

RESEARCH ARTICLE

Long divergent haplotypes introgressed from wild sheep are associated with distinct morphological and adaptive characteristics in domestic sheep

Hong Cheng^{1,2}, Zhuangbiao Zhang¹, Jiayue Wen¹, Johannes A. Lenstra³, Rasmus Heller⁴, Yudong Cai¹, Yingwei Guo¹, Ming Li¹, Ran Li¹, Wenrong Li^{5,6}, Sangang He^{5,6}, Jintao Wang¹, Junjie Shao¹, Yuxuan Song¹, Lei Zhang¹, Masum Billah¹, Xihong Wang¹, Mingjun Liu^{5,6*}, Yu Jiang^{1*}

1 Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, China, **2** Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China, **3** Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands, **4** Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, Copenhagen, Denmark, **5** Key Laboratory of Ruminant Genetics, Breeding & Reproduction, Ministry of Agriculture, China, **6** Key Laboratory of Animal Biotechnology of Xinjiang, Institute of Biotechnology, Xinjiang Academy of Animal Science, Urumqi, Xinjiang, China

☯ These authors contributed equally to this work.

* xjlmj2006@126.com (ML); yu.jiang@nwafu.edu.cn (YJ)



OPEN ACCESS

Citation: Cheng H, Zhang Z, Wen J, Lenstra JA, Heller R, Cai Y, et al. (2023) Long divergent haplotypes introgressed from wild sheep are associated with distinct morphological and adaptive characteristics in domestic sheep. *PLoS Genet* 19(2): e1010615. <https://doi.org/10.1371/journal.pgen.1010615>

Editor: Mikkel H. Schierup, Aarhus University, DENMARK

Received: April 15, 2022

Accepted: January 13, 2023

Published: February 23, 2023

Copyright: © 2023 Cheng et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Raw sequencing data generated in this study have been deposited to the NCBI BioProject database under accession numbers PRJNA814428 and PRJNA521847 (Tibetan sheep). The custom code and scripts have deposited in GitHub https://github.com/Chenghong412/sheep_introgression.

Funding: This work was supported by grants from the National Natural Science Foundation of China (U21A20247 and 31822052) Shaanxi Innovation

Abstract

The worldwide sheep population comprises more than 1000 breeds. Together, these exhibit a considerable morphological diversity, which has not been extensively investigated at the molecular level. Here, we analyze whole-genome sequencing individuals of 1,098 domestic sheep from 154 breeds, and 69 wild sheep from seven *Ovis* species. On average, we detected 6.8%, 1.0% and 0.2% introgressed sequence in domestic sheep originating from Iranian mouflon, urial and argali, respectively, with rare introgressions from other wild species. Interestingly, several introgressed haplotypes contributed to the morphological differentiations across sheep breeds, such as a *RXFP2* haplotype from Iranian mouflon conferring the spiral horn trait, a *MSRB3* haplotype from argali strongly associated with ear morphology, and a *VPS13B* haplotype probably originating from urial and mouflon possibly associated with facial traits. Our results reveal that introgression events from wild *Ovis* species contributed to the high rate of morphological differentiation in sheep breeds, but also to individual variation within breeds. We propose that long divergent haplotypes are a ubiquitous source of phenotypic variation that allows adaptation to a variable environment, and that these remain intact in the receiving population probably due to reduced recombination.

Team Project (2022TD-10), Natural Science Basic Research Program of Shaanxi (2021JCW-11) to Y.J., Research on High-efficiency and Healthy Breeding Technology for Both Milk and Meat Sheep, the Jinchang Meat Sheep Test Demonstration Base (TGZX202137) to Y.S. and Genetic Improvement and Breeding of Xinjiang Native Sheep Breeds, Key Program of Science & Technology of Xinjiang (2022A02001-2) to M.L. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors declare no competing interests.

Author summary

Introgression, introducing beneficial alleles from wild relatives, has been repeatedly proved to play important roles in adaptation to various environments during animal domestication and subsequent breeds formation. Several adaptive introgressions have been reported in domestic sheep, however, the systemic exploration is still lacking. Using a collection of more than 1000 individuals including 154 domestic breeds and seven wild relatives, we describe the genomic introgression spectrum across worldwide domestic sheep. Interestingly, we found several functional genes associated with morphological traits located in the introgressed regions. Besides, the frequency of highly divergent haplotypes in numerous introgressed regions are quite different among domestic breeds, indicating selection acted on these fragments. We speculated that the introgressed fragments contribute to the fast morphological differentiation of ~1,400 distinct domestic breeds. Deciphering the phenotypic variations in response to different introgressed haplotypes can reflect underlying breeding mechanisms. Overall, our work proves the important effect of introgression on sheep breeding, and shows how introgression, selection and recombination shape the genetic and phenotypic diversity.

Introduction

The importance of introgression has been recognized as the mounting evidence accumulate. Candidate genes with introgression signatures in human, animals and plants involved broad functional categories [1, 2], including defence against pathogen [3–5], pigmentation [6–8], altitude adaptation [9–11], substance metabolism [3–5, 12] and other uncharacterized functions. Among them, adaptive introgression that facilitates the adaptation to a diversity of environment is particularly important. Recent studies have identified several intriguing examples of adaptive introgression in domestic animals, that probably have an important influence on their domestication and evolution, such as goat, sheep and cattle [4, 5, 13]. Adaptive introgression generally undergone selection, resulting in high frequency or fixed in particular populations but absent or low frequency in other populations. Such selection against introgressed fragments in functional regions facilitated the phenotypic differences among distinct populations.

The *Ovis* genus, including seven recognized wild species (snow sheep, *O. nivicola*; bighorn, *O. canadensis*; thinhorn, *O. dalli*; argali, *O. ammon*; urial, *O. vignei*; Asiatic mouflon, *O. orientalis* and European mouflon, *O. musimon*) and only one domestic species, possess intricate evolutionary history and pronounced gene flow events. The admixture events among these *Ovis* species have been repeatedly documented previously [5, 11, 14–16]. However, most reports focused on isolated cases of gene flow between two sympatric *Ovis* species, e.g. the introgression from European mouflon into European domestic breeds [5, 14], from Iranian mouflon into domestic sheep [15], and from argali to Tibetan sheep [11]. Moreover, some of these studies based on the sheep 50K SNP BeadChip considered only a limited number of variants to evaluate the introgression proportions [11, 14]. Given these recurrent findings of interspecies introgression, it would be preferable to jointly infer the magnitude of such introgression across the whole genus, as pairwise introgression results can be biased by ignoring the presence of other introgression events in such reticulated evolution scenarios. Nonetheless, these studies have yielded interesting evidence for introgression of functional genes, such as the *HBB* locus as adaptation to the high-altitude of the Qinghai-Tibetan plateau [11].

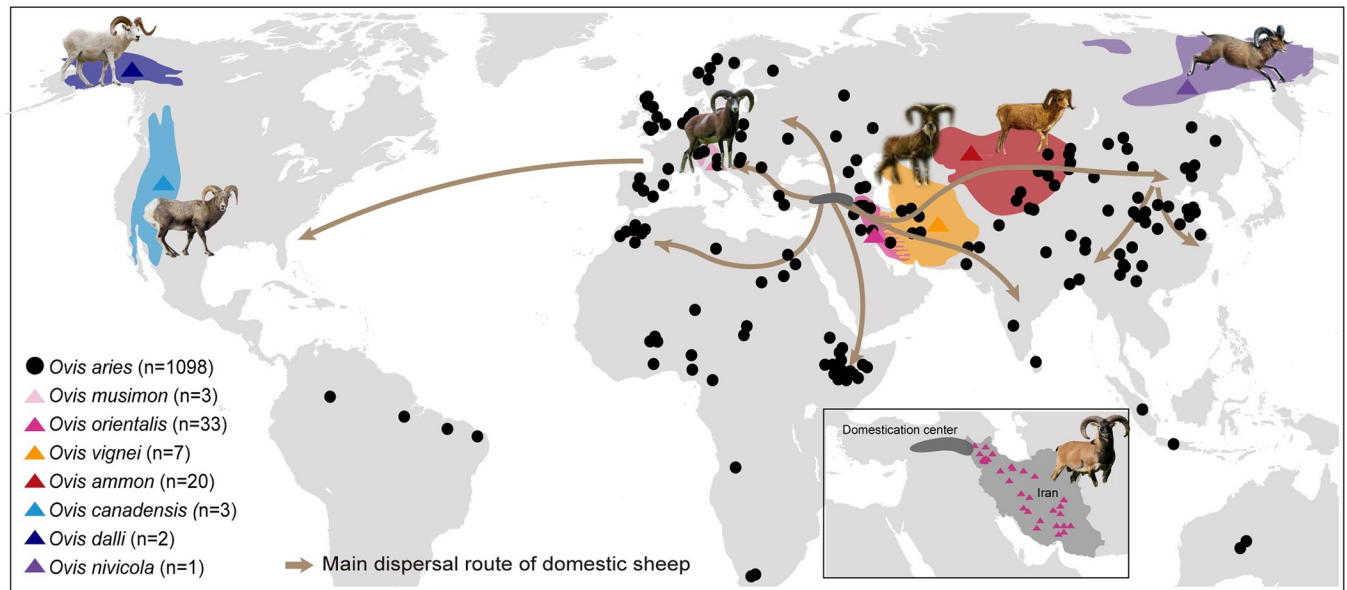


Fig 1. Locations of different geographically wild *Ovis* species and diverse domestic sheep populations used in this study. The colored blocks show the geographic distributions of the wild species. And each black dot represents a domestic breed. The dark grey block means the domestication center of sheep, and the solid lines represent dispersal routes of domestic sheep out of their domestication area [20–22]. The map base layer and photo credits using in this figure are showed in S7 Table. https://commons.wikimedia.org/wiki/File:Polio_worldwide_2012.svg.

<https://doi.org/10.1371/journal.pgen.1010615.g001>

Domestic sheep (*Ovis aries*) descends from Asiatic mouflon [17,18] approximately 11,000 years ago in southeastern Anatolia of Turkey. As many as 1,400 different breeds [19] exhibit a remarkable phenotypic diversity in response to selection pressures from various environments as well as to human selection. How did these distinct phenotypes form in such a short period after domestication, and how were these phenotypic variations affected by introgression remain largely explored. A comprehensive understanding of the genome-wide influence of introgression from wild relatives into domestic sheep on phenotypes or traits is lacking.

To excavate the impact of introgression on diverse phenotypic traits of sheep, here we build a collection of 1,167 whole-genome resequenced sheep (Fig 1 and S1 Table) with 156 samples were newly generated (S2 Table). We phased the genomes into haplotypes for an integrative analysis of the introgression from different wild sheep species. We further collected genotypes and phenotypes from East-Friesian sheep × Hu Sheep F2 hybrids to annotate the potential functional impact of various introgression signals. Our results provide further insight into the reticulated history of sheep evolution and particularly into the role of divergent haplotypes in the phenotypic diversity.

Results

Genetic variant data and phylogenetic relationships of *Ovis* genus

To investigate the phylogeny and population differentiation of *Ovis* species, we collected and generated a whole genome SNP dataset from 1,167 individuals comprising 1,098 domestic sheep across the geographic distribution of 154 breeds and 69 samples of their seven wild relatives (S1 Table). After aligning reads to the Oar_v4.0 (GCF_000298735.2) and quality control, a total of 83,386,953 SNPs were detected.

A whole-genome maximum likelihood (ML) phylogenetic tree revealed that European mouflon is intermediate between Asiatic mouflon and domestic sheep (Figs 2A and S1). This

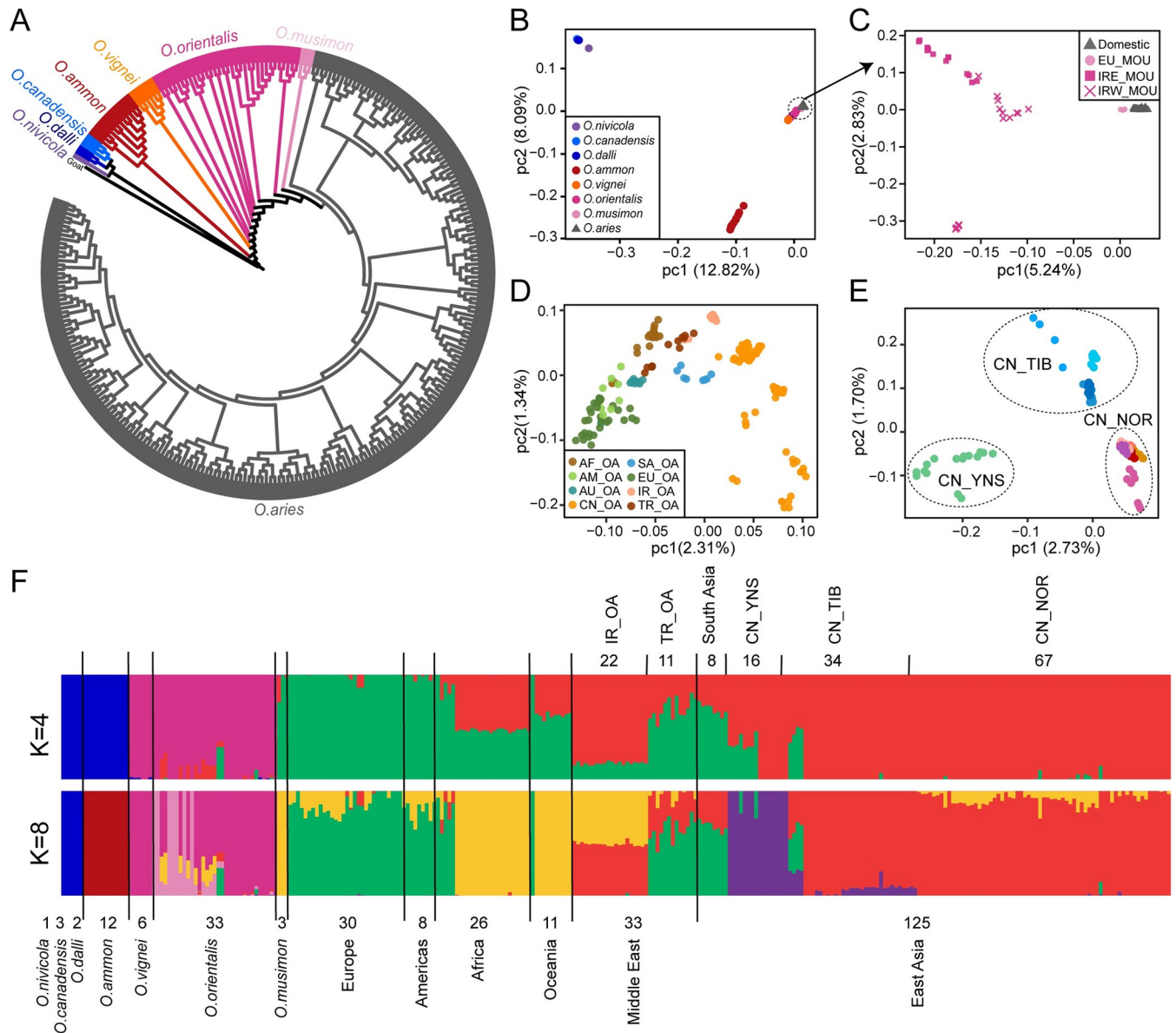


Fig 2. Phylogenetic analysis and population genetic structure. (A) A maximum likelihood (ML) phylogenetic tree of 293 representative samples covering all species of *Ovis* genus with Goat (GCA_000317765.2) as an outgroup. The tree was built with 100 bootstraps using a total of 332,990 4DV sites. (B-E) PCA analysis of wild and domestic sheep (B), Iranian and European mouflon and domestic sheep (C), domestic (D) and Chinese sheep (E), respectively. EU_MOU, European mouflon; IR-MOU, mouflon sheep from Iran. AF_OA, AM_OA, AU_OA, CN_OA, SA_OA, EU_OA, IR_OA, TR_OA, separately represent domestic sheep from Africa, America, Oceania, China, south Asia, Europe, Iran and Turkey. CN_TIB, CN_YNS, CN_NOR for domestic sheep from Tibet, Yunnan, Northern China. (F) ADMIXTURE results for k = 4 and k = 8.

<https://doi.org/10.1371/journal.pgen.1010615.g002>

is in agreement with their descent from the ancestral population of European domestic sheep, which were then subsequently replaced by the first domestic wool sheep populations. Domestic sheep was much closer to the Iranian mouflon located in western Iran (S2 Fig), near to the domestic center. The evolutionary relationships among other wild sheep were consistent with the topology inferred by mtDNA sequences [23]. Principal component analysis (PCA) further divided *Ovis* species into three separate clusters (1) *O. nivicola*, *O. canadensis* and *O. dalli*; (2) *O. ammon*; (3) *O. vignei*, *O. orientalis*, *O. musimon* and *O. aries* (Fig 2B). The PCA of mouflon

and domestic sheep as well as the ADMIXTURE pattern at $k \geq 7$ (Figs 2F and S3) shows a differentiation of eastern and western Iranian mouflon according to their geographic origin (Fig 2B). Moreover, the PCA confirms the relatively close relationship of western Iranian mouflons and domestic sheep. Both PCA and ADMIXTURE at $k = 8$ reveal a correlation of genetic clustering and geographic distances for domestic sheep (Figs 2D, 2F and S3). Samples from China were subdivided into three groups (Fig 2E and 2F), CN_YNS (Yunnan sheep), CN_TIB (Oula, Prairie Tibetan, Valley Tibetan) and CN_NOR (Small tailed Han sheep, Cele black sheep, Hu sheep, Tan sheep, Bayinbuluke sheep and Ujimqin Sheep).

Introgressions from wild relatives into domestic sheep

To evaluate the admixture proportion and locate the putative introgressive fragments in domestic populations from their wild relatives, we performed local ancestry inference (LAI) method program LOTER for each fully phased sheep genome. The bighorn, thinhorn, argali, urial, Iranian mouflon and European domestic sheep were used as source populations. European domestic sheep, which has not been in contact with the Asian wild sheep populations following their divergence, shared only few alleles with wild species (S4B Fig) and was used as the non-introgressed reference population. The European mouflon was not tested as a source population due to its close relationship with domestic sheep (Fig 2), which would confound the detection of introgression from the other wild sheep species.

In order to distinguish the putative signals of introgression from shared ancestral polymorphisms (incomplete lineage sorting, ILS), we calculated the expected length L of ILS tracts (see Materials and Methods) using variable recombination rates and removed the inferred introgressed segments with a length $< L$. This could remove some short introgressed regions, but is justified by the expectation that introgressed regions are considerably longer as they had less time to be broken up by recombination. In addition, we were mostly concerned about long introgressed haplotypes in the present study.

Using the filtered results, we calculated the genome-wide proportions of admixture. We detected an average of 10,036 segments (5,600–13,057), in total corresponding to an average of 180 Mb of wild *Ovis* sequence (range 96–224 Mb, SD = 23Mb) for each haploid domestic sheep genome. The average proportions of domestic sheep genome from Iranian mouflon, urial, argali, bighorn and thinhorn sheep were 6.8% (3.8–8.5%), 1.0% (0.5–1.4%), 0.2% (0.07–0.3%), 0.03% (0.01–0.05%) and 0.01% (0.006–0.022%), respectively (Figs 3A, S6 and S7), values that are similar to those previously reported for sympatric wild-to-domestic introgression. The introgressed proportions varied considerably across wild donor species, in particular between Iranian mouflon and the other wild species (Fig 3A). The domestic sheep from Asia had a significant higher percentage of Iranian mouflon lineage than those from Americas (S8 Fig). More detailed statistics revealed that domestic sheep from Iran shared more variations with Iranian mouflon compared with domestic sheep from Turkey (S8 Fig). Such significant statistical differences across the whole genome were less likely produced by ILS, supporting the introgression from Iranian mouflon to domestic sheep. East Asian domestic sheep has a relatively strong introgression from urial and argali (Fig 3A), consistent with their biogeographic history.

We further computed the modified f_d -statistic value [24] for each 50-kb window with a 20-kb step across the genomes in the form f_d (European domestic sheep, domestic population; wild source of introgression, goat) (Fig 3C–3E). We grouped the domestic samples into 16 focal populations (see Materials and Methods). For each population, the regions with significant f_d values ($P < 0.001$) were defined as potentially introgressed regions [11,25]. We further

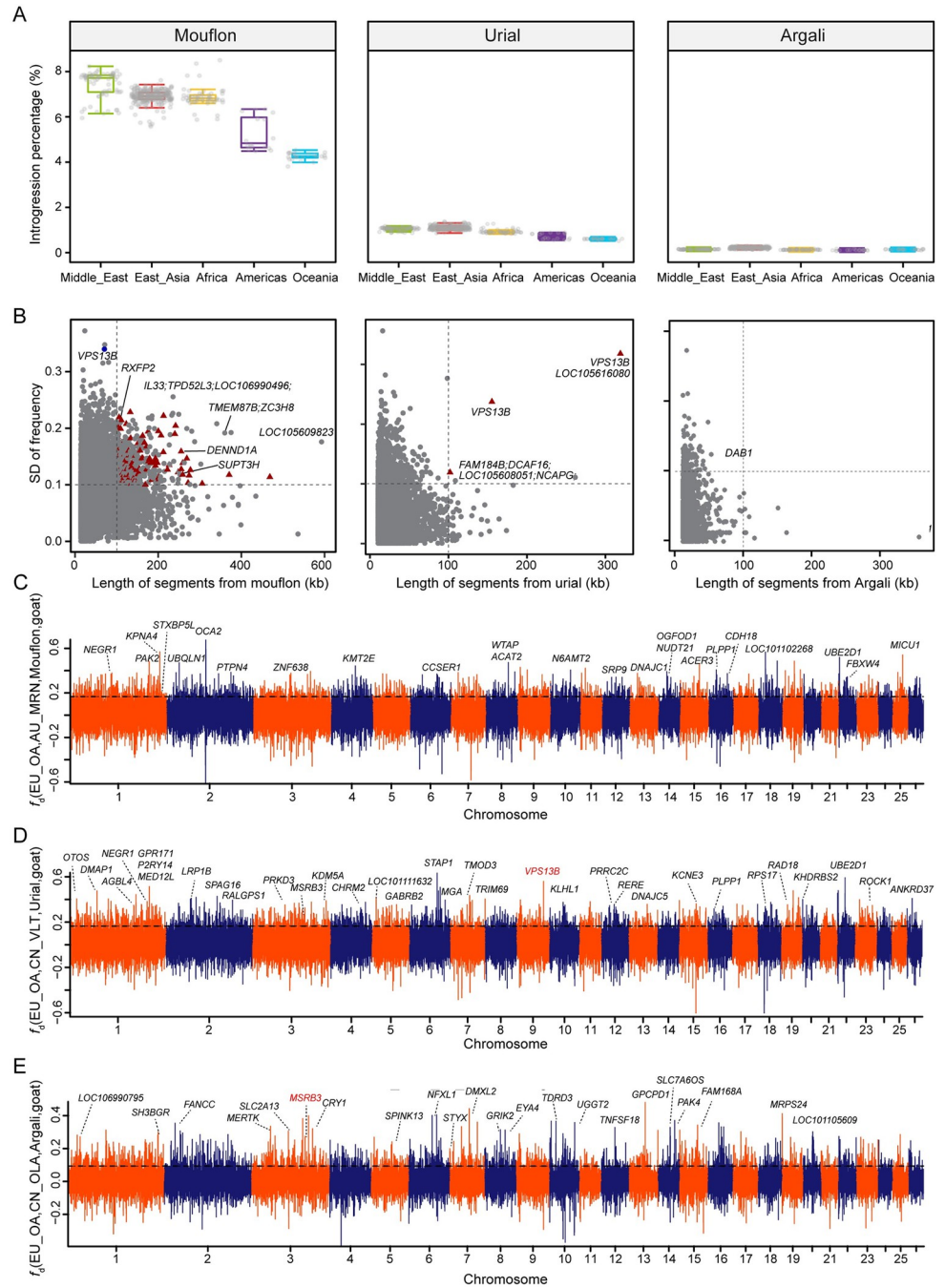


Fig 3. Genome-wide introgression evaluation and potential adaptive introgression. (A) The proportions of introgressed sequences from wild relatives (Iranian mouflon, urial and argali) identified in each domestic population. Each dot indicates a phased haploid. (B) Joint distribution of length for introgressed tracks (x axis), SD for introgressed haplotype frequency among distinct populations (y axis). Red triangles indicate tracts with significant high F_{ST} values ($P < 0.001$, Z test) between at least one of the 16 domestic meta-populations and Iranian mouflon. (C-E) Manhattan plot of values showing the introgression signals from Iranian mouflon to Australian Merino (C), from Urial to Valley Tibetan sheep (D) and from Argali to Tibetan Oula sheep (E). The horizontal dashed line indicated the $P < 0.05$ cutoff.

<https://doi.org/10.1371/journal.pgen.1010615.g003>

estimated d_{XY} , phylogenetic trees and haplotype networks to corroborate the signals of introgression in specific regions.

Selection and adaptive signatures for introgressed segments

We focused on those introgressed haplotype blocks that are conserved within but not across populations, since they are most likely involved in population differentiation and adaptation to local habitats or selection [2]. For this, we calculated allele frequencies of the introgressed fragments in 16 domestic meta-populations (see [Materials and Methods](#)). Next, we defined 483 mouflon, 5 urial and no argali outlier haplotypes, putatively introgressed on the basis of their length (≥ 100 kb), their total frequency (≥ 0.05) and frequency variation in the 16 meta-populations (> 0.1 standard deviation) ([Fig 3B](#)).

In order to detect fragments that are significantly differentiated between domestic and wild populations, we plotted F_{ST} for each of the 16 meta-populations to the Iranian mouflon across the genome in 50-kb windows with a 20-kb step size ([S5 Fig](#)). With a $P < 0.001$ (Z test) F_{ST} cut-off and joining the windows that were separated by a distance of ≤ 50 kb, we obtained 2,305 non-overlapping regions. These blocks were slightly but significantly longer than the general blocks ([S9 Fig](#)). For the 488 mouflon and urial introgression outliers above, 116 and 3 overlapped with these highly differentiated regions ([Fig 3B](#)) and were here studied in more detail.

As expected, introgressed haplotype blocks are unevenly distributed among the domestic populations with a clear geographic signal. For instance, in region chr2: 109,998,387–110,183,036, the frequency of the Iranian mouflon derived haplotype is high in Iran local breeds (0.60) and Tan sheep (0.61), but very low (0.00) in Australian Merino and several Chinese breeds such as Hu sheep, Ujimqin Sheep, and Valley Tibetan sheep. The longest introgressed urial haplotype chr9:77,117,407–77,437,296 has a high frequency in Tibetan sheep including Oula (0.75), Prairie (0.90) and Valley Tibetan (0.80) and is almost entirely absent in sheep from Africa, America, Oceania and the Middle East. Overall, these putatively introgressed regions contained 891 genes, of which 883 and 8 were within the haplotypes derived from Iranian mouflon and Urial, respectively. It is noteworthy that within subset of introgressed regions that we identify as highly differentiated regions, several genes have been associated with morphological traits, particularly in facial shape. *RXFP2* was strongly associated with sheep horn morphology [26,27], and *SUPT3H* was reported to be associated with nose bridge breadth [28], nose morphology [29], chin dimples [30] and forehead protrusion [31]. *MSTRB3* had been identified as a candidate gene for external ear morphology in pig, dog, goat and sheep [32–39]. Furthermore, several other genes (e.g., *STXBP5L*, *DENND1A*, *VPS13B*) were identified in GWAS studies of human facial shape analyses [29,31,40].

Introgressed *RXFP2* affects horn status

There are three main types of horn status in sheep (1) horned males and females (“horned”); (2) horned males, polled females (“sex-specific”); (3) polled males and females (“polled”) [41,42]. A previous study indicated that the “horned” haplotypes in Tibetan sheep within *RXFP2* was most likely introgressed from argali [11]. However, in our study the same region (chr10: 29,435,112–29,481,215) was detected as introgression from Iranian mouflon not argali ([Fig 3B](#)). Furthermore, we found that this introgressed region showed significant high F_{ST} ($P < 0.001$) in breeds with different horn status, and arose long-stretched LD block ([Figs 4A, S10, and S11](#)). LAI indicates that most haplotypes in breeds with horn status (3) contain haplotypes most closely related to those of Iranian mouflon ([Fig 4B–4D](#)), pointing to a possible origin of this phenotype from this wild sheep species. We further investigated in detail the pattern of haplotype variation in this region.

Haplotype patterns in this region across all 1,167 sheep showed three major highly divergent haplogroups (hap-a, hap-b and hap-c), with a few other diverse or recombinant haplotypes at low frequency ([Figs 4C and S12 Fig](#)). Hap-a is the dominant haplotype in domestic

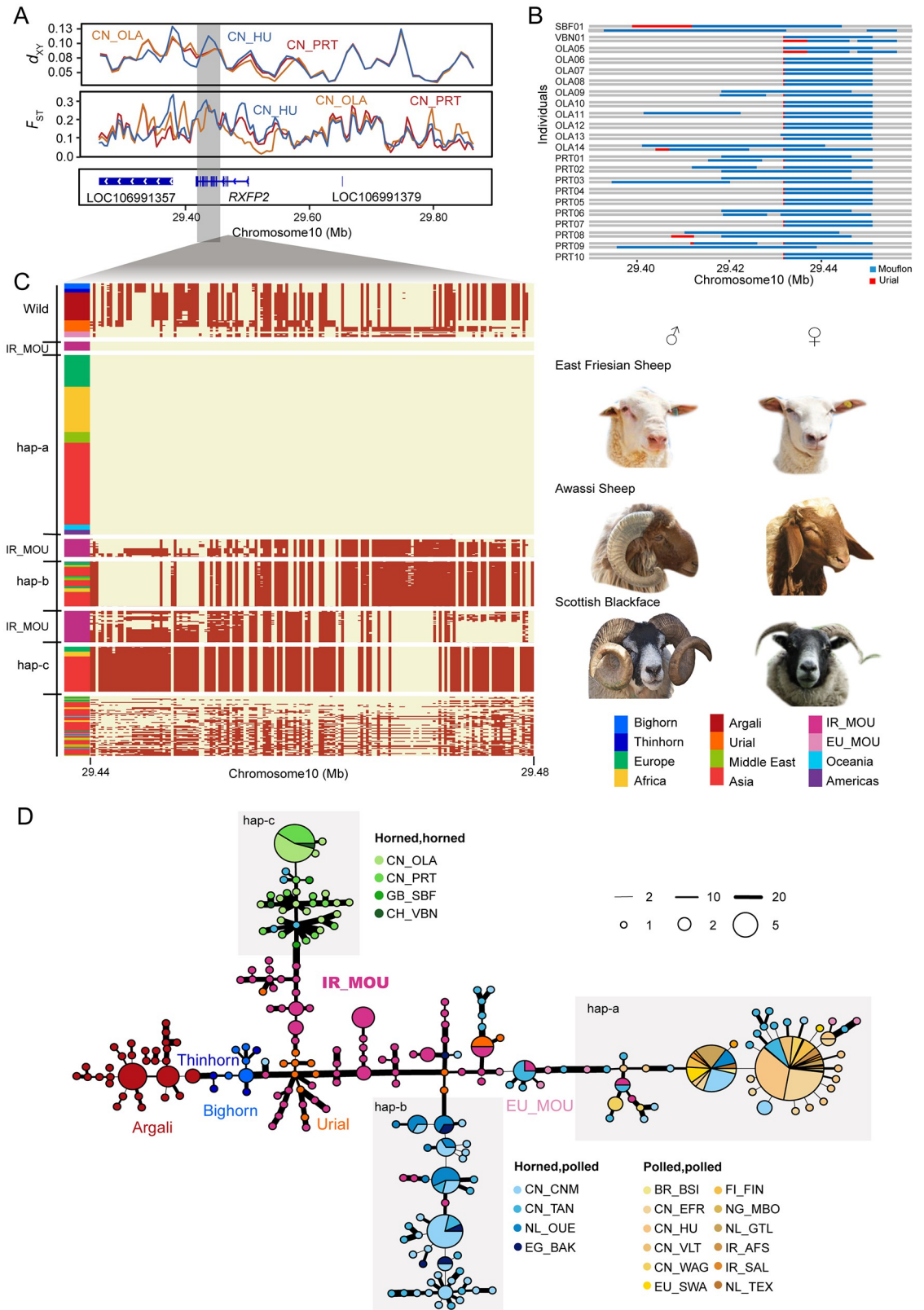


Fig 4. Identification and haplotype resolution of introgressed fragment at the *RXFP2* locus. (A) Distributions of mean pairwise sequence divergence (d_{xy}) values and pairwise fixation index (F_{ST}) values calculated by Iranian mouflon and domestic sheep

populations for each 50-kb window. Gene annotations in the selected region in Oar_v4.0 are indicated at the bottom. (B) LAI within *RXFP2* in Valais Blacknose, Scottish Blackface, Oula and Prairie Tibetan sheep illustrating mosaic patterns of source population. (C) The haplotype patterns of *RXFP2* introgressed region (Oar_v4.0 chr10: 29,436,086–29,466,717). Each row is a phased haplotype, and each column is a polymorphic SNP variant. The reference and alternative alleles are indicated by light yellow and red, respectively. The haplotypes present in Iranian mouflons are indicated separately. Photo credits are shown in S7 Table. (D) A haplotype network generated by the R software package PEGAS based on 221 SNPs of 666 haplotypes.

<https://doi.org/10.1371/journal.pgen.1010615.g004>

sheep (Figs 4C and S12) and is completely fixed (frequency = 1) in Finnsheep (n = 12), Gotland (n = 10), Waggir (n = 9), Afshari (n = 6), and East Friesian sheep (n = 10) (S13 Fig and S4 Table), all of which have the “polled” phenotype. Intriguingly, hap-a is present as heterozygotes in two Iranian mouflon samples (Fig 4C). We speculated that polledness had likely occurred in wild sheep progenitors, possibly as recessive trait, and rapidly became widespread in domestic sheep because it was under strong selection in a domesticated setting.

Hap-b is generally found at high frequency in breeds with the “sex-specific” horn phenotype, including Chinese Merino (25/40, 0.625), Ouessant (12/20, 0.6) and Barki sheep (5/6, 0.83). Hap-c in contrast is usually at high frequency in breeds with the “horned” phenotype, including Oula, Prairie Tibetan, Valais Blacknose and Scottish Blackface sheep (S13 Fig).

Fig 4D shows a network of intact non-recombined haplotypes in the ~46-kb region around *RXFP2* from wild and domestic sheep. The network suggests that haplogroups corresponding to Hap-a, Hap-b and Hap-c, respectively, are all linked to haplotypes that occur in Iranian mouflon. Two possible explanations remained either by introgression from Iranian mouflon or variation that has persisted since domestication. The former was more likely as the probability to be ILS for a 46,103 bp haplotype was extremely low (2.03×10^{-5}). In the network and in the ML tree (S14 Fig), Hap-a and Hap-b haplotypes of mouflon are intermingled with those of urial, so the introgressed fragments possibly originated from urial and were introgressed into domestic sheep via the mouflons. The Valais Blacknose and Scottish Blackface (European domestic breeds) haplotypes were assigned to the “horned” phenotype cluster, validating the earlier introgressed time for this locus as well. Nucleotide difference between the Iranian mouflon haplotypes and hap-c (Fig 2B) suggest that hap-c was introgressed from a mouflon sub-population of Asiatic mouflon that has not yet been sampled or were the new mutations since the time of introgression. In spite of these facts, there was still possibility that the direct ancestor of domestic sheep had all three haplotypes (a, b and c), because it is really difficult to distinguish introgression from ILS. Analysis of non-silent mutations (S15 Fig) did not reveal a single causative mutation, but variant chr10: 29,439,011 has the highest correlation with the phenotype. A ~1.8-kb insertion in the 3' UTR region of *RXFP2* has also been identified to be a putatively causal mutation for horn status [43,44]. But it was not completely linked with horn phenotypes in all the domestic breeds or populations [43]. More efforts are needed to construct the relationship between these mutations and horn status, and to explore how these variations regulate horn status.

Ear morphology influenced by introgressed *MSRB3*

Another prominent introgressed region with high F_{ST} contains *MSRB3*, encoding methionine sulfoxide reductase B3 (Figs 5A–5C and S16). Interestingly, ear morphology has been mapped to *MSRB3* in sheep using breeds fixed for divergent ear types [39], designated as ear size (large-eared vs. small-eared) and ear erectness (drop-eared vs. prick-eared). This gene yielded significant f_d values in 9 pairwise comparisons of argali vs. domestic population, encompassing chr3:154,000,001–154,090,000 (Fig 5A). This was confirmed by the absolute divergence d_{XY} of argali and Oula, and of argali and Prairie Tibetan populations. (Fig 5B), which indicated introgression rather than shared ancestry (ILS) [24]. By contrast, the d_{XY} of Iranian mouflon and

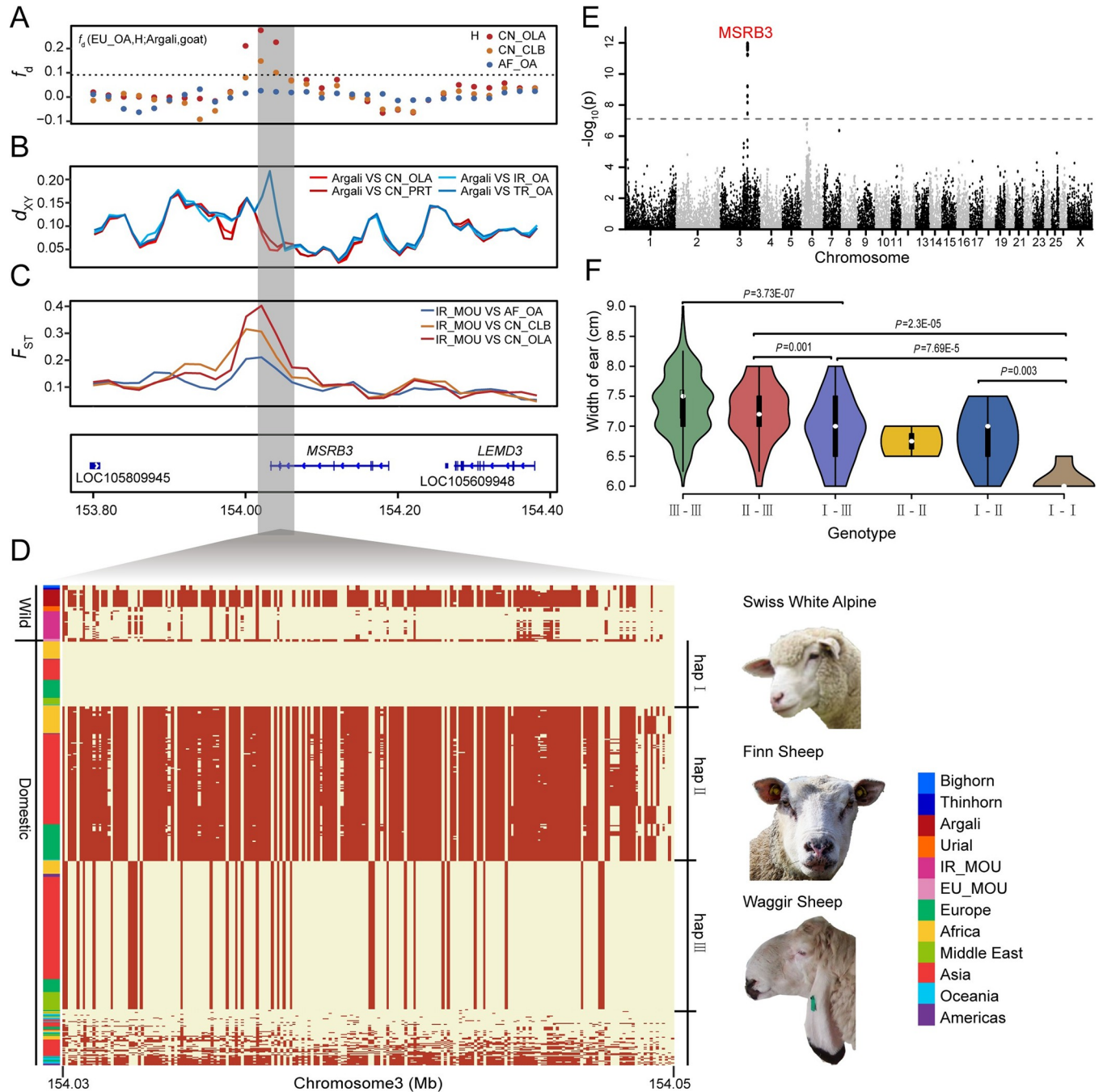


Fig 5. Identification and functional annotation of introgression segment at the *MSRB3* locus from argali to domestic sheep. (A-C) Distributions of f_d (EU_OA, H; Argali, Goat), d_{XY} values and F_{ST} surrounding the introgressed region (Oar_v4.0 chr3: 153,800,001–154,380,001). The horizontal dashed line in f_d track indicate the significance cutoff ($P < 0.001$, Z test). (D) Haplotype pattern in the potential introgressed region (chr3: 154,030,048–154,062,195) of *MSRB3* gene. Each row is a phased haplotype, and each column is a polymorphic SNP variant. The reference and alternative allele are indicated in light yellow and red, respectively. Hap-II was the introgressed haplotype from argali. Photo credits are showed in S7 Table. (E) GWAS $-\log_{10} P$ values for the width of the external ear of East-Friesian \times Hu F2 crossbreds are plotted against position on the chromosomes. The gray horizontal dashed lines indicate the genome-wide significance threshold of the GWAS (7.72×10^{-8}). (F) The violin plots refer to width of external ear for East-Friesian \times Hu hybrids with different genotypes. III-III refers to homozygote of hap III defined in Fig 5D, the other five genotypes are denoted accordingly.

<https://doi.org/10.1371/journal.pgen.1010615.g005>

either Oula or Prairie Tibetan populations are elevated (**S17 Fig**), indicating that the phylogenetic relationship of this region deviates from the phylogeny of the *Ovis* species.

A haplotype plot of the ~32-kb *MSRB3* region across 1,167 individuals (**Figs 5D and S18**) group into three main haplogroups, denoted as hap-I, hap-II and hap-III. These three haplogroups were corroborated by haplotype network and ML tree, in which domestic sheep haplotypes assigned to three clusters (**S20 and S21 Figs**). The hap-II cluster is close to argali haplotypes (**S20 and S21 Figs**), consistent with it being introgressed from argali. Furthermore, hap-II has the highest frequency in domestic sheep (1109/1831, 0.605), and is fixed in Finn-sheep ($n = 24$), Hanzhong ($n = 10$), Tibetan Oula ($n = 28$), Feral ($n = 6$) and Old Spael sheep, and nearly fixed (≥ 0.95) in Cameroon, Gotland, Ouessant sheep (**S19 Fig and S5 Table**). Due to its high frequency among sheep breeds, we speculated this haplogroup was likely to confer an adaptive advantage over the other two groups. Intriguingly, all the European mouflon ($n = 3$) in this study were likewise fixed for hap-II, suggesting that this introgression probably occurred before the first wave of migration of sheep into European breeds [20,45]. The frequency of hap-I is relatively low across all domestic sheep (155/1831, 0.085), but has a high frequency in Swiss White Alpine (8/8, 1), Mossi sheep (4/6, 0.67) and Diqing sheep (14/18, 0.78), all of which generally have small ears. Hap-III is found in breeds with exceptionally large and floppy ears (567/1831, 0.309), including Waggir (18/18, 1), Karakul (6/6, 1) and Duolang (63/68, 0.926), but also in Solognote (3/16, 0.188), Shetland (3/14, 0.214), Norwegian White (3/6, 0.5), Drente Heath (4/8, 0.5), East Friesian (15/20, 0.75) and Texel (6/6, 1) sheep that have small ears, suggesting that in addition to *MSRB3* other genes are involved in ear morphology.

For a more controlled analysis of a link between *MSRB3* variants and ear size, we used an F2 East-Friesian \times Hu sheep hybrid population. We performed genome-wide association study (GWAS) of all the external ear traits, including measured width and length, in F2 hybrids ($n = 323$) (**Figs 5E and S22**). The analysis of ear width revealed a single significant association peak located in *MSRB3* (**Fig 5E and S6 Table**), but there was no significant signal associated with ear length (**S22 Fig**) or the other ear traits. Crossbred individuals with different haplotype combinations (**Fig 5F**) or different genotypes of diagnostic SNPs (**S23 and S24 Figs**) displayed significant difference in ear width.

Complex patterns of introgressed regions within *VPS13B*

Another strong introgression signal was found in *VPS13B* (vacuolar protein sorting 13 homolog B), which showed the most significant introgressed signals from urial according to LAI (**Figs 3B and S25**), as well as several consecutive outlier windows in the top f_d values ($P < 0.001$) (**Figs 3D and 6A**). *VPS13B* is a large gene spanning about 800 kb and has a complex structure with 50 exons and 6 alternatively spliced transcripts. It encodes a large protein with more than 4000 amino acids. The LAI results showed that there were two urial introgressed regions located in *VPS13B* chr9:77,117,407–77,437,296 (319.8 kb) and chr9:77,511,156–77,666,735 (155.5 kb) (**Fig 3B**), comprising 4 and 3 major haplogroups respectively and covering about 59% of the gene (**Figs 6E, 6F, S27 and S28**). In addition, there is another separate introgression signal derived from mouflon in this gene chr9:76,946,737–77,016,847, with three haplogroups (**Figs 3B, 6D and S26**). The haplogroups in these three regions form five major haplotype combinations, at least one of which is a recombinant (**Fig 6D–6F**), and the dominant haplotype in the first region is tightly linked to one of the five more downstream haplogroup combinations. The compound introgressed haplotype appears to have high frequency in Tibetan sheep (Oula: 0.9, Prairie Tibetan: 1; Valley Tibetan: 0.75), while at low frequency in domestic sheep from Iran (0.125), Turkey (0.045), America (0.063) and Australia (0).

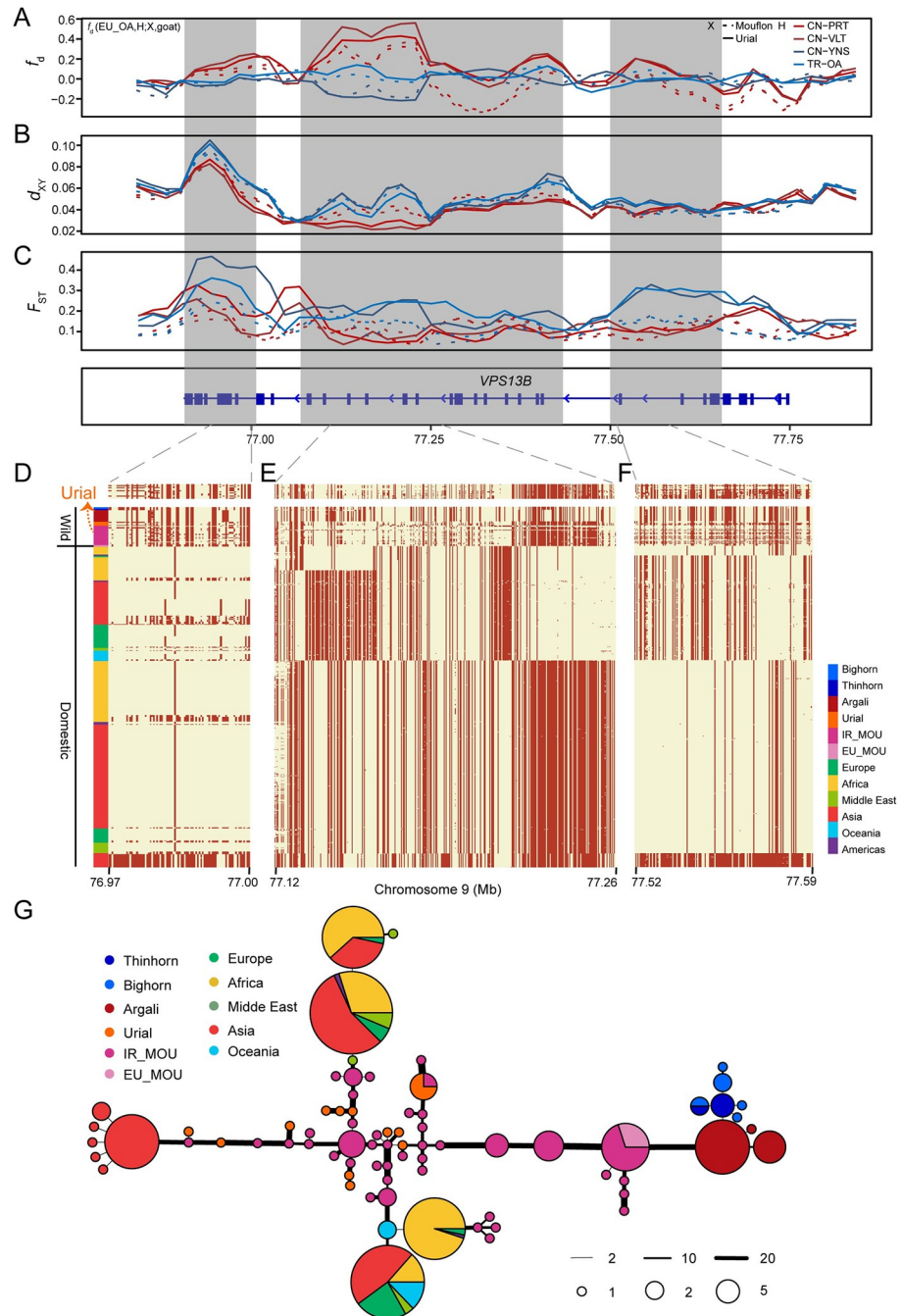


Fig 6. Signals of introgression in the *VPS13B* gene. (A-C) Distributions of f_d (EU_OA, X; Urrial, Goat), d_{XY} , and F_{ST} extending the three introgressed regions. The gene structures of *VPS13B* are indicated at the bottom track. (D-F) The patterns of haplotype sharing for part of introgressed regions. Each row is a phased haplotype, and each column is a polymorphic SNP variant. The reference and alternative alleles are indicated in light yellow and red, respectively.

<https://doi.org/10.1371/journal.pgen.1010615.g006>

These signals were also supported by d_{XY} and F_{ST} values which were lower between introgressed haplotypes and urrial than mouflon, despite the closer phylogenetic position of the later to domestic sheep (Figs 6B, 6C and S29). Whereas, this pattern was almost undetectable in partial introgressed region. We built haplotype networks of each region to investigate in detail

the donor of introgressed haplotypes, but due to intermixed haplotypes we cannot distinguish whether the donor was urial or Iranian mouflon (Figs 6G and S30–S32). Similar to *RXFP2*, urial and Asiatic mouflon probably share the *VPS13B* haplotypes, which precludes an identification of the origin of the introgressions into domestic sheep.

VPS13B is functionally relevant to numerous phenotypes and diseases. It causes Cohen syndrome in humans with diverse manifestations including microcephaly, craniofacial and limb anomalies. In addition, variations in *VPS13B* affect face morphology, particularly nose morphology in human and mice [31]. Although the role of the introgressed fragments in *VPS13B* in sheep cannot at this stage be functionally verified, observations in other species suggest it may play a role in the development of facial shape.

Discussion

In the present study, we performed a detailed investigation of introgression in sheep and evaluated the amounts of sequence introgressed from each wild relative into domestic sheep. We have focused on the most consequential breed-specific variants by selection of fragments of >100 kb with a significant high F_{ST} (Z test, $P < 0.001$). We also present an in-depth investigation of three regions containing the genes *RXFP2*, *MSRB3* and *VPS13B*, which have been introgressed from wild sheep and now occur in a substantial proportion of the global sheep population. We show how these haplotypes are associated with variation in several morphological variations in domestic sheep.

Wild-domestic introgressions

Consistent with the previous studies, we detected the pronounced gene flow from wild relative to domestic populations, and the pattern that average proportions of wild relative sequence decrease with the phylogenetic distance between wild sheep species and domestic sheep [5,16]. Whereas, the location of introgressed fragments and the range of introgression percentages showed differences between this and previous studies [5,16], we speculated such differences were mainly caused by the distinct statistics or software that we used to identify introgressions. Besides, the focused breeds and their geographic distribution were not exactly the same.

The strong signal of early sympatric gene flow of the Iranian mouflon into ancestral domestic sheep is geographically plausible and explains the high proportion of mouflon-derived sequences in domestic sheep. Domestic sheep have acquired urial DNA segments either directly from sheep breeds in the eastern distribution range of the urial (Fig 1) or indirectly via the Iranian mouflon population [46]. Subsequent dispersal has brought domestic sheep into contact with argali. The introgression from snow sheep had also been proved previously [47]. A considerable genetic overlap of Asiatic mouflon and urial [47,48] indicated incomplete speciation and/or mutual introgression. This has resulted in an incomplete differentiation of these species and does not allow a clear differentiation of Asiatic mouflon and urial as source of introgression of *RXFP2* and *VPS13B*.

Accurate identification of donor species may depend on the availability of whole genome sequencing (WGS) data from wild species candidates, and on method used to infer it. Hu et al. proposed argali introgression into *RXFP2*, but did not test the Asiatic mouflon [11]. Our data, especially the haplotype network (Fig 4D), clearly indicate that, although the argali haplotype does resemble the introgressed haplotype (hap-c), hap-c has much closer affinity with haplotypes in Iranian mouflon. Moreover, hap-c is actually geographically widespread among domestic sheep, being found in both Tibetan, European and African sheep breeds, which has not been shown before (Figs 4D and S12 and S4 Table). This supports that introgression of

this haplotype predated the global dispersal from the sheep domestication rather than much later and localized to the Tibetan Plateau.

Long divergent haplotypes contribute to diversity of sheep

We found that introgressed wild haplotypes covered about 8% of the sheep genome, and therefore contributed substantially to the diversity of domestic sheep, on the level of either individual or breed-specific variation. As indicated by Fig 3B, we focused on a small proportion of all introgressed regions, but fragments that are shorter than 100 kb have a more random distribution across the breeds (low SD of within-breed frequencies) and do not appear to have high F_{ST} between all 16 breed groups and Asiatic mouflon. Despite this, they may still contribute to the overall diversity of domestic sheep.

Breed-specific introgression may well be related to local adaptation through their link to sheep phenotypes, e.g. hypoxia responses and high-altitude adaptation [11,16,23], resistance to pneumonia [5] and reproduction [49]. It would be a reasonable expectation that traits resulting from human selection [50,51] were only indirectly influenced by wild introgression, such as the different wool [52] and tail types [53]. However, the absence of horns, a typical domestic feature, corresponding to *RXFP2* haplotypes is also detected in “horned” Asiatic mouflon. A testable hypothesis is that *RXFP2* of wild sheep is involved in balanced selection controlling the size of the horns.

A common feature of this study and comparable studies of cattle and goats is the observation of introgressed long (50 kb or longer) divergent haplotypes [4,13,54–56]. Divergence of homologous sequences inhibits recombination [57–60] which explains the absence of intermediates of the diverged haplotypes and allows to retain the divergence of the haplotypes. Structural variants (SVs) residing in these long extended introgressed haplotypes have played an important role in local adaptation in human [61–63]. More analyses were necessary in sheep and other domestic animals to dig up the introgressed adaptive SVs.

In conclusion, using whole-genome sequencing data of large-scale individuals, we clarified the phylogenetic relationship among the eight extant species in the *Ovis* genus. In addition, we generated a global admixture graph of wild relative in diverse domestic sheep populations and determined whether positive selection had acted on these fragments. We also highlighted three introgressive regions in *RXFP2*, *MSRB3* and *VPS13B*. Through detailed haplotype and functional analyses, we evaluated the role of long divergent haplotypes from wild relatives in shaping the morphological traits of domestic sheep, which may be a ubiquitous phenomenon in animal evolution.

Materials and methods

Ethics statement

Blood samples were taken by conforming with the Helsinki Declaration of 1975 (as revised in 2008) concerning Animal Rights, and this study was reviewed and approved by the Animal Ethical and Welfare Committee (DK2021019), Northwest A&F University, China.

Sample collection

We newly sequenced 156 samples of WGS comprising 147 domestic sheep (*O. aries*) and 9 wild relatives (7 argali [*O. ammon*], 1 urial [*O. vignei*] and 1 European mouflon [*O. musimon*]) (Fig 1 and S1 and S2 Tables). Following standard library preparation protocols, we used at least 0.5 µg of genomic DNA for each sample to construct paired-end library with insert sizes from 300 to 500 bp. Sequencing was performed on the Illumina HiSeq X Ten platform with a

mean coverage of 13.30×. WGS data for 60 wild species (1 snow sheep [*O. nivicola*], 3 bighorn [*O. Canadensis*], 2 thinhorn [*Ovis dalli*], 13 argali, 6 urial, 33 Asiatic mouflon [*Ovis orientalis*] and 2 European mouflon), and 951 domestic individuals were obtained from previous studies [11,23,48,51,64–68] (NCBI <https://www.ncbi.nlm.nih.gov/>, Nextgen <http://projects.ensembl.org/nextgen/>). The domestic samples originated from 154 different breeds with a geographic origin from Asia to the Middle East, Europe, Africa, Oceania and America (S1 Table).

Read alignment and variant calling

We firstly removed low-quality sequence reads of combined dataset by TRIMOMMATIC v.0.39 [69]. Next, we aligned cleaned reads to Oar v.4.0 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000298735.2) using the program BURROWS-WHEELER ALIGNER v.0.7.17 (BWA-MEM) algorithm [70] with the default parameters. Duplicate reads were excluded using PICARD Markduplicates and bam files were sorted using PICARD SORTSAM (Picard v2.18.2 <http://broadinstitute.github.io/picard/>). Then, Genome Analysis Toolkit (GATK version 4.2.0.0) [71] was performed to realign the reads around indels with REALIGNERTARGETCREATOR and INDELREALIGNER modules. To obtain the candidate SNPs from bam files, we used the workflow adapted from GATK HAPLOTYPECALLER to create genomic variant call format (gVCF) file for each sample. After merging all gVCF files, we implemented following criteria to SNPs using GATK VariantFiltration to avoid false-positive calls " Quality by Depth (QD) <2.0 || FS (Fisher Strand) > 60.0 || MQRankSum (MappingQualityRankSumTest) <-12.5 || ReadPosRankSum (ReadPosRankSumTest) <-8.0 || SOR (StrandOddsRatio) >3.0 || MQ (root mean square of Mapping Quality) <40.0 ". SNPs not meeting the following criterias were further excluded: (1) biallelic variation; (2) missing rate < 0.1; (3) mean reads depth (DP) > 1/3× and < 3×. For remaining SNPs, imputation and phasing were simultaneous performed using BEAGLE v4.1 (Browning and Browning 2007; Browning and Browning 2016) with default parameters. SNPs and indels were annotated using the software ANNOVAR [72].

Population structure and phylogenetic analysis

We analyzed the population structure using 293 representative samples (S3 Table), including all wild species and 56 domestic breeds. We used genome-wide 332,990 fourfold degenerate (4DV) sites to construct a maximum likelihood (ML) phylogenetic tree using RAXML v8.2.9 [73] with the following parameters: `-f a -x 123 -p 23 -# 100 -k -m GTRGAMMA` ('f a' to run a search for the ML tree and a rapid bootstrap analysis in one run, '-x' a random number seed for the ML search, '-p' a random number seed for the parsimony inference, '-#' the number of bootstrap, '-m' substitution models). The robustness of specific tree topology was tested by 100 bootstraps. The final tree topology was visualized using INTERACTIVE TREE OF LIFE (iTOL) [74], and rooted at the branch of goat (Figs 2A and S1).

Principal component analysis (PCA) of whole-genome SNPs using was performed with the SMARTPCA program in the package of EIGENSOFT v.6.0beta [75]. To clarify the relationship between wild populations and domestic sheep, we performed four separate PCA using different dataset: (1) 293 samples, with the first two principal components cumulatively explaining 20.91% of the total variance. (2) 267 individuals including 3 European mouflon, 31 Asiatic mouflon and 233 domestic sheep; (3) 233 domestic individuals sampling from eight different regions, including Africa, Americans, Australian, China, South Asia, European, Iran and Turkey; (4) 117 individuals from 11 Chinese breeds (Fig 2B–2D).

We used ADMIXTURE v1.23 [76] to infer $K = 2$ to $K = 9$ clusters of related individuals to estimate the ancestry of each individual and quantify genome-wide admixture. For each K , we

ran ADMIXTURE 20 times and calculated the mean cross-validation (CV) error to determine the optimal group number, the minimum CV value among 20 repetitions of each K was taken as the final result (Figs 2F and S3).

Selective sweep analysis

To detect potential selective signals, we calculated F_{ST} in the pairwise comparisons between Iranian mouflon and each of the domestic populations. The 233 domestic sheep were divided into 16 groups according to breeds and region of origin: EU_OA (Europe), AM_OA (America), AF_OA (Africa), TR_OA (Turkey), IR_OA (Iran), CN_YNS (Yunnan), CN_WZM (Ujimqin), CN_TAN (Tan), CN_STH (Small-tailed Han), CN_PRT (Prairie Tibetan), CN_VLT (Valley Tibetan), CN_OLA (Tibetan Oula sheep), CN_HU (Hu sheep), CN_CLB (Cele Black sheep), CN_BYK (Bayinbuluke sheep) and AU_MRN (Australian Merino). F_{ST} was calculated in 50-kb sliding windows with 20-kb step size (S11, S16 and S29 Figs) using vcfTools v.0.1.13 [77]. In each comparison, the top 1% genomic regions with the highest scores overlapped were considered to be potential selective signatures. We performed Z test, and then focused putatively selected regions on the windows with a significance level of $P < 0.001$. To further verify whether this threshold was appropriate, permutations were performed in F_{ST} analysis (S5 Fig). We first randomly selected 33 individuals from all the domestic sheep dataset and mixed these samples with 33 Iranian mouflon together, and then randomly divided them into two new populations. Next, we calculated windowed F_{ST} values between these two new populations (in sliding 50-kb windows with 20-kb steps). The maximum value across all windows were recorded and this process was repeated 100 times. Finally, we sorted the 100 maximum values from largest to smallest and selected the largest value (0.155) as the final permutation result.

Whole-genome analysis of genomic introgression

Estimation of introgression on population scale. We implemented *D*-statistics with DSUITE [78] across all combinations of the 16 species/populations defines as described above. The species/population tree required for DSUITE was constructed using Treemix [79] without assuming gene flow (-m 0) using goat as outgroup (S4 Fig). Then, *D* and *f*₄-ratios of all populations were calculated with the DTRIOS module, and the results for each chromosome were combined with Dtrioscombine module. After that, *D* and *f* statistics were calculated for each branch of the population tree using the FBRANCH module, and visualized the statistical results using dtools.py script provided in the DSUITE software (S4 Fig). Because there were few alleles sharing between domestic sheep from Europe and the other wild species and domestic populations (S4 Fig), domestic European samples were identified as non-introgressed reference population.

Identification and localization of genomic introgression. We used local ancestry inference (LAI) implemented in LOTER [80], which uses phased data and has been shown to outperform other tools for more ancient admixture. We specified seven wild relatives (1 snow sheep, 3 bighorn and 2 thinhorn, 12 argali, 6 urial, 31 Iranian mouflon, 3 European mouflon), and European domestic sheep as reference population, in which European domestic sheep ($n = 30$) was the control population for domestic component. It was assumed that a haplotype of an admixed domestic individual consists of a mosaic of existing haplotypes from the eight reference populations. For each fragment, LOTER derives the most likely ancestral origin on the basis of allele frequencies of reference populations and the selected populations. We calculated introgression percentages from each of the wild relatives into the haploid genomes (Fig 3A) and merge overlapping introgressed regions from the same source. Then, the frequencies

in the 16 domestic groups with their standard deviations (SD) and ranges (max-frequency minus min-frequency) were calculated for each selected fragment (Fig 3B).

f_d in sliding windows. We computed the modified f -statistic (f_d) value [24] using a 50-kb sliding window with 20-kb step size in the form of f_d (EU_OA, domestic populations; wild species, goat), where EU_OA represents the European domestic sheep ($n = 30$) and domestic populations include 16 populations described above. We evaluated the statistical significance using two-tailed Z-test. We calculated the P values according to Z-transformed f_d values, and the windows with $P < 0.001$ was defined as potential introgressed regions (Fig 3C–3E). Mean pairwise sequence divergence (d_{XY}) [24] was also calculated for 50-kb windows with 20-kb steps across whole genome using same populations above (Figs 5B, 6B and S17).

Incomplete lineage sorting (ILS). In order to exclude common ancestry as explanation for the presence of introgressed fragments, we calculated the expected length of ancestral sequence shared by domestic sheep and each wild relative, respectively. The expected shared ancestral sequence length (L) is calculated as $L = 1/(r/t)$, in which r is the recombination rate per generation per base pair (bp), and t is the length between wild relatives and domestic sheep since divergence. The probability of a length of at least m is $1 - \text{GammaCDF}(m, \text{shape} = 2, \text{rate} = 1/L)$, in which GammaCDF is the Gamma distribution function and the numbers within the parenthesis are its arguments [9]. We used a generation time of 4 years [81], a recombination rate of 1.0×10^{-8} per base pair (bp) per generation [27] and the following divergence times: 0.032 Ma between Iranian mouflon and domestic sheep, 1.26 Ma for urial and domestic sheep, 2.36 Ma for argali and domestic sheep, and 3.12 Ma for bighorn (or thinhorn) and domestic sheep [17,23,82–83]. This gives expected lengths of L (Iranian mouflon) = 6,192 bp, L (uria) = 159 bp, L (argali) = 85 bp, and L (bighorn/thinhorn) = 64 bp. We then removed inferred introgressed fragments shorter than L , and calculated the total length of remaining introgressed tracks. The length distributions are showed in S7 and S8 Figs. Probabilities of length of observed introgressed regions were calculated by the R function `pgamma` using the local recombination rates estimated in previous study [84]. The probabilities are 2.03×10^{-5} for 46.10 kb (*RXFP2* introgressed region) and zero for 31.70 kb (*MSRB3* argali-introgressed region), 319.89 and 155.58 (*VPS13B* urial-introgressed regions), and 1.37×10^{-5} for 70.11 kb (*VPS13B* mouflon-introgressed) (S8 Table).

Haplotype patterns and network. To view the specific genotypes patterns of the prominent introgressed regions, including *RXFP2*, *MSRB3* and *VPS13B*, we extracted the phased SNPs in these regions from 1,167 whole-genome sequencing individuals and visualized specific genotypes patterns in a heatmap (S12, S18 and S26–S28 Figs). We also constructed haplotype networks of *RXFP2*, *MSRB3* and *VPX13B* using R package PEGAS [85] based on the pairwise differences (Figs 4D, 6G, S21, and S30–S32). We screened and eliminated samples whose haplotypes were interrupted due to recombination and removed SNPs with minor allele frequency $\leq 5\%$. In the 46.3 kb *RXFP2* introgressed region we retained 333 samples from 6 wild species and 20 domestic sheep breeds and 221 SNPs. In the 23 kb *MSRB3* introgressed region we retained 202 SNPs in 201 individuals. We analyzed the sheep ear shapes of the corresponding varieties of three haplotypes, and defined three main haplotypes as hap-I, hap-II and hap-III.

Genome-wide association study. From an East-Friesian sheep \times Hu Sheep F2 generation bred by Gansu Yuansheng Agriculture and Animal Husbandry Technology Co.,Ltd. (Jinchang, Gansu) 323 samples were collected. Phenotypes include ear length and width, birth weight and age. DNA was collected from blood samples and whole genomes were sequenced by Shijiazhuang Boruidi Biotechnology Co., Ltd (Shijiazhuang, Hebei) using 40K liquid chip generated by genotyping by target sequencing [86]. Raw fastq files were filtered using `fastp` [87], reads were mapped to `Oar_rambouillet_v1.0` and the variation were summarized in a VCF file. We

used PLINK1.9 [88] to remove samples with > 10% missing genotypes and SNPs with minor allele frequency < 0.05 and >10% missing scores, retaining 317 sheep and 209,625 SNPs. To improve variant density, we used BEAGLE5.0 [89] to impute genotype using reference panel size of 43 East Friesen sheep and 8 Hu sheep with default settings and removed SNP with DR² (dosage R-squared) ≤ 0.8, resulting in a total of 647,471 SNPs.

GWAS was conducted using GEMMA(0.98.3) [90] with the linear mixed model:

$$y = W\alpha + x\beta + u + \varepsilon$$

y is the phenotype of $n \times 1$ vector; W the $n \times c$ matrix of covariates including fixed effects; α the c -vector of the corresponding coefficients including the intercept; x is the n -vector of markers; β the effect size of the markers; u the n -vector of the random effect with $u \sim \text{MVN}_n(0, \lambda\tau^{-1}K)$; MVN_n the n -dimensional multivariate normal distribution; λ the ratio between the two variance components; τ^{-1} is the variance of the residual errors; K represents the known $n \times n$ relatedness matrix calculated by SNP markers; ε the random error n -vector with $\varepsilon \sim \text{MVN}_n(0, \tau^{-1}I_n)$, where I_n denotes $n \times n$ identity matrix. To decrease false positive signals, the genome-wide significance threshold was set to be 7.72×10^{-8} ($0.05/647,471$) after the Bonferroni correction.

Supporting information

S1 Fig. ML tree of domestic sheep and their seven wild relatives (293 individuals) using genome-wide 332,990 fourfold degenerate (4DV) sites. Goat was used as outgroup. (TIF)

S2 Fig. Phylogeography between Iranian mouflon and domestic sheep. The map shows the geographic distribution of 33 Iranian mouflon samples, which can be divided into two groups according to their geographical distribution, western Iran (IRW_MOU) and eastern Iran (IRE_MOU). https://d-maps.com/carte.php?num_car=5494&lang=zh (TIF)

S3 Fig. ADMIXTURE results of 293 worldwide sheep. (a) ADMIXTURE results for $k = 2$ to $k = 9$. For each k value, the run with the lowest cross validation (CV) error out of 20 replicates is plotted. The number of samples and population names are listed both at the top and bottom. (b) CV error for varying k in the ADMIXTURE analysis. (TIF)

S4 Fig. (a) The species/population tree constructed by Treemix ignored gene flow, and goat was used as outgroup. (b) Results of Fbranch in Dsuite for wild sheep species and different domestic sheep populations. The species/populations tree is shown along the left and upper sides, and the tree on the y axis is displayed in “expanded” form. The color-block in the matrix refer to the excess allele sharing between the branch identified on the expanded tree on the y axis and the species/populations on the x-axis. The darker the color means the higher ratio in allele sharing, the lighter the ratio is lower. (TIF)

S5 Fig. Distribution of the pairwise fixation index (F_{ST}) values between Iranian mouflon and domestic sheep populations for each 50-kb window. In addition, *MSRB3* (chromosome 3), *VPS13B* (chromosome 9) and *RXFP2* (chromosome 10) were also highlight in purple. The blue dashed lines show the F_{ST} threshold ($P < 0.001$, Z test), and the shallow blue dashed lines show the permutation threshold with 100 bootstrap. (TIF)

S6 Fig. Distribution of total length of segments inferred to be introgressive from different wild relatives into per haploid. Incomplete lineage sorting (ILS) is precluded. Different color means domestic sheep populations originated from divergent region.

(TIF)

S7 Fig. Length distribution of introgressed fragments from different wild relatives to domestic sheep. The vertical black lines indicate the mean segments length.

(TIF)

S8 Fig. The total length of introgressed fragments from Iranian mouflon across different geographic populations of domestic sheep. (a) The indicated P values are based on t-test. (b) Different colors represent different geographic regions displayed in a.

(TIF)

S9 Fig. The length of introgressed segments from different wild relatives. Purple: all introgressed fragments; green: segments with F_{ST} values above the threshold. The **** represents significant difference ($P < 0.05$).

(TIF)

S10 Fig. The length distribution of LD blocks on chromosomes 3, 9 and 10. Red dots indicate the blocks located in introgressed segments of *MSRB3*, *VPS13B* and *RXFP2*.

(TIF)

S11 Fig. Distributions of F_{ST} values surrounding *RXFP2* gene between Iranian mouflon and domestic sheep populations. The F_{ST} was calculated in 50-kb sliding window with 20-kb step size. Triangles beside the population labels indicate that the population showed selective signals (F_{ST} top 1%) in this region. The grey box means the location of windows showing selective signal.

(TIF)

S12 Fig. Haplotype patterns of the introgressed region (chr10: 29,435,112–29,481,215) at the *RXFP2* locus in 1,167 sheep based on 221 SNPs (MAF > 0.05). Each column represents a SNP variant, and each row represents a phased haplotype. Yellow predicted alleles identical to reference genome; red predicted alleles differ to reference genome.

(TIF)

S13 Fig. Frequency distribution of different haplotypes among various domestic sheep breeds in the introgressed *RXFP2* region (10:29,435,112–29,481,215).

(TIF)

S14 Fig. ML tree of introgressed region located in *RXFP2*. The ML tree was built using 221 SNPs with minor allele frequency (MAF) > 0.05 from 29,435,112 to 29,481,215 on chromosome 10 with 100 bootstraps.

(TIF)

S15 Fig. Nonsynonymous SNPs residing in *RXFP2* region. The species or population names are listed at the bottom, the corresponding horn types of domestic sheep are showed on the top.

(TIF)

S16 Fig. Distribution of F_{ST} values in the region of *MSRB3* gene between Iranian mouflon and domestic sheep populations. The F_{ST} was calculated in 50-kb sliding window with 20-kb step size. Triangles beside the population labels indicate that the population showed selective signals (F_{ST} top 1%) in *MSRB3*. The grey box means the location of windows showing selective signal.

(TIF)

S17 Fig. The distribution of mean sequence divergence (d_{XY}) surrounding the introgressed region in *MSRB3* between Iranian mouflon and domestic sheep populations. Introgressed region (chr3: 154,030,492–154,062,195) are gray-shaded. The d_{XY} values between Iranian mouflon and Oula (CN-OLA)/ Prairie Tibetan (CN-PRT) showed a marked increase in the introgressed region, compared with Iran (IR-OA) and Turkey (TR-OA) sheep. The d_{XY} values are calculated in 50-kb sliding window with 20-kb step size.
(TIF)

S18 Fig. Haplotype patterns of the introgressed region at the *MSRB3* locus (chr3: 154,030,762–154,053,023) in 1,167 sheep based on 202 SNPs. Each column represents a SNP variant, and each row represents a phased haplotype. Yellow predicted alleles identical to reference genome; red predicted alleles differ to reference genome.
(TIF)

S19 Fig. Frequency distribution of different haplotypes among domestic sheep species in introgressive region of *MSRB3* (chr3: 154,030,492–154,062,195).
(TIF)

S20 Fig. ML tree of introgressed region located in *MSRB3*. The ML tree was built using 327 SNPs with minor allele frequency ($MAF > 0.05$) from 154,030,492 to 154,053,023 on chromosome 3 by 100 bootstraps.
(TIF)

S21 Fig. A haplotype network based on 202 SNPs ($MAF > 0.05$) of *MSRB3* introgressed region (chr3: 154,030,762–154,053,023). Different color shows different wild species or different regional sources of domestic sheep. Three major haplotypes and some haplotypes with lower frequencies in domestic sheep were identified. The R software package PEGAS were used to generate the network.
(TIF)

S22 Fig. Genome-wide association analysis of external ear length using the F2 hybrids of East Friesian×Hu sheep. The gray horizontal dashed line indicates the significance threshold of the GWAS ($P = 7.72e-08$).
(TIF)

S23 Fig. Genotype patterns within *MSRB3* for F2 hybrids from East Friesian×Hu sheep. These SNPs are genotyped by target sequencing. Each column indicates a variation significantly associated with ear width, III_III refers to homozygous hapIII; other genotypes are denoted accordingly. Homozygous reference, heterozygous variant and homozygous variant are indicated in light beige, orange and brick red, respectively.
(TIF)

S24 Fig. Phenotypic differences among different genotypes of SNPs (top 7) significantly associated with external ear width. The x axial represents different genotypes, and the ordinate represents the ear width of the corresponding samples. The SNP on position chr3:154,039,306 is nonsynonymous.
(TIF)

S25 Fig. Schematic representation of introgressed haplotypes within *VPS13B* in Valley Tibetan, Wuzhumuqin, Small Tail Han, Tan sheep, Oula, Prairie Tibetan sheep, Tibetan sheep, Yunnan sheep, local breeds from Iran and Turkey, Bayinbuluke and Cele Black

sheep, which illustrates mosaic patterns of source population inferred by LOTER software. (TIF)

S26 Fig. Haplotype patterns of fragments (chr9:76,974,289–77,004,488) introgressed from mouflon at *VPS13B* locus in 1,167 sheep based on 84 SNPs (MAF > 0.05). Each column represents a SNP variant, and each row represents a phased haplotype. The different color strips on the left show different groups. Yellow and red indicates the reference and the alternative alleles, respectively.

(TIF)

S27 Fig. Haplotype patterns of fragments (chr9: 77,117,452–77,259,733) introgressed from urial at *VPS13B* locus in 1,167 sheep based on 560 SNPs (MAF > 0.05). Each column represents a SNP variant, and each row represents a phased haplotype. The different color strips on the left show different groups. Yellow and red indicates the reference and the alternative alleles, respectively.

(TIF)

S28 Fig. Haplotype patterns of fragments (chr9: 77,516,138–77,585,410) introgressed from urial at *VPS13B* locus in 1,167 sheep based on 165 SNPs (MAF > 0.05). Each column represents a SNP variant, and each row represents a phased haplotype. The different color strips on the left show different groups. Yellow and red indicates the reference and the alternative alleles, respectively.

(TIF)

S29 Fig. Distribution of F_{ST} values in the region of *VPS13B* gene between Iranian mouflon and domestic sheep populations. The F_{ST} was calculated in 50-kb sliding window with 20-kb step size. Triangles beside the population labels indicate that the population showed selective signals (F_{ST} top 1%) in *VPS13B*. The grey box means the location of windows showing selective signal.

(TIF)

S30 Fig. A haplotype network based on 85 SNPs (MAF > 0.05) of *VPS13B* introgressed region (chr9: 76,978,748–77,012,389). Different color shows different wild species or different regional sources of domestic sheep. The R software package PEGAS were used to generate the network.

(TIF)

S31 Fig. A haplotype network based on 121 SNPs (MAF > 0.05) of *VPS13B* introgressed region (chr9: 77,219,494–77,235,075). Different color shows different wild species or different regional sources of domestic sheep. The R software package PEGAS were used to generate the network.

(TIF)

S32 Fig. A haplotype network based on 165 SNPs (MAF > 0.05) of *VPS13B* introgressed region (chr9: 77,516,138–77,585,410). Different color shows different wild species or different regional sources of domestic sheep. The R software package PEGAS were used to generate the network.

(TIF)

S1 Table. Summary information of worldwide 1,167 wild and domestic sheep.

(XLSX)

S2 Table. Sampling information of 156 domestic and wild sheep individuals in this study.
(XLSX)

S3 Table. Sampling information of 293 individuals used in phylogenetic analysis and genome-wide introgression estimation.
(XLSX)

S4 Table. Numbers of different haplotypes of introgressed region in *RXFP2* gene among various domestic sheep breeds.
(XLSX)

S5 Table. Numbers of different haplotypes of introgressed region in *MSRB3* gene among various domestic sheep breeds.
(XLSX)

S6 Table. Significant loci in GWAS of ear width.
(XLSX)

S7 Table. The map base layers and photos were downloaded with the Creative Commons license or taken by ourself.
(XLSX)

S8 Table. The simulations of ILS length using local recombination rates surrounding *RXFP2*, *MSRB3* and *VPS13B*. Length is the observed length of the shared haplotype, and ILS is the expected length estimated using local recombination rate given no introgression. Probability is the P value for how likely the observed distance would be under a no introgression scenario, which are calculated using Gamma distribution function $1 - \text{GammaCDF}(m, \text{shape} = 2, \text{rate} = 1/L)$.
(XLSX)

Acknowledgments

We thank the High-Performance Computing platform of Northwest A&F University for providing computing resources. We express our thanks to the owners of the sheep for providing samples (for sampling information see [S2 Table](#)).

Author Contributions

Conceptualization: Yu Jiang.

Data curation: Mingjun Liu.

Formal analysis: Hong Cheng, Zhuangbiao Zhang, Jiayue Wen.

Funding acquisition: Yuxuan Song, Mingjun Liu, Yu Jiang.

Investigation: Zhuangbiao Zhang, Yingwei Guo, Jintao Wang, Junjie Shao.

Methodology: Hong Cheng, Zhuangbiao Zhang, Jiayue Wen, Yudong Cai, Ming Li, Ran Li.

Project administration: Yu Jiang.

Resources: Zhuangbiao Zhang, Ming Li, Wenrong Li, Sangang He, Yuxuan Song, Lei Zhang, Mingjun Liu.

Software: Hong Cheng, Zhuangbiao Zhang, Jiayue Wen, Yudong Cai, Yingwei Guo, Ming Li, Masum Billah.

Supervision: Yu Jiang.

Visualization: Zhuangbiao Zhang, Jiayue Wen.

Writing – original draft: Hong Cheng.

Writing – review & editing: Hong Cheng, Zhuangbiao Zhang, Jiayue Wen, Johannes A. Lenstra, Rasmus Heller, Ran Li, Xihong Wang, Mingjun Liu, Yu Jiang.

References

1. Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. Evidence for archaic adaptive introgression in humans. *Nat Rev Genet.* 2015; 16:359–371. <https://doi.org/10.1038/nrg3936> PMID: 25963373
2. Janzen GM, Wang L, Hufford MB. The extent of adaptive wild introgression in crops. *New Phytol.* 2019; 221:1279–1288. <https://doi.org/10.1111/nph.15457> PMID: 30368812
3. Mendez L, Fernando, Watkins C, Joseph, F Hammer, Michael. A Haplotype at *STAT2* introgressed from Neanderthals and serves as a candidate of positive selection in Papua New Guinea. *Am J Hum Genet.* 2012; 91:265–274. <https://doi.org/10.1016/j.ajhg.2012.06.015> PMID: 22883142
4. Zheng Z, Wang X, Li M, Li Y, Yang Z, Wang X, et al. The origin of domestication genes in goats. *Sci Adv.* 2020; 6:eaa5216. <https://doi.org/10.1126/sciadv.aaz5216> PMID: 32671210
5. Cao YH, Xu SS, Shen M, Chen ZH, Gao L, Lv FH, et al. Historical introgression from wild relatives enhanced climatic adaptation and resistance to pneumonia in Sheep. *Mol Biol Evol.* 2021; 38:838–855. <https://doi.org/10.1093/molbev/msaa236> PMID: 32941615
6. Baiz MD, Wood AW, Brelsford A, Lovette IJ, Toews DPL. Pigmentation genes show evidence of repeated divergence and multiple bouts of introgression in *Setophaga* Warblers. *Curr Biol.* 2021; 31:643–649. <https://doi.org/10.1016/j.cub.2020.10.094> PMID: 33259789.
7. Vernot B, Akey JM. Resurrecting surviving Neandertal lineages from modern human genomes. *Science.* 2014; 343:1017–1021. <https://doi.org/10.1126/science.1245938> PMID: 24476670.
8. Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, et al. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature.* 2014; 507:354–357. <https://doi.org/10.1038/nature12961> PMID: 24476815
9. Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature.* 2014; 512:194–197. <https://doi.org/10.1038/nature13408> PMID: 25043035
10. Ai H, Fang X, Yang B, Huang Z, Chen H, Mao L, et al. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat Genet.* 2015; 47:217–225. <https://doi.org/10.1038/ng.3199> PMID: 25621459.
11. Hu XJ, Yang J, Xie XL, Lv FH, Cao YH, Li WR, et al. The genome landscape of tibetan sheep reveals adaptive introgression from Argali and the history of early human settlements on the Qinghai-Tibetan plateau. *Mol Biol Evol.* 2019; 36:283–303. <https://doi.org/10.1093/molbev/msy208> PMID: 30445533
12. Khrameeva EE, Bozek K, He L, Yan Z, Jiang X, Wei Y, et al. Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans. *Nat Commun.* 2014; 5:3584. <https://doi.org/10.1038/ncomms4584> PMID: 24690587
13. Wu DD, Ding XD, Wang S, Wojcik JM, Zhang Y, Tokarska M, et al. Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nat Ecol Evol.* 2018; 2:1139–1145. <https://doi.org/10.1038/s41559-018-0562-y> PMID: 29784979.
14. Barbato M, Hailer F, Orozco-Terwengel P, Kijas J, Mereu P, Cabras P, et al. Genomic signatures of adaptive introgression from European mouflon into domestic sheep. *Sci Rep.* 2017; 7:7623. <https://doi.org/10.1038/s41598-017-07382-7> PMID: 28790322
15. Li R, Yang P, Li M, Fang W, Yue X, Nanaei HA, et al. A Hu sheep genome with the first ovine Y chromosome reveal introgression history after sheep domestication. *Sci China Life Sci.* 2020; 64:1116–1130. <https://doi.org/10.1007/s11427-020-1807-0> PMID: 32997330
16. Lv F-H, Cao Y-H, Liu G-J, Luo L-Y, Lu R, Liu M-J, et al. Whole-genome resequencing of worldwide wild and domestic sheep elucidates genetic diversity, introgression, and agronomically important loci. *Mol Biol Evol.* 2022;39:msab353. <https://doi.org/10.1093/molbev/msab353> PMID: 34893856
17. Zeder MA. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc Natl Acad Sci U S A.* 2008; 105:11597–11604. <https://doi.org/10.1073/pnas.0801317105> PMID: 18697943
18. Mason IL. Evolution of domesticated animals. Longman, London and New York.1984.

19. Scherf BD. World watch list for domestic animal diversity. 3rd ed. Food and Agriculture Organization (FAO); 2000.
20. Lv FH, Peng WF, Yang J, Zhao YX, Li WR, Liu MJ, et al. Mitogenomic meta-analysis identifies two phases of migration in the history of eastern eurasian sheep. *Mol Biol Evol.* 2015; 32:2515–2533. <https://doi.org/10.1093/molbev/msv139> PMID: 26085518
21. Singh S, Kumar S Jr., Kolte AP, Kumar S. Extensive variation and sub-structuring in lineage A mtDNA in Indian sheep: genetic evidence for domestication of sheep in India. *PLoS One.* 2013; 8:e77858. <https://doi.org/10.1371/journal.pone.0077858> PMID: 24244282
22. Zhao YX, Yang J, Lv FH, Hu XJ, Xie XL, Zhang M, et al. Genomic reconstruction of the history of native sheep reveals the peopling patterns of nomads and the expansion of early pastoralism in east Asia. *Mol Biol Evol.* 2017; 34:2380–2395. <https://doi.org/10.1093/molbev/msx181> PMID: 28645168
23. Rezaei HR, Naderi S, Chintauan-Marquier IC, Taberlet P, Virk AT, Naghash HR, et al. Evolution and taxonomy of the wild species of the *genus Ovis* (*Mammalia, Artiodactyla, Bovidae*). *Mol Phylogenet Evol.* 2010; 54:315–326. <https://doi.org/10.1016/j.ympev.2009.10.037> PMID: 19897045
24. Martin SH, Davey JW, Jiggins CD. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Mol Biol Evol.* 2015; 32:244–257. <https://doi.org/10.1093/molbev/msu269> PMID: 25246699
25. Teng H, Zhang Y, Shi C, Mao F, Cai W, Lu L, et al. Population genomics reveals speciation and introgression between Brown Norway Rats and their sibling species. *Mol Biol Evol.* 2017; 34:2214–2228. <https://doi.org/10.1093/molbev/msx157> PMID: 28482038
26. Johnston SE, Mcewan JC, Pickering NK, Kijas JW, Beraldi D, Pilkington JG, et al. Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Mol Ecol.* 2011; 20:2555–2566. <https://doi.org/10.1111/j.1365-294X.2011.05076.x> PMID: 21651634
27. Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, et al. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol.* 2012; 10:e1001258. <https://doi.org/10.1371/journal.pbio.1001258> PMID: 22346734
28. Adhikari K, Reales G, Smith AJP, Konka E, Palmen J, Quinto-Sanchez M, et al. A genome-wide association study identifies multiple loci for variation in human ear morphology. *Nat Commun.* 2015; 6:7500. <https://doi.org/10.1038/ncomms8500> PMID: 26105758
29. Claes P, Roosenboom J, White JD, Swigut T, Sero D, Li J, et al. Genome-wide mapping of global-to-local genetic effects on human facial shape. *Nat Genet.* 2018; 50:414–423. <https://doi.org/10.1038/s41588-018-0057-4> PMID: 29459680
30. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet.* 2016; 48:709–717. <https://doi.org/10.1038/ng.3570> PMID: 27182965
31. Bonfante B, Faux P, Navarro N, Mendoza-Revilla J, Dubied M, Montillot C, et al. A GWAS in Latin Americans identifies novel face shape loci, implicating VPS13B and a Denisovan introgressed region in facial variation. *Sci Adv.* 2021; 7:eabc6160. <https://doi.org/10.1126/sciadv.abc6160> PMID: 33547071
32. Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, et al. A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol.* 2010; 8:e1000451. <https://doi.org/10.1371/journal.pbio.1000451> PMID: 20711490
33. Webster MT, Kamgari N, Perloski M, Hoepfner MP, Axelsson E, Hedhammar A, et al. Linked genetic variants on chromosome 10 control ear morphology and body mass among dog breeds. *BMC Genomics.* 2015; 16:474. <https://doi.org/10.1186/s12864-015-1702-2> PMID: 26100605
34. Wei C, Wang H, Liu G, Wu M, Cao J, Liu Z, et al. Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds. *BMC Genomics.* 2015; 16:194. <https://doi.org/10.1186/s12864-015-1384-9> PMID: 25888314
35. Zhang LC, Liang J, Pu L, Zhang YB, Wang LG, Liu X, et al. mRNA and protein expression levels of four candidate genes for ear size in Erhualian and Large White pigs. *Genet Mol Res.* 2017; 16: gmr16029252. <https://doi.org/10.4238/gmr16029252> PMID: 28407177.
36. Zhang Y, Liang J, Zhang L, Wang L, Liu X, Yan H, et al. Porcine methionine sulfoxide reductase B3: molecular cloning, tissue-specific expression profiles, and polymorphisms associated with ear size in *Sus scrofa*. *J Anim Sci Biotechnol.* 2015; 6:60. <https://doi.org/10.1186/s40104-015-0060-x> PMID: 26719797
37. Chen N, Cai Y, Chen Q, Li R, Wang K, Huang Y, et al. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat Commun.* 2018; 9:2337. <https://doi.org/10.1038/s41467-018-04737-0> PMID: 29904051

38. Kumar C, Song S, Dewani P, Kumar M, Parkash O, Ma Y, et al. Population structure, genetic diversity and selection signatures within seven indigenous Pakistani goat populations. *Anim Genet.* 2018; 49:592–604. <https://doi.org/10.1111/age.12722> PMID: 30229969
39. Paris JM, Letko A, Häfliger IM, Ammann P, Drögemüller C. Ear type in sheep is associated with the *MSRB3* locus. *Anim Genet.* 2020; 51:968–972. <https://doi.org/10.1111/age.12994> PMID: 32805068
40. White JD, Indencleef K, Naqvi S, Eller RJ, Hoskens H, Roosenboom J, et al. Insights into the genetic architecture of the human face. *Nat Genet.* 2021; 53:45–53. <https://doi.org/10.1038/s41588-020-00741-7> PMID: 33288918
41. Darlington CD. The origin of iso-chromosomes. *J Hered.* 1940; 39:351–361. <https://doi.org/10.1007/bf02982848>
42. Dolling CHS. Hornedness and polledness in sheep. IV. Triple alleles affecting horn development in the Merino. *Aust J Agric Res.* 1961; 12:535–361. <https://doi.org/10.1071/ar9610535>
43. Lühken G, Krebs S, Rothhammer S, Küpper J, Mioč B, Russ I, et al. The 1.78-kb insertion in the 3'-untranslated region of *RXFP2* does not segregate with horn status in sheep breeds with variable horn status. *Genet Sel Evol.* 2016; 48(1):78. <https://doi.org/10.1186/s12711-016-0256-3> PMID: 27760516
44. Wiedemar N, Drogemüller C. A 1.8-kb insertion in the 3'-UTR of *RXFP2* is associated with polledness in sheep. *Anim Genet.* 2015; 46:457–461. <https://doi.org/10.1111/age.12309> PMID: 26103004
45. Chessa B, Pereira F, Arnaud F, Amorim A, Goyache F, Mainland I, et al. Revealing the history of sheep domestication using retrovirus integrations. *Science.* 2009; 324:532–536. <https://doi.org/10.1126/science.1170587> PMID: 19390051
46. Demirci S, Koban Baştanlar E, Dağtaş ND, Pişkin E, Engin A, Özer F, et al. Mitochondrial DNA Diversity of modern, ancient and wild sheep (*Ovis gmelinii anatolica*) from Turkey: new insights on the evolutionary history of sheep. *PLoS One.* 2013; 8:e81952. <https://doi.org/10.1371/journal.pone.0081952> PMID: 24349158
47. Chen ZH, Xu YX, Xie XL, Wang DF, Aguilar-Gómez D, Liu GJ, et al. Whole-genome sequence analysis unveils different origins of European and Asiatic mouflon and domestication-related genes in sheep. *Commun Biol.* 2021; 4:1307. <https://doi.org/10.1038/s42003-021-02817-4> PMID: 34795381
48. Deng J, Xie XL, Wang DF, Zhao C, Lv FH, Li X, et al. Paternal origins and migratory episodes of domestic sheep. *Curr Biol.* 2020; 30:4085–4095. <https://doi.org/10.1016/j.cub.2020.07.077> PMID: 32822607
49. Liu Z, Ji Z, Wang G, Chao T, Hou L, Wang J. Genome-wide analysis reveals signatures of selection for important traits in domestic sheep from different ecoregions. *BMC Genomics.* 2016; 17: 863. <https://doi.org/10.1186/s12864-016-3212-2> PMID: 27809776
50. Xu SS, Li MH. Recent advances in understanding genetic variants associated with economically important traits in sheep (*Ovis aries*) revealed by high-throughput screening technologies. *Front Agr Sci Eng.* 2017; 4:279–288. <https://doi.org/10.15302/j-fase-2017151>
51. Li X, Yang J, Shen M, Xie XL, Liu GJ, Xu YX, et al. Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nat Commun.* 2020; 11:2815. <https://doi.org/10.1038/s41467-020-16485-1> PMID: 32499537
52. Jackson N, Maddocks IG, Watts JE, Scobie D, Mason RS, Gordon-Thomson C, et al. Evolution of the sheep coat: the impact of domestication on its structure and development. *Genet Res (Camb).* 2020; 102:e4. <https://doi.org/10.1017/S0016672320000063> PMID: 32517826
53. Kalds P, Luo Q, Sun K, Zhou S, Chen Y, Wang X. Trends towards revealing the genetic architecture of sheep tail patterning: promising genes and investigatory pathways. *Anim Genet.* 2021; 52:799–812. <https://doi.org/10.1111/age.13133> PMID: 34472112
54. Mei C, Wang H, Liao Q, Wang L, Cheng G, Wang H, et al. Genetic architecture and selection of chinese cattle revealed by whole genome resequencing. *Mol Biol Evol.* 2018; 35:688–699. <https://doi.org/10.1093/molbev/msx322> PMID: 29294071
55. Chen Q, Shen J, Hanif Q, Chen N, Huang Y, Dang R, et al. Whole genome analyses revealed genomic difference between European taurine and East Asian taurine. *J Anim Breed Genet.* 2021; 138:56–68. <https://doi.org/10.1111/jbg.12501> PMID: 32770713
56. Via S. Divergence hitchhiking and the spread of genomic isolation during ecological speciation with gene flow. *Philos Trans R Soc Lond B Biol Sci.* 2012; 367:451–460. <https://doi.org/10.1098/rstb.2011.0260> PMID: 22201174
57. Metzberg AB, Wurzer G, Huisman TH, Smithies O. Homology requirements for unequal crossing over in humans. *Genetics.* 1991; 128:143–161. <https://doi.org/10.1093/genetics/128.1.143> PMID: 2060774
58. Opperman R, Emmanuel E, Levy AA. The effect of sequence divergence on recombination between direct repeats in *Arabidopsis*. *Genetics.* 2004; 168:2207–2215. <https://doi.org/10.1534/genetics.104.032896> PMID: 15611187

59. Dreissig S, Maurer A, Sharma R, Milne L, Flavell AJ, Schmutzer T, et al. Natural variation in meiotic recombination rate shapes introgression patterns in intraspecific hybrids between wild and domesticated barley. *New Phytol.* 2020; 228:1852–1863. <https://doi.org/10.1111/nph.16810> PMID: 32659029
60. Veller C, Edelman NB, Muralidhar P, Nowak MA. Recombination and selection against introgressed DNA. *bioRxiv.* 2021. <https://doi.org/10.1101/846147>
61. Yan SM, Sherman RM, Taylor DJ, Nair DR, Bortvin AN, Schatz MC, et al. Local adaptation and archaic introgression shape global diversity at human structural variant loci. *Elife.* 2021; 10:e67615. <https://doi.org/10.7554/eLife.67615> PMID: 34528508
62. Hsieh P, Vollger MR, Dang V, Porubsky D, Baker C, Cantsilieris S, et al. Adaptive archaic introgression of copy number variants and the discovery of previously unknown human genes. *Science.* 2019; 366:eaax2083. <https://doi.org/10.1126/science.aax2083> PMID: 31624180
63. Quan C, Li Y, Liu X, Wang Y, Ping J, Lu Y, et al. Characterization of structural variation in Tibetans reveals new evidence of high-altitude adaptation and introgression. *Genome Biol.* 2021; 22:159. <https://doi.org/10.1186/s13059-021-02382-3> PMID: 34034800
64. Alberto FJ, Boyer F, Orozco-terWengel P, Streeter I, Servin B, de Villemereuil P, et al. Convergent genomic signatures of domestication in sheep and goats. *Nat Commun.* 2018; 9:813. <https://doi.org/10.1038/s41467-018-03206-y> PMID: 29511174
65. Naval-Sanchez M, Nguyen Q, McWilliam S, Porto-Neto LR, Tellam R, Vuocolo T, et al. Sheep genome functional annotation reveals proximal regulatory elements contributed to the evolution of modern breeds. *Nat Commun.* 2018; 9:859. <https://doi.org/10.1038/s41467-017-02809-1> PMID: 29491421
66. Pan Z, Li S, Liu Q, Wang Z, Zhou Z, Di R, et al. Whole-genome sequences of 89 Chinese sheep suggest role of *RXFP2* in the development of unique horn phenotype as response to semi-feralization. *Giga-science.* 2018; 7:gjy019. <https://doi.org/10.1093/gigascience/gjy019> PMID: 29668959
67. Wang X, Liu J, Niu Y, Li Y, Zhou S, Li C, et al. Low incidence of SNVs and indels in trio genomes of Cas9-mediated multiplex edited sheep. *BMC Genomics.* 2018; 19:397. <https://doi.org/10.1186/s12864-018-4712-z> PMID: 29801435
68. Upadhyay M, Hauser A, Kunz E, Krebs S, Blum H, Dotsev A, et al. The first draft genome assembly of snow sheep (*Ovis nivicola*). *Genome Biol Evol.* 2020; 12:1330–1336. <https://doi.org/10.1093/gbe/evaa124> PMID: 32592471
69. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014; 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
70. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2010; 26:589–595. <https://doi.org/10.1093/bioinformatics/btp698> PMID: 20080505
71. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010; 20:1297–1303. <https://doi.org/10.1101/gr.107524.110> PMID: 20644199
72. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010; 38(16):e164. <https://doi.org/10.1093/nar/gkq603> PMID: 20601685
73. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014; 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033> PMID: 24451623
74. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 2016; 44:W242–W245. <https://doi.org/10.1093/nar/gkw290> PMID: 27095192
75. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006; 2:e190. <https://doi.org/10.1371/journal.pgen.0020190> PMID: 17194218
76. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009; 19:1655–1664. <https://doi.org/10.1101/gr.094052.109> PMID: 19648217
77. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 2011; 27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330> PMID: 21653522
78. Malinsky M, Matschiner M, Svardal H. Dsuite—Fast D-statistics and related admixture evidence from VCF files. *Mol Ecol Resour.* 2021; 21:584–595. <https://doi.org/10.1111/1755-0998.13265> PMID: 33012121
79. Bunch TD, Wu C, Zhang Y-P, Wang S. Phylogenetic analysis of Snow Sheep (*Ovis nivicola*) and closely related Taxa. *J Hered.* 2006; 97:21–30. <https://doi.org/10.1093/jhered/esi127> PMID: 16267166
80. Dias-Alves T, Mairal J, Blum MGB. Loter: A software package to infer local ancestry for a wide range of species. *Mol Biol Evol.* 2018; 35:2318–2326. <https://doi.org/10.1093/molbev/msy126> PMID: 29931083

81. Guerrini M, Forcina G, Panayides P, Lorenzini R, Garel M, Anayiotos P, et al. Molecular DNA identity of the mouflon of Cyprus (*Ovis orientalis ophion*, Bovidae): near eastern origin and divergence from Western Mediterranean conspecific populations. *Syst Biodivers*. 2015; 13:472–483. <https://doi.org/10.1080/14772000.2015.1046409>
82. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet*. 2012; 8:e1002967. <https://doi.org/10.1371/journal.pgen.1002967> PMID: 23166502
83. Yang J, Li WR, Lv FH, He SG, Tian SL, Peng WF, et al. Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Mol Biol Evol*. 2016; 33:2576–2592. <https://doi.org/10.1093/molbev/msw129> PMID: 27401233
84. Petit M, Astruc JM, Sarry J, Drouilhet L, Fabre S, Moreno CR, et al. Variation in recombination rate and its genetic determinism in sheep populations. *Genetics*. 2017; 207:767–784. <https://doi.org/10.1534/genetics.117.300123> PMID: 28978774
85. Paradis E. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*. 2010; 26:419–420. <https://doi.org/10.1093/bioinformatics/btp696> PMID: 20080509
86. Guo Y, Bai F, Wang J, Fu S, Zhang Y, Liu X, et al. Design and characterization of a high-resolution multiple-SNP capture array by target sequencing for sheep. *J Anim Sci*. 2023;101. <https://doi.org/10.1093/jas/skac383> PMID: 36402741
87. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018; 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560> PMID: 30423086
88. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81:559–575. <https://doi.org/10.1086/519795> PMID: 17701901
89. Browning BL, Zhou Y, Browning SR. A one-penny imputed genome from next-generation reference panels. *Am J Hum Genet*. 2018; 103:338–348. <https://doi.org/10.1016/j.ajhg.2018.07.015> PMID: 30100085
90. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet*. 2012; 44:821–824. <https://doi.org/10.1038/ng.2310> PMID: 22706312