

# Whole-Genome Resequencing Reveals Adaptation Prior to the Divergence of Buffalo Subspecies

Mostafa Rafiepour<sup>1,2,3</sup>, Esmail Ebrahimie<sup>1,4,5,\*</sup>, Mohammad Farhad Vahidi<sup>6</sup>, Ghasem Hosseini Salekdeh<sup>2</sup>, Ali Niazi<sup>1</sup>, Mohammad Dadpasand<sup>7</sup>, Dong Liang<sup>3</sup>, Jingfang Si<sup>3</sup>, Xiangdong Ding<sup>3</sup>, Jianlin Han<sup>8,9</sup>, Yi Zhang<sup>3,\*</sup>, and Saber Qanbari<sup>10,\*</sup>

<sup>1</sup>Institute of Biotechnology, School of Agriculture, Shiraz University, Iran

<sup>2</sup>Department of System Biology, Agricultural Biotechnology Research Institute of Iran (ABRII), Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran

<sup>3</sup>Key Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture, National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, Agricultural University, Beijing, China

<sup>4</sup>Genomics Research Platform, School of Life Sciences, Melbourne, Victoria, Australia

<sup>5</sup>School of Animal and Veterinary Sciences, The University of Adelaide, South Australia, Australia

<sup>6</sup>Department of Animal Biotechnology, Agricultural Biotechnology Research Institute of Iran (ABRII), Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran

<sup>7</sup>Department of Animal Science, School of Agriculture, Shiraz University, Iran

<sup>8</sup>CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing, China

<sup>9</sup>International Livestock Research Institute (ILR), Nairobi, Kenya

<sup>10</sup>Institute of Genetics and Biometry, Leibniz Institute for Farm Animal Biology (FBN), Dummerstorf, Germany

\*Corresponding authors: E-mails: qanbari@fbn-dummerstorf.de; yizhang@cau.edu.cn; esmaeil.ebrahimie@adelaide.edu.au.

Accepted: 28 October 2020

## Abstract

The application of high-throughput genotyping or sequencing data helps us to understand the genomic response to natural and artificial selection. In this study, we scanned the genomes of five indigenous buffalo populations belong to three recognized breeds, adapted to different geographical and agro-ecological zones in Iran, to unravel the extent of genomic diversity and to localize genomic regions and genes underwent past selection. A total of 46 river buffalo whole genomes, from West and East Azerbaijan, Gilan, Mazandaran, and Khuzestan provinces, were resequenced. Our sequencing data reached to a coverage above 99% of the river buffalo reference genome and an average read depth around 9.2× per sample. We identified 20.55 million SNPs, including 63,097 missense, 707 stop-gain, and 159 stop-loss mutations that might have functional consequences. Genomic diversity analyses showed modest structuring among Iranian buffalo populations following frequent gene flow or admixture in the recent past. Evidence of positive selection was investigated using both differentiation (*F<sub>st</sub>*) and fixation (*P<sub>i</sub>*) metrics. Analysis of fixation revealed three genomic regions in all three breeds with aberrant polymorphism contents on BBU2, 20, and 21. Fixation signal on BBU2 overlapped with the *OCA2-HERC2* genes, suggestive of adaptation to UV exposure through pigmentation mechanism. Further validation using resequencing data from other five bovine species as well as the Axiom Buffalo Genotyping Array 90K data of river and swamp buffaloes indicated that these fixation signals persisted across river and swamp buffaloes and extended to taurine cattle, implying an ancient evolutionary event occurred before the speciation of buffalo and taurine cattle. These results contributed to our understanding of major genetic switches that took place during the evolution of modern buffaloes.

**Key words:** Iranian indigenous buffalo, selection signature, differentiation index, nucleotide diversity, admixture.

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits noncommercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

## Significance

This study describes the results of a high-resolution genome scan to investigate the footprints of historical adaptation in the genome of water buffalo. To the best of our knowledge, this is the most comprehensive effort to study the adaptation in river buffalo populations indigenous to Iran. To this end, we sequenced the whole genomes of 46 Iranian buffaloes and employed available whole-genome sequencing data from five bovine species, including cattle, gayal, bison, banteng, and yak. Additionally, the 90K array genotyping data of river and swamp buffaloes were further used to validate our results. Our findings provide a genome-wide map of past selection in water buffalo, exemplified with several striking selective sweeps during their speciation. We believe that these results will be relevant to a growing number of genomic studies in evolutionary biology and a broad range of Genome Biology and Evolution readers.

## Introduction

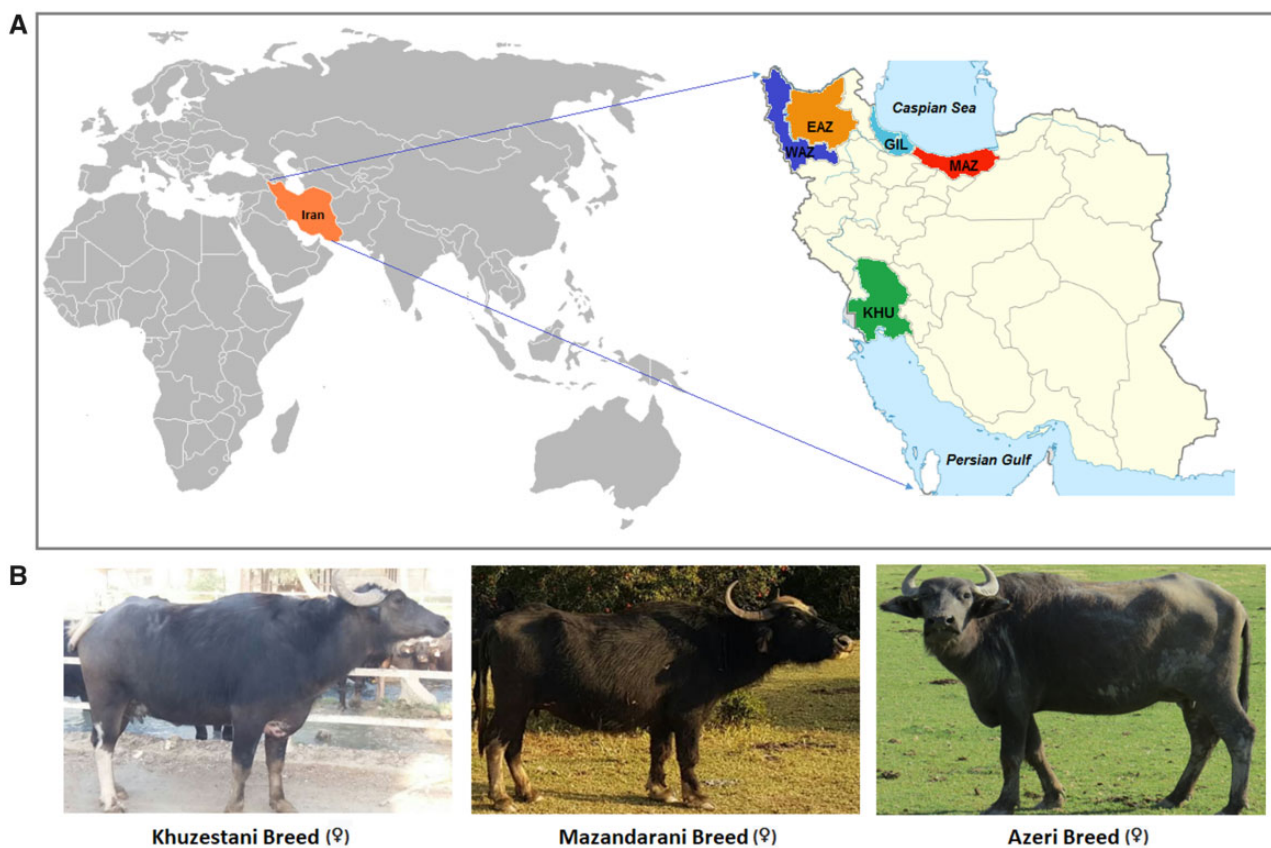
The domestic water buffalo (*Bubalus bubalis*) is an important livestock species for livelihoods in tropical and subtropical regions with hot and humid climates. Water buffaloes are a source of milk, meat, and draught power, having a significant impact on the rural economy (Zhang et al. 2016, 2020; lamartino et al. 2017; Mokhber et al. 2018). They utilize less digestible feeds than cattle, making them easier to be maintained with locally available roughages (Bartocci et al. 1997; Agarwal et al. 2008; Sarwar et al. 2009). The world population of domestic water buffaloes are around 202 million (<http://www.fao.org/faostat/>), of which 174 million are present in Asia, 5 million in Africa (only in Egypt), 3.5 million in America (mostly in Brazil), and some small populations in Europe and Australia (Borghese 2013; Williams et al. 2017; Zhang et al. 2020).

Domestic water buffalo is classified into two major categories: river buffalo (*Bubalus bubalis bubalis*,  $2n = 50$ ) and swamp buffalo (*Bubalus bubalis carabanensis*,  $2n = 48$ ). Cytogenetically, a fusion between river buffalo BBU4 and BBU9 is comparable to swamp buffalo BBU1 (Iannuzzi and Di Meo 2009), otherwise all chromosomes and chromosomal arms are conserved between these two subspecies. Although their taxonomical status is still being debated, a growing body of molecular evidence supports the scenario of two independent domestication events (Kumar et al. 2007; Colli et al. 2018). River buffalo was likely domesticated in the Indian sub-continent and has been spreading to the west across Middle East, Eastern Europe, and the Italian peninsula (Kumar et al. 2007; Mokhber et al. 2018). The swamp buffalo was probably domesticated in a region close to the border between China and Indochina (Zhang et al. 2011, 2016, 2020; Wang et al. 2017) and has been migrating to Thailand, Indonesia, Philippines, central and eastern China (Zhang et al. 2011; Colli et al. 2018; Zhang et al. 2020). The river and swamp buffaloes display distinct morphological and behavioral traits. The swamp buffalo has primarily been used for draught power with no formally recognized breed worldwide. The river buffalo, however, has been selected for dairy production with several recognized breeds which are distributed in a wide geography and diverse agro-

ecosystems across many countries (Yang et al. 2007; Zhang et al. 2020).

Several studies have been conducted to characterize the level and distribution of molecular diversity in domestic water buffalo populations using microsatellite markers (Barker et al. 1997; Arora et al. 2004; Sukla et al. 2006; Triwitayakorn et al. 2006; El-Kholy et al. 2007; Zhang et al. 2007, 2011; Elbeltagy et al. 2008; Mishra et al. 2010; Joshi et al. 2012; Jaayid and Dragh 2013; Özkan Ünal et al. 2014; Uffo et al. 2017). Recent studies on buffaloes have used SNP array genotypes to identify the footprints of past selection (lamartino et al. 2017; Colli et al. 2018; Mokhber et al. 2018; Li et al. 2019). Array genotypes, however, suffer from ascertainment biases (AB) caused by the origin of samples and process applied to discover the SNPs. For example, the Axiom Buffalo Genotyping Array 90K was overrepresented by river buffalo breeds in the SNP discovery panel, therefore this array was affected by certain degree of AB when it was used for genotyping swamp buffaloes (lamartino et al. 2017). The development of a swamp buffalo specific array was then recommended (Colli et al. 2018). Though this array contains 90,000 putative SNPs (lamartino et al. 2017), the final data sets retained only 20,463 SNPs for swamp buffaloes and 52,637 SNPs for both river and swamp buffaloes after quality control (Colli et al. 2018), resulting in a limited resolution of genome coverage by informative SNPs. Moreover, in the absence of a buffalo genome reference assembly at the time, these studies aligned the arrayed SNPs or resequencing reads to the cattle genome assembly as a reference to map the detected variations (lamartino et al. 2017; Colli et al. 2018; Mokhber et al. 2018; Li et al. 2019). The first reference genome assembly of a Mediterranean river buffalo (UOA\_WB\_1) based on the PacBio long-read sequencing data provides a new opportunity to map such genomic variations at chromosomal scale, which allows a depth investigation of the genomic diversity and adaptation of domestic water buffaloes at high resolution of genome coverage (Low et al. 2019).

The buffaloes are one of the vital domestic animals distributed throughout north, northwest, south, and southwest of Iran. This livestock species is raised in several agro-ecological regions with different temperatures, humidity, and altitudes.



**Fig. 1.**—Sampling information of buffaloes in Iran. (A) Geographic location of Iran in the world map (southwest of Asia) and country-wide geographic distributions of the buffaloes. (B) Photos of the three breeds of Iranian indigenous buffaloes with different phenotypes.

Iranian buffalo populations are assumed to be all of river type and divided into three main breeds of Azeri, Khuzestan, and Mezendrani (Ghavi et al. 2012; Pournourali et al. 2015; Safari et al. 2018).

Azeri buffaloes are distributed in the north to northwest of the country (e.g., Gilan, Ardabil, and Azerbaijan provinces) with a population of 119,000 heads. This breed has a small body size and gray body color with white spots (fig. 1) (Safari et al. 2018). Azeri buffaloes are phenotypically similar to Mediterranean river buffaloes, believed to be descended from the same ancestors (Borghese 2013). The lactation duration persists for 200–220 days with an average milk yield of 1,300 kg and fat content of 6.8%. The body weight is 400–600 kg for adult males and 350–500 kg for females. Age at the first calving is around 34 months, and birth weight is 37 kg for male and 30 kg for female calves. Azeri buffaloes are well adapted to average temperatures of 35 °C in the summer and 5 °C in the winter (Borghese 2013; Mokhber et al. 2018; Safari et al. 2018).

Khuzestani buffaloes are distributed in the west and southwest, mainly in Khuzestan province, with a population of 81,000 animals (Mokhber et al. 2018). Khuzestan buffaloes have a large body size of 600–800 kg and are likely the biggest buffaloes in the world (Borghese 2013). The skin color is

often black or dark brown with some white spots on the frontal part or the legs and feet (fig. 1). The lactation period persists for 210–250 days with an average milk yield of 2,107 kg and fat content of 6.23%. The age at the first calving is around 26 months, and birth weight is 42 kg for male and 35 kg for female calves (Borghese 2013; Mokhber et al. 2018; Safari et al. 2018). The breed is adapted to hot climate with an average temperature of 45 °C in the summer and 15 °C in the winter.

Mazandarani breed is distributed in the north of the country, mainly in Mazandaran province and the Miankaleh protected peninsula with a relatively small population of 4,000 buffaloes (fig. 1) (Borghese 2013; Mokhber et al. 2018; Safari et al. 2018). The lactation duration is 220–230 days with an average milk yield of 1,300 kg and fat content of 6.75%. The adult males weigh up to 600 kg whereas adult females may reach to 500 kg. Age at the first calving is around 36 months, and birth weights for male and female calves are 40 and 32 kg, respectively. Mazandarani buffaloes are present in the area with an average temperature of 30 °C in the summer with high humidity and 5 °C in the winter (Aminafshar et al. 2008; Sanjabi et al. 2009).

The farming systems differ among these breeds. In Khuzestan, buffaloes are raised outdoors throughout the

**Table 1**Summary of Resequencing and Mapping Data and Estimates of Nucleotide Diversity ( $\pi$ ) and Coefficient of Inbreeding ( $F_{is}$ ) within Iranian Buffalo Populations

Population	Code	No. of Samples	Depth	Total Reads	Mapping Ratio (%)	Mean $\pi \pm SD$	Mean $F_{is}$
West Azerbaijan	WAZ	7	8.2	230,167,476	99.69	0.0023 $\pm$ 0.0009	-0.0856
East Azerbaijan	EAZ	9	9.3	256,304,521	99.65	0.0023 $\pm$ 0.0009	-0.0384
Gilan	GIL	10	9.0	251,652,834	99.60	0.0022 $\pm$ 0.0009	-0.0066
Mazandaran	MAZ	10	9.1	256,655,094	99.61	0.0021 $\pm$ 0.0009	-0.0319
Khuzestan	KHU	10	10.3	264,462,875	99.57	0.0022 $\pm$ 0.0009	-0.0672

year, often living in the marshes, whereas buffaloes are housed in Azerbaijan and Mazandaran in the autumn and winter (Kianzad 2000; Borghese 2013). Azeri and Khuzestani buffaloes are dominant breeds in Iran in terms of their population sizes (Aminafshar et al. 2008; Mokhber et al. 2018).

In this study, we resequenced and mapped the whole genomes from 46 Iranian river buffaloes to the river buffalo reference genome (UOA\_WB\_1), to study their genomic variability and population genomic structure as well as to localize genomic signatures of past selection. Using whole-genome sequencing information rather than genotypes from preselected SNP panels avoids the problems caused by AB and thus provides additional power to detect selection signatures which were probably missed in previous studies. Also, by using the available sequencing data from other bovine species as well as re-mapping of the SNP genotypes from large populations of swamp and river buffaloes to the river buffalo reference genome, we further validated our results. This study reports a genome-wide map of selection signatures in the buffalo genomes exemplified by several striking selective sweeps during their speciation and domestication.

## Materials and Methods

### Genetic Materials

#### Sequencing Data

Blood samples of 46 buffaloes were collected from West Azerbaijan (7), East Azerbaijan (9), Gilan (10), Mazandaran (10), and Khuzestan (10) provinces (table 1 and fig. 1). Population acronyms are defined as WAZ for West Azerbaijan, EAZ for East Azerbaijan, GIL for Gilan, MAZ for Mazandaran, and KHU for Khuzestan. The buffalo populations in the country are kept by smallholders, as such most of the herds consisted of 2–5 animals whereas some herds have 20–50 and very few own 200–300 buffaloes. Samples were collected from multiple herds, for example, maximal two unrelated animals per herd selected based on available pedigree and owner's information, to avoid inbreeding and to represent most of the available genetic diversity (supplementary table S1, Supplementary Material online).

A further set of sequencing data from other five bovine species (Wu et al. 2018), including cattle ( $N = 74$ ), gayal (23),

bison (12), banteng (8), and yak (7), were retrieved from the Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>; accession codes: PRJNA396672, PRJNA427536 and PRJNA422979) and used in this study.

### Array Genotyping Data

We also used SNP array data from Colli et al. (2018) with 165 river and 181 swamp buffaloes available from the Dryad Repository (<https://datadryad.org/resource/doi:10.5061/dryad.h0cc7>). These individuals were sampled from 15 river and 16 swamp buffalo populations and genotyped with the Axiom Buffalo Genotyping Array 90K from Affymetrix (Iamartino et al. 2017).

### Genome Resequencing and Bioinformatics Pipeline

DNA was extracted from whole blood samples using the DNeasy Blood & Tissue Kit (Qiagen Kit). Following the Illumina standard protocol, genomic libraries with insert size ~350 bp were constructed and sequenced using the HiSeq 2500 instrument.

Raw sequence reads were filtered by removing adaptors using the AdapterRemoval program version 2.1.1 (<https://adapterremoval.readthedocs.io/en/latest/>; Schubert et al. 2016) and low-quality bases using the QualityTrim program (<https://bitbucket.org/arobinson/qualitytrim>) with following parameters “-s -a 20 -l 50 -N 3 -z 1,” where -s was the trim start of sequence, -a the minimum read average quality, -l the minimum read length, -N the maximum N base, and -z the input file type. We kept all default parameters but only set the maximum N base to a higher value at 3 to ensure our data at very stringent quality. Qualified reads were then aligned to the river buffalo reference genome UOA\_WB\_1 (Low et al. 2019) using the MEM algorithm implemented in the Burrows-Wheeler Aligner version 0.7.15-r1140 (<http://bio-bwa.sourceforge.net/bwa.shtml>; Li and Durbin 2009). Reads were sorted and duplicates were marked with MarkDuplicates module implemented in the Picard Tools program version 2.9.0 (<http://broadinstitute.github.io/picard/index.html>). Local realignment was carried out using the RealignerTargetCreator and IndelRealigner tools in the Genome Analysis Toolkit (GATK, version 3.7) (<https://software.broadinstitute.org/gatk/>;



McKenna et al. 2010). Variants were then called using the UnifiedGenotyper in the GATK.

To remove possible false positive SNPs, the hard filters were applied to the raw SNPs using the GATK (supplementary fig. S1, Supplementary Material online) according to the criteria applied by Wu et al. (2018) as follows: QUAL < 30; QualByDepth (QD) < 2.0; RMSMappingQuality (MQ) < 40.0; MappingQualityRankSumTest (MQRankSum) < -12.5; ReadPosRankSumTest (ReadPosRankSum) < -8.0; and HaplotypeScore > 13.0. It is well-known that sex chromosomes (X and Y) follow different evolutionary models because they are not the same in sizes and only share very small portions in genomic homology, therefore the variants on sex chromosomes are usually handled separately from the autosomal SNPs which show very dynamic evolution determined by recombination between homologous chromosomes. In the reference buffalo genome *Bubalus bubalis* UOA\_WB\_1, there is no Y chromosome, it is impossible to align and map the Y-chromosome linked resequencing data. Therefore, in our current analysis, only autosomal bi-allelic SNPs that passed the quality control were used for following analyses.

The same quality control procedures and parameters were applied to the sequencing reads and SNP calling of the 124 bovine animals based on alignment to the *Bos taurus* reference genome assembly UMD3.1 (Zimin et al. 2009), as what was done by Wu et al. (2018). The same quality control procedures and parameters as implemented by Colli et al. (2018) were replicated for the SNP array data of the 165 river and 181 swamp buffaloes whereas the qualified SNPs were mapped based on their coordinates to river buffalo reference genome UOA\_WB\_1 (Low et al. 2019).

#### Common/Shared SNPs across the Five Populations and Population-Specific SNPs

Because sample size affects the number of SNPs per population, we accounted for this effect by random sampling of seven individuals from each population except for WAZ with only seven samples and an average value was taken over 10 times of the sampling using a Python script (<https://github.com/JingfangSi/SnpCountCU>).

#### Annotation of Genetic Variants

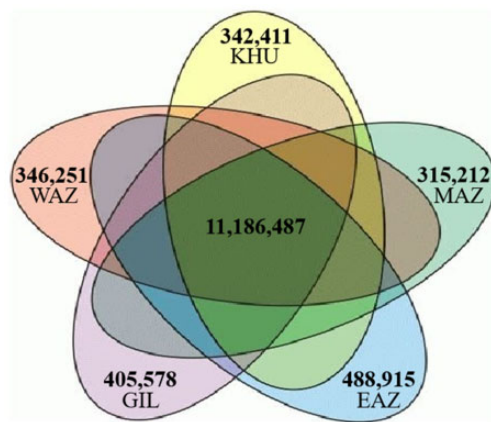
SNPs were functionally classified as InDels, synonymous, missense, upstream, downstream, intergenic, intronic, UTRs, stop-gain, and stop-loss variants in reference to the NCBI *Bubalus bubalis* Annotation Release 101 using the SnpEff version 4.3T (<http://snpeff.sourceforge.net/SnpEff.html>, Cingolani et al. 2012).

#### Estimation of Average Nucleotide Diversity ( $\pi$ ) and Coefficient of Inbreeding ( $F_{is}$ ) within Populations

The VCFtools ([https://vcftools.github.io/man\\_latest.html](https://vcftools.github.io/man_latest.html), Danecek et al. 2011) were used to estimate  $\pi$  in overlapping windows of 40 kb with steps of 20 kb (Rubin et al. 2010). We used a filtration strategy similar to previous studies (Jones, et al. 2018; Wang et al. 2020a; 2020b). To avoid inflated estimates, windows with fewer than 50 or more than 500 SNPs were removed (14,669 out of 619,705 [2.37%] total windows) as these likely consisted of larger structural variants. This quality control procedure did not affect the distribution pattern of SNPs (supplementary fig. S2, Supplementary Material online). The remaining 605,036 windows (97.63%) were used to calculate the average and standard deviation of  $\pi$  values per population. The coefficient of inbreeding ( $F_{is}$ ; Wright 1949), as a measure of lost heterozygosity after sub-structuring in every buffalo population from their common ancestors, was also estimated for individual Iranian buffalo populations. For this, we compared the heterozygosity in each buffalo population on the basis of all SNPs segregating in their common ancestors. By using the command line “-hardy gz” implemented in the PLINK program version 1.9 (<https://www.cog-genomics.org/plink2/>, Purcell et al. 2007; Chang et al. 2015), the observed (Ohet) and expected (Ehet) heterozygosity were obtained, which were then used to estimate  $F_{is}$  following the formula  $F_{is} = 1 - Ohet/Ehet$  for each SNP and the final averaged  $F_{is}$  for each population. The same method was applied to calculate the average observed heterozygosity of each of the 20,463 qualified and mapped SNPs in 327 buffaloes (Colli et al. 2018), which were used to build the plot of heterozygosity distributions of these SNPs along the buffalo chromosomes.

#### Analysis of Population Genetic Structure and Relatedness

The autosomal SNPs were further filtered for missing genotypes and pruning of genotypes at high linkage disequilibrium (LD) with options ‘-indep-pairwise 50 10 0.2 -geno 0’ using the PLINK program, where 50 was as the window size and 10 as the step of each sliding window in kb, 0.2 as the  $r^2$  for LD while 0 meant no missing genotype allowed. The GCTA program version 1.26 (<http://cns.genomics.com/software/gcta/#Overview>, Yang et al. 2011) was used to perform principal component analysis (PCA) from qualified autosomal SNPs to examine the genetic relationship among the five Iranian buffalo populations. Also, the identity-by-state (IBS) matrix was generated based on the proportion of alleles shared between 46 buffaloes after the transformation of SNPs data into PED format for the option ‘-distance square ibs’ using the PLINK program. We further examined the genetic structure among the five populations using the ADMIXTURE program version 1.3 (<http://software.genetics.ucla.edu/admixture>). The model built in the program includes a cross-validation method, allowing the identification of an optimal value of potential



**Fig. 2.**—Venn diagram showing the numbers of shared SNPs across the five populations and private SNPs for each population of Iranian buffaloes along with a tabled summary of the variants detected.

Type of variants	No. of total variants
SNPs	20,546,654
InDels	3,687,860
Synonymous	103,184
Missense	63,097
Upstream/downstream	2,022,922
Intergenic	11,119,079
Intronic	8,948,819
UTRs	170,012
Stop gain	707
Stop loss	159

ancestries ( $K$ ) among all samples (Alexander et al. 2009). We also employed the TreeMix software version 1.13 (<https://bitbucket.org/nygcresearch/treemix/wiki/Home>, Pickrell and Pritchard 2012) to construct a ML phylogenetic tree for all the five Iranian buffalo populations with 1 to 4 migration events to examine their split and subsequent gene flow.

### Detection of Selective Sweeps

We employed the fixation index ( $F_{st}$ ) as a measure of population differentiation due to genetic substructuring (Weir and Cockerham 1984) and also nucleotide diversity ( $\pi$ ) metrics to detect selection signatures using the VCFtools. In the first step, we explored the differentiation of SNPs between the five Iranian buffalo populations and also between KHZ and other four populations (MAZ, GIL, WAZ, and EAZ). Then to reduce locus-to-locus variation in the inference of selection,  $F_{st}$  values of individual SNPs were averaged in 40 kb windows with steps of 20 kb across the genomes (Rubin et al. 2010) with following parameters “–weir-fst-pop pop1.txt –weir-fst-pop pop2.txt –fst-window-size 40000 –fst-window-step 20000 –out Fst\_pop1\_pop2\_sliding.txt.” The window-based  $F_{st}$  values were then normalized using Z-transformation (ZFst) for plotting their distributions along the chromosomes. In the second step, we searched the genomic regions with high degrees of fixation. For this, we compared  $\pi$  along the chromosomes.  $\pi$  values were estimated in overlapping windows of 40 kb with steps of 20 kb (Rubin et al. 2010) with following parameters “–window-pi 40000 –window-pi-step 20000” and were then Z-transformed for plotting their distributions along the chromosomes. Based on the alignment of sequence synteny between the river buffalo chromosomes with the bovine genome sequence assembly (Btau\_4.0) (Amaral et al. 2008) and their gene coordinates, the  $\pi$  values of qualified SNPs of the five bovine species mapped onto BTA2 and BTA23 corresponding to BBU2 (Wu et al. 2018) were calculated and the Z-transformed  $\pi$  ( $Z\pi$ ) were used to plot for their

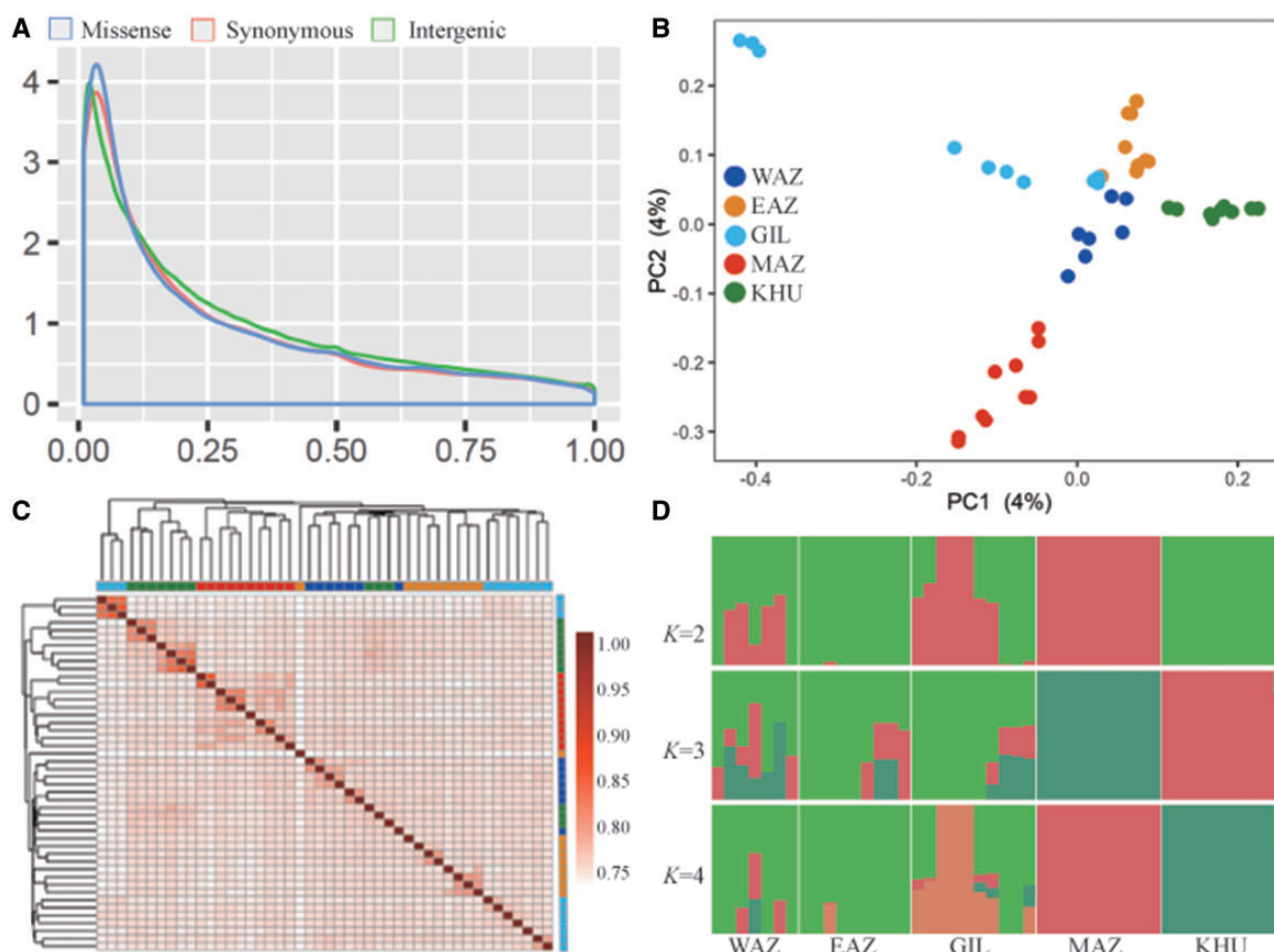
distributions. To confirm the result of the five bovine species, the corresponding genomic regions on BTA23 from taurine cattle reference genome UMD3.1 (Zimin et al. 2009) released in 2014 (from 2,475,803 at the start of *PRIM2* to 3,058,878 at the end of *ZNF451*) and ARS-UCD1.2 (Rosen et al. 2020) released in 2018 (from 2,560,544 to 3,144,946) were aligned to examine their genomic completeness and the order of annotated genes.

### Results

Our clean data ranged from 203,518,115 to 323,251,933 reads per sample, reached to a coverage above 99% of the river buffalo reference genome UOA\_WB\_1 and an average read depth around  $9.2\times$  per sample (table 1). After the first filtering process, 20,546,654 autosomal SNPs were obtained from all 46 Iranian buffaloes. These SNPs were distributed in an average density of one SNP in around every 150 bp of the river buffalo reference genome. The number of common/shared SNPs were 11.186 million (54.44%) across the five Iranian buffalo populations while the numbers of population specific SNPs varied between 315,212 (1.53%) in MAZ and 488,915 (2.38%) in EAZ. Most SNPs were located in intergenic (54.12%) and intronic (43.55%) regions. Furthermore, 9.85% of the SNPs were mapped within the regions in upstream or downstream of transcription starting or ending sites while only 0.31% were annotated as missense and stop gain/loss mutations (fig. 2).

### Analysis of Genomic Diversity and Differentiation

The comparison of SNP frequency profiles revealed no substantial difference among the Iranian buffalo populations (data not shown). In general, the distribution of SNP frequencies across all five populations showed a marked overrepresentation of infrequent alleles (fig. 3A), consistent with the patterns observed for high-quality genotyping or



**Fig. 3.**—The site frequency spectrum (SFS) and genomic differentiation based on the resequencing data of 46 Iranian buffaloes. (A) SFS is represented for missense, synonymous, and intergenic SNPs based on the annotations of 20,546,654 autosomal SNPs. (B) PCA plot of five Iranian buffalo populations. (C) Visualization of the IBS distances between 46 Iranian buffaloes. (D) Genetic structuring of five Iranian buffalo populations. Acronyms are defined as West Azerbaijan (WAZ), East Azerbaijan (EAZ), Gilan (GIL), Mazandaran (MAZ), and Khuzestan (KHU).

resequencing data in many other organisms including humans (Li et al. 2010), cattle, and chickens (Qanbari et al. 2014, 2019). Genome-wide nucleotide diversity ( $\pi$ ) estimates also showed no difference among the five populations while very low and negative coefficients of inbreeding ( $F_{is}$ ) indicated the presence of little outbreeding among the five populations, an indication of traditional indiscriminate mating practiced in Iranian buffaloes (table 1).

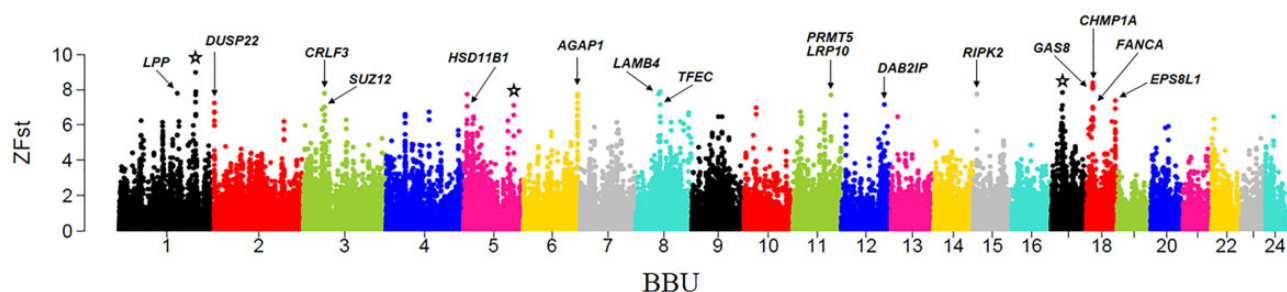
After the second filtering process, 1,409,993 SNPs with missing genotypes and 17,702,880 SNPs with highly linked genotypes were removed, only 1,433,841 SNPs were retained for following analyses. PCA detected a moderate separation of KHU from other four populations at PC1 while MAZ further differentiated from the remaining three populations at PC2 (fig. 3B). The IBS distances grouped most buffaloes into the expected sources of their corresponding breeds (fig. 3C). The population-genetic substructuring based on the admixture analysis identified a very frequent gene flow between the

two populations sampled from East (EAZ) and West (WAZ) Azerbaijan provinces, whereas most GIL individuals were highly mixed with EAZ and WAZ, validating all these three populations to be of Azeri breed. Similar to the PCA result, MAZ and KHU carried rather homogenous but independent genetic backgrounds with no gene flow from Azeri breed (fig. 3D). The TreeMix result indicated the direction of gene flow from MAZ to GIL and KHU and from EAZ to WAZ and KHU (supplementary fig. S3, Supplementary Material online).

### Genome-Wide Selection Signatures

#### Analysis of Differentiation

In the first step, we sought to estimate the magnitude of inter-population differentiation using information from SNP frequencies. However, all genome-wide scans based on the  $Z_{Fst}$  values between the five Iranian buffalo populations revealed no clear signal (supplementary fig. S4,



**Fig. 4.**—Genome-wide distributions of ZFst values between KHU and other four populations (WAZ, EAZ, MAZ, and GIL). Signals marked by stars represent regions lacking annotated genes.

**Table 2**

A Summary of Candidate Genes Identified in Differentiation Analysis

BBU	Bin_Start	Bin_End	No. of SNPs	Fst	ZFst	Genes	Function/Association <sup>a</sup>
1	123,360,001	123,400,000	186	0.370	7.775	<i>LPP</i>	Cell adhesion
2	200,001	240,000	146	0.347	7.239	<i>DUSP22</i>	Cellular response to epidermal growth factor
2	200,001	240,000	146	0.347	7.239	<i>IRF4</i>	Defense response to cell differentiation
3	45,240,001	45,280,000	73	0.338	7.034	<i>SUZ12</i>	Chromatin DNA binding
3	45,260,001	45,300,000	124	0.371	7.782	<i>CRLF3</i>	Transition of mitotic cell cycle
5	7,460,001	7,500,000	151	0.370	7.765	<i>TRAF3IP3</i>	Adapter molecule
5	7,440,001	7,480,000	113	0.339	7.053	<i>HSD11B1</i>	Lung development
6	114,620,001	114,720,000	1,313	0.370	7.626	<i>AGAP1</i>	Actin cytoskeleton
8	49,040,001	49,080,000	348	0.370	7.756	<i>LAMB4</i>	Cell adhesion—embryonic development
8	52,140,001	52,180,000	40	0.343	7.150	<i>TFEC</i>	Cellular response to heat
11	80,880,001	80,920,000	138	0.367	7.688	<i>LRP10</i>	Inner ear development
11	80,880,001	80,920,000	138	0.367	7.688	<i>PRMT5</i>	Endothelial cell activation—liver regeneration
11	80,880,001	80,920,000	138	0.367	7.688	<i>RBM23</i>	Pre-mRNA splicing process
11	80,880,001	80,920,000	138	0.367	7.688	<i>REM2</i>	Gated calcium channel activity
12	92,220,001	92,260,000	279	0.344	7.170	<i>DAB2IP</i>	Layer formation in cerebral cortex—neuron projection morphogenesis
15	8,280,001	8,320,000	162	0.370	7.758	<i>RIPK2</i>	Adaptive immune response—apoptotic process
18	14,140,001	14,180,000	68	0.396	8.372	<i>CDK10</i>	Cell projection organization—regulation of cilium assembly
18	14,140,001	14,180,000	68	0.396	8.372	<i>CHMP1A</i>	Endosome transport—midbody abscission—nucleus organization
18	14,320,001	14,360,000	104	0.388	8.172	<i>DEF8</i>	Intracellular signal transduction—regulation of bone resorption
18	14,200,001	14,240,000	100	0.337	7.017	<i>FANCA</i>	Female and male gonad development
18	14,200,001	14,240,000	100	0.337	7.017	<i>SPIRE2</i>	Cleavage furrow formation
18	14,160,001	14,200,000	78	0.390	8.223	<i>ZNF276</i>	Zinc ion binding
18	14,340,001	14,380,000	123	0.384	8.087	<i>GAS8</i>	Brain development—sperm motility
18	62,260,001	62,300,000	202	0.353	7.367	<i>EPS8L1</i>	Regulation of ruffle assembly
18	62,260,001	62,300,000	202	0.353	7.367	<i>PPP1R12C</i>	Actin cytoskeleton

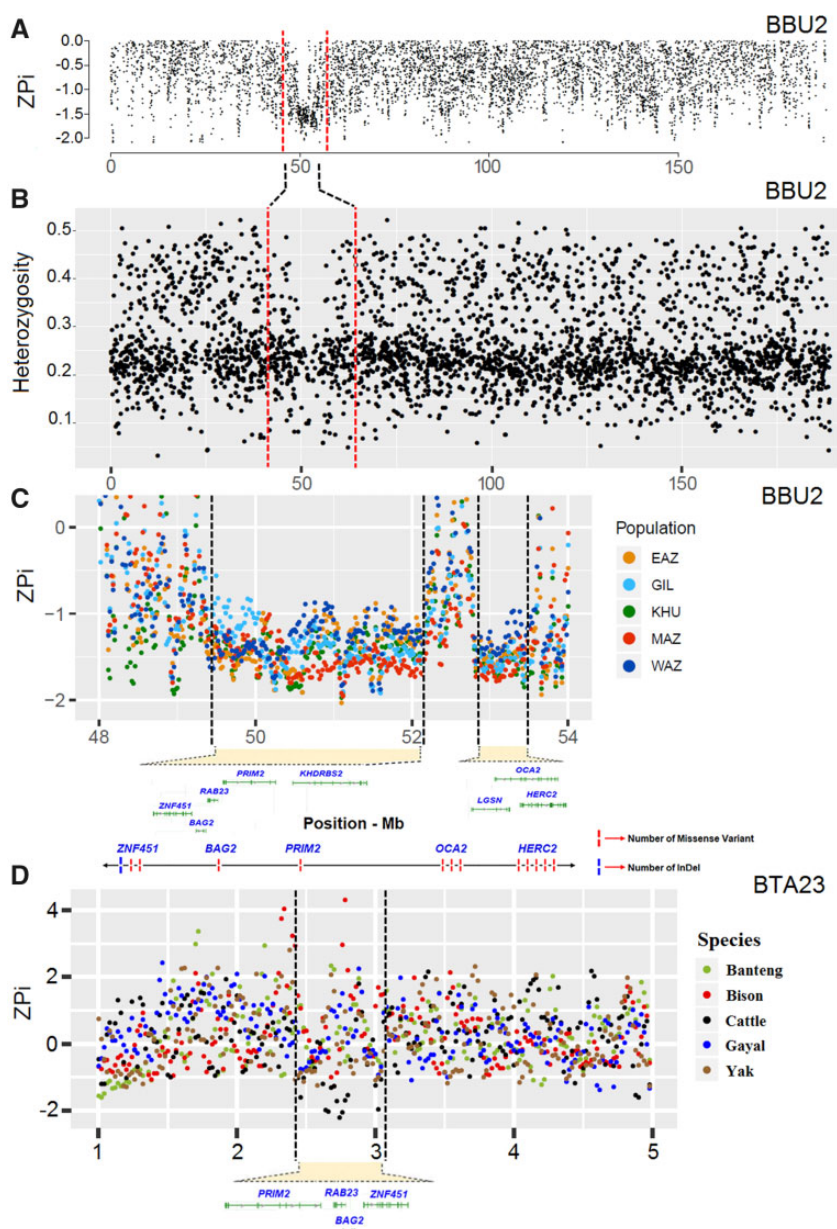
<sup>a</sup>Gene functions derived from either <https://www.uniprot.org>, <https://www.genecards.org>, or <https://www.ncbi.nlm.nih.gov/gene>.

Supplementary Material online). KHU has an isolated geographic distribution and distinct profile of phenotypes relative to other four buffalo populations/breeds, and was also clearly separated in PCA, IBS distances and admixture analysis at  $K = 4$  (fig. 3). We therefore examined the differentiation between KHU and other four populations/breeds (MAZ, GIL, WAZ, and EAZ). Given the small sample size ( $n = 10$ ) available in KHU, this analysis was designated only to detect stand-alone Fst signals emerging from possible cases of extreme divergences. To this end, Fst values were estimated locally

for each SNP and then averaged in 40 kb windows with steps of 20 kb along the chromosomes (fig. 4).

Genome-wide analysis of differentiation identified 18 regions with the highest ZFst values  $> 7.0$  and they distributed nonuniformly across the genomes. Annotation of these regions using the UOA\_WB\_1 river buffalo reference genome (NCBI *Bubalus bubalis* Annotation Release 101) was performed to locate candidate genes within the selected regions. List of candidate genes identified is presented in table 2.





**Fig. 5.**—(A) The pattern of nucleotide diversity for the candidate region localized on BBU2 in Iranian indigenous buffaloes. (B) The pattern of observed heterozygosity in the candidate region of BBU2 derived from 90K SNP array genotypes from both river and swamp buffaloes. (C) A detailed graphical representation of the candidate region of BBU2 in Iranian indigenous buffaloes with the location of candidate genes and mutations listed underneath. (D) The pattern of nucleotide diversity based on sequencing data in the homologous candidate region localized on BTA23 of *Bos taurus*.

### Analysis of fixation

In a further step, we employed nucleotide diversity ( $\pi$ ) metric to measure the degree of fixation across the genomes. In accordance with differentiation analysis,  $\pi$  values were estimated in overlapping windows of 40 kb with steps of 20 kb. Based on the plot of window-based Z $\pi$  values along all chromosomes, we observed the most strikingly contrasted distribution patterns in three genomic regions located on BBU2, 20, and 21 displaying aberrant polymorphism contents with the Z $\pi$  values less than  $-2$  (fig. 5A, supplementary figs. S5

and S6, Supplementary Material online). In total, 363 windows exceeded this threshold representing 0.3% of all windows across the genome. Based on the annotation of these regions on the UOA\_WB\_1 river buffalo reference genome (NCBI *Bubalus bubalis* Annotation Release 101), 19 candidate genes were detected within the overlapping fixation signals (table 3, Tables S2 and S3, Supplementary Material online).

The strongest signal of fixation was localized on BBU2 spanning over 49.3–53.6 Mb (fig. 5A). The distributions of observed heterozygosity of SNPs from the 327 buffaloes

**Table 3**

List of Candidate Genes Overlapping the Fixed Regions on BBU2

Bin_Start	Bin_End	Size (bp)	No. of SNPs	Pi (%)	ZPi	Genes	Function/Association <sup>a</sup>
49,280,001	49,320,000	40,000	185	0.08	-0.484	<i>ZNF451</i>	Growth factor beta receptor signaling pathway—cellular response to heat
49,300,001	49,340,000	40,000	120	0.04	-1.260	<i>BAG2</i>	Protein stabilization—protein folding
49,320,001	49,360,000	40,000	138	0.05	-0.992	<i>RAB23</i>	Craniofacial suture morphogenesis
49,460,001	49,820,000	360,000	940	0.03	-1.459	<i>PRIM2</i>	DNA replication
51,420,001	52,160,000	740,000	2,291	0.02	-1.577	<i>KHDRBS2</i>	Regulation of mRNA splicing
52,900,001	52,980,000	80,000	109	0.02	-1.696	<i>LGSN</i>	Nitrogen metabolic process—glutamine biosynthetic process
53,000,001	53,340,000	40,000	682	0.02	-1.629	<i>OCA2</i>	Eye pigment biosynthetic process—melanin biosynthetic process—melanocyte differentiation
53,320,001	53,580,000	26,000	588	0.03	-1.556	<i>HERC2</i>	Spermatogenesis—like growth factor receptor—blue eye color and blond hair

<sup>a</sup>Gene functions derived from either <https://www.uniprot.org>, <https://www.genecards.org> or <https://www.ncbi.nlm.nih.gov/gene>.

(Colli et al. 2018) mapped to this region also displayed a very similar pattern (fig. 5B). The candidate genes underlying this region are oculocutaneous albinism type 2 (*OCA2*) and HECT domain and RCC1-like domain-containing protein 2 (*HERC2*) (fig. 5C). *OCA2* is known as the melanocyte-specific transporter protein, a major determinant of melanogenesis and mammalian pigmentary system. *HERC2* gene functions as the repressor of *OCA2* expression (Kayser et al. 2008). The polymorphisms in *OCA2-HERC2* genes were reported to control skin, hair, and eye pigmentation in humans, mice, and zebrafish models (reviewed in Sturm 2009; Donnelly et al. 2012; Yang et al. 2016). The *OCA2* was also suggested to be involved in meat color and quality traits in pigs (Steibel et al. 2011). The fixation signal on BBU2 further overlapped the KH RNA Binding Domain Containing, Signal Transduction Associated 2 (*KHDRBS2*) and RAS-associated protein RAB23 (*RAB23*) genes. Fixation signals on BBU20 (supplementary fig. S4, Supplementary Material online) and BBU21 (supplementary fig. S5, Supplementary Material online) were linked with genes of small nuclear ribonucleoprotein polypeptide N (SNRPN) upstream reading frame (SNURF) (*SNRPN/SNURF*) involved in RNA processing and ELKS/RAB6-interacting/CAST family member 2 (*ERC2*) functioning as regulators of neurotransmitter release, respectively.

The distributions of Z-transformed Pi (ZPi) values of qualified SNPs of the five bovine species (Wu et al. 2018) mapped to the homologous regions on BTA2 (BTA2: 0.5–1.0 Mb corresponding to BBU2: 52.7–53.6 Mb) and BTA23 (BTA23: ~2.4–3.1 Mb to BBU2: ~49.3–52.0 Mb and ~52.7–53.6 Mb) identified reduced variability only at the larger sweep in taurine cattle (BBU2: ~49.3–52.0 Mb) (fig. 5D). A close examination of the homologous genomic DNA sequences of the river buffalo reference genome UOA\_WB\_1 (BBU2: ~49.3–52.0 Mb) and the two versions of *Bos taurus* reference genome UMD3.1 (Zimin et al. 2009) and ARS-UCD1.2 (BTA23: ~2.4–3.1 Mb, Rosen et al. 2020) indicated no gap in this region for UOA\_WB\_1 and ARS-UCD1.2 but 13 gaps in

UMD3.1, however, the three genomic regions shared the same order of the four annotated genes (*PRIM2*, *RAB23*, *BAG2*, and *ZNF451*, fig. 5D). The alignment of this 583-kb long genomic region from UMD3.1 (from 2,475,803 at the start of *PRIM2* to 3,058,878 at the end of *ZNF451*) and ARS-UCD1.2 (from 2,560,544 to 3,144,946) on BTA23 indicated that all the 14 fragments in UMD3.1 were completely mapped to ARS-UCD1.2 in the same orientations and the 13 gaps in 17–1,048 bp in UMD3.1 were all closed in ARS-UCD1.2. This finding suggested that the shared pattern of reduced nucleotide diversity in this sweep between the buffaloes and taurine cattle, which had the largest sample size (74) among the five bovine species, was rather reliable. Nevertheless, this observation warrants further validation as we could not completely rule out possible InDel artifacts in the long-read-based ARS-UCD1.2 reference genome.

## Discussion

This study primarily aimed at assessing the magnitude of genomic diversity among the Iranian indigenous buffalo populations/breeds which were found to carry 20,546,654 autosomal SNPs throughout their genomes, of which 11.186 million were shared by all Iranian buffaloes (fig. 2 and table 1). Their population-genetic substructuring was also evaluated by various methodologies, including PCA, IBS distance, and admixture. These analyses based on the SNP information derived from the whole-genome resequencing data indicated a moderate genetic differentiation and substructuring among the Iranian indigenous buffaloes with Azeri, Khuzestani, and Mazandrani breeds being clearly separated from each other. Nevertheless, there was frequent gene flow among the three populations belonging to Azeri buffaloes distributed in northwestern Iran (fig. 3). This observation was consistent with phenotypic differences among the Iranian buffaloes (Ghavi et al. 2012; Pournourali et al. 2015; Safari et al. 2018) and also with a recent study on two Iranian

breeds of Azeri and Khuzestani buffaloes (Mokhber et al. 2018) genotyped using the Affymetrix Axiom Buffalo Genotyping Array 90K (Iamartino et al. 2017).

A further objective of the study was to identify genomic signatures of adaptation to diverse geographical and climatic conditions. The Khuzestan buffaloes are dispersed in the southwest of Iran, mainly in Khuzestan province with very high average temperatures of 45 and 15 °C in the summer and winter, respectively. However, Azeri and Mazandaran breeds live in the north and north-west with almost identical climatic conditions and at average temperature of 30 °C in the summer and 5 °C in the winter (Sanjabi et al. 2009; Mokhber et al. 2018). These three breeds also differ in terms of morphological and physiological characteristics, for example, Khuzestan buffaloes have larger body and higher milk yield than Azeri and Mazandarani breeds (fig. 1) (Safari et al. 2018).

Evidence of positive selection was investigated through two approaches. First, we measured local divergence using *F<sub>st</sub>* metric (Weir and Cockerham 1984) to identify genomic regions evolved for diverse adaptation (Bonin et al. 2007). Considering the geographic distribution, phenotypic, and phylogenetic patterns among the five Iranian populations, we designed the *F<sub>st</sub>* analysis to visualize the differentiation between KHZ and other four populations including WAZ, WAZ, GIL, and MAZ. We focused only on the top 18 genomic regions with the highest Z*F<sub>st</sub>* values (fig. 4). Although such a pattern of differentiation was expected, it was different from the motivation of this analysis to locate the stand-alone cases of differentiation. After annotating these regions on the UOA\_WB\_1 genome assembly, a total of 25 candidate genes were recognized with diverse biological functions involved in cellular responses to heat and epidermal growth factor, immune responses, embryonic development, cell adhesion, and cytoskeleton (table 2). This experiment, however, failed to locate a standalone case of gene divergence due probably to the confounding effects of small sample size and genetic drift in shaping the emerged differentiation profile.

Second, nucleotide diversity metric (*P<sub>i</sub>*) was employed to measure the degree of fixation along the chromosomes. Fixation analysis revealed genomic regions with aberrant patterns of polymorphisms on BBU2, 20, and 21 (fig. 5, supplementary figs. S5 and S6, tables S2 and S3, Supplementary Material online). The observed signal on BBU2 (fig. 5A) showed the strongest fixation pattern, therefore we decided to explore it further. To this end, we first examined the degree of differentiation locally among the Iranian buffaloes to learn if these populations have adapted to different agro-ecological zones by diverse haplotypes in these genomic regions. *F<sub>st</sub>* analysis revealed no indication of local divergence among the Iranian buffalo populations (supplementary fig. S7A, Supplementary Material online), likely suggesting an adaptation prior to domestication and migration of riverine buffalo to the Iranian Plateau. We further explored the candidate regions by employing the data reported by Colli et al. (2018)

and Deng et al. (2019). The SNPs from Colli et al. (2018) were originally mapped against the cattle reference genome whereas Deng et al. (2019) aligned the SNPs to the recently published river buffalo UOA\_WB\_1 genome assembly (Low et al. 2019). This enabled us to assess the map coordinates of the SNPs through intersecting the identical IDs in both studies. We then estimated the observed heterozygosity per SNP as an indicator of local nucleotide diversity and compared these SNPs of both river and swamp buffaloes with our sequence data. This analysis confirmed an elevated homozygosity pattern perfectly overlapping the candidate region on BBU2 (fig. 5B). This may suggest either convergent evolution in both subspecies or a unique evolutionary event occurred before the divergence of river and swamp buffalo subspecies. To explore this further, we examined the degree of local differentiation between the two subspecies and found some local haplotype similarity (supplementary fig. S7B, Supplementary Material online), most likely suggesting an adaptation took place prior to the divergence of river and swamp buffalo subspecies.

A close look at the BBU2 sweep (fig. 5C) revealed two separate signals emerging from the region. An extensively swept region spanning ~49.3–52.0 Mb overlaps the genes of *RAB23*, *PRIM2*, and *KHDRBS2*, which play roles in the regulation of diverse cellular functions associated with intracellular membrane trafficking, the replication of DNA and the regulation of alternative splicing, respectively. The second sweep in the region is placed adjacently over ~52.7–53.6 Mb and centralized over the *OCA2* and *HERC2* genes. The *OCA2-HERC2* system is a bold example of convergent evolution in mammalian species. Through a close interaction with MC1R, they play a key role in skin and hair pigmentation for adaptive response to UV exposure (reviewed in Sturm 2009; Donnelly et al. 2012; Yang et al. 2016). Consistent with the human model, molecular signatures of adaptation for pigmentation traits have been extensively reported in domestic animals (Rubin et al. 2012; Qanbari et al. 2014 and 2019; Liang et al. 2020), we therefore suggested the *OCA2-HERC2* mechanism as one of the candidates for driving the evolution of modern buffaloes.

The strong fixation signal or significantly reduced DNA variability in the homologous genomic regions between buffalo (BBU2: ~49.3–52.0 Mb) and taurine cattle (BTA23: ~2.4–3.1 Mb) implies that both buffalo and taurine cattle have likely adapted to an unknown quality locally. This adaptation may have evolved before their speciation and was conserved even after their chromosomal rearrangements. Alternatively, this shared pattern may be the result of a convergent evolution after they diverged into separate species. In contrast, the *OCA2-HERC2* sweep (BBU2: 52.7–53.6 Mb corresponding to BTA2: 0.5–1.0 Mb) was not observed in any of the five bovine species (supplementary fig. S8, Supplementary Material online), denoting a locally divergent evolution of the modern buffaloes. However, given the persistence of



the *OCA2-HERC2* sweep across all swamp and river buffaloes, our results likely indicated an evolutionary event occurred prior to the divergence of swamp and river buffalo subspecies.

## Conclusions

This is the first genetic variant discovery study with an average read depth of 9.2× in Iranian indigenous buffaloes. More than 20.55 million SNPs were identified, including 63,097 missense, 707 stop-gain, and 159 stop-loss mutations. Patterns of genetic differentiation and substructuring among five Iranian buffalo populations indicated a clear separation of Khuzestani, Azeri, and Mazandrani breeds from each other. Fixation analysis revealed regions with aberrant patterns of polymorphisms across the genomic regions on BBU2, 20, and 21. We found molecular signatures of strong fixation in the region of *OCA2-HERC2* genes, suggestive of an adaptation through the pigmentation mechanism. Given the persistence of the fixation pattern observed locally at *OCA2-HERC2* genes across all swamp and river buffaloes, we postulated that an ancient evolutionary event has taken place before the divergence of the two domestic water buffalo subspecies. These results contributed to the understanding of major genetic switches that took place during the evolution of modern buffaloes.

## Consent for Publication

All authors have read and approved the article.

## Ethics Approval and Consent to Participate

All activities and procedures involving the animals were approved by the Animal Care and Use committee and was done according to the local guidelines in the Agricultural Biotechnology Research Institute of ABRII-North Branch (Rasht), Iran and with consent of the animal owners.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

We thank to F. Behfarjam, M. R. Ghaffari, and N.A.K. Sima for providing technical support and to S. Ardestani and A.H. Sadri for their assistance in analyzing data. This project was financially supported by National Natural Scientific Foundation of China (31561143010) and fund from the Agricultural Biotechnology Research Institute of Iran (ABRII). Yi Zhang was partially supported by the Program for Changjiang Scholars and Innovative Research in University (IRT1191).

## Author Contributions

M.R. conducted the study, analyzed, and visualized the results. S.Q., Y.Z., M.R., and J.H. conceptualized and supervised the study, undertook the project management and wrote the manuscript. S.Q., J.H., Y.Z., M.F.V., and G.H.S. contributed to the overall design and the development of the methods, participated in data interpretation and revised the manuscript. D.L. and J.S. participated in data analysis. M.F.V., G.H.S., E.E., A.N., M.D., and X.D. contributed in data provision.

## Data Availability

The resequencing data of the 46 Iranian river buffaloes have been deposited at the NCBI Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>; accession code: PRJNA633724). The link for reviewers' examination is at: <https://dataview.ncbi.nlm.nih.gov/object/PRJNA633724?reviewer=k71innjrcunl48s5oardg6mmkc>. The SNP data in vcf format have been deposited to the Dryad at: [https://datadryad.org/stash/share/TDDOy7l45dw7JWmUE9LQf0Ez5emZiMkA5-i\\_VptYjNk](https://datadryad.org/stash/share/TDDOy7l45dw7JWmUE9LQf0Ez5emZiMkA5-i_VptYjNk).

## Literature Cited

- Agarwal N, Kamra DN, Chatterjee PN, Kumar R, Chaudhary LC. 2008. *In vitro* methanogenesis, microbial profile and fermentation of green forages with buffalo rumen liquor as influenced by 2-bromoethanesulphonic acid. *Asian Australas J Anim Sci.* 21(6):818–823.
- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9):1655–1664.
- Amaral ME, et al. 2008. A first generation whole genome RH map of the river buffalo with comparison to domestic cattle. *BMC Genomics.* 9:631.
- Aminafshar M, Amirinia C, Torshizi RV. 2008. Genetic diversity in buffalo population of Guilan using microsatellite markers. *J Ani Vet Adv.* 7:1499–1502.
- Arora R, Lakhchaura BD, Prasad RB, Tantia MS, Vijn RK. 2004. Genetic diversity analysis of two buffalo populations of northern India using microsatellite markers. *J Anim Breed Genet.* 121(2):111–118.
- Barker JS, et al. 1997. Genetic diversity of Asian water buffalo (*Bubalus bubalis*): Microsatellite variation and a comparison with protein-coding loci. *Anim Genet.* 28(2):103–115.
- Bartocci S, Amici A, Verna M, Terramocchia S, Martillotti F. 1997. Solid and fluid passage rate in buffalo, cattle and sheep fed diets with different forage to concentrate ratios. *Livestock Prod Sci.* 52(3):201–208.
- Bonin A, Nicole F, Pompanon F, Miaud C, Taberlet P. 2007. Population adaptive index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation. *Conserv Biol.* 21(3):697–708.
- Borghese A. 2013. Buffalo livestock and products in Europe. *Buffalo Bull.* 32(Special Issue 1):50–74.
- Chang CC, et al. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4(1):7.
- Cingolani P, et al. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; Iso-2; Iso-3. *Fly* 6(2):80–92.
- Colli L, et al. 2018. New insights on water buffalo genomic diversity and post-domestication migration routes from medium density SNP chip data. *Front Genet.* 9:53.



- Danecek P, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.
- Deng T, et al. 2019. Genome-wide SNP data revealed the extent of linkage disequilibrium, persistence of phase and effective population size in purebred and crossbred buffalo populations. *Front Genet.* 9:688.
- Donnelly MP, et al. 2012. A global view of the OCA2-HERC2 region and pigmentation. *Hum Genet.* 131(5):683–696. p
- Elbeltagy AR, et al. 2008. Biodiversity in Mediterranean buffalo using two microsatellite multiplexes. *Livestock Sci.* 114(2–3):341–346.
- El-Kholy AF, Hassan HZ, Amin AM, Hassanane MS. 2007. Genetic diversity in Egyptian buffalo using microsatellite markers. *Arab J Biotechnol.* 10:219–232.
- Ghavi HZ, Madad M, Shadparvar AA, Kianzad D. 2012. An observational analysis of secondary sex ratio, stillbirth and birth weight in Iranian buffaloes (*Bubalus bubalis*). *J Agric Sci Technol.* 14:1477–1484.
- Iamartino D, et al. 2017. Design and validation of a 90K SNP genotyping assay for the water buffalo (*Bubalus bubalis*). *PLoS One.* 12(10):e0185220.
- Iannuzzi L, Di Meo GP. 2009. Water buffalo. In: *Genome Mapping and Genomics in Domestic Animals*. Berlin, Heidelberg: Springer. p. 19–31.
- Jaayid TA, Dragh MA. 2013. Genetic biodiversity in buffalo population of Iraq using microsatellites markers. *J Agric Sci Technol.* 3(A):297–301.
- Jones MR, et al. 2018. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* 360(6395):1355–1358.
- Joshi J, et al. 2012. Comparative evaluation of Murrah breeds with buffaloes of Indo-Gangetic Plains. *DHR Int J Biomed Life Sci.* 3:93–105.
- Kayser M, et al. 2008. Three genome-wide association studies and a linkage analysis identify *HERC2* as a human iris color gene. *Am J Hum Genet.* 82(2):411–423.
- Kianzad D. 2000. A case study on buffalo recording and breeding in Iran. *ICAR Tech Ser.* 4:37–44.
- Kumar S, et al. 2007. Mitochondrial DNA analyses of Indian water buffalo support a distinct genetic origin of river and swamp buffalo. *Anim Genet.* 38(3):227–232.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- Li W, et al. 2019. Comparative sequence alignment reveals river buffalo genomic structural differences compared with cattle. *Genomics* 111(3):418–425.
- Li Y, et al. 2010. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet.* 42(11):969–972.
- Liang D, et al. Forthcoming 2020. Genomic analysis revealed a convergent evolution of LINE-1 in coat color: a case study in water buffaloes (*Bubalus bubalis*). *Mol Biol Evol.* doi: 10.1093/molbev/msaa279.
- Low WY, et al. 2019. Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity. *Nat Commun.* 10(1):260.
- McKenna A, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20(9):1297–1303.
- Mishra BP, et al. 2010. Microsatellite based genetic structuring reveals unique identity of Banni among river buffaloes of Western India. *Livestock Sci.* 127(2–3):257–261.
- Mokhber M, et al. 2018. A genome-wide scan for signatures of selection in Azeri and Khuzestani buffalo breeds. *BMC Genomics.* 19(1):Article number: 449.
- Özkan Ünal E, Soysal Mİ, Yüncü E, Dağtaş ND, Togan İ. 2014. Microsatellite based genetic diversity among the three water buffalo (*Bubalus bubalis*) populations in Turkey. *Arch Anim Breed.* 57(1):1–12.
- Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8(11):e1002967.
- Pournourali M, Tarang A, Mashayekhi F. 2015. Chromosomal analysis of two buffalo breeds of Mazani and Azeri from Iran. *Iran J Vet Sci Technol.* 7:22–31.
- Purcell S, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81(3):559–575.
- Qanbari S, et al. 2014. Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet.* 10(2):e1004148.
- Qanbari S, et al. 2019. Genetics of adaptation in modern chicken. *PLoS Genet.* 15(4):e1007989.
- Rosen BD, et al. 2020. *De novo* assembly of the cattle reference genome with single-molecule sequencing. *GigaScience* 9(3):giaa021.
- Rubin CJ, et al. 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464(7288):587–591.
- Rubin CJ, et al. 2012. Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci U S A.* 109(48):19529–19536.
- Safari A, Hossein-Zadeh NG, Shadparvar AA, Arpanahi RA. 2018. A review on breeding and genetic strategies in Iranian buffaloes (*Bubalus bubalis*). *Trop Anim Health Prod.* 50(4):707–714.
- Sanjabi MR, Naderfard HR, Moeini MM, Lavaf A, Ahadi AH. 2009. Potential of milk production of Iranian water buffaloes. In: *EAAP-60th Annual Meeting, Barcelona*; Vol. 1. p. 1–21.
- Sarwar M, Khan MA, Nisa M, Bhatti SA, Shahzad MA. 2009. Nutritional management for buffalo production. *Asian Australas J Anim Sci.* 22(7):1060–1068.
- Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes.* 9:88.
- Steibel JP, et al. 2011. Genome-wide linkage analysis of global gene expression in loin muscle tissue identifies candidate genes in pigs. *PLoS One.* 6(2):e16766.
- Sturm RA. 2009. Molecular genetics of human pigmentation diversity. *Hum Mol Genet.* 18(R1):R9–17.
- Sukla S, Yadav BR, Bhattacharya TK. 2006. Characterization of Indian riverine buffaloes by microsatellite markers. *Asian Australas J Anim Sci.* 19(11):1556–1560.
- Triwityakorn K, et al. 2006. Analysis of genetic diversity of the Thai swamp buffalo (*Bubalus bubalis*) using cattle microsatellite DNA markers. *Asian Australas J Anim Sci.* 19(5):617–621.
- Uffo O, et al. 2017. Analysis of microsatellite markers in a Cuban water buffalo breed. *J Dairy Res.* 84(3):289–292.
- Wang MS, et al. 2020a. 863 genomes reveal the origin and domestication of chicken. *Cell Res.* 30(8):693–701.
- Wang MS, et al. 2020b. Ancient hybridization with an unknown population facilitated high-altitude adaptation of canids. *Mol Biol Evol.* 37(9):2616–2629.
- Wang S, et al. 2017. Whole mitogenomes reveal the history of swamp buffalo: initially shaped by glacial periods and eventually modelled by domestication. *Sci Rep.* 7(1):4708.
- Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38(6):1358–1370.
- Williams JL, et al. 2017. Genome assembly and transcriptome resource for river buffalo, *Bubalus bubalis* (2n = 50). *GigaScience* 6(10):gix088.
- Wright S. 1949. The genetical structure of populations. *Ann Eugen.* 15(1):323–354.
- Wu DD, et al. 2018. Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nat Ecol Evol.* 2(7):1139–1145.
- Yang B, Zeng XLQ, Qin J, Yang C. 2007. Dairy buffalo breeding in countryside of China. *Italian J Anim Sci.* 6(Suppl 2):25–29.
- Yang J, Hong Lee S, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 88(1):76–82.
- Yang Z, et al. 2016. A genetic mechanism for convergent skin lightening during recent human evolution. *Mol Biol Evol.* 33(5):1177–1187.

- Zhang Y, et al. 2016. Strong and stable geographic differentiation of swamp buffalo maternal and paternal lineages indicates domestication in the China/Indochina border region. *Mol Ecol.* 25(7):1530–1550.
- Zhang Y, Colli L, Barker JSF. 2020. Asian water buffalo: domestication, history and genetics. *Anim Genet.* 51(2):177–191.
- Zhang Y, Vankan D, Zhang Y, Barker JS. 2011. Genetic differentiation of water buffalo (*Bubalus bubalis*) populations in China, Nepal and south-east Asia: inferences on the region of domestication of the swamp buffalo. *Anim Genet.* 42(4):366–377.
- Zhang Y, Sun D, Yu Y, Zhang Y. 2007. Genetic diversity and differentiation of Chinese domestic buffalo based on 30 microsatellite markers. *Anim Genet.* 38(6):569–575.
- Zimin AV, et al. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 10(4):R42.

**Associate editor:** Dorothee Huchon