GLIS1, a Potential Candidate Gene Affect Fat Deposition in Sheep Tail

Rongsong Luo (✉ luorongsong@mail.kiz.ac.cn)

Chinese Academy of Science

Xiaoran Zhang

Inner Mongolia University

Likai Wang

China Agricultural University

Li Zhang

Inner Mongolia University

Guangpeng Li

Inner Mongolia University

Zhong Zheng

Inner Mongolia University

Research Article

Keywords: population genomics, fat tail, Mongolian sheep, GLIS1, PDGFD

Posted Date: February 24th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-249943/v1>

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

[Read Full License](#)

1 ***GLIS1*, a Potential Candidate Gene Affect Fat Deposition in**
2 **Sheep Tail**

3 Rongsong Luo^{1,2}, Xiaoran Zhang¹, Likai Wang³, Li Zhang¹, Guangpeng Li^{1*}, Zhong
4 Zheng^{1*}

5
6 ¹*State Key Laboratory of Reproductive Regulation & Breeding of Grassland Livestock,*
7 *School of Life Sciences, Inner Mongolia University, Hohhot 010070, China*

8 ²*State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of*
9 *Zoology, Chinese Academy of Sciences, Kunming 650223, China*

10 ³*State Key Laboratory for Agrobiotechnology, College of Biological Sciences, China*
11 *Agricultural University, Beijing 100193, China*

12 **Corresponding author:*

13 *Guangpeng Li, State Key Laboratory of Reproductive Regulation & Breeding of*
14 *Grassland Livestock, School of Life Sciences, Inner Mongolia University, Hohhot*
15 *010070, China. E-mail: gpengli@imu.edu.cn*

16 *Zhong Zheng, State Key Laboratory of Reproductive Regulation & Breeding of*
17 *Grassland Livestock, School of Life Sciences, Inner Mongolia University, Hohhot*
18 *010070, China. E-mail: zhengzhong@imu.edu.cn*

19
20
21
22
23

Abstract

Fat deposition in sheep tails has complicated mechanisms. In this study, the population genomics analysis has been applied to identify candidate genes associated with fat tails based on high depth whole-genome sequencing of Mongolia sheep (MG, fat-tailed), Small Tail Han sheep (STH, fat-tailed) and two dairy sheep breeds DairyMeade and East Friesian (DS, thin-tailed). The selective signature analysis demonstrated that *GLIS1*, *LOC101117953*, *PDGFD* and *T* were in the significant divergent regions between DS and STH-MG. A nonsynonymous point mutation (g. 27807636G>T) was found within *GLIS1* in STH-MG and resulting in a Pro to Thr substitution. As a pro-adipogenic factor, *GLIS1* may play critical roles in the mesodermal cell differentiation during sheep fetal development and affect the fat deposition in sheep tails. This study provides a new insight into the genetic basis of species-specific traits of fat tails.

Keywords: population genomics; fat tail; Mongolian sheep; *GLIS1*; *PDGFD*

50 **Introduction**

51 The initial domestication, nature and artificial selection have driven the species to have
52 remarkable phenotypic diversifications in appearance, growth, local adaptability,
53 fertility, etc [1]. Sheep (*Ovis aries*) is one of the first domesticated livestock, which
54 could provide meat, milk, fur, and wool products for humans. China has diverse
55 landscapes and climatic features. Indigenous sheep breeds have developed well
56 adaptation to various environmental conditions, such as harsh winter, drought, food
57 scarcity, and high altitude, and become essential livestock for the animal husbandry
58 industry [2]. These breeds with different traits have already adapted in various
59 production systems in the vast geographical regions of China, which also provides us
60 the opportunities to elucidate the genetic basis of adaptation.

61 The wild ancestors of domestic sheep had thin tails, and it has been suggested that fat
62 tails be developed following domestication as an adaptive response to store energy
63 during migration and harsh winter [3]. Chinese indigenous fat-tailed sheep breeds are
64 mainly originated from an ancestral lineage Mongolia sheep, which are widely
65 distributed in northern China and Mongolian People's Republic. The over-deposition of
66 fat in the tails could help fat-tailed sheep overcome harsh environments characterized
67 by extreme cold, drought, and food scarcity. However, it may also compromise
68 reproduction and fattening performance under characterized by full housing or half
69 housing, thus reducing their economic values [4, 5]. Fortunately, the fat-tailed sheep
70 provide us an ideal model to study the mechanism of fat deposition in animals. In recent
71 years, population genomics have been applied extensively and effectively to identify
72 candidate genes associated with phenotypic diversity and important agronomic traits in
73 domestic animals. Previous studies provided evidence of promising candidate genes
74 influencing tail types based on single nucleotide polymorphism (SNP) markers [6-9].
75 However, the fat-tailed trait may be caused by a combination of multiple genes and had
76 a complicated co-regulation mechanism [10-12]. DairyMeade and East Friesian are the
77 two dairy sheep breed recently introduced into China which are large frame, fast growth,

and lean sheep types with typical thin tails. DairyMeade is a new dairy sheep breeds developed in New Zealand and originated from East Friesian [13, 14]. These two breeds provide us new materials to study the mechanism of fat deposition in sheep tails. In this study, we conducted high depth whole-genome sequencing of two typical fat-tailed breeds (Mongolian sheep and Small Tail Han sheep) and two typical thin-tailed breeds (DairyMeade sheep and East Friesian sheep) and provided new insights into the genetic basis of species-specific adaptive traits of the fat tail.

Materials and Methods

Sampling, DNA extraction and sequencing

Ear tissues of 13 dairy sheep (including 9 DairyMeade sheep, 2 East Friesian sheep, 1 East Friesian x Small Tail Han sheep F1 sheep and 1 DairyMeade x F1 F2 sheep), 7 Small Tail Han sheep and 9 Mongolia sheep were collected at different locations in Inner Mongolia Autonomous Region, China, for whole-genome resequencing (Fig. S1 and Table S1). All the ear tissues were collected and stored in liquid nitrogen immediately. The animal experimental procedures were performed according to the guidelines approved by the Ethics Committee of Inner Mongolia University.

Genomic DNA was extracted from the ear tissues using the standard phenol-chloroform method and checked for quality and quantity on the Qubit 2.0 fluorometer (Invitrogen). Next-generation sequence library construction for resequencing was performed with 3µg of genomic DNA according to the standard Illumina library preparation protocols and insert sizes from 300 to 500 bp. All libraries were sequenced on an Illumina Hiseq 2500 platform to generate paired-end reads. The resequencing depth ranged from 12.3x to 35.5x fold coverage, with an average depth of 18.14x.

Reads mapping and SNP calling

The adaptors and low-quality sequences of raw reads were trimmed and filtered to

obtain clean reads using FastQC (version 0.11.7) [15] and Trimmomatic (version 0.36) [16]. High-quality paired-end reads were mapped to the sheep reference genome OAR4.0 using the BWA-MEM alignment tool [17] implemented in BWA software with the command 'mem -t 10 -M'. Alignment of bam files were sorted and duplicated reads were removed using the SORTSAM and MARKDUPLICATES functions in the PICARDS package (picard-tools-2.18, <http://picard.sourceforge.net>). SAMTOOLS [18] was used to create index for bam files. Then SNPs were called using bcftools (mpileup) and filtered by vcftools (-minQ 30 --min-alleles 2 --max-alleles 2 --min-meanDP 4.0 -max-meanDP 72.0 --max-missing 0.9 --non-ref-ac 2 --remove-indels --recode --recode-INFO-all) [19]. Finally, all SNPs were annotated with ANNOVAR [20] according to NCBI's gene annotation database.

Population structure and genomic diversity analysis

Based on the genetic variants from autosomal, PLINK v1.9 [21] was used to calculate the genetic distance of the sheep individuals, followed by MEGA v7.0 [22] to construct the Neighbor-Joining (NJ) tree for the genetic distance matrix. The fourfold degenerate sites were also used to build ML and NJ tree, respectively. The principal component analysis of all sheep was conducted by using vcftools and PLINK with parameters '--maf 0.05 --max-missing 0.9 --chr-set 26'. The nucleotide diversity (in terms of nucleotide diversity π) was calculated using vcftools with parameters '--window-pi50000 --window-pi-step 25000'. The PopLDdecay software [23] was used to calculate r^2 (-minMAF 0.05 -hwcutoff 0.001 -Het 0.88 -Miss 0.25) for the pairs of SNPs and plot the LD curves. To exclude the bias introduced by the difference in the number of samples in different populations, we randomly sampled individuals from each population to keep the consistency of sample size during the calculation (7 individuals per group). Only SNPs with minor allele frequency (MAF) greater than 0.05 were considered.

Genomic selective sweep analysis

We identified potentially selective sweep signals using population differentiation index (F_{ST} , the DS group vs. the STH and MG groups) and locus-specific branch lengths (LSBL) [24, 25] based on the sliding window strategy (window size: 50 kb; step size: 25 kb). We estimated the LSBL based on the pairwise F_{ST} values [26] of each polymorphic site among three groups: Target (DS), Control (STH), and Background (MG). The formula $LSBL = (F_{ST} (DS-STH) + F_{ST} (DS-MG) - F_{ST} (STH-MG))/2$. The threshold for identifying the putative selection regions in the F_{ST} and LSBL analyses was empirically set to the top 1% percentile outliers. The genes putatively under selection were submitted to DAVID [27] for enrichment analysis of the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway categories. Fisher's Exact Test was used for correcting the p -value. Only terms with a p -value less than 0.05 were considered as significant and listed. The data analysis and visualization were carried out with our customized R scripts.

Results

Population structure and Genomic Diversity

Ear tissues of sheep were collected from different regions of Inner Mongolia, China, and conducted whole-genome sequencing at an average depth of 18.14x coverage (Table S2). After rigorously filtering, a total of 25, 375, 422 high-quality SNPs were obtained for further analysis. Among them, 15,525,859 SNPs were in intergenic regions, and 171, 462 SNPs were in exonic regions (Table S3). We then explored the genetic relationships between these sheep individuals based on all the genetic variants and four-fold degenerate sites. The phylogenetic tree was constructed by the neighbor-joining (NJ) method showed each sheep breed population clusters into a distinct clade (DairyMeade sheep and East Friesian sheep, DS; Small Tail Han sheep, STH; Mongolia

sheep, MG) (Fig. S2a). The same genetic affinities were obtained in phylogenetic trees constructed by the neighbor-joining (Fig. S3a) and maximum-likelihood (ML) (Fig. S3b) using four-fold degenerate sites. Principle component analysis (PCA) also uncovered different population structuring among DS, MG and STH, and the PC1 (4.06%) divided sheep individuals into fat-tailed sheep and thin-tailed sheep (Fig. S2b). ADMIXTURE analysis revealed similar population affinities, that fat-tailed sheep were separated from thin-tailed sheep (when $K = 2$), and there is no genetic exchange (Fig. S4).

Then, the genetic diversity index was calculated based on the whole-genome genetic variants. When compared with STH and MG, DS showed a lower level of nucleotide diversity (DS, $\pi=2.533\text{e-}3$; STH, $\pi=2.79\text{e-}3$; MG, $\pi=2.87\text{e-}3$) (Fig. S2c and Fig. S6) and slower decay rates of linkage disequilibrium (LD) (dropped to half of its maximum at 79 kb, followed by STH group (62 kb) and MG group (46 kb)) (Fig. S2d). These results suggested that indigenous breeds MG and STH have higher genetic diversity while bottlenecking and/or inbreeding occurred in the two dairy sheep breeds.

Selective signatures in fat- and thin-tailed sheep

The prominent phenotypic difference between DS and MG/STH is the tail shape. We then analyzed the inter- and intra-population diversities of the highly significant sweep regions to explore the genetic basis underlying fat deposition in the tail. The population differentiation index (F_{ST}) and the lineage-specific branch length (LSBL) of DS, STH and MG on a sliding-window basis (50 kb sliding window with 25 kb step increment) were calculated to detect the candidate divergent regions. We therefore found 798 genomic regions displayed an increased level of differentiation index between DS and STH-MG ($F_{ST} > 0.42$; LSBL > 0.435 ; both were top 1% threshold) (Fig. 2a and Table S4). In total, 510 shared protein-coding genes (619 and 614 genes were identified by F_{ST} and LSBL, respectively) were identified with signatures of selection (Table S5), which account for 1.96% of the whole-genome annotated genes (a total of 26076). The

functional enrichment analysis (in terms of KEGG) for the detected selective genes revealed overrepresented functional categories being associated with cell growth and immunity, such as focal adhesion (adjusted p -value = 0.00086) and T cell receptor signaling pathway (adjusted p -value = 0.0013) (Table S6).

Among these candidate divergent regions, two putative sweeps had the highest population differentiation scores. One located on chromosome 1 (LSBL = 0.86 and F_{ST} = 0.79) as displayed in the Manhattan plots (Fig. 1a). This region, from 27.75 Mb to 27.86 Mb, only harbors *GLIS1* gene (Fig. 1b). Further haplotype analysis showed DS carrying a haplotype pattern that differs strikingly from STH and MG (Fig. 1c and Fig. S6). A nonsynonymous point mutation (g. 27807636G>T) was found within *GLIS1* in STH-MG, which resulting in a nonsynonymous Pro107→Thr (P107T) substitution, making STH-MG different from DM and other thin tail mammals in this position (Fig. 2). The second putative sweep appeared at the loci on chromosome 13 (LSBL = 0.82 and F_{ST} = 0.78) harboring several pseudogenes, including *LOC101117953*, *LOC101118207* and *LOC101110166* (Fig. S7). Another genomic region (from 3.825 to 3.90 Mb) on chromosome 15 also exhibits strong selection signatures (LSBL = 0.92, 0.93) between DS and STH-MG (Fig. S8), in which harboring *PDGFD* gene, a member of the platelet-derived growth factor family. Other genes related to sheep tail traits were also found in our study, such as *T* (LSBL = 1.02, F_{ST} = 0.53).

Fig. 1 Selective-sweep analysis by comparing genomes between thin-tailed DS (dairy sheep, DairyMeade and East Friesian) and fat-tailed STH-MG (Small Tail Han sheep and Mongolian sheep). (a) Distribution of population differentiation index (F_{ST} , top panel) and the lineage-specific branch length (LSBL, bottom panel) between DS and STH-MG in a 50 kb sliding window with a 25 kb step increment across all autosomes. (b) π and LSBL values around the genomic region on chromosome 1 (from 27.4 Mb to 28.4 Mb) between DS and STH-MG populations. *GLIS1* is located in this genomic sweep region. The red, green and blue cells represent DS, STH and MG population, respectively. (c) Haplotype pattern of the selective-sweep region. Haplotype pattern in a region defined by SNPs that are at high frequency in DS and at low frequency in STH-MG. Each column is a polymorphic genomic location, each row is a phased haplotype, and the colored column on the left denotes the population identity of the individuals. The reference/alternative allele is indicated in light yellow/green.

						
	70	80	90	100	110	120
<i>Ovis aries</i>	GLKRCCLLGL	PPTSSVSVSP	CAASDITSII	RSSQTALVTC	VNGLRSTPLP	GDTGGPPKRA
<i>Capra hircus</i>	GLKRCCLLGL	PPTSSVSVSP	CAASDITSII	RSSQTALVTC	VNGLRSPPLP	GDTGGPSKRA
<i>Bos indicus</i>	GLKRCCLLGL	PPTSSVSVSP	CAASDITSII	RSSQTALVTC	VNGLRSPMP	GDTGGPPKRA
<i>Bos taurus</i>	GLKRCCLLGL	PPTSSVSVSP	CAASDITSII	RSSQTALVTC	VNGLRSPMP	GDTGGPPKRA
<i>Bos mutus</i>	GLKRCCLLGL	PPTSSVSVSP	CAASDITAI	RSSQTALVTC	VNGLRSPMP	GDTGGPPKRA
<i>Bubalus bubalis</i>	GLKRCCLLGL	PPTSSVSVSP	CAASDITSII	RSSQTALVTC	VNGLRSPMP	GDTGGPPKRA
<i>Sus scrofa</i>	GLKRCCLLGL	PPTSSASASP	CASSDITSIV	RASQTALVTC	VNGLRSPPLP	GDLGGPPRR
<i>Odocoileus virginianus</i>	GLKRCCLLGL	PPTSSVSVSP	CAASDITSII	RSSQTALVTC	VNGLRSPPLP	GDTGGPPKRA
<i>Vicugna pacos</i>	GLKRCCLLGL	PPTSSVSASP	CASSDVTSII	RSSQTALVTC	VNGLRSPPLP	GDLGGPPKRA
<i>Camelus dromedarius</i>	GLKRCCLLGL	PPTSSISASP	CASSDVTSII	RSSQTALVTC	VNGLRSPPLP	GDLGVPPKRA

Fig. 2 Alignment of the amino acid sequences of *GLIS1* protein in different mammals.

Position where the amino acid differ are highlighted in gray.

Discussion

The fat tail phenotype in sheep is probably the result of a combination of multiple genes.

The study suggested that the ovine genome have encountered a recent selective sweep

at *GLIS1* loci. *GLIS1* is a zinc finger protein that acts as both activator and repressor of

transcription [28]. During mouse embryonic development, it starts to express in the

forelimb, hindlimb and tail at 10.0 days post coitus (dpc), then it expresses in the

anterior region of the forelimb, ventral part of the body and tail at 10.5 dpc and the

expression gets ever stronger at 11.0 dpc, consistent with mesoderm differentiation [29].

In a recent study, *GLIS1* was recognized as a novel pro-adipogenic transcription factor.

It expresses at a high level in bipotent muscle satellite cells. But when overexpressed,

increased occupancy of *GLIS1* is observed at the promoters of adipogenic genes *Adipoq*,

Cebpa and *Ucp1*, and drives brown adipogenesis *in vitro* and *in vivo* [30]. The role of

GLIS1 is rarely studied in sheep, but it was reported that SNP in *GLIS1* affects the feed

efficiency in Dual Purpose and Blackface rams [31] which may also be related to

different muscle and fat ratio in the carcass. DS and MG/STH had a remarkable

difference in growth speed and tail phenotype. For DS, almost no fat deposition could

be found inside the tail from newborn lamb to adult sheep. While in MG and STH, a

large amount of fat deposition accumulated in the ventral region of the tail, and

subcutaneously. It is worth noting that fat deposition in their tail's ventral region could be observed as early as the postnatal stage, indicating that the tail phenotype is already determined during fetal development. Thus, it could be an innate feature of adaptation for MG and STH to face the challenge of cold and food scarcity lambing season (March to April) in northern China. Combined with this information together, we hypothesized that, as a pro-adipogenic factor, *GLIS1* may play key roles in the mesodermal cell differentiation during fetal development of fat-tailed sheep and initiate the accumulation and differentiation of preadipocytes in the tails.

Previous studies suggested that the *LOC101117953* and the *BMP2* (bone morphogenetic protein 2, which locates in chromosome 13 from 48387181 to 48400679 bp, upstream of the current sweep region) were related to the fat deposition in fat-tailed sheep [6, 8, 9]. Since *LOC101117953* is a retro-copy of *PPP1CC* (protein phosphatase PP1-gamma catalytic subunit gamma) which lacking the promoter regions and not expressing in adult tissues, it is less likely to be the causative gene for tail phenotypes [9]. Previous studies showed *PDGFD* is a likely causal gene for fat deposition in the sheep tail, which promotes proliferation and inhibits differentiation of preadipocyte [12, 32-34]. Two SNPs of *PDGFD* significantly affect the tail length and tail width, respectively [35]. *T* is the key regulator of mesoderm formation during early development and was reported related to short-tail phenotype in Hulunbuir sheep, a subpopulation of Mongolia sheep [36]. It may be also related to the caudal vertebra phenotype differences between DS and STH/MG, since DS has long and straight tails and STH/MG has relatively shorter tails with a slightly curved tail tip.

Our results demonstrated that ovine genome encountered a recent selective sweep at *GLIS1* loci. As a novel-pro-adipogenic transcription factor, *GLIS1* may initiate the accumulation and differentiation of preadipocytes in the tails during fetal development and affect the tail phenotypes in sheep.

Data accessibility.

The whole-genome resequencing datasets used in this study were submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive with the accession code PRJNA531155. The additional data supporting the conclusions in this paper can be found in the additional information.

Acknowledgments.

We would like to thank Gong Zhang and Meng Tian Ran Animal Husbandry Technical Development Co., Ltd. for the help during DairyMeade and East Friesian sheep sampling.

Funding.

This work was supported by the Science and Technology Major Project of Inner Mongolia Autonomous Region of China to the State Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock (ZDZX2018065) and to Li Zhang (30900-517390), the Inner Mongolia scientific and technological innovation, leading & reward fund (30500-5173203, to Li Zhang) and the Inner Mongolia natural science foundation fund (2019MS03024, to Zhong Zheng).

Authors' contributions.

ZZ, RSL and GPL conceived and designed the experiments. ZZ, LZ, XRZ, and LKW completed sampling and performed the experiments. RSL and ZZ completed data analysis and visualized it. ZZ and GPL supervised the project. RSL and ZZ wrote the manuscript. All authors read and approved the final manuscript.

Conflict of interest.

The authors declare no conflicts of interest.

References:

- 291 1. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X et al (2007) Genome-wide
 292 detection and characterization of positive selection in human populations. *Nature* 449:913-918.
 293 <https://doi.org/10.1038/nature06250>
- 294 2. Yang J, Li W, Lv F, He S, Tian S, Peng W, Sun Y, Zhao Y, Tu X, Zhang M, Xie X, Wang Y, Li J,
 295 Liu Y, Shen Z, Wang F, Liu G, Lu H, Kantanen J, Han J, Li M, Liu M (2016) Whole-Genome
 296 Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments.
 297 *Molecular biology and evolution* 33:2576-2592. <https://doi.org/10.1093/molbev/msw129>
- 298 3. Moradi MH, Nejati-Javaremi A, Moradi-Shahrbabak M, Dodds KG, McEwan JC (2012) Genomic
 299 scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated
 300 with fat deposition. *BMC Genet* 13:10. <https://doi.org/10.1186/1471-2156-13-10>
- 301 4. Kilminster TF, Greeff JC (2011) A note on the reproductive performance of Damara, Dorper and
 302 Merino sheep under optimum management and nutrition for Merino ewes in the eastern wheatbelt of
 303 Western Australia. *Trop Anim Health Prod* 43:1459-1464. [https://doi.org/10.1007/s11250-011-9871-](https://doi.org/10.1007/s11250-011-9871-8)
 304 8
- 305 5. Frisch RE (1987) Body fat, menarche, fitness and fertility. *Hum Reprod* 2:521-533.
 306 <https://doi.org/10.1093/oxfordjournals.humrep.a136582>
- 307 6. Moioli B, Pilla F, Ciani E (2015) Signatures of selection identify loci associated with fat tail in sheep.
 308 *J Anim Sci* 93:4660-4669. <https://doi.org/10.2527/jas.2015-9389>
- 309 7. Moradi MH, Nejati-Javaremi A, Moradi-Shahrbabak M, Dodds KG, McEwan JC (2012) Genomic
 310 scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated
 311 with fat deposition. *BMC Genet* 13:10. <https://doi.org/10.1186/1471-2156-13-10>
- 312 8. Wei C, Wang H, Liu G, Wu M, Cao J, Liu Z, Liu R, Zhao F, Zhang L, Lu J, Liu C, Du L (2015)
 313 Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds.
 314 *BMC Genomics* 16:194. <https://doi.org/10.1186/s12864-015-1384-9>
- 315 9. Pan Z, Li S, Liu Q, Wang Z, Zhou Z, Di R, An X, Miao B, Wang X, Hu W, Guo X, Lv S, Li F, Ding
 316 G, Chu M, Li Y (2019) Rapid evolution of a retro-transposable hotspot of ovine genome underlies the
 317 alteration of BMP2 expression and development of fat tails. *BMC Genomics* 20:261.
 318 <https://doi.org/10.1186/s12864-019-5620-6>
- 319 10. Zhao F, Deng T, Shi L, Wang W, Zhang Q, Du L, Wang L (2020) Genomic Scan for Selection
 320 Signature Reveals Fat Deposition in Chinese Indigenous Sheep with Extreme Tail Types. *Animals*
 321 10:773. <https://doi.org/10.3390/ani10050773>
- 322 11. Xu SS, Ren X, Yang GL, Xie XL, Zhao YX, Zhang M, Shen ZQ, Ren YL, Gao L, Shen M, Kantanen
 323 J, Li MH (2017) Genome-wide association analysis identifies the genetic basis of fat deposition in the
 324 tails of sheep (*Ovis aries*). *Animal Genetics* 48:560-569. <https://doi.org/10.1111/age.12572>
- 325 12. Dong K, Yang M, Han J, Ma Q, Han J, Song Z, Luosang C, Gorkhali NA, Yang B, He X, Ma Y,
 326 Jiang L (2020) Genomic analysis of worldwide sheep breeds reveals PDGFD as a major target of fat-
 327 tail selection in sheep. *BMC Genomics* 21 <https://doi.org/10.1186/s12864-020-07210-9>
- 328 13. ME K, JE K, CB P (2014) Sheep Dairying in New Zealand - The Kingsmeade Story. *Proceedings of*
 329 *the New Zealand Society of Animal Production* 74:58-61.

14. Allison AJ (1995) Importing a sheep which offers more - the East Friesian. *Proceeding of the New Zealand Society of Animal Production* 55:321-333.
15. Bioinformatics B (2011) FastQC: a quality control tool for high throughput sequence data. Cambridge, UK: Babraham Institute
16. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.
17. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*
18. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078-2079.
19. Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987-2993. <https://doi.org/10.1093/bioinformatics/btr509>
20. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* 38:e164.
21. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics* 81:559-575.
22. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution* 33:1870-1874.
23. Zhang C, Dong S, Xu J, He W, Yang T (2018) PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35:1786-1788.
24. Ai H, Fang X, Yang B, Huang Z, Chen H, Mao L, Zhang F, Zhang L, Cui L, He W (2015) Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nature genetics* 47:217.
25. Shriver MD, Kennedy GC, Parra EJ, Lawson HA, Sonpar V, Huang J, Akey JM, Jones KW (2004) The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Human genomics* 1:274.
26. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome research* 12:1805-1814.
27. Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 4:44-57.
28. Kim Y, Lewandoski M, Perantoni AO, Kurebayashi S, Nakanishi G, Jetten AM (2002) Identification of Glis1, a Novel Gli-related, Krüppel-like Zinc Finger Protein Containing Transactivation and Repressor Functions. *Journal of Biological Chemistry* 277:30901-30913. <https://doi.org/10.1074/jbc.M203563200>
29. Nakashima M, Tanese N, Ito M, Auerbach W, Bai C, Furukawa T, Toyono T, Akamine A, Joyner AL (2002) A novel gene, GliH1, with homology to the Gli zinc finger domain not required for mouse development. *Mech Dev* 119:21-34. [https://doi.org/10.1016/s0925-4773\(02\)00291-5](https://doi.org/10.1016/s0925-4773(02)00291-5)
30. Tosić M, Allen A, Willmann D, Lepper C, Kim J, Duteil D, Schüle R (2018) Lsd1 regulates skeletal muscle regeneration and directs the fate of satellite cells. *Nature Communications* 9 <https://doi.org/10.1038/s41467-017-02740-5>

31. Cockrum RR, Pickering NK, Anderson RM, Hyndman DL, Bixley MJ, Dodds KG, Stobart RH, McEwan JC, Cammack KM (2012) Identification of single nucleotide polymorphisms associated with feed efficiency in rams.:79.
32. Zhao F, Deng T, Shi L, Wang W, Zhang Q, Du L, Wang L (2020) Genomic Scan for Selection Signature Reveals Fat Deposition in Chinese Indigenous Sheep with Extreme Tail Types. *Animals* (Basel) 10<https://doi.org/10.3390/ani10050000>
33. Li X, Yang J, Shen M, Xie X, Liu G, Xu Y, Lv F, Yang H, Yang Y, Liu C, Zhou P, Wan P, Zhang Y, Gao L, Yang J, Pi W, Ren Y, Shen Z, Wang F, Deng J, Xu S, Salehian-Dehkordi H, Hehua E, Esmailzadeh A, Dehghani-Qanatqestani M, Štěpánek O, Weimann C, Erhardt G, Amane A, Mwacharo JM, Han J, Hanotte O, Lenstra JA, Kantanen J, Coltman DW, Kijas JW, Bruford MW, Periasamy K, Wang X, Li M (2020) Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nature Communications* 11<https://doi.org/10.1038/s41467-020-16485-1>
34. Wei C, Wang H, Liu G, Wu M, Cao J, Liu Z, Liu R, Zhao F, Zhang L, Lu J, Liu C, Du L (2015) Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds. *BMC Genomics* 16<https://doi.org/10.1186/s12864-015-1384-9>
35. Li Q, Lu Z, Jin M, Fei X, Quan K, Liu Y, Ma L, Chu M, Wang H, Wei C (2020) Verification and Analysis of Sheep Tail Type-Associated PDGF-D Gene Polymorphisms. *Animals* 10:89. <https://doi.org/10.3390/ani10010089>
36. Zhi D, Da L, Liu M, Cheng C, Zhang Y, Wang X, Li X, Tian Z, Yang Y, He T, Long X, Wei W, Cao G (2018) Whole Genome Sequencing of Hulunbuir Short-Tailed Sheep for Identifying Candidate Genes Related to the Short-Tail Phenotype. *G3* (Bethesda) 8:377-383. <https://doi.org/10.1534/g3.117.300307>

Figures

Figure 1

Selective-sweep analysis by comparing genomes between thin-tailed DS (dairy sheep, DairyMeade and East Friesian) and fat-tailed STH-MG (Small Tail Han sheep and Mongolian sheep). (a) Distribution of population differentiation index (F_{ST} , top panel) and the lineage-specific branch length (LSBL, bottom panel) between DS and STH-MG in a 50 kb sliding window with a 25 kb step increment across all autosomes. (b) π and LSBL values around the genomic region on chromosome 1 (from 27.4 Mb to 28.4 Mb) between DS and STH-MG populations. *GLIS1* is located in this genomic sweep region. The red, green and blue cells represent DS, STH and MG population, respectively. (c) Haplotype pattern of the selective-sweep region. Haplotype pattern in a region defined by SNPs that are at high frequency in DS and at low frequency in STH-MG. Each column is a polymorphic genomic location, each row is a phased haplotype,

and the colored column on the left denotes the population identity of the individuals. The reference/alternative allele is indicated in light yellow/green.

											
		70		80		90		100		110	120
<i>Ovis aries</i>	GLKRCCLLGL	PPTSSSVSVSP	CAASDITSII	RSSQTALVTC	VNGLRSTPLP	GDTGGPPKRA					
<i>Capra hircus</i>	GLKRCCLLGL	PPTSSSVSVSP	CAASDITSII	RSSQTALVTC	VNGLRSPPLP	GDTGGPSKRA					
<i>Bos indicus</i>	GLKRCCLLGL	PPTSSSVSVSP	CAASDITSII	RSSQTALVTC	VNGLRSPMP	GDTGGPPKRA					
<i>Bos taurus</i>	GLKRCCLLGL	PPTSSSVSVSP	CAASDITSII	RSSQTALVTC	VNGLRSPMP	GDTGGPPKRA					
<i>Bos mutus</i>	GLKRCCLLGL	PPTSSSVSVSP	CAASDITAI	RSSQTALVTC	VNGLRSPMP	GDTGGPPKRA					
<i>Bubalus bubalis</i>	GLKRCCLLGL	PPTSSSVSVSP	CAASDITSII	RSSQTALVTC	VNGLRSPMP	GDTGGPPKRA					
<i>Sus scrofa</i>	GLKRCCLLGL	PPTSSASASP	CASSDITSIV	RASQTALVTC	VNGLRSPPLP	GDLGGPPRA					
<i>Odocoileus virginianus</i>	GLKRCCLLGL	PPTSSSVSVSP	CAASDITSII	RSSQTALVTC	VNGLRSPPLP	GDTGGPPKRA					
<i>Vicugna pacos</i>	GLKRCCLLGL	PPTSSVSASP	CASSDVTSII	RSSQTALVTC	VNGLRSPPLP	GDLGGPPKRA					
<i>Camelus dromedarius</i>	GLKRCCLLGL	PPTSSISASP	CASSDVTSII	RSSQTALVTC	VNGLRSPPLP	GDLGVPPKRA					

Figure 2

Alignment of the amino acid sequences of GLIS1 protein in different mammals. Position where the amino acid differ are highlighted in gray.

Supplementary Files

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

- [SupplementaryTables.xlsx](#)
- [SupplementaryFigures.docx](#)