Fine mapping of candidate regions associated with fat deposition in thin and fat tail sheep breeds suggests new insights into molecular aspects of fat tail selection

Mohammad Hossein Moradi (✉ Hoseinmoradi@ut.ac.ir)

Arak University

Ardeshir Nejati-Javaremi

University of Tehran

Mohammad Moradi-Shahrbabak

University of Tehran

Ken G. Dodds

AgResearch

Rudiger Brauning

AgResearch

John C. McEwan

AgResearch

Research Article

Keywords: Genomic scan, Selection signature, Lipid metabolisms, Candidate genes, Fat tail sheep

Posted Date: August 23rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-832298/v1>

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

Abstract

Background

Fatness related traits are economically very important in sheep production and are associated with serious diseases in humans. The fat tail is a phenotype that divides domesticated sheep into two major groups. The objective of the present study is to refine the map location of candidate regions associated with fat deposition, obtained via two separate whole genome scans contrasting thin and fat tail breeds, and to determine the nature of the selection occurring in these regions using hitchhiking approach.

Results

Zel (thin tail) and Lori-Bakhtiari (fat tail) breed samples that had previously been run on the Illumina Ovine 50k BeadChip, were genotyped with a denser set of SNPs in the three candidate regions using a Sequenom Mass ARRAY platform. Statistical tests were then performed using different and complementary methods based on either site frequency (F_{ST} and Median homozygosity) or haplotype (iHS and XP-EHH). Results from candidate regions on chromosome 5 and X revealed clear evidence of selection with the derived haplotypes that were consistent with selection to near fixation for the haplotypes affecting fat tail size in the fat tail breed. Analysis of the candidate region on chromosome 7 indicated that selection differentiated the beneficial alleles between breeds and homozygosity has increased in the thin tail breed which also had the ancestral haplotype. These results enabled us to confirm the signature of selection in these regions and refine the critical intervals from 113kb, 201kb and 2,831kb to 28kb, 142kb and 1,006kb on chromosome 5, 7 and X respectively. These regions contain several genes associated with fat metabolism or developmental processes consisting *TCF7* and *PPP2CA* (*OAR5*), *PTGDR* and *NID2* (*OAR7*), *AR*, *EBP*, *CACNA1F*, *HSD15B*, *SLC35A2*, *BMP15*, *WDR13* and *RBM3* (*OAR X*), each of which could potentially be the actual target of selection. Study of core haplotypes alleles in our regions of interest also supported the hypothesis that the first domesticated sheep were thin tailed and fat tail animals were developed later.

Conclusions

Our results provide a comprehensive assessment of how and where selection has affected the patterns of variation in candidate regions associated with fat deposition in thin and fat tail sheep breeds. The hitchhiking mapping approach in this study was novel in the sense that most of the exploratory genome scan studies in domestic animals have not clarified the signal from the candidate regions, probably due to the lack of suitable genomic resources.

Background

Identifying regions of the genome that have experienced substantial selective pressure can provide a powerful tool into the location of functionally important polymorphisms and can help prioritize targets for association mapping [1, 2]. Fat tailed breeds comprise approximately 25% of the world sheep population and are grazing in a wide range of countries, especially in northern parts of Africa, the Middle East and Asia. The major biological role of the fat tail is to serve as an energy store against periodic food scarcity such as in drought and winter [3]. It has also been used as a source of fat (or ghee) for human consumption for millennia [4]. Historically, the climatic conditions in these areas as well as the nutritional and religious food requirements of the people encouraged sheep producers to select for higher fat tail weight [5]. However, to date the genetic mechanism underlying this particular trait has not been fully elucidated and the genes associated with fat deposition in these breeds are unknown. Therefore, finding the regions associated with fat deposition

and the nature of selection occurred in these locations is one of the most important and challenging areas of research in the countries grazing these breeds.

Genome scan approaches, such as quantitative trait locus (QTL) and hitchhiking mapping, provide the opportunity to investigate the genetic regions associated with different traits. QTL mapping studies have been applied to examine the genetic basis of various economically important traits [e.g. 6–8]. However, a standard QTL mapping approach requires the segregating populations and to the association between the phenotype and genotype can be established [9]. Also, many traits underlying adaptive divergence are not always easily detectable at the phenotypic level and not well suited to QTL mapping. In this case, one approach to discover the genetic basis of different traits is to use hitchhiking mapping [10, 11]. Hitchhiking mapping is a population genetics approach to identify genomic regions influenced by selection [12, 13]. A great advantage of this approach is that it can be performed using molecular markers alone [14, 15]. This approach is especially suitable for the traits like fat tail where the phenotypic records in almost all countries grazing these breeds are sparse and their measurement is difficult.

Hitchhiking mapping starts with a genome scan using an approximately uniformly spaced set of molecular markers followed by a fine scale analysis of the candidate regions [10]. For example, marker based analysis of chromosome 1 in rats from warfarin-resistant populations revealed a 0.5 centimorgan (cM) region that was the likely locus for this gene [16]. This information was used later to identify the warfarin resistance gene through more traditional candidate gene approaches and association mapping techniques [17].

Genomic scan for finding candidate regions associated with fat deposition in thin and fat tailed breeds have been described previously [18]. In brief, two independent experiments, including Iranian and ovine HapMap genotyping data contrasting thin and fat tailed breeds were analyzed, and using different statistics, especially F_{ST} , three regions on chromosomes 5 (between 47,149 – 47,263 kb), 7 (between 46,642 – 46,843 kb) and X (between 58,621 – 61,452 kb) were confirmed in both data sets. Interestingly, these regions have also been confirmed in recently published studies through genome-wide analysis of transcriptomic data [19, 20] and selective sweeps [21, 22] in different thin and fat tailed sheep breeds.

Unlike F_{ST} that has been used in the previous study, tests based on linkage disequilibrium (LD) like iHS [23] and XP-EHH [24] are multi marker tests. These tests are commonly used for human SNP data, where there are now millions of SNPs available (approximately one marker per 1kb), and depend on SNP spacing and frequency [25]. Densely spaced SNPs give greater power when using statistical tests that rely on LD, as signals of selection are less likely to be lost [26]. The Illumina Ovine SNP50k BeadChip, while providing uniform genome wide coverage, has a marker about every 56kb [18]. Fine mapping, where more SNPs are genotyped in an area of interest, improves the ability to localize causal variants.

In this study, three regions previously recognized as candidate regions associated with fat deposition were chosen for further analysis. This was achieved through fine-mapping with additional SNPs, using Sequenom technology. The aim of this study was to confirm the signature of selection in the candidate loci, narrow down the chromosomal region showing the selective imprint and to determine the nature of the selection process occurred in these regions. Our analyses focused on two site frequency (F_{ST} and Median homozygosity) and haplotype based (iHS and XP-EHH) statistics and using these variety of selection sweep tests we were able to obtain some novel results associated with fat deposition in these breeds.

Results

SNP Genotyping and data mining

A total of 37, 30 and 31 SNPs passed the filtering criteria in our regions of interest on chromosome 5, 7 and X respectively. Most of these SNPs that could be used for further analysis were not selected for the Illumina Ovine SNP50k BeadChip due

to the lower expected minor allele frequency (MAF) and reliability; the validation rate for SNPs selected for the 50K BeadChip was 93%, whereas for these SNPs previous experience suggested validation reduced to around 50% in a separate population.

Table 1 presents a descriptive summary of the characteristics of the SNPs used in the final analysis. The distribution of SNPs varied among the regions, especially regarding to those used in statistical tests (Additional file 1: Table S1); however, the average SNP intervals were relatively consistent and the overall average distance between adjacent SNPs was about 21kb, 15kb and 80kb for candidate regions on chromosome 5, 7 and X respectively, compared to about 60kb and 115kb for autosomes and chromosome X on the Illumina Ovine SNP50k BeadChip respectively.

Table 1: Summary of SNP characteristics for different regions after data cleaning

Chromosome	No. of measured SNPs through Sequenom	No. of SNPs passed quality control	All used SNPs for final analysis	Average distance between SNPs (kb)
5	71	37	46 (15+31) ¹	21.05
7	62	30	34 (10+24)	15.61
X	57	31	38 (20+18)	80.67

¹SNPs included Ovine SNP50k Bead Chip + non repeated Sequenom SNPs

Distribution of F_{ST} and median homozygosity

To identify loci that have been targets of selection across thin and fat tailed breeds, the windowed F_{ST} was plotted against genomic location for candidate regions (Figure 1-a). Variants with unusually large F_{ST} values are typically interpreted as being the targets of local selective pressures due to the hitch-hiking effect [12].

The average of differentiation between Zel (thin tail) and Lori Bakhtiari (fat tail) for the whole genome was 0.024 (SD=0.036), while this parameter was 0.093 (SD=0.136), 0.172 (SD=0.184) and 0.279 (SD=0.217) at the areas of interest on chromosome 5, 7 and X respectively. As shown in figure 1, in these regions we found evidence of selection across relatively short distances with windowed F_{ST} values>0.30 on chromosomes 5 (between 47,149,400-47,245,841 bp), 7 (between 46,587,943-46,843,356 bp) and values>0.40 on chromosome X (between 59,257,971-59,984,949 bp). The score of 0.30 is in the 99.9 percentile of autosomal SNPs (n=44,558) and a score of 0.40 is in the 99.0 percentile of chromosome X SNPs (n=1126).

When there is a selection for a causal mutation in one breed and not the other, the breed under selection shows high homozygosity in a genomic interval while the other does not [27]. To further test this hypothesis, median homozygosity was calculated and plotted against genomic position for candidate regions (Figure 1-b). The results indicate that homozygosity increased over the areas of interest on chromosome 5 and X for the fat tailed and at the candidate region on chromosome 7 for the thin tailed breed. The largest differences of median homozygosity were for chromosome X and homozygosity was present for a longer distance as well, whereas these parameters are less on Chromosome 5.

Calculation of integrated haplotype score (iHS)

iHS detects signatures of strong selection in favor of alleles that have not yet reached fixation [23]. The results revealed that regions on chromosomes 5 and X had no iHS peak (Additional file 2: Figure S2 and Additional file 3: Figure S3). This observation may suggest that selected alleles have already been fixed in these locations. However, analysis of the region

on chromosome 7 displayed an obvious peak in both thin and fat tail breeds (Figure 2). For most locations of the region the values are higher in the thin tail breed, however, there is also some even stronger indication of selection in fat tail breed especially at the end of the region.

Cross-population EHH (XP-EHH)

To identify selective sweeps in which the selected allele has approached or achieved fixation in a subpopulation, but remains polymorphic in the population as a whole, the standardized cross population extent of haplotype homozygosity scores (XP-EHH) were plotted against genomic locations (Figure 3). The results revealed clear peaks in all of our regions of interest, suggesting that selected alleles approached fixation or have risen to near fixation in favour of fat tail breed on chromosome 5 and X while on chromosome 7 the frequency of alleles have risen close to fixation in thin tail breed. These results are based on linkage disequilibrium around given SNP and are in agreement with the results of median homozygosity plots.

Sabeti et al. 2007 [24] considered a region as a candidate for selection in the human HapMap Phase 2 dataset when the two population XP-EHH was above 4.34. This score is in the 99.9 percentile for thin versus fat tail breeds. As shown in figure 3, we found evidence of selection with |XP-EHH| value > 4.34 on chromosomes 5 (47,141,229-47,171,110 bp), 7 (46,604,500-46,642,359 bp) and X (59,187,456-60,264,325 bp). These regions are almost identical to the positions with highest F_{ST} .

Study of core haplotypes and their ancestral status in candidate regions

To evaluate the haplotype frequencies and the status of selected alleles (derived or ancestral), the core SNP alleles and their haplotype frequencies were investigated in the selected regions (Table 2).

Table 2: Core SNP alleles and haplotype frequencies in candidate regions for fat tail (Lori) and thin tail (Zel) breeds: The haplotypes with higher frequencies have been highlighted.

a) Chromosome 5

Core SNP alleles						Core haplotype frequencies	
SNPs	C/T	A/G	T/C	A/G	A/C	Fat tail breed	Thin tail breed
Genomic position (bp)	47149354	47149400	47165900	47171110	47175489		
Ancestral allele	T	G	T	G	C		
Haplotype 1	C	A	C	A	A	0.90	0.15
Haplotype 2	-	G	-	-	-	0.01	0.06
Haplotype 3	-	G	T	-	-	0.01	-
Haplotype 4	-	G	-	G	C	-	0.12
Haplotype 5	T	G	-	-	-	-	0.13
Haplotype 6	T	-	T	-	-	0.01	-
Haplotype 7	T	G	T	-	-	0.04	0.11
Haplotype 8	T	G	T	G	C	0.02	0.31
Other Haplotypes						-	0.14(6 Haplotypes)

b) Chromosome 7

Core SNP alleles					Core haplotype frequencies	
SNPs	C/T	A/C	T/C	A/C	Fat tail breed	Thin tail breed
Genomic position (bp)	46604500	46604644	46604722	46642359		
Ancestral allele	C	C	C	A		
Haplotype 1	C	C	C	C	0.06	0.80
Haplotype 2	-	-	-	A	0.20	0.14
Haplotype 3	T	A	-	-	0.02	0.03
Haplotype 4	T	A	-	A	0.14	-
Haplotype 5	T	A	T	-	0.31	0.02
Haplotype 6	T	A	T	A	0.27	-

c) Chromosome X

Core SNP alleles							Core haplotype frequencies	
SNPs	A/G	C/T	T/C	T/A	G/C	G/A	Fat tail breed	Thin tail breed
Position (bp)	59742181	59750338	59912586	59971891	59971909	59984949		
Ancestral allele	A	T	T	T	C	A		
Haplotype 1	G	C	T	T	G	G	0.89	0.12
Haplotype 2	-	-	C	-	C	A	-	0.01
Haplotype 3	-	-	C	A	C	A	-	0.18
Haplotype 4	A	T	-	A	C	A	-	0.04
Haplotype 5	A	T	C	-	-	-	0.11	0.08
Haplotype 6	A	T	C	-	C	A	-	0.01
Haplotype 7	A	T	C	A	C	A	-	0.54

For chromosome 5 (Table 2-a), we defined a core region of 26k where both F_{ST} and XP-EHH statistics were at their highest. There are 5 genotyped SNPs in this region. The SNPs defined 14 core haplotypes (denoted haplotype 1 to 14) in thin tailed but only 6 core haplotypes in fat tailed sheep. As shown in table 2, there is a common haplotype (haplotype 1) with a frequency of 90% in fat tailed sheep whereas its frequency for thin tailed is 15%. In contrast, the common haplotype in thin tail breed is haplotype 8 with a frequency of 31%, while it is almost absent (2%) in the fat tail breed. The interesting result in this region is that all the SNPs in the common haplotype for the fat tail breed are derived SNPs whereas all SNPs in the common haplotype for the thin tail breed are ancestral.

For chromosome 7 (Table 2-b) we defined a core region of 37k in the selected area with 4 genotyped SNPs, where all statistics showed strong evidence of selection. There were only 4 core haplotypes (denoted haplotype 1 to 4) in thin tailed and 6 haplotypes in fat tailed sheep. The common haplotypes of fat and thin tail breeds in this region were haplotype 5 and 1 with frequency of 31% (2% in thin tailed) and 80% (6% in fat tailed), respectively. Once again, while the common haplotype in the fat tail breed had all derived alleles the common haplotype in the thin tail breed were almost all ancestral alleles (3 out of 4 SNPs). However, it is notable that in this region the haplotypes appear more polymorphic in the fat tail breed and the frequency of the common haplotype of the thin tail breed is not as high as for the other candidate regions.

Subsequently, this approach was performed in the candidate region on chromosome X. We defined a core region of 242 k corresponding to 6 genotyped SNPs. The longer length of the core region, in this case is due to the wider SNP spacing on chromosome X. The genotyped SNPs defined seven core haplotypes in the thin tail breed and only two haplotypes in the fat tail breed (Table 2-c). The common haplotype in the fat tail breed was haplotype 1 with a frequency of 89%, whereas its frequency in thin tailed was 12 %. In this region the common haplotype in the thin tail breed (54%) was absent in the fat tail breed. The results in this region were similar to the previous regions in that the derived alleles were more prevalent

in common haplotype of fat tailed (4 out of 6 core SNPs) with the same result for ancestral alleles in the common haplotype of the thin tail breed. Together, these results could suggest that ancestral alleles have been under selection in the thin tail breed while the contrary happened for the fat tail breed where derived alleles have undergone selection.

Study of the identified genes associated with fat metabolisms in candidate regions

The regions of interest were investigated to determine if any genes related to fat deposition could be identified in sheep or their corresponding areas of the cow genome. The genes obtained in *O.aries* or by orthology with *B. taurus* and their functions are presented in table 3.

Table 3: Fat metabolism related genes located within the candidate regions in *O. aries* and their orthologous area in *B.taurus*

Species	Chromosome	RefSeq Number	Gene Name	Gene Symbol	Function
<i>Ovis aries</i>	X	NM_001037811	hydroxysteroid (17-beta) dehydrogenase 10	<i>HSD17B10</i>	lipid metabolic process
	X	NM_000044	androgen receptor	<i>Ar</i>	lipid binding
	X	NM_173963	synaptophysin	<i>Syp</i>	lipid binding (Cholestrol binding)
<i>Bos taurus</i>	5	NM_002715	protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform	<i>PPP2CA</i>	cellular lipid metabolic process
					lipid metabolic process
					membrane lipid metabolic process
					sphingolipid metabolic process
	X	NM_001034500	emopamil binding protein (sterol isomerase)	<i>EBP</i>	lipid metabolic process

There is also the hypothesis that some genes selected for, in fat tail sheep breeds are likely to be also associated with developmental defects or ectopic expression of organs [29, 30]. Our results revealed that these regions contain many

genes, having some known biological functions associated with the developmental process in *O.aries* (Table 4) and their orthologous coordinates in *B. taurus* (Additional file 4: Table S).

Table 4: Developmental process or gene expression related genes, located within the candidate regions in *O. aries*

Chr.	Gene symbol	Gene Name	Functions*
------	-------------	-----------	------------

5	<i>TCF7</i>	Transcription factor 7 (T-cell specific, HMG-box)	regulation of gene expression
7	<i>PTGDR</i>	Prostaglandin D2 receptor (DP)	developmental process
	<i>NID2</i>	Nidogen 2 (osteonidogen)	cellular macromolecule metabolic process
X	<i>AR</i>	Androgen receptor	gland development, organ development, system development, anatomical structure development, regulation of gene expression
	<i>FOXP3</i>	Forkhead box P3	organ development, system development, anatomical structure development, regulation of developmental process, regulation of gene expression
	<i>FGD1</i>	FYVE, RhoGEF and PH domain containing 1	organ development, system development, anatomical structure development, regulation of developmental process
	<i>BMP15</i>	Bone morphogenetic proteins 15	
	<i>HSD17B10</i>	Prostaglandin D2 receptor (DP)	organ development, system development, anatomical structure development
	<i>TIMP1</i>	TIMP metalloproteinase inhibitor 1	organ development, system development, anatomical structure development
	<i>PFKFB1</i>	Hydroxysteroid (17-beta) dehydrogenase 10	organ development, system development, anatomical structure development
	<i>SLC35A2</i>	Solute Carrier Family 35 Member A2	organ development, system development, anatomical structure development
	<i>ALAS2</i>	Aminolevulinatase, delta-, synthase 2	organ development, system development, anatomical structure development
	<i>HEPH</i>	Hephaestin	organ development, system development, anatomical structure development
	<i>PCSK1N</i>	Proprotein convertase subtilisin/kexin type 1 inhibitor	organ development, system development, anatomical structure development
	<i>SHROOM4</i>	Shroom family member 4	organ development, system development, anatomical structure development
	<i>CACNA1F</i>	Calcium channel, voltage-dependent, L type, alpha 1F subunit	system development, anatomical structure development
	<i>TFE3</i>	Transcription factor binding to IGDM enhancer 3	regulation of developmental process, regulation of gene expression
	<i>ELK1</i>	ELK1, member of ETS oncogene family	regulation of gene expression
	<i>KDM5C</i>	Lysine (K)-specific demethylase 5C	regulation of gene expression
	<i>ZNF41, 81</i>	Zinc finger protein 41, 81	regulation of gene expression

***gland development:** The process whose specific outcome is the progression of a gland over time, from its formation to the mature structure. A gland is an organ specialized for secretion.

organ development: Development of a tissue or tissues that work together to perform a specific function or functions. Development pertains to the process whose specific outcome is the progression of a structure over time, from its

formation to the mature structure. Organs are commonly observed as visibly distinct structures, but may also exist as loosely associated clusters of cells that work together to perform a specific function or functions.

system development: The process whose specific outcome is the progression of an organismal system over time, from its formation to the mature structure. A system is a regularly interacting or interdependent group of organs or tissues that work together to carry out a given biological process.

anatomical structure development: The biological process whose specific outcome is the progression of an anatomical structure from an initial condition to its mature state. This process begins with the formation of the structure and ends with the mature structure, whatever form that may be including its natural destruction. An anatomical structure is any biological entity that occupies space and is distinguished from its surroundings. Anatomical structures can be macroscopic such as a carpel, or microscopic such as an acrosome.

developmental process: A biological process whose specific outcome is the progression of an integrated living unit: an anatomical structure (which may be a subcellular structure, cell, tissue, or organ), or organism over time from an initial condition to a later condition.

regulation of developmental process: Any process that modulates the frequency, rate or extent of development, the biological process whose specific outcome is the progression of a multicellular organism over time from an initial condition (e.g. a zygote, or a young adult) to a later condition (e.g. a multicellular animal or an aged adult).

regulation of gene expression: Any process that modulates the frequency, rate or extent of gene expression. Gene expression is the process in which a gene's coding sequence is converted into a mature gene product or products (proteins or RNA). This includes the production of an RNA transcript as well as any processing to produce a mature RNA product or an mRNA (for protein-coding genes) and the translation of that mRNA into protein. Some protein processing events may be included when they are required to form an active form of a product from an inactive precursor form.

Discussion

Directional selection can leave a set of signatures in the genes under its influence, such as the rapid divergence of functional sites among populations, the depression of polymorphism within populations and a large amount of linkage disequilibrium (LD) existing over an extended region harboring the selected allele [1, 21, 30]. To animal breeders, these patterns could be useful in the process of identifying genomic regions or genes underlying traits of interest, since locations of selection in livestock populations should be correlated with QTL affecting production traits [5, 14]. The “fat tail” is a phenotype that divides domesticated sheep into two major groups. In the past, this trait has been desirable for both associated requirements of humans and as a store of energy for animals during feed scarcity that led to selection for higher fat tail weight [31]. Although, in recent times due to the improved forage availability and decreased price for the product, most of the advantages of a large fat tail have reduced importance in genetic improvement programs [3, 32]. An increasing number of studies have been conducted to detect signals of recent positive selection on a genome-wide scale in different domestic animals [e.g. 34–37]; however, there are relatively few genomic regions identified that have been subject to selection for a specific mutations underlying evolutionary shifts in a trait. In this paper, in order to fine map the genomic regions potentially affecting the fat deposition in thin and fat tail breeds and get more insight into the genomic basis of that, we have investigated three candidate regions using this approach.

These candidate regions were analyzed using a variety of statistics to clarify the signals of selection observed in these regions. Our analyses focused on two allele frequency spectrum (F_{ST} and median homozygosity) and haplotype based (|iHS| and XP-EHH) statistics. These tests were chosen because previous power analyses suggested that these are largely

complementary [27]. F_{ST} and median homozygosity tests detect highly differentiated alleles between populations, where positive selection in one area causes larger frequency differences compared to neutrally evolving alleles. Akey et al. 2002 [37] suggested using the loci in the tails of the empirical distribution of F_{ST} as candidate targets of selection. iHS, a measure of within breed evidence for selection, has good power to detect partial selective sweeps. Voight et al. 2006 [23] demonstrated that the iHS method was more powerful than Tajima's D [38] or Fay and Wu's H [39] tests. A disadvantage of this approach is that it loses power when the beneficial allele is close to fixation, because fixation will eliminate variation at and near the selected site. To overcome this problem, we next applied XP-EHH, to detect selected alleles that have risen to near fixation in one but not another population, as for this purpose it is more powerful than iHS test. iHS and XP-EHH tests have been combined to search for recent positive selection in humans [24, 40, 41], and other species such as *Plasmodium falciparum* [42]. Grossman et al. 2010 [43] showed that the F_{ST} and XP-EHH signals peaked more narrowly around the causal variant, making them useful for spatial localization, while iHS is better to distinguish causal variants and contributed little to spatial resolution. These tests were relatively uncorrelated in neutral regions, and only weakly correlated for neutral variants within selected regions [43].

Study of candidate regions on chromosome 5 and X revealed obvious evidence of selection using F_{ST} , median homozygosity and XP-EHH test in a relatively narrow region; while an examination of these regions identified no particular |iHS| peak. With a hypothesis that historically different selection pressures operated in thin and fat tail breeds and somehow selection acted on a variant that was advantageous only in one breed, these results suggest that selection in these regions occurred for mutations affecting fat tail size as the beneficial mutations have risen to near fixation in fat tailed breeds. This suggestion is also supported by the core haplotype frequencies observed in candidate regions on these chromosomes as the common core haplotypes in fat tail breeds were near fixation (Table 2).

Analysis of the region on chromosome 7 indicated strong evidence of selection using all selective sweep statistics. The results of F_{ST} and median homozygosity suggested that the selection differentiated the beneficial alleles between breeds and that homozygosity has been increased in favor of thin tailed in this region. However, |iHS| and XP-EHH revealed additional information. |iHS| provided evidence of partial selection in both breeds, while the XP-EHH results showed that the selected alleles have approached fixation in the thin tail breed.

As discussed earlier, the power of the |iHS| statistic to detect selective sweeps is greatest at a moderate allele frequency (~40-60%), while XP-EHH test is more powerful for detecting selective sweeps close to fixation (>80%) [23, 40]. However, both of these methods do have power outside their optimal ranges. Sabeti et al. 2007 [24] demonstrated that the iHS statistic could detect signals over the range of 20-80% and XP-EHH do have power between 60-100%. Overall, these results suggest that while the frequency of selected alleles has been raised to fixation in thin tail breed, its frequency should be ~ 60-80% to be picked up by both methods. Simultaneously, the favorable allele should be increased to mid-range frequency in the fat tail breed. The results of core haplotype frequencies in this region (Table 2-b) are in consistent to this point as the common haplotype in thin tail breed has a frequency of 80%, whereas all haplotypes observed in this region appears polymorphic in the fat tail breed. One inference is that there has been ongoing infusion of fat tailed haplotypes into this breed, but also selection for the thin-tailed phenotype.

To further test this hypothesis, we constructed the pattern of haplotype blocks in this region and the decay of LD (pairwise r^2) was visualized using Haploview [44]. The effect of a selective sweep on patterns of variation is expected to decline with time (due to recombination) and if a selective sweep is still ongoing in a subpopulation, the hitchhiking haplotype is expected to be rather long [27, 30]. Our results (Additional file 5: Figure S5) revealed that although there were haplotype blocks in both breeds, they extended for longer distances in the fat tailed compare to the thin tail breed. This suggests that as the prevalence of selected alleles increased in thin tail breed, the LD around variants decayed due to recombination, while the selection in the fat tail breed is younger (longer haplotype blocks). These results confirm our observations for evidence of selection in this region with both |iHS| and XP-EHH tests.

The earliest known depiction of a fat tail sheep is on an Uruk III stone vessel about 5000 years before present, approximately 4000 years after initial domestication [45]. Given that fat tailed breeds are now prevalent in the Fertile Crescent, where sheep were originally domesticated, while thin tailed sheep breeds are predominant in peripheral areas and that the wild ancestor of sheep is thin tail, it has been assumed that the first domesticated sheep were thin tailed and fat tail was developed later [45, 46]. We have investigated this hypothesis through the classification of selected alleles in core haplotypes of our regions of interest as ancestral or derived. Our results provide the preliminary molecular evidence to confirm this assumption since, we observed that in almost all cases, derived alleles have been under selection pressure in the fat tail breed, and this is consistent with the selection of a new mutation in these breeds (Table 2).

Population demographic history can also cause similar patterns on DNA sequence variation and could be a source of error in making inferences on genomic targets of selection. This caveat can be avoided by screening a large number of markers (as has previously been performed in this research) spaced across whole genome [18], as selection will result in regional patterns compared to the genome-wide effects of population history and demographic events [10, 47]. Similarly conducting this type of fine-scale analysis at the candidate genomic regions using a dense set of markers and multiple statistical tests, reduces the chance that a signature of selection will be a false positive if it is detected in more than one marker locus and statistical test [43].

The candidate regions of interest have previously been studied using OAR v1.0 in *O. aries* and their corresponding area of *B. taurus* and no particular candidate genes associated with fat deposition were identified [18]. In this study, reinvestigation of these regions using the newly available sheep genome OAR v3.1 [48], defined some genes associated with fat metabolism in *O. aries* or their orthologous areas of *B. taurus* (Table 3). **Protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform** (*PPP2CA*) has a variety of roles in different biological process such as cellular lipid metabolic, membrane lipid metabolic and **sphingolipid metabolic process**, while hydroxysteroid (17-beta) dehydrogenase 10 (*HSD17B10*) and emopamil binding protein (*EBP*) genes play some roles in **lipid metabolic process** and androgen receptor (*Ar*) and **synaptophysin** (*Syp*) genes get participate in lipid binding [49]. Interestingly, most of the genes identified here, have been recently reported as candidate genes associated with lipid metabolisms using various molecular techniques in sheep. Yuan et al. 2019 [28] implemented differential expression analysis using RNA-seq technology in longissimus dorsi muscle tissue (MUT), perirenal adipose tissue (PAT) and tail adipose tissue (TAT) of different Chinese short and fat tailed sheep breeds and revealed that *PPP1CA* is highly expressed in TAT. Also, *PPP1CA* was identified as plausible genes associated with the fat-tailed or fat-rumped phenotype by comparing copy number variations (CNVs) with different tail types [50]. Moreover, protein phosphatase 1, catalytic subunit, gamma isozyme (*PPP1CC*) have been under selection signature in Chinese thin and fat-tailed sheep breed and reported to be associated with tail type [51]. Kang et al. 2017 [19] profiled transcriptomes from Tan sheep, a Chinese indigenous breed with notable fat tail, using RNA-seq to unravel the potential underlying mechanisms governing accumulation of adipose tissues in various regions of the body including subcutaneous, visceral and tail. Their results exhibited marked changes in expression among the three adipose depots for *AR* gene. Hormone receptor density differences in the adipose tissues influence regional fat distribution, for example, *AR* contributes to the control of adipocyte development by interacting with its own ligand androgen [19]. *HSD17B12* has been also reported to be associated with fat tail metabolism in thin and fat tailed sheep breeds using deep transcriptome analysis obtained by RNA-Seq data [52]. It is reported that *HSD17B12* (act as elongates) are important genes for controlling the overall balance of fatty acid composition [52].

To confirm the results of our study, the exon 1 of *PPP2CA* gene was amplified and its variation patterns were sequenced in an independent study on Zel and Lori-Bakhtiari sheep breeds [53]. Two patterns were identified and the results of sequencing showed that in Lori-Bakhtiari, Del/Del genotype resulted in heavier fat tail than T/T genotype (5.20 ± 0.21 kg vs 3.28 ± 0.12 kg) ($P < 0.05$) while, in Zel, the effect of genotypes on carcass fat percentage and triglyceride was significant, so that the T/T genotype had more carcass fat percentage comparing to Del/Del genotypes ($P < 0.05$). Overall, it seems as the

annotation of the ovine genome becomes more complete, all genes located in the candidate regions will be identified and promising targets can then be verified by further experimentation.

A result which is irrelevant to the inheritance of the trait, but provides an insight into a possible mechanism of fat deposition in this organ, are the results of Gökdal et al. [29] who examined the effects of docking in fat tail breeds. The carcasses of the docked group contained more kidney, pelvic and internal fat than the intact lambs as well as a higher percentage of subcutaneous and intramuscular fat. The weights of the different carcass cuts of the docked lambs were also heavier than those of the intact group. However, there was little change in overall carcass composition, suggesting that the genes affecting the fat tail phenotype are associated with the localization of fat stores to a regional depot rather than control of the overall level of fat deposition. This observation also may provide support to the suggestion that some genes selected for in fat tail sheep breeds in these regions are likely to be also associated with developmental defects or ectopic expression of organs. Our results revealed that these regions contain many genes, having some known biological functions associated with developmental process (Table 4 and Additional file 4: Table S4). Several earlier studies provide evidence for this issue. For example, the transcription factor (*TCF*) genes has been among highly differentially expressed genes in perirenal adipose tissue (PAT) and identified as being the most likely to account for the fat-tailed phenotype of sheep [28]. *TCF7* is involved in the Wnt/ β -catenin signaling pathway, and this pathway plays a critical role in regulating sheep [52] and porcine [54] adipogenesis genes expression. Zhu et al. 2019 [55] studied copy number variations (CNVs) and selection signatures on the X chromosome of Chinese indigenous sheep with different tail types and revealed that the regions harboring CNVs and selective sweeps in different sheep breeds overlapped with [calcium channel, voltage-dependent, L type, alpha 1F subunit](#) (*CACNA1F*) gene that could be as associated with tail type in these breeds. In addition, *HSD15B* [52], *SLC35A2* [51], *AR* and *TIMP1* [19], have been identified as candidate genes that affect fat tail development. Moreover, it is important to note that our regions of interest overlapped with some genes, for example *BMP15*, *WDR13* and *RBM3* that belong to the gene families that their closely related genes consisting *BMP2* [21, 51, 56–58], *WDR92* [20, 51] and *RBM11* [5] have recently reported to be associated with fat tail formation and adipose tissue gene expression in sheep.

Finally, fine mapping of candidate regions using different sweep statistical tests has enabled us to confirm the signature of selection in these chromosomal regions and better refine the critical regions from 113 kb (47,149,400-47,263,230) to 28 kb (47,146,931- 47,175,489) on chromosome 5, from 201 kb (46,642,359-46,843,356) to 142 kb (consisting to shorter intervals: 46,587,943-46,642,359 and 46,765,080-46,852,870) on chromosome 7 and from 2,831 kb (58,621,412-61,452,816) to 1,006 kb (59,257,971-60,264,325) on chromosome X. These regions were refined considering all statistical test results (Additional file 6: Figure 6). Acquiring of the genes located within the regions of interest after fine mapping revealed some genes consisting *TCF7* on chromosome 5, *PTGDR* and *NID2* on chromosome 7 and finally *AR* on chromosome X that have a multiple effect on lipid metabolisms, macromolecule metabolic process, organ/gland development or associated with ectopic expression of organs simultaneously.

Recently published study on the origin of European sheep as revealed by the diversity of the Balkan breeds and by optimizing population-genetic analysis tools [59] using a variety of sheep breed samples from Southwest-Asian, Mediterranean, Central-European and North-European showed that the thin-tailed Zel sheep is found to be in the same genetic cluster as the fat-tailed Iranian sheep, whereas the fat-tailed Italian Laticauda is related to other breeds in central Italy. This implies that the tail phenotype is encoded by a limited number of genes. By combining information of the present study, previously reported and annotated biological functional genes, we suggest *PPP2CA* and *TCF7* (*OAR5*), *PTGDR* and *NID2* (*OAR7*), *AR*, *EBP*, *CACNA1F*, *HSD15B*, *SLC35A2*, *BMP15*, *WDR13* and *RBM3* (*OAR X*) as the most promising candidate genes for type of tail traits. It is obvious that understanding the mechanisms that underlie fat tail inheritance in sheep is difficult to verify solely by selective sweep profiling. Therefore, further work based on the results from this study are required to uncover the exact genetic mechanisms of fat deposition in the tail of sheep. Also, these regions may still be too large to efficiently implement technologies such as marker assisted selection or positional

cloning. More detailed and larger scale experiments from these and other thin and fat tailed breeds may allow us to refine the location of the causal mutations. Likewise, it does not exclude contemporaneous selection for other traits, and any regions identified still need to be tested for a functional genetic relationship via trait measurement in contrasting genotypes or phenotypes. Specifically, future studies should be conducted in reciprocal F_2 crosses to provide independent and causal evidence and verify the mode of inheritance. If these areas are shown to have a significant effect, then further work sequencing either side of the regions can help the search for causal variants.

Conclusion

Our results provide a comprehensive assessment of how and where selection has affected the patterns of variation in candidate regions associated with fat deposition in thin and fat tail sheep breeds. These results enabled us to confirm the signature of selection in these regions, refine the critical intervals regions, and to identify the most promising candidate genes associated with fat deposition in thin and fat tail sheep. These results paved a solid foundation for studying the regulation of fat deposition in sheep and do offer hope that the causal mutations and the mode of inheritance of this trait will soon be discovered by further experimentation.

Methods

Animal sampling and genotyping

The animal samples and genotyping with the Illumina Ovine SNP50k bead chip have been described previously [18]. Briefly, two independent data sets consisting of Zel-Lori Bakhtiari and Ovine HapMap samples were used for genome scan analysis of selective sweeps in thin and fat tail breeds. These samples were genotyped for approximately 5×10^4 SNPs derived from the [International Sheep Genomics Consortium](#) (ISGC). Because, the latter project did not formally phenotype the individuals concerned and we did not have access to their DNA samples, in this research further analysis was performed for the Zel-Lori Bakhtiari data. This data set consisted of 45 samples each of the thin (Zel) and fat tailed (Lori Bakhtiari) breeds (37 Females and 8 Males per each breed). The animals were collected to be as unrelated as possible. Three regions (Table 5), identified as being under selection based on F_{ST} and median homozygosity results from the first genome scan, were chosen for further analysis.

Once the regions of interest were defined, all known SNPs in each region was examined for suitability for genotyping. These included SNPs discovered on both the Solexa and 454 platforms (<http://www.sheephapmap.org/genseq.php>). A total of 156, 89 and 78 putative SNPs flanked by 150 bp on each side (301 bp total) were submitted to Sequenom's primer design software (Mass ARRAY Assay Design 3.1) for the regions on chromosomes 5, 7 and X, respectively and primers and probes were designed in multiplex format. Of these SNPs, assays were successfully designed for 140 (90%), 75 (84%) and 66 (84%) of all SNPs initially selected for study, and were grouped into 6, 3 and 4 multiplex assays for the three regions respectively. The remaining SNPs failed primer design, primarily due to high repeated content. The majority of these multiplex assays were located in the first and second plex which contained 71, 62 and 57 SNPs for chromosomes 5, 7 and X respectively. These two SNP assays were used for genotyping. Selected SNPs, their locations, PCR primers, unextended primer and its extension masses are represented in additional table S7. All SNPs were genotyped by use of the mass-spectrometry based iPLEX Mass Array platform provided by Sequenom (Sequenom, San Diego, California, USA. <http://www.sequenom.com/>). Experiments were conducted following the Sequenom iPLEX Assay application note [60].

Table 5: Regions chosen for fine mapping in this study

Chromosome	Region (bp)*	Length (bp)	Ovine SNP50k BeadChip SNPs
5	46971979- 47919440	947461	15
7	46392398- 46852870	460472	10
X	58424602- 61409447	2984845	20

* These regions cover the areas under selection based on the first genome scan [18]

Quality control

filters

All samples with more than 30% missing data and subsequently all loci with more than 15% missing data were excluded. These rejection thresholds were chosen by plotting numbers of animals or loci against percent missing data and the cutoff point was determined as the curve inflection point where the rate of change in the number of excluded loci became linear with every increased percent of missing data [61]. For the remaining SNPs those with MAF less than 2% and outlier departure from Hardy-Weinberg equilibrium over all animals of a breed ($P\text{-value} < 10^{-6}$) were excluded as described previously [18, 55]. We combined the SNP genotyping results from both the Ovine SNP50k Bead Chip (in the regions of interest) and the Sequenom Mass ARRAY platform for final analysis. For the few SNPs that had been genotyped on both platforms, the Ovine SNP50k Bead Chip results were used due to their greater accuracy.

Estimates of F_{ST} and median homozygosity

To determine the pattern of positive selection, the basic form of Wright's fixation index (F_{ST}) was calculated as described by [62]. The value of F_{ST} can theoretically range from 0 (showing no differentiation) to 1 (indicating complete differentiation, i.e. populations are fixed for different alleles). For each set of 5 adjacent SNPs, the average of F_{ST} values was calculated and termed windowed F_{ST} . This is an approximate method of looking for regions where selection is apparent over multiple markers, rather than one-off high values. A window of 5 markers was chosen as it appeared to provide the better signal compare to other arbitrary window sizes [18]. The windowed F_{ST} values were then plotted against genome location. All analyses presented in this work were also performed using the Weir and Cockerham method [63]. The results were almost identical for both methods, so that, we have only presented results based on F_{ST} in the present study. All scripts for calculating Wright (F_{ST}) and Weir and Cockerham fixation indices were written and performed in R v 4.0.2.

Selection for a new beneficial mutation increases the level of homozygosity in the selected allele and neighboring regions due to the hitchhiking effect. Therefore, one method of looking for the region where selection has taken place is to compare runs of homozygosity between breeds. To do this, the median run of homozygosity for each SNP was calculated following [18] and the values plotted against the genome. It should be noted that since the historical effective population sizes of males and females are not the same for sexual and autosomal chromosomes [64], all analyses presented in this paper for chromosome X were performed by using only females (36 animals per each breed), although analysis with both sexes produced similar results.

Determining of ancestral alleles

To calculate iHS and XP-EHH, the ancestral allele state of each SNP must be specified. Ancestral alleles for the ovine chip SNP were obtained from international sheep hapmap project (ISGC), and then to obtain the ancestral alleles for the additional Sequenom SNPs, 301 base pairs of sequence (1 bp of the alleles plus 150 bases either side of the SNP) were

aligned against *Bos taurus* (cattle), *Sus scrofa* (pig), *Equus caballus* (horse), *Canis familiaris* (dog) and *Homo sapiens* (human) genomes using BLAST (<http://blast.ncbi.nlm.nih.gov/Blast>). The ancestral allele was taken as the base in the genome sequence at the resulting SNP position. For loci where only one SNP allele was represented in the other species, that allele was determined as ancestral. For other SNPs, as an additional tool in determining the ancestral allele, a phylogenetic tree of the 5 species used was constructed. This provided a crude tree, which could be used in decision making; for example, if all the animals had the same allele (C) apart from humans (T), then the ancestral was more likely to be C, as humans are more distantly related than the other species. In this research, we were able to determine the ancestral status of 36, 29 and 38 from the 46, 35 and 38 SNPs which passed quality control in the regions on chromosome 5, 7 and X respectively.

Reconstruction of haplotypes

A pair of haplotypes was reconstructed for each animal in the sample using fastPHASE version 1.2.3 [65]. This software implements an Expectation-Maximization strategy for estimating missing genotypes and for reconstructing haplotypes from unphased SNP genotypes data of unrelated individuals [65].

Calculation of integrated haplotype score (iHS) and cross-population EHH (XP-EHH)

iHS was calculated as in Voight et al. [23] and XP-EHH as in Sabeti et al. [24]. Scripts for computing these tests were obtained from the Pritchard lab website (<http://hgdp.uchicago.edu/Software/>).

Briefly and following Voight et al [23], the iHS was computed for every SNP with ancestral state information and MAF above 5% (Additional file 1: Table S1). This test is based on the extended haplotype homozygosity (EHH) statistic [27], which measures the decay of identity, as a function of distance, of haplotypes that carry a specified core allele at one end. The integral of the observed decay of EHH with distance from the core allele is calculated until EHH reaches 0.05. This integrated EHH (iHH) (summed over both directions away from the core SNP) is denoted as iHH_A or iHH_D , depending on whether it is computed for the ancestral or derived core allele. The unstandardized iHS (uiHS) was then calculated as $\ln(iHH_A/iHH_D)$. The uiHS is thus adjusted so that the final statistic has a mean of 0 and a variance of 1, regardless of allele frequency at the core SNP. To do this, the results for each breed were split into 18 equally sized allele frequency bins, from which a mean and standard deviation were calculated. These were used in the following equation [23]: **see equation in the supplementary files section.**

Where the expectation and standard deviation are calculated using SNPs from the same bin (p). Due to the different demographic histories of the X chromosome and the autosomes (e.g., due to smaller effective population size), we normalized the iHS scores of this chromosome separately from those of the other chromosomes. Results of iHS are presented here as |iHS| for a window of 10 SNPs and then plotted against genome location. We chose a window of 10 SNPs because of the longer extent of LD and SNP spacing in sheep compared to humans, in which the window length used is commonly around 40 SNPs [23]. Large positive and negative values of iHS indicate unusually long haplotypes carrying the ancestral and derived allele, respectively; by taking the absolute iHS value, interesting variants will be shown by large positive values.

While iHS detects partial selective sweeps, XP-EHH detects selective sweeps in which the selected allele has risen to high frequency in one population, but remains polymorphic in the population as a whole. To do this, XP-EHH compares haplotypes between populations to control for local variation in recombination rates [24]. Briefly, XP-EHH is defined relative to a given SNP i in two populations, A and B. In each population, the expected haplotype homozygosity

(EHH) [27] was integrated with respect to genetic distance in both directions from i . The log of the ratio of these integrals, $\ln(I_A/I_B)$, is the unnormalized XP-EHH (for more details see [24]). XP-EHH must also be normalized for genome-wide differences in haplotype length between populations, so that there is a mean of zero and unit variance. This was done by subtracting the mean and dividing by the standard deviation of all scores. In this study, we defined the XP-EHH test with respect to thin and fat tail breeds and an unusually positive value suggests selection in fat tailed population and a negative value selection in thin tailed.

Core SNP alleles and haplotype frequencies in candidate regions

After the aforementioned filtering process and reconstruction of haplotypes for candidate regions using PHASE 2.1 [66], the haplotypes were fed into SWEEP v.1.1 [27] to detect core regions based on the EHH statistic in candidate regions, which is fully described by Sabeti et al. 2002 [27]. PHASE was chosen here over fastPHASE as haplotype estimates are slightly more accurate using PHASE (<http://stephenslab.uchicago.edu/software.html>), although substantially slower to compute. For selection of core regions and study of haplotype frequencies in selected area of our interested regions, the results of F_{ST} , iHS and XP-EHH tests were also used as additional information.

Study of identified genes in candidate regions

Genes located in the genomic regions significantly differentiated between sheep breeds were acquired by the use of the data mining tool Biomart (<http://asia.ensembl.org/biomart/martview>), with the latest reference assembly of the *O. aries* genome OAR v3.1 [48]. The regions of interest in *O. aries* were also compared to the corresponding areas in *B. taurus* as its genome is better annotated. Regions chosen for fine mapping onto OARv3.1 and their orthologous coordinates in *B. taurus* is shown in additional table S8. It should be noted that, due to the easier comparison of coordinates obtained by the current study with the previous article [18], the coordinates have been presented for different statistics in this study based on OAR v1.0. However, the coordinate of candidate regions associated with fat deposition reported in article [18] and the fine mapped results in this study have been also shown based on different OAR versions in additional table S9. Nearby genes within a flanking distance of 500 kbs from each region were acquired. This distance has been selected as previously considered by Moradi et al. 2012 [18] for sheep and by Do et al. 2017 [67] for Holstein dairy cattle. The Biological Process (BP), Molecular Functions (MF) and KEGG pathway of the entire identified genes in candidate regions were determined using DAVID [68]. Furthermore, a comprehensive literature review was conducted to verify whether these genes have some relevance with fat deposition or developmental process in sheep or other mammals.

Abbreviations

QTL: Quantitative Trait Loci, SNP: Single Nucleotide Polymorphism, LD: Linkage Disequilibrium, F_{ST} : Population differentiation index; MAF: Minor Allele Frequency, bp: Base Pair, HWE: Hardy-Weinberg Equilibrium; iHS: integrated Haplotype Score, XP-EHH: cross population Extent of Haplotype Homozygosity.

Declarations

Authors' contributions

MHM planned and performed the analyses, and drafted the manuscript, ANJ coordinated the study and supervised the analysis, MMS coordinated the study and sample collection, KGD and RB provided statistical and analysis support and JCM supervised the analysis and participated in the design of the study. All authors have contributed to the editing of the article, and approved the final manuscript.

Funding

This study was supported by Animal Science Research Institute of Iran, 2016/11547 research grants from Iran's Ministry of Agriculture-Jahad, Mobarakandish Institute and AgResearch, New Zealand. The funding organizations played no role in the design of the study, data collection, analysis, and interpretation or writing the manuscript.

Availability of data and materials

All data generated or analyzed during this study are included in this manuscript and its supplementary information files. In addition, more detail data are available from the corresponding author on reasonable request

Ethics approval and consent to participate

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

Consent for publication

Not applicable.

Competing interests

The authors declare they have no competing interests.

Acknowledgements

The authors gratefully acknowledge the International Sheep Genomics Consortium for access to the Ovine HapMap genotypes and the Animal Breeding Center of Iran (ABCI) for access to the records and animals of the Iranian breeds. Thanks to the staff of the University of Tehran, Animal Science Research Institute of Iran and AgResearch, especially Mr. Tim Manley, Ms. Rayna Anderson and Ms. Kathryn McRae who helped and supported this research. The authors also acknowledge the financial contributions of Animal Science Research Institute of Iran, Mobarakandish Institute and AgResearch, New Zealand. We thank Dr. Pardis Sabeti and Dr. Benjamin Voight for their invaluable advises during this research.

References

1. 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.
2. Fariello M-I, Servin B, Tosser-Klopp G, Rupp R, Moreno C, Consortium ISG, et al. Selection signatures in worldwide sheep populations. *PLoS One*. 2014;9:e103813.
3. Khaldari M, Azarfar A, Pahlavan R. The size of fat tail does not have an effect on growth performance and carcass characteristics in Lori-Bakhtiari lambs. *Small Ruminant Research*. 2020;187:106088.
4. Negussie E, Rottmann OJ, Pirchner F, Rege JEO. Patterns of growth and partitioning of fat depots in tropical fat-tailed Menz and Horro sheep breeds. *Meat Science*. 2003;64:491–8.
5. Zhang T, Gao H, Sahana G, Zan Y, Fan H, Liu J, et al. Genome-wide association studies revealed candidate genes for tail fat deposition and body size in the Hulun Buir sheep. *Journal of Animal Breeding and Genetics*. 2019;136:362–70.
6. MacNeil MD, Grosz MD. Genome-wide scans for QTL affecting carcass traits in Hereford× composite double backcross populations. *Journal of Animal Science*. 2002;80:2316–24.

7. Schrooten C, Bink M, Bovenhuis H. Whole genome scan to detect chromosomal regions affecting multiple traits in dairy cattle. *Journal of Dairy Science*. 2004;87:3550–60.
8. McClure MC, Morsci NS, Schnabel RD, Kim JW, Yao P, Rolf MM, et al. A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. *Animal Genetics*. 2010;41:597–607.
9. Mackay TFC. The genetic architecture of quantitative traits. *Annual Review of Genetics*. 2001;35:303–39.
10. Makinen HS, Shikano T, Cano JM, Merila J. Hitchhiking mapping reveals a candidate genomic region for natural selection in three-spined stickleback chromosome VIII. *Genetics*. 2008;178:453–65.
11. Moioli B, Scatà MC, Steri R, Napolitano F, Catillo G. Signatures of selection identify loci associated with milk yield in sheep. *BMC Genetics*. 2013;14:1–7.
12. Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. *Genetics Research*. 1974;23:23–35.
13. Moon S, Kim T-H, Lee K-T, Kwak W, Lee T, Lee S-W, et al. A genome-wide scan for signatures of directional selection in domesticated pigs. *BMC Genomics*. 2015;16:1–12.
14. Akey JM. Constructing genomic maps of positive selection in humans: where do we go from here? *Genome Research*. 2009;19:711–22.
15. Shikano T, Ramadevi J, Merilä J. Identification of local-and habitat-dependent selection: scanning functionally important genes in nine-spined sticklebacks (*Pungitius pungitius*). *Molecular Biology and Evolution*. 2010;27:2775–89.
16. Kohn MH, Pelz H-J, Wayne RK. Natural selection mapping of the warfarin-resistance gene. *Proceedings of the National Academy of Sciences*. 2000;97:7911–5.
17. Rost S, Fregin A, Ivaskevicius V, Conzelmann E, Hörtnagel K, Pelz H-J, et al. Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2. *Nature*. 2004;427:537–41.
18. Moradi MH, Nejati-Javaremi A, Moradi-Shahrabak M, Dodds KG, McEwan JC. Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC Genetics*. 2012;e13.
19. Kang D, Zhou G, Zhou S, Zeng J, Wang X, Jiang Y, et al. Comparative transcriptome analysis reveals potentially novel roles of Homeobox genes in adipose deposition in fat-tailed sheep. *Scientific Reports*. 2017;7:1–17.
20. Ma L, Li Z, Cai Y, Xu H, Yang R, Lan X. Genetic variants in fat-and short-tailed sheep from high-throughput RNA-sequencing data. *Animal Genetics*. 2018;49:483–7.
21. Mastrangelo S, Moioli B, Ahbara A, Latairish S, Portolano B, Pilla F, et al. Genome-wide scan of fat-tail sheep identifies signals of selection for fat deposition and adaptation. *Animal Production Science*. 2018;59:835–48.
22. Ahbara A, Bahbahani H, Almathen F, Al Abri M, Agoub MO, Abeba A, et al. Genome-wide variation, candidate regions and genes associated with fat deposition and tail morphology in Ethiopian indigenous sheep. *Frontiers in Genetics*. 2019;9:699.
23. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biology*. 2006;4:e72.
24. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature*. 2007;449:913–8.
25. McRae KM, McEwan JC, Dodds KG, Gemmell NJ. Signatures of selection in sheep bred for resistance or susceptibility to gastrointestinal nematodes. *BMC Genomics*. 2014;15:1–13.
26. Qanbari S. On the extent of linkage disequilibrium in the genome of farm animals. *Frontiers in Genetics*. 2020;10:1304.

27. Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature*. 2002;419:832–7.
28. Yuan Z, Xiang R, Li W, Li F, Yue X. Transcriptomic analyses revealed common tailed and perirenal adipose differentially expressed genes in four Chinese indigenous sheep breeds. *Livestock Science*. 2019;230:103832.
29. Gokdal O, Ulker H, Karakus F, Cengiz F, Temur C, Handil H. Growth, feedlot performance and carcass characteristics of Karakas and crossbred lambs (F1)(Ile de France x Akkaraman (G1) x Karakas) under rural farm conditions in Turkey. *South African Journal of Animal Science*. 2004;34:223–32.
30. Qanbari S, Gianola D, Hayes B, Schenkel F, Miller S, Moore S, et al. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics*. 2011;12:1–12.
31. Moradi MH, Khaltabadi-Farahani AH, Khodaei-Motlagh M, Kazemi-Bonchenari M, McEwan J. Genome-wide selection of discriminant SNP markers for breed assignment in indigenous sheep breeds. *Annals of Animal Science*. 2021;21:807–31.
32. Nejati-Javaremi A, Izadi F, Rahmati GH, Moradi M, Izadi F. Selection in fat-tailed sheep based on two traits of fat-tail and body weight versus single-trait total body weight. *International Journal of Agricultural Biology*. 2007;9:645–8.
33. Rothhammer S, Seichter D, Förster M, Medugorac I. A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. *BMC Genomics*. 2013;14:1–17.
34. Stainton JJ, Haley CS, Charlesworth B, Kranis A, Watson K, Wiener P. Detecting signatures of selection in nine distinct lines of broiler chickens. *Animal Genetics*. 2015;46:37–49.
35. Ma Y, Wei J, Zhang Q, Chen L, Wang J, Liu J, et al. A genome scan for selection signatures in pigs. *PLoS One*. 2015;10:e0116850.
36. Gurgul A, Jasielczuk I, Semik-Gurgul E, Pawlina-Tyszko K, Stefaniuk-Szmukier M, Szmatoła T, et al. A genome-wide scan for diversifying selection signatures in selected horse breeds. *PLoS One*. 2019;14:e0210751.
37. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. *Genome Research*. 2002;12:1805–14.
38. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123:585–95.
39. Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. *Genetics*. 2000;155:1405–13.
40. Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, et al. Signals of recent positive selection in a worldwide sample of human populations. *Genome Research*. 2009;19:826–37.
41. Khrunin A V, Khvorykh G V, Fedorov AN, Limborska SA. Genomic landscape of the signals of positive natural selection in populations of Northern Eurasia: A view from Northern Russia. *Plos One*. 2020;15:e0228778.
42. Mu J, Myers RA, Jiang H, Liu S, Ricklefs S, Waisberg M, et al. Plasmodium falciparum genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. *Nature Genetics*. 2010;42:268–71.
43. Grossman SR, Shylakhter I, Karlsson EK, Byrne EH, Morales S, Frieden G, et al. A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science*. 2010;327:883–6.
44. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21:263–5.
45. Davidson A. The Oxford companion to food. OUP Oxford; 2014.
46. Ryder ML. Sheep & man. Longman: London and New York; 1983.
47. Hayes BJ, Chamberlain AJ, Maceachern S, Savin K, McPartlan H, MacLeod I, et al. A genome map of divergent artificial selection between Bos taurus dairy cattle and Bos taurus beef cattle. *Animal Genetics*. 2009;40:176–84.
48. Cunningham F, Achuthan P, Akanni W, Allen J, Amode MR, Armean IM, et al. Ensembl 2019. *Nucleic Acids Research*. 2019;47:D745–51.

49. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*. 2009;4:1.
50. Zhu C, Fan H, Yuan Z, Hu S, Ma X, Xuan J, et al. Genome-wide detection of CNVs in Chinese indigenous sheep with different types of tails using ovine high-density 600K SNP arrays. *Scientific Reports*. 2016;6:1–9.
51. Yuan Z, Liu E, Liu Z, Kijas JW, Zhu C, Hu S, et al. Selection signature analysis reveals genes associated with tail type in Chinese indigenous sheep. *Animal Genetics*. 2017;48:55–66.
52. Bakhtiarizadeh MR, Salehi A, Alamouti AA, Abdollahi-Arpanahi R, Salami SA. Deep transcriptome analysis using RNA-Seq suggests novel insights into molecular aspects of fat-tail metabolism in sheep. *Scientific Reports*. 2019;9:1–14.
53. Zare M, Sadeghi M, Moradi-Shahrbabak M. Identification of single nucleotide polymorphisms of PPP2CA gene and its effect on carcass, growth and fat storage traits in Lori Bakhtiari and Zel sheep. *Animal Production Research*. 2016;5.
54. Xu Z, Sun H, Zhang Z, Zhang C-Y, Zhao Q, Xiao Q, et al. Selection signature reveals genes associated with susceptibility loci affecting respiratory disease due to pleiotropic and hitchhiking effect in Chinese indigenous pigs. *Asian-Australasian Journal of Animal Sciences*. 2020;33:187.
55. Zhu C, Huang X, Li M, Qin S, Fang S, Ma Y. GWAS and post-GWAS to identification of genes associated with sheep tail fat deposition. *Animal Science and Zoology*. 2019;06.
56. Moioli B, Pilla F, Ciani E. Signatures of selection identify loci associated with fat tail in sheep. *Journal of Animal Science*. 2015;93:4660–9.
57. Mastrangelo S, Bahbahani H, Moioli B, Ahbara A, Al Abri M, Almathen F, et al. Novel and known signals of selection for fat deposition in domestic sheep breeds from Africa and Eurasia. *PLoS One*. 2019;14:e0209632.
58. Zhu C, Li N, Cheng H, Ma Y. Genome wide association study for the identification of genes associated with tail fat deposition in Chinese sheep breeds. *Biology Open*. 2021;10:bio054932.
59. Ciani E, Mastrangelo S, Da Silva A, Marroni F, Ferencaković M, Ajmone-Marsan P, et al. On the origin of European sheep as revealed by the diversity of the Balkan breeds and by optimizing population-genetic analysis tools. *Genetics Selection Evolution*. 2020;52:1–14.
60. Oeth P, Beaulieu M, Park C, Kosman D, del Mistro G, van den Boom D, et al. iPLEX assay: Increased plexing efficiency and flexibility for MassArray system through single base primer extension with mass-modified terminators. *Sequenom application note*. 2005;27.
61. Barendse W, Harrison BE, Bunch RJ, Thomas MB, Turner LB. Genome wide signatures of positive selection: the comparison of independent samples and the identification of regions associated to traits. *BMC Genomics*. 2009;10:1–15.
62. MacEachern S, Hayes B, McEwan J, Goddard M. An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (*Bos taurus*) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in Domestic cattle. *BMC Genomics*. 2009;10:1–19.
63. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Source: Evolution*. 1984;36:1357–70.
64. Sayres MAW. Genetic diversity on the sex chromosomes. *Genome Biology and Evolution*. 2018;10:1064.
65. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. *Science*. 2002;296:2225–9.
66. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics*. 2001;68:978–89.

67. Do DN, Bissonnette N, Lacasse P, Miglior F, Sargolzaei M, Zhao X, et al. Genome-wide association analysis and pathways enrichment for lactation persistency in Canadian Holstein cattle. *Journal of Dairy Science*. 2017;100:1955–70.
68. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biology*. 2003;4:1–11.

Figures

Figure 1

Plots of windowed FST (a) and run of median homozygosity (b) in relation to genomic position for thin and fat tail breeds in candidate regions: SNP positions in the genome (bp) are shown on the X-axis, and windowed FST or median homozygosity are plotted on the Y-axis. Fat and thin tailed breeds are shown by blue diamonds and red squares respectively on median homozygosity plot.

Figure 2

Plot of $|iHS|$ in relation to the genomic position (bp) for thin and fat tail breeds on chromosome 7 (Upper) and our candidate region (lower): Fat and thin tail breeds are shown by blue circles and red diamonds respectively, and $|iHS|$ statistic averaged over 10 SNPs.

Figure 3

Plot of XP-EHH relation to genomic position (bp) for thin and fat tailed breeds on whole chromosome (upper) and our candidate region (lower) in different chromosomes: High positive values suggest selection in fat tailed population and negative values selection in thin tailed population.

Supplementary Files

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

- [AdditionalFile1.pdf](#)
- [AdditionalFile2.pdf](#)
- [AdditionalFile3.pdf](#)
- [AdditionalFile4.pdf](#)
- [AdditionalFile5.pdf](#)
- [AdditionalFile6.pdf](#)
- [AdditionalFile7.xlsx](#)
- [AdditionalFile8.pdf](#)

- [AdditionalFile9.pdf](#)
- [equation.docx](#)