2*, *CSN1S1*, and *CSN1S2* are members of the casein protein gene family and are highly and specifically expressed in the mammary gland (Figure 3c; Figure S13, Supporting Information). Interestingly, there are nine SNPs in *CSN2* (g.32125,570G>C; g.32125,819G>T; g.32125,945C>T; and g.32127,633G>T), *CSN1S1* (g.32148,856A>G; g.32148,918A>G; and g.32158,292T>C), and *CSN1S2* (g.32020,897T>C and g.32026,256T>C) with significant allele specific expressions (ASE) in the mammary gland (Figure 3d; Figures S14 and S15 and Table S9, Supporting Information), indicating potential regulations by *cis*-acting genetic variants. Notably, the alleles with the highest expression levels are dominant in the dairy relative to non-dairy buffaloes for all of the nine SNPs (Figure 3e; Figure S14, Supporting Information), indicating these casein genes to be highly expressed in the dairy buffaloes following positive selection.

The mammary gland is a complex tissue with diverse cell types. To examine the key cell types underlying the divergence between the dairy and non-dairy buffaloes, we constructed a

single-cell landscape of the mammary gland of the dairy and non-dairy buffaloes over their early lactation period (Figure 3f, Figure S16a, Supporting Information). Three genes of *CSN2*, *CSN1S1*, and *CSN1S2*, being positively selected in dairy buffaloes, are predominantly expressed in the luminal cells of the dairy and non-dairy buffaloes (Figure S16b and c, Supporting Information), suggesting the importance of these cells for milk secretion. Notably, *CSN2* and *CSN1S1* exhibit significantly higher expressions in the luminal cells of the dairy than non-dairy buffaloes (Figure 3g; Figure S16d, Supporting Information). This observation further corroborates the findings from the ASE analysis based on bulk RNA-seq and suggests that the casein genes and luminal cells have undergone positive selection during the development of dairy buffaloes. Further, we found that the luminal cells of the dairy buffaloes have higher CytoTRACE scores (Figure 3h) and are all located at the initial position of the pseudo-time developmental tree (Figure 3i), indicating that the luminal cells of the dairy buffaloes have a higher developmental potential than those of the non-dairy buffaloes. We then clustered the gene expression patterns and observed specific changes in two sets of genes over the pseudo-time, where the casein genes clustered with the genes that are switched at an early stage and decrease their expressions across the pseudo-time (Figure S17, Supporting Information). These genes are enriched

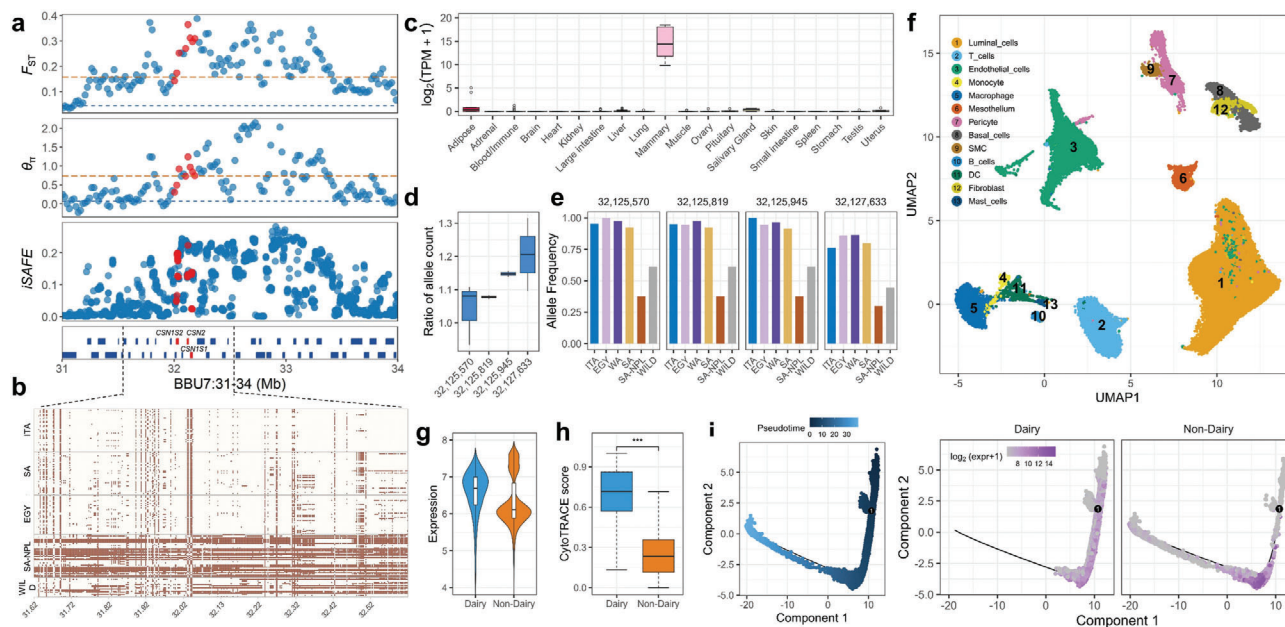


Figure 3. Directional selective sweeps in the casein gene family of the river buffaloes. a) The selected sweeps, gene annotations, and haplotype structures of genomic regions (BBU7:31 000 000–34 000 000) around the casein gene family (*CSN1S2*, *CSN2*, and *CSN1S1*). The selected sweeps are measured by the F_{ST} and θ_{π} ($\pi_{\text{Wild-Indigenous}}/\pi_{\text{Dairy}}$) estimates per 50 kb window and the *iSAFE* per site. The windows and SNPs within genomic regions of the three casein genes are highlighted in red. b) Haplotype structures of the SNPs with maximum F_{ST} values per 2 kb window and with 1 Mb flanking the casein gene family. c) Gene expression atlas of *CSN2*. d) The ratios of allele counts (ratio = derived allele reads / ancestral allele reads) of four allele-specific expression (ASE) loci identified in *CSN2*, where the reference allele is the derived allele. e) The frequencies of reference (derived) alleles of the four ASE loci in the river buffaloes. f) UMAP projection plot of 12 cell types. g) Gene expression of *CSN2* in the luminal cells between the dairy (blue) and non-dairy (orange) buffaloes. h) Predicted ordering by the CytoTRACE in the luminal cells, which orders the luminal cells based on their developmental potentials from the most mature (the lowest values) to the most immature (the highest values). i) Pseudo-time trajectory analysis corresponding to the luminal cells in the dairy and non-dairy buffaloes. Pseudo-time corresponds to the reconstruction of dynamic gene expression programs of the underlying biological process. Cells are colored by pseudo-time (left panel) and gene expression level of *CSN2* (right panel). The distribution of cells by the dairy and non-dairy buffalo is shown in the right panel.

in the pathways related to protein synthesis and localization, ribosomes and translation, and RNA and protein degradation (Figure S17 and Table S10, Supporting Information), indicating their associations with milk secretion. Altogether, these results suggest that the luminal cells in the dairy buffaloes acquired a relatively high differentiation potential following long-term artificial selection, which may enhance the expression of the genes related to milk production of dairy buffaloes.

2.6. A Potential Causative Mutation of *MC1R*

There are 15 genes in the BBU18:14160001–14470000 region under the selection (Figure 4a), including *MC1R* with the highest F_{ST} and *iSAFE* values (Table S11, Supporting Information). This gene encodes the melanocortin receptor 1 to control mammalian coat color by regulating the formation of melanin.^[40,41] Haplotyping of this region shows that the dairy buffaloes are less variable than the Egyptian buffaloes and Nepalese wild and domestic buffaloes (Figure 4b), providing additional evidence for strong selection acting on *MC1R* in the dairy buffaloes. There are two missense mutations at g.14298,614G>A (p.Gly104Ser) and g.14298,688G>T (p.Met128Ile) (Figure 4c). The ancestral alleles of these two mutations are highly conserved among mammals (Figure 4c). Although the frequencies of derived alleles of these two mutations are higher in the river buffaloes with

solid black coat color (ITA and SA), only g.14298,688T has a low frequency in the non-black buffaloes (SA-NPL, EGY, WILD, Surti, and Bhadawari) (Figure 4d). In an additional panel of 115 solid black and 31 non-black buffaloes, g.14298,688G>T is more significantly associated with black coat color than g.14298,614G>A ($P = 2.27 \cdot 10^{-15}$ for g.14298,688G>T and $P = 1.17 \cdot 10^{-9}$ for g.14298,614G>A; Table S12, Supporting Information). These combined results suggest g.14298,688G>T to be the most likely causative mutation for the black coat color in the river buffaloes, while g.14298,614G>A is a partially linked marker.

To elucidate the functional effects of these two mutations on the *MC1R* protein, we performed in vitro assays in HEK 293T cells expressing the wild-type (WT) and derived alleles. The cells expressing the g.14298,688T allele had a higher level of constitutive cAMP accumulation than the cells expressing the WT (g.14298,614G and g.14298,688G) and g.14298,614A ($P < 0.05$) alleles before and after the stimulation by α -MSH (Figure 4e). This result indicates that the gain-of-function mutation of the g.14298,688T allele increases basal *MC1R* activity via cAMP formation and elevated synthesis of eumelanin to confer a solid black coat color in the river buffaloes (Figure 4f).^[40,42] The corresponding p.Gly104Ser replacement of g.14298,614G>A is within the third transmembrane domain where melanism-causing point mutations are also reported for other mammals, including fox,^[43] pig,^[44] and sheep,^[45] illustrating a convergent evolution of melanism across a range of species.^[46–49]

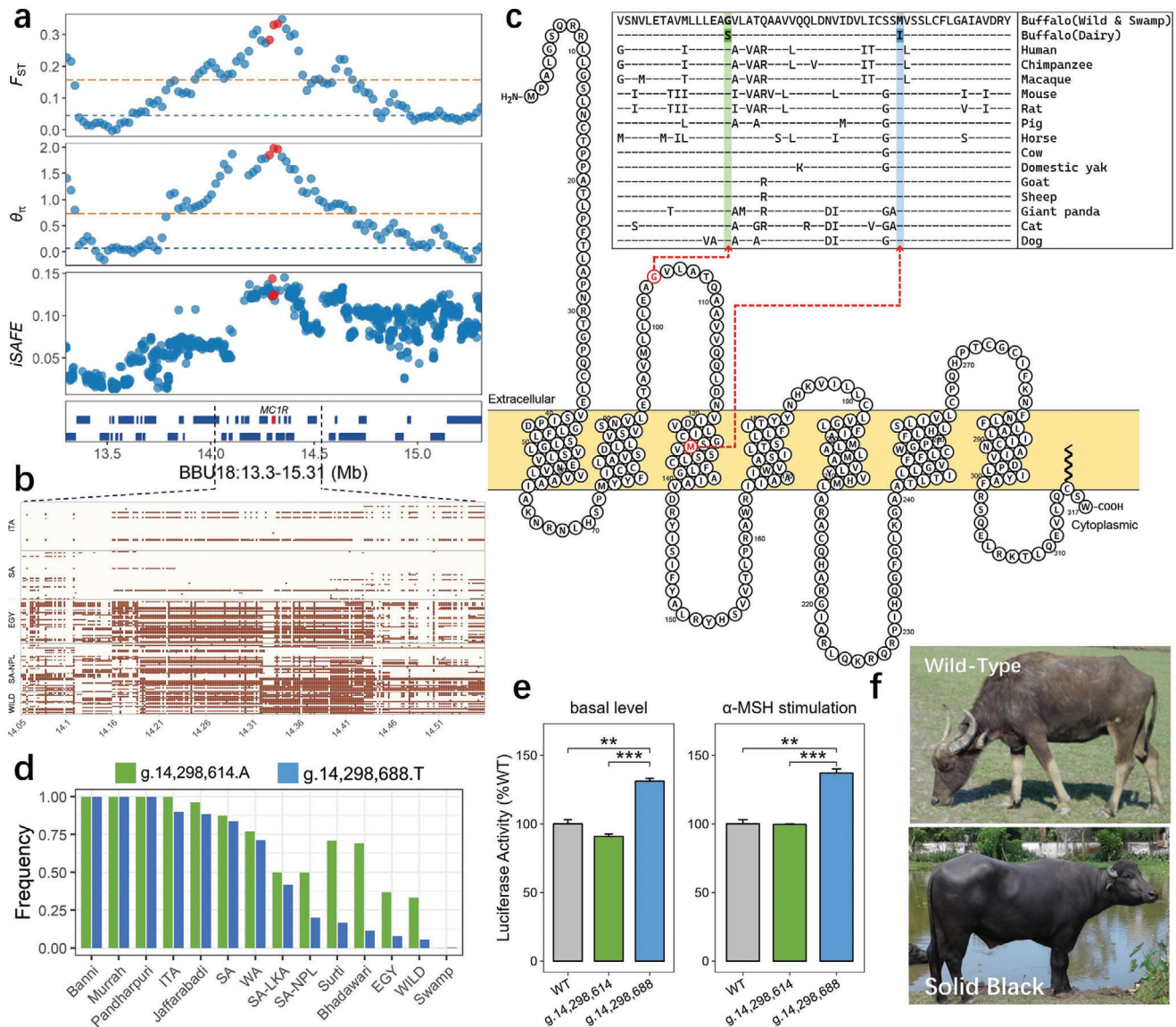


Figure 4. The selection signal, protein structure, and causal mutation of the *MC1R* gene. **a**) The selected sweep, gene annotation, and haplotype structure of genomic region around the *MC1R* gene. The selected sweeps are measured by the F_{ST} and θ_{π} estimates ($\pi_{\text{Non-dairy}}/\pi_{\text{Dairy}}$) per 50 kb window and the $iSAFE$ per site. The windows and SNPs within genomic region of the *MC1R* gene are highlighted in red. **b**) Haplotype structure based on the SNPs with maximum F_{ST} values per 2 kb window and in 0.5 Mb flanking the *MC1R* gene. **c**) Amino acid sequence and transmembrane α -helices of *MC1R*. The two missense variants identified in buffaloes and their homologues across mammalian species are highlighted in the top. The wild-type (WT) alleles are fixed in the swamp buffaloes but segregated in both the wild and river buffaloes. **d**) Frequencies of the two missense variants in eight buffalo populations and six Indian buffalo breeds. In the eight buffalo populations, the ITA, SA, and WA primarily consist of solid-black buffaloes, while the SA-LKA, SA-NPL, WILD, EGY, and Swamp are predominantly non-black (wild-type) buffaloes. Among six Indian breeds, Banni, Jaffarabadi, Murrah, and Pandharpuri have solid-black coat color, while Surti and Bhadawari have white marking at their knees. **e**) Signaling intensities measured by the luciferase activity of the WT and mutants of the *MC1R* gene under base level and α -MSH stimulation (0.1 μM). Intracellular cAMP levels are measured with the luciferase reporter gene assays. Luciferase activities (mean $\pm$ SEM) are expressed as the percentages of individual mutants against the WT. **: $P < 0.01$; ***: $P < 0.001$. **f**) Photos of solid black (bottom) and wild-type (upper) buffaloes.

2.7. Differentiation of Linkage Disequilibrium (LD) in the Regions around *CSN2* and *MC1R*

In the region around *CSN2*, the LD levels are relatively low, but many sites are fixed under the selection (Figure S18a, Supporting Information). This pattern is consistent with the high number of recombination events observed in the casein genes of other dairy animals^[50–52] and recently confirmed in the Mediterranean

buffaloes.^[53] Conversely, high LD levels are observed in the region around *MC1R* (Figure S18b, Supporting Information). Furthermore, the extended haplotypes around *CSN2* are shorter and more heterogeneous than those around *MC1R* (Figure S18, Supporting Information). These contrasting patterns indicate that these two genes may have experienced different selection modes. The H2/H1 statistic, a haplotype-based method proposed to differentiate hard and soft selective sweeps,^[54] is higher around

CSN2 than *MC1R* (Figure S19, Supporting Information), indicating that the former exhibits greater haplotype diversity than the latter. Therefore, we believe that the genomic pattern of the region around *CSN2* is shaped by a gradual, soft sweep, while the pattern around *MC1R* is driven by a fast, hard sweep.

3. Discussion

In this study, we have intensively explored the genomic landscape of Asian water buffaloes of diverse origins and a near-complete coverage of their global range. We present some novel insights into the genetic differentiation and the rapid evolution of reproduction-related genes following divergent selections between the river and swamp buffaloes. We find the selective sweeps related to milk production and coat color genetics of the highly selected dairy and uniquely color-coated river buffaloes. These findings illustrate the impacts of artificial selection on the diversification and specialization of morphology and productivity of Asian water buffaloes.

Our newly sequenced 301 genomes contribute a substantial number of novel genomic variants (e.g., 27% to 74% per genome) to those reported by previous studies.^[19–21] According to the population genetic structuring, we identify several populations with distinct genetic backgrounds that are not captured in previous whole-genome studies, including the wild water buffaloes in Nepal and the domesticated buffaloes in Nepal, Egypt, and Indonesia. Among the river buffaloes, we find some of the Nepalese indigenous buffaloes (e.g., Lime and Parkote) to be closely related to the wild water buffaloes, suggesting these Nepalese buffaloes to be the most-ancient domesticated genetic entity. However, we cannot exclude the possibility of interspecific introgression from the wild water buffaloes distributed in Nepal and from other unknown wild buffalo species. Furthermore, we discover that the Egyptian buffaloes carry typical haplotypes of the casein genes of the dairy buffaloes but their haplotypes of *MC1R* resemble those of the wild water buffaloes, suggesting the artificial selections for milk production and coat color traits to have been undertaken at different times. Based on the migration history of the river buffaloes, we speculate that the selection for milk production traits was likely implemented first due to their economic importance for buffalo stakeholders. The artificial selection and breed improvement of the river buffaloes in general and of those unique breeds/populations in particular require further investigation. Among the swamp buffaloes, populations from southwestern China and the Indochina Peninsula exhibit the highest levels of polymorphisms, supporting the observation of an early study^[29] and thus suggesting this geographic region to be a center for their domestication. However, the genetic introgression from the river buffaloes into some of the swamp buffalo populations complicates the domestication process and migration of the swamp buffaloes. Further in-depth analysis involving the wild water buffaloes from this geographic region is needed. Additionally, we find the strongest evidence of reciprocal gene flow between the two types of buffaloes with a historical sympatry in Bangladesh, providing a valuable opportunity for us to investigate the evolutionary divergence and subsequent genetic exchange between them.

By exploring the genomic differentiations between the dairy and non-dairy buffaloes, we identify a unique selection signature

in the genomic region harboring the casein genes (e.g., *CSN1S1*, *CSN2*, and *CSN1S2*). These genes are found to be significantly associated with the lactation traits in mammals.^[50,55–57] They are located in one casein cluster in dairy cattle and show highly significant associations with both protein and milk yields.^[58] We detect nine SNPs in the three genes, all of them with significant ASE patterns in the mammary gland and two of them to be associated with the milk traits of the Mediterranean buffaloes.^[59] In particular, the mutation of g.32148,856A>G (p.Leu178Ser) in *CSN1S1* is associated with protein percentage,^[59] whereas the variant of g.32026,256T>C (p.Thr190Ala) in *CSN1S2* is associated with high palmitic acid content in buffalo milk.^[60] Furthermore, favorable genetic variants in the casein genes (e.g., *CSN1S1* and *CSN3*) have been identified in buffaloes by independent studies^[19,53,57,61] and they are believed to have significant economic impacts for dairy purposes.^[53,62] As a result, the genotypes at these loci are included in the sire selection of the Mediterranean buffaloes (https://www.risbufala.it/?page_id=58841). Given that the casein genes are tightly clustered^[50] and likely regulated by *cis*-regulatory elements,^[63] further investigation of their regulatory variants will help to elucidate their biological functions and then their application in future genetic improvement of domestic buffaloes.

By constructing the single-cell transcriptomic maps of mammary glands from dairy and non-dairy buffaloes, we discover that the luminal cells are the key cell type, which is also reported in humans.^[64] The three casein genes are mainly expressed in the luminal cells in both the dairy and non-dairy buffaloes; however, the *CSN2* and *CSN1S1* genes show upregulated expressions in the dairy buffaloes. Importantly, the luminal cells in the dairy buffaloes show higher CytoTRACE scores than those in the non-dairy buffaloes. A high CytoTRACE score represents the stemness to be gradually decreased from the cells in the embryonic glands to those in the adult glands.^[65] This indicates that the dairy buffaloes have a relatively high proportion of luminal progenitor cells, which may enhance their milk production capacity throughout the lactation period.^[66,67]

Another notable selection signature in the dairy buffaloes is found around the genomic region of *MC1R*. This gene plays a central role in the regulation of eumelanin (black/brown) and pheomelanin (red/yellow) syntheses in mammalian melanocytes responsible for the coat color switch.^[41] We find two missense variants (e.g., g.14298,614G>A (p.Gly104Ser) and g.14298,688G>T (p.Met128Ile), of which the replacement of p.Met128Ile is validated to be the potential causative mutation because the gain-of-function variants are found to result in constitutive activation of the *MC1R* receptor that upregulates the synthesis of eumelanin and causes solid black color (melanism).^[40,42] Before our study, the true effect of this p.Met128Ile replacement on pigmentation is unclear. A previous study claims that both p.Gly104Ser and p.Met128Ile significantly impact the function of *MC1R* based on an in silico functional analysis, while the river buffalo allele of p.Gly104Ser is associated with the black coat color.^[68] In fact, the p.Gly104Ser is considered to be a natural mutation without impact in humans,^[69] whereas this p.Gly104Ser replacement is species specific in affecting the basal activity of *MC1R* in Sulawesi macaques alone.^[70] Nevertheless, we have to be aware that the amino acid replacements realized in all protein compartments (intra- and extra-cellular loops and transmem-

brane regions) are shown to affect the pigmentation.^[69,71–73] Based on our association analysis and functional validation, we provide strong evidence for the p.Met128Ile replacement to be the causative mutation responsible for the solid black color in the river buffaloes. We propose that the switch from the ancestral coat pattern to the solid black color in dairy buffaloes is due to the artificial selection force acting on the *MC1R* gene.

In conclusion, domesticated Asian water buffaloes have significant social-economic values for developing countries in the tropical and subtropical zones. They are also unique and promising animal genetic resources in the context of global warming. Our findings based on the complete genomic landscape of domestic river and swamp buffaloes illustrate how the natural and artificial selections drive the evolutionary divergence and genetic differentiation in morphology and productivity of the domesticated buffaloes. The genomic resources generated in this study will definitely facilitate their genetic improvement towards sustainability and high efficiency.

4. Experimental Section

Sample Collection: It was sampled 446 buffaloes from 12 countries, including 13 wild water buffaloes, 176 domesticated river buffaloes, and 257 domesticated swamp buffaloes. Of these, 79 river and 66 swamp buffaloes were included in the previous studies.^[74,75] In addition, it was downloaded previously published genomes of 24 domesticated river buffaloes^[76–79] and one African buffalo (*Syncerus caffer*).^[30] The final dataset included 470 Asian water buffaloes and one African buffalo (outgroup) (Table S1, Supporting Information).

Both the river and swamp buffaloes were separated into sub-groups according to their geographical distributions^[29] (Figure 1). The river buffalo populations were divided into six groups: South Bangladesh (BGD-S), South Asia (SA), Nepal (SA-NPL), West Asia (WA), Egypt (EGY), and Italy (ITA). The swamp buffalo populations were separated into eight groups: Northeast Bangladesh (BGD-NE), Southwestern most China (CHN-SWM), Southwest China (CHN-SW), middle and upper reaches of the Yangtze River in China (CHN-YR), Southeast China (CHN-SE), South of Southeast Asia (SEA-S), North of Southeast Asia (SEA-N), and Indonesia (IDN).

Sequencing and Variant Calling: Genomic DNA was extracted from blood hair follicles, or ear skin tissues. For each sample, paired-end libraries with insert size of 350 bp were constructed and sequenced using the Illumina HiSeq X Ten system (Illumina, San Diego, CA, USA). All the sequencing reads were trimmed using the Quality-trim (<https://bitbucket.org/arobinson/qualitytrim/src/master/>) with parameters “-s -a 20 -l 50 -N 3 -z 1” and then mapped to the river buffalo reference genome (UOA_WB_1, GCF_0 031 21 395.1) using the BWA-MEM algorithm^[80] with the default parameters. For the mapped reads, after sorting using the Samtools,^[81] potential PCR duplicates were identified using the “MarkDuplicates” of the Picard v2.9.0 and local realignment around indels was performed by the Genome Analysis Toolkit (GATK, version 3.7).^[82] Then it was called variants using the UnifiedGenotyper of the GATK. Finally, hard filters were applied to the raw SNPs using the GATK program based on previous studies,^[6,7,10,83] with the following parameters: QUAL < 30; QualByDepth (QD) < 2.0; RMSMappingQuality (MQ) < 40.0; MappingQualityRankSumTest (MQRankSum) < -12.5; ReadPosRankSumTest (ReadPosRankSum) < -8.0; and HaplotypeScore > 13.0.

After filtering, the remaining high-quality biallelic SNPs were annotated by the SnpEff v5.0^[84] with the database built based on the GTF files from the river buffalo reference genome (UOA_WB_1, GCF_0 031 21 395.1). The data used in the downstream analyses were filtered in order to remove

SNPs with missing genotype rates > 0.02 and minor allele frequencies < 0.01.

Population Differentiation and Genetic Structure: For PCA and Admixture analysis, LD-based pruning using the PLINK v1.90^[85] with “-indep pairwise 50 10 0.2” option was performed. Next, the eigenvalues and eigenvectors using the Genome-wide Complex Trait Analysis (GCTA) tool v1.93.0 was estimated.^[86] Admixture v1.3.0^[87] was run by increasing the assumed numbers of ancestral components of *K* from 2 to 8 (*K* = 6 was the best model based on the cross-validation). For phylogenetic analysis, we constructed a ML tree using the FastTree v2.1.^[88] The BCFtools/ROH^[89] to estimate the number and length of the ROHs in each population after excluding the ROHs < 100 kb was used.

For measuring the mutational load, the SnpEff to identify putative deleterious variants in three different categories was performed: loss of function variants (LoFSNPs), high-impact variants (hSNPs, including start codon lost, stop codon lost, and stop codon gained, splice acceptor variant, and splice donor variant) and missense variants (misSNPs). It was then calculated the ratio of the number of variants in these four categories to the number of synSNPs per genomes to assess the genetic load of individual populations. We also calculated the relative number of derived alleles R_{xy} , introduced by Xue et al.^[90] and Do et al.,^[91] to estimate the difference in frequency of the variants in all the four categories between the Indonesian and mainland (including SEA-S, SEA-N, CHN-SWM, CHN-SW, CHN-YR, CHN-SE) buffaloes. A random set of 615713 intergenic SNPs for standardization was used, which made the R_{xy} robust against sampling effects and population sub-structuring, and estimated the variance in the R_{xy} by running a Jackknife approach in blocks of 1000 from the set of sites in each category of the variants.

Genomic Divergence Between the River and Swamp Buffaloes: The samples with inferred ancestral compositions $\geq 99.999\%$ based on the Admixture analysis at *K* = 2 were considered as pure ancestry (146 river and 205 swamp buffaloes) and used for genomic differentiation analysis.

For gene-based analysis, the F_{ST} to assess the level of differentiation for each gene was initially employed. It was then utilized the HKA (Hudson-Kreitman-Aguadé) test^[92,93] to independently evaluate the level of selection acting upon each gene in the river and swamp buffaloes. To perform the HKA test, it was calculated four values for the protein-coding genes annotated in the river buffalo genome: A) Number of polymorphic sites in the river buffaloes; B) Number of polymorphic sites in the swamp buffaloes; C) Number of fixed differences between the river buffalo and each of the swamp and African buffaloes (e.g., 0/0 in the river buffaloes and 1/1 in the swamp and African buffaloes); D) Number of fixed differences between the swamp buffaloes and each of the river and African buffaloes (e.g., 0/0 in the swamp buffaloes and 1/1 in the river and African buffaloes).

Then, for instance, the HKA test in the river buffaloes by comparing the ratio of A to C for each gene to the ratio of A to C summed across all genes analyzed ($A/C[gene] = A/C[genome]$) using the Fisher exact test was performed. For the final set of putative positively selected genes, it was retained those with a significant HKA test score ($P < 0.05$) in both the river and swamp buffaloes.

For estimating the genomic differentiation between the river and swamp buffaloes, F_{ST} values were measured using the VCFtools v0.1.17.^[94] The analysis was performed using a sliding window of 50 kb with an increment of 20 kb. Windows with extremely high or low numbers of SNPs ($Mean_{SNP} \pm 3 \times SD_{SNP}$) were excluded. All statistics were Z-transformed and a window with FDR adjusted *p*-value of < 0.05 was regarded as a significantly different region. To further detect the direction of selections, π_{river} and π_{swamp} were measured using the VCFtools with the same sliding windows.

Detection of Positive Selection Signature in the River Buffaloes: To identify positive selection signatures in the river buffaloes selected for milk production, the F_{ST} (target versus reference) and θ_{π} ($\pi_{reference}/\pi_{target}$) values using the VCFtools v0.1.17^[94] in a sliding window of 50 kb length with an increment of 20 kb was calculated, where the target and reference populations referred to intensively selected dairy buffaloes (*n* = 40, from South Asia and Italy) and un-selected (*n* = 29, wild water buffaloes and indigenous buffaloes in Nepal) buffaloes, respectively. Windows with extremely high or low numbers of SNPs ($Mean_{SNP} \pm 3 \times SD_{SNP}$) were

excluded. The mean F_{ST} and θ_{π} values were Z-transformed to calculate the P -value in each window. A window with FDR adjusted P -values of < 0.05 in both statistics was regarded as a significant selection signal. Genes within these significant genomic regions were annotated using the ANNOVAR.^[95] Gene Ontology (GO) enrichment was retrieved using the clusterProfiler.^[96]

The iSAFE (integrated Selection of Allele Favored by Evolution)^[97] to pinpoint the top signals and identify the key SNPs within the top region was used. The Haploview^[98] to perform LD block analysis and the REHH^[99] to calculate the extended haplotype homozygosity was used. It was applied two haplotype-based statistics, the H12 statistic and haplotype homozygosity statistic ($H2/H1$)^[54] to differentiate hard from soft selective sweeps within the top region. These methods hypothesize that the hard sweeps result in the fixation of a single haplotype close to the target of selection, whereas the soft sweeps lead to multiple common haplotypes.^[100]

Construction of Gene Expression Atlas and Quantification of Allele Specific Expression: Gene expression in multiple tissues and lists of tissue-specific expression genes were obtained from a multi-tissue gene expression atlas of water buffalo.^[101] The same RNA-seq data generated above were further genotyped using the GATK v4.1.5.0^[82] following the pipeline described in Liu et al. (2020).^[102] Low quality SNPs were filtered using the command “-filter-expression FS > 30.0, and QD < 2.0”. The ASE analysis was conducted using the ASEReadCounter with the commands “-U ALLOW_N_CIGAR_READS -minDepth 10 -minMappingQuality 255 -minBaseQuality 10”. SNPs were further filtered by the following criteria: at least 30 reads cover a SNP, at least 5 reads per allele, and at least 1% of all reads supporting a minor allele. A binomial P -value was calculated by comparing to the expected ratio under the null hypothesis and multiple-test correction was done with the Benjamini–Hochberg (BH) approach. SNPs with FDR < 0.05 were considered as significant ASE.^[103]

Single-Cell Transcriptome Sequencing and Data Analysis: The mammary gland samples were collected from four adult female buffaloes (two river and two swamp buffaloes) at early lactation. The DNBelab C Series Single-Cell Library Prep Set (MGI) was used for single-cell RNA-seq library preparation. The sequencing libraries were quantified by the Qubit ssDNA Assay Kit (Thermo Fisher Scientific) and sequenced by the DNBSEQ T7. The DNBC4tools and Cell Ranger (10x Genomics) were used to analyze the raw snRNA-seq dataset with default parameters and to generate the gene-cell matrices. The Seurat package^[104] was applied for further analyses. Genes expressed in fewer than three cells and cells with featured RNA number lower than 200 were discarded. Then, the “NormalizeData” (Log-Normalize method, scale factor 10 000) function was performed to normalize the data. The “FindVariableFeatures” (vst method, 2000 features) function was performed to detect variable genes. The “RunPCA” (30 principal components) function was applied for reducing dimensions. The “FindNeighbors” (dims set 1:25) and “FindClusters” (resolution 0.6) functions were used to construct the SNN graph and to cluster the cells. The “RunTSNE” and “RunUMAP” functions were performed to visualize the result with nonlinear dimensional reduction algorithms. The clusters were classified and annotated according to the known functions and expression patterns of the genes enriched in each cluster.^[105–107] To construct a trajectory of the luminal cells, pseudo-time analysis was performed using the Monocle2.^[108] The marker genes with adjusted P -value lower than 0.05 were selected as the developmental progress specific genes. The “reduceDimension” (DDR Tree method, max_components = 2) function was used to reduce the dimensionality of the data into two components. The “orderCells” function was used to order the cells in pseudo-time according to the transcriptome correlation. The CytoTRACE package^[109] was used to determine developmental potential, which was further employed to determine the starting point of a differentiation trajectory.

Identification and Functional Validation of Causative Mutations in MC1R: A genetic association test with the PLINK v1.90^[85] using 115 solid black and 31 non-black buffaloes was conducted. It was further integrated 62 genomes of six river buffalo breeds (Banni, Bhadawari, Jaffarabadi, Murrah, Pandharpuri, and Surti) from a previous study^[19] to analyze the frequency spectrum of two missense mutations (g.14298,614G>A and g.14298,688G>T) in these buffalo populations.

Finally, it was performed functional validation of the two mutations at the cellular level. Levels of intracellular cAMP were quantified using the Dual-luciferase Reporter assay system (Promega). Wild type (WT) and the two mutants in the buffalo *MC1R* (*bMC1R*) cDNA were ligated into the pcDNA3.1(+) vector. The forward amplification primer included a Kozak (GCCACC) sequence 5' to the start codon. The HEK293T cells were co-transfected with 0.1 μ g *bMC1R* plasmid, pGL4.29[luc2P/CRE/Hygro], and pRL-TK vector at a ratio of 2:10:1 using the Lipofectamine 2000 (Invitrogen). After 4 h, the transfection mixture was replaced with growth medium. After incubation over 24 h, the basal luciferase activity of *bMC1R* was measured according to manufacturer's instructions without stimulation of the alpha-melanocyte stimulating hormone (α MSH). Afterwards, the cells were stimulated with 1 nM α MSH for 6 h at 37 °C and the luciferase was measured again. The experiment was repeated in triplicate and the mean and the standard error of mean (SEM) were calculated.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This study was supported by the National Key Research and Development Program of China (2021YFD1200904), National Natural Scientific Foundation of China (31561143010), Yunnan Fundamental Research Projects (202301AW070012, 202401AV070007, and 202305AH340006), the earmarked fund for CARS36, and the High-performance Computing Platform of China Agricultural University. The authors were grateful to the participants and all staff members who supported this study and the anonymous reviewers for their critical review of our manuscript.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

J.S., D.D., and N.A.G. contributed equally to this work. D.W., J.H. and Y.Z. conceived the study. N.A.G., S.S., R.C.K., A.S., A.D., M.O.F., A.I.O., E.M.S., H.A., M.I.A.D., M.Y., H.E.R., H.E., A.A., M.A.R., L.D.P., W.M.M.P.H., G.L.L.P.S., W.Z., S.M., M.B.A., F.V. and S.A.A. prepared the samples. J.S. performed analyses with suggestions from M.W., S.W. and L.F. D.D. performed experiments. J.S. and Y.Z. drafted the manuscript with contributions from M.W., A.P., J.A.L., J.S.F.B., L.F., H.E.R., D.W. and J.H. All authors critically revised the manuscript for important intellectual content and approved the final version.

Data Availability Statement

All data are available in the main text or the supplementary materials. The newly generated sequence data are available from the Sequence Read Archive (SRA) with the Bioproject accession numbers of PRJNA1057008, PRJNA267486, PRJNA633724, and PRJNA633919. The publicly available sequences were downloaded from the SRA following project accession numbers of PRJNA207334 (Italian Mediterranean), PRJNA350833 (Italian Mediterranean), PRJNA267486 (Egyptian), PRJNA40113 (Jafarabadi), PRJNA33659 (Murrah), and PRJNA341313 (African buffalo).

Keywords

buffalo, *CSN2*, *MC1R*, population genomics, selection, single-cell RNA-seq

Received: July 6, 2024
Revised: October 18, 2024
Published online: December 4, 2024

- [1] M. D. Purugganan, D. Q. Fuller, *Nature* **2009**, 457, 843.
- [2] G. Larson, D. Q. Fuller, *Annu. Rev. Ecol. Evol. Syst.* **2014**, 45, 115.
- [3] C. J. Rubin, M. C. Zody, J. Eriksson, J. R. Meadows, E. Sherwood, M. T. Webster, L. Jiang, M. Ingman, T. Sharpe, S. Ka, F. Hallbook, F. Besnier, O. Carlborg, B. Bed'hom, M. Tixier-Boichard, P. Jensen, P. Siegel, K. Lindblad-Toh, L. Andersson, *Nature* **2010**, 464, 587.
- [4] E. Axelsson, A. Ratnakumar, M. L. Arendt, K. Maqbool, M. T. Webster, M. Perloski, O. Liberg, J. M. Arnemo, A. Hedhammar, K. Lindblad-Toh, *Nature* **2013**, 495, 360.
- [5] L. A. Frantz, J. G. Schraiber, O. Madsen, H. J. Megens, A. Cagan, M. Bosse, Y. Paudel, R. P. Crooijmans, G. Larson, M. A. Groenen, *Nat. Genet.* **2015**, 47, 1141.
- [6] D. D. Wu, X. D. Ding, S. Wang, J. M. Wojcik, Y. Zhang, M. Tokarska, Y. Li, M. S. Wang, O. Faruque, R. Nielsen, Q. Zhang, Y. P. Zhang, *Nat. Ecol. Evol.* **2018**, 2, 1139.
- [7] K. Kim, T. Kwon, T. Dessie, D. Yoo, O. A. Mwai, J. Jang, S. Sung, S. Lee, B. Salim, J. Jung, H. Jeong, G. M. Tarekegn, A. Tijjani, D. Lim, S. Cho, S. J. Oh, H. K. Lee, J. Kim, C. Jeong, S. Kemp, O. Hanotte, H. Kim, *Nat. Genet.* **2020**, 52, 1099.
- [8] Y. T. Utsunomiya, M. Milanese, M. R. S. Fortes, L. R. Porto-Neto, A. T. H. Utsunomiya, M. Silva, J. F. Garcia, P. Ajmone-Marsan, *Anim. Genet.* **2019**, 50, 557.
- [9] Q. Qiu, L. Wang, K. Wang, Y. Yang, T. Ma, Z. Wang, X. Zhang, Z. Ni, F. Hou, R. Long, R. Abbott, J. Lenstra, J. Liu, *Nat. Commun.* **2015**, 6, 10283.
- [10] Z. Zheng, X. Wang, M. Li, Y. Li, Z. Yang, X. Wang, X. Pan, M. Gong, Y. Zhang, Y. Guo, Y. Wang, J. Liu, Y. Cai, Q. Chen, M. Okpeku, L. Colli, D. Cai, K. Wang, S. Huang, T. S. Sonstegard, A. Esmailzadeh, W. Zhang, T. Zhang, Y. Xu, N. Xu, Y. Yang, J. Han, L. Chen, J. Lesur, K. G. Daly, et al., *Sci. Adv.* **2020**, 6, eaaz5216.
- [11] X. Li, J. Yang, M. Shen, X. L. Xie, G. J. Liu, Y. X. Xu, F. H. Lv, H. Yang, Y. L. Yang, C. B. Liu, P. C. Wan, Y. S. Zhang, L. Gao, J. Q. Yang, W. H. Pi, Y. L. Ren, Z. Q. Shen, F. Wang, J. Deng, S. S. Xu, H. Salehian-Dehkordi, E. Hehua, A. Esmailzadeh, M. Dehghani-Qanatqestani, O. Stepanek, C. Weimann, G. Erhardt, A. Amane, J. M. Mwacharo, et al., *Nat. Commun.* **2020**, 11, 2815.
- [12] A. O. Rubio, K. Summers, *PLoS One* **2022**, 17, e0263830.
- [13] F. J. Alberto, F. Boyer, P. Orozco-terWengel, I. Streeter, B. Servin, P. de Villemereuil, B. Benjelloun, P. Librado, F. Biscarini, L. Colli, M. Barbato, W. Zamani, A. Alberti, S. Engelen, A. Stella, S. Joost, P. Ajmone-Marsan, R. Negrini, L. Orlando, H. R. Rezaei, S. Naderi, L. Clarke, P. Flicek, P. Wincker, E. Coissac, J. Kijas, G. Tosser-Klopp, A. Chikhi, M. W. Bruford, P. Taberlet, et al., *Nat. Commun.* **2018**, 9, 813.
- [14] K. E. Witt, E. Huerta-Sanchez, *Philos. Trans. R Soc. Lond. B Biol. Sci.* **2019**, 374, 20180235.
- [15] D. D. Wu, C. P. Yang, M. S. Wang, K. Z. Dong, D. W. Yan, Z. Q. Hao, S. Q. Fan, S. Z. Chu, Q. S. Shen, L. P. Jiang, Y. Li, L. Zeng, H. Q. Liu, H. B. Xie, Y. F. Ma, X. Y. Kong, S. L. Yang, X. X. Dong, A. Esmailzadeh, D. M. Irwin, X. Xiao, M. Li, Y. Dong, W. Wang, P. Shi, H. P. Li, Y. H. Ma, X. Gou, Y. B. Chen, Y. P. Zhang, *Natl. Sci. Rev.* **2020**, 7, 952.
- [16] M. Cieslak, M. Reissmann, M. Hofreiter, A. Ludwig, *Biol. Rev. Camb. Philos. Soc.* **2011**, 86, 885.
- [17] F. W. Nicholas, *Anim. Genet.* **2021**, 52, 3.
- [18] Y. Zhang, L. Colli, J. S. F. Barker, *Anim. Genet.* **2020**, 51, 177.
- [19] P. Dutta, A. Talenti, R. Young, S. Jayaraman, R. Callaby, S. K. Jadhav, V. Dhanikachalam, M. Manikandan, B. B. Biswa, W. Y. Low, J. L. Williams, E. Cook, P. Toye, E. Wall, A. Djikeng, K. Marshall, A. L. Archibald, S. Gokhale, S. Kumar, D. A. Hume, J. G. D. Prendergast, *Nat. Commun.* **2020**, 11, 4739.
- [20] T. Sun, J. Shen, A. Achilli, N. Chen, Q. Chen, R. Dang, Z. Zheng, H. Zhang, X. Zhang, S. Wang, T. Zhang, H. Lu, Y. Ma, Y. Jia, M. R. Capodiferro, Y. Huang, X. Lan, H. Chen, Y. Jiang, C. Lei, *GigaScience* **2020**, 9, giz166.
- [21] X. Luo, Y. Zhou, B. Zhang, Y. Zhang, X. Wang, T. Feng, Z. Li, K. Cui, Z. Wang, C. Luo, H. Li, Y. Deng, F. Lu, J. Han, Y. Miao, H. Mao, X. Yi, C. Ai, S. Wu, A. Li, Z. Wu, Z. Zhuo, D. D. Giang, B. Mitra, M. F. Vahidi, S. Mansoor, S. A. Al-Bayatti, E. M. Sari, N. A. Gorkhali, S. Prastowo, et al., *Natl. Sci. Rev.* **2020**, 7, 686.
- [22] R. Macgregor, *Vet. Rec.* **1941**, 53, 443.
- [23] F. Ulbrich, H. Fischer, *Z. Tierz. Zuchtungsbiol.* **1966**, 83, 219.
- [24] L. Iannuzzi, W. A. King, D. Di Berardino, *Cytogenet. Genome Res.* **2009**, 126, 49.
- [25] A. Pauciullo, C. Versace, A. Perucatti, G. Gaspa, L.-Y. Li, C.-Y. Yang, H.-Y. Zheng, Q. Liu, J.-H. Shang, *Sci. Rep.* **2022**, 12, 8440.
- [26] M. Curaudeau, R. Rozzi, A. Hassanin, *Mol. Phylogenet. Evol.* **2021**, 161, 107170.
- [27] M. Nagarajan, K. Nimisha, S. Kumar, *Genome Biol. Evol.* **2015**, 7, 1252.
- [28] S. Wang, N. Chen, M. R. Capodiferro, T. Zhang, H. Lancioni, H. Zhang, Y. Miao, V. Chanthakhoun, M. Wanapat, M. Yindee, Y. Zhang, H. Lu, L. Caporali, R. Dang, Y. Huang, X. Lan, M. Plath, H. Chen, J. A. Lenstra, A. Achilli, C. Lei, *Sci. Rep.* **2017**, 7, 4708.
- [29] Y. Zhang, Y. Lu, M. Yindee, K. Y. Li, H. Y. Kuo, Y. T. Ju, S. Ye, M. O. Faruque, Q. Li, Y. Wang, V. C. Cuong, L. D. Pham, B. Bouahom, B. Yang, X. Liang, Z. Cai, D. Vankan, W. Manatchaiworakul, N. Kowlim, S. Duangchantrasiri, W. Wajjwalku, B. Colenbrander, Y. Zhang, P. Beerli, J. A. Lenstra, J. S. Barker, *Mol. Ecol.* **2016**, 25, 1530.
- [30] B. Glanzmann, M. Möller, N. le Roex, G. Tromp, E. G. Hoal, P. D. van Helden, *BMC Genomics* **2016**, 17, 1001.
- [31] W. Y. Low, R. Tearle, D. M. Bickhart, B. D. Rosen, S. B. Kingan, T. Swale, F. Thibaud-Nissen, T. D. Murphy, R. Young, L. Lefevre, D. A. Hume, A. Collins, P. Ajmone-Marsan, T. P. L. Smith, J. L. Williams, *Nat. Commun.* **2019**, 10, 260.
- [32] M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. Del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernysky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, *Nat. Genet.* **2011**, 43, 491.
- [33] L. Colli, M. Milanese, E. Vajana, D. Iamartino, L. Bomba, F. Puglisi, M. Del Corvo, E. L. Nicolazzi, S. S. E. Ahmed, J. R. V. Herrera, L. Cruz, S. Zhang, A. Liang, G. Hua, L. Yang, X. Hao, F. Zuo, S. J. Lai, S. Wang, R. Liu, Y. Gong, M. Mokher, Y. Mao, F. Guan, A. Vlaic, B. Vlaic, L. Ramunno, G. Cosenza, A. Ahmad, I. Soysal, et al., *Front. Genet.* **2018**, 9, 53.
- [34] A. Caballero, I. Bravo, J. Wang, *Heredity* **2017**, 118, 177.
- [35] N. Dussex, T. van der Valk, H. E. Morales, C. W. Wheat, D. Diez-Del-Molino, J. von Seth, Y. Foster, V. E. Kutschera, K. Guschanski, A. Rhie, A. M. Phillippy, J. Korlach, K. Howe, W. Chow, S. Pelan, J. D. Mendes Damas, H. A. Lewin, A. R. Hastie, G. Formenti, O. Fedrigo, J. Guhlin, T. W. R. Harrop, M. F. Le Lec, P. K. Dearden, L. Haggerty, F. J. Martin, V. Kodali, F. Thibaud-Nissen, D. Iorns, M. Knapp, et al., *Cell Genom.* **2021**, 1, 100002.
- [36] D. Wang, H. Salehian-Dehkordi, L. Suo, F. Lv, *Genes* **2023**, 14, 1977.
- [37] Y. Zhang, D. Vankan, Y. Zhang, J. S. F. Barker, *Anim. Genet.* **2011**, 42, 366.
- [38] H. S. Wang, G. S. Wang, Y. B. Dai, Z. H. Li, Y. Zhu, F. Sun, *Mol. Reprod. Dev.* **2019**, 86, 1199.
- [39] F. Murat, N. Mbengue, S. B. Winge, T. Trefzer, E. Leushkin, M. Sepp, M. Cardoso-Moreira, J. Schmidt, C. Schneider, K. Mossinger, T. Bruning, F. Lamanna, M. R. Belles, C. Conrad, I. Kondova, R. Bontrop, R. Behr, P. Khaitovich, S. Paabo, T. Marques-Bonet, F.

- Grutzner, K. Almstrup, M. H. Schierup, H. Kaessmann, *Nature* **2023**, 613, 308.
- [40] M. E. Majerus, N. I. Mundy, *Trends Genet.* **2003**, 19, 585.
- [41] E. M. Wolf Horrell, M. C. Boulanger, J. A. D'Orazio, *Front. Genet.* **2016**, 7, 95.
- [42] E. Eizirik, N. Yuhki, W. E. Johnson, M. Menotti-Raymond, S. S. Hannah, S. J. O'Brien, *Curr. Biol.* **2003**, 13, 448.
- [43] D. I. Vage, D. Lu, H. Klungland, S. Lien, S. Adalsteinsson, R. D. Cone, *Nat. Genet.* **1997**, 15, 311.
- [44] J. M. Kijas, R. Wales, A. Tornsten, P. Chardon, M. Moller, L. Andersson, *Genetics* **1998**, 150, 1177.
- [45] D. I. Vage, H. Klungland, D. Lu, R. D. Cone, *Mamm. Genome* **1999**, 10, 39.
- [46] N. I. Mundy, *PLoS Genet.* **2009**, 5, e1000388.
- [47] E. B. Rosenblum, H. Rompler, T. Schoneberg, H. E. Hoekstra, *Proc. Natl. Acad. Sci. USA* **2010**, 107, 2113.
- [48] A. Schneider, C. Henegar, K. Day, D. Absher, C. Napolitano, L. Silveira, V. A. David, S. J. O'Brien, M. Menotti-Raymond, G. S. Barsh, E. Eizirik, *PLoS Genet.* **2015**, 11, e1004892.
- [49] H. R. McRobie, N. D. Moncrief, N. I. Mundy, *BMC Evol. Biol.* **2019**, 19, 140.
- [50] A. M. Caroli, S. Chessa, G. J. Erhardt, *J. Dairy Sci.* **2009**, 92, 5335.
- [51] G. Cosenza, A. Paucullo, D. Gallo, L. Colimoro, A. D'Avino, A. Mancusi, L. Ramunno, *Small Ruminant Res.* **2008**, 74, 84.
- [52] A. Paucullo, E. T. Shuiep, M. D. Ogah, G. Cosenza, L. Di Stasio, G. Erhardt, *Front. Genet.* **2019**, 10, 748.
- [53] A. Paucullo, S. Martorello, K. Carku, C. Versace, A. Coletta, G. Cosenza, *Ital. J. Anim. Sci.* **2021**, 20, 1264.
- [54] N. R. Garud, P. W. Messer, E. O. Buzbas, D. A. Petrov, *PLoS Genet.* **2015**, 11, e1005004.
- [55] J. C. Jiang, J. B. Cole, E. Freebern, Y. Da, P. M. VanRaden, L. Ma, *Commun. Biol.* **2019**, 2, 212.
- [56] V. Bonfatti, G. Di Martino, A. Cecchinato, D. Vicario, P. Carnier, *J. Dairy Sci.* **2010**, 93, 3797.
- [57] V. Bonfatti, M. Giantin, M. Gervaso, A. Coletta, M. Dacasto, P. Carnier, *J. Dairy Sci.* **2012**, 95, 3435.
- [58] H. Nilsen, H. G. Olsen, B. Hayes, E. Sehested, M. Svendsen, T. Nome, T. Meuwissen, S. Lien, *Genet. Sel. Evol.* **2009**, 41, 24.
- [59] G. Cosenza, A. Paucullo, N. P. P. Macciotta, E. Apicella, R. Steri, A. La Battaglia, L. Jemma, A. Coletta, D. Di Berardino, L. Ramunno, *Anim. Prod. Sci.* **2015**, 55, 654.
- [60] G. Cosenza, D. Gallo, B. Auzino, G. Gaspa, A. Paucullo, *Front. Genet.* **2021**, 11, 622494.
- [61] L. Zicarelli, R. Di Palo, R. Napolano, H. Tonhati, E. De Carlo, R. Gagliardi, A. Di Luccia, B. la Gatta, *Int. Dairy J.* **2020**, 100, 104559.
- [62] A. Paucullo, G. Gaspa, Y. Zhang, Q. Liu, G. Cosenza, *Animals* **2024**, 14, 1414.
- [63] H. K. Lee, M. Willi, C. Liu, L. Hennighausen, *Nat. Commun.* **2023**, 14, 4992.
- [64] P. Bhat-Nakshatri, H. Gao, L. Sheng, P. C. McGuire, X. Xuei, J. Wan, Y. Liu, S. K. Althouse, A. Colter, G. Sandusky, A. M. Stornio, H. Nakshatri, *Cell Rep. Med.* **2021**, 2, 100219.
- [65] K. Saeki, G. Chang, N. Kanaya, X. Wu, J. Wang, L. Bernal, D. Ha, S. L. Neuhausen, S. Chen, *Commun. Biol.* **2021**, 4, 660.
- [66] M. S. Holland, R. E. Holland, *J. Dairy Sci.* **2005**, 88, E1.
- [67] K. Bach, S. Pensa, M. Grzelak, J. Hadfield, D. J. Adams, J. C. Marioni, W. T. Khaled, *Nat. Commun.* **2017**, 8, 2128.
- [68] Y. Miao, G. Wu, L. Wang, D. Li, S. Tang, J. Liang, H. Mao, H. Luo, Y. Zhang, *Sci. China-Life Sci.* **2010**, 53, 267.
- [69] J. C. García-Borrón, B. L. Sánchez-Laorden, C. Jiménez-Cervantes, *Pigm. Cell Res.* **2005**, 18, 393.
- [70] X. Yan, Y. Terai, K. A. Widayati, A. Itoigawa, L. H. P. S. Purba, F. Fahri, B. Suryobroto, H. Imai, *Sci. Rep.* **2022**, 12, 7593.
- [71] V. Chhajlani, X. Xu, J. Blauw, S. Sudarshi, *Biochem. Biophys. Res. Commun.* **1996**, 219, 521.
- [72] P.-A. Frändberg, M. Doufexis, S. Kapas, V. Chhajlani, *Biochem. Biophys. Res. Commun.* **2001**, 281, 851.
- [73] B. Holst, T. W. Schwartz, *Ann. N. Y. Acad. Sci.* **2003**, 994, 1.
- [74] M. Rafiepour, E. Ebrahimie, M. F. Vahidi, G. H. Salekdeh, A. Niazi, M. Dadpasand, D. Liang, J. Si, X. Ding, J. Han, Y. Zhang, S. Qanbari, *Genome Biol. Evol.* **2021**, 13, evaa231.
- [75] D. Liang, P. Zhao, J. Si, L. Fang, E. Pairo-Castineira, X. Hu, Q. Xu, Y. Hou, Y. Gong, Z. Liang, B. Tian, H. Mao, M. Yindee, M. O. Faruque, S. Kongvongxay, S. Khamphoumee, G. E. Liu, D.-D. Wu, J. S. F. Barker, J. Han, Y. Zhang, M. Wilson, *Mol. Biol. Evol.* **2021**, 38, 1122.
- [76] L. K. Whitacre, J. L. Hoff, R. D. Schnabel, S. Albarella, F. Ciotola, V. Peretti, F. Strozzi, C. Ferrandi, L. Ramunno, T. S. Sonstegard, J. L. Williams, J. F. Taylor, J. E. Decker, *Sci. Rep.* **2017**, 7, 39719.
- [77] J. L. Williams, D. Iamartino, K. D. Pruitt, T. Sonstegard, T. P. L. Smith, W. Y. Low, T. Biagini, L. Bomba, S. Capomaccio, B. Castiglioni, A. Coletta, F. Corrado, F. Ferré, L. Iannuzzi, C. Lawley, N. Macciotta, M. McClure, G. Mancini, D. Matassino, R. Mazza, M. Milanese, B. Moiola, N. Morandi, L. Ramunno, V. Peretti, F. Pilla, P. Ramelli, S. Schroeder, F. Strozzi, F. Thibaud-Nissen, et al., *GigaScience* **2017**, 6, gix088.
- [78] D. A. El-Khishin, A. Ageez, M. E. Saad, A. Ibrahim, M. Shokrof, L. R. Hassan, M. I. Abouelhoda, *PLoS One* **2020**, 15, e0237087.
- [79] M. Tania, R. Vijh, V. Bhasin, P. Sikka, P. Vij, R. Kataria, B. Mishra, S. Yadav, A. Pandey, R. Sethi, *Indian J. Anim. Sci.* **2011**, 81, 38.
- [80] H. Li, Preprint at <https://arxiv.org/abs/1303.3997>, **2013**.
- [81] H. Li, *Bioinformatics* **2011**, 27, 2987.
- [82] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernysky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, *Genome Res.* **2010**, 20, 1297.
- [83] N. Chen, Y. Cai, Q. Chen, R. Li, K. Wang, Y. Huang, S. Hu, S. Huang, H. Zhang, Z. Zheng, W. Song, Z. Ma, Y. Ma, R. Dang, Z. Zhang, L. Xu, Y. Jia, S. Liu, X. Yue, W. Deng, X. Zhang, Z. Sun, X. Lan, J. Han, H. Chen, D. G. Bradley, Y. Jiang, C. Lei, *Nat. Commun.* **2018**, 9, 2337.
- [84] P. Cingolani, A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, D. M. Ruden, *Fly* **2012**, 6, 80.
- [85] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, P. C. Sham, *Am. J. Hum. Genet.* **2007**, 81, 559.
- [86] J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, *Am. J. Hum. Genet.* **2011**, 88, 76.
- [87] D. H. Alexander, J. Novembre, K. Lange, *Genome Res.* **2009**, 19, 1655.
- [88] M. N. Price, P. S. Dehal, A. P. Arkin, *PLoS One* **2010**, 5, e9490.
- [89] V. Narasimhan, P. Danecek, A. Scally, Y. Xue, C. Tyler-Smith, R. Durbin, *Bioinformatics* **2016**, 32, 1749.
- [90] Y. Xue, J. Prado-Martinez, P. H. Sudmant, V. Narasimhan, Q. Ayub, M. Szpak, P. Frandsen, Y. Chen, B. Yngvadottir, D. N. Cooper, M. de Manuel, J. Hernandez-Rodriguez, I. Lobon, H. R. Siegmund, L. Pagani, M. A. Quail, C. Hvilson, A. Mudakikwa, E. E. Eichler, M. R. Cranfield, T. Marques-Bonet, C. Tyler-Smith, A. Scally, *Science* **2015**, 348, 242.
- [91] R. Do, D. Balick, H. Li, I. Adzhubei, S. Sunyaev, D. Reich, *Nat. Genet.* **2015**, 47, 126.
- [92] S. Liu, E. D. Lorenzen, M. Fumagalli, B. Li, K. Harris, Z. Xiong, L. Zhou, T. S. Korneliussen, M. Somel, C. Babbitt, G. Wray, J. Li, W. He, Z. Wang, W. Fu, X. Xiang, C. C. Morgan, A. Doherty, M. J. O'Connell, J. O. McInerney, E. W. Born, L. Dalen, R. Dietz, L. Orlando, C. Sonne, G. Zhang, R. Nielsen, E. Willerslev, J. Wang, *Cell* **2014**, 157, 785.
- [93] R. R. Fitak, E. Mohandesan, J. Corander, A. Yadamsuren, B. Chuluunbat, O. Abdelhadi, A. Raziq, P. Nagy, C. Walzer, B. Faye, P. A. Burger, *Commun. Biol.* **2020**, 3, 316.

- [94] P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, *Bioinformatics* **2011**, 27, 2156.
- [95] K. Wang, M. Li, H. Hakonarson, *Nucleic Acids Res.* **2010**, 38, e164.
- [96] T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, G. Yu, *Innovation* **2021**, 2, 100141.
- [97] A. Akbari, J. J. Vitti, A. Iranmehr, M. Bakhtiari, P. C. Sabeti, S. Mirarab, V. Bafna, *Nat. Methods* **2018**, 15, 279.
- [98] J. C. Barrett, B. Fry, J. Maller, M. J. Daly, *Bioinformatics* **2005**, 21, 263.
- [99] M. Gautier, A. Klassmann, R. Vitalis, *Mol. Ecol. Resour.* **2017**, 17, 78.
- [100] B. Charlesworth, J. D. Jensen, *Annu. Rev. Ecol. Evol. Syst.* **2021**, 52, 177.
- [101] J. Si, D. Dai, K. Li, L. Fang, Y. Zhang, *Genes* **2023**, 14, 890.
- [102] S. Liu, Y. Gao, O. Canela-Xandri, S. Wang, Y. Yu, W. Cai, B. Li, R. Xiang, A. J. Chamberlain, E. Pairo-Castineira, K. D'Mellow, K. Rawlik, C. Xia, Y. Yao, P. Navarro, D. Rocha, X. Li, Z. Yan, C. Li, B. D. Rosen, C. P. Van Tassell, P. M. Vanraden, S. Zhang, L. Ma, J. B. Cole, G. E. Liu, A. Tenesa, L. Fang, *bioRxiv* **2020**, 2020.
- [103] S. Liu, Y. Gao, O. Canela-Xandri, S. Wang, Y. Yu, W. Cai, B. Li, R. Xiang, A. J. Chamberlain, E. Pairo-Castineira, K. D'Mellow, K. Rawlik, C. Xia, Y. Yao, P. Navarro, D. Rocha, X. Li, Z. Yan, C. Li, B. D. Rosen, C. P. Van Tassell, P. M. Vanraden, S. Zhang, L. Ma, J. B. Cole, G. E. Liu, A. Tenesa, L. Fang, *Nat. Genet.* **2022**, 54, 1438.
- [104] T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck 3rd, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, *Cell* **2019**, 177, 1888.
- [105] X. Zhang, Y. Lan, J. Xu, F. Quan, E. Zhao, C. Deng, T. Luo, L. Xu, G. Liao, M. Yan, Y. Ping, F. Li, A. Shi, J. Bai, T. Zhao, X. Li, Y. Xiao, *Nucleic Acids Res.* **2019**, 47, D721.
- [106] J. J. Wu, S. Zhu, F. Gu, T. G. Valencak, J. X. Liu, H. Z. Sun, *J. Adv. Res.* **2022**, 37, 1.
- [107] T. Kumar, K. Nee, R. Wei, S. He, Q. H. Nguyen, S. Bai, K. Blake, M. Pein, Y. Gong, E. Sei, M. Hu, A. K. Casasent, A. Thennavan, J. Li, T. Tran, K. Chen, B. Nilges, N. Kashikar, O. Braubach, B. Ben Cheikh, N. Nikulina, H. Chen, M. Teshome, B. Menegaz, H. Javaid, C. Nagi, J. Montalvan, T. Lev, S. Mallya, D. F. Tifrea, et al., *Nature* **2023**, 620, 181.
- [108] X. Qiu, A. Hill, J. Packer, D. Lin, Y. A. Ma, C. Trapnell, *Nat. Methods* **2017**, 14, 309.
- [109] G. S. Gulati, S. S. Sikandar, D. J. Wesche, A. Manjunath, A. Bharadwaj, M. J. Berger, F. Ilagan, A. H. Kuo, R. W. Hsieh, S. Cai, M. Zabala, F. A. Scheeren, N. A. Lobo, D. Qian, F. B. Yu, F. M. Dirbas, M. F. Clarke, A. M. Newman, *Science* **2020**, 367, 405.