

# Inbreeding estimations based on runs of homozygosity revealed genes associated with local adaptation and dairy traits in Iranian water buffaloes, *Bubalus bubalis*

**Seyed Mohammad Ghoreishifar**

University of Tehran

**Hossein Moradi-shahrbabak** (✉ [hmoradis@ut.ac.ir](mailto:hmoradis@ut.ac.ir))

University of Tehran <https://orcid.org/0000-0002-6680-7662>

**Mohammad Hosein Fallahi**

University of Tehran

**Ali Jalil Sarghale**

University of Tehran

**Mohammad Moradi-Shahrbabak**

University of Tehran

**Rostam Abdollahi-Arpanahi**

University of Tehran

---

## Research article

**Keywords:** Buffalo Genotyping 90K; genetic diversity; genomic regions; gene enrichment; runs of homozygosity stretches

**Posted Date:** June 10th, 2019

**DOI:** <https://doi.org/10.21203/rs.2.10180/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

Background Consecutive homozygous fragments of the genome that inherited from a common ancestor to offspring are known as runs of homozygosity (ROH). The ROH stretches can be used to estimate the genomic inbreeding and also for identifying genomic regions under extensive selection pressure. The aims of this study were (i) to estimate inbreeding based on ROH (FROH); (ii) to compare FROH with FHOM (excess of homozygosity), FUNI (correlation between uniting gametes), and FGRM (genomic relationship matrix); and (iii) to identify frequently occurred ROH islands (i.e., shared by >20% of the population) for gene enrichment analyses. A dataset consisted of 396 buffalo including Azeri (N = 281) and Khuzestani (N = 115) breeds genotyped for ~65000 SNP were used. Results In total, 10,245 ROH were identified with an average number of  $23.3 \pm 14.1$  and  $33.2 \pm 15.9$  segments per animal for Azeri (AZ) and Khuzestani (KHZ) breeds, respectively. Our results showed that on average 4.71% ( $117.9 \pm 120.7$  Mb) and 5.96% ( $149.15 \pm 107.7$  Mb) of the genome of AZ and KHZ is autozygous, respectively. The amount of FHOM, FUNI and FGRM were 0.03, 0.03 and 0.03 in Azeri breed and 0.02, 0.02 and 0.02 in Khuzestani breed, respectively. Strong correlations between FROH-FHOM (0.93) and between FROH-FUNI (0.98 and 0.94) and moderate to strong correlations between FROH-FGRM (0.86 and 0.78) were obtained. Overall, we identified 13 frequently occurred ROH regions. The GO analyses revealed numerous enriched terms for the ROH gene list in KHZ breed. Genes associated with adaptation to hot and arid climate (BMP2), immune response (CYP27B1), milk production, protein, and fat percentage (MARS and ADRA1A, KCTD16), coat color and pigmentation (PMEL and MYO1A), reproductive traits (INHBE and INHBC, STAT6, PCNA), and bone development (SUOX) were identified. Conclusion Genes within common ROH spots present a modest phenotypic selection for milk production traits in Khuzestani buffalo. Further, our results suggest that natural selection may have a role in forming the genome of Iranian buffaloes in order to adapt them to the harsh environment of the country.

## Background

There are two main species of buffalo: Asian buffalo which is often called water (or river) buffalo, and African wild buffalo (*Syncerus*) [1, 2]. In many tropical and subtropical countries, water buffalo is raised for both meat and milk production [3]. The worldwide buffalo population is only about 11% that of cattle. However, compared to other domesticated species, the world's buffalo population has raised in the last five decades by ~1.65% annually [4], for which India, Pakistan, and Europe with 5.3%, 4.8%, and 4.5% have the highest rates of annual increase, respectively. [5].

Domestication of *Bubalus bubalis* occurred around 5000 years ago [1]. The three major Iranian buffalo breeds are Azeri (AZ), Khuzestani (KHZ), and Mazandarani (MZ), each belongs to different geographical zone [1, 4]. The AZ, KHZ, and MZ are common in the north-west (and north), west (and southwest), and north of the country, respectively. Buffalo is of great economic and cultural importance in Iran and other developing nations for their milk production because of having such unique properties as resistance to local parasites, high ability to make use of low-quality feed, and long productive lifespan. However, according to Safari, Hossein-Zadeh, Shadparvar and Arpanahi [1] the total number of buffalo population in Iran in 1930 was estimated 1500 thousand heads which decreased to 48 thousand by 1995, and this trend is not similar with the global buffalo population growth.

The inbreeding coefficient (F) is the most prevalent parameter in use in breeding programs because it can be estimated from pedigree records. However, the access to high-density SNP markers makes it possible to estimate the F based on genome-wide autozygosity [6]. Autozygosity occurs when parents pass identical chromosomal fragments, which they inherited from a common ancestor, to their offspring [7], and this is described as runs of homozygosity (ROH) [8]. The ROH has been suggested as a reliable way as it can distinguish between homozygous tracks that are IBD (identical by descent) from those non-autozygous IBS (identical by state) segments in the genome [6]. So, it has previously been reported that molecular methods of calculating the inbreeding such as  $F_{ROH}$  can eliminate the drawbacks of using pedigrees to estimate inbreeding [9, 10].

Also, studying ROH may give us a better understanding about the directions in which genetic selection has occurred as it leaves footprints on the genome. [11]. ROH islands that frequently occurred among individuals may contain genes associated with different traits of interest so that they can be interesting in animal genetics [11, 12]. In other words, recombination occurs randomly, and the distribution of ROH is expected to be random over the genome. So, frequently occurring ROH in some regions

of the genome may be a representative of selection pressure on those spots. The distribution and the occurrence of ROH have been studied in humans [13], cattle [9, 14], pigs [15], sheep [16-18] but rarely in species such as buffaloes. In the current study, we aimed to estimate autozygosity in the genome of Azeri and Khuzestani buffalo breeds and to study ROH spots that frequently occurred among the individuals. Moreover, we provided a comparison of  $F_{ROH}$  with other molecular inbreeding estimation methods such as  $F_{GRM}$ ,  $F_{HOM}$ , and  $F_{UNI}$  in these populations.

## Material & Methods

### *Sample collection, ethical statement, and data quality control*

Hair roots and blood samples were obtained from 159 milk-recorded herds including 112 herds of AZ and 47 of KHZ. All practices regarding data collection were reviewed and confirmed by the research ethics committee of the College of Agriculture and Natural Resources of the University of Tehran and by the Animal breeding Center of Iran (ABCI). Three hundred and ninety-six buffaloes (i.e., 281 sample from AZ and 115 from KHZ breeds) were genotyped using the 90K SNPChip (Axiom® Buffalo 90K Genotyping Array) including 89,988 evenly distributed SNPs throughout the buffalo genome. This dataset has previously been used by Mokhber, Moradi-Shahrbabak, Sadeghi, Moradi-Shahrbabak, Stella, Nicolzzi, Rahmaninia and Williams [4].

For the 90K chip, SNPs were selected using buffalo DNA sequence [2, 19]. However, since the draft of buffalo genome [20] does not have the appropriate quality [2, 19], the SNPs were aligned to the cattle reference genome (UMD\_3.1) [21]. Genotypes were obtained through AffyPipe [22]; then all the monomorphic and polymorphic SNPs with high-resolution ( $n=64,750$ ) were first maintained. According to the filtration criteria, samples with more than 5% missing genotype and SNPs with 5% missing rate were eliminated from further analyses. SNPs with unidentified position in the UMD\_3.1 assembly, with  $MAF > 2\%$ , and with HWE chi-square test  $P$ -value of lower than 0.000001 were filtered out. In total, 64,711 variants and 396 samples with an average call rate of 99.6% passed filters and QC.

### *Genetic distance among breeds*

The genetic distance which is based on the IBS matrix was estimated through the `-ibs` command implemented in PLINK v1.09 [23]. Principal Component (PC) analysis of genetic distances was performed to visualize the genetic relationship among samples, and the graphical representation was displayed using R (<http://www.R-project.org/>). We removed four samples including two from each breed which were placed outside their expected breed cluster.

### *Runs of homozygosity analyses*

Analyses of ROH were conducted for Azeri ( $n=279$ ) and Khuzestani ( $n=113$ ) populations, separately. For each individual within their breed cluster, ROH segments were identified using PLINK v1.09 according to the following criteria: (1) In order to avoid short ROH that may originate from linkage disequilibrium, they should have at least 1 Mb length; (2) only one heterozygous and one missing SNP were permitted in an ROH; (3) To decrease the amount of false positive ROH, the smallest number of SNPs were allowed in an ROH using the following equation:

$$l = \log_e \alpha n_a \log_e (1 - het), \text{ (Eq. 1)}$$

where  $l$  is the number of SNP allowed in ROH,  $\alpha$  set to be 0.5 which represent the false positive ROH;  $n_a$  and  $n_s$  represent the numbers of animal per breed and the number of SNP per individual, respectively; and  $het$  is the average heterozygosity for each breed. In the current study,  $l$  was calculated to be 40 and 38 for Azeri and Khuzestani breed; (4) each ROH should have at least 1 SNP over 100 Kb; (5) max gap between two neighboring SNPs that were in ROH was set 1 Mb. Then we categorized detected ROH into five groups including 1 to 2, 2 to 4, 4 to 8, 8 to 16 and >16 Mb using the same classification that reported by other authors [9, 24]. After that, the frequency and the mean length (Mb) of ROHs within each category, the percentage of each ROH category, and the percentage of genome coverage by each ROH category were calculated, separately for each breed, through R software (<http://www.R-project.org/>).

### *Inbreeding coefficient estimations*

The coefficient of inbreeding were estimated through ROH ( $F_{ROH}$ ), excess of homozygous genotypes ( $F_{HOM}$ ), correlation between uniting gametes ( $F_{UNI}$ ) and genomic relationship matrix ( $F_{GRM}$ ).  $F_{ROH}$  (Eq. 2) was calculated per individual as following [25]:

$$F_{ROH}^i = \frac{\sum_{j=1}^n L_{ROH}^j}{L_{aut}},$$

(Eq. 2)

in which  $F_{ROH}^i$  is the inbreeding for the  $i^{th}$  animal;  $n$  and  $L_{ROH}^j$  are the total number of ROHs and the length of the  $j^{th}$  ROH for the  $i^{th}$  animal, respectively; and  $L_{aut}$  is the total autosome length covered by the SNP markers (i.e., 2.5 Gb in the present study). Alternatively, we calculated genomic inbreeding through the `-ibc` command implemented in GCTA software [26] which returns three different inbreeding estimations including  $F_{GRM}$  (Eq. 3),  $F_{HOM}$  (Eq. 4) and  $F_{UNI}$  (Eq. 5)

$$F_{GRM} = x_i - 2p_i \hat{h}_i - 1, \text{ (Eq. 3)}$$

$$F_{HOM} = 1 - x_i (2 - x_i) h_i, \text{ (Eq. 4)}$$

$$F_{UNI} = x_i^2 - (1 + 2p_i) x_i + 2p_i^2 h_i, \text{ (Eq. 5)}$$

where  $x_i$  and  $p_i$  are the number of copies and the frequency of the first allele for the  $i^{th}$  genotype, respectively; and  $h_i$  is  $2p_i(1-2p_i)$ . The Pearson's correlation coefficient among these four inbreeding measures were also calculated.

#### *Frequently appeared ROH islands and gene enrichment analyses*

To identify genomic regions which were frequently covered with ROH, the frequency that each SNP occurred in ROH was calculated for each breed. The percentage that each SNPs has occurred in ROH were depicted against their position along the autosomes which were mapped against the cattle reference genome (UMD3.1). The ROHs that had a frequency of > 20% were considered as prospective genomic regions. To identify genes inside these spots, we used the UMD3.1 map viewer from NCBI website (<https://www.ncbi.nlm.nih.gov/mapview/>). Then to identify significantly enriched gene ontology terms ( $P < 0.05$ ) using the list of genes that had previously been obtained based on the *Bos taurus* annotation file, we used DAVID v6.8 tool [27, 28]. Finally, we did an extensive literature review to explore the biological function of annotated genes inside the ROH regions.

## Results And Discussion

The objective of this research was to estimate genome-wide inbreeding in the genome of AZ and KHZ buffalo breeds using different methods, as well as to identify ROHs which shared by more than 20% of samples within each breed for further gene enrichment analyses. AZ and KHZ, which are the two major buffalo breeds in Iran, adapted to distinct geographical areas. [4]. As shown in Fig. 1 principal component analyses of the IBS matrix derived from SNP data shows two separate populations with no overlap between them which means that they are genetically distinct. However, population structure analyses reported by Mokhber, Moradi-Shahrbabak, Sadeghi, Moradi-Shahrbabak, Stella, Nicolzzi, Rahmaninia and Williams [4] has already shown a significant admixture between two breeds.

#### *Runs of homozygosity*

In the present study, we defined ROHs as lengths of homozygous genotypes that had >1Mb length but only up to one heterozygous genotype. Due to linkage disequilibrium, short homozygous haplotypes are prevalent in the bovine genome. Hence, we set a min length of 1 Mb and a min number of 40 (AZ) and 38 (KHZ) SNP for ROH detection in order not to consider those small and prevalent haplotypes in ROH. Unlike human populations, livestock species have higher levels of autozygosity and longer ROHs [9, 14, 25]. In addition, genotyping error can always occur. Therefore, allowing at least one heterozygous SNP seems more reasonable [16, 17, 24] as it avoids underestimating of long ROHs.

In total, 10,247 ROH were detected, from which 6496 ROH were in the AZ genome and 3751 in the KHZ genome (Table 1). The average  $F_{ROH>1Mb}$  was  $0.047 \pm 0.048$  in the AZ and  $0.06 \pm 0.04$  in the KHZ breed. The average number of ROH per individual was

23.3 in AZ (ranging from 4 to 88) and 33.2 in KHZ (ranging from 4 to 132). These results illustrate that all the individuals of AZ and KHZ breed had at least 4 ROH longer than 1Mb (data not shown). As illustrated in Table (1), more than 53% of the detected ROH were of length 2-4 Mb. The percent of different length of ROH can be a reliable indicator of the number of past generations in which inbreeding has occurred. This is due to the recombination event. In fact, recent inbreeding results in longer ROH because the recombination event requires sufficient time (generations) to disintegrate those long IBD stretches. On the other hand, short ROHs are due to ancient inbreeding because during continuous meiosis long IBD segment would have been broken down [13]. Surprisingly, 100% of the samples involved in our study displayed not less than one ROH with a length between 2-4 Mb (data not shown), representing that inbreeding event has occurred around 20 generations ago [6]. However, this results should be interpreted with cautious. As reported by Ferenčaković, Sölkner and Curik [29], a medium-density chip can result in the overestimation of the number of short-length ROHs (>4 Mb) probably because the searches cannot find the heterozygous genotype on the medium-density panel that appeared on the denser one.

#### *Evaluation of different methods of genomic inbreeding*

The mean inbreeding coefficient estimated through different methods were described in Table 2. While the KHZ presented lower inbreeding values for  $F_{HOM}$ ,  $F_{UNI}$ , and  $F_{GRM}$  (ranging from 0.019 to 0.021) than did AZ breed (0.03) for the same methods,  $F_{ROH}$  was higher for KHZ (0.059) than AZ breed (0.047) (Table 2). Compared to  $F_{ROH}$ , inbreeding estimators such as  $F_{HOM}$ ,  $F_{UNI}$ , and  $F_{GRM}$  show noticeable variation among different breeds as they depend on allele frequency [10]. Our result was not surprising because there are differences in the minor allele frequency distribution among two buffalo breeds, and SNPs with low MAF (<0.2) are more frequent in KHZ genome (Figure S1).

Pearson's correlation between  $F_{ROH}$  and the other inbreeding estimators were represented in Table 3. High correlations were observed between  $F_{ROH}$  and  $F_{UNI}$  (0.94 and 0.98 for KHZ and AZ, respectively), as well as between  $F_{ROH}$  and  $F_{HOM}$  (0.93 for both breeds) (Table 3). In the literature, low to high correlation between  $F_{ROH}$  and  $F_{UNI}$  (ranging from 0.15 to 0.80) was reported by Zhang, Calus, Guldbandsen, Lund and Sahana [10]. In addition, close correlations between  $F_{ROH}$ - $F_{HOM}$  ranging from 0.83 to 0.95 was described by Mastrangelo, Tolone, Di Gerlando, Fontanesi, Sardina and Portolano [14] for Italian dairy cattle breeds. Another high correlation (0.84) between these two inbreeding estimator was reported by Zhang, Young, Wang, Sun, Wolc and Dekkers [15] which was based on ROHs >5Mb. There are also considerable variation in correlations between  $F_{ROH}$  and  $F_{HOM}$  reported in the literature (0.06, 0.35 and 0.61) for three dairy cattle breeds [10].

As shown in Table 3, we obtained modest to strong correlations between  $F_{ROH}$ - $F_{GRM}$  (0.78 and 0.86 for KHZ and AZ, respectively) which was the lowest, among others. In Holstein cattle, the correlation of 0.81 was reported by Bjelland, Weigel, Vukasinovic and Nkrumah [30]. However, small to modest correlations between  $F_{ROH}$ - $F_{GRM}$  (ranging from 0.17 to 0.42) was also reported by Mastrangelo, Tolone, Di Gerlando, Fontanesi, Sardina and Portolano [14] in Italian dairy breeds, and a moderate correlation (0.62) by Pryce, Haile-Mariam, Goddard and Hayes [31] in Holstein and Jersey populations. The variation among different results reported in the literature for correlation between  $F_{ROH}$  and  $F_{GRM}$  is most likely due to the fact that  $F_{GRM}$  depends strongly on the allele frequencies, and especially populations with different allele frequency are expected to illustrate misleading IBD results.

$F_{PED}$  of 0.03 in Iranian buffalo was previously reported [32] that was lower than  $F_{ROH}$  estimated in this study. Since pedigree data were not available in the present study, comparing  $F_{PED}$  and  $F_{ROH}$  was not possible. However, Zhang, Calus, Guldbandsen, Lund and Sahana [10] and Purfield, McParland, Wall and Berry [18] reported a moderate to high (from 0.47 to 0.82) and low to moderate (from 0.12 to 0.76) correlations between  $F_{PED}$  and  $F_{ROH}$  in cattle and sheep breeds, respectively. Low to moderate correlation between  $F_{PED}$  and  $F_{ROH}$  was also reported by Peripolli, Stafuzza, Munari, Lima, Irgang, Machado, do Carmo Panetto, Ventura, Baldi and da Silva [9] suggesting that  $F_{PED}$  may not capture small IBD segment that are the result of ancient inbreeding. It is apparent that while inbreeding based on pedigree has its own drawbacks such as the inability to capture ancient inbreeding, and that accurate and in-depth pedigree records are not always available, inbreeding methods based on allele frequency has also shown a considerable variation among different studies. So, it can be concluded that since ROH does not depend on allele frequencies and can capture both recent and ancient inbreeding, it is a more suitable approach to assess inbreeding in livestock genomics.

It should be noted that ROH in an animal could occur as a result of either inbreeding or such events as selection. Indeed, some haplotypes that frequently occur on the genome are a direct result of selection pressure on the regions of genome harboring functional role [33]. Iranian buffalo breeds have not been exposed to intensive artificial selection; hence, the existence of ROH in these two breeds can be due to both natural selection and the inbreeding itself. Pryce, Haile-Mariam, Goddard and Hayes [31] studying the association of homozygosity on the genome of Holstein cattle with inbreeding depression reported associations of ROH<sub>>3.5Mb</sub> with decreased milk production. In the current study, around 35% of the total detected ROHs were of length > 3.5 Mb. Therefore, our results suggest drawing up conservation plans to preserve the local livestock genetic resources.

Variations between samples with respect to their total ROH number and total ROH length were seen (Fig. 2). Individuals with an equal portion of the genome covered by ROH revealed different numbers of segments, which is likely due to the fact that they have a distinct distance from the common ancestor. On the whole, the portion of the genome that was autozygot (4.71% and 5.96% in AZ and KHZ, respectively) was high compared with the result reported in Nellore cattle (4.58%) [34], and was low compared to the result reported in Marchigiana beef breed (7%) [11], Austrian dual purpose breeds (9%) [35], and in Holstein cattle (10%) [36].

#### *Candidate genes inside frequently occurred ROH regions*

A genome-wide search for SNPs that have frequently occurred within ROH revealed 13 regions on BTA1, BTA2, BTA5, BTA7, BTA13, BTA14, BTA19, and BTA29 (Figure 3 & Table S1). From among these ROH islands, those on BTA7, BTA13, and BTA14 were seen in both AZ and KHZ breeds (Figure 3). The strongest peak was viewed on BTA19 (19:411,773-3,701,223 bp) for AZ and on BTA5 (5:55,217,391-57,476,442 bp) for KHZ, which appeared in about 30% of the individuals (Fig. 3).

In this study, all the gene ontologies that significantly enriched ( $P < 0.05$ ) were not discussed in detail, but we concentrated on those regions that displayed associations with particular traits of interest regarding animal genetics and breeding. In total, 220 genes were identified; of which 22 in AZ and 198 in KHZ genome (Table S2). Surprisingly, no significant GO term ( $P < 0.05$ ) was found for the AZ population. The total length of frequently occurred ROH islands were 6,254,221 and 15,356,992 kb in the AZ and KHZ genome, respectively. That is more likely why only a few genes with no significant GO was identified for the AZ. However, the GO analyses revealed numerous significant process for the in KHZ breed. Five genes identified with respect to positive regulation of DNA metabolic development (GO:0051054) in biological process, in which we highlight the STAT6 (signal transducer and activator of transcription 6) gene on BTA 5. Interestingly, in feedlot cattle, it was recognized to have significant effects on the growth efficiency and on the quality of carcass [37]. In addition, Nguyen, Reverter, Cánovas, Venus, Anderson, Islas-Trejo, Dias, Crawford, Lehnert and Medrano [38] using a co-expression network analysis reported the critical role for transcription factors: STAT6, PBX homeobox2, and polybromo1 in regulating pubertal development in Brahman heifers.

Twelve genes recognized in ROH islands that were related to lipid metabolic process (GO:0006629). Of these, BMP2 (bone morphogenetic protein 2, on BTA13) may play a role in the rebuilding of the hair follicle [39]. In porcine, this gene was reported to have an important role in regulating body dimension and muscle developments [40, 41]. Kim, Elbeltagy, Aboul-Naga, Rischkowsky, Sayre, Mwacharo and Rothschild [42], studying signatures of selection between goats and sheep native to Egypt, reported associations between several genes including BMP2 with body size and development and concluded that these genes regulate adaptation to hot arid habitat. According to Moioli, Pilla and Ciani [43] BMP2 an VNRT genes are most likely associated with the fat tail phenotype in sheep. Yuan, Liu, Liu, Kijas, Zhu, Hu, Ma, Zhang, Du and Wang [44] reported some genes including BMP2 as a candidate gene associate with tail type in Chinese indigenous sheep. Another gene related to this ontology identified on BTA 5 is CYP27B1 (1- $\alpha$ -hydroxylase), for which gene expression was reported to be up-regulated as a result of bacterial infection suggesting that this gene has a role in modulating innate immune responses [45]. On BTA13 of KHZ breed, three genes identified with regard to positive regulation of DNA replication (GO:0045740) in biological process; from which, the PCNA (proliferating cell nuclear antigen) gene has earlier been reported to associate with follicular development and growth in buffalo ovary [46].

Single-organism cellular process (GO:0009987) with 64 genes was significantly enriched ( $p = 0.05$ ), which contains the MARS (methionyl-tRNA synthetase) and ADRA1D (adrenoceptor alpha 1D also known as ADRA1A) genes on BTA5 and 13, respectively. Surprisingly, MARS has previously been reported to be in association with milk and protein production in Chines [47] and

Portuguese [48] Holstein cattle, and ADRA1D with milk protein in Murrah dairy buffaloes [49]. Other interesting genes within this category were INHBC and INHBE (inhibin beta C and E subunits) on BTA 5 reported as candidate genes associate with reproductive ability of tropical young bulls [50], and with composite reproductive traits of Lori-Bakhtiari sheep [51]. KCTD16 gene on BTA7 were reported as candidate gene for fat yield in Nordic Holstein cattle [52], with meat quality in Simmental beef cattle [53]. And with residual feed intake in Junmu White pigs [54]. The PMEL (premelanosome protein) and MYO1A (myosin IA) on BTA 5 are two genes within this category that have been reported as putative candidate genes related to coat color phenotypes in cattle by [55] and [56], respectively. PMEL is required for the melanin biosynthesis process; for which the original function is the pigmentation of hair, mucous membranes, and eyes [57]. In cattle, this gene is reported as a candidate gene associated with the dilution of coat color, which has an effect on color intensity [58, 59]. SUOX (Sulfite Oxidase), the other important gene within this category was reported to associate with bone development in cattle [60].

Basically, ROH pattern differs from one to another population [14, 61]; as seen in the recent study between AZ and KHZ. Investigation of the pattern of ROH within breeds may help us discover the direction of selection and trace the impact of artificial and natural selection on the genome over time [36]. For buffalo breeding in Iran, selection practice on dairy traits including milk production, and fat and protein percentage have been carrying out through the government and under the supervision of ABCI [1]. Candidate bulls are selected from rural herds based upon their body conformation and maternal performance, and semen of selected bulls are distributed in individual herds [1]. On the other hand, Iranian buffalo breeds display noticeable adaptability to the rigid environment of the country. For example, according to Mokhber, Moradi-Shahrababak, Sadeghi, Moradi-Shahrababak, Stella, Nicolzzi, Rahmaninia and Williams [4], AZ breed in the northwest of Iran is subject to dry summers with temperatures approaching 35 °C, and cold and below-freezing winters with heavy snowfall. The climate of Khuzestan province placed in the Southwest of the country is usually sweltering and sometimes sultry with summers' temperature regularly reaching 45 °C. Therefore, it is not astonishing that the detected ROH regions with significant GO term harboring candidate genes are associated with adaptability to the harsh environments of the country and with milk production related traits which deserve further studies.

## Conclusion

In this work, we assessed runs of homozygosity (ROH) in the genome of Iranian local buffalo breeds. Several genes were identified in hotspot ROH zones, some of which were associated with growth efficiency, immune response, reproduction, and milk and protein production. Our result suggest that genes detected in the present study are related to adaptation of the animal to the harsh environment of the country, as well as to milk production related traits that have been under phenotypic selection. Thus, they deserve further investigations.

## List Of Abbreviations

ROH: Runs of homozygosity; SNP: Single nucleotide polymorphism; AZ: Azeri breed; KHZ: Khuzestani breed; ABCI: Animal Breeding Centre of Iran.

## Declarations

### Acknowledgments

The authors gratefully acknowledge the Animal Breeding Center of Iran (ABCI) and Towsee Kesht va Dam Noandish Alborz Co (Takdna) for access to the animals and recording.

### Funding

We have received no specific funding for the current study.

### Competing interest

The authors declare that they have no competing interests.

## Authors' contributions

SMG HMS MHF AJS RAA MMS conceived and designed the study, SMG MHF AJS analyzed the data and SMG wrote the paper. RAA AJS revised the manuscript. All authors read and approved the final manuscript.

## Ethics approval

Sample collection from studied animals were performed in accordance with animal ethics and approved by the Animal Use Committee in University of Tehran and Animal Breeding Center of Iran (ABCI). In addition, all the samples were collected from commercial farms and the animal owners agreed to be involved in the project through their respective associations with Animal Breeding Center of Iran.

## Consent for publication

Not applicable

## Availability of data and materials

The datasets used used for the current study are available from the corresponding author on reasonable request.

## References

- [1] Safari A., Hossein-Zadeh N.G., Shadparvar A.A., Arpanahi R.A., A review on breeding and genetic strategies in Iranian buffaloes (*Bubalus bubalis*), *Tropical animal health and production*. 50 (2018) 707-714.
- [2] Iamartino D., Nicolazzi E.L., Van Tassell C.P., Reecy J.M., Fritz-Waters E.R., Koltjes J.E., Biffani S., Sonstegard T.S., Schroeder S.G., Ajmone-Marsan P., Design and validation of a 90K SNP genotyping assay for the water buffalo (*Bubalus bubalis*), *PloS one*. 12 (2017) e0185220.
- [3] De Camargo G., Aspilcueta-Borquis R.R., Fortes M., Porto-Neto R., Cardoso D.F., Santos D., Lehnert S., Reverter A., Moore S., Tonhati H., Prospecting major genes in dairy buffaloes, *BMC genomics*. 16 (2015) 872.
- [4] Mokhber M., Moradi-Shahrbabak M., Sadeghi M., Moradi-Shahrbabak H., Stella A., Nicolizzi E., Rahmaninia J., Williams J.L., A genome-wide scan for signatures of selection in Azeri and Khuzestani buffalo breeds, *BMC genomics*. 19 (2018) 449.
- [5] Dhanda O., Developments in water buffalo in Asia and Oceania, in *Proc. of the Seventh World Buffalo Congress, Manila, Philippines, 2004*, Vol. 20, pp. 17-28.
- [6] Howrigan D.P., Simonson M.A., Keller M.C., Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms, *BMC genomics*. 12 (2011) 460.
- [7] Broman K.W., Weber J.L., Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain, *The American Journal of Human Genetics*. 65 (1999) 1493-1500.
- [8] Gibson J., Morton N.E., Collins A., Extended tracts of homozygosity in outbred human populations, *Human molecular genetics*. 15 (2006) 789-795.
- [9] Peripolli E., Stafuzza N.B., Munari D.P., Lima A.L.F., Irgang R., Machado M.A., do Carmo Panetto J.C., Ventura R.V., Baldi F., da Silva M.V.G.B., Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (*Bos indicus*) dairy cattle, *BMC genomics*. 19 (2018) 34.
- [10] Zhang Q., Calus M.P., Guldbbrandtsen B., Lund M.S., Sahana G., Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds, *BMC genetics*. 16 (2015) 88.

- [11] Marras G., Gaspa G., Sorbolini S., Dimauro C., Ajmone-Marsan P., Valentini A., Williams J.L., Macciotta N.P., Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy, *Animal genetics*. 46 (2015) 110-121.
- [12] Purfield D.C., Berry D.P., McParland S., Bradley D.G., Runs of homozygosity and population history in cattle, *BMC genetics*. 13 (2012) 70.
- [13] Kirin M., McQuillan R., Franklin C.S., Campbell H., McKeigue P.M., Wilson J.F., Genomic runs of homozygosity record population history and consanguinity, *PloS one*. 5 (2010) e13996.
- [14] Mastrangelo S., Tolone M., Di Gerlando R., Fontanesi L., Sardina M., Portolano B., Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds, *Animal*. 10 (2016) 746-754.
- [15] Zhang Y., Young J., Wang C., Sun X., Wolc A., Dekkers J., Inbreeding by pedigree and genomic markers in selection lines of pigs, in *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production: WCGALP-2014*, 2014.
- [16] Ghoreishifar S.M., Moradi-Shahrbabak H., Parna N., Davoudi P., Khansefid M., Linkage disequilibrium and within-breed genetic diversity in Iranian Zandi sheep, *Archives Animal Breeding*. 62 (2019) 143-151.
- [17] Mastrangelo S., Portolano B., Di Gerlando R., Ciampolini R., Tolone M., Sardina M., Consortium I.S.G., Genome-wide analysis in endangered populations: a case study in Barbaresca sheep, *Animal*. 11 (2017) 1107-1116.
- [18] Purfield D.C., McParland S., Wall E., Berry D.P., The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds, *PLoS One*. 12 (2017) e0176780.
- [19] Ghoreishifar S.M., Moradi-Shahrbabak H., Moradi-Shahrbabak M., Nicolazzi E.L., Williams J.L., Iamartino D., Nejati-Javaremi A., Accuracy of imputation of single-nucleotide polymorphism marker genotypes for water buffaloes (*Bubalus bubalis*) using different reference population sizes and imputation tools, *Livestock Science*. 216 (2018) 174-182.
- [20] Williams J.L., Iamartino D., Pruitt K.D., Sonstegard T., Smith T.P., Low W.Y., Biagini T., Bomba L., Capomaccio S., Castiglioni B., Genome assembly and transcriptome resource for river buffalo, *Bubalus bubalis* (2 n= 50), *GigaScience*. 6 (2017) gix088.
- [21] Zimin A.V., Delcher A.L., Florea L., Kelley D.R., Schatz M.C., Puiu D., Hanrahan F., Pertea G., Van Tassel C.P., Sonstegard T.S., A whole-genome assembly of the domestic cow, *Bos taurus*, *Genome biology*. 10 (2009) R42.
- [22] Nicolazzi E.L., Iamartino D., Williams J.L., AffyPipe: an open-source pipeline for Affymetrix Axiom genotyping workflow, *Bioinformatics*. 30 (2014) 3118-3119.
- [23] Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M., Lee J.J., Second-generation PLINK: rising to the challenge of larger and richer datasets, *GigaScience*. 4 (2015).
- [24] Szmatoła T., Gurgul A., Ropka-Molik K., Jasielczuk I., Ząbek T., Bugno-Poniewierska M., Characteristics of runs of homozygosity in selected cattle breeds maintained in Poland, *Livestock Science*. 188 (2016) 72-80.
- [25] McQuillan R., Leutenegger A.-L., Abdel-Rahman R., Franklin C.S., Pericic M., Barac-Lauc L., Smolej-Narancic N., Janicijevic B., Polasek O., Tenesa A., Runs of homozygosity in European populations, *The American Journal of Human Genetics*. 83 (2008) 359-372.
- [26] Yang J., Lee S.H., Goddard M.E., Visscher P.M., GCTA: a tool for genome-wide complex trait analysis, *The American Journal of Human Genetics*. 88 (2011) 76-82.
- [27] Huang D.W., Sherman B.T., Lempicki R.A., Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists, *Nucleic acids research*. 37 (2008) 1-13.

- [28] Huang D.W., Sherman B.T., Lempicki R.A., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nature protocols*. 4 (2008) 44.
- [29] Ferenčaković M., Sölkner J., Curik I., Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors, *Genetics Selection Evolution*. 45 (2013) 42.
- [30] Bjelland D., Weigel K., Vukasinovic N., Nkrumah J., Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding, *Journal of Dairy Science*. 96 (2013) 4697-4706.
- [31] Pryce J.E., Haile-Mariam M., Goddard M.E., Hayes B.J., Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle, *Genetics Selection Evolution*. 46 (2014) 71.
- [32] Ghavi Hossein-Zadeh N., Analysis of population structure and genetic variability in Iranian buffaloes (*Bubalus bubalis*) using pedigree information, *Animal Production Science*. 56 (2016) 1130-1135.
- [33] Gaspa G., Marras G., Sorbolini S., Ajmone Marsan P., Williams J., Valentini A., Dimauro C., Macciotta N., Genome-wide homozygosity in Italian Holstein cattle using HD panel, in *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production, 2014, Vol. 17*.
- [34] Zavarez L.B., Utsunomiya Y.T., Carmo A.S., Neves H.H., Carneiro R., Ferenčaković M., Pérez O'Brien A.M., Curik I., Cole J.B., Van Tassell C.P., Assessment of autozygosity in Nelore cows (*Bos indicus*) through high-density SNP genotypes, *Frontiers in genetics*. 6 (2015) 5.
- [35] Ferenčaković M., Hamzić E., Gredler B., Curik I., Sölkner J., Runs of homozygosity reveal genome-wide autozygosity in the Austrian Fleckvieh cattle, *Agriculturae Conspectus Scientificus*. 76 (2011) 325-329.
- [36] Kim E.-S., Cole J.B., Huson H., Wiggans G.R., Van Tassell C.P., Crooker B.A., Liu G., Da Y., Sonstegard T.S., Effect of artificial selection on runs of homozygosity in US Holstein cattle, *PloS one*. 8 (2013) e80813.
- [37] Rincon G., Farber E., Farber C., Nkrumah J., Medrano J., Polymorphisms in the STAT6 gene and their association with carcass traits in feedlot cattle, *Animal genetics*. 40 (2009) 878-882.
- [38] Nguyen L.T., Reverter A., Cánovas A., Venus B., Anderson S.T., Islas-Trejo A., Dias M.M., Crawford N.F., Lehnert S.A., Medrano J.F., STAT6, PBX2, and PBRM1 emerge as predicted regulators of 452 differentially expressed genes associated with puberty in Brahman heifers, *Frontiers in genetics*. 9 (2018) 87.
- [39] Su R., Zhang W.-G., Sharma R., Chang Z.-L., Yin J., Li J.-Q., Characterization of BMP 2 gene expression in embryonic and adult Inner Mongolia Cashmere goat (*Capra hircus*) hair follicles, *Canadian journal of animal science*. 89 (2009) 457-462.
- [40] Fan B., Onteru S.K., Du Z.-Q., Garrick D.J., Stalder K.J., Rothschild M.F., Genome-wide association study identifies loci for body composition and structural soundness traits in pigs, *PloS one*. 6 (2011) e14726.
- [41] Lee J., Kang J.-H., Kim J.-M., Bayes Factor-Based Regulatory Gene Network Analysis of Genome-Wide Association Study of Economic Traits in a Purebred Swine Population, *Genes*. 10 (2019) 293.
- [42] Kim E.-S., Elbeltagy A., Aboul-Naga A., Rischkowsky B., Sayre B., Mwacharo J.M., Rothschild M.F., Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment, *Heredity*. 116 (2016) 255.
- [43] Moioli B., Pilla F., Ciani E., Signatures of selection identify loci associated with fat tail in sheep, *Journal of Animal Science*. 93 (2015) 4660-4669.
- [44] Yuan Z., Liu E., Liu Z., Kijas J., Zhu C., Hu S., Ma X., Zhang L., Du L., Wang H., Selection signature analysis reveals genes associated with tail type in Chinese indigenous sheep, *Animal genetics*. 48 (2017) 55-66.

- [45] Nelson C.D., Reinhardt T.A., Beitz D.C., Lippolis J.D., In vivo activation of the intracrine vitamin D pathway in innate immune cells and mammary tissue during a bacterial infection, *PloS one*. 5 (2010) e15469.
- [46] Sharma G.T., Dubey P.K., Katiyar A., Kumar G.S., Localization and expression of proliferating cell nuclear antigen (PCNA) and cyclin B1 in buffalo (*Bubalus bubalis*) ovary during different stages of follicular development, *Indian Journal of Animal Sciences*. 81 (2011) 29-00.
- [47] Li C., Cai W., Zhou C., Yin H., Zhang Z., Loo J.J., Sun D., Zhang Q., Liu J., Zhang S., RNA-Seq reveals 10 novel promising candidate genes affecting milk protein concentration in the Chinese Holstein population, *Scientific reports*. 6 (2016) 26813.
- [48] Carvalheira J., Salem M., Thompson G., Chen S., Beja-Pereira A., Genome-wide association study for milk and protein yields in Portuguese Holstein cattle, *MARS*. 131 (2014) 131-183.
- [49] Araújo D.N., de Camargo G.M.F., Fonseca P.D.d.S., Cardoso D.F., Hurtado-Lugo N.A., Aspilcueta-Borquis R.R., Tonhati H., Polymorphisms in Oxytocin and  $\alpha 1a$  Adrenergic Receptor Genes and Their Effects on Production Traits in Dairy Buffaloes, *Animal biotechnology*. 26 (2015) 165-168.
- [50] Fortes M., Reverter A., Kelly M., McCulloch R., Lehnert S., Genome-wide association study for inhibin, luteinizing hormone, insulin-like growth factor 1, testicular size and semen traits in bovine species, *Andrology*. 1 (2013) 644-650.
- [51] Abdoli R., Mirhoseini S.Z., Ghavi Hossein-Zadeh N., Zamani P., Ferdosi M.H., Gondro C., Genome-wide association study of four composite reproductive traits in Iranian fat-tailed sheep, *Reprod Fertil Dev*. (2019). <https://doi.org/10.1071/RD18282>
- [52] Cai Z., Guldbbrandtsen B., Lund M.S., Sahana G., Dissecting closely linked association signals in combination with the mammalian phenotype database can identify candidate genes in dairy cattle, *BMC genetics*. 19 (2018) 30.
- [53] Xia J., Qi X., Wu Y., Zhu B., Xu L., Zhang L., Gao X., Chen Y., Li J., Gao H., Genome-wide association study identifies loci and candidate genes for meat quality traits in Simmental beef cattle, *Mammalian genome*. 27 (2016) 246-255.
- [54] Bai C., Pan Y., Wang D., Cai F., Yan S., Zhao Z., Sun B., Genome-wide association analysis of residual feed intake in Junmu No. 1 White pigs, *Animal genetics*. 48 (2017) 686-690.
- [55] Mészáros G., Petautschnig E., Schwarzenbacher H., Sölkner J., Genomic regions influencing coat color saturation and facial markings in Fleckvieh cattle, *Animal genetics*. 46 (2015) 65-68.
- [56] Edea Z., Dadi H., Dessie T., Uzzaman M., Rothschild M., Kim E.S., Sonstegard T., Kim K.S., Genome-wide scan reveals divergent selection among taurine and zebu cattle populations from different regions, *Animal genetics*. 49 (2018) 550-563.
- [57] Urbinati I., Stafuzza N.B., Oliveira M.T., Chud T.C.S., Higa R.H., de Almeida Regitano L.C., de Alencar M.M., Buzanskas M.E., Munari D.P., Selection signatures in Canchim beef cattle, *Journal of animal science and biotechnology*. 7 (2016) 29.
- [58] Kühn C., Weikard R., An investigation into the genetic background of coat colour dilution in a Charolais× German Holstein F2 resource population, *Animal genetics*. 38 (2007) 109-113.
- [59] Gutiérrez-Gil B., Wiener P., Williams J.L., Genetic effects on coat colour in cattle: dilution of eumelanin and phaeomelanin pigments in an F2-Backcross Charolais× Holstein population, *BMC genetics*. 8 (2007) 56.
- [60] Drögemüller C., Tetens J., Sigurdsson S., Gentile A., Testoni S., Lindblad-Toh K., Leeb T., Identification of the bovine Arachnomelia mutation by massively parallel sequencing implicates sulfite oxidase (SUOX) in bone development, *PLoS genetics*. 6 (2010) e1001079.
- [61] Mastrangelo S., Sardina M., Tolone M., Di Gerlando R., Suter A., Fontanesi L., Portolano B., Genome-wide identification of runs of homozygosity islands and associated genes in local dairy cattle breeds, *animal*. 12 (2018) 2480-2488.

## Tables

Table 1. Descriptive statistics for detected runs of homozygosity (ROH) with different lengths in megabase pair (Mb).

	nROH1		Percentage		Mean length (Mb)		SD (Mb)		Coverage%	
	AZ	KHZ	AZ	KHZ	AZ	KHZ	AZ	KHZ	AZ	KHZ
ROH1-2	991	632	15.26	16.85	1.83	1.81	0.13	0.15	0.07	0.05
ROH2-4	3466	2117	53.36	56.44	2.70	2.71	0.53	0.53	0.37	0.23
ROH4-8	1107	609	17.04	16.24	5.46	5.38	1.08	1.08	0.24	0.13
ROH8-16	590	241	9.08	6.42	11.21	11.22	2.26	2.21	0.26	0.11
ROH>16	342	152	5.26	4.05	26.44	26.21	10.74	9.31	0.36	0.16

1. Number of runs of homozygosity segments

Table 2. Inbreeding estimations obtained through FGRM, FHOM, FUNI, and FROH>1Mb1

Breed	F2GRM	F HOM	FUNI	FROH>1Mb
<b>Azeri</b>	0.03±0.05	0.03±0.05	0.03±0.05	0.047±0.05
<b>Khuzestani</b>	0.019±0.04	0.021±0.06	0.021±0.05	0.059±0.04

1. FROH: Runs of homozygosity, FHOM: excess of homozygous genotypes, FUNI: correlation between uniting gametes, FGRM: genomic relationship matrix

2. Inbreeding estimation ± standard deviation

Table 3. Correlation coefficients between FROH and other estimators of inbreeding coefficient1

Breed	Correlation		
	FGRM -FROH	FHOM-FROH	FUNI-FROH
<b>Azeri</b>	0.86	0.93	0.98
<b>Khuzestani</b>	0.78	0.93	0.94

1. For abbreviations see footnote of Table 2.

## Figures

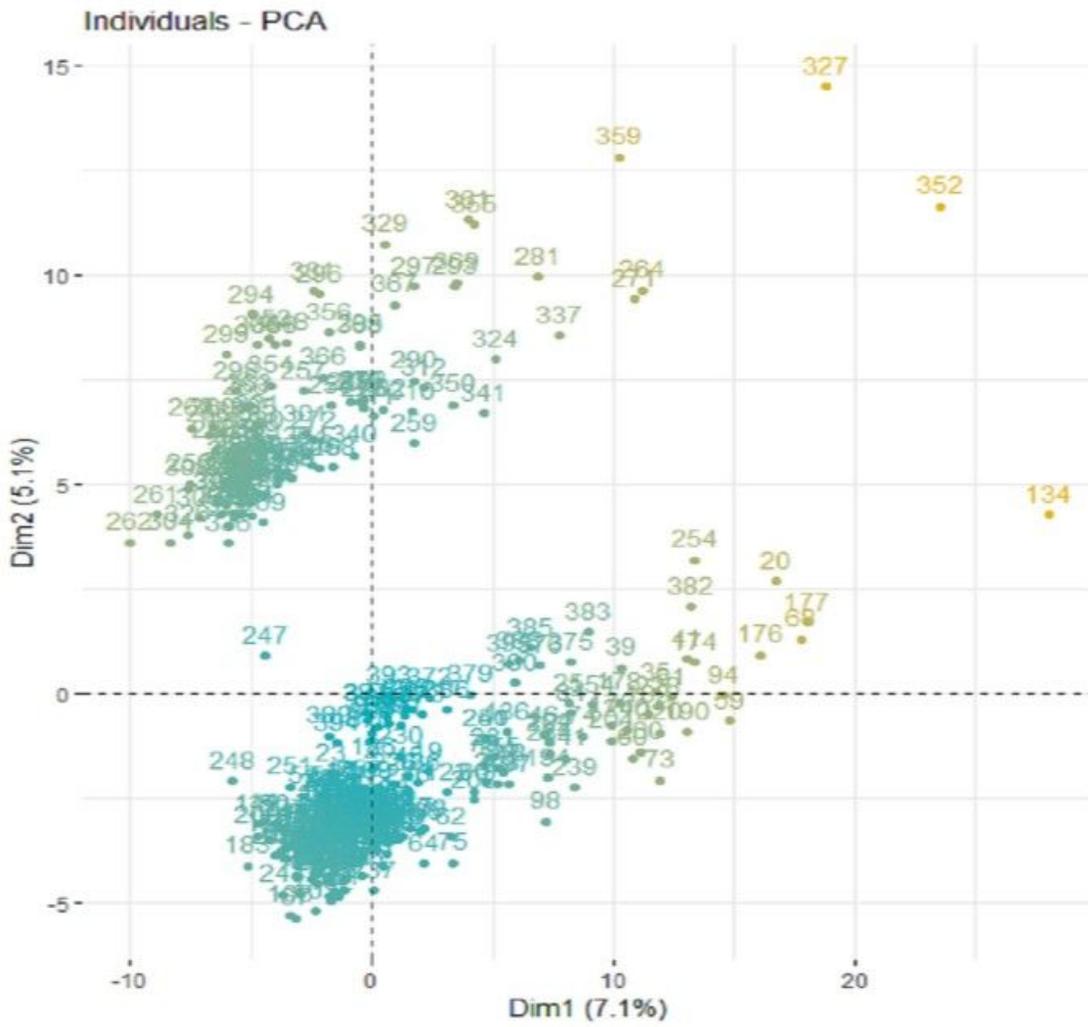


Figure 1

Principal Component analysis using IBS distance matrix

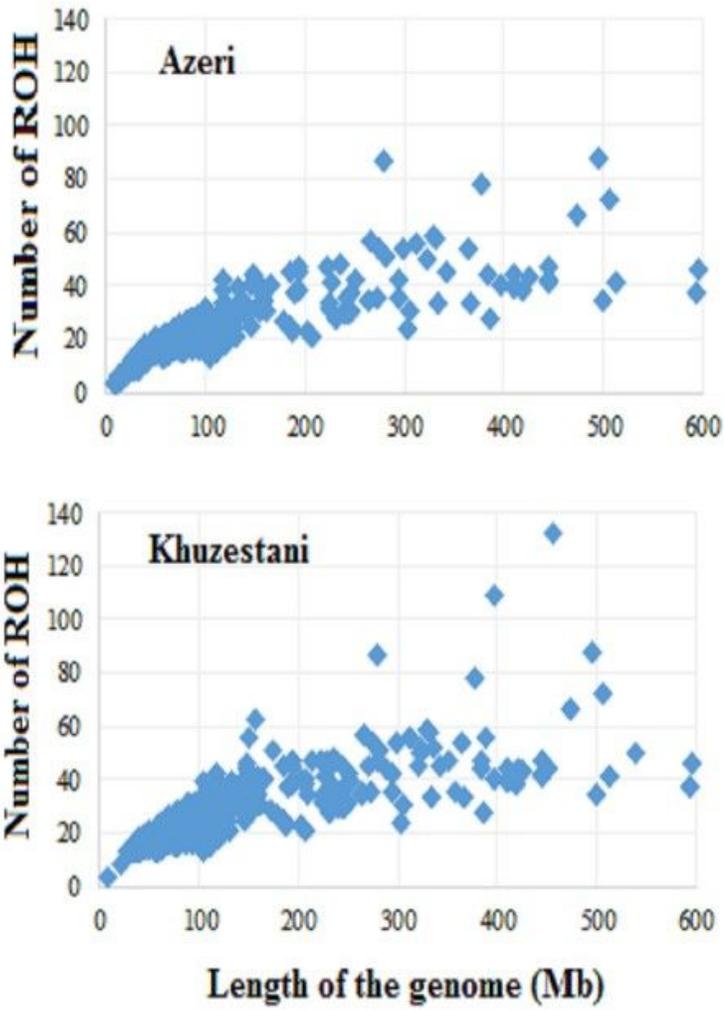


Figure 2

Number and the length of the genome covered by runs of homozygosity (ROH) per animal

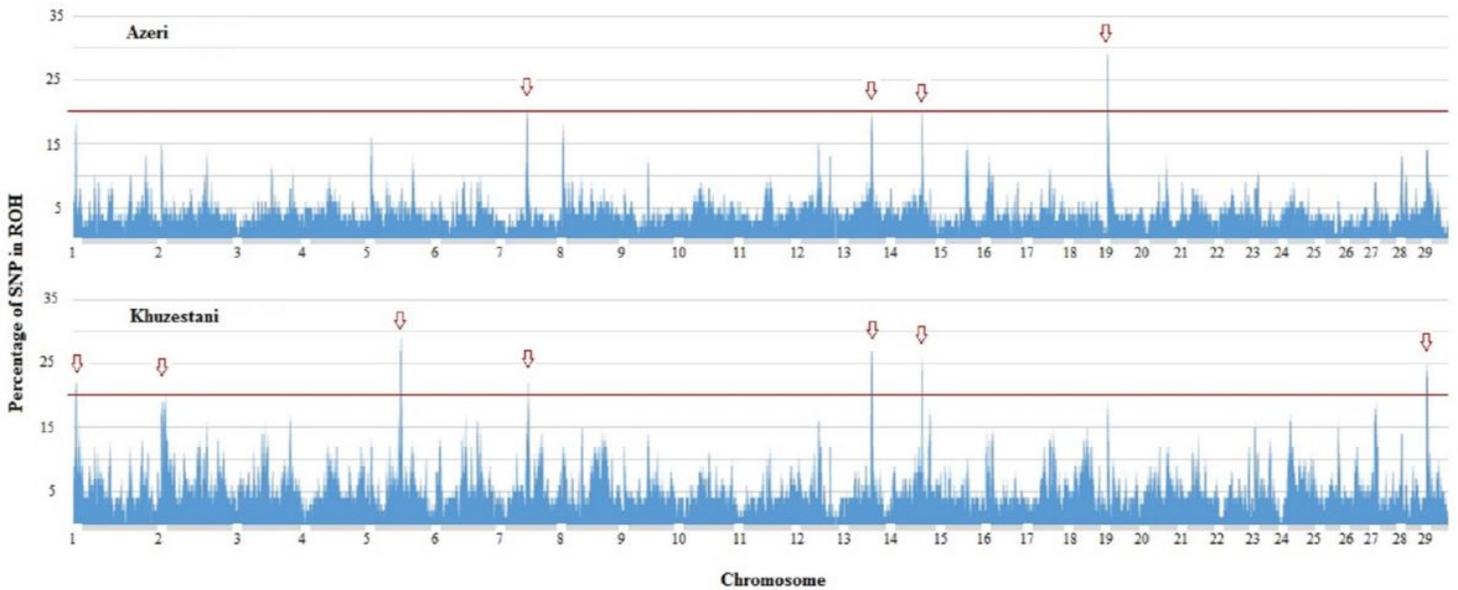


Figure 3

Distribution of frequently occurred runs of homozygosity (ROH) in the Iranian buffalo genome. The X-axis shows the distribution of ROH over the genome, and the Y-axis displays the percentage of overlapping ROH shared among animals for each breed

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS2.xlsx](#)
- [TableS1.xlsx](#)
- [FigureS1.tif](#)