

Genomic measures of inbreeding coefficients and genome-wide scan for runs of homozygosity islands in Iranian river buffalo, *Bubalus bubalis*

Seyed Mohammad Ghoreishifar

University of Tehran

Hossein Moradi-Shahrbabak (✉ hmoradis@ut.ac.ir)

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

Mohammad Hossein Fallahi

University of Tehran

Ali Jalil Sarghale

University of Tehran

Mohammad Moradi-Shahrbabak

University of Tehran

Rostam Abdollahi-Arpanahi

University of Tehran

Majid Khansefid

AgriBio

Research article

Keywords: water buffalo; river buffalo; genetic diversity; inbreeding; gene enrichment; runs of homozygosity; selection signatures

Posted Date: February 5th, 2020

DOI: <https://doi.org/10.21203/rs.2.17561/v5>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on February 10th, 2020. See the published version at <https://doi.org/10.1186/s12863-020-0824-y>.

Abstract

Background: Consecutive homozygous fragments of a genome inherited by offspring from a common ancestor are known as runs of homozygosity (ROH). ROH can be used to calculate genomic inbreeding and to identify genomic regions that are potentially under historical selection pressure. The dataset of our study consisted of 254 Azeri (AZ) and 115 Khuzestani (KHZ) river buffalo genotyped for ~65000 SNPs for the following two purposes: 1) to estimate and compare inbreeding calculated using ROH (FROH), excess of homozygosity (FHOM), correlation between uniting gametes (FUNI), and diagonal elements of the genomic relationship matrix (FGRM); 2) to identify frequently occurring ROH (i.e. ROH islands) for our selection signature and gene enrichment studies. **Results:** In this study, 9102 ROH were identified, with an average number of 21.2 ± 13.1 and 33.2 ± 15.9 segments per animal in AZ and KHZ breeds, respectively. On average in AZ, 4.35% (108.8 ± 120.3 Mb), and in KHZ, 5.96% (149.1 ± 107.7 Mb) of the genome was autozygous. The estimated inbreeding values based on FHOM, FUNI and FGRM were higher in AZ than they were in KHZ, which was in contrast to the FROH estimates. We identified 11 ROH islands (four in AZ and seven in KHZ). In the KHZ breed, the genes located in ROH islands were enriched for multiple Gene Ontology (GO) terms ($P \leq 0.05$). The genes located in ROH islands were associated with diverse biological functions and traits such as body size and muscle development (BMP2), immune response (CYP27B1), milk production and components (MARS, ADRA1A, and KCTD16), coat colour and pigmentation (PMEL and MYO1A), reproductive traits (INHBC, INHBE, STAT6 and PCNA), and bone development (SUOX). **Conclusion:** The calculated FROH was in line with expected higher inbreeding in KHZ than in AZ because of the smaller effective population size of KHZ. Thus, we find that FROH can be used as a robust estimate of genomic inbreeding. Further, the majority of ROH peaks were overlapped with or in close proximity to the previously reported genomic regions with signatures of selection. This tells us that it is likely that the genes in the ROH islands have been subject to artificial or natural selection.

Background

There are two main species of buffalo: the Asian water buffalo (*Bubalus bubalis*) and the African wild buffalo (*Syncerus caffer*), the second of which is also known as the cape buffalo [1, 2]. Domestication of *B. bubalis*, including of the river (*B. bubalis bubalis*, $2n=50$) and swamp (*B. bubalis carabanensis*, $2n=48$) subspecies, occurred approximately 3000–6000 years ago [3]. The domestication of river buffalo occurred in the Indo–Pakistani area, and domestication of swamp buffalo occurred close to the border of China [4]. River buffalo expanded broadly from India, Egypt and Southeast Asia to Europe, and the swamp buffalo is the most common type of buffalo in China and Southeast Asia [3-5]. The worldwide water buffalo population accounts for only approximately 11% of the entire cattle population. However, the population of water buffalo has increased in the past five decades by approximately 1.65% annually [5]. India, Pakistan and Europe (with 5.3%, 4.8% and 4.5% rates of increase, respectively) have the highest rates of annual increase [6].

In many tropical and subtropical countries, river buffalo are raised for both meat and milk production [7]. In Iran and in other developing nations, river buffalo production is of great economic importance because

of the ability of buffalo to make the best use of low-quality feed in the production of their valuable milk, which has a unique taste and curd properties, high resistance to local parasites, high adaptation to harsh climate conditions, and long productive lifespan. The three major Iranian river buffalo breeds are Azeri (AZ), Khuzestani (KHZ) and Mazandarani (MZ), and each of these breeds belongs to different geographical zone [2]. The AZ, KHZ and MZ are common in the north-west and north, west and south-west, and north of the country, respectively. In Iran, the recording of milk and meat production, and the selective breeding of buffalo for better dairy performance (i.e. in milk production, and fat and protein percentage) and better meat production are performed by the Animal Breeding Centre of Iran (ABCI) [2]. Following performance and pedigree recording in some herds, and genetic analysis, candidate bulls are selected from rural herds based on their genetic merits, and the semen of these selected bulls is collected and distributed to all herds [2]. However, despite buffalo production being important in Iran, particularly in rural regions, controlling inbreeding and ensuring genetic improvement of desired traits through traditional breeding programmes are difficult because of a shortage of reliable pedigree and performance records for water buffalo in the country.

The inbreeding coefficient measured from pedigree information (F_{PED}) has been the most common parameter for describing the level of inbreeding since Wright [8]. However, the reliability of the estimated F_{PED} depends on the completeness and correctness of pedigree. With the availability of high-density SNP-chip markers, inbreeding can also be defined according to genomic information such as genome-wide autozygosity [9]. Autozygosity occurs when parents pass identical chromosomal fragments, which they already inherited from a common ancestor, on to their offspring [10]; these genomic regions of homozygosity are known as runs of homozygosity (ROH) [11, 12]. Estimated inbreeding based on ROH (F_{ROH}) can discriminate between homozygous (i.e. identical by descent [IBD]) and non-autozygous (i.e. identical by state [IBS]) positions in the genome [9]. Further, well-recorded pedigree information is not required to have reliable F_{ROH} . Thus, using genetic markers instead of pedigree information to calculate inbreeding can produce more robust estimates [13, 14].

Identifying ROH can also help to find the footprints of genetic selection on the genome [15-17]. However, ROH are suggestive, but not conclusive, of genomic regions under natural or artificial selection because the incidence, extent and distribution of ROH across the genome are influenced by many factors other than ROH, such as recombination rate, population structure, mutation rate and inbreeding [16]. Nevertheless, ROH that frequently occur among individuals may contain genes associated with different traits that have been under historical selection, so that the genes located in ROH islands can be important for selective breeding [13, 15, 18]. ROH can also provide detailed information on the genetic relatedness of animals, which allows breeders to better control inbreeding in the population [16]. This allows mate allocation aiming to minimise inbreeding at the genome level to be achieved more precisely, and the individual animals that have high proportions of ROH coverage to be excluded or used less frequently in mating [16]. The distribution and the occurrence of ROH have been studied in humans [10, 11, 19, 20], cattle [13-15, 18, 21-26], pigs [27-29] and sheep [17, 30-32] but are poorly studied in some species, for example, in water buffalo.

The current study aims to estimate autozygosity in the genome of AZ and KHZ buffalo breeds, and identify ROH spots that frequently occur among the individuals. The study also examines the function of the genes located in ROH islands to identify potential selection signature regions. Moreover, the study compares F_{ROH} with other genomic methods of inbreeding estimation.

Results

Runs of homozygosity

The AZ and KHZ are two major buffalo breeds adapted to distinct geographical areas in Iran [2, 5] (Fig. 1). The PC analysis of the IBS matrix derived from SNP data confirmed two separate populations with no overlap, which means that the samples from the AZ and KHZ breeds were genetically different (Fig. 2). Although Mokhber et al. [5] reported that AZ and KHZ are two distinct populations, they reported a moderate level of admixture between the AZ and MZ breeds. Thus, we excluded the MZ breed from our study. In total, 9102 ROH were detected, 5352 ROH in the AZ genome and 3750 in the KHZ genome (Table 1; Additional file 1). The average number of ROH per individual was 21.23 ± 13.06 in the AZ breed (ranging from 4 to 88) and 33.2 ± 15.92 in the KHZ breed (ranging from 4 to 132). Moreover, all of the individuals in our study had at least four ROH longer than 1 Mb. The variation between samples in total number of ROH and total length of ROH are presented in Fig. 3. Individuals with an almost equal portion of the genome covered by ROH had different numbers and lengths of ROH, which could be an indication of different combinations of recent and distant inbreeding events in the samples.

Evaluation of different methods of genomic inbreeding

Table 2 presents the averages of the estimated inbreeding coefficients using different methods (see also Additional file 2). The average F_{ROH} calculated from ROH >1 Mb in length was 0.043 ± 0.05 in the AZ breed and 0.059 ± 0.04 in the KHZ breed (Table 2; Additional file 1). The estimated inbreeding values based on F_{HOM} , F_{UNI} and F_{GRM} were higher in AZ than they were in KHZ, which was in contrast to the F_{ROH} estimates. However, the Pearson's correlations between F_{ROH} and the estimated inbreeding with other methods were high (Table 3).

Candidate genes inside frequently occurring runs of homozygosity regions

A genome-wide search for SNPs that have frequently occurred within ROH hotspots revealed 11 regions on BTA1, BTA2, BTA5, BTA7, BTA13, BTA14, BTA19 and BTA29 (Fig. 4; Additional file 3). The detected ROH islands on BTA7, BTA13 and BTA14 were partially overlapped in AZ and KHZ. The strongest peaks detected in approximately 30% of the individuals were located on BTA19 (19:411,773–3,701,223 bp) in AZ and on BTA5 (5:55,217,391–57,476,442 bp) in KHZ. In the KHZ breed, the genes located in the ROH islands were significantly enriched ($P \leq 0.05$) in 40 GO terms. These GO terms belonged to 23 biological processes (BP), 12 cellular component (CC) and 5 molecular function (MF) groups (Additional file 4).

Co-location of ROH islands and the identified selection signatures using iHS

The majority of ROH hotspots detected in our study (Fig. 4; Additional file 3) overlapped with selection signature regions reported by Mokhber et al. [5] for the AZ and KHZ breeds using the haplotype-based method (i.e. iHS) (Additional files 5, 6). For example, the SNP Affx-79610232 on BTA5 (55,271,590 bp) with the highest iHS was located in our detected ROH island in the KHZ breed. On BTA13, the eight SNPs with the largest iHS values were located in our detected ROH region. The SNP Affx-79540796 on BTA14 (52,933,269 bp) had the second-highest iHS value in our reported ROH hotspot. On BTA29, the SNP Affx-79545556 (3,274,219 bp) was the SNP with the sixth-highest iHS value that was located in our reported ROH islands.

Discussion

We defined ROH as the lengths of homozygous genotypes that were >1 Mb and contained only up to one heterozygous genotype. Given the strong linkage disequilibrium (LD) between SNPs with a distance up to 100 Kb [33], short homozygous haplotypes are expected to be prevalent in the buffalo genome. Thus, we set a minimum length of 1 Mb and a minimum number of 40 (AZ) and 38 (KHZ) SNP (as described in methods section) to avoid detecting small and prevalent haplotypes as ROH. Unlike human populations, livestock species generally have higher levels of autozygosity and longer ROHs [13, 20, 21]. However, genotyping errors can always affect the quality of ROH calling [12]. Therefore, we allowed one heterozygous SNP in ROH [25, 30, 31] to avoid losing particularly long ROH because of a single genotyping error.

As presented in Table 1, more than 53% of the detected ROH were 2–4 Mb in length. The proportion of different lengths of ROH can be used as an indicator of the number of past generations in which inbreeding has occurred, because the recombination events can rearrange the chromosomes and reduce the length of ROH. Thus, recent inbreeding results in longer ROH because of long IBD stretches. In contrast, short ROHs arise as a result of ancient inbreeding because in meiosis across generations, the long IBD segments are broken down [19]. We detected ROH with a length from 2 to 4 Mb in all of the samples (Additional file 1), which might indicate that some inbreeding events occurred about 20 generations ago [9]. However, our results should be interpreted with caution. As reported by Ferenčaković et al. [34], a medium-density chip could result in overestimation of the number of long-length ROHs (>4 Mb), probably because some heterozygous genotypes tend to appear in these ROHs by increasing the density of markers. Nevertheless, our results were in line with a previous report of a relatively sharp decrease in the effective population size (N_e) of AZ and KHZ breeds and the consequent increased rate of inbreeding since 20 generations ago [33].

The portion of the genome that was autozygote in the AZ and KHZ breeds was lower than the reported ROH coverage in the Marchigiana beef breed (7%) [15], Austrian dual purpose breeds (9%) [35], and Holstein cattle (10%) [36]. This could be because of lower inbreeding in Iranian water buffalo or because we ignored ROH of <1 Mb in length in our study.

On average, F_{HOM} , F_{UNI} and F_{GRM} were higher in AZ than they were in KHZ. However, the previously reported N_e for AZ (477) was larger than it was for KHZ (212) [33]. Therefore, we expected a lower inbreeding level in AZ. The only comparable estimated inbreeding with our expectation was F_{ROH} , which showed lower inbreeding for AZ (0.043) than for KHZ (0.059).

The highest correlation was observed between F_{UNI} and F_{ROH} (AZ=0.98 and KHZ=0.94). Literature has reported different correlation coefficients between F_{UNI} and F_{ROH} (0.15–0.80) [14], between F_{HOM} and F_{ROH} (0.06–0.95) [14, 21, 27], and between F_{GRM} and F_{ROH} (0.17–0.81) [21, 37, 38]. The considerable variation among different studies may be because of a strong dependency of F_{HOM} , F_{UNI} and F_{GRM} on allelic frequencies [39].

The F_{PED} of 0.03 previously reported in Iranian buffalo [40] was lower than the estimated F_{ROH} in the current study. Given that pedigree data were not available for our study, we could not calculate F_{PED} and compare it with F_{ROH} . However, previous studies reported moderate to high (0.47–0.82) and low to moderate (0.12–0.76) correlations between F_{PED} and F_{ROH} in cattle and sheep, respectively [14, 17]. A low to moderate correlation between F_{PED} and F_{ROH} was also reported by Peripolli et al. [13] in Gyr cattle, suggesting that F_{PED} may not accurately capture small IBD segments that result from ancient inbreeding. Further, accurate and in-depth pedigree records are required to measure F_{PED} . Additionally, methods based on allelic frequency have demonstrated considerable variation among different breeds [14]. Given that ROH does not depend on allele frequencies, and can capture recent and ancient inbreeding, it seems to be a suitable method for measuring inbreeding.

The total length of ROH islands were about 6 and 15 Mb in the AZ and KHZ breeds, respectively (Additional file 3). Consequently, fewer genes were identified in ROH islands in the AZ breed than in the KHZ breed; that is probably why the genes located in ROH islands of the AZ breed were not enriched in any GO terms ($P > 0.05$). In the KHZ breed, however, the genes located in the ROH islands were significantly enriched ($P \leq 0.05$) in 40 GO terms (Additional file 4). These GO terms belonged to 23 biological processes (BP), 12 cellular component (CC) and 5 molecular function (MF) groups. In this paper, we focused principally on the GO terms that include the genes with known large effects on important traits in livestock.

Five genes were identified with positive regulation of DNA metabolic development (GO:0051054) in the BP group. Among these genes, STAT6 (signal transducer and activator of transcription 6, on BTA5) has been reported to have large effects on the growth efficiency and the quality of carcass in cattle [41]. Additionally, using co-expression network analysis, Nguyen et al. [42]. reported the critical role of STAT6, PBX2 (PBX homeobox2) and PBRM1 (Protein polybromo1) as transcription factors in regulating pubertal development in Brahman heifers.

Twelve genes in ROH islands were associated with lipid metabolic process (GO:0006629) in the BP group. Of these genes, BMP2 (bone morphogenetic protein 2, on BTA13) plays a major role in rebuilding

hair follicles in goats [43]. Further, BMP2 in porcine, cattle and sheep has been reported to have an influence on regulating body size and muscle development [44-47]. Kim et al. [48] found several signatures of selection containing genes such as BMP2 associated with body size and development in goats and sheep native to Egypt. These researchers concluded that the genes influencing body size may be important in regulating adaptation to hot, arid habitats because efficiency in thermoregulation can be associated with body size. Supporting their conclusion is the fact that most breeds in tropical zones have smaller body size than breeds in temperate zones because tropical breeds can regulate their body temperature more efficiently [49]. However, other factors that differ between temperate and arid zones may also contribute to variations in the body sizes of breeds living in different climates.

CYP27B1 (cytochrome P450 family 27 subfamily B member 1) located on BTA5 was also one of the genes enriched in the lipid metabolic process (GO:0006629). This gene is important for making 1- α -hydroxylase, which is required in vitamin D bio-activation, and has been reported to be up-regulated as a result of bacterial infection, suggesting that this gene plays a role in modulating innate immune responses [50].

In ROH islands on BTA13, three genes were associated with the positive regulation of DNA replication (GO:0045740) in the BP group. Proliferating cell nuclear antigen (PCNA) has been reported to be associated with follicular development and growth in buffalo ovaries [51], and may therefore be related to fertility performance.

Single-organism cellular process (GO:0044763) with 64 genes was significantly enriched ($P=0.05$) in ROH islands, including MARS (methionyl-tRNA synthetase, on BTA5), and ADRA1D (adrenoceptor alpha 1D, on BTA13). MARS has been reported to influence milk and protein production in Chinese [52] and Portuguese [53] Holstein cattle, and ADRA1D largely affects milk protein in Murrah dairy buffalo [54]. INHBC and INHBE (inhibin beta C and E subunits, on BTA5) have been reported as candidate genes associated with reproductive performance in tropical young bulls [55], and composite reproductive traits in Lori-Bakhtiari sheep [56]. KCTD16 (potassium channel tetramerization domain containing 16, on BTA7) was reported as a candidate gene for meat quality in Simmental beef cattle [57], for residual feed intake in Junmu White pigs [58], and for fat yield in Nordic Holstein cattle [59]. PMEL (premelanosome protein) and MYO1A (myosin IA) on BTA5 have been reported as putative candidate genes related to coat colour phenotypes in cattle [60, 61]. PMEL is required for the melanin biosynthesis process in the pigmentation of hair, mucous membranes and eyes [62]. In cattle, PMEL is reported as a candidate gene associated with the dilution of coat colour and consequently colour intensity [63, 64]. Light coat colouring can be beneficial for animals in adapting to hot climates because it can help them to reduce sunlight absorption [65]. However, most of the AZ and KHZ buffalo have a dark coat, which could be a result of some other favourable traits associated with a darker coat colour or the result of artificial selection caused by human interference. SUOX (Sulphite oxidase, on BTA5), within this BP category, was reported to be associated with bone development in cattle [66].

The average LD (r^2) between adjacent SNPs in ROH islands was higher than the r^2 of adjacent SNPs located on the same chromosome (Additional file 3). Thus, the recombination rates in the ROH islands were lower than those in the rest of the genome. These results are in line with some previous studies [13, 17]. However, a moderate recombination rate has been reported between the SNPs in ROH islands in Valle del Belice sheep [67]. Additionally, ROH hotspots can result from a wide range of underlying causes such as inbreeding and selection [12]. Peripolli et al. [13] argued that the high LD observed in most ROH hotspots is not necessarily caused by selection or conserved IBD haplotypes, but can be an indication of a lower recombination rate in those regions. Nevertheless, most of the ROH hotspots in our study overlapped with selection signature regions found with iHS, which supports the theory that ROH can be used to find genomic regions that have been under natural and/or artificial selection.

Buffalo species have a relatively lower heat tolerance capability than some other livestock species because of their inadequately dispersed sweat glands and their dark coat colour [68]. However, Iranian buffalo breeds have historically been raised in a hot climate [69]. Therefore, selection for higher heat tolerance may have occurred in Iranian buffalo for better adaptation to heat stress [5]. It has been reported that combined networks of multiple genes are often involved in the regulation of complex traits such as adaptation to hot climates [48, 70, 71]. Thus, selection for complex traits would leave only minor footprints because of the selection for numerous regions with lower intensity across the genome [70]. Therefore, we expected to find several genes directly or indirectly influencing different traits that were under artificial selection or important for adaptation and survival in hot areas. We found genes influencing energy and digestive metabolism (KCTD16), autoimmune response (CYP27B1), thermoregulation (BMP2), embryonic development and reproduction (STAT6, PCNA, INHBC and INHBE). These genes seem to be important for species such as water buffalo that have adapted to a hot climate [48].

Conclusion

The inbreeding coefficients based on F_{HOM} , F_{UNI} and F_{GRM} were higher in the AZ breed than they were in the KHZ breed, which contradicted our expectations according to higher N_e in AZ breed. Given that F_{ROH} was the only measurement of inbreeding in our study that showed AZ water buffalo were more inbred, this measurement seems to be a suitable measure of genomic inbreeding. This is most likely because it is less affected by allele frequencies. Further, knowing the distribution of ROH across the genome, inbreeding can be avoided more efficiently through mating allocation. Additionally, frequently occurring ROH can be used as suggestive evidence of historical selection. In our study, we found some overlap between ROH islands and genomic regions showing signatures of selection in previous studies of AZ and KHZ breeds. Therefore, the genes located in ROH islands could be under the influence of artificial and/or natural selection. We found that the genes located in ROH islands were associated with biological pathways such as adaptation to a hot climate, immune response, milk production, growth efficiency, reproduction performance and bone development.

Methods

Sample collection, ethical statement, and data quality control

Hair roots and blood samples were obtained from 112 herds of AZ and 47 herds of KHZ breeds. Samples of the AZ breed were gathered from East and West Azerbaijan, Gilan and Ardabil (37.02° – 38.78° N, 44.81° 49.52°E), which are north-western provinces of Iran. Samples of the KHZ breed were obtained from Kermanshah (34.54°N, 45.60°E) and Khuzestan (30.68–32.55° N, 48.02°–48.97° E), which are the south and south-western provinces of Iran, respectively (Fig. 1). All practices relating to data collection were reviewed and confirmed by the research ethics committee of the College of Agriculture and Natural Resources of the University of Tehran, Iran and by the ABCI. Three hundred and sixty-nine buffalo (254 AZ and 115 KHZ) were genotyped using 90K SNPChip (Axiom[®] Buffalo 90K Genotyping Array), which consisted of 89 988 almost evenly distributed SNPs throughout the genome. The same dataset was previously used by Mokhber et al. [5], and it partially overlapped with the dataset used by Colli et al. [4] and by Fallahi et al. [72].

The SNPs in the 90K SNPChip were selected using buffalo DNA sequence, but similar to methods used in previous studies [1, 5, 33, 72-75], were reported according to the location on the cattle reference genome assembly (UMD3.1 [76]) Although chromosome-level assembly of the water buffalo genome (UOA_WB_1) has been published recently [77], we used the UMD3.1 assembly in our study because it is more reliable and has better gene annotation information. Genotypes were obtained through AffyPipe [78], and all the monomorphic and polymorphic SNPs with high resolution (n=64 750) were stored. According to the filtration criteria, samples with more than 5% missing genotype and SNPs with 5% missing rate were eliminated from further analyses. We also filtered out SNPs with unidentified position in the UMD3.1 assembly, positioned on the sex chromosomes, with minor allele frequency of <2%, and with p-value for the Hardy–Weinberg equilibrium chi-square test <10⁻⁶. In total, 62 122 SNPs and 369 samples with an average call rate of 99.6% passed the quality-control filters.

Genetic distance between breeds

Genetic distance, which is based on the IBS matrix, was estimated through the `-ibs-matrix` command in PLINK v1.9 [79]. Principal component (PC) analysis of genetic distances was performed to visualise the genetic diversity of the samples, and was depicted using R (<http://www.R-project.org/>). According to the first and second PCs, we removed four samples: two from each breed that were placed outside their expected breed cluster.

Runs of homozygosity analyses

ROH can be detected in the genome through two main approaches: 1) genotype-counting algorithms in which the genome is scanned to identify long stretches of consecutive homozygous genotypes like the one implemented in PLINK v1.9 [79], and 2) model-based methods that utilize Hidden Markov Models (HMM) like the one implemented in RzoRoH [80]. This package can enable a better assessment of the

contribution of various generations to the current level of inbreeding, estimating inbreeding at both genome-wide and local scales, and classifying homozygous-by-descent (HBD) segments into age-based classes [81]. However, we used PLINK in our study because of the simplicity of running the sliding-window approach to detect ROH with sufficiently high assurance [82].

A genome scan for ROH was conducted for the AZ (n=252) and KHZ (n=113) breeds, separately. For each individual, ROH segments with the following attributes were identified: 1) each ROH stretch was at least 1 Mb in length; 2) there was at the most only one heterozygous and one missing SNP in each ROH; 3) there was a minimum number of SNPs that could form ROH in each breed, calculated according to Eq. 1 to control the false positive rate of the identified ROH. (see Equation 1 in the Supplementary Files)

where l is the minimum number of SNPs in ROH, α is the false positive rate of the identified ROH (set at 0.5); n_a and n_s are the number of individuals and the number of SNPs per individual, respectively; and het is the average heterozygosity across individuals. l was calculated to be 40 and 38 in AZ and KHZ breeds, respectively; 4) each ROH contained at least one SNP over 100 Kb; and 5) the maximum gap between two neighbouring SNPs in ROH had to be less than 1 Mb.

The ROH that had these five attributes were divided into the following five groups: 1-2, 2-4, 4-8, 8-16 and >16 Mb, as suggested in the literature [13, 15, 25]. Then for each breed, the frequency and the average length (Mb) of ROH within each category, the percentage of each ROH category, and the percentage of genome coverage by each ROH category were calculated, using R (<http://www.R-project.org/>).

Inbreeding coefficient estimations

The coefficient of inbreeding was estimated using ROH (F_{ROH}), excess of homozygosity (F_{HOM}), correlation between uniting gametes (F_{UNI}) and diagonal elements of the genomic relationship matrix (F_{GRM}).

F_{ROH} was calculated for each individual using Eq. 2 [20]: (see Equation 2 in the Supplementary Files)

where f_i is the inbreeding coefficient of animal i ; n is the total number of ROH; and l_j is the length of the j^{th} ROH in animal i ; L_{aut} is the total autosome length covered by the SNP markers (2.5 Gb in our study).

We also calculated the following three different genomic inbreeding estimations: F_{GRM} (Eq. 3), F_{HOM} (Eq. 4) and F_{UNI} (Eq. 5) using `-ibc` command in GCTA software [39]. (see Equations 3-5 in the Supplementary Files)

where x_i and p_i are the number of copies and the frequency of the reference allele for SNP i , respectively; h_i is $2p_i(1-2p_i)$; and n is the total number of SNPs. The Pearson's correlation coefficient between F_{ROH} and the other genomic inbreeding estimates was also calculated.

Frequently appearing runs of homozygosity and gene enrichment analyses

To detect the genomic regions frequently covered with ROH in the AZ and KHZ populations, the number of times each SNP occurred in ROH was calculated separately in each breed. The ROH repeated in more than 20% of the individuals in each breed (approximately less than 1% of the SNPs) were nominated ROH islands, as suggested in previous studies [25, 67]. Further, the frequency of ROHs were plotted against their physical position along UMD3.1.

To identify genes in ROH islands, we used UMD3.1 map viewer from the NCBI website (<https://www.ncbi.nlm.nih.gov/mapview/>). Additionally, to find significantly enriched Gene Ontology (GO) terms ($P \leq 0.05$) of the genes located in ROH peaks, we used DAVID v6.8 tool [83, 84]. Finally, we performed an extensive literature review to explore the biological function of the annotated genes in ROH islands.

To discover whether ROH islands were associated with regions of the genome with a low recombination rate, the average LD of all the adjacent SNPs across each chromosome was compared with the average LD between adjacent SNPs inside the ROH islands located on the same chromosome. Additionally, to discover whether the ROH hotspots were associated with genomic regions that showed signatures of selection through other methods, we compared ROH islands with integrated haplotype homozygosity scores (iHS) that had already been published for AZ and KHZ breeds [5]. An iHS is a measure of haplotype homozygosity based on the difference between observed LD structure around a selected allele relative to the expected LD pattern according to the whole genome [85]. Therefore, it can be used to detect regions under historical selection [85].

List Of Abbreviations

ROH: runs of homozygosity; SNP: single nucleotide polymorphism; AZ: Azeri breed; KHZ: Khuzestani breed; MZ: Mazandarani breed; GO: Gene Ontology; IBD: identical by descent; IBS: identical by state; N_e : effective population size; FROH: inbreeding coefficient calculated using ROH; FHOM: inbreeding coefficient calculated using excess of homozygosity; FUNI: inbreeding coefficient calculated using correlation between uniting gametes; FGRM: inbreeding coefficient calculated using diagonal elements of the genomic relationship matrix; LD: linkage disequilibrium; ABCI: Animal Breeding Centre of Iran; PC: principal component; iHS: integrated haplotype homozygosity score.

Declarations

Ethics approval and consent to participate

The procedure was in accordance with animal ethics and approved by the University of Tehran and Animal Breeding Centre of Iran (ABCI) authorized representatives. Indications, risks, and benefits explained to animal owners and depending on his/her literacy, written or verbal informed consent obtained before the initiation of sampling.

Consent for publication

Not applicable.

Availability of data and material

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Competing interest

The authors declare that they have no competing interests.

Funding

We have received no specific funding for the current study.

Authors' contributions

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

Acknowledgments

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

References

1. Iamartino D, Nicolazzi EL, Van Tassell CP, Reecy JM, Fritz-Waters ER, Koltjes JE, et al. Design and validation of a 90K SNP genotyping assay for the water buffalo (*Bubalus bubalis*). PLoS One. 2017;12(10):e0185220.
2. Safari A, Ghavi Hossein-Zadeh N, Shadparvar AA, Abdollahi Arpanahi R. A review on breeding and genetic strategies in Iranian buffaloes (*Bubalus bubalis*). Trop Anim Health Prod. 2018;50(4):707-14.
3. Yindee M, Vlamings B, Wajjwalku W, Techakumphu M, Lohachit C, Sirivaidyapong S, et al. Y-chromosomal variation confirms independent domestications of swamp and river buffalo. Anim Genet. 2010;41(4):433-5.
4. Colli L, Milanesi M, Vajana E, Iamartino D, Bomba L, Puglisi F, et al. New insights on water buffalo genomic diversity and post-domestication migration routes from medium density SNP chip data. Front Genet. 2018;9:53.
5. Mokhber M, Moradi-Shahrbabak M, Sadeghi M, Moradi-Shahrbabak H, Stella A, Nicolazzi E, et al. A genome-wide scan for signatures of selection in Azeri and Khuzestani buffalo breeds. BMC Genom. 2018;19(1):449.

6. Dhanda O: Developments in water buffalo in Asia and Oceania. Manila, Philippines: 7th World Buffalo Congr. 2004;17-28.
7. De Camargo G, Aspilcueta-Borquis RR, Fortes M, Porto-Neto R, Cardoso DF, Santos D, et al. Prospecting major genes in dairy buffaloes. *BMC Genom.* 2015;16(1):872.
8. Wright S. Coefficients of inbreeding and relationship. *Am Nat.* 1922;56(645):330-8.
9. Howrigan DP, Simonson MA, Keller MC. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genom.* 2011;12(1):460.
10. Broman KW, Weber JL. Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *Am J Hum Genet.* 1999;65(6):1493-500.
11. Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. *Hum Mol Genet.* 2006;15(5):789-95.
12. Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: windows into population history and trait architecture. *Nat Rev Genet.* 2018;19(4):220.
13. Peripolli E, Stafuzza NB, Munari DP, Lima ALF, Irgang R, Machado MA, et al. Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (*Bos indicus*) dairy cattle. *BMC Genom.* 2018;19(1):34.
14. Zhang Q, Calus MP, Guldbbrandtsen B, Lund MS, Sahana G. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC Genet.* 2015;16(1):88.
15. Marras G, Gaspa G, Sorbolini S, Dimauro C, Ajmone-Marsan P, Valentini A, et al. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Anim Genet.* 2015;46(2):110-21.
16. Peripolli E, Munari D, Silva M, Lima A, Irgang R, Baldi F. Runs of homozygosity: current knowledge and applications in livestock. *Anim Genet.* 2017;48(3):255-71.
17. Purfield DC, McParland S, Wall E, Berry DP. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PLoS One.* 2017;12(5):e0176780.
18. Purfield DC, Berry DP, McParland S, Bradley DG. Runs of homozygosity and population history in cattle. *BMC Genet.* 2012;13(1):70.
19. Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF. Genomic runs of homozygosity record population history and consanguinity. *PloS One.* 2010;5(11):e13996.
20. McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of homozygosity in European populations. *Am J Hum Genet.* 2008;83(3):359-72.
21. 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.* 2016;10(5):746-54.
22. Gaspa G, Marras G, Sorbolini S, Ajmone Marsan P, Williams J, Valentini A, et al. Genome-wide homozygosity in Italian Holstein cattle using HD panel. Vancouver, Canada: 10th World Congr. Genet. Appl. to Livest. Prod; 2014.

23. Gurgul A, Szmatoła T, Topolski P, Jasielczuk I, Żukowski K, Bugno-Poniewierska M. The use of runs of homozygosity for estimation of recent inbreeding in Holstein cattle. *J applied Genet.* 2016;57(4):527-30.
24. Mastrangelo S, Sardina M, Tolone M, Di Gerlando R, Sutera A, Fontanesi L, et al. Genome-wide identification of runs of homozygosity islands and associated genes in local dairy cattle breeds. *animal.* 2018;12(12):2480-8.
25. 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. *Livest Sci.* 2016;188:72-80.
26. Zavarez LB, Utsunomiya YT, Carmo AS, Neves HH, Carvalheiro R, Ferenčaković M, et al. Assessment of autozygosity in Nellore cows (*Bos indicus*) through high-density SNP genotypes. *Front Genet.* 2015;6:5.
27. Zhang Y, Young J, Wang C, Sun X, Wolc A, Dekkers J: Inbreeding by pedigree and genomic markers in selection lines of pigs. Vancouver, Canada: 10th World Congr. Genet. Appl. to Livest. Prod; 2014.
28. Bosse M, Megens H-J, Madsen O, Paudel Y, Frantz LA, Schook LB, et al. Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genet.* 2012;8(11):e1003100.
29. Xu Z, Sun H, Zhang Z, Zhao Q, Olasege BS, Li Q, et al. Assessment of Autozygosity Derived From Runs of Homozygosity in Jinhua Pigs Disclosed by Sequencing Data. *Front Genet.* 2019;10(274).
30. Ghoreishifar SM, Moradi-Shahrbabak H, Parna N, Davoudi P, Khansefid M. Linkage disequilibrium and within-breed genetic diversity in Iranian Zandi sheep. *Arch Anim Breed.* 2019;62(1):143-51.
31. Mastrangelo S, Portolano B, Di Gerlando R, Ciampolini R, Tolone M, Sardina M, et al. Genome-wide analysis in endangered populations: a case study in Barbaresca sheep. *animal.* 2017;11(7):1107-16.
32. Al-Mamun HA, Clark SA, Kwan P, Gondro C. Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. *Genet Sel Evol.* 2015;47(1):90.
33. Mokhber M, Moradi-Shahrbabak M, Sadeghi M, Moradi-Shahrbabak H, Stella A, Nicolazzi E, et al. Study of whole genome linkage disequilibrium patterns of Iranian water buffalo breeds using the Axiom Buffalo Genotyping 90K Array. *PLoS One.* 2019;14(5):e0217687.
34. Ferenčaković M, Sölkner J, Curik I. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genet Sel Evol.* 2013;45(1):42.
35. Ferenčaković M, Hamzić E, Gredler B, Curik I, Sölkner J. Runs of homozygosity reveal genome-wide autozygosity in the Austrian Fleckvieh cattle. *Agric Conspec Sci.* 2011;76(4):325-9.
36. Kim E-S, Cole JB, Huson H, Wiggans GR, Van Tassell CP, Crooker BA, et al. Effect of artificial selection on runs of homozygosity in US Holstein cattle. *PloS One.* 2013;8(11):e80813.
37. 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. *J Dairy Sci.* 2013;96(7):4697-706.

38. Pryce JE, Haile-Mariam M, Goddard ME, Hayes BJ. Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genet Sel Evol.* 2014;46(1):71.
39. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88(1):76-82.
40. Ghavi Hossein-Zadeh N. Analysis of population structure and genetic variability in Iranian buffaloes (*Bubalus bubalis*) using pedigree information. *Anim Prod Sci.* 2016;56(7):1130-5.
41. Rincon G, Farber E, Farber C, Nkrumah J, Medrano J. Polymorphisms in the STAT6 gene and their association with carcass traits in feedlot cattle. *Anim Genet.* 2009;40(6):878-82.
42. Nguyen LT, Reverter A, Cánovas A, Venus B, Anderson ST, Islas-Trejo A, et al. STAT6, PBX2, and PBRM1 emerge as predicted regulators of 452 differentially expressed genes associated with puberty in Brahman heifers. *Front Genet.* 2018;9:87.
43. 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. *Can J Anim Sci.* 2009;89(4):457-62.
44. Fan B, Onteru SK, Du Z-Q, Garrick DJ, Stalder KJ, Rothschild MF. Genome-wide association study identifies loci for body composition and structural soundness traits in pigs. *PLoS One.* 2011;6(2):e14726.
45. 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.* 2019;10(4):293.
46. Kijas JW, Lenstra JA, Hayes B, Boitard S, Neto LRP, San Cristobal M, et al. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol.* 2012;10(2):e1001258.
47. Xu L, Bickhart DM, Cole JB, Schroeder SG, Song J, Tassell CPV, et al. Genomic signatures reveal new evidences for selection of important traits in domestic cattle. *Mol Biol Evol.* 2014;32(3):711-25.
48. Kim E-S, Elbeltagy A, Aboul-Naga A, Rischkowsky B, Sayre B, Mwacharo JM, et al. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity.* 2016;116(3):255.
49. McManus C, Louvandini H, Gugel R, Sasaki LCB, Bianchini E, Bernal FEM, et al. Skin and coat traits in sheep in Brazil and their relation with heat tolerance. *Trop Anim Health Prod.* 2011;43(1):121-6.
50. Nelson CD, Reinhardt TA, Beitz DC, Lippolis JD. In vivo activation of the intracrine vitamin D pathway in innate immune cells and mammary tissue during a bacterial infection. *PLoS One.* 2010;5(11):e15469.
51. Sharma GT, Dubey PK, Katiyar A, Kumar GS. Localization and expression of proliferating cell nuclear antigen (PCNA) and cyclin B1 in buffalo (*Bubalus bubalis*) ovary during different stages of follicular development. *Indian J Anim Sci.* 2011;81(3):231-4.
52. Li C, Cai W, Zhou C, Yin H, Zhang Z, Loo JJ, et al. RNA-Seq reveals 10 novel promising candidate genes affecting milk protein concentration in the Chinese Holstein population. *Sci Rep.* 2016;6:26813.

53. 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*. 2014;131(131.83):131-83.
54. Araújo DN, de Camargo GMF, Fonseca PDdS, Cardoso DF, Hurtado-Lugo NA, Aspilcueta-Borquis RR, et al. Polymorphisms in Oxytocin and $\alpha 1a$ Adrenergic Receptor Genes and Their Effects on Production Traits in Dairy Buffaloes. *Anim Biotechnol*. 2015;26(3):165-8.
55. 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*. 2013;1(4):644-50.
56. Abdoli R, Mirhoseini SZ, Ghavi Hosseini-Zadeh N, Zamani P, Ferdosi MH, Gondro C. Genome-wide association study of four composite reproductive traits in Iranian fat-tailed sheep. *Reprod Fert Devleop*. 2019;31(6),1127-33.
57. Xia J, Qi X, Wu Y, Zhu B, Xu L, Zhang L, et al. Genome-wide association study identifies loci and candidate genes for meat quality traits in Simmental beef cattle. *Mamm Genom*. 2016;27(5-6):246-55.
58. Bai C, Pan Y, Wang D, Cai F, Yan S, Zhao Z, et al. Genome-wide association analysis of residual feed intake in Junmu No. 1 White pigs. *Anim Genet*. 2017;48(6):686-90.
59. Cai Z, Guldbbrandtsen B, Lund MS, Sahana G. Dissecting closely linked association signals in combination with the mammalian phenotype database can identify candidate genes in dairy cattle. *BMC Genet*. 2018;19(1):30.
60. Mészáros G, Petautschnig E, Schwarzenbacher H, Sölkner J. Genomic regions influencing coat color saturation and facial markings in Fleckvieh cattle. *Anim Genet*. 2015;46(1):65-8.
61. Edea Z, Dadi H, Dessie T, Uzzaman M, Rothschild M, Kim ES, et al. Genome-wide scan reveals divergent selection among taurine and zebu cattle populations from different regions. *Anim Genet*. 2018;49(6):550-63.
62. Urbinati I, Stafuzza NB, Oliveira MT, Chud TCS, Higa RH, de Almeida Regitano LC, et al. Selection signatures in Canchim beef cattle. *J Anim Sci Biotechnol*. 2016;7(1):29.
63. Kühn C, Weikard R. An investigation into the genetic background of coat colour dilution in a Charolais× German Holstein F2 resource population. *Anim Genet*. 2007;38(2):109-13.
64. Gutiérrez-Gil B, Wiener P, Williams JL. Genetic effects on coat colour in cattle: dilution of eumelanin and phaeomelanin pigments in an F2-Backcross Charolais× Holstein population. *BMC Genet*. 2007;8(1):56.
65. Asres A, Amha N. Physiological adaptation of animals to the change of environment: a review. *J Biol Agric Healthc*. 2014;4(25):146-51.
66. Drögemüller C, Tetens J, Sigurdsson S, Gentile A, Testoni S, Lindblad-Toh K, et al. Identification of the bovine Arachnomelia mutation by massively parallel sequencing implicates sulfite oxidase (SUOX) in bone development. *PLoS Genet*. 2010;6(8):e1001079.
67. Mastrangelo S, Tolone M, Sardina MT, Sottile G, Suter AM, Di Gerlando R, et al. Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in

- Valle del Belice sheep. *Genet Sel Evol.* 2017;49(1):84.
68. Silanikove N, Maltz E, Halevi A, Shinder D. Metabolism of water, sodium, potassium, and chlorine by high yielding dairy cows at the onset of lactation. *J Dairy Sci.* 1997;80(5):949-56.
69. Das S, Upadhyay R, Madan M. Heat stress in Murrah buffalo calves. *Livest Prod Sci.* 1999;61(1):71-8.
70. Kemper KE, Saxton SJ, Bolormaa S, Hayes BJ, Goddard ME. Selection for complex traits leaves little or no classic signatures of selection. *BMC Genom.* 2014;15(1):246.
71. Lv F-H, Agha S, Kantanen J, Colli L, Stucki S, Kijas JW, et al. Adaptations to climate-mediated selective pressures in sheep. *Mol Biol Evol.* 2014;31(12):3324-43.
72. Fallahi MH, Moradi-Shahrbabak H, Moradi-Shahrbabak M, Abdollahi-Arpanahi R, Gholami S. Detection of Haplotypic Structure for Genome of Azerbaijani Buffalo Using High Density SNP Markers. *Russian J Genet.* 2019;55(8):1000-7.
73. Ghoreishifar SM, Moradi-Shahrbabak H, Moradi-Shahrbabak M, Nicolazzi EL, Williams JL, lamartino D, et al. Accuracy of imputation of single-nucleotide polymorphism marker genotypes for water buffaloes (*Bubalus bubalis*) using different reference population sizes and imputation tools. *Livest Sci.* 2018;216:174-82.
74. El-Halawany N, Abdel-Shafy H, Abd-El-Monsif AS, Abdel-Latif MA, Al-Tohamy AF, El-Moneim OMA. Genome-wide association study for milk production in Egyptian buffalo. *Livest Sci.* 2017;198:10-6.
75. Li J, Liu J, Campanile G, Plastow G, Zhang C, Wang Z, et al. Novel insights into the genetic basis of buffalo reproductive performance. *BMC Genom.* 2018;19(1):814.
76. Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, et al. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 2009;10(4):R42-10.1186/gb-2009-10-4-r42.
77. Low WY, Tearle R, Bickhart DM, Rosen BD, Kingan SB, Swale T, et al. Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity. *Nat Commun.* 2019;10(1):260.
78. Nicolazzi EL, lamartino D, Williams JL. AffyPipe: an open-source pipeline for Affymetrix Axiom genotyping workflow. *Bioinformatics.* 2014;30(21):3118-9.
79. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4(1):7.
80. Bertrand A, Kadri NK, Flori L, Gautier M, Druet T. RZooRoH: An R package to characterize individual genomic autozygosity and identify homozygous-by-descent segments. *Methods Ecol Evol.* 2019;10:860-6.
81. Druet T, Gautier M. A model-based approach to characterize individual inbreeding at both global and local genomic scales. *Mol Ecol.* 2017;26(20):5820-41.
82. Gusev A, Lowe JK, Stoffel M, Daly MJ, Altshuler D, Breslow JL, et al. Whole population, genome-wide mapping of hidden relatedness. *Genome Res.* 2009;19(2):318-26.
83. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2008;37(1):1-13.

84. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2008;4(1):44.
85. Voight BF, Kudravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol.* 2006;4(3):e72.

Tables

Table 1. Summary of the detected runs of homozygosity (ROH) grouped according to their length (Mb).

ROH group	nROH ¹		Percentage		Average length (Mb)		Standard deviation (Mb)		Percentage of genome coverage	
	AZ	KHZ	AZ	KHZ	AZ	KHZ	AZ	KHZ	AZ	KHZ
ROH ₁₋₂	841	629	15.71	16.77	1.83	1.81	0.13	0.15	0.06	0.05
ROH ₂₋₄	2870	2119	53.62	56.51	2.70	2.71	0.53	0.53	0.31	0.23
ROH ₄₋₈	855	609	15.98	16.24	5.44	5.38	1.09	1.08	0.19	0.13
ROH ₈₋₁₆	484	241	9.04	6.42	11.23	11.22	2.25	2.21	0.22	0.11
ROH _{>16}	302	152	5.64	4.05	26.70	26.19	11.06	9.32	0.32	0.16

1. Number of runs of homozygosity segments

Table 2. Average inbreeding coefficients (\pm standard error) estimated using diagonal elements of genomic relationship matrix (F_{GRM}), excess of homozygosity (F_{HOM}), correlation between uniting gametes (F_{UNI}) and runs of homozygosity (ROH) >1 Mb (F_{ROH}) in Azeri (AZ) and Khuzestani (KHZ) breeds.

Breed	F_{GRM}	F_{HOM}	F_{UNI}	F_{ROH}
AZ	0.026 \pm 0.05	0.026 \pm 0.05	0.026 \pm 0.05	0.043 \pm 0.05
KHZ	0.019 \pm 0.04	0.021 \pm 0.06	0.021 \pm 0.05	0.059 \pm 0.04

Table 3. Correlation between inbreeding coefficients calculated using runs of homozygosity (ROH) >1 Mb (F_{ROH}) and estimated using diagonal elements of genomic relationship matrix (F_{GRM}), excess of homozygosity (F_{HOM}), and correlation between uniting gametes (F_{UNI}) in Azeri (AZ) and Khuzestani (KHZ) breeds.

Breed	Correlation coefficient		
	F_{GRM} - F_{ROH}	F_{HOM} - F_{ROH}	F_{UNI} - F_{ROH}
AZ	0.88	0.92	0.98
KHZ	0.78	0.93	0.94

Additional File Legends

Additional file 1: List of the detected runs of homozygosity (ROH) in Azeri (AZ) and Khuzestani (KHZ) breeds.

Additional file 2. The estimated inbreeding coefficient using runs of homozygosity (ROH) >1 Mb (F_{ROH}), diagonal elements of genomic relationship matrix (F_{GRM}), excess of homozygosity (F_{HOM}) and correlation between uniting gametes (F_{UNI}) in Azeri (AZ) and Khuzestani (KHZ) breeds.

Additional file 3. Frequently occurring runs of homozygosity (ROH) regions (i.e. ROH islands) in Azeri (AZ) and Khuzestani (KHZ) breeds. The last column represents the average r^2 of ROH islands divided by the average r^2 of each chromosome.

Additional file 4. The genes located in detected runs of homozygosity (ROH) islands in the Khuzestani (KHZ) breed that were significantly enriched ($P \leq 0.05$) in biological processes (BP), cellular component (CC) and molecular function (MF) Gene Ontology (GO) terms.

Additional file 5. List of integrated haplotype homozygosity scores (iHS) for all SNPs in Azeri (AZ) and Khuzestani (KHZ) breeds.

Additional file 6. Manhattan plot of integrated haplotype homozygosity score (iHS) across the genome.

Figures

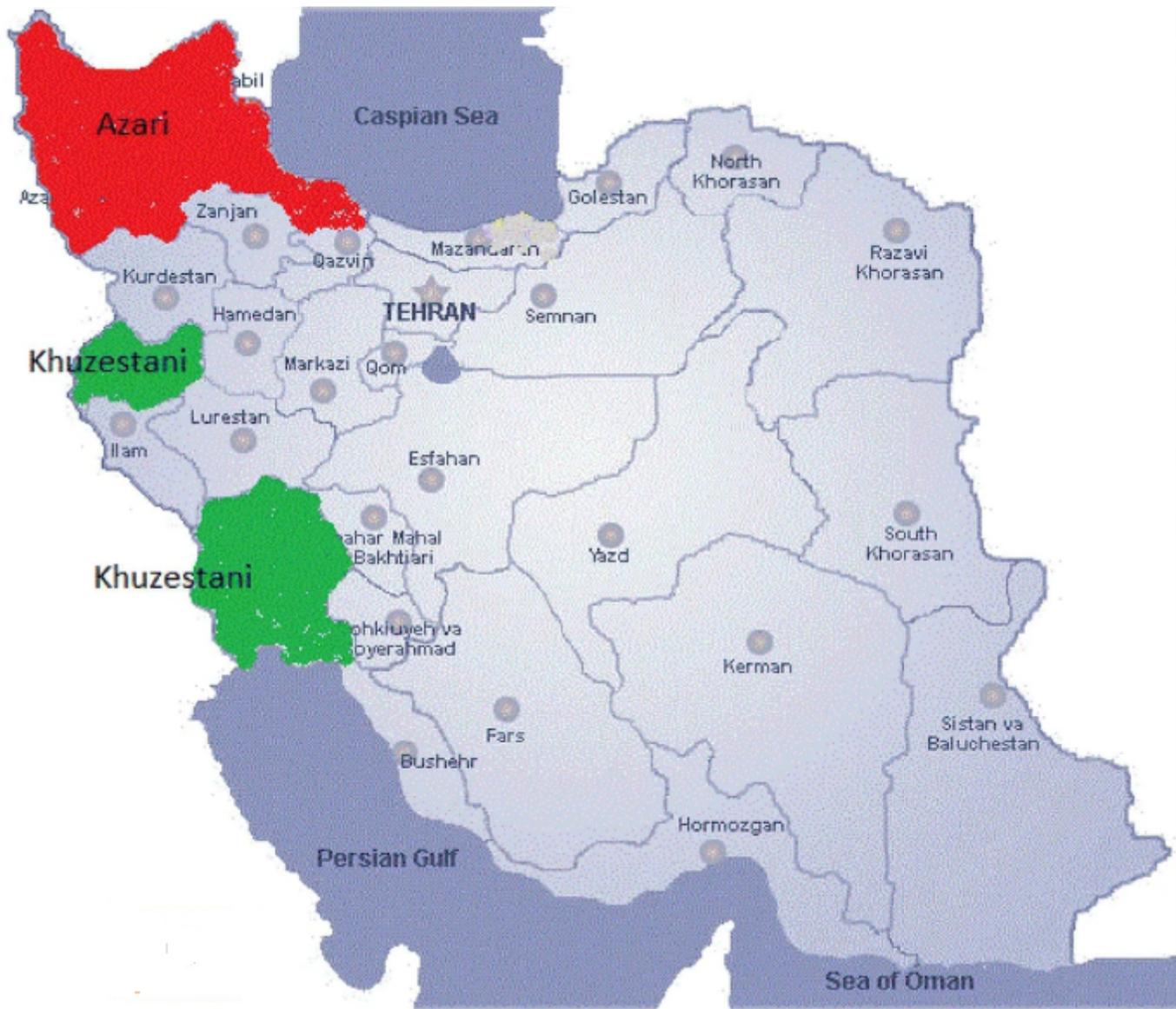


Figure 1

Geographic distribution of Azeri (AZ) and Khuzestani (KHZ) breeds used in this study. The samples of the AZ breed were obtained from the provinces shown in red (located in north and north-western part of Iran i.e. East and West Azerbaijan, Ardabil and Gilan). The samples for the Khuzestani (KHZ) breed were taken from the provinces shown in green (located in the west and south-western part of Iran i.e. Khuzestan and Kermanshah). Reprinted from “A genome-wide scan for signatures of selection in Azeri and Khuzestani buffalo breeds,” by Mahdi Mokhber et al., 2018; BMC Genom., 19(1), 449. Copyright 2018 by the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>). Reprinted with permission.

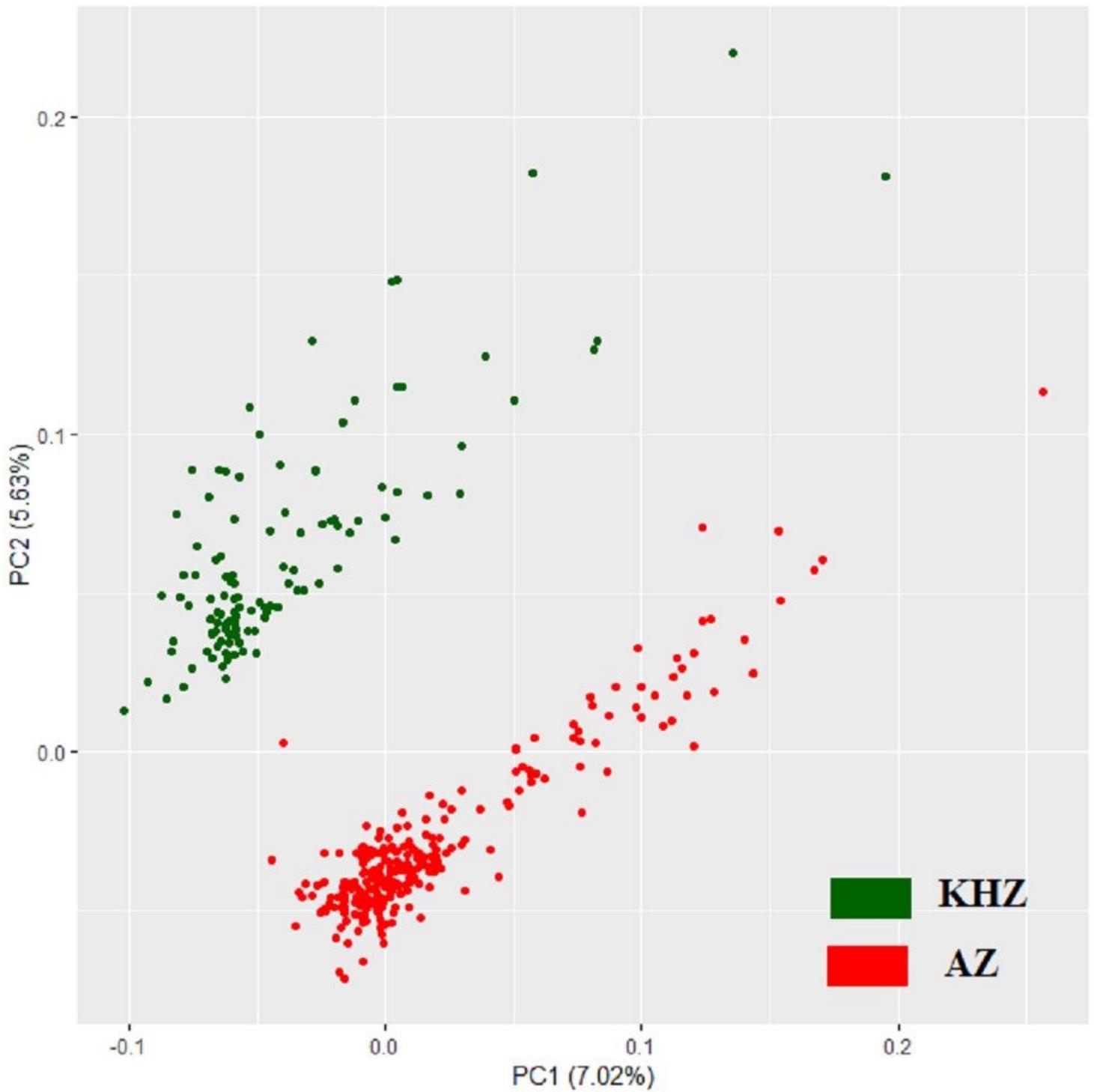


Figure 2

Azeri (AZ) and Khuzestani (KHZ) breeds clustered according to principal component (PC) analysis of identical by state (IBS) distance matrix. The first and second PCs explain 7.02% and 5.63% of the total variance, respectively.

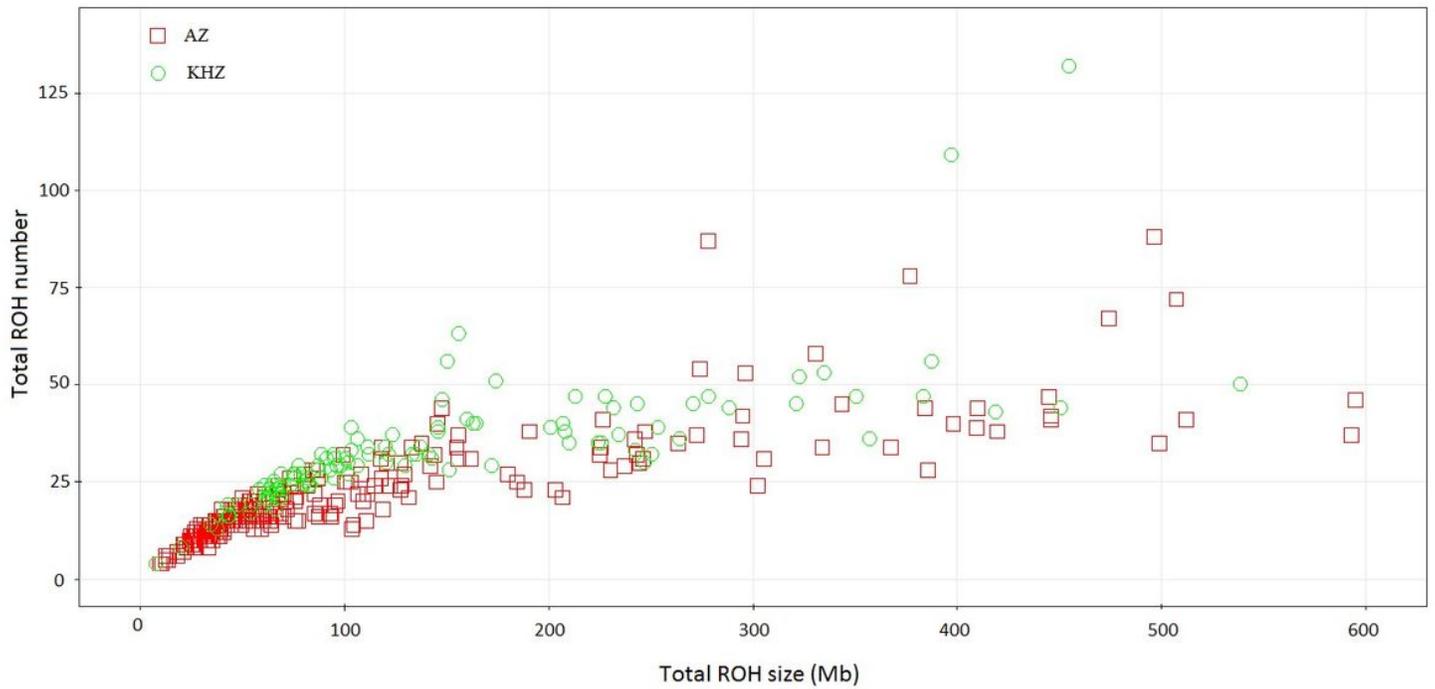


Figure 3

Number of runs of homozygosity (ROH) and the length of the genome covered by ROH in the samples taken from the Azeri (AZ) and Khuzestani (KHZ) breeds.

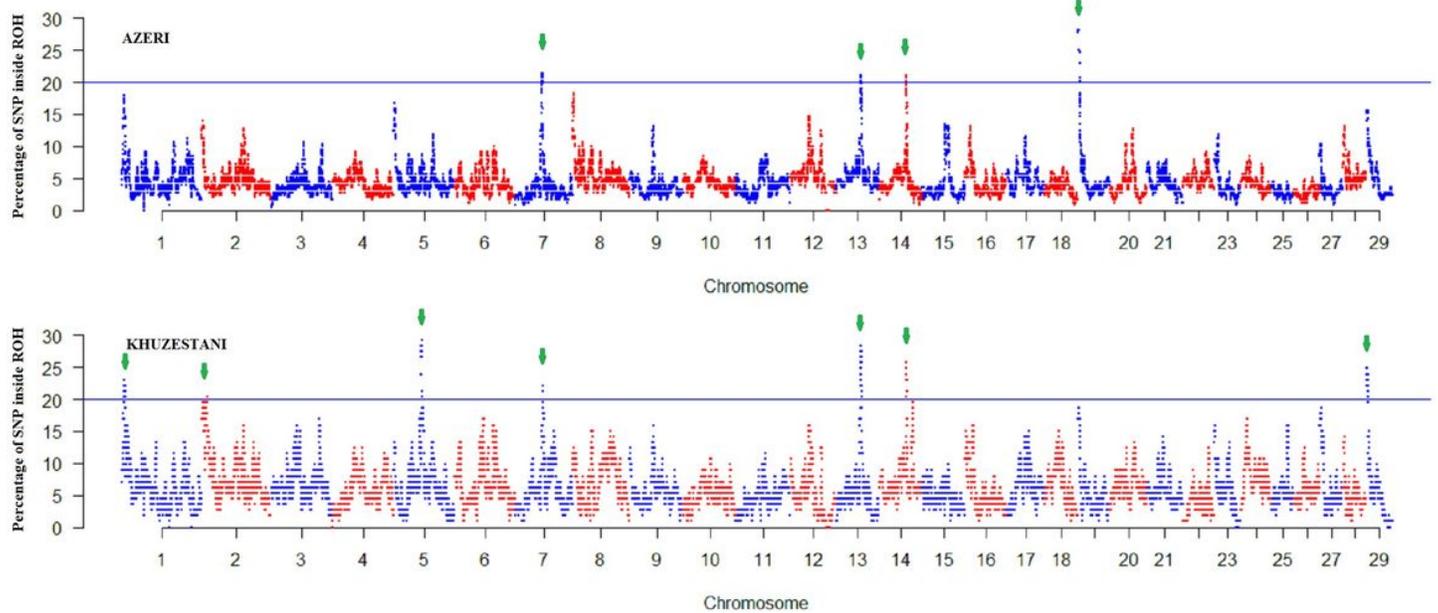


Figure 4

Manhattan plot of the distribution of frequently occurring runs of homozygosity (ROH) in Azeri (AZ) and Khuzestani (KHZ) Iranian water buffalo breeds. The X-axis shows the distribution of ROH over the

genome, and the Y-axis shows the percentage of ROH shared among animals within each breed. The significance threshold of 20% (less than 1% of all SNPs) shown as a blue line is used for detecting ROH islands (green arrows).

Supplementary Files

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

- [AdditionalFile6.tif](#)
- [Additionalfile1.csv](#)
- [additionalfile5.csv](#)
- [Equations.pdf](#)
- [additionalfile4.csv](#)
- [additionalfile3.csv](#)
- [additionalfile2.csv](#)