

RESEARCH

Genomic analyses reveal distinct genetic architectures and selective pressures in buffaloes

Ting Sun^{1,†}, Jiafei Shen^{1,†}, Alessandro Achilli ^{2,†}, Ningbo Chen¹, Qiuming Chen¹, Ruihua Dang¹, Zhuqing Zheng¹, Hucai Zhang³, Xiaoming Zhang⁴, Shaoqiang Wang¹, Tao Zhang⁵, Hongzhao Lu⁵, Yun Ma⁶, Yutang Jia⁷, Marco Rosario Capodiferro², Yongzhen Huang¹, Xianyong Lan¹, Hong Chen¹, Yu Jiang ^{1,*} and Chuzhao Lei ^{1,*}

¹Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi 712100, China; ²Dipartimento di Biologia e Biotechnologie “L. Spallanzani,” Università di Pavia, Pavia 27100, Italy; ³Key Laboratory of Plateau Lake Ecology and Environment Change, Yunnan University, Kunming 650504, China; ⁴State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; ⁵School of Bioscience and Engineering, Shaanxi University of Technology, Hanzhong, Shaanxi 723000, China; ⁶Agricultural College, Ningxia University, Yinchuan 750021, China and ⁷Institute of Animal Science and Veterinary Medicine, Anhui Academy of Agriculture Science, Hefei 230001, China

*Correspondence address. Yu Jiang, Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi 712100, China. E-mail: yu.jiang@nwfufu.edu.cn  <http://orcid.org/0000-0003-4821-3585> and Chuzhao Lei, Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi 712100, China. leichuzhao1118@126.com  <http://orcid.org/0000-0003-1647-1037>

[†]These authors contributed equally to this work.

Abstract

Background: The domestic buffalo (*Bubalus bubalis*) is an essential farm animal in tropical and subtropical regions, whose genomic diversity is yet to be fully discovered. **Results:** In this study, we describe the demographic events and selective pressures of buffalo by analyzing 121 whole genomes (98 newly reported) from 25 swamp and river buffalo breeds. Both uniparental and biparental markers were investigated to provide the final scenario. The ancestors of swamp and river buffalo diverged ~0.23 million years ago and then experienced independent demographic histories. They were domesticated in different regions, the swamp buffalo at the border between southwest China and southeast Asia, while the river buffalo in south Asia. The domestic stocks migrated to other regions and further differentiated, as testified by (at least) 2 ancestral components identified in each subspecies. Different signals of selective pressures were also detected in these 2 types of buffalo. The swamp buffalo, historically used as a draft animal, shows selection signatures in genes associated with the nervous system, while in river dairy breeds, genes under selection are related to heat stress and immunity. **Conclusions:** Our findings substantially expand the catalogue of genetic variants in buffalo and reveal new insights into the evolutionary history and distinct selective pressures in river and swamp buffalo.

Received: 22 May 2019; Revised: 26 November 2019; Accepted: 27 December 2019

© The Author(s) 2020. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Keywords: buffalo; whole-genome resequencing; genetic history; selection

Background

The domestic buffalo is an important farm animal in tropical and subtropical regions, which can provide milk, meat, and draught power for rice cultivation. The domestic buffalo can be divided into 2 types: swamp and river buffalo. These 2 types show differences concerning body size, outward appearance, biological characteristics, and chromosome karyotype ($2n = 48$ in swamp buffalo; $2n = 50$ in river buffalo) [1, 2]. The swamp buffalo is mainly bred in extensive rural areas in northeast India, southeast Asia, and south China, while the river buffalo is distributed from western India to Mediterranean areas. Swamp buffalo was traditionally raised as a draught animal for rice cultivation, while the river type was mainly selected for milk production [3].

There is a large agreement on the common ancestor of river and swamp buffaloes, both descending from the wild Asian buffalo (*Bubalus arnee*) [3]. However, details on the domestication process and its consequences are still missing. The oldest domestic buffalo remains of southeast Asia were found in northern Thailand and dated to 2,900–2,300 years ago [4], while other archeozoological evidence is scarce [5]. Therefore, the genomic screening of current breeds was often used to clarify the overall scenario. In particular, both uniparental markers were initially studied [6–13], while further details were eventually provided by autosomal analyses [14]. The current data on river buffalo point to an initial domestication in the Indian subcontinent 6,300–4,600 years ago and a following migration westwards into southern Europe [6, 10, 14]. The swamp buffalo was probably domesticated in the China/Indochina border. Then, the original stocks migrated to other regions: northward to China and then bending southwards into the Philippines; southward initially across the Mekong, then to Sumatra (Mekong colonization), and finally eastwards to the rest of Indonesia [12–14].

Considering that the entire genome variation of buffalo was largely unexplored [15], we sequenced the whole genome of 98 buffaloes from 21 swamp and 4 river breeds (Supplementary Tables 1 and 2) with different geographic origins in order to fully describe the genomic diversity, population structure, and demographic history of this important livestock species and to reveal possible signs of natural and artificial selection.

Data Description

We sampled a total of 98 domestic buffaloes (NCBI:txid89462) from different locations: China (81), Laos (5), Vietnam (4), and India (1 Nili-Ravi, 2 Murrah, and 5 Indian buffaloes). All genomic data have been submitted to the NCBI SRA under the BioProject accession number PRJNA547460. After adding 23 available genomes (Fig. 1a, Supplementary Tables 1 and 2) [15], the final dataset of 121 genomes was subdivided into 6 geographic groups: Upper Yangtze, Middle-Lower Yangtze, Southwest China, Southeast Asia, South Asia, and Italy (Fig. 1a, Supplementary Fig. 1).

Analyses

Population genetic structure and relationships

Neighbour-joining (NJ) trees, principal component analysis (PCA), and ADMIXTURE were used to explore the genetic

relationships among the examined 121 buffaloes. The NJ tree, rooted with *Syncerus caffer*, showed a deep division between swamp and river buffaloes; then buffaloes from adjacent geographical regions form distinctive clades (Fig. 1b). The same geographic/genomic proximity was also confirmed by the maximum-likelihood (ML) tree (Supplementary Fig. 2). The PCA (Fig. 1c and d) showed that the first principal component (PC1) was driven by differences between swamp and river buffaloes. The PC2 (Fig. 1c) separates ITA and SA river buffaloes. This structure was also confirmed (Supplementary Note 2, Supplementary Fig. 3) when whole-genome data were merged with samples genotyped using the 90 K Axiom™ Buffalo Genotyping Array [14]. The PC3 (Fig. 1d, Supplementary Fig. 4, Supplementary Table 7) highlights the variability among swamp breeds, separating the southwest Chinese and Vietnamese buffaloes from other swamp breeds. The ADMIXTURE analysis confirmed this genetic structure (Fig. 1e, Supplementary Table 8, Supplementary Fig. 5). At $K = 4$, all individuals were unambiguously assigned to 2 ancestries in swamp buffalo (cold colors: SC, SEA) and 2 in river buffalo (warm colors: SA, ITA) (Fig. 1e). Within swamp buffalo, the SC ancestral component (blue) characterizes most of Upper and Middle-Lower Yangtze Valley buffaloes, but it was also detected in southwest Chinese and Laotian (LA) breeds. The SEA ancestry (green) was shared between Vietnamese and 3 southwest Chinese breeds, and found at a low level in Middle-Lower Yangtze, showing evidence of recent admixture, probably due to the frequent trading of buffaloes among these regions. As for the river buffalo, the SA ancestry (orange) was abundant in Murrah and Nili-Ravi buffaloes, while the ITA ancestry (red) is unique to Italian breeds. Some swamp buffaloes showed evidence of admixture, which may be attributable to introgression events by way of recent crossbreeding with river buffalo for improving milk production traits [16].

Uniparental phylogenies

Y-chromosome and mitochondrial DNA (mtDNA) are very useful to investigate genetic origins and ancient migrations (Supplementary Note 3). After quality control and filtering, 520 Y-chromosome single-nucleotide polymorphisms (SNPs) were retrieved from 89 male buffaloes and used to build a phylogenetic tree that clearly divides swamp (YS) and river (YR) clades. Most of the variants defined the branch that connects swamp and river common ancestors. Two haplogroups (YS1 and YS2) were identified in the swamp branch, both retrieved in all geographic regions. The haplogroup YS1 dominated the buffaloes from Upper and Middle-Lower Yangtze (76.09%), while the haplogroup YS2 was extremely frequent (84.62%) in southwest China and southeast Asia (Fig. 2a, Supplementary Fig. 6, and Supplementary Table 9).

We also inferred the maternal history of buffalo, combining the novel 118 mitogenomes from this study with 107 sequences from previous studies (Supplementary Table 10) [12]. Swamp buffaloes can be assigned to 5 previously defined lineages: 2 major haplogroups (SA haplogroup and SB haplogroup with various sub-clades) and 3 rare ones (SC, SD, and SE) (Fig. 2b, Supplementary Fig. 7, Supplementary Table 10). This larger dataset confirmed the geographic differentiation of current swamp buffalo populations (Fig. 1e), as previously reported by analyzing partial and complete mtDNA data [12, 13]. The Upper

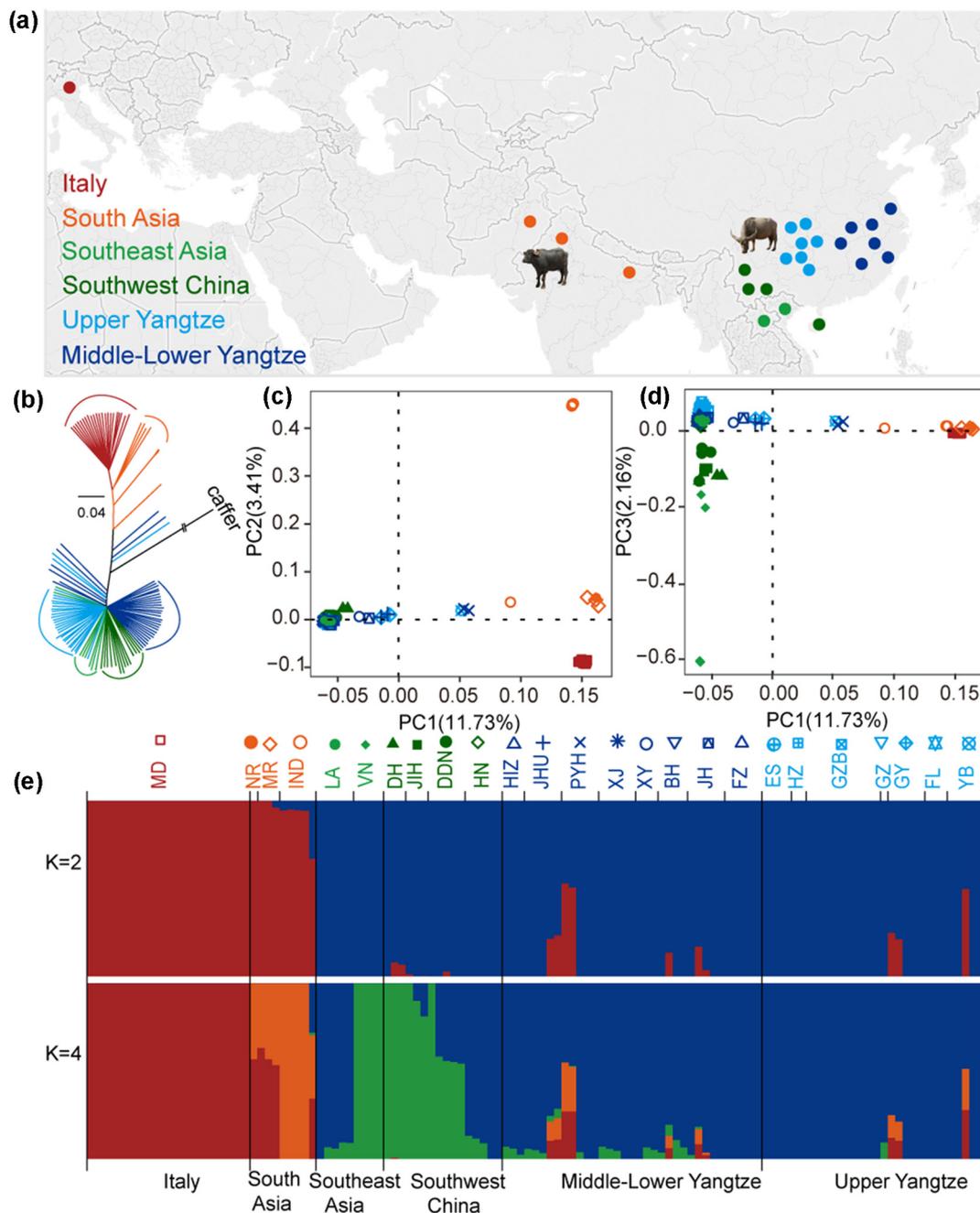


Figure 1: Population structure and relationships among buffaloes. (a) Geographic map indicating the origins of the buffalo breeds. (b) Neighbour-joining tree of buffaloes constructed using whole-genome autosomal SNP data. (c, d) Principal component analyses (PCA) showing PC1 against PC2 and PC1 against PC3, respectively. Each breed was labelled with different colors and shapes as shown in the top of Fig. 1e. (e) Genetic structure of buffalo breeds using ADMIXTURE program with $K = 2, 4$. Population acronyms are explained in Supplementary Tables 1 and 5.

and Middle-Lower Yangtze buffalo breeds primarily belonged to lineage SA1. Buffaloes from southwest China and southeast Asia harbor almost all lineages, except for the rare lineage SE, and showed high frequencies of SA2 and SB2. The greatest variety of lineages was identified in southwest China and southeast Asia, thus confirming the hypothesized domestication of swamp buffalo at the border of the 2 regions [12, 14].

As for the Y-chromosome variation of river buffalo, we identified 1 ancestral node (YR) and 2 haplogroups (YR1 and YR2). The YR haplotype was found in 1 Indian, 2 south Asian, and

9 southern Chinese buffaloes. The latter finding was probably due to recent importation of bulls in China through cross-breeding programs [16], which is consistent with the autosomal analyses (PCA, ADMIXTURE, and NJ tree, Fig. 1b–e). The haplogroups YR1 and YR2 were found in south Asia and Italy, respectively (Fig. 2a and 2b and Supplementary Fig. 4). Four different mtDNA haplogroups have been identified in river buffaloes (Fig. 2b), thus adding a new lineage (R4) to the 3 previously defined ones (R1, R2, and R3). However, considering the low number (and coverage) of river mitogenomes (22 ITA and 10 SA buffaloes)

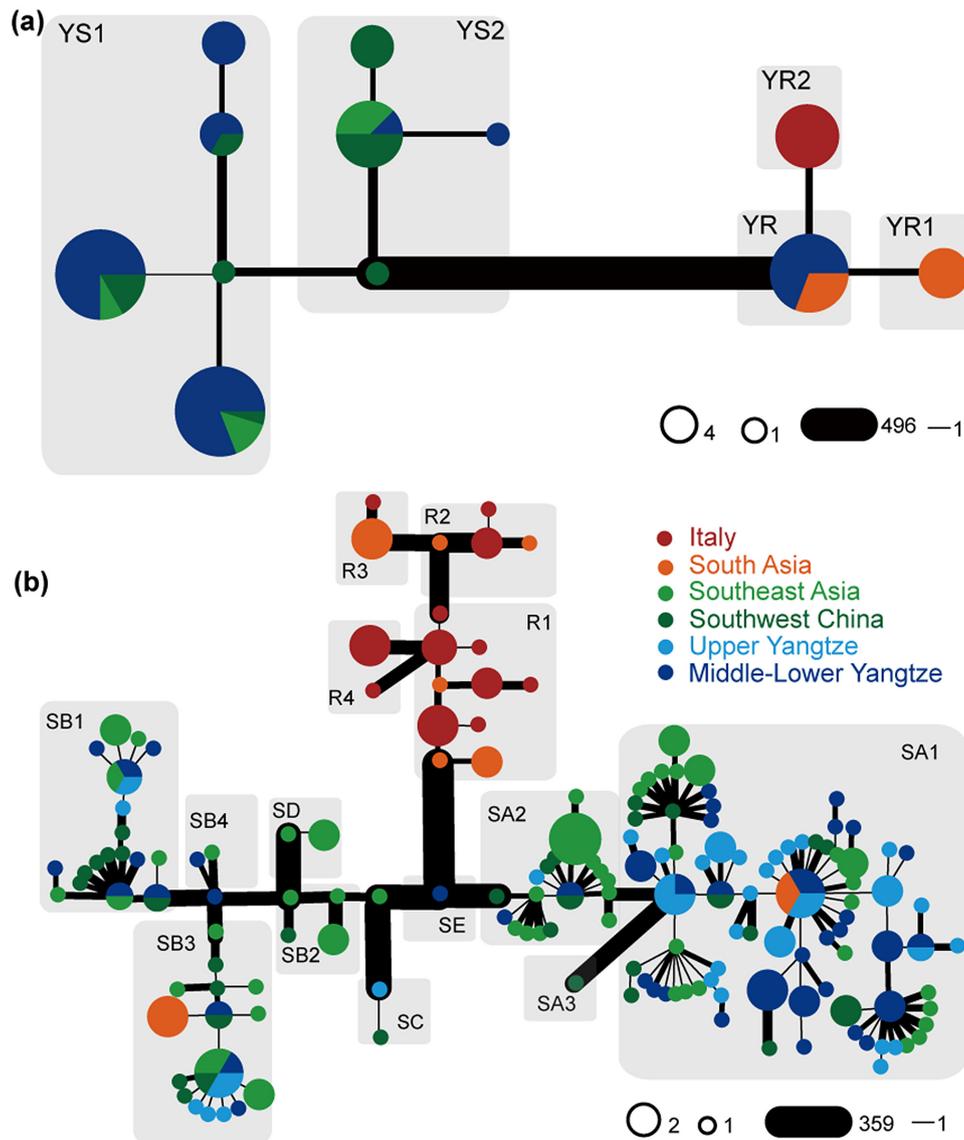


Figure 2: Y-chromosome and mitogenome phylogenies. The width of the edges is proportional to the number of pairwise differences between the joined haplotypes. (a) Y-chromosome network using 520 SNPs. (b) Mitogenome network of swamp buffalo. Different haplogroups of Y-chromosome and mitogenomes were labelled in each gray shadow.

no further phylogeographic analyses have been carried out (Fig. 2).

Demographic history

We used the multiple sequentially Markovian coalescent (MSMC) method to detect the changes in the effective population size (N_e) of 4 “ancestral” buffalo groups. River and swamp buffaloes underwent 2 apparent expansions and 2 bottlenecks that seemed to overlap with 3 major glacial cycles (Fig. 3a). Initially, the N_e of river and swamp buffaloes showed similar demographic trajectories, peaking at ~ 0.8 Mya and then quickly declining during the Naynayxungla glaciation (NG, 0.78–0.50 Mya), which was the most extensive glaciation during the Quaternary Period. The ancestral N_e of river buffalo recovered very quickly and reached the highest peak at ~ 70 kya after a short bottleneck ~ 0.23 Mya. On the contrary, the ancestors of the swamp buffalo went through a long period of population decline until

the retreat of the penultimate glaciation (PG, ~ 0.30 – 0.13 Mya), and then, the N_e slightly increased starting from ~ 0.10 Mya. During the interglacial period, both river and swamp buffalo populations reached another peak and quickly declined during the last glaciation (LG). These results confirmed that the glaciations had a strong effect on the demographic history of swamp buffalo, as already observed by analyzing complete mtDNAs [12]. The decline from ~ 6.0 to ~ 4.5 kya is consistent with the onset of domestication, before the final increase until the present time.

The MSMC approach was also used to calculate the divergence time among 4 buffalo ancestral populations: SC, SEA, SA, and ITA (Fig. 3b). We observed a decrease in the cross-coalescence rate between river and swamp buffalo to 0.5 at ~ 0.21 – 0.23 Mya (from 0.25 at ~ 0.15 – 0.18 Mya to 0.75 at ~ 0.28 – 0.38 Mya). The splitting time of SC and SA ancestors was observed at ~ 28 kya, while a decline to 0.5 between ITA and SA was detected later at ~ 11 kya.

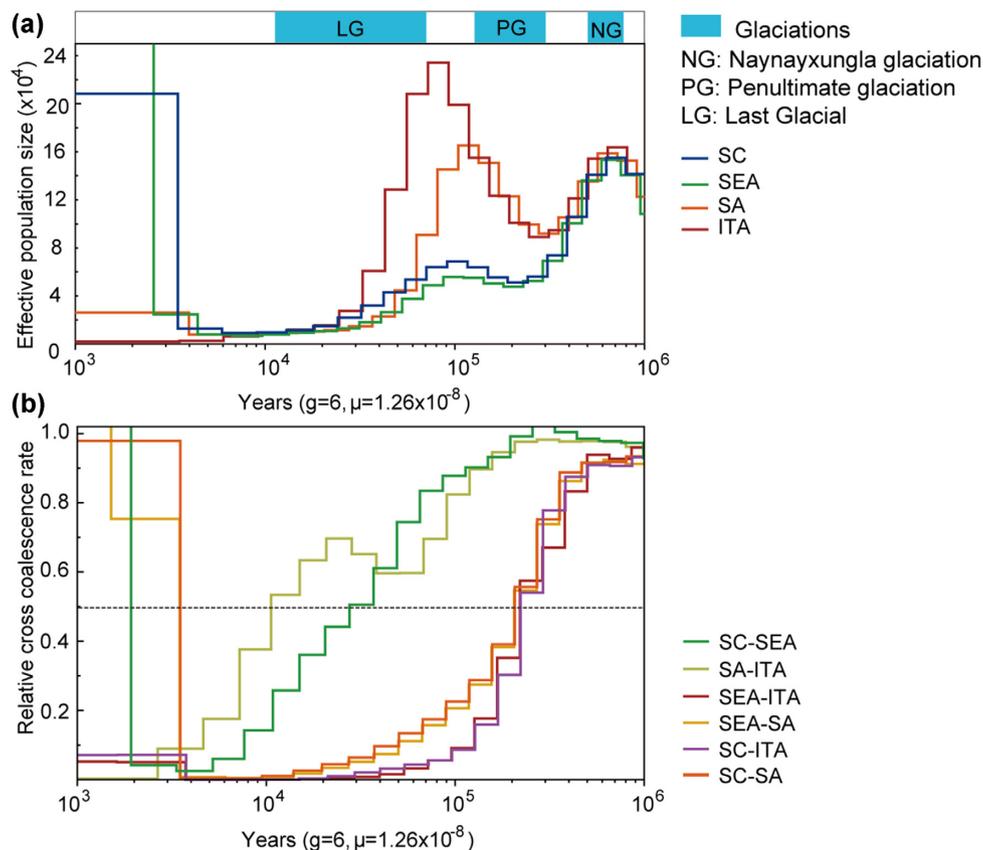


Figure 3: Demographic history and divergence of buffalo populations using MSMC. (a) Population size history inference of swamp and river buffalo based on 4 high-coverage haplotypes from southwest China (SC), southeast Asia (SEA), south Asia (SA), and Italy individuals (ITA). (b) Inferred relative cross-coalescence rates between pairs of populations over time based on the same 4 haplotypes.

Genome-wide differential selection in river and swamp buffalo

We applied 4 methods (F_{ST} , π ln ratio, cross-population composite likelihood ratio (XP-CLR), cross-population extended haplotype homozygosity (XP-EHH) to detect genomic regions related to selection in river and swamp buffalo. Two or more methods showed outlier signals (P -value < 0.005) in overlapping regions and were therefore considered as candidate selective regions. Functional gene set enrichment was used to identify KEGG pathways and Gene Ontology (GO) terms that are statistically significantly associated with the genes.

In river buffalo, a total of 502 candidate regions under selection containing 569 genes were detected (Supplementary Tables 11–14). Candidate genes in river buffalo are significantly over-represented (corrected P -value < 0.05) in the Jak-STAT signaling pathway, glioma, and pathways associated with cancer (Supplementary Tables 15 and 16). The Jak-STAT pathway plays a crucial role in prolactin signal transduction of the mammary gland [17] and control of immune responses [18, 19]. We also identified GO terms associated with immunity and others with DNA damage and repair (Supplementary Table 16, Supplementary Fig. 8a). In particular, 4 candidate genes (*AP4B1*, *BCL2L15*, *PHTF1*, and *PTPN22*) (Fig. a) are associated with the immune system response, while *MMS22L* (Fig. 4c) may be associated with heat stress. Moreover, we detected few non-synonymous variants that are completely fixed at *PTPN22* and *BCL2L15* in river buffaloes (Fig. 4b, Supplementary Fig. 8b), as well as 1 at *MMS22L* (Fig. 4d). Additional signs of selection were identified

in genes coding for productive and economically significant traits, such as *NUMB* [20] and *SGMS2* [21] associated with milk production, while others related to growth (*NRF1*) [22] and feed efficiency (*TNPO3*) [23].

In swamp buffalo, a total of 171 candidate regions under selection containing 209 genes were detected (Supplementary Tables 11 and 17–19). These genes can be significantly associated (corrected P -value < 0.05) to 4 KEGG pathways (Supplementary Tables 20 and 21). The most significant one was “Glutamatergic synapse” involving 5 genes (*HOMER1*, *GRIK2*, *DLGAP1*, *GNG7*, *LOC102398542*) (Fig. 5, Supplementary Table 20, Supplementary Fig. 9). We also found significantly over-represented GO terms associated with dendritic spines (*TIAM1*, *RELN*, *DISC1*, *NLGN1*, *LOC102398542*) and nervous system (e.g., neuron, dendritic spine, synapse) involving 42 genes (*HDAC9*, *HOMER1*, *BIN1*, and *GRIK2* showed higher values in the detection methods) (Supplementary Table 21, Supplementary Fig. 9). Among these detected candidate genes, *HDAC9*, *HOMER1*, and *GRIK2* (Fig. F5a–c) may be associated with the development of the nervous system in swamp buffalo (Fig. 5d).

Discussion

In this study, we analyzed the whole-genome sequence of 121 buffaloes (91 swamp and 30 river buffaloes). Our autosomal data reveal an ancient separation between river and swamp buffaloes ~0.23 Mya, overlapping with previously reported data [9, 11, 12] and certainly predating buffalo domestication. Therefore, we

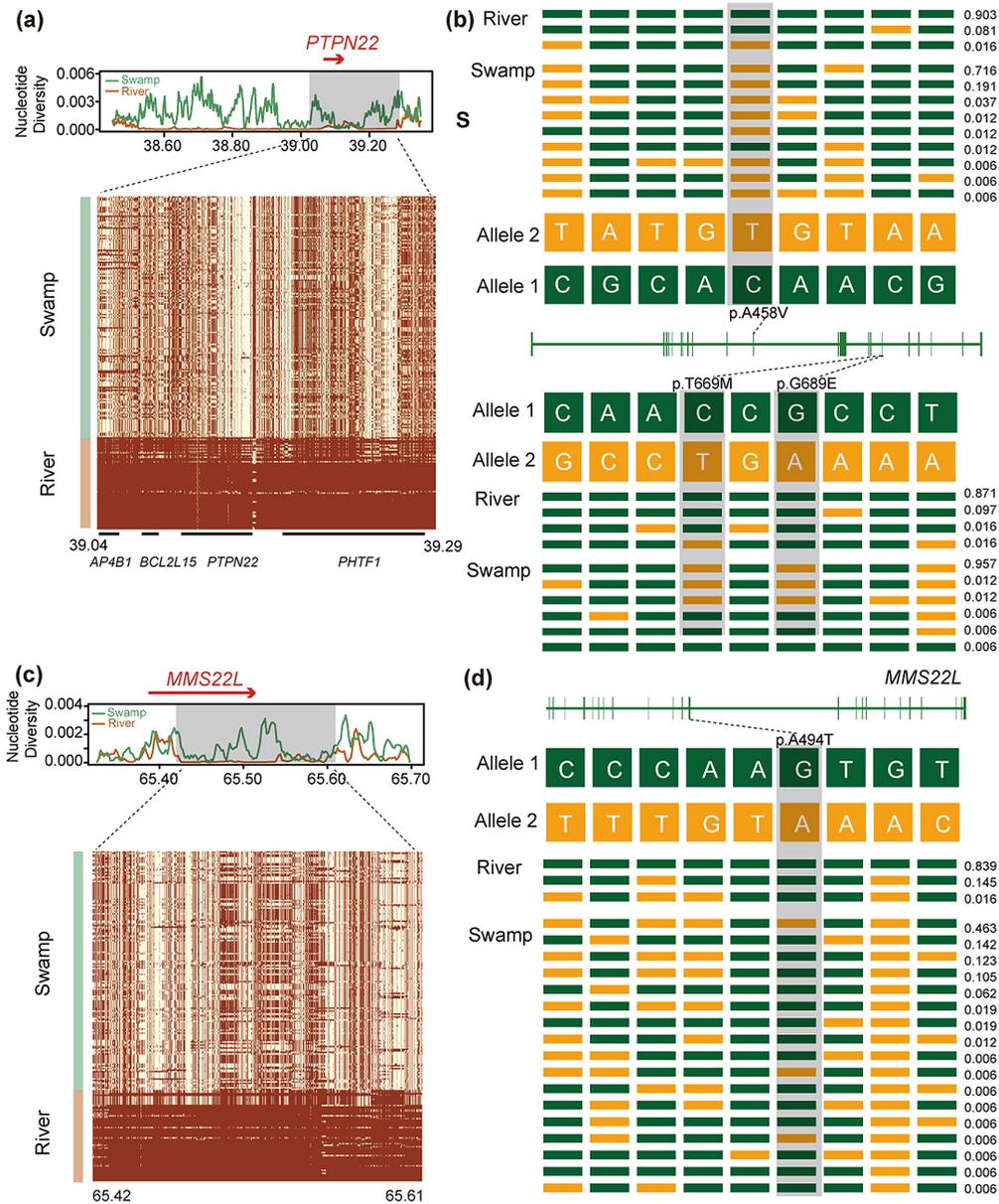


Figure 4: Signatures of selective sweep regions at *PTPN22* and *MMS22L* genes in river buffalo. Different parameters were estimated for each gene (*PTPN22* and *MMS22L*): nucleotide diversity, degree of haplotype sharing across populations (a and c). A red arrow notes the specific gene region. A schematic structure of each gene (b and d) is also depicted with exons indicated by vertical bars and reference/alternative alleles noted with different colors (green/yellow) and combined to form different haplotypes (each with a specific haplotype frequency next to it). Non-synonymous SNPs are highlighted in gray.

can assume that river and swamp buffaloes probably descended from different wild populations. The demographic histories of swamp and river buffaloes were differentially linked to climatic changes. A similar pattern has been observed in taurine and indicine cattle, suggesting similar habitat requirements [24]. After diverging, the 2 types of buffalo evolved independently. In fact, 2 different ancestral components were identified for each of them (Fig. 1e). Distinctive Y-chromosome lineages were also revealed in river buffalo, with a basal haplogroup (YR) unique to breeds from India and Pakistan, YR1 typical of south Asia, and YR2 identified only in Italy. The most likely scenario based on previous studies [14] points to an early river buffalo domestication in the Indo-Pakistan region, followed by westward migration. Later, YR1 remained in south Asia, where it is still highly dif-

fused, whereas YR2 became unique to the buffaloes bred in Italy. In swamp buffalo, YS2 is found mostly in southwest China and southeast Asia (84.62%), while YS1 dominates buffaloes from Upper and Middle-Lower Yangtze (76.09%). Considering that YS2 diverge earlier in the phylogeny, we might speculate that the swamp buffalo population migrated from the southern regions towards the north, where the YS1 experienced a population expansion. A clear geographic pattern was also identified in the mitochondrial gene pool of swamp buffalo. Taking into account the frequencies of uniparental haplogroups in swamp buffalo (Fig. 2), we could observe a correlation between Y-chromosome and mtDNA haplogroups (i.e., YS1 with SA and YS2 with SB), which could mark some similarities between maternal and paternal histories.

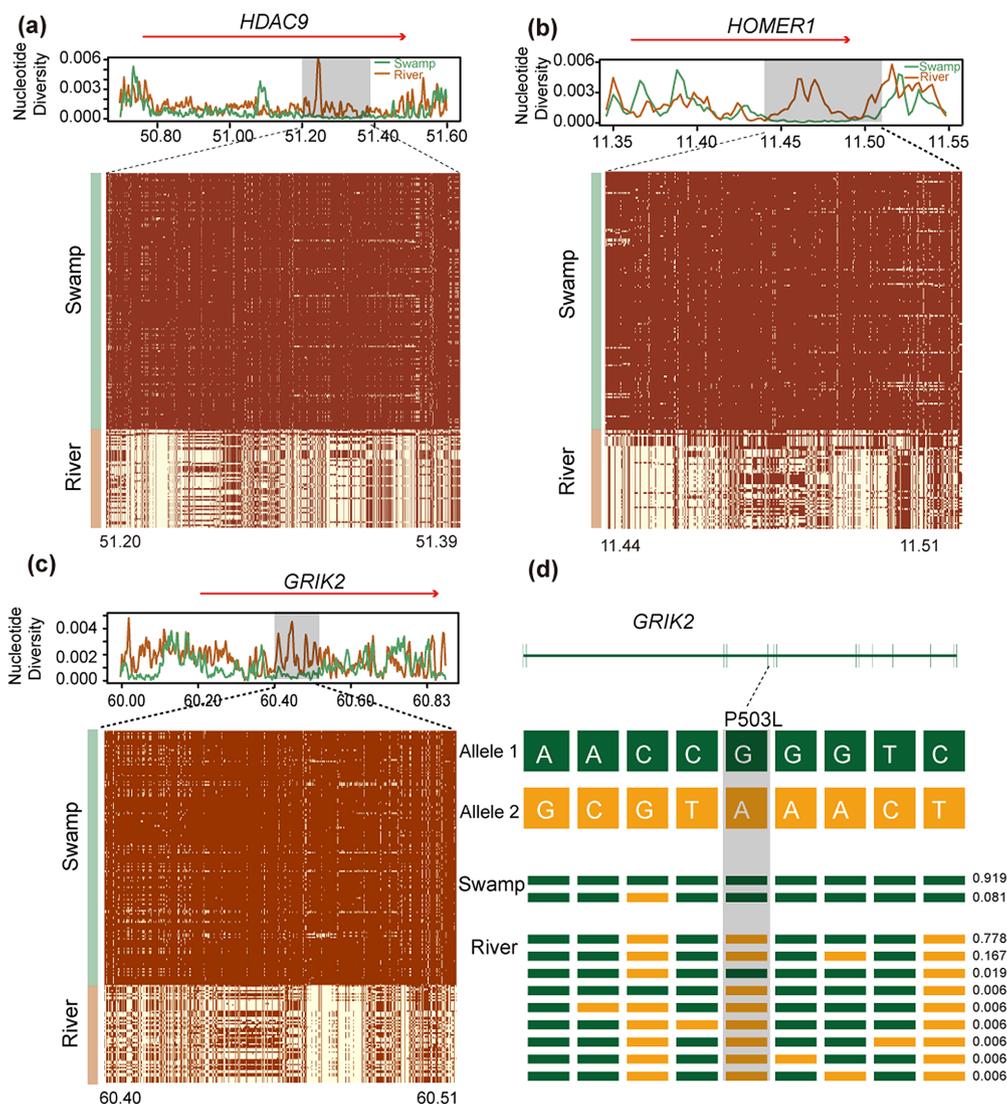


Figure 5: Signatures of selective sweep regions at HDAC9, HOMER1, and GRIK2 genes in swamp buffalo. See the legend of Figure 4 for further details.

We identified significant and distinctive signatures of selective sweeps in these 2 types of buffalo. To better understand possible explanations for the selective pressures, we have explored the most likely biological functions of these genes. River buffaloes are mainly distributed in western India to Mediterranean areas, where domestic herds are usually more resistant to various diseases than in tropical regions [16]. In river genomes, positively selected genes are significantly over-represented in GO terms associated with immunity. Among these, *PTPN22* encodes a negative regulator of T-cell receptor (TCR), which has been associated with human autoimmune diseases [25–28], bovine leukemia virus [29], and milk somatic cell counts of cow [29]. Among the other genes (*PTPN22*, *AP4B1*, *BCL2L15*, and *PHTF1*) were those related to bovine leukemia virus [30] and human autoimmune diseases [28]. The *BCL2L15* SNP A226G (acid changed: T76A) is a conserved region and almost fixed with the allele A in river buffalo ($P > 0.90$) (Supplementary Fig. 8b). In addition, heat stress is an important issue for many livestock species, particularly for dairy animals, leading to different problems in the phenotypic features (e.g., impairment of reproduction and slower growth), as well as in cellular stability, causing a reduced efficiency of DNA synthesis [31]. We found

3 GO terms associated with DNA damage and repair (cellular response to DNA damage stimulus; DNA repair; double-strand break repair) that are significantly over-represented in river buffalo. *MMS22L*, a component of the *MMS22L-TONSL* complex important for the DNA repair system [32, 33], was involved in these 3 GO terms. Actually, heat stress can induce the formation of double-stranded DNA breaks [34] and inhibit the functionality of the homologous recombination (HR) system [35]. Double-stranded DNA breaks can be repaired by HR, allowing DNA replication to continue at stalled or broken forks [36, 37]. *MMS22L* can facilitate HR-mediated maintenance of genome stability during DNA replication [38]. Taking into account that river buffalo is mainly selected for milk production and well adapted to a hot climate [39], these genes might play an important role in the heat adaptability of river buffalo. Finally, we were also able to identify signatures of selection on some genes for important economic and reproductive traits, which is not unexpected considering the great effort undertaken by the herders to improve the river breeds.

The swamp buffalo has historically been used as a draft animal to provide farm power in rice cultivation. Therefore, these animals are very docile and easy to handle and train. The swamp

buffalo genome showed signs of selection in some genes of the “Glutamatergic synapse” pathway (*HOMER1*, *GRIK2*, *DLGAP1*, *GNG7*, and *LOC102398542*), which plays an important role in behavior, particularly concerning adaptation to stress and fear responses (Supplementary Table 20) [40]. We also found 42 genes that were over-represented in GO categories associated with the nervous system. Among them *HDAC9*, *HOMER1*, *BIN1*, and *GRIK2* seem the best candidates we have identified (Supplementary Table 21). *HDAC9* (Fig. 5a), a member of the class II HDAC proteins, plays a crucial role in neuronal differentiation during cortical development [41] and muscle development [42–45]. A study showed that degradation of class II HDAC proteins can activate myocyte enhancer factor 2, which enhances muscle endurance and fatigue resistance [46]. *HOMER1* (Fig. 5b) encodes a member of the homer family of dendritic proteins, involved in several psychiatric disorders, such as schizophrenia [47, 48] and major depression [49]. *HOMER1* plays an important role in brain development and behavior; *HOMER1* knockout mice showed deficits in learning and memorizing, and impairment of pain perception [47, 50–53]. *HOMER1* is also an important scaffold for TRP channels and regulates mechanotransduction in skeletal muscle [54]. Mice lacking *HOMER1* showed myopathy with decreased muscle fiber cross-sectional area and reduced skeletal muscle strength generation [54]. Studies have showed that *BIN1* is associated with Alzheimer disease [55–57]. *BIN1* is also involved in the biogenesis of T-tubules, which are responsible for the plasma membrane invaginations that allow for the excitation-contraction coupling machinery in cardiac and skeletal muscles [58–60]. *GRIK2* (Fig. 5c and 5d) encodes for GluR6, a kainite receptor that is highly expressed in the brain and is associated with autosomal recessive mental retardation [61]. *GRIK2* knockout mice exhibited reduction in fear memory [62], less anxiety, and more risk-taking type than despair-type behavior [63]. Notably, *GRIK2* was also identified as a candidate selective gene in domestic rabbits [64]. In addition, we also identified over-represented GO categories associated with dendritic spines (*TIAM1*, *RELN*, *DISC1*, *NLGN1*, *LOC102398542*) (Supplementary Fig. 9). The structural and functional plasticity of dendritic spines is crucial for learning and memorizing [65]. *TIAM1* plays an important role in the formation and morphogenesis of dendritic spines [66–68]. *RELN*, *DISC1*, and *NLGN1* were related to schizophrenia, mood disorders, and memory [69–77]. Therefore, these genes could be associated with nervous system development in swamp buffalo.

Potential implications

This is the first population genetics study on buffalo using a large amount of whole-genome resequencing data. We reconstructed the genetic history and population structure of buffalo from all genetic perspectives using both uniparental and biparental markers. The final scenario indicates that the ancestors of swamp and river buffalo diverged ~0.23 Mya. The swamp buffalo was then domesticated at the border between southwest China and southeast Asia, while the river buffalo was domesticated in south Asia (between northern India and Pakistan). The domestic herds then migrated to other regions and further differentiated. In fact, we were able to identify 2 ancestral and distinctive components in the current genomes of swamp (south China and southeast Asia components) and river (South Asia and Italy) populations. The different genetic history of these 2 subspecies is also evident by 2 distinct selection patterns identified in their genomes. River buffalo was selected to improve milk production, while the swamp buffalo was mainly raised to

provide power for rice cultivation. We were able to intercept distinctive signature of selection in genes associated with nervous and muscle development in swamp buffaloes and in genes related to economic and reproductive traits in river breeds.

In summary, this is the first study providing a large amount of genomic data on river and swamp buffaloes, which was needed to describe their current genetic diversities and population structures, to reconstruct their demographic histories, and to scan for distinctive selective pressures.

Methods

Sample collection and sequencing

We sampled a total of 98 buffaloes from different locations: China (81), Laos (5), Vietnam (4), and India (1 Nili-Ravi, 2 Murrah, and 5 Indian buffaloes). Genomic DNA was extracted from ear tissue or blood samples using the standard phenol-chloroform protocol [78], amplified in genomic libraries with an average insert size of 500 bp, and sequenced (150-bp paired-end reads) on an Illumina HiSeq 2000. We also considered 23 available genome sequences from river buffalo, including 22 Mediterranean and 1 Murrah buffaloes. Additional details are provided in Supplementary Tables 1 and 2. This study was approved by Institutional Animal Care and Use Committee of Northwest A&F University (Permit No. NWAFA1019).

Alignments and variant identification

All cleaned reads were aligned to the reference genome (GCA.000471725.1) linked to “24+X+unplaced” pseudo-chromosomes (Supplementary Note 1, Supplementary Table 4) using BWA-MEM with default settings [79]. Duplicate reads were filtered using Picard tools. The SNPs were detected with GATK, version 3.6-0-g89b7209 [80], and filtered using the “VariantFiltration” tool, as described in Supplementary Note 1.

Phylogenetic and population structure analyses

NJ tree, PCA, and ADMIXTURE methods were used to explore the genetic relationships among buffalo populations (Supplementary Note 2). An individual-based NJ tree based on the matrix of pairwise genetic distances from the autosomal SNP data of 121 buffaloes was constructed with PLINK version 1.9 (PLINK, [RRID:SCR.001757](#)) and visualized with FigTree (FigTree, [RRID:SCR.008515](#)). TreeMix software was used to construct a population-level phylogeny [81]. The PCA was performed using SmartPCA in the package EIGENSOFT v5.0 (Eigensoft, [RRID:SCR.004965](#)) [82] and the eigenvectors’ significance was detected by the Tracy-Widom test. The population genetic structure was estimated using ADMIXTURE v. 1.3.0 (ADMIXTURE, [RRID:SCR.001263](#)) [83] considering from 2 to 5 clusters (K).

Y-chromosome and mitogenome phylogenies

After removing sites shared with female buffaloes, heterozygous sites, and sites with a genotyping rate <5%, a total of 520 male-specific SNPs were used to construct the phylogenetic tree with BEAST 1.8.0 (BEAST, [RRID:SCR.010228](#)) (Supplementary Note 3). A total of 98 mitochondrial genomes with an average coverage >100× were assembled from the whole-genome resequencing data. An additional 107 whole mtDNA sequences were obtained from GenBank. A phylogenetic tree based on the final alignment was constructed using RaxML (RaxML, [RRID:SCR.006086](#)) with

the following parameters: -f a -x 123 -p 23 -# 100 -k -m 132 GTRGAMMA. The phylogenies were built using pegas [84].

Estimates of the effective population size and divergence time

An MSMC was used to infer effective population sizes (N_e) and divergence times considering 2 samples with average coverage $>16\times$ for each population. Autosomal SNPs of each sample were identified using GATK (GATK, [RRID:SCR.001876](#)). After removing variant outliers (with extremely low or high coverage), all sites were phased using BEAGLE v. 4.1 (BEAGLE, [RRID:SCR.001789](#)) [85]. The same high-coverage samples were also used to infer relative cross-coalescence rate, considering a value of 0.5 as a reference to extrapolate split times between populations (samples). The time scale is calculated using an average generation time of 6 years ($g = 6$) and a mutation rate of $\mu_g = 1.26 \times 10^{-8}$ [86].

Genome-wide selective sweep test

To detect selective sweeps in swamp and river buffalo, we performed the following comparisons: (i) the swamp buffalo as a reference and the river buffalo as the object population and (ii) the river buffalo as a reference and the swamp buffalo as the object population. A total of 4 methods were used. (i) The fixation index (F_{ST}) values [87] were calculated in sliding 50-kb windows with 20-kb steps along the autosomes using VCFtools (VCFtools, [RRID:SCR.001235](#)) [88]. (ii) High differences in genetic diversity (π ln ratio) were calculated with 50-kb sliding windows and 20-kb steps along the autosomes using VCFtools and in-house scripts. (iii) The XP-CLR is a likelihood method that was used to detect whether the change in allele frequency at the locus likely occurred too quickly to be due to random drift between 2 populations [89]. We used such conditions to perform XP-CLR: non-overlapping sliding windows of 50 kb, maximum number of SNPs within each window of 600, and correlation level from which the SNP contribution to the XP-CLR result was down-weighted of 0.95. (iv) We also performed the XP-EHH test for every SNP using the default settings of selscan v1.1 [90], which was designed to detect ongoing or nearly fixed selective sweeps by comparing haplotypes from 2 populations [91]. For the XP-EHH selection scan, our test statistic was the average normalized XP-EHH score in each 50-kb region. Significant genomic regions were identified by P -value < 0.005 . Two or more methods showed significant signals (P -value < 0.005) in overlapping regions and were therefore considered as the candidate regions affected by selection. The KOBAS 3.0 tool (KOBAS, [RRID:SCR.006350](#)) [92] was used to gain a better understanding of their biological functions and involved pathways as enriched GO terms and KEGG pathways.

Availability of Supporting Data and Materials

Raw sequencing data are available from EBI European Nucleotide Archive Bioproject number PRJNA547460. All supporting genotype data and additional materials are available in the Giga-Science GigaDB database [93].

Additional Files

Supplementary Note 1. Linking pseudo-chromosomes.
Supplementary Note 2. The whole-genome diversity of buffalo.
Supplementary Note 3. Population structure analysis.

Supplementary Note 4. Y-chromosome and whole mitochondrial genome phylogeny.

Supplementary Figure 1. Genome-wide distribution of nucleotide diversity of buffaloes in 6 geographical regions in 50-kb sliding windows with 20-kb steps.

Supplementary Figure 2. TreeMix relationships between 25 buffalo breeds.

Supplementary Figure 3. PCA of river buffaloes, with PC1 plotted against PC2.

Supplementary Figure 4. PCA of swamp buffaloes, with PC1 plotted against PC2.

Supplementary Figure 5. Model-based clustering of buffalo using the ADMIXTURE program with K from 2 to 5.

Supplementary Figure 6. ML phylogeny of the Y-chromosome using 520 SNPs for 89 buffaloes.

Supplementary Figure 7. ML phylogeny of the mitochondrial genome.

Supplementary Figure 8. (a) Hierarchical graph of the over-represented (with significant corrected P -values < 0.05) GO terms associated with immunity. The color intensity is positively correlated to the corrected P -value of the GO term. (b) Amino acid conservation of *BCL2L15* (first exon) in mammals. Most amino acids are highly conserved with only a few exceptions, including SNP A226G (acid changed: T76A).

Supplementary Figure 9. (a) Hierarchical graph of the over-represented (with significant corrected P -values < 0.05) GO terms associated with the nervous system. The color intensity is positively correlated to the corrected P -values of the GO term. (b) Amino acid conservation of *LOC102398542*. Nonsynonymous SNP C455T (acid changed: P152L) located in the first exon. Amino acids at this site are highly conserved in other mammals.

Supplementary Table 1. Overview of sample information and sequencing statistics concerning the 121 buffaloes analyzed in this study.

Supplementary Table 2. Summary information on 25 buffalo breeds.

Supplementary Table 3. Distribution of SNPs within various genomic regions.

Supplementary Table 4. Summary information of the linked pseudo-chromosomes.

Supplementary Table 5. The $\theta\pi$ value for the buffalo population groups.

Supplementary Table 6. Pairwise F_{ST} values.

Supplementary Table 7. Tracy-Widom (TW) statistics and P -value for the 10 first eigenvalues in the PCA of buffaloes.

Supplementary Table 8. Cross-validation (CV) errors for ADMIXTURE ancestry models with K ranging from 2 to 5.

Supplementary Table 9. The genotype of 520 SNPs in the Y chromosome.

Supplementary Table 10. Mapping results for 118 novel buffalo mitochondrial genomes plus 107 published mitogenomes.

Supplementary Table 11. A summary of genes from F_{ST} .

Supplementary Table 12. A summary of genes from XP-CLR (P -value $< 0.5\%$) in river buffalo.

Supplementary Table 13. A summary of genes from \ln ratio ($\pi_{\text{swamp}}/\pi_{\text{river}}$) (P -value $< 0.5\%$) in river buffalo.

Supplementary Table 14. A summary of genes from XP-EHH in river buffalo.

Supplementary Table 15. KEGG pathway analysis of candidate genes in river buffalo.

Supplementary Table 16. GO enrichment of candidate genes in river buffalo.

Supplementary Table 17. A summary of genes from XP-CLR (P-value < 0.5%) in swamp buffalo.

Supplementary Table 18. A summary of genes from $\ln(\pi_{river}/\pi_{swamp})$ (P-value < 0.5%) in swamp buffalo.

Supplementary Table 19. A summary of genes from XP-EHH in swamp buffalo.

Supplementary Table 20. KEGG pathway analysis of candidate genes in swamp buffalo.

Supplementary Table 21. GO enrichment of candidate genes in swamp buffalo.

Abbreviations

BWA: Burrows-Wheeler Aligner; bp: base pairs; GATK: Genome Analysis Toolkit; GO: Gene Ontology; HR: homologous recombination; ITA: Italy; kb: kilobase pairs; KEGG: Kyoto Encyclopedia of Genes and Genomes; kya: thousand years ago; LG: last glacialiation; MSMC: multiple sequentially Markovian coalescent; mtDNA: mitochondrial DNA; Mya: million years ago; NG: Naynayxungla glacialiation; NCBI: National Center for Biotechnology Information; NJ: neighbour-joining; PCA: principal component analysis; PG: penultimate glacialiation; RAxML: Randomized Accelerated Maximum Likelihood; SA: south Asia; SC: southwest China; SEA: southeast Asia; SNP: single-nucleotide polymorphism; SRA: Sequence Read Archive; XP-CLR: cross-population composite likelihood ratio; XP-EHH: cross-population extended haplotype homozygosity.

Competing Interests

The authors declare that they have no competing interests.

Funding

The work was supported by the National Beef Cattle and Yak Industrial Technology System (CARS-37), Natural Science Foundation of China (31872317) to C.Z.L., and National Thousand Youth Talents Plan to Y.J.; and the Italian Ministry of Education, University and Research (MIUR), i.e., Dipartimenti di Eccellenza Program (2018–2022)-Dept. of Biology and Biotechnology “L. Spalanzani,” University of Pavia (to A.A.).

Authors' Contributions

Y.J. and C.Z.L. conceived and supervised the experiments. T.S. and J.F.S.H. performed the majority of the analysis with contributions from Q.M.CH., N.B.CH., and ZH.Q.ZH. T.S. and A.A. wrote and revised the manuscript. R.H.D., H.C.ZH., X.M.ZH., M.R.C., Y.ZH.H., X.Y.L., and H.CH. provided and prepared the samples. All authors reviewed the manuscript and gave final approval for publication.

Acknowledgements

We thank Wen Wang for sharing the genome data of *Syncerus caffer*.

References

- Fischer H, Ulbrich F. Chromosomes of the Murrah buffalo and its crossbreeds with the Asiatic swamp buffalo (*Bubalus bubalis*). *J Anim Breed Genet* 1967;**84**(1-4):110–4.
- Iannuzzi L. Standard karyotype of the river buffalo (*Bubalus bubalis* L., 2n = 50). Report of the committee for the standardization of banded karyotypes of the river buffalo. *Cytogenet Cell Genet* 1994;**67**:102–13.
- Cockrill WR. The water buffalo: a review. *Br Vet J* 1981;**137**(1):8–16.
- Pietruszewsky M. The people of Ban Chiang: Bioarchaeology of the 1974 and 1975 skeletons. In: *The International Conference on the Anniversary of the Discovery of the Ban Chiang Site, Udon Thani, Thailand*. 2016.
- Cluttonbrock J. A natural history of domesticated mammals. *Zoolog Afr* 1990;**36**(1):113–20.
- Nagarajan M, Nimisha K, Kumar S. Mitochondrial DNA variability of domestic river buffalo (*Bubalus bubalis*) populations: genetic evidence for domestication of river buffalo in Indian subcontinent. *Genome Biol Evol* 2015;**7**(5):496–503.
- Yindee M, Vlamings BH, Wajjwalku W, et al. Y-chromosomal variation confirms independent domestications of swamp and river buffalo. *Anim Genet* 2010;**41**(4):433–5.
- Lei CZ, Zhang W, Chen H, et al. Two maternal lineages revealed by mitochondrial DNA D-loop sequences in Chinese native water buffaloes (*Bubalus bubalis*). *Asian Australas J Anim Sci* 2007;**20**(4):471–6.
- Lei CZ, Zhang W, Chen H, et al. Independent maternal origin of Chinese swamp buffalo (*Bubalus bubalis*). *Anim Genet* 2007;**38**(2):97–102.
- Kumar S, Nagarajan M, Sandhu JS, et al. Phylogeography and domestication of Indian river buffalo. *BMC Evol Biol* 2007;**7**(1):186.
- Kumar S, Nagarajan M, Sandhu J, et al. Mitochondrial DNA analyses of Indian water buffalo support a distinct genetic origin of river and swamp buffalo. *Anim Genet* 2007;**38**(3):227–32.
- Wang S, Chen N, Capodiferro MR, et al. Whole mitogenomes reveal the history of swamp buffalo: initially shaped by glacial periods and eventually modelled by domestication. *Sci Rep* 2017;**7**(1):4708.
- Zhang Y, Lu Y, Yindee M, et al. Strong and stable geographic differentiation of swamp buffalo maternal and paternal lineages indicates domestication in the China/Indochina border region. *Mol Ecol* 2016;**25**(7):1530–50.
- Colli L, Milanese M, Vajana E, et al. New insights on water buffalo genomic diversity and post-domestication migration routes from medium density SNP chip data. *Front Genet* 2018;**9**:53.
- Whitacre LK, Hoff JL, Schnabel RD, et al. Elucidating the genetic basis of an oligogenic birth defect using whole genome sequence data in a non-model organism, *Bubalus bubalis*. *Sci Rep* 2017;**7**:39719.
- Borghese A. Buffalo production and research. *Ital J Anim Sci* 2005;**5**:2.
- Watson CJ, Burdon TG. Prolactin signal transduction mechanisms in the mammary gland: the role of the Jak/Stat pathway. *Rev Reprod* 1996;**1**(1):1–5.
- Shuai K, Liu B. Regulation of JAK-STAT signalling in the immune system. *Nat Rev Immunol* 2003;**3**(11):900–11.
- O'Shea JJ, Plenge R. JAK and STAT signaling molecules in immunoregulation and immune-mediated disease. *Immunity* 2012;**36**(4):542–50.
- Liu L-L, Fang C, Liu W-J. Identification on novel locus of dairy traits of Kazakh horse in Xinjiang. *Gene* 2018;**677**:105–10.
- Li H, Wang Z, Moore SS, et al. Genome-wide scan for positional and functional candidate genes affecting milk production traits in Canadian Holstein Cattle. 2010. Proc 9th WCGALP, Genetics of Trait Complexes: Lactation - Lecture Sessions: 0535 (Leipzig).

22. Wei X, Li H, Yang J, et al. Circular RNA profiling reveals an abundant circLMO7 that regulates myoblasts differentiation and survival by sponging miR-378a-3p. *Cell Death Dis* 2017;**8**:e3153.
23. Hardie LC, Vandehaar MJ, Tempelman RJ, et al. The genetic and biological basis of feed efficiency in mid-lactation Holstein dairy cows. *J Dairy Sci* 2017;**100**(11):9061–75.
24. Mei C, Wang H, Liao Q, et al. Genetic architecture and selection of Chinese cattle revealed by whole genome resequencing. *Mol Biol Evol* 2018;**35**(3):688–99.
25. Begovich AB, Carlton VEH, Honigberg LA, et al. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 2004;**75**(2):330–7.
26. Bottini N, Musumeci L, Alonso A, et al. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nat Genet* 2004;**36**:337.
27. Kyogoku C, Langefeld CD, Ortmann WA, et al. Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. *Am J Hum Genet* 2004;**75**(3):504–7.
28. Ban Y, Tozaki T, Nakano Y. Association studies of the GPR103 and BCL2L15 genes in autoimmune thyroid disease in the Japanese population. *Front Endocrinol* 2016;**7**:92.
29. Ibeagha-Awemu EM, Peters SO, Akwanji KA, et al. High density genome wide genotyping-by-sequencing and association identifies common and low frequency SNPs, and novel candidate genes influencing cow milk traits. *Sci Rep* 2016;**6**:31109.
30. Brym P, Bojarojćnosowicz B, Oleński K, et al. Genome-wide association study for host response to bovine leukemia virus in Holstein cows. *Vet Immunol Immunopathol* 2016;**175**:24–35.
31. Belhadj Slimen I, Najar T, Ghram A, et al. Heat stress effects on livestock: molecular, cellular and metabolic aspects, a review. *J Anim Physiol Anim Nutr (Berl)* 2016;**100**(3):401–12.
32. Saredi G, Huang H, Hammond CM, et al. H4K20me0 marks post-replicative chromatin and recruits the TONSL–MMS22L DNA repair complex. *Nature* 2016;**534**:714.
33. Ben-Aroya S, Agmon N, Yuen K, et al. Proteasome nuclear activity affects chromosome stability by controlling the turnover of Mms22, a protein important for DNA repair. *PLoS Genet* 2010;**6**(2):e1000852.
34. George I, Wenqi W, Minli W. DNA double strand break repair inhibition as a cause of heat radiosensitization: re-evaluation considering backup pathways of NHEJ. *Int J Hyperthermia* 2008;**24**(1):17.
35. Kantidze OL, Velichko AK, Luzhin AV, et al. Heat stress-induced DNA damage. *Acta Naturae* 2016;**8**(2):75–8.
36. Branzei D, Foiani M. Maintaining genome stability at the replication fork. *Nat Rev Mol Cell Biol* 2010;**11**:208.
37. Filippo JS, Sung P, Klein H. Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem* 2008;**77**(1):229–57.
38. Duro E, Lundin C, Ask K, et al. Identification of the MMS22L–TONSL complex that promotes homologous recombination. *Mol Cell* 2010;**40**(4):632–44.
39. Marai IFM, Haebe AAM. Buffalo's biological functions as affected by heat stress-A review. *Livest Sci* 2010;**127**(2):89–109.
40. Kamprath K, Plendl W, Marsicano G, et al. Endocannabinoids mediate acute fear adaptation via glutamatergic neurons independently of corticotropin-releasing hormone signaling. *Genes Brain Behav* 2009;**8**(2):203–11.
41. Sugo N, Oshiro H, Takemura M, et al. Nucleocytoplasmic translocation of HDAC9 regulates gene expression and dendritic growth in developing cortical neurons. *Eur J Neurosci* 2010;**31**(9):1521–32.
42. Zhang S, Xu H, Liu X, et al. The muscle development transcriptome landscape of ovariectomized goat. *R Soc Open Sci* 2017;**4**(12):171415.
43. Haberland M, Arnold MA, McAnally J, et al. Regulation of HDAC9 gene expression by MEF2 establishes a negative-feedback loop in the transcriptional circuitry of muscle differentiation. *Mol Cell Biol* 2007;**27**(2):518.
44. Mei C, Wang H, Liao Q, et al. Genome-wide analysis reveals the effects of artificial selection on production and meat quality traits in Qinchuan cattle. *Genomics* 2019;**111**(6):1201–8.
45. Haberland M, Montgomery RL, Olson EN. The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat Rev Genet* 2009;**10**:32.
46. Potthoff MJ, Wu H, Arnold MA, et al. Histone deacetylase degradation and MEF2 activation promote the formation of slow-twitch myofibers. *J Clin Invest* 2007;**117**(9):2459–67.
47. Szumlinski KK, Lominac KD, Kleschen MJ, et al. Behavioral and neurochemical phenotyping of Homer1 mutant mice: possible relevance to schizophrenia. *Genes Brain Behav* 2005;**4**(5):273–88.
48. Spellmann I, Rujescu D, Musil R, et al. Homer-1 polymorphisms are associated with psychopathology and response to treatment in schizophrenic patients. *J Psychiatr Res* 2011;**45**(2):234–41.
49. Rietschel M, Mattheisen M, Frank J, et al. Genome-wide association-, replication-, and neuroimaging study implicates HOMER1 in the etiology of major depression. *Biol Psychiatry* 2010;**68**(6):578–85.
50. Jaubert PJ, Golub MS, Lo YY, et al. Complex, multimodal behavioral profile of the Homer1 knockout mouse. *Genes Brain Behav* 2007;**6**(2):141–54.
51. Gerstein H, O'Riordan K, Osting S, et al. Rescue of synaptic plasticity and spatial learning deficits in the hippocampus of Homer1 knockout mice by recombinant Adeno-associated viral gene delivery of Homer1c. *Neurobiol Learn Mem* 2012;**97**(1):17–29.
52. Inoue N, Nakao H, Migishima R, et al. Requirement of the immediate early gene vesl-1S/homer-1a for fear memory formation. *Mol Brain* 2009;**2**(1):7.
53. Klugmann M, Szumlinski KK. Targeting Homer genes using adeno-associated viral vector: lessons learned from behavioural and neurochemical studies. *Behav Pharmacol* 2008;**19**(5–6):485–500.
54. Stiber JA, Zhang Z-S, Burch J, et al. Mice lacking homer 1 exhibit a skeletal myopathy characterized by abnormal transient receptor potential channel activity. *Mol Cell Biol* 2008;**28**(8):2637–47.
55. Chapuis J, Hansmannel F, Gistelincq M, et al. Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. *Mol Psychiatry* 2013;**18**:1225.
56. Wijsman EM, Pankratz ND, Choi Y, et al. Genome-wide association of familial late-onset Alzheimer's disease replicates BIN1 and CLU and nominates CUGBP2 in interaction with APOE. *PLoS Genet* 2011;**7**(2):e1001308.
57. Yu L, Chibnik LB, Srivastava GP, et al. Association of brain DNA methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 with pathological diagnosis of Alzheimer disease. *JAMA Neurol* 2015;**72**(1):15–24.
58. Lee E, Marcucci M, Daniell L, et al. Amphiphysin 2 (Bin1) and T-Tubule biogenesis in muscle. *Science* 2002;**297**(5584):1193.

59. Razzaq A, Robinson IM, McMahon HT, et al. Amphiphysin is necessary for organization of the excitation-contraction coupling machinery of muscles, but not for synaptic vesicle endocytosis in *Drosophila*. *Genes Dev* 2001;15(22):2967–79.
60. Butler MH, David C, Ochoa G-C, et al. Amphiphysin II (SH3P9; BIN1), a member of the Amphiphysin/Rvs family, is concentrated in the cortical cytomatrix of axon initial segments and nodes of ranvier in brain and around T tubules in skeletal muscle. *J Cell Biol* 1997;137(6):1355.
61. Motazacker MM, Rost BR, Hucho T, et al. A defect in the ionotropic glutamate receptor 6 gene (GRIK2) is associated with autosomal recessive mental retardation. *Am J Hum Genet* 2007;81(4):792–8.
62. Ko S, Zhao M-G, Toyoda H, et al. Altered behavioral responses to noxious stimuli and fear in glutamate receptor 5 (GluR5)- or GluR6-deficient mice. *J Neurosci* 2005;25(4):977.
63. Shaltiel G, Maeng S, Malkesman O, et al. Evidence for the involvement of the kainate receptor subunit GluR6 (GRIK2) in mediating behavioral displays related to behavioral symptoms of mania. *Mol Psychiatr* 2008;13:858.
64. Carneiro M, Rubin C-J, Di Palma F, et al. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* 2014;345(6200):1074.
65. Kasai H, Matsuzaki M, Noguchi J, et al. Structure–stability–function relationships of dendritic spines. *Trends Neurosci* 2003;26(7):360–8.
66. Zhang H, Macara IG. The polarity protein PAR-3 and TIAM1 cooperate in dendritic spine morphogenesis. *Nat Cell Biol* 2006;8(3):227–37.
67. Toliaf KF, Bikoff JB, Kane CG, et al. The Rac1 guanine nucleotide exchange factor Tiam1 mediates EphB receptor-dependent dendritic spine development. *Proc Natl Acad Sci U S A* 2007;104(17):7265.
68. Toliaf KF, Bikoff JB, Burette A, et al. The Rac1-GEF Tiam1 couples the NMDA receptor to the activity-dependent development of dendritic arbors and spines. *Neuron* 2005;45(4):525–38.
69. Li M, Luo X-J, Xiao X, et al. Analysis of common genetic variants identifies RELN as a risk gene for schizophrenia in Chinese population. *World J Biol Psychiatr* 2013;14(2):91–9.
70. Abdolmaleky HM, Cheng K-h, Russo A, et al. Hypermethylation of the reelin (RELN) promoter in the brain of schizophrenic patients: a preliminary report. *Am J Med Genet B Neuropsychiatr Genet* 2005;134B(1):60–6.
71. Zhou Z, Hu Z, Zhang L, et al. Identification of RELN variation p.Thr3192Ser in a Chinese family with schizophrenia. *Sci Rep* 2016;6:24327.
72. Hennah W, Thomson P, Peltonen L, et al. Genes and schizophrenia: Beyond schizophrenia: The role of DISC1 in major mental illness. *Schizophr Bull* 2006;32(3):409–16.
73. Mackie S, Millar JK, Porteous DJ. Role of DISC1 in neural development and schizophrenia. *Curr Opin Neurobiol* 2007;17(1):95–102.
74. Kamiya A, Kubo K-i, Tomoda T, et al. A schizophrenia-associated mutation of DISC1 perturbs cerebral cortex development. *Nat Cell Biol* 2005;7(12):1167–78.
75. Katzman A, Alberini CM. NLGN1 and NLGN2 in the prefrontal cortex: their role in memory consolidation and strengthening. *Curr Opin Neurobiol* 2018;48:122–30.
76. Balan S, Yamada K, Hattori E, et al. Population-specific haplotype association of the postsynaptic density gene DLG4 with schizophrenia, in family-based association studies. *PLoS One* 2013;8(7):e70302.
77. Cheng M-C, Lu C-L, Luu S-U, et al. Genetic and functional analysis of the DLG4 gene encoding the post-synaptic density protein 95 in schizophrenia. *PLoS One* 2010;5(12):e15107.
78. Green MR, Sambrook J. *Molecular Cloning: A Laboratory Manual*. 4th ed.. Cold Spring Harbor Laboratory Press; 2012.
79. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009;25(14):1754–60.
80. Nekrutenko A, Taylor J. Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. *Nat Rev Genet* 2012;13(9):667–72.
81. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 2012;8(11):e1002967.
82. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006;2(12):e190.
83. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009;19(9):1655–64.
84. Paradis E. Pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* 2010;26(3):419–20.
85. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* 2007;81(5):1084–97.
86. Chen N, Cai Y, Chen Q, et al. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat Commun* 2018;9(1):2337.
87. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population-structure. *Evolution* 1984;38(6):1358–70.
88. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics* 2011;27(15):2156–8.
89. Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. *Genome Res* 2010;20(3):393–402.
90. Szpiech ZA, Hernandez RD. selscan: An efficient multi-threaded program to perform EHH-based scans for positive selection. *Mol Biol Evol* 2014;31(10):2824–7.
91. Sabeti PC, Varilly P, Fry B, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature* 2007;449:913.
92. Xie C, Mao X, Huang J, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res* 2011;39(suppl 2):W316–W22.
93. Sun T, Shen J, Chen N, et al. Supporting data for “Genomic analyses reveal distinct genetic architectures and selective pressures in buffaloes.” *GigaScience Database* 2019. <http://dx.doi.org/10.5524/100682>.