



OPEN

Identification of selection signatures in Iranian dromedary and Bactrian camels using whole genome sequencing data

Reza Khalkhali-Evrigh¹, Nemat Hedayat^{1✉}, Liang Ming² & Jirimutu²

The Old World camels play an important role as one of the main food sources in large parts of Asia and Africa. Natural selection combined with artificial selection by human has affected parts of the domestic animal genome for adapting them to their habitats and meeting human needs. Here, we used whole genome sequencing data of 34 camels (including 14 dromedaries and 20 Bactrian camels) to identify the genomic signature of selection in the Iranian dromedary (ID) and Bactrian camels (IB). To detect the mentioned regions, we used two methods including population differentiation index (F_{st}) and cross-population extended haplotype homozygosity (XP-EHH) with 50 kb sliding window and 25 kb step size. Based on gene ontology analysis on the candidate genes identified for IB camels, we found GO terms associated with lung development, nervous system development, immune system and behavior. Also, we identified several genes related to body thermoregulation (ZNF516), meat quality (ANK1 and HSPA13), and high-altitude adaptation (OPA1) for IB camels. In the list of detected candidate genes under selection in ID camels, the genes related to energy metabolism (BDH1), reproduction (DLG1, IMMP2L and FRAS1), long-term memory (GRIA1), kidney (SLC12A1), lung development (EMILIN2 and FBN1) and immunity (SOCS2, JAK1, NRROS and SENP1) were found. Our findings, along with further studies in this field, will strengthen our knowledge about the effect of selection on the camelid genome under different geographical, climatic and even cultural conditions.

Human beings started to lead a settled life based on agriculture about 12,000 years ago (ya) in Near East¹. The shift of the human lifestyle from hunter-gatherer to farmer changed the human attitude about the environment. One of the most important and influential events in the history of our species was the attempt to domesticate plants and animals. Following dogs as the first domesticated animal (~ 23,000 ya in Siberia)² which helped increase human security, animals such as sheep (~ 11,000 ya in the Middle East), goat (~ 10,500 ya in the Middle East), cattle (~ 10,300 ya in the Middle East), and horse (~ 5500 ya in central Asia) were domesticated to feed the farmer human³. It can be mentioned that Old World camels including dromedaries (1100–1800 BCE in Arabian Peninsula)⁴ and Bactrian camels (4450 ya in Central Asia)⁵ are the last domesticated animals among large mammals.

After domestication, artificial selection by humans along with climate-mediated selection were the most important factors for shaping the genetic and phenotypic structure of animals across different regions of the world. These factors have left their footprints on the specific region of the organism genome as the signature of selection. A decline in genetic variation has occurred in these regions followed by fixation of the selectively favorable alleles in the population after several generations⁶. Today, high-throughput techniques, availability of genomic data and advances in computational methods have made it possible to investigate the genomes of different organisms from various aspects.

Iran is one of the few countries that is home to both dromedary and Bactrian camels. From the past to the present, Bactrian camels have played a very important role in the lives of nomads inhabiting northwestern Iran (especially Ardabil province), as this area is one of the possible places for the domestication of Bactrian camels⁵. On the other hand, dromedary camels are spread in the central deserts of Iran and supply protein to many people in these areas. The devastating effects of climate change on the agricultural and livestock industry⁷ have necessitated the need for extensive studies on camels as one of the most resistant animals to harsh environmental

¹Department of Animal Science, Faculty of Agriculture and Natural Recourses, University of Mohaghegh Ardabili, Ardabil, Iran. ²College of Food Science and Engineering, Inner Mongolia Agricultural University, Huhhot, China. ✉email: nhedayat@uma.ac.ir

Genomic region	Dromedary	Bactrian
Intron	1,558,384	3,071,665
Intergenic	1,438,870	2,782,614
Up/Downstream	407,780	753,105
Intragenic	94,139	189,288
UTR	47,158	89,849
EXON (Synonymous)	23,949	44,148
EXON (Non-Synonymous)	18,696	30,913
EXON (Start/Stop altering)	292	534
Spicing-Site	3879	6890
Others	13,105	23,159
Total	3,606,252	6,992,165

Table 1. Proportion of annotated SNPs in different genomic regions for dromedary and Bactrian camels.

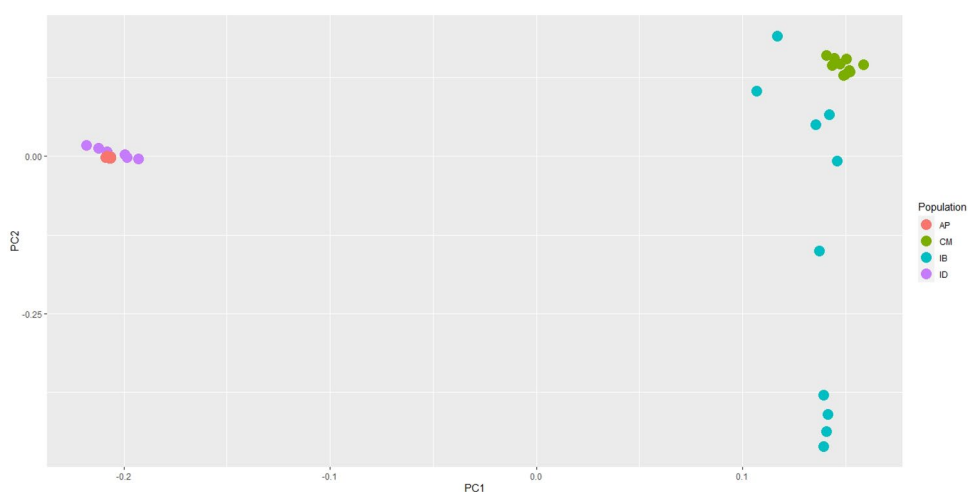


Figure 1. PCA plot visualized by R package ggplot2 (<https://cran.r-project.org/web/packages/ggplot2/>); showing the two first PCs on all used dromedary and Bactrian camels in this study.

conditions. It is a species that in the not-too-distant future could become a main source of food for large parts of Africa and Asia by becoming a farm animal.

In recent years, although many studies have been done to identify the selection signs in different species of domestic animals, including cattle^{8,9}, sheep^{10,11}, goat^{12,13} and horse^{14,15}, but the studies of Bahbahani et al.¹⁶ on Sudanese dromedary camels and Liu et al.¹⁷ on Chinese Bactrian camels are among the few cases in this field for camels. In the present study, we used whole genome sequencing (WGS) data to identify the selection footprints on the genome of Iranian dromedary (ID) and Iranian Bactrian (IB) camels as a step to better understand these valuable species. Here, WGS data of dromedary camels from Arabian Peninsula (AP) and Bactrian camels from China-Mongolia (CM) were also used.

Results

Genomic data and SNP calling. Whole genome resequencing analysis was applied on 14 dromedary and 20 Bactrian camels to identify the signature of selection for ID and IB camels. A total of 2.23 and 2.67 billion cleaned reads were generated for dromedary and Bactrian camels, respectively. After removing duplicated reads, the mean depth was 15.64X and 14.27X for dromedary and Bactrian camels, respectively. We detected a total of 3,606,252 autosomal SNPs for dromedary and 6,992,165 autosomal SNPs for Bactrian camels (Table 1).

Population genetic structure. We used LD-pruned SNPs to obtain a general view of the genetic relationship between dromedary and Bactrian camels. As displayed in Fig. 1, PC1 separated two species and the PC2 reflects the genetic diversity among IB camels. Separation of IB from CM Bactrian camels is more clearly visible than ID from AP dromedary camels. In addition, PCA results for each species were presented in Supplementary Figs. S1 (dromedary camels) and S2 (Bactrian camels) separately.

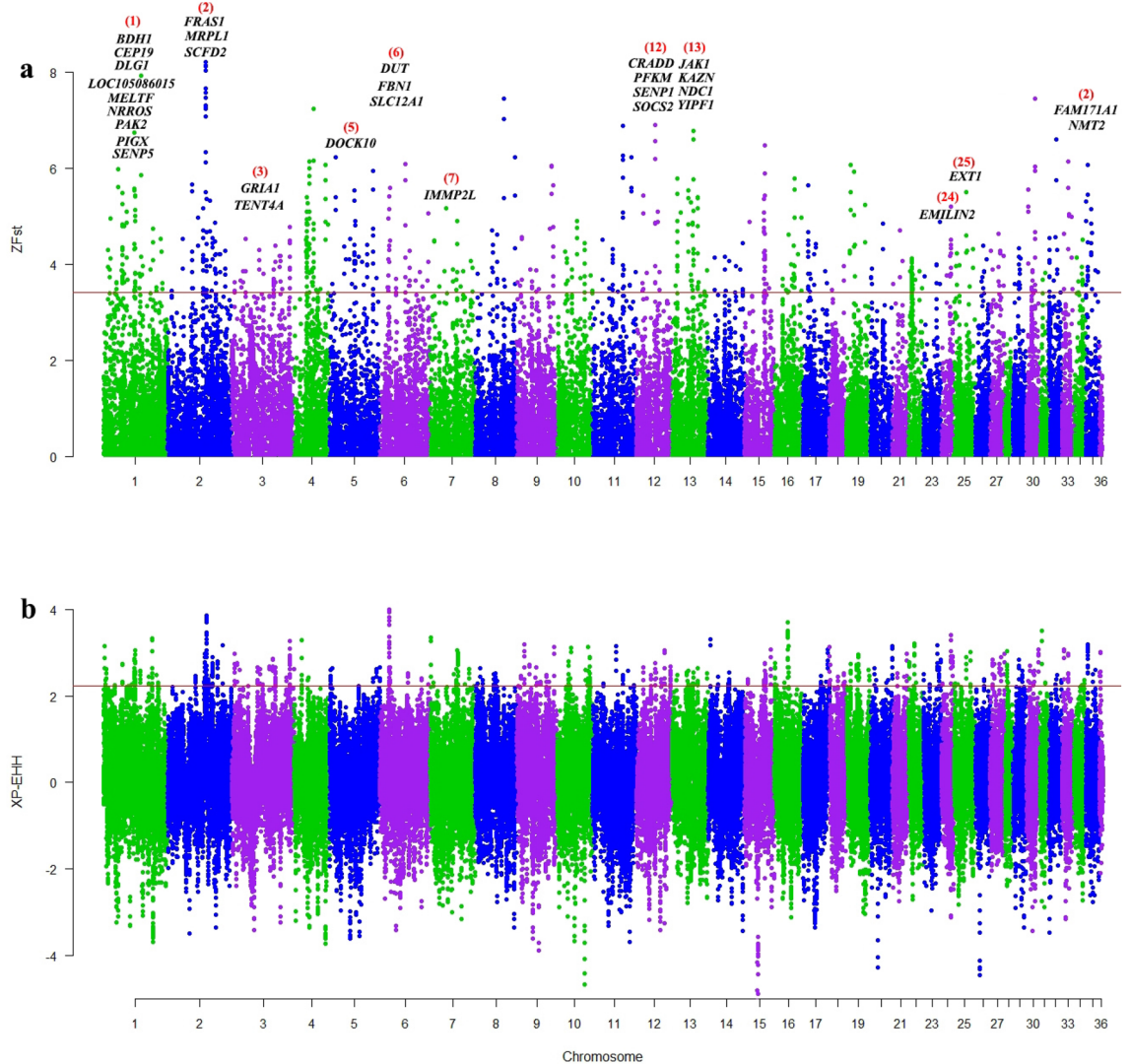


Figure 2. Manhattan plot visualized by R package qqman (<https://cran.r-project.org/web/packages/qqman/>); showing the genome-wide distribution of ZFst (a) and XP-EHH (b) between ID and AP camels. The red line denotes a threshold of ZFst > 3.41 and XP-EHH > 2.24. The gene symbols inserted in the figure were identified as positively selected genes using both methods. The red numbers in parentheses represent the chromosome number.

Genome-wide signature of selection. Here we used two comparative sets including ID vs AP dromedary camels and IB vs MC Bactrian camels to identify the signature of selection for ID and IB camels, respectively. For both sets, genomic windows showing extremely high ZFst value and XP-EHH scores (top 1% percentile of distribution as threshold) were considered to be potential candidate regions under selection. Based on the results, we found 1525 (765 for ZFst and 760 for XP-EHH) and 1540 (769 for ZFst and 771 for XP-EHH) genomic windows with values higher than the 99th percentile of ZFst (ZFst > 3.41 for ID and ZFst > 3.61 for IB camels) and XPEHH distribution (XP-EHH > 2.24 for ID and XP-EHH > 2.23 for IB camels) as selective sweeps in ID (Fig. 2) and IB camels (Fig. 3), respectively.

Also, to identify selection signatures in AP and CM camels, Fst methods were applied on AP vs ID and CM vs IB sets. Then, genomic windows with scores higher than 99th percentile in ZFst distribution (AP vs ID and CM vs IB) and lower than first percentile in XP-EHH distribution (ID vs AP and IB vs CM) were defined as selection signatures for AP and CM camels. Based on the defined thresholds (ZFst > 3.41, XP-EHH < -2.19 for AP camels and ZFst > 3.61, XP-EHH < -2.17 for CM camels), we found 1534 (765 for ZFst and 769 for XP-EHH) genomic windows for AP (Supplementary Fig. S3) and 1549 (769 for ZFst and 780 for XP-EHH) genomic windows for CM (Supplementary Fig. S4) camels.

Genes and GO analysis. In total, we identified 649 (containing 450 protein-coding genes) genes from Fst and 521 (containing 369 protein-coding genes) genes from XP-EHH as candidate positively selected genes for ID camels. Also, 480 (containing 329 protein-coding genes) and 527 (containing 379 protein-coding genes) candidate genes were found for IB camels from Fst and XP-EHH, respectively. Furthermore, 164 (containing 31

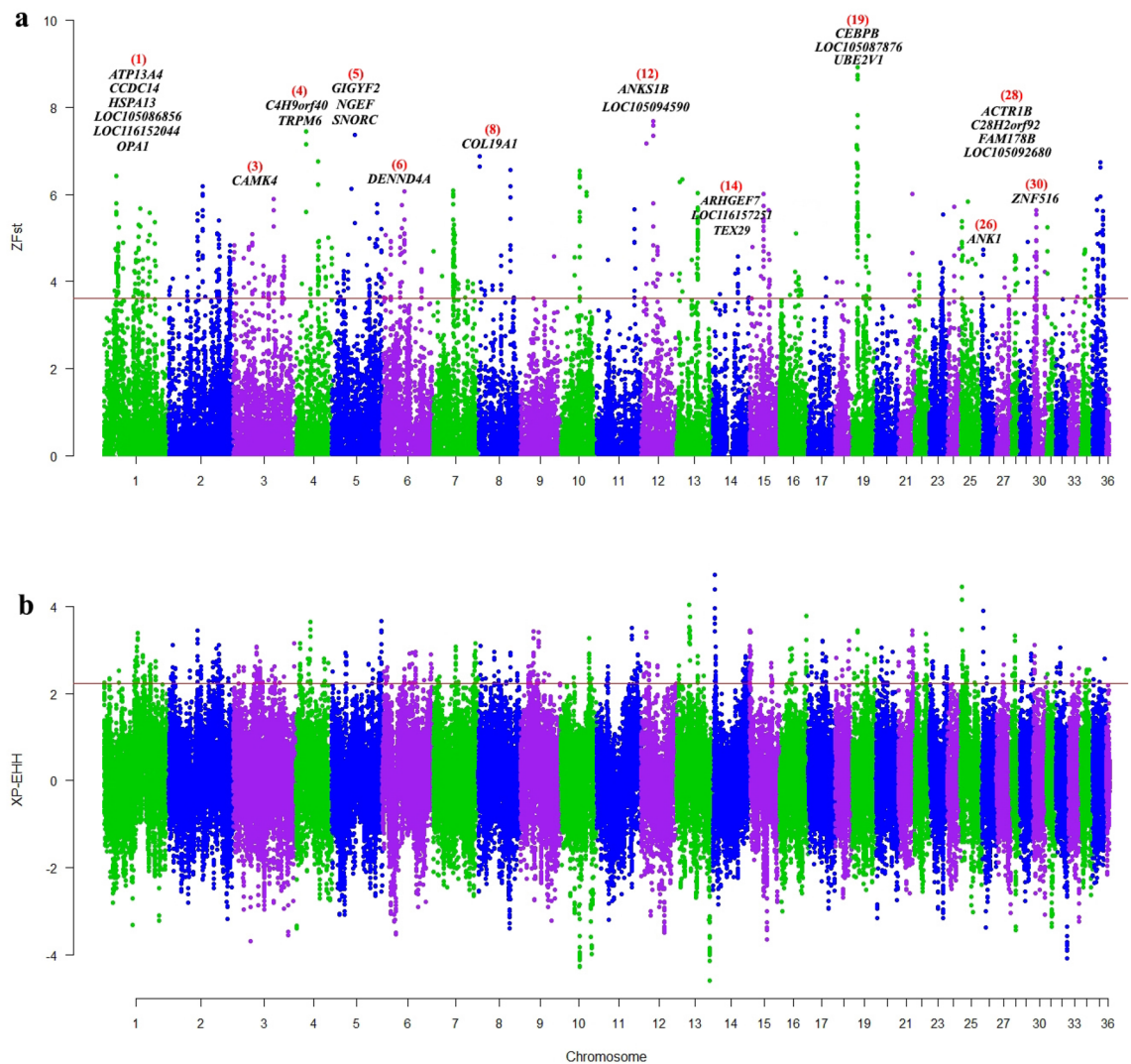


Figure 3. Manhattan plot visualized by R package qqman (<https://cran.r-project.org/web/packages/qqman/>); showing the genome-wide distribution of ZFst (a) and XP-EHH (b) between IB and CM camels. The red line denotes a threshold of ZFst > 3.61 and XP-EHH > 2.23. The gene symbols inserted in the figure were identified as positively selected genes using both methods. The red numbers in parentheses represent the chromosome number.

protein-coding genes) and 104 (containing 28 protein-coding genes) overlapped genomic regions between the Fst and XP-EHH approaches were identified for ID and IB, respectively (Fig. 4). The genes located within those regions were considered as potentially under positive selection in Iranian camels. We performed GO analysis for combined unique protein-coding genes identified through both ZFst and XP-EHH statistics for ID vs AP (787) and IB vs MC (677) camels as well as for common genes identified by both methods for ID (31) (Supplementary Table S1) and IB (28) (Supplementary Table S2) camels. Based on the results, 18 (3 BP and 15 CC) (Supplementary Table S3) and 29 (26 BP and 3 CC) (Supplementary Table S4) GO terms were significantly enriched for combined unique genes for ID and IB camels, respectively. We also performed GO analysis on the genes shared between two used methods for ID and IB; 8 (7 BP and 1 CC) (Supplementary Table S5) and 21 (18 MF and 3 CC) (Supplementary Table S6) GO terms were found to be significantly enriched for ID and IB camels, respectively. Note that no significantly enriched pathways were found for any genes sets in the present study. Additionally, we found 34 (Supplementary Table S7) and 17 (Supplementary Table S8) genes as positively selected genes for AP and CM camels, respectively. The results of the GO analysis on these genes are presented in Supplementary Table S9 for AP and Supplementary Table S10 for CM.

Discussion

As the most intelligent as well as the most ambitious species on earth, humans have tamed a number of animals, dominated them, and made significant changes in their phenotypes and genome architecture. Factors such as geography, climate, culture and even religion influence how humans interact with domestic animals. Thus, various domesticated animal breeds and ecotypes have been created by humans in different parts of the world for specific purposes. Now, the development of technologies such as WGS and high-density SNP arrays along with

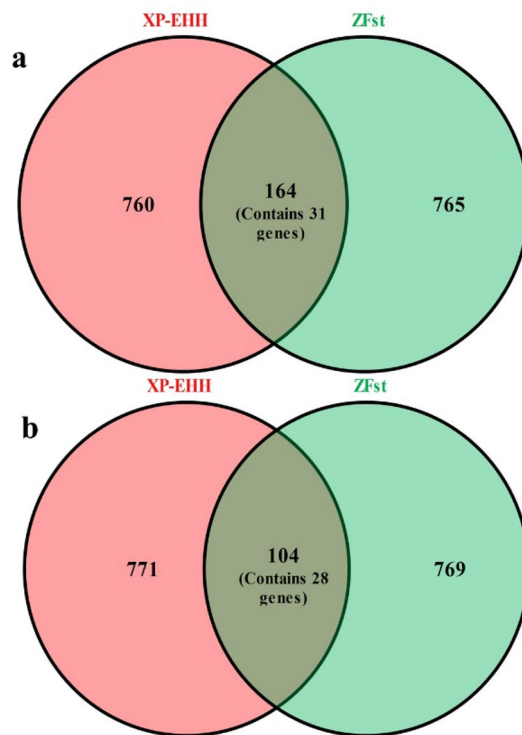


Figure 4. Venn diagram showing the overlapped genomic windows between the top 1% percentile of ZFst and XP-EHH distributions in ID (a) and IB (b) camels.

advances in the field of bioinformatics have enabled to identify footprints of positive selection in the genome of different species¹⁸.

Based on the PCA result, we observed high genetic diversity for IB camels compared with other populations in the PCA plot. This result was consistent with the findings based on mitochondrial sequence analysis which revealed that despite the small number of IB camels, they showed higher genetic diversity¹⁹. Also, based on microsatellite and WGS data, high heterozygosity was calculated for IB camels^{5,20,21}.

Here, we used two inter-population approaches for detecting signatures of selection. Fst is one of the commonly used methods for identifying selective sweeps in domestic animals¹⁸ and considered as one of the effective approaches to detecting positively selected genes for populations without phenotypic data²². Indeed, by focusing on identifying differentiated parts of the genome between population, this method can detect recent signatures of selection¹⁰. According to Sabeti et al.²³, XP-EHH identifies alleles that have been fixed in one population (target population) while being polymorphic in another (reference population) as selective sweeps. Generally, the XP-EHH scores range from negative to positive values, so a positive value suggests selection in the target population, while a negative value shows selection in the reference population.

Generally, the most of significantly enriched GO terms for IB camels were associated with development such as anatomical structure development (GO:0048856), lung alveolus development (GO:0048286) and central nervous system development (GO:0007417), along with immune system such as leukocyte activation (GO:0002366), natural killer cell cytokine production (GO:0002370), and lymphocyte activation (GO:0002285), as well as behavior such as adult behavior (GO:0030534), and adult locomotory behavior (GO:0008344). The only habitat of IB camels in Iran is in the area around Sabalan Mountain (4811 in elevation) located in Ardabil province. The nomads of this region (Shahsevan) as Bactrian camel owners, migrate to higher altitudes of the mountain in the warm season to provide suitable pasture for their sheep herds. In this nomadic lifestyle, most of the lifespan of IB camels is spent at high altitudes (up to 2500 m above sea level), which has probably led to the positive selection of genes involved in lung alveolus development for optimal use of oxygen. Of 26 GO terms in BP categories, six terms were associated with immune system in IB camels. In accordance with our findings for IB camels, several studies on different species^{10,24,25} have shown that immune-related genes have been among the targets of natural and artificial selection. Exposure to diverse pathogens in different regions may be the reason for the selection of different immune-related genes²⁶.

Among the 28 genes located in the overlapped genomic windows between top 1% percentile of ZFst and XP-EHH scores, as the signature of selection for IB camels, the product of the ZNF516 gene as a cold-inducible factor activates UCP1 transcription to promote browning of white fat and the development of brown fat in mice²⁷. Since, brown adipose tissue acts as a thermal regulator and heat generator for the body upon cold exposure^{28,29}, the positive selection of gene involved in the development of this tissue can be vital for Bactrian camels that have adapted to cold climates. The CAMK4 gene is known as a domestication-related gene in sheep³⁰ and has been shown to play an important role in the reproduction of mice³¹, sheep³², and chicken^{33,34}. This gene is also associated with

the formation of long-term memory^{35,36}. One of the interesting genes was the ANK1 which was found to have a link between polymorphism in its promoter region and the meat quality of porcine³⁷ and cattle^{38,39}. Additionally, the HSPA13 gene influences meat quality by affecting the water holding capacity of meat⁴⁰. The results of some studies^{41,42} have indicated an association between the ANKS1B gene and growth plus body size in sheep. COL19A1 and NGEF play roles in muscle differentiation⁴³ and fat deposition⁴⁴, respectively. Evidence suggests that the TRPM6 gene is involved in magnesium homeostasis and absorption in the intestine and kidneys⁴⁵. The results of a study showed that a magnesium-deficient diet causes hypomagnesemia and enhances the expression of the TRPM6 gene in mice kidney and intestine⁴⁶. One of the possible reasons for the selection of this gene in IB camels may be related to the composition of soil and vegetation in the habitat of camels; however, to confirm this claim we need accurate information about these parameters. Previous studies have found that the OPA1 gene, as one of the candidate genes under selection for IB camels, is involved in the adaptation of snow and Tibetan sheep to survive at high altitudes with low oxygen and temperature^{47,48}, that is somewhat similar to the habitat of IB camels. Also, Tang et al.⁴⁹ reported that OPA1 potentially enhances the ability of Tibetan chicken to survive under the hypoxic condition of the Qinghai-Tibet Plateau.

Among the 31 genes located in the overlapped genomic windows between top 1% percentile of ZFst and XP-EHH scores, as the signature of selection for ID camels, the BDH1 plays an important role in energy metabolism through synthesis and degradation of ketone bodies. During starvation, the ketone bodies are considered one of the most important sources of energy for the body⁵⁰. The protein encoded by this gene as one of the ketone body metabolic enzymes, can be a key factor in the efficiency of camels for the optimal use of ketones. Another candidate under selection genes in ID camels was CEP19. Yang et al.⁵¹ found that this gene as a protein synthesis related gene was one of the positively selected genes in fox and probably plays a role in adaptation to strong seasonal fluctuation and energy requirements. Inactivation of this gene could lead to morbid obesity in humans and mice⁵². The CRADD is associated with muscle compactness and is well-known as a meat quality related gene^{53,54}. DLG1, IMMP2L and, FRAS1 were three reproduction-related genes that we found in this study as the putative signature of selection in ID camels. Hernández-Montiel et al.⁵⁵ reported that the DLG1 gene may be associated with the ovary development process and litter size in ewes. The IMMP2L gene plays a role in fertility in mice^{56,57}. Also, it has been found that the FRAS1 gene is one of the candidate genes for cattle⁵⁸ and pig⁵⁹ fertility. EMILIN2 (belonging to elastin protein family) and FBN1 (belong to fibrillin protein family) are involved in fibrogenesis in the yak lung which along with several other genes can improve the efficiency of oxygen utilization in yaks at high altitude⁶⁰. It has been found that GRIA1 is one of the domestication-related genes in the dog⁶¹ and cat⁶². The protein encoded by this gene, known as neurotransmitter receptor, plays an important role in the long-term memory formation⁶³. We hypothesized that the identification of this gene as one of the candidate genes under selection in ID camels is probably due to their rearing system in Iran. ID camels in the central region of Iran are left in the desert every spring to graze⁶⁴, and a memory-related gene (GRIA1) can be linked to finding pre-identified food sources as well as returning home. The SLC12A1 gene is one of the important kidney-related genes that we found as one of the positively selected genes in ID camels. Mara et al.⁶⁵ reported that this gene is a key determinant of physiological water conservation in desert rodents, due to its role in the reabsorption of sodium and subsequent water reabsorption in the kidney. Several genes including SOCS2, JAK1, NRROS, and SENP1 were immune-related genes^{66–69}. However, evidence suggests that the SOCS2 gene is potentially associated with milk production traits⁷⁰, growth, body weight⁷¹, as well as efficient energy metabolism⁷².

The GO analysis to functionally categorize the positively selected genes for CM camels led to the identification of three GO terms related to embryo development including embryo implantation (GO:0007566), female pregnancy (GO:0007565) and positive regulation of endodermal cell differentiation (GO:1903226). The obtained results showed differences in the selection pattern in IB and CM camels. This difference in selection pattern can be related to various factors such as geography, climate and also the culture of camel breeders in Iran and China-Mongolia. Investigation of selection signatures in AP camels revealed that three olfactory-related genes (LOC105087657, LOC105105401 and LOC105105413) are on the list of identified genes. Presumably, the role of olfactory receptors in locating food and mate, avoiding threats (e.g. predators and toxins) and feeding behaviors (e.g. food choice)⁷³ has made these genes a candidate to help AP camels adapt to harsh desert environments.

Conclusion

In the present study, we used WGS data of dromedary (from Iran and AP) and Bactrian camels (from Iran and MC) to detect the genome-wide signature of selection in the ID and IB camels. Using two methods, we identified genomic regions under positive selection that harbor some considerable genes in ID and IB camels. Despite the interesting results, the small sample size can be considered as a limitation in the present study. Therefore, further studies with a higher number of samples along with other levels of omics data (e.g. RNA-Seq data) can help to better understand the genetic basis of production-related phenotypes of camels. It can be mentioned that camels have not been subjected to intense artificial selection, unlike domestic animals such as cattle, sheep, horse, chicken and pig. Therefore, knowledge about genomic features of camels can be an important step in designing and implementing effective breeding programs for camels in the future. Indeed, understanding the mechanism of selection, genetic background of adaptation processes to local environment as well as identification of candidate genes of interest to breeding programs can help us to draw a roadmap for camel breeding in the near future.

Materials and methods

Samples and whole genome resequencing. WGS data of 34 Old World camels including 14 dromedaries and 20 Bactrian camels were used in the present study. Of these samples, 6 and 10 samples belonged to ID and IB camels, respectively. ID samples were from Semnan (latitude: 35.429288 and longitude: 55.007100), Yazd (latitude: 31.610649 and longitude: 55.447436) and Ardabil (latitude: 38.249364 and longitude: 48.297162)

provinces, while samples of IB were from Ardabil province as the only habitat of Bactrian camels in Iran. Additionally, to use as reference population against ID and IB camels to identify signature of selection, WGS data of 8 dromedary camels belonging to AP and 10 Bactrian camels belonging to CM were obtained from Sequence Read Archive at the NCBI (Supplementary Table S11).

For sequencing, after shearing of genomic DNA, fragments with 300–500 bp were generated after which these fragments were repaired, 'A'-tailed, and ligated to Illumina sequencing adapters. DNA fragments with 370–470 bp were selected on 2% agarose gels and after amplification using PCR, were used for sequencing. Finally, one lane of 150 bp paired-end sequencing was performed by Illumina HiSeq 2500 system.

Raw reads filtration. Before performing any analysis, the quality of the sequenced reads was assessed by FastQC v.0.11.9 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). We used Trimmomatic v.0.39⁷⁴ to apply quality filtration on raw reads using the sliding windows (5:20) approach. The employed trimming options were not the same for all samples as HEADCROP and ILLUMINACLIP were applied to some samples for removing base content noises and adapter contamination, respectively. Finally, reads with less than 40 bp were excluded from downstream analysis.

Alignment and variant calling. Reads that passed the quality filtration step, were mapped to the dromedary camel reference genome (GCF_000803125.2) using BWA v.0.7.17⁷⁵ with the “mem” algorithm. The outputted SAM file was converted to BAM format using Samtools where sorting of BAM file in coordinate order was done by Picard “SortSam” (<http://broadinstitute.github.io/picard>). Mark duplication was performed using Picard to remove duplicated reads during sample preparation and sequencing steps. Two approaches of Genome Analysis Toolkit (GATK) v.3.7⁷⁶ including “RealignerTargetCreator” and “IndelRealigner” were performed for local realignment around INDELS.

In the present study, to identify single nucleotide polymorphisms (SNPs), GATK “HaplotypeCaller” algorithm with “-ERC GVCF” mode was used to create “gVCF” for each sample separately. Then, we used “GenotypeGVCFs” approach of GATK to jointly genotype all gVCFs for creating a VCF file containing raw called variants (SNPs and INDELS) for all samples in each group (dromedary and Bactrian camels). Called SNPs were extracted from the VCF file by GATK “SelectVariants” for further analysis. The identified SNPs were filtered using BCFtools, VCFtools⁷⁷, and GATK with the following criteria: SNPs with $QD < 2$, $FS > 60$, $MQ < 40$, $MQRankSum < -12.5$, $ReadPosRankSum < -8$ and $SOR > 3$ were removed. We also filtered those SNPs with a cluster windows size of 10 bp and a cluster size of 3 SNPs. Finally, SNPs with only two alleles, sites with minor allele frequency greater than or equal to 0.05 and loci with genotypes called in at least 80% of individuals were retained for downstream analysis. We then excluded chromosome X and unplaced contigs from the final VCF file. Annotation of final SNPs was carried out using SnpEff v.5.0e⁷⁸.

Population genetic analysis. The genetic relationship between dromedary and Bactrian camels was investigated using principal component analysis (PCA) by PLINK1.9⁷⁹. Before performing PCA, we pruned identified SNPs using the “indep-pairwise 50 10 0.2” option in PLINK1.9 to minimize the effects of SNPs contributed by regions of extensive strong linkage disequilibrium (LD). This would remove one of the pairs of SNPs with a square correlation greater than 0.2 in windows with 50 and a step size of 10 SNPs.

Detection of selection signatures. To identify regions under positive selection in the genome of ID and IB camels, we used two methods including population differentiation index (F_{st}) and cross-population extended haplotype homozygosity (XP-EHH) with 50 kb sliding window and 25 kb step size. The window size and step size were selected for this study based on previous studies on the genomes of dromedary¹⁶ and Bactrian¹⁷ camels. The F_{st} in each window and XP-EHH score for each SNP were estimated using VCFtools and Selscan v.2.0⁸⁰, respectively. In the present study, Beagle v.5.2⁸¹ was performed with argument burnin = 5 and iterations = 20 for haplotype phasing. For dromedaries, ID camels constituted the target population while dromedaries from AP formed the reference population. In the second comparative set, IB camels were regarded as the target population compared to CM Bactrian camels as the reference population.

Normalization of each XP-EHH score across all chromosomes was performed using norm program distributed along with Selscan. We transformed F_{st} values to Z F_{st} : $ZF_{st} = (F_{st} - \mu F_{st}) / \sigma F_{st}$, where μF_{st} and σF_{st} are the mean and standard deviation of F_{st} values in all windows. Note that windows containing ≤ 5 SNPs were removed to reduce the probability of identifying false positive selected regions. We estimated the average normalized XP-EHH score in each window for both comparative sets (ID vs AP and IB vs CM). The overlapped genomic windows in the top 1% percentile of Z F_{st} values and normalized XP-EHH scores were considered as the signature of selection in the ID and IB genomes. For a comparative view, a similar process was performed for the other two sets, including AP vs ID and CM vs IB.

Annotation and gene ontology. The putative selective sweeps identified by F_{st} and XP-EHH methods were annotated using GTF (gene transfer format) file related to dromedary camel genome and BEDtools v.2.27.1⁸². In each comparative set, for gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, we submitted the combined list of the identified genes by F_{st} and XP-EHH approaches, as well as a list of genes shared between the two mentioned approaches to “g:Profiler”. GO analysis allows us to better understand and express the biological function of genes and classifies them into three categories including biological process (BP), molecular function (MF), and cellular component (CC). The calculated p-values were corrected using the Benjamini–Hochberg FDR (false positive rate) and enriched terms were considered statistically significant at p-adjusted ≤ 0.05 .

Data availability

Three new whole genome sequence data related to Iranian dromedary (one sample) and Bactrian (two samples) camels analyzed in this study are available in the Sequence Read Archive (SRA) with the BioProject accession number PRJNA801784 (including SRR17817908, SRR17817909 and SRR17817910).

Received: 27 January 2022; Accepted: 6 June 2022

Published online: 10 June 2022

References

1. Salamini, F., Özkan, H., Brandolini, A., Schäfer-Pregl, R. & Martin, W. Genetics and geography of wild cereal domestication in the near east. *Nat. Rev. Genet.* **3**, 429–441 (2002).
2. Perri, A. R. *et al.* Dog domestication and the dual dispersal of people and dogs into the Americas. *PNAS* **118**, e2010083118 (2021).
3. Ahmad, H. I. *et al.* The domestication makeup: Evolution, survival, and challenges. *Front. Ecol. Evol.* **8**, 103 (2020).
4. Almathen, F. *et al.* Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. *PNAS* **113**, 6707–6712 (2016).
5. Ming, L. *et al.* Whole-genome sequencing of 128 camels across Asia reveals origin and migration of domestic Bactrian camels. *Commun. Biol.* **3**, 1–9 (2020).
6. Maynard, J. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **89**, 391–403 (2007).
7. Rojas-Downing, M. M., Nejadhashemi, A. P., Harrigan, T. & Woznicki, S. A. Climate change and livestock: Impacts, adaptation, and mitigation. *Clim. Risk Manag.* **16**, 145–163 (2017).
8. Nanaei, H. A., Qanatqestani, M. D. & Esmailzadeh, A. Whole-genome resequencing reveals selection signatures associated with milk production traits in African Kenana dairy zebu cattle. *Genomics* **112**, 880–885 (2020).
9. Peripolli, E. *et al.* Genome-wide detection of signatures of selection in indicine and Brazilian locally adapted taurine cattle breeds using whole-genome re-sequencing data. *BMC Genomics* **21**, 1–16 (2020).
10. Ruiz-Larrañaga, O. *et al.* Genomic selection signatures in sheep from the Western Pyrenees. *Genet. Sel. Evol.* **50**, 1–12 (2018).
11. Wang, W. *et al.* Deep genome resequencing reveals artificial and natural selection for visual deterioration, plateau adaptability and high prolificacy in Chinese domestic sheep. *Front. Genet.* **10**, 300 (2019).
12. Guo, J. *et al.* Whole-genome sequencing reveals selection signatures associated with important traits in six goat breeds. *Sci. Rep.* **8**, 1–11 (2018).
13. Berihulay, H. *et al.* Whole genome resequencing reveals selection signatures associated with important traits in Ethiopian indigenous goat populations. *Front. Genet.* **10**, 1190 (2019).
14. Ablondi, M., Viklund, Å., Lindgren, G., Eriksson, S. & Mikko, S. Signatures of selection in the genome of Swedish warmblood horses selected for sport performance. *BMC Genomics* **20**, 1–12 (2019).
15. Salek Ardestani, S. *et al.* Signatures of selection analysis using whole-genome sequence data reveals novel candidate genes for pony and light horse types. *Genome* **63**, 387–396 (2020).
16. Bahbahani, H. *et al.* Genome diversity and signatures of selection for production and performance traits in dromedary camels. *Front. Genet.* **10**, 893 (2019).
17. Liu, C., Chen, H., Yang, X., Zhang, C. & Ren, Z. Exploring the genomic resources of seven domestic Bactrian camel populations in China through restriction site-associated DNA sequencing. *PLoS ONE* **16**, e0250168 (2021).
18. Saravanan, K. A. *et al.* Selection signatures in livestock genome: A review of concepts, approaches and applications. *Livest. Sci.* **241**, 104257–104268 (2020).
19. Hedayat-Evrigh, N., Khalkhali-Evrigh, R., Vahedi, V., Pourasad, K. & Sharifi, R. S. Genetic diversity and population structure of old world camelids (*Camelus dromedaries* and *Camelus bactrianus*) in Iran using mitochondrial DNA. *Kafkas Üniv. Vet. Fak. Derg.* **24**, 467–471 (2018).
20. Hedayat-Evrigh, N., Miraei-Ashtiani, S. R., Khalkhali-Evrigh, R. & Pourasad, K. Molecular assessment of genetic diversity in dromedaries and Bactrian camel using microsatellite markers. *J. Agric. Sci. Technol.* **20**, 1137–1148 (2018).
21. Hedayat-Evrigh, N., Khalkhali-Evrigh, R. & Bakhtiarzadeh, M. R. Genome-wide identification and analysis of variants in domestic and wild bactrian camels using whole-genome sequencing data. *Int. J. Genomics* **2020**, 1–9 (2020).
22. Narum, S. R. & Hess, J. E. Comparison of FST outlier tests for SNP loci under selection. *Mol. Ecol. Resour.* **11**, 184–194 (2011).
23. Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
24. Li, M. *et al.* Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat. Genet.* **45**, 1431–1438 (2013).
25. Weng, Z. *et al.* Genomic variations and signatures of selection in Wuhua yellow chicken. *PLoS ONE* **15**, e0241137 (2020).
26. Kjærner-Semb, E. *et al.* Atlantic salmon populations reveal adaptive divergence of immune related genes—a duplicated genome under selection. *BMC Genomics* **17**, 1–12 (2016).
27. Dempersmier, J. *et al.* Cold-inducible Zfp516 activates UCP1 transcription to promote browning of white fat and development of brown fat. *Mol. Cell.* **57**, 235–246 (2015).
28. Chondronikola, M. *et al.* Brown adipose tissue is linked to a distinct thermoregulatory response to mild cold in people. *Front. Physiol.* **7**, 129 (2016).
29. Lundh, M. *et al.* Cold-induction of afadin in brown fat supports its thermogenic capacity. *Sci. Rep.* **11**, 1–11 (2021).
30. Chen, Z. H. *et al.* Whole-genome sequence analysis unveils different origins of European and Asiatic mouflon and domestication-related genes in sheep. *Commun Biol.* **4**, 1–15 (2021).
31. Wu, J. Y., Gonzalez-Robayna, I. J., Richards, J. S. & Means, A. R. Female fertility is reduced in mice lacking Ca²⁺/calmodulin-dependent protein kinase IV. *Endocrinology* **141**, 4777–4783 (2000).
32. Li, X. *et al.* Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nat. Commun.* **11**, 1–16 (2020).
33. Nguyen, T. M. D., Combarrous, Y., Praud, C., Duittoz, A. & Blesbois, E. Ca²⁺/calmodulin-dependent protein kinase kinases (CaMKKs) effects on AMP-activated protein kinase (AMPK) regulation of chicken sperm functions. *PLoS ONE* **11**, e0147559 (2016).
34. Zhao, X. *et al.* Identification of candidate genomic regions for chicken egg production traits based on genome wide association study. *BMC Genomics* **22**, 610 (2021).
35. Tonegawa, S., Nakazawa, K. & Wilson, M. A. Genetic neuroscience of mammalian learning and memory. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358**, 787–795 (2003).
36. Fukushima, H. *et al.* Upregulation of calcium/calmodulin-dependent protein kinase IV improves memory formation and rescues memory loss with aging. *J. Neurosci.* **28**, 9910–9919 (2008).
37. Aslan, O. *et al.* Association between promoter polymorphisms in a key cytoskeletal gene (Ankyrin 1) and intramuscular fat and water-holding capacity in porcine muscle. *Mol. Biol. Rep.* **39**, 3903–3914 (2012).

38. Aslan, O., Sweeney, T., Mullen, A. M. & Hamill, R. M. Regulatory polymorphisms in the bovine Ankyrin 1 gene promoter are associated with tenderness and intramuscular fat content. *BMC Genet.* **11**, 1–14 (2010).
39. Shin, Y. *et al.* Discovery of gene sources for economic traits in Hanwoo by whole-genome resequencing. *Asian Austral J. Anim. Sci.* **29**, 1353 (2016).
40. Du, L. *et al.* Transcriptome profiling analysis of muscle tissue reveals potential candidate genes affecting water holding capacity in Chinese Simmental beef cattle. *Sci. Rep.* **11**, 1–14 (2021).
41. Zhang, L. *et al.* Genome-wide genetic diversity and differentially selected regions among Suffolk, Rambouillet, Columbia, Polypay, and Targhee sheep. *PLoS ONE* **8**, e65942 (2013).
42. Kominakis, A. *et al.* Combined GWAS and ‘guilt by association’-based prioritization analysis identifies functional candidate genes for body size in sheep. *Genet. Sel. Evol.* **49**, 1–16 (2017).
43. Sumiyoshi, H. *et al.* Esophageal muscle physiology and morphogenesis require assembly of a collagen XIX-rich basement membrane zone. *J. Cell. Biol.* **166**, 591–600 (2004).
44. Wang, L., Li, X., Ma, J., Zhang, Y. & Zhang, H. Integrating genome and transcriptome profiling for elucidating the mechanism of muscle growth and lipid deposition in Pekin ducks. *Sci. Rep.* **7**, 1–11 (2017).
45. Schlingmann, K. P. *et al.* Hypomagnesemia with secondary hypocalcemia is caused by mutations in TRPM6, a new member of the TRPM gene family. *Nat. Genet.* **31**, 166–170 (2002).
46. Rondón, L. J., Groenestege, W. M. T., Rayssiguier, Y. & Mazur, A. Relationship between low magnesium status and TRPM6 expression in the kidney and large intestine. *Am. J. Physiol.* **294**, R2001–R2007 (2008).
47. Wang, G., He, Y. & Luo, Y. Expression of OPA1 and Mic60 genes and their association with mitochondrial cristae morphology in Tibetan sheep. *Cell Tissue Res.* **376**, 273–279 (2019).
48. Upadhyay, M. *et al.* Whole genome sequencing reveals a complex introgression history and the basis of adaptation to subarctic climate in wild sheep. *Mol. Ecol.* **30**, 6701–6717 (2021).
49. Tang, Q. *et al.* Mitochondrial Fusion potentially regulates a metabolic change in Tibetan chicken embryonic brain during hypoxia. *Front. Cell Dev. Biol.* **9**, 166 (2021).
50. Iizuka, K. *Dietary Interventions in Liver Disease* 263–274 (Academic Press, 2019).
51. Yang, X. *et al.* Comparative transcriptome provides insights into the selection adaptation between wild and farmed foxes. *Ecol. Evol.* **11**, 13475–13486 (2021).
52. Shalata, A. *et al.* Morbid obesity resulting from inactivation of the ciliary protein CEP19 in humans and mice. *Am. J. Hum. Genet.* **93**, 1061–1071 (2013).
53. Raschetti, M., Rignanesi, D., Pagnacco, G., Castiglioni, B. & Chessa, S. Polymorphisms in swine candidate genes for meat quality detected by PCR-SSCP. *Ital. J. Anim. Sci.* **8**, 129–131 (2009).
54. Raschetti, M. *et al.* SNP identification in swine candidate genes for meat quality. *Livest. Sci.* **155**, 165–171 (2013).
55. Hernández-Montiel, W. *et al.* Genome-wide association study reveals candidate genes for litter size traits in pelibuey sheep. *Animals* **10**, 434 (2020).
56. Lu, B. *et al.* A mutation in the inner mitochondrial membrane peptidase 2-like gene (*Immp2l*) affects mitochondrial function and impairs fertility in mice. *Biol. Reprod.* **78**, 601–610 (2008).
57. He, Q. *et al.* The *Immp2l* mutation causes ovarian aging through ROS-Wnt/ β -catenin-estrogen pathway: preventive effect of melatonin. *Endocrinology* **161**, 119 (2020).
58. Cai, Z., Guldbrandsen, B., Lund, M. S. & Sahana, G. Prioritizing candidate genes for fertility in dairy cows using gene-based analysis, functional annotation and differential gene expression. *BMC Genomics* **20**, 1–9 (2019).
59. Fischer, D., Laiho, A., Gyenesei, A. & Sironen, A. Identification of reproduction-related gene polymorphisms using whole transcriptome sequencing in the Large White pig population. *G3: Genes Genomes Genet.* **5**, 1351–1360 (2015).
60. Li, J. *et al.* Changes in the expression levels of elastic fibres in yak lungs at different growth stages. *BMC Dev. Biol.* **21**, 1–11 (2021).
61. Wang, G. D. *et al.* Out of southern East Asia: The natural history of domestic dogs across the world. *Cell Res.* **26**, 21–33 (2016).
62. Montague, M. J. *et al.* Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *PNAS* **111**, 17230–17235 (2014).
63. Mead, A. N. & Stephens, D. N. Selective disruption of stimulus-reward learning in glutamate receptor *gria1* knock-out mice. *J. Neurosci.* **23**, 1041–1048 (2003).
64. Khalkhali-Evrigh, R., Hafezian, S. H., Hedayat-Evrigh, N., Farhadi, A. & Bakhtiarzadeh, M. R. Genetic variants analysis of three dromedary camels using whole genome sequencing data. *PLoS ONE* **13**, e0204028 (2018).
65. Marra, N. J., Eo, S. H., Hale, M. C., Waser, P. M. & DeWoody, J. A. A priori and a posteriori approaches for finding genes of evolutionary interest in non-model species: osmoregulatory genes in the kidney transcriptome of the desert rodent *Dipodomys spectabilis* (banner-tailed kangaroo rat). *Comp. Biochem. Physiol. Part D Genomics Proteomics* **7**, 328–339 (2012).
66. Starr, R. *et al.* A family of cytokine-inducible inhibitors of signalling. *Nature* **387**, 917–921 (1997).
67. Killick, K. E. *et al.* Genome-wide transcriptional profiling of peripheral blood leukocytes from cattle infected with *Mycobacterium bovis* reveals suppression of host immune genes. *BMC Genomics* **12**, 1–18 (2011).
68. Noubade, R. *et al.* NRROS negatively regulates reactive oxygen species during host defence and autoimmunity. *Nature* **509**, 235–239 (2014).
69. Zheng, C., Li, D., Zhan, W., He, K. & Yang, H. Downregulation of SENP1 suppresses LPS-induced macrophage inflammation by elevating Sp3 SUMOylation and disturbing Sp3-NF- κ B interaction. *Am. J. Transl. Res.* **12**, 7439 (2020).
70. Arun, S. J. *et al.* Targeted analysis reveals an important role of JAK-STAT-SOCS genes for milk production traits in Australian dairy cattle. *Front. Genet.* **6**, 342 (2015).
71. Rupp, R. *et al.* A point mutation in suppressor of cytokine signalling 2 (*Socs2*) increases the susceptibility to inflammation of the mammary gland while associated with higher body weight and size and higher milk production in a sheep model. *PLoS Genet.* **11**, e1005629 (2015).
72. Yang, J. I. *et al.* Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Mol. Biol. Evol.* **33**, 2576–2592 (2016).
73. Connor, E. E., Zhou, Y. & Liu, G. E. The essence of appetite: does olfactory receptor variation play a role? *J. Anim. Sci.* **96**, 1551–1558 (2018).
74. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
75. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
76. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
77. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
78. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
79. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
80. Szpiech, Z. A. & Hernandez, R. D. selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* **31**, 2824–2827 (2014).

81. Browning, B. L., Tian, X., Zhou, Y. & Browning, S. R. Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* **108**, 1880–1890 (2021).
82. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

Acknowledgements

We would to thank the Iran National Science Foundation (INSF) for supporting our project under number 99023297.

Author contributions

R.K-E. and N.H. designed the project. N.H. collected the Iranian camel's blood samples. R.K-E. performed analysis on WGS data and interpreted the results. R.K-E. and N.H. wrote the manuscript. L.M. and J. performed the sequencing process of genomes and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-14376-7>.

Correspondence and requests for materials should be addressed to N.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022