

Genomic signatures of divergent selection for production and adaptation in domestic goats (*Capra hircus*)

Joram M Mwacharo (✉ j.mwacharo@cgiar.org)

International Center for Agricultural Research in the Dry Areas

Eui-Soo Kim

Recombinetics

Ahmed R Elbeltagy

Animal Production Research Institute ARC Egypt

Adel M Aboul-Naga

Animal Production Research Institute, ARC Egypt

Barbara A Rischkowsky

International Center for Agricultural Research in the Dry Areas

Max F Rothschild

Iowa State University

Research article

Keywords: Arid environments, Climate change, Extreme environments, SNP genotypes, Selection signatures

Posted Date: October 11th, 2019

DOI: <https://doi.org/10.21203/rs.2.15925/v1>

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

[Read Full License](#)

Abstract

Background : Goats are distributed worldwide and include many breeds with marked phenotypic variation. Both natural and human driven processes have shaped their genomes. Here, 52K genome-wide SNP genotypes were used to investigate signatures of divergent selection between indigenous goats from an arid hot environment in Egypt and breeds developed under temperate environments in Europe. Three selection signature approaches were used, the di derivative of F_{ST}, iHS and Rsb. Results: Out of a total of 36 candidate regions that were detected to be under selection by the three approaches, the analysis detected nine regions with the strongest selection signatures spanning 134 genes of which 68 were the most significantly functionally enriched. In addition to well-known genes affecting dairy traits (LAP3, ABCG2, MEPE, IBSP, MED28) and body weight and stature (LCORL, NCAPG, CCSER1, DCAF16) in several mammalian species, we found evidence for selection in regions spanning genes implicated in the regulation of fatty acid synthesis and metabolic pathways (PI4K2B and PPARGC1A), inflammatory gene expression and autoimmune response pathways (CXCL8, HERC5, RGS18, TROVE2) and spermatogenesis (SPATA18, LAP3). Other genes not reported before in goats included SLIT2, PACRGL, GRP125, DHX15, SOD3 and KCNIP4 which have been associated with thermal nociception in mice. We also detected a paralog of TBC1D12, namely TBC1D14, in one of the candidate regions. TBC1D12 has been linked to environmental adaptation in sheep. Gene ontology analysis revealed the 68 candidate genes were highly enriched in two biological processes viz regulation of G protein-coupled receptor signaling pathway (GO:0045744) and cellular response to stress (GO:0033554). Conclusions: Our results provide evidence that goat genomes have evolved, in part, via diverse positive selection directed at breed formation, at past and recent genetic improvement and adaptation to environments. Based on their uniqueness, these breed-group specific signatures of selection can be considered to be footprints of divergence which may be useful in characterizing genome architecture and diversity in domestic animals and for targeted genetic improvement.

Background

Humans and the environment have played a fundamental role in the evolutionary history of domestic livestock [1]. Genetic changes in the caprine genome driven by the quest to adapt to new circumstances dates to about 11,000 years ago [2]. Domestication and subsequent human-mediated dispersion across the globe, and socio-cultural and economic activities, exposed domestic goat genomes to new and challenging natural and/or artificial environments and husbandry practices. For some of the exposures, the species evolved to adapt to the novel environments but in other instances, the resulting outcomes were not favourable. The favourable adaptations saw the species being established in a wide range of habitats within the tropics, sub-tropics and temperate environments, where locally adapted populations and breeds were differentiated by various factors including selection, genetic drift and reproductive isolation. More recently, artificial selection for various unique and aesthetic morphological (coat colours, presence of horns etc.) and economically important (growth rates, milk production etc.) traits geared

towards the development of specialized and distinct breeds must have left signatures of divergent selection in their genome.

The introduction of goats to Africa may have commenced with the movement of herders from the Near East into North Africa (Egypt, Libya, Algeria) around 7,000 years ago [3]. Egypt has approximately 4.4 million goats [4], that are predominantly of indigenous types. The hot climate and arid environment in Egypt present a major positive selection pressure and livestock keepers prefer indigenous stocks over their exotic counterparts, due to their better adaptive survival and low husbandry requirements. The combined effects of natural selection, traditional management, and ancient and contemporary intermixing may have resulted in a uniquely adapted indigenous goat gene pool which can be used to examine the phenotypic and genetic consequences of adaptive evolution.

The detection of genomic regions that have been targeted by recent positive selection has proved a powerful tool for delineating genes contributing to adaptation to environmental variables and for informing functions accounting for phenotypic diversity. Over the last decade, many genome-wide scans for selection have been reported in domestic animals, fuelled by the advent of single nucleotide polymorphism (SNP) data sets and whole-genome sequences. These studies have made use of various statistical methods based on the predictable effects of positive selection on patterns of genetic variation. These effects include a decrease in haplotype diversity, high fraction of rare alleles, or major shifts of allele frequency between populations [see reviews by 5, 6]. These approaches have led to the identification of several hundred genomic regions displaying selection footprints, suggesting the presence in these regions of new beneficial mutations that have spread rapidly through the population.

Several candidate selection signatures associated with adaptation to arid environments were reported in the genomes of Barki goats and sheep from Egypt [7]. The authors however, analysed only a small number of individuals (68 goats and 59 sheep) with a restricted geographic range. Furthermore, selection signatures associated with meat and milk production traits have been reported mostly for breeds of sheep and cattle in the developed countries [8, 9, 10, 11]. To advance our knowledge and understanding of the genome dynamics of indigenous goats *vis a vis* well-defined commercial breeds, we analysed 52K SNP genotype data of breeds of goats from a sub-tropical arid environment in Egypt versus temperate European environment, and exposed to contrasting selection pressures, natural versus artificial. The comparative analysis of the goat breeds from Africa and Europe, identified 36 candidate regions spanning 258 genes under positive selection. Sixty-eight functionally enriched genes were found across nine candidate regions that had the strongest signatures of selection. We considered the 68 genes to be the likely primary candidate genes constituting the first line of adaptive evolutionary divergence amongst the different goat breeds.

Results

Genome-wide patterns of signatures of selection. We investigated candidate regions under selection using allele frequencies estimated within (*iHS*) and between (*di* and *Rsb*) Egyptian and European breed-groups of goats (Supplementary Table S1). The *di* identified four candidate regions showing significant differentiation between the two breed-groups on four chromosomes (CHI5, CHI6, CHI12 and CHI17; Figure 1a; Supplementary Table S2). CHI6 and CHI12 had the strongest signatures of differential selection spanning 5 and 20 significant SNPs, respectively. The *Rsb* revealed 19 candidate regions spread across 12 chromosomes (Figure 1b; Supplementary Table S2). The strongest and most significant regions occurred on CHI3, CHI6 and CHI16 and spanned 15, 11 and 81 significant SNPs, respectively. Thirteen candidate regions found across eight chromosomes were revealed by *iHS* (Figure 1c; Supplementary Table S2). The strongest of these regions were found on CHI6 and CHI12 and were defined by 19 and 18 significant SNPs, respectively.

Overlap between candidate regions. All the three approaches identified candidate regions under selection on CHI6 and CHI12 (Supplementary Table S2). On CHI6, one region was identified by *di* (68,327,191-68,495,816 bp), four by *Rsb* (75,072,253-76,237,641 bp; 78,852,104-79,060,435 bp; 82,782,072-82,791,242 bp; 89,340,232-89,433,167 bp) and two by *iHS* (30,246,187-47,044,407 bp; 82,852,913-82,907,013 bp). One of the regions identified by *Rsb* (82,782,072-82,791,242 bp) overlapped with one of the two *iHS* regions (82,852,913-82,907,013 bp). On CHI12, one region was revealed by *di* (34,478,328-35,402,998 bp), and two each by *Rsb* (6,942,093-7,015,304 bp; 34,091,464-34,143,365 bp) and *iHS* (34,295,846-36,783,669 bp; 41,241,216-41,433,623 bp) (Supplementary Table S2), respectively. The candidate region identified by *di*, overlapped with one of the two *iHS* regions (34,295,846-36,783,669 bp). In addition to the regions identified on CHI6 and CHI12 by *iHS* and *Rsb*, respectively, the two tests also identified candidate regions on CHI11 and CHI16 (Supplementary Table S2). On CHI11, the single candidate region (68,478,589-69,137,181 bp) identified by *Rsb* overlapped with the one revealed by *iHS* (68,478,589-69,208,064 bp). On CHI16, none of the two regions identified by *Rsb* (6,511,375-13,049,713 bp; 46,186,034-46,690,120 bp) overlapped with the one identified with *iHS* (53,386,905-53,459,228 bp).

Gene content and functional annotation of the strongest candidate regions. There were five (*Rsb* = 3, *iHS* = 2) candidate regions that spanned no annotated genes (Supplementary Table S2). Such regions have been designated as “gene deserts” and have also been reported by other investigators [12, 13, 14, 15]. Genome-wide, we identified 258 genes (*di* = 13, *Rsb* = 115, *iHS* = 130) mapping to 36 (*di* = 4, *Rsb* = 19, *iHS* = 13) candidate selection regions that were defined by 287 significant SNPs (Supplementary Table S2). Amongst the 36 candidate regions, there were seven that were the strongest (*di* = 2, *Rsb* = 3, *iHS* = 2) and spanned 169 significant SNPs (Table 1). In total, 134 genes mapped to the nine strongest candidate regions (Table 1). We regard the positive selection acting on these nine strongest candidate regions to be the primary ones driving the evolutionary divergence between the two groups of goats analysed.

The two strongest signals of selection revealed by *di* were respectively, on CHI6 and CHI12. The region on CHI6 spanned four genes and the one on CHI12 extended across two genes (Table 1). The strongest signals revealed by *Rsb* were on CHI3, CHI6 and CHI16, respectively. The region on CHI3 spanned four genes, the one on CHI6 five genes and the one on CHI16, 29 genes (Table 1). Chromosomes CHI6 and CHI12 also had the strongest *iHS* candidate regions. The one on CHI6 spanned 80 genes, while 10 genes occurred in the region on CHI12 (Table 1).

Enrichment analysis was performed with the 134 genes present in the nine strongest candidate regions. We limited the functional annotation to *B. taurus* and used it as the background species because, in both cases, using *C. hircus* returned no results, most likely due to either the incomplete annotation of its genome or its exclusion in the DAVID database. Based on the bovine RefSeq gene annotation, of the 134 genes, 68 were significantly enriched. From DAVID analysis, we found seven functional annotation clusters that were significantly enriched (Supplementary Table S3); the top three were regulator of G protein-coupled receptor (GPCR) protein signaling pathway (GO:0008277; Enrichment score = 4.14), protein ubiquitination involved in ubiquitin-dependent protein catabolic process (GO:0042787; Enrichment score = 2.02) and Leucine-rich repeats (LRR) typical subfamily (IPR003591; Enrichment score = 1.42).

The functions of the 68 genes were further investigated using the text-mining tool in STRING. We detected several genes that have been reported before to be strongly linked to positive selection in other domestic livestock. These genes included *LAP3*, *IBSP*, *MED28*, *MEPE*, *ABCG2*, *CCSER1*, *SPP1*, that have been associated with milk production related traits [8, 12, 16, 17, 18, 19, 20] and *PLAG1*, *NCAPG*, *IBSP*, *DCAF16*, *CCSER1*, *FAM184B*, *CCKAR* and *LCORL* which are body size-related genes [12, 21, 22] in mammals. There were also genes associated with male fertility and spermatogenesis (*ANAPC5*, *GPR125*, *SMARCAD1*, *SPATA18*) and prolificacy (*BMPR1B*). Other genes were involved in fatty acid synthesis (adipogenesis) and metabolic pathways (*PI4K2B*, *PPARGC1A*, *RGS1*, *RGS2*, *RGS13*), and with pro-inflammation and immune response (*SPP1*, *CXCL8/IL8*, *HERC5*, *HERC3*) [23]. There were several genes that have been reported to be of functional significance in other species. *SLIT2*, *PACRGL*, *GPR125*, *DHX15*, *SOD3* and *KCNIP4* were identified in a mice QTL interval linked to thermal nociception [24, 25]. *TBC1D14* has a critical development role for RAB-GAP-mediated protein transport in early embryogenesis in mouse, while its paralog, *TBC1D12*, has been implicated as a candidate gene for environmental genetic adaptation [26]. *TROVE2*, *UCHL5* and *RGS1* were identified as autoimmune disease response candidate genes in activated T cells in human subjects and in interactions with ubiquitin and proteasome pathways [27] while RGS proteins play an essential role in B-lymphocytes and thus in immune response in mice [28].

Protein–protein interactions (PPI), as well as GO enrichment, were also investigated for the 68 most plausible candidate genes using STRING. The network proteins encoded by the 68 genes had significantly more interactions amongst themselves than was expected for a random set of proteins of similar size, drawn from the genome (85 edges identified; PPI enrichment P -value = 1.0×10^{-16} ; Figure 2). STRING also revealed two GO terms that were the most significantly enriched (FDR < 0.05) *viz.* negative regulation of G protein-coupled receptor (GPCR) signaling (GO: GO:0045744) and cellular response to stress (GO:0033554).

GPCR's represent the largest family of transmembrane receptors in the mammalian genome and occur in nearly all multicellular organisms [29, 30]. They play a major role in signal transduction and perception of external environments in mammalian cells. In vertebrates, GPCR signaling networks have been associated with neurotransmission, cellular metabolism, secretion, cellular differentiation and growth, inflammatory and immune responses, smell, taste and vision [31]. Cellular stress response encompasses a wide range of molecular changes in response to environmental stressors, including extremes of temperature, exposure to toxins, and mechanical damage. The various processes involved in cellular stress response serve the adaptive function of protection against unfavourable environmental conditions, through either short-term mechanisms that minimize acute damage to cell integrity, and/or long-term mechanisms which provide a measure of resiliency against adverse conditions.

Discussion

Goats have played multiple and crucial socio-cultural and economic roles in ancient and modern societies. Accurate characterization of the species genome architecture could help infer the genetic and functional variation underlying the population structure and the design of appropriate conservation, and genomic selection and breeding programs. Using 708 samples we provide an assessment of the genome structure amongst goats of African descent from a hot non-tropical arid environment and those of European ancestry selected artificially for production traits in temperate environments.

Comparing the results of the three analyses, there were differences in the number of signals detected. The *di* analysis detected four candidate regions, compared with 19 and 13 that were detected by *Rsb* and *iHS*, respectively. The differences can be attributed to the configuration of the three selection mapping statistics that capture different patterns in the genome [32, 33, 34]. Selection signatures based on genetic differentiation persists for a longer period than signals based on haplotype structure. The, *di* therefore can detect older signatures of selection [32, 34] which are most likely associated with long term adaptive divergence. The signatures that are based on haplotype structure arise from short- and/or medium-term events. They are likely to reveal outcomes of artificial selection and formation of breeds given that systematic selection and animal breeding based on predefined traits, which resulted in well-defined

breeds, came into effect in the mid-1800's [35]. The combination of the three methods therefore provides a comprehensive picture of the signatures of selection acting in the genomes of goats that are exposed to different environments and selection pressures. Although few overlapping selection signature regions have been identified using computationally different approaches, when they occur, they do provide support for credibility of the results [36] and in our case, they increase the reliability of the nine primary candidate regions reported here to result from genuine selection events. Furthermore, our selection signature scans used allele frequencies that were derived from multiple populations, and individual chromosome segments have a low probability of drifting, simultaneously, towards fixation in several populations.

In performing the *di* and *Rsb*, we pooled four Egyptian breeds of goats into one test group (Egyptian predominantly under natural selection) and 10 breeds from Europe into another group (non-Egyptian predominantly under artificial selection). The grouping was based on *a priori* knowledge of the breeding history and geographic-environmental origin of the breeds. Such a pooling strategy has been implemented previously [13, 15, 37]. Sample pooling offers several advantages. Apart from the fact that it results in stronger and more distinct signals of selection, it also reduces the effects of population specific linkage disequilibrium arising from genetic drift, smoothens out population specific effects that may arise from specific genetic backgrounds and averages out the effects of within group structure on haplotype homozygosity [37].

Five of the candidate regions under selection were gene deserts. This can be due to the lack of genes with functional significance in the candidate regions or more likely due to poor annotation. In observing such regions in cattle, Zhao et al. [21] suggested that such regions are likely to play important roles in adaptation which may be elucidated in future with improvements in genome annotations. It is possible that there may be an uncharacterized non-coding RNA in the candidate regions that is mutated. There is one such reported case of an ENU-induced mutation in a microRNA gene resulting in progressive hearing loss in mice [38]. Furthermore, functional genes may also affect long range regulatory sequences several base pairs away through hitchhiking. This appears to be the case with an ENU-induced allele of the quaking gene that is located 40-640 Kb upstream of a candidate region [39] or the Callipyge point mutation in sheep that affects transcription of genes up to several hundred kilobases away [40, 41]. Given these observations, we investigated the flanking regions of the five gene deserts which revealed the presence of functional genes downstream and/or upstream of the regions. For instance, *TMC05A* and *MEIS2* were 156,249 bp and 598,985 bp downstream and upstream of the region on chromosome 10, respectively, *KCTD12* occurred 61,090 bp downstream of the region on chromosome 12, and *ADAMTS20* was 585,283 bp downstream of the region on CHI05. *TMC05A* is related to spermatogenesis and testis development processes with significant glycolysis and pyruvate metabolism [42, 43, 44]. *KCTD12* is a component of the G-protein-coupled receptors for Y-aminobutyric acid receptor and using B16 murine

melanoma cells it has been shown that it may play a role in hypoxic responses [45, 46]. Several studies have implicated mutations in *ADAMTS20* in skin pigmentation [47, 48]. The current data therefore raises the possibility that, by virtue of their functions, mutations in the genes flanking the gene deserts may affect long range regulatory sequences and this may explain our finding. Alternatively, the mutations could disrupt local chromatin structures, or interaction of chromatin with the nuclear matrix, in a manner that affects the expression of other genes. However, these postulations require further investigations. Given the limited marker density in our data and the finding of Kemper et al. [49] that response to selection is usually based on small changes in frequency at many loci, it is likely that we may have only detected loci with major effects due to strong selection. The collection and analysis of many phenotypes together with high-density genotype data or full genome sequences, may be necessary to disentangle the genetic basis and functional significance of the gene deserts.

Out of the 36 candidate regions, nine were defined by the strongest signals and spanned 68 functional genes. We hypothesize that these nine candidate regions and the associated 68 genes could be the primary candidates driving genomic adaptive divergence and evolution in the study breeds and could therefore be potential targets for functional studies. A large number of the 68 candidate genes are linked to multiple functions. This suggests that the underlying genetic mechanism for optimal productivity and performance and adaptation to arid environments are complex and may involve multiple genomic regions and genes with pleiotropic functions. For instance, *PPARGC1A* mediates expression of genes involved in oxidative metabolism, adipogenesis, and gluconeogenesis [50]. Expression of this gene has been suggested to be required for the initiation and development of lactation in dairy cattle [18] and has also been shown to be associated with milk composition [18, 51], reproduction [52], growth [53], carcass traits [54] and meat quality [55]. *LAP3* and *CCSER1* have been associated with milk production and body size in Charolais cattle [12, 21, 56]. Furthermore, DAVID revealed 12 gene clusters that were functionally enriched while STRING revealed more significant protein-protein interactions than were expected for the 68 primary genes suggesting the proteins are, at least partially biologically linked as a functional group.

The variability in signatures of selection can shed light on genome function and evolution. Our analysis revealed several genes associated with production traits, and variation in the latter can provide evidence of phenotypic response to artificial selection [12]. Genes underlying variation in quantitative and qualitative traits are of interest in breed formation, breeding improvement and tracing trajectories of genome evolution [12]. Several causal variants identified within *PLAG1*, *DCAF16*, *NCAPG* and *LCORL* genes, which correspond to the same QTL, have been reported to affect calving ease through their impacts on animal development (growth and weight, body size, height and stature) and birth weights in sheep [9], cattle [16, 57], pigs [58], horse [59] and dogs [60]. *RGS* proteins have been implicated in efficiency in differentiation during adipogenesis and osteogenesis and thus in bone development and therefore body size and fat deposition [61]. In addition to *LAP3*, *LCORL* and *NCAPG*, *IBSP*, *MED28*, *MEPE*

and *ABCG2* have been associated with dairy traits and explain a large percentage of the variation in milk yields, protein content and fat composition [62, 63, 64]. Through differential expression analysis, it was revealed that *ABCG2* also acts as a regulator of lactation [65]. *NCAPG* also plays a role in decreasing subcutaneous fat thickness and increasing carcass yields while *PI4K2B* and *PPARGC1A* are involved in fatty acid synthesis and metabolic pathways. The considerable overlap between our results and those of other studies, appear to suggest that the genetic mechanisms underlying several production traits (milk and body size) might be the same across domestic livestock and it is unlikely that they are false associations, or our selection signatures are false positives. It also implies that the incorporation of any identified causative variants into genomic selection strategies could improve their accuracy and robustness, while allowing targeted selection to achieve more rapid genetic improvement across domestic livestock. Further studies may therefore be required to assess how mutations in these candidate genes affect protein functions and phenotypes.

Our results provided indirect insights on possible mechanisms underlying adaptation to arid environments in indigenous goats as evidenced by the two most significant biological processes, negative regulation of G protein-coupled receptor (GPCR) signaling and cellular response to stress as revealed by STRING. Arid environments are characterized by a number of bio-physical stressors including high temperatures, solar radiation, feed and water stress. Indigenous domestic animals are known to be well adapted to these stressors and several studies on signatures of selection have been undertaken to unravel the genetic basis of adaptations to environments in a number of livestock species [7, 13, 15, 26, 66]. A paralog of *TBC1D14*, namely *TBC1D12*, was present in one of our candidate regions and was identified previously by Lv et al. (26) to be candidate gene for environmental adaptation in sheep. Other genes such as *SLIT2*, *PACRGL*, *GPR125*, *DHX15*, *SOD3* and *KCNIP4* were identified in a mice QTL interval linked to thermal nociception [24, 25]. Environmental stressors can trigger a variety of responses at the cellular and subsequently organism level. Depending on the severity and duration of the stress factors, cells either re-establish cellular homeostasis to the former state or adopt an altered state in the new environments. Signal transduction, a fundamental biological process that is required to maintain cellular homeostasis and ensure coordinated cellular activity in all organisms, is important in this regard. Cell surface membrane proteins provide the communication interface between the cell's external and internal environments. GPCRs are one of the largest and most diverse membrane protein families [67]. They function by detecting a wide spectrum of extracellular signals and following ligand binding, undergo conformational changes, causing the activation of complex cytosolic signalling networks, resulting in a cellular response [68]. Based on our results, it is likely that an interplay between the genes modulating cellular stress response and GPCRs constitutes a "stress surveillance and response system" which functions to detect biotic and abiotic stressors to trigger appropriate responses resulting in either short- and/or long-term adaptation. We hypothesize that the "stress surveillance and response system" could be highly conserved in indigenous goats and may reflect shared mechanisms of environmental response related to GPCR signaling across domestic animals.

Conclusions

We carried out a genome-wide scan of selection signatures using the Caprine 52K SNP BeadChip array that contrasted breeds of goats from an arid environment in Egypt and a temperate one in Europe. The two genetic stocks represented diverse gene pools based on location and breeding history. The analysis of these two diverse genetic stocks identified a number of genes under positive selection that are related to milk production, growth and body size, and other functions such as fatty acid synthesis and response to stress and adaptation. Although some of the genes were previously reported in other domestic animals, we confirmed the occurrence of positive selection in the goat breeds using more diverse samples. We also revealed other genes not reported before in domestic animals that are linked to skeletal structure and thus body size, reproduction and nociception, aspects which are important in adaptation to stressful environments. Our results clearly indicate that diverse genomic selection during the process of genetic improvement and breed formation, and adaptation to arid environments have contributed to the sculpturing of the caprine genome.

Methods

Sampling, genotyping and data quality control: Whole blood was collected via jugular venipuncture from 366 individuals of four breeds of Egyptian indigenous goats (Barki, Saidi, Wahati, Nubian; Supplementary Table S1), using EDTA coated vacutainer tubes. All the animals were released immediately following sampling for grazing. The animals were selected at random from flocks maintained by livestock producers and research farm. Efforts were made to ensure the animals represented, as much as possible, the target breeds of goats. Genotyping was performed on the Illumina CaprineSNP52K BeadChip at GeneSeek Inc. The raw genotypes for the 53,347 SNPs included in the genotyping platform were first analysed with GenomeStudio (Illumina; GenCall score for raw genotypes > 0.15) and used to extract the genotypes in standard format for PLINK 1.09 [69]. Amongst the 53,347 SNPs that were genotyped, 1483 that failed genotyping in at least 80% of the individuals in each population and 1756 that were not polymorphic (minor allele frequency (MAF) ≤ 0.03) were dropped. We included in this study, a subset of SNP genotypes that were generated with the same BeadChip array that were collated within the framework of the ADAPTmap project [70]. The extracted subset included 447 individuals of 12 European breeds of goats that have been bred and selected for meat and milk production under temperate environments (Supplementary Table S1). These two sets of genetic stocks present divergent but ideal, gene pools for assessing genome regions under different selection pressures (mainly natural versus artificial). The two datasets were merged and the SNPs that mapped on autosomes, with positions based on the goat genome assembly v1.0 [71], were chosen for analysis. In total, 708 individuals with genotypes for at least 95% of the SNPs were available for analysis. This dataset was subjected to the following quality control filters: individual call rates $\geq 95\%$, SNP call rates $\geq 90\%$, MAF ≤ 0.03 and genotype frequencies had to be in Hardy–Weinberg equilibrium with $P \geq 10^{-4}$. The number of SNPs retained in the final dataset following quality control filtering was 49,056.

Detection of selection signatures: The genotypes were pooled into two groups comprised of Egyptian and European (non-Egyptian) breeds of goats based on *a priori* knowledge of their breeding history and geographic origin. Three analyses were performed to detect genomic regions under positive selection. A genetic differentiation analysis was used to contrast the Egyptian and European breeds by calculating the unbiased estimate of F_{ST} [72] as described by Akey et al. [73], for each of the 49,056 SNPs that passed quality thresholds. The F_{ST} values were averaged across a sliding window of 20 SNPs which overlapped by five SNPs following Kim et al. [7]. Identification of candidate signatures of selection was based on window estimates at the extreme of the empirical distributions following Akey et al. [73]. This approach has been used to detect selection sweeps that are specific to a breed/population, or a subset of breeds/populations [7, 12, 74]. Specifically, we considered that a position carried a signature of selection if it occurred in the top 1 percentile of the empirical distribution for genetic differentiation.

Two extended haplotype homozygosity (EHH) statistics, iHS [75] and Rsb [76] were also used to map selection signatures. The tests detect regions with high levels of haplotype homozygosity, either within (iHS) or between (Rsb) breeds, over an unexpected extended genomic region relative to neutral expectations. The intra-population Integrated Haplotype Score (iHS score) was computed using the `rehh` package [77] in R [78]. The natural log of the ratio between the integrated EHH for the ancient (iHH_A) and derived (iHH_D) alleles for the Egyptian goats breed-group were calculated for each SNP. The allelic states (ancient/derived) of each SNP were inferred using two approaches: i) the allele with the highest frequency was inferred as the ancestral one as was determined for humans [79]; ii) the allelic states were assigned at random after permutating 100 times to ensure consistency in the assigned allelic states. A two tailed Z-test was used to identify statistically significant SNPs under selection with an unusual extended haplotype of ancient (positive iHS) or derived (negative iHS) alleles. Assuming the standardized iHS values were normally distributed, as is expected under neutrality, two-sided P-values that are associated with the neutral hypothesis of no selection were derived as $-\log_{10}(1-2|\Phi(iHS)-0.5|)$, where $\Phi(iHS)$ is the Gaussian cumulative distribution function.

The inter-population Rsb scores were also computed between the Egyptian and European breed-groups of goats. The Rsb statistic was defined as the natural logarithm of the ratio between iES_{Egyptian} and iES_{European} . For this purpose, the integrated EHHS (site-specific EHH) for each SNP and group of goats (iES) was calculated. A Z-test was used to identify the significant SNPs under selection in the Egyptian group of goats. One sided P-values were derived as $-\log_{10}(1-\Phi(Rsb))$, where $\Phi(Rsb)$ represents the Gaussian cumulative distribution function.

For the two EHH statistics, haplotypes were reconstructed by phasing the genotyped SNPs using Beagle [80]. Sliding windows of 20 SNPs that overlapped by five SNPs as per Kim et al. [7] were used to calculate haplotype frequencies. The $-\log_{10}(\text{P-value}) > 4.0$ ($\text{P-value} < 0.0001$), was used as the threshold to define the significant iHS and Rsb scores. Candidate regions were identified if at least three SNPs passed the threshold. In instances where multiple SNPs passed the threshold, a distance of 0.25 Kb up- and down-stream of the most significant SNP was used to define candidate selection region intervals. This distance was chosen based on the rate of change in the mean pairwise linkage disequilibrium statistic (r^2) calculated using the *r2fast* function of GenABEL [81], binned over distance intervals across autosomes of the Egyptian goats (Supplementary Figure S1).

Functional annotation and protein-protein interactions: All variants found within the confidence intervals of the candidate selection regions were annotated. To avoid missing important genes while including potential effects of regulatory changes on loci, as well as reduce the risk of excluding the outermost portions of the selected haplotypes, as may arise due to using sliding windows of fixed sizes [82], the confidence intervals were extended by 100 kb on each side of the candidate region interval. We retrieved genes within the candidate region intervals from the NCBI database based on the caprine ARS1 reference genome assembly [71], *Capra hircus* Annotation Release 102. Functional annotation and gene ontology (GO) enrichment analysis were performed with the function annotation clustering tool of DAVID 6.8 [83, 84]. Each gene was analyzed and an enrichment analysis was performed using the *B. taurus* genome background supplied with DAVID 6.8. Corrections for multiple testing were performed by applying the Benjamini and Hochberg [85] approach. The GO terms and the KEGG pathways were considered significant at P values lower than 0.05. In the case of the functional annotation clusters, an enrichment score of 1.3 was used as the threshold following Huang et al. [83, 84].

Functional protein–protein interaction (PPI) networks encoded by the possible candidate genes were also investigated using STRING Genomics 11.0 [86]. STRING provides known PPI from curated databases or experiments and PPI predicted based on gene neighborhood, fusions, and co-occurrences, text mining in literature, co-expression, or protein homology. A global PPI network which retained only interactions with a high level of confidence (score > 0.4) was constructed.

Declarations

Acknowledgements

Special thanks to flock owners in Egypt whose goats were sampled and to ADAPMap consortium for providing genotypes of the European breeds. This study was carried out as a component of the ARC-Egypt/ICARDA Small Ruminant Research Program.

Funding statement

ICARDA thanks the donors supporting the Egypt/ICARDA Small Ruminant Research Program and the CGIAR Research Program on Livestock (Livestock CRP) and its predecessor Livestock and Fish (Livestock and Fish CRP). The genotyping was supported by the generous donation through the Illumina Greater Good Initiative and additionally by discounted prices from Gene Seek Inc. Activities involving Iowa State University were supported by the Ensminger Endowment, the Department of Animal Science and the College of Agriculture and Life Sciences and the State of Iowa. The funding bodies did not participate in or influence any part of this study and in drafting the manuscript.

Ethics approval and consent to participate

The animals used in this study are owned by farmers and pastoralists and 35 Barki individuals came from Egypt's Animal Production Research Institute (APRI), research farm. The farmers provided informed consent to sample their animals after the objectives of the study were explained to them in local languages. Given the nature of sampling from the small holder farmers and pastoralists no names or addresses were obtained at the time of sampling and therefore cannot be published. No specific permissions were needed to sample the animals maintained in APRI's research farm. Blood sampling was performed by a licensed veterinarian following the guidelines of the General Organization for Veterinary Service (GOVS), Egypt. The study and all the procedures described here were approved by the Department of Veterinary Services of the Government of Egypt and by the Institute Ethics Committee of the International Centre for Agricultural Research in the Dry Areas (ICARDA).

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and/or analysed during the current study are available from the ADAPTMap Project consortium. <http://www.goatadaptmap.org/>.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MJM, AAM, RB, and RMF conceptualized and designed the study. AAM performed the sampling. RMF provided genotyping reagents. MJM, KE-S and EAR analysed the data and interpreted the results. MJM supervised the study and wrote the manuscript. All authors read and approved the manuscript.

References

1. Diamond J. Evolution, consequences and future of plant and animal domestication. *Nature*. 2002;418:700–7.
2. Zeder MA, Hesse B. The initial domestication of goats (*Capra hircus*) in the Zagros Mountains 10,000 years ago. *Science*. 2000;287:2254-7.
3. FAOSTAT. <http://www.fao.org/faostat/en/#data/QA>. Accessed 19th August 2019.
4. Pereira F, Queirós S, Gusmão L, Nijman IJ, Cuppen E, Lenstra JA, Econogene Consortium, Davis SJM, Nejmeddine F, Amorim A. Tracing the history of goat pastoralism: New clues from mitochondrial and Y chromosome DNA in North Africa. *Mol Biol Evol*. 2009;26:2765-73.
5. Oleksyk, TK, Smith MW, O'Brien SJ. Genome-wide scans for footprints of natural selection. *Philos. Trans. R. Soc. London B*. 2010;365:185–205.
6. Gouveia JJS, Barbosa da Silva MVG, Paiva SR, de Oliveira SMP. Identification of selection signatures in livestock species. *Genet Mol Biol*. 2014;37:330-42.
7. Kim E-S, Elbeltagy AR, Aboul-Naga AM, Rischkowsky B, Sayre B, Mwacharo JM, Rothschild MF. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity*. 2016;116:255-64.
8. Gutiérrez-Gil B, Arranz JJ, Pong-Wong R, García-Gámez E, Kijas J, Wiener P. Application of selection mapping to identify genomic regions associated with dairy production in sheep. *PLoS One* 2014;9:e94623.
9. Matika O, Riggio V, Anselme-Moizan M, Law AS, Pong-Wong R, Archibald AL, Bishop SC. Genome-wide association reveals QTL for growth, bone and in vivo carcass traits as assessed by computed tomography in Scottish Blackface lambs. *Genet Sel Evol*. 2016;48:11.
10. Xia J, Fan H, Chang T, Xu L, Zhang W, Song Y, Zhu B, Zhang L, Gao X, Chen Y, Li J, Gao H. Searching for new loci and candidate genes for economically important traits through gene-based association analysis of Simmental cattle. *Sci Rep*. 2017;7:42048.
11. Chung NC, Szyda J, Fraszczak M. Population structure analysis of bull genomes of European and western ancestry. *Sci. Rep*. 2017;7:40688.
12. Xu L, Bickhart DM, Cole JB, Schroeder SG, Song J, Tassell CP, et al. Genomic signatures reveal new evidences for selection of important traits in domestic cattle. *Mol Biol Evol*. 2015;32:711–25.

13. Bahbahani H, Salim B, Almathen F, Al Enezi F, Mwacharo JM, Hanotte O. Signatures of positive selection in African Butana and Kenana dairy zebu cattle. *PLoS ONE*. 2018;13:e0190446.
14. Iso-Touru T, Tapio M, Vilkki J, Kiseleva T, Amimosov I, Ivanova Z, Popov R, Ozerov M, Kantanen J. Genetic diversity and genomic signatures of selection among cattle breeds from Siberia, eastern and northern Europe. *Anim Genet*. 2016;47:647–57.
15. Kim J, Hanotte O, Mwai OA, Dessie T, Bashir S, Diallo B, Agaba M, Kim K, Kwak W, Sung S, Seo M, Jeong H, Kwon T, Taye M, Song K-D. et al. The genome landscape of indigenous African cattle. *Genome Biol*. 2017;18:34.
16. Olsen HG, Meuwissen TH, Nilsen H, Svendsen M, Lien S. Fine mapping of quantitative trait loci on bovine chromosome 6 affecting calving difficulty. *J. Dairy Sci*. 2008;91:4312-22.
17. García-Fernández M, Gutiérrez-Gil B, Sánchez JP, Morán JA, García-Gámez E, Alvarez L, et al. The role of bovine causal genes underlying dairy traits in Spanish Churra sheep. *Anim Genet*. 2011;42:415–20.
18. Weikard R, Widmann P, Buitkamp J, Emmerling R, Kuehn C. Revisiting the quantitative trait loci for milk production traits on BTA6. *Anim Genet*. 2012;43:318–23.
19. Pan D, Zhang S, Jiang J, Jiang L, Zhang Q, Liu J. Genome-wide detection of selective signature in Chinese Holstein. *PLoS One*. 2013;8:e60440.
20. Ruiz-Larrañaga O, Langa J, Rendo F, Manzano C, Iriando M, Estonba A. Genomic selection signatures in sheep from the Western Pyrenees. *Genet Sel Evol*. 2018;50:9.
21. Zhao F, McParland S, Kearney F, Du L, Berry DP. Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genet Sel Evol*. 2015;47:49.
22. Al-Mamun HA, Kwan P, Clark SA, Ferdosi MH, Tellam R, Gondro C. Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight. *Genet Sel Evol*. 2015;47:66.
23. Dantoft W, Martinez-Vicente P, Jafali J, Perez-Martinez L, Martin K, Kotzamanis K, Craigon M, Auer M, Young NT, Walsh P, Marchant A, Angulo A, Forster T, Ghazal P. Genomic programming of human neonatal dendritic cells in congenital systemic and in vitro /I cytomegalovirus infection reveal plastic and robust immune pathway biology process. *Front Immunol*. 2017;8:1146.
24. Philip VM, Sokoloff G, Ackert-Bicknell CL, Stritz M, Branstetter L, Beckmann MA, Spence JS, Jackson BL, Galloway LD, Barker P, Wymore AM, Hunsicker PR, Durtschi DC, Shaw GS, Shipcock S, Manly KF, Miller DR, Donohue KD, Culiati CT, Churchill GA, Lariviere WR, Palmer AA, O'Hara BF, Voy BH, Chesler EJ. Genetic analysis in the collaborative cross breeding population. *Genome Res*. 2011;21:1223-38.
25. Baud A, Flint J. Identifying genes for neurobehavioural traits in rodents: progress and pitfalls. *Dis. Models Mech*. 2017;10:373-83.
26. Lv F-H, Agha S, Kantanen J, Colli L, Stucki S, Kijas JW, Joost S, Li M-H, Ajmone-Marsan P. Adaptations to climate-mediated selective pressures in sheep. *Mol Biol Evol*. 2014;31:3324–43.
27. Burren OS, Garcia AR, Javierre B-M, Rainbow DB, Cairns J, Cooper Nj, Lambourne JJ, Schofield E, Dopico XC, Ferreira RC, Coulson R, Burden F, Rowlston SP, Downes K, Wingett SW, Frontini M,

- Ouwehand WH, Fraser P, Spivakov M, Todd JA, Wicker LS, Cutler AJ, Wallace C. Chromosome contacts in activated T cells identify autoimmune disease candidate genes. *Genome Biol.* 2017;18:165.
28. Hwang IY, Park C, Harrison K, Boularan C, Gales C, Kehrl JH. An essential role for RGS protein/GAlfai2 interactions in B lymphocyte-directed cell migration and trafficking. *J. Immunol.* 2015;194:2128-39.
 29. Bockaert J, Pin JP. Molecular tinkering of G protein-coupled receptors: an evolutionary success. *EMBO J.* 1999;18:1723-29.
 30. Fredriksson R, Schioth HB. The repertoire of G-protein-coupled receptors in fully sequenced genomes. *Mol Pharmacol.* 2005;67:1414-25.
 31. Qian B, Soyer OS, Neubig RR, Goldstein RA. Depicting a protein's two faces: GPCR classification by phylogenetic tree-based HMMs. *FEBS Lett.* 2003;554:95-9.
 32. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, et al. Positive natural selection in the human lineage. *Science.* 2006;312:1614–20.
 33. Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics.* 2013;193:929–41.
 34. González-Rodríguez A, Munilla S, Mouresan EF, Cañas-Álvarez JJ, Díaz C, Piedrafita J, et al. On the performance of tests for the detection of signatures of selection: a case study with the Spanish autochthonous beef cattle populations. *Genet Sel Evol.* 2016;48:81.
 35. Porter V. *Masons World Dictionary of Livestock Breeds, Types and Varieties.* 5th Edition. Wallingford: CABI Publishing. 2002.
 36. Yang J, Li W-R, Lv F-H, He S-G, Tian S-L, Peng W-F, Sun Y-W, Zhao Y-X, Tu X-L, Zhang M, Xie X-L, Wang Y-T, Li J-Q, Liu Y-G, Shen Z-Q, Wang F, Liu G-J, Lu H-F, Kantanen J, Han J-L, Li M-H, Liu M-J. Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Mol Biol Evol.* 2016;33:2576-92.
 37. Gautier M, Naves M. Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Mol Ecol.* 2011;20:3128-43.
 38. Lewis MA, Quint E, Glazier AM, Fuchs H, De Angelis MH, Langford C, van Dongen S, Abreu-Goodger C, Piipari M, Redshaw N. et al. An ENU-induced mutation of miR-96 associated with progressive hearing loss in mice. *Nat Genet.* 2009;41:614–18.
 39. Noveroske JK, Hardy R, Dapper JD, Vogel H, Justice MJ. A new ENU-induced allele of mouse quaking causes severe CNS dysmyelination. *Mamm Genome.* 2005;16:672–82.
 40. Freking BA, Murphy SK, Wylie AA, Rhodes SJ, Keele JW, Leymaster KA, Jirtle RL, Smith TP. Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Res.* 2002;12:1496–1506.
 41. Takeda H, Caiment F, Smit M, Hiard S, Tordoir X, Cockett N, Georges M, Charlier C. The callipyge mutation enhances bidirectional long-range DLK1-GTL2 intergenic transcription in cis. *Proc Natl Acad Sci USA.* 2006;103:8119–24.

42. Liu F-J, Jin S-H, Li N, Liu X, Wang HY, Li J-Y. Comparative and functional analysis of testis-specific genes. *Biol. Pharm. Bull.* 2011;34:28-35.
43. Djureinovic D, Fagerberg L, Hallström B, Danielsson A, Lindskog C, Uhlén M, Pontén F. The human testis-specific proteome defined by transcriptomic and antibody-based profiling. *Mol Hum Reprod.* 2014;20:476-88.
44. Kaneko T, Minohara T, Shima S, Yoshida K, Fukuda A, Iwamori N, Inai T, Iida H. A membrane protein, TMC05A, has a close relationship with manchette microtubules in rat spermatids during spermiogenesis. *Mol Reprod Dev.* 2019;86:330-41.
45. Zhou L, Wang LM, Song HM, Shen YQ, Xu WJ et al. Expression profiling analysis of hypoxia pulmonary disease. *Genet. Mol. Res.* 2013;12:4162-70.
46. Hu H, Petousi N, Glusman G, Yu Y, Bohlender R, Tashi T, Downie JM, Roach JC, Cole AM et al. Evolutionary history of Tibetans inferred from whole-genome sequencing. *PLoS Genet.* 2017;13:e1006675.
47. Rao C, Foernzler D, Loftus SK, Liu S, McPherson JD, Jungers KA, Apte SS, Pavan WJ, Beier DR. A defect in a novel ADAMTS family member is the cause of the belted white-spotting mutation. *Development.* 2003;19:4665-72.
48. Silver DL, Hou L, Somerville R, Young ME, Apte SS, Pavan WJ. The secreted metalloprotease ADAMTS20 is required for melanoblast survival. *PLoS Genet.* 2008;4:e1000003.
49. Kemper KE, Saxton SJ, Bolormaa S, Hayes BJ, Goddard ME. Selection for complex traits leaves little or no classic signatures of selection. *BMC Genomics.* 2014;15:246.
50. Puigserver P, Spiegelman BM. Peroxisome proliferator-activated receptor-gamma coactivator 1 alpha (PGC-1 alpha): transcriptional coactivator and metabolic regulator. *Endocr Rev.* 2003;24:78-90.
51. Schennink A, Bovenhuis H, Léon-Kloosterziel KM, van Arendonk JA, Visker MH. Effect of polymorphisms in the FASN, OLR1, PPARGC1A, PRL, and STAT5A genes on bovine milk-fat composition. *Anim Genet.* 2009;40:909-16.
52. Komisarek J, Walendowska A. Analysis of the PPARGC1A gene as a potential marker for productive and reproductive traits in cattle. *Folia Biol.* 2012;60:171-4.
53. Li Q, Wang Z, Zhang B, Lu Y, Yang Y, Ban D et al. Single nucleotide polymorphism scanning and expression of the pig PPARGC1A gene in different breeds. *Lipids.* 2014;49:1047-55.
54. Ramayo-Caldas Y, Fortes MR, Hudson NJ, Porto-Neto LR, Bolormaa S, Barendse W et al. A marker-derived gene network reveals the regulatory role of PPARGC1A, HNF4G and FOXP3 in intramuscular fat deposition of beef cattle. *J Anim Sci.* 2014;92:2832-45.
55. Sevane N, Armstrong E, Cortés O, Wiener P, Wong RP, Dunner S, et al. Association of bovine meat quality traits with genes included in the PPARG and PPARGC1A networks. *Meat Sci.* 2013;94:328-35.
56. Randhawa IA, Khatkar MS, Thomson PC, Raadsma HW. A meta-assembly of selection signatures in cattle. *PLoS One.* 2016;11:e0153013.

57. Bongiorno S, Mancini G, Chillemi G, Pariset L, Valentini A. Identification of a short region on chromosome 6 affecting direct calving ease in Piedmontese cattle breed. *PLoS One*. 2012;7:e50137.
58. Rubin CJ, Megens HJ, Martinez BA, Maqbool K, Sayyab S, Schwochow D, Wang C, Carlborg O, Jern P, Jorgensen CB et al. Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci USA*. 2012;109:19529-536.
59. Metzger J, Philipp U, Lopes MS, Machado da Camara A, Felicetti M, Silvestrelli M, Distl O. Analysis of copy number variants by three detection algorithms and their association with body size in horses. *BMC Genomics*. 2013;14:487.
60. Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Rosengren Pielberg G, Sigurdsson S, Fall T, Seppälä EH, Hansen MST, Lawley CT, et al. Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet*. 2011;7:e1002316.
61. Madrigal A, Tan L, Zhao Y. Expression regulation and functional analysis of RGS2 and RGS4 in adipogenic and osteogenic differentiation of human mesenchymal stem cells. *Biol Res*. 2017;50:43.
62. Zheng X, Ju Z, Wang J, Li Q, Huang J, Zhang A, Zhong J, Wang C. Single nucleotide polymorphisms, haplotypes and combined genotypes of LAP3 gene in bovine and their association with milk production traits. *Mol. Biol. Rep*. 2011;38:4053-61.
63. Rothammer S, Seichter D, Martin Förster M, Medugorac I. A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. *BMC Genomics*. 2013;14:908.
64. Sanchez MP, Govignon-Gion A, Croiseau P, Fritz S, Hozé C, Miranda G, Martin P, Barbat-Leterrier A, Letaiéf R, Rocha D, Brochard M, Boussaha M, Boichard D. Within-breed and multi-breed GWAS on imputed whole-genome sequence variants reveal candidate mutations affecting milk protein composition in dairy cattle. *Genet Sel Evol*. 2017;49:68.
65. Sheehy PA, Riley LG, Raadsma HW, Williamson P, Wynn PC. A functional genomics approach to evaluate candidate genes located in a QTL interval for milk production traits on BTA6. *Anim Genet*. 2009;40:492–8.
66. Ahbara A, Bahbahani H, Almathen F, Al Abri M, Agoub MO, Abeba A, Kebede A, Musa HH, Mastrangelo S, Pilla F, Ciani E, Hanotte O, Mwacharo JM. Genome-wide variation, candidate regions and genes associated with fat deposition and tail morphology in Ethiopian indigenous sheep. *Front Genet*. 2019;9:699.
67. Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB (2003). The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol*. 2003;63:1256–72.
68. Venkatakrisnan AJ, Deupi X, Lebon G, Tate CG, Schertler GF, Babu MM (2013). Molecular signatures of G-protein-coupled receptors. *Nature*. 2013;494:185-94.
69. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559 –75.
70. ADAPTmap Project. <http://www.goatadaptmap.org/>. Accessed 30th April 2019.

71. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, Lee J, Lam ET, Liachko I, Sullivan ST, Burton JN, Huson HJ, Nystrom JC, Kelley CM, Hutchison JL, Zhou Y, Sun J, Crisà A, Ponce de León FA, Schwartz JC, Hammond JA, Waldbieser GC, Schroeder SG, Liu GE, Dunham MJ, Shendure J, Sonstegard TS, Phillippy AM, Van Tassell CP, Smith TP. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat Genet.* 2017;49:643–50
72. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution.* 1984;38:1358–70.
73. Akey JM, Ruhe AL, Akey DT, Wong AK, Connelly CF, Madeoy J et al. Tracking footprints of artificial selection in the dog genome. *Proc Natl Acad Sci USA.* 2010;107:1160–65.
74. Petersen JL, Mickelson JR, Rendahl AK, Valberg SJ, Andersson LS, Axelsson J, Bailey E, Bannasch D, Binns MM, Borges AA et al. Genomewide analysis reveals selection for important traits in domestic horse breeds. *PLoS Genet.* 2013;9:e1003211.
75. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol.* 2006;4:e72.
76. Tang K, Thornton KR, Stoneking M. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* 2007;5:e171.
77. Gautier M, Vitalis R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics.* 2012;28:1176–77.
78. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013:URL <http://www.R-project.org/>.
79. Hacia JG, Fan J-B, Ryder O, Jin L, Edgemon K, Ghandour G, Mayer RA, Sun B, Hsie L, Robbins CM, Brody LC, Wang D, Lander ES, Lipshutz R, Fodor SPA, Collins FS *et al.* Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. *Nat Genet.* 1999;22:164–67.
80. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. *Am J Hum Genet.* 2007;81:1084-97.
81. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics.* 2007;23:1294-96.
82. Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M. The Genomic signatures of dog domestication reveals adaptation to a starch-rich diet. *Nature.* 2013;495:360-64.
83. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc.* 2009a;4:44-57.
84. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009b;e37:1-13.
85. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol.* 1995;57:289–300.

86. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, von Mering C. STRING V11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47:D607-D13.

Table

Table 1. Nine strongest potential selection regions identified by *di*, *Rsb* and *iHS*, genes mapping to them, the 27 primary candidate genes

a) <i>Di</i> analysis					
Chromosome	From	To	Di/Rsb/iHS value	No. of significant SNPs	Genes
6	68,327,191	68,495,816	3.50	6	<i>LRR66, SGCB, LOC106502204, SPATA18</i>
12	34,478,328	35,402,998	4.57	20	<i>LOC108637258, LOC108637259</i>
b) <i>Rsb</i> analysis					
3	60,098,471	61,041,488	14.75	15	<i>ENSCHIG00000006864 (lincRNA), ENSCHIG00000006347 (rRNA), TRNAC-ACA, TLL7</i>
6	78,852,104	89,433,167	6.78	11	<i>LOC102184984, CXCL8, LOC108636281, LOC102181582, PPBP</i>
16	6,511,375	13,049,713	15.75	81	<i>ENSCHIG0000000891 (snRNA), ENSCHIG0000000823 (lincRNA), TRNAW-CCA, LOC102188613, LOC102190155, ENSCHIG0000000829 (lincRNA), TRNAK-UUU, CDC73, LOC102170075, LOC102170344, B3GALT2, GLRX2, TROVE2, UCHL5, RGS2, RGS13, RGS1, RGS21, RGS18, LOC108637735, TRNAY-GUA, TRNAS-GGA, ENSCHIG0000000832 (lincRNA), ENSCHIG0000000834 (lincRNA), ENSCHIG0000000836 (lincRNA), ENSCHIG0000000838 (lincRNA), ENSCHIG0000000846 (lincRNA), ENSCHIG0000000849 (lincRNA), ENSCHIG0000004663 (rRNA)</i>
c) <i>iHS</i> analysis					
6	30,246,187	47,044,407	7.66	19	<i>BMPR1B, PDLIM5, HPGDS, SMARCA1, LOC102188664, LOC102191506, ATOH1, TRNAS-GGA, TRNAE-UUC, GRID2, LOC102187307, TRNAC-GCA, CCSER1, TRNAW-CCA, LOC108636213, MMRN1, LOC108636214, SNCA, GPRIN3, TIGD2, FAM13A, HERC3, NAP1L5, PIGY, LOC106502196, LOC108636216, HERC5, HERC6, PPM1K, ABCG2, PKD2, SPP1, MEPE, IBSP, LOC102176111, TRNAA-CGC, LAP3, MED28, FAM184B, LOC106502187, DCAF16, NCAPG, LCORL, TRNAC-GCA, LOC102178056, LOC102190868, SLIT2, PACRGL, KCNIP4, TRNAC-GCA, LOC102191411, ADGRA3, LOC102191687, LOC102180474, LOC102168617, LOC106502188, PPARGC1A, DHX15, LOC102169170, SOD3, CCDC149, LOC106502189, LGI2, SEPSECS, PI4K2B, ZCCHC4, LOC106502190, ANAPC4, SLC34A2, SEL1L3, LOC102183775, LOC106502199, SMIM20, TRNAA-UGC, LOC102184606, RBPJ, LOC102184796, CCKAR, TBC1D19, STIM2</i>
12	34,295,846	36,783,669	6.63	17	<i>LOC108637259, LOC108637258, LMO7, COMMD6, LOC102189877, TBC1D4, UCHL3, LOC102181790, TRNAE-CUC, LOC102189597</i>

Supplementary Material Legends

Supplementary Figure S1. Distribution of r^2 values against genomic distances across the genome of indigenous goats.

Supplementary Table S1. Sample sizes and information on the goat populations and breeds used in the current study.

Supplementary Table S2. Potential candidate selection sweep regions and their genes determined from genome wide *di*, *Rsb* and *iHS* analysis of Egyptian indigenous goats

Figures

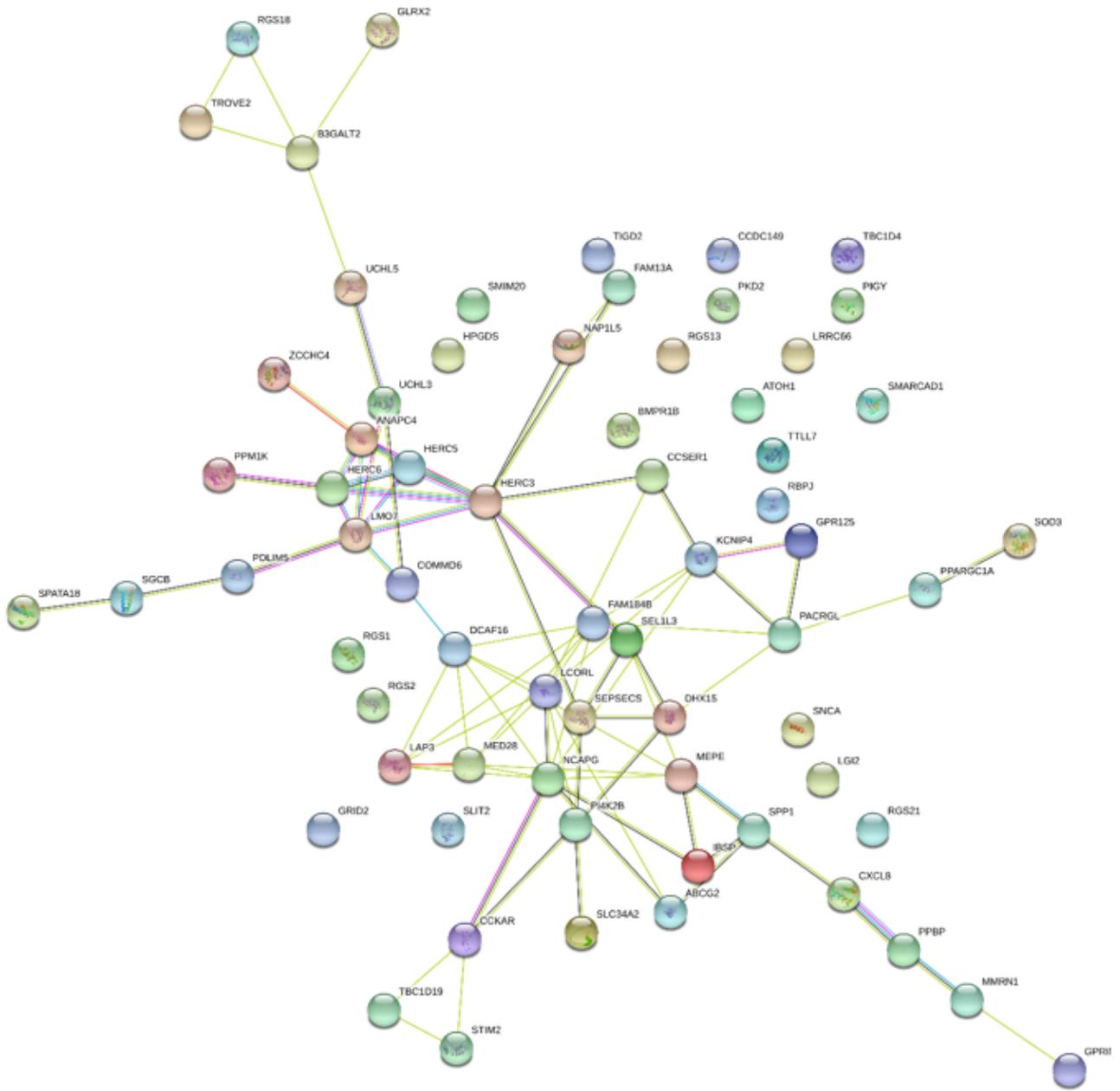


Figure 1

The Protein-Protein Interaction (PPI) network for the 68 primary candidate genes generated using STRING.

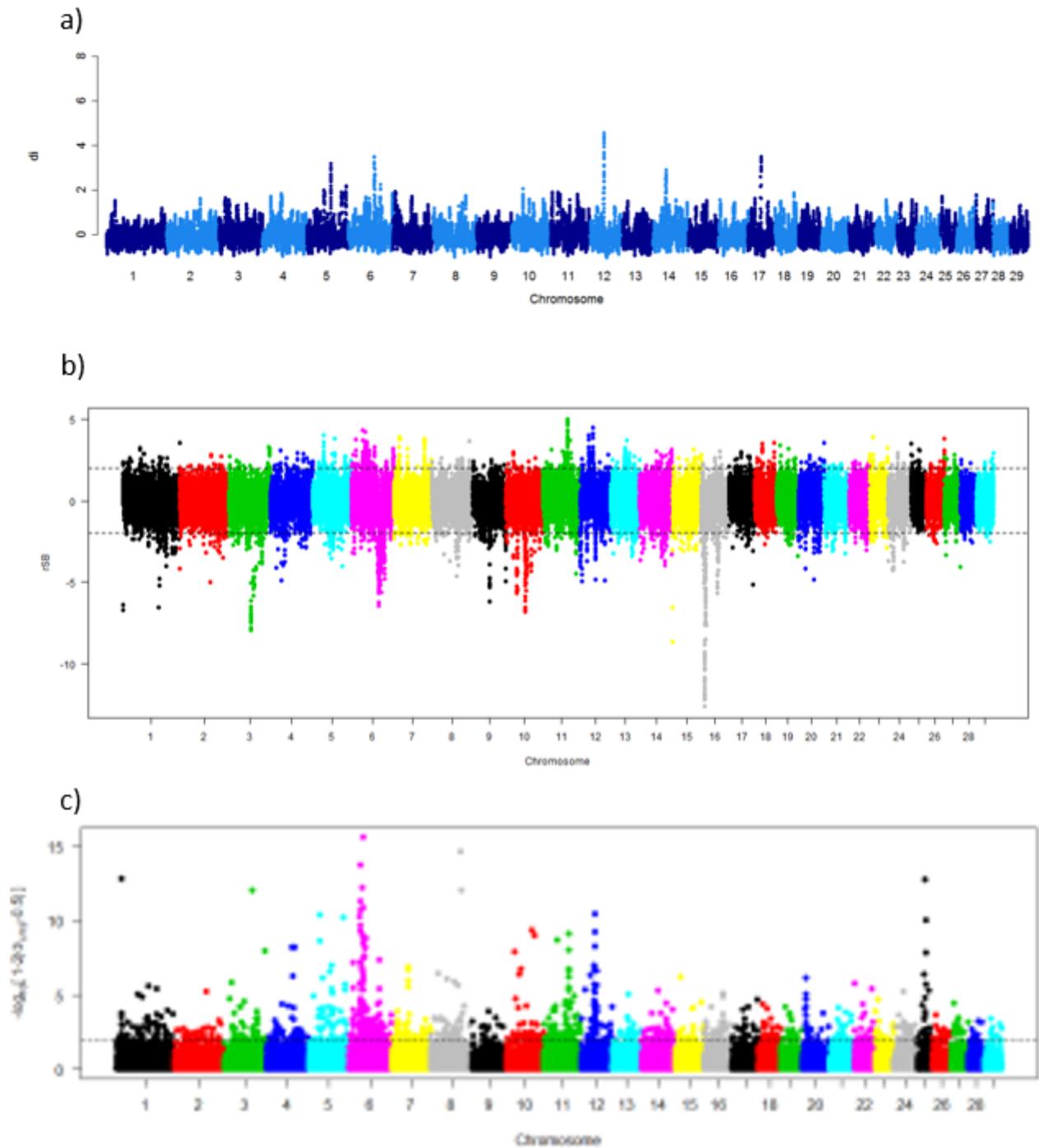


Figure 2

Manhattan plots for the genome-wide distribution of SNPs following selection sweep analysis based on d_i (1a), R_{sb} (1b) and iHS (1c) analysis.

Supplementary Files

This is a list of supplementary files associated with this preprint. [Click to download.](#)

- [TableS1andS2.pdf](#)
- [FigureS1.jpg](#)
- [SupplementaryTableS3.xlsx](#)