Whole-genome sequence analysis unveils different origins of European and Asiatic mouflon and domestication-related genes in sheep

Ze-Hui Chen[1,2,16], Ya-Xi Xu[3,16], Xing-Long Xie[1,2], Dong-Feng Wang[1,2], Diana Aguilar-Gómez[4], Guang-Jian Liu[5], Xin Li[1,2], Ali Esmailizadeh[6], Vahideh Rezaei[6], Juha Kantanen[7], Innokentyi Ammosov[8], Maryam Nosrati[9], Kathiravan Periasamy[10], David W. Coltman[11], Johannes A. Lenstra[12], Rasmus Nielsen[13,14,15*], Meng-Hua Li[1,3*]

[1]CAS Key Laboratory of Animal Ecology and Conservation Biology, Institution of Zoology, Chinese Academy of Sciences (CAS), Beijing, China

[2]University of Chinese Academy of Sciences (UCAS), Beijing, China

[3]College of Animal Science and Technology, China Agricultural University, Beijing, China

[4]Center for Computational Biology, University of California at Berkeley, Berkeley, CA 94720, USA

[5]Novogene Co., Ltd, Tianjin, China

[6]Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran

[7]Natural Resources, Natural Resources Institute Finland (Luke), Jokioinen, Finland

[8]Board of Agricultural Office of Eveno-Bytantaj Region, Batagay-Alyta, Russia

[9]Department of Agriculture, Payame Noor University, Tehran, Iran

[10]Animal Production and Health Laboratory, Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, International Atomic Energy Agency, Vienna, Austria

[11]Department of Biological Sciences, University of Alberta, Edmonton, AB T6G2E9, Canada

[12]Faculty of Veterinary Medicine, Utrecht University, Utrecht, the Netherlands

[13]Department of Integrative Biology, University of California at Berkeley, Berkeley, CA 94720, USA

[14]Department of Statistics, UC Berkeley, Berkeley, CA 94707, USA.

[15]Globe Institute, University of Copenhagen, 1350 København K, Denmark.

[16]These authors contributed equally to this work.

* e-mail: menghua.li@cau.edu.cn, rasmus_nielsen@berkeley.edu

Abstract

The domestication and subsequent development of sheep are crucial events in the history of human civilization and the agricultural revolution. However, the impact of interspecific introgression on the genomic regions under domestication and subsequent selection remains unclear. Here, we analyze the whole genomes of domestic sheep and all their wild relative species. We found introgression from wild sheep such as the snow sheep and its American relatives (bighorn and thinhorn sheep) into urial, Asiatic and European mouflons. We observed independent events of adaptive introgression from wild sheep into the Asiatic and European mouflons, as well as shared introgressed regions from both snow sheep and argali into Asiatic mouflon before or during the domestication process. We revealed European mouflons arose through hybridization events between a now extinct sheep in Europe and feral domesticated sheep around 6,000 – 5,000 years BP. We also unveiled later introgressions from wild sheep to their sympatric domestic sheep after domestication. Several of the introgression events contain loci with candidate domestication genes (e.g., *PAPPA2*, *NR6A1*, *SH3GL3*, *RFX3* and *CAMK4*), associated with morphological, immune, reproduction or production traits (wool/meat/milk). We also detected introgression events that introduced genes related to nervous response (*NEURL1*), neurogenesis (*PRUNE2*), hearing ability (*USH2A*) and placental viability (*PAG11* and *PAG3*) to domestic sheep and their ancestral wild species from other wild species.

***Key worlds*:** *Ovis* genus, introgression, domestication, adaptation

***Running title*:** The origins of domestication-related genes in sheep

Introduction

The genus *Ovis* spans ~8.31 million years of evolution and comprises eight extant species: domestic sheep *O. aries*, argali *O. ammon*, Asiatic mounflon *O. orientalis*, European mouflon *O. musimon*, urial *O. vignei*, bighorn sheep *O. canadensis*, thinhorn sheep *O. dalli* and snow sheep *O. nivicola*[1]. Earlier archeological and genetic studies have provided strong evidence for that sheep have been domesticated from their wild ancestor Asiatic mouflon (*O. orientalis*) in the Fertile Crescent ~12,000 – 10,000 years BP [2-4]. The domestication during the Neolithic agricultural revolution had contributed significantly to human civilization by providing a stable source of meat, wool, leather and dairy.

In spite of varying diploid number of chromosomes ($2n$ = 52 - 58) [1], hybridization between wild and domestic sheep, as well as between wild sheep species, has been documented to produce viable and fertile interspecific hybrids [5-9]. Previous studies have shown genetic evidence for introgression [10-14], including adaptive introgression from wild relatives to domestic sheep [15,16]. However, the importance of introgression in the entire *Ovis* genus and its contribution to the sheep domestication process remains largely unexplored.

Because wild sheep have adapted to different biogeographic ranges resulting in them being resilient to many biotic and abiotic stresses, the existing genetic variation of

wild sheep provide an important genetic resource for improving domestic sheep in response to increased food production demands, animal disease occurrence and rapid global climate change. Elucidating the evolutionary and genetic connection between wild and domesticated sheep is therefore important for understanding the potential for using wild sheep genetic material for improvement of domesticated sheep.

In this study, we use high-depth whole genome sequences (average coverage = ~21×) of 72 individuals from the eight *Ovis* species, most of which were understudied in previous genomic studies [7,17-19]. We reconstructed the phylogeny and evolutionary history of these species. In addition, we explored gene flow between species and selection signatures of domestication. These findings add to our understanding of the origins of the Asian and European mouflons and the emergence of domestic sheep.

Results

Sequencing and variant calling

High-depth resequencing of 72 individuals from eight *Ovis* species (Fig.1a and Supplementary Table 1) generated a total of 35.91 billion 150-bp paired-end reads (5.39 Tb), and 35.84 billion clean reads (5.28 Tb) with an average depth of 20.7× (12.2 – 36.9×) per individual and average genome coverage of 97.2% (96.5% – 98.3%) after filtering. The average sequence coverage was 19.3× for *O. aries*, 17.8× for *O. ammon*, 18.9× for *O. canadensis*, 19.8× for *O. vignei*, 18.9× for *O. musimon*, 17.8×

for *O. nivicola*, 19.4× for *O. dalli*, and 27.1× for *O. orientalis*. On average, 95.83% individuals had ≥ 4× coverage, 90.11% had ≥ 10× coverage, and 46.68% had ≥ 20× coverage. Of all the individual sequencing reads, 91.86% were mapped to the *O. aries* reference genome Oar_v4.0 (Supplementary Table 2). Summed over all samples, 125,982,209 SNPs, 13,043,920 INDELs (insertions and deletions ≤ 50bp; ~0.89 million common indels shared by all sheep species and on average 2,605,718 per individual) (Table 1 and Supplementary Tables 2 – 6) and genome-wide structural variations (SVs, 51bp – 997.369 kb: inversions, insertions, deletions, duplications and translocations, on average 41,965 per individual), including copy number variations (CNVs, deletions and duplications of 51bp to 997.369 kb, on average 31,124 per individual) (Supplementary Table 7) were detected. The number of SVs shared by two species ranged from 29,884 to 91,186 (Supplementary Table 8 and Supplementary Fig. 1a). On average, 1.88% SVs were located in exonic regions, 65.2% SVs were located in intergenic regions, and 29.9% SVs were located in intronic regions, while 67.0%, 31.0% and 0.66% SNPs were in intergenic, intronic and exonic regions, respectively (Supplementary Tables 9 and 10).

The percentage of SNPs that was present in public databases [e.g., NCBI sheep dbSNP database v150 and European Variation Archive (EVA)] ranges from 78.1% in argali to 94.3% in domestic sheep (Supplementary Table 11). Our dataset added 2,139,962 novel SNPs (an increase of 7.04%) to the NCBI and EVA database of

sheep genetic variants (Supplementary Table 6). Of the 176,403 common sites between detected SNPs and the Ovine BeadChip, an average of 288,638 genotypes observed here were validated by the Ovine Infinium HD SNP BeadChip data available for 14 individuals of the samples sequenced (97.1% validation rate, and 297,115 common SNPs), and an average of 23,220 genotypes were validated by the Ovine SNP50K BeadChip data available for another 12 individuals of the samples (96.64% validation rate, and 22,583 common SNPs) (Supplementary Table 11).

Moreover, 74 randomly selected SNPs, which are from the NCBI sheep dbSNP database and the candidate genes identified below, were inspected in 4-12 individuals by Sanger sequencing and produced an overall validation rate of 95.5% (Supplementary Table 12). For PCR and qPCR validation of CNVs (deletions and duplications), 14 randomly selected CNVs with 85.4% concordant genotypes (38/42 deletions and 32/40 duplications; Supplementary Table 13 and Supplementary Fig. 2) were successfully validated. The validation rates observed here are higher than those in previous studies[17,20], which could be due to more efficient and precise CNV detection methods used here. The high validation rate indicated high reliability of the genetic variants created in this study.

Patterns of variation

The 126 million SNPs were detected across all eight species. The number of SNPs varied from 11.3 to 20.1 million per individual and from13.4 to 53.6 million (0.6 – 18.2 unique) per species (Supplementary Table 2, Supplementary Table 6 and Supplementary Fig. 3c). We observed 4,431,063 SNPs shared among all the eight species, with the shared SNPs for pairwise comparisons varying from 6,241,176 between European mouflon and snow sheep to 25,195,033 between Asiatic mouflon and urial (Supplementary Table 4 and Supplementary Table 6). More comparisons of structural variants (SVs) and copy number variants (CNVs) among species for uniqueness and sharing are shown in Supplementary Fig. 1c.

Using pairwise genome-wide F_{ST}, the species with highest genetic differentiation were snow sheep and European mouflon, and the ones with least differentiation were urial and Asiatic mouflon (Supplementary Table 14). The species with the highest genomic diversity (π), when only including SNPs with < 10% missing data, were domestic sheep, Asiatic mouflon, and urial (0.0032 – 0.0044), and the ones with lowest diversity were snow sheep, bighorn sheep and thinhorn sheep (0.00075 – 0.00078) (Supplementary Fig. 4b). On average, 67.0% of SNPs were located in intergenic regions, 31.0% in introns, and 0.7% SNPs in exons. The ratio of non-synonymous to synonymous substitutions ranged from 0.72 in urial and 0.77 in domestic sheep to 0.88 in European mouflon (Supplementary Table 10). We pooled the SVs and CNVs across all eight species yielding a high depth of coverage for the

shared and unique SVs and CNVs among them (Supplementary Figs. 1a, b and Supplementary Tables 6 and 7). Annotation of genes overlapped with SVs were summarized in Supplementary Table 15 (see Supplementary Information)

Phylogenomic reconstruction among the *Ovis* species

We generated eight high-depth whole pseudo-haploid genomes (see Online Methods), representing the eight *Ovis* species. Phylogenetic trees were then constructed from concatenated protein coding regions (CDSs) of autosomes, the X chromosome and the mitogenome of the assembled genomes, separately (Supplementary Fig. 5). These trees showed different phylogenetic patterns, but a consistent split between the three Pachyceriform species (i.e., bighorn, thinhorn and snow sheep) and the others, consistent with earlier genetic studies [1,21]. Together with the observation that the first fossil evidence of caprinae is in the Upper Vallesian in Spain [21], these trees confirmed a Eurasian origin of the ovine species [15,22].

We split the whole genome (one high-depth genome per species, see Materials and Methods) into 2,462 autosomal and 136 X-chromosomal 1Mb non-overlapping windows of each species, and estimated Maximum likelihood (ML) trees for these windows. Three topologies (A, B and C) were observed for 46.1%, 29.1% and 17.8% of the autosomal trees, 33.8%, 50.0% and 7.4% of the X chromosomal trees, respectively (Supplementary Fig. 6). The main topologies A and B were also found

126 using the maximum likelihood estimation on the concatenated CDSs (topology A, Fig.
127 2b) and using consensus methods of the Densitree on the non-overlapping fragments
128 for autosomes (topology A, Fig. 2a) and X-chromosome (topologies B, Fig. 2a). We
129 also estimated trees using high-depth individual autosomes and X-chromosome
130 (Supplementary Fig. 5), which also support topologies A and B, respectively, while
131 the individual mtDNA tree did not resemble any of the nuclear topologies.
132
133 The minor topologies (e.g., B and C for autosomes; and A and C for X-chromosome)
134 may reflect local introgression between the species or incomplete lineage sorting (ILS)
135 of ancestral phylogenies. The three phylogenies of all the 72 individuals using
136 concatenated CDSs of autosomes (Fig. 2b and Supplementary Fig. 7a), X-
137 chromosome and mitogenomes (Supplementary Fig. 7) showed seven major clades of
138 individuals, with European mouflon sequences located among domestic sheep,
139 compatible with the assumption that European mouflon is a domestic sheep
140 subspecies [15]. Also, European mouflon and domestic sheep show the same diploid
141 number of chromosomes ($2n = 54$) [1].
142
143 The phylogenetic trees (Supplementary Figs. 5, 7) and pairwise F_{ST} (Supplementary
144 Fig. 8b) showed two clear clusters, one comprised of the European mouflon and
145 domestic sheep, and another of the Asiatic mouflon and the urial sheep. This is again
146 compatible with the hypothesis that the European mouflon is a feral derivative of

147 domestic sheep, but it also suggested that the Asiatic mouflons, sampled in Iran, have
148 diverged considerably from the mouflon ancestors of both the early domestic hair
149 sheep and the domestic wool sheep of more recent origin [22].
150
151 A coalescent hidden Markov model (CoalHMM) based on autosomal sequences
152 indicated a divergence time of domestic sheep and the three Pachyceriform species of
153 0.244 to 0.270 mya. The argali and the urial were estimated to have diverged from
154 domestic sheep ~0.124 – 0.150 mya and ~0.077 – 0.092 mya, respectively
155 (Supplementary Figs. 9 – 11). The divergence time of the Asiatic mouflon and the
156 urial was estimated to have occurred 0.073 – 0.083 mya, which is earlier than the
157 divergence between bighorn and thinhorn sheep (~0.036 – 0.052 mya). An Isolation
158 with Migration (IM) model [23], which incorporates the impact of migration among
159 species, gave a similar estimation with the Isolation (I) model [23].
160
161 The relatively recent divergence of the European mouflon from domestic sheep 5,550
162 – 5,450 BP (Supplementary Figs. 9 – 11) is concordant with the paleontological
163 evidence of teeth and bone for a divergence of the Corsican mouflon and domestic
164 sheep dated at 6,000 – 5,000 BP [24]. Moreover, the coalHMM, IM and I models, with a
165 filtering thresholds of < 1,000 years and > 20,000 years for the split time [14], showed a
166 split time of 12,800 - 8,800 BP between domestic sheep and the Asiatic mouflon. This
167 estimate is congruent with the estimated domestication time of sheep from the Asiatic

mouflon around 9,000 – 11,000 BP, based on archaeological data [4,24], and also is in agreement with the time range 12,000 BP- 8,000 BP from the start of exploitation to the end of domestication [25].

Demographic history

The pairwise sequentially Markovian coalescent (PSMC) model found a dramatic decline in population sizes of these species ~80 – 250 thousand years ago (kya) with a bottleneck for urial and Asiatic mouflon during 30,000 – 10,000 BP (Fig. 3a), coinciding with the glacial periods. The subsequent increase in their population sizes can be ascribed to the prosperity of animal husbandry, agriculture and sedentarism [26]. The SMC++ analysis showed a decline of all species 10,000 – 1,000 BP. In particular, we noted European mouflon has a more dramatic decline of *N*e than domestic sheep 6,000 – 5,000 BP, which probably corresponds to the feralization of the European mouflon (Fig. 3c). The split between domestic sheep and the Asiatic mouflon occurred during 15,000 – 9,000 BP. During this time period, the Asiatic mouflon showed an increased *Ne*, whereas domestic sheep experienced a severe bottleneck because of domestication.

Genetic structure and differentiation

PCA clusters individuals according to the recognized eight species. The cluster of argali showed significant within-species genetic divergence (Fig. 1b), which was also

189 observed in the admixture pattern at high *K* values (Supplementary Fig. 12). The
190 Asiatic mouflon cluster was dispersed and overlaps partially with the urial cluster (Fig.
191 1c and Supplementary Fig. 12). The population tree was compatible with the inferred
192 genetic clustering at *K* =11 (Supplementary Figs. 12a, b), in which each species is
193 assigned its own components. The admixture plot may suggest gene flow from argali
194 (0.06 – 0.76%) and urial (1.4% – 15%) to Asiatic mouflon and possibly from wild
195 relatives to domestic sheep, such as from European mouflon (5.4 – 5.7% of the
196 genomes at *K*=6) to Ouessant sheep, which was an isolated island domestic breed
197 (Supplementary Fig. 12). However, we noted that admixture proportions cannot be
198 interpreted a direct evidence of admixture.
199
200 We observed higher levels of linkage disequilibrium (LD) in European mouflon and
201 domestic sheep than in other species (Fig. 3b). This may be explained by a strong
202 bottleneck during domestication. The Ouessant sheep [27] clearly had a higher LD than
203 other domestic breeds, which was consistent with their low genomic diversity (Fig. 3b
204 and Supplementary Fig. 4). Likewise, the high LD in European mouflon could be
205 explained by a small population size and possible bottleneck during its reintroduction
206 from Corsica island to continental Europe [15].
207
208 **Genomic introgression between wild species**

209 The ABBA-BABA analysis (*D*-statistic) was implemented using ANGSD-based on
210 alignments, which suggested introgressions from bighorn, thinhorn and snow sheep
211 into their Eurasian relatives such as urial and Asian and European mouflon.
212 (Supplementary Table 16). Statistical analyses based on variants using Admixtools
213 (Supplementary Tables 17, 18), TreeMix (Supplementary Figs. 13 and 14) and f_d
214 statistics (Supplementary Fig. 15 and Supplementary Table 19) consistently showed
215 significant introgression of snow, bighorn and thinhorn sheep into urial, Asiatic and
216 European mouflon. Bighorn and thinhorn sheep showed similar patterns of
217 introgression as snow sheep in terms of several statistic indices, such as percentage
218 (urial: 6.23 – 6.33%, Asiatic mouflon: 3.63 – 3.7%, European mouflon: 1.43 – 1.47%),
219 length (urial: 152.68 –155.1Mb, Asiatic mouflon: 88.96 – 90.7 Mb, European
220 mouflon: 35.08 – 35.96Mb) and shared genes (urial: 720 –744, Asiatic mouflon: 449
221 – 468, European mouflon: 151– 155) of introgression (Supplementary Fig. 15). For
222 simplicity, we will focus only on the snow sheep introgression. The introgression
223 events into urial and Asiatic mouflon had a lot of overlap in terms of genomic regions,
224 while there was very minimal overlap between Asiatic and European mouflon
225 introgression segments (Supplementary Fig. 15 and Supplementary Table 17).
226 Furthermore, admixture graph fitting based on f_4 statistics was carried out using the R
227 package *admixturegraph* (Fig. 4), indicating a very close relationship between wild
228 sheep of Pachyceriforms and European mouflon.
229

Signatures of introgression were detected in candidate regions overlapping 892 genes from snow sheep to urial sheep, these genes were significantly (False Discovery Rate, FDR of 0.05 by the method of Benjamini-Hochberg[28]) enriched for nerve conduction, energy metabolism, membrane signal transduction, bile secretion, drug addiction and motor activity using DAVID annotation tools. From snow sheep to Asiatic mouflon or European mouflon, we found candidate introgression regions covering 497 and 179 genes, respectively (Supplementary Fig. 15a). In European mouflon, the introgressed genes were enriched for nerve regulation, locomotory behavior, cardiac disease, insulin secretion, serotonin metabolic process and calcium signaling pathway, while in Asiatic mouflon the genes were enriched in walking behavior, regulation of cell differentiation, ovarian steroidogenesis and platelet activation. Noteworthy, we observed three shared GO terms for the genes involved in the inter-species introgression events, such as motor, iron channel activity, and dendrite development. (Supplementary Table 20).

Among the three sets of introgressed genes between wild species, we observed 12 shared genes (*CYP2J, PRUNE2, ZNF385B, IMMP2L, GRIK2, HS6ST3, USH2A, LOC101111335, TMEM132D, PAG11, PAG3* and *CTNNA3*), which have functions associated with reproduction and production traits such as follicular development (*CYP2J*, *IMMP2L*), prolificacy (*GRIK2*), growth (*HS6ST3*), wool and body weight (*TMEM132D*) [29-34], and nervous response such as hearing ability evolution (*USH2A*)

and nerve development (*PRUNE2*) [35,36]. In particular, shared signatures of introgression were observed in the PAG gene family, which is involved in pregnancy detection and placental viability evaluation [37]. Moreover, these genes were significantly enriched in a GO term (GO:0004190), which consists of pregnancy-associated glycoproteins (*PAG3* and *PAG11*) related to aspartic-type endopeptidase activity. We also observed one marginally significantly enriched KEGG pathway of protein digestion and absorption (oas04974) including the two genes from the PAG gene family.

Dating the introgression from wild relatives to Asiatic mouflon

In addition to the introgression from snow sheep to Asiatic mouflon (Figs. 5a – d) mentioned above, *D*-statistics, f_3 statistics and TreeMix analysis also detected signatures of introgression from argali into Asiatic mouflon (Supplementary Tables 16 – 21 and Supplementary Figs. 13 and 15a). Across-genome f_d values detect 670 and 734 segments introgressed by snow sheep and argali, respectively, corresponding to a genomic coverage of 3.68% and 3.98%, containing 497 and 540 genes. (Supplementary Tables 19, 21). The program DATES yielded time estimates for the snow sheep and argali introgression events of 3,481 and 2,493 generations ago, respectively. Similar estimates were obtained with Ancestry_hmm: 3,096 and 2,545 generations ago (Supplementary Fig. 16). With a generation time of 4 years for Asiatic mouflon, both methods indicated that the introgression from snow sheep, as

well as of bighorn and thinhorn sheep, occurred before the domestication 13,924 – 11,580 years BP. In contrast, the introgression from argali to Asiatic mouflon at 9,972 – 10,180 years BP coincides with the domestication process. Because introgression of argali in domestic sheep is confined to sympatric populations [16,38], we believe that the gene flow between argali and Asiatic mouflon did not take place until after the domestication process, resulting in the first domestic sheep lacking gene flow from argali. The gene flow from argali was probably also absent in the mouflon population ancestral to domestic sheep. GO categories and KEGG pathway of snow sheep and argali introgression into Asiatic mouflon were reported in the Supplementary Table 20.

Selection signatures in domestic sheep

To detect selection signatures in domestic sheep, we used pairwise differences π ratio (π_w/π_d) > 2.36 and F_{ST} between domestic sheep and Asiatic mouflon. We selected the overlap of the top 10% outliers in both methods, identifying 340 windows as candidate regions for selection. These regions contained a set of 131 selective functional genes (Supplementary Table 22 and Figs. 5e, f) which were significantly (*P* 0.05) enriched for GO terms involved in the activation of the innate immune response, positive regulation of defense response to virus by host, ectoderm development, membrane transport and enzyme activity (Fig. 5g).Of the 131 genes, 62 (47%) overlapped with domestication-related genes of previous studies, and were

293 defined as the candidate domestication genes in sheep (Supplementary Table 23).
294 Remarkably, from these candidate domestication genes 11 and 13 (15,365 genes on
295 autosomes, significantly overlapped between two gene lists with Fisher's exact test, P
296 < 0.01) have been introgressed into Asiatic mouflon from snow sheep and argali (Figs.
297 5e, f). These genes were functionally involved in immune response (*HERC3* and
298 *NFYA*), visual evolution (e.g., *RNF24*), resistance to virus (e.g., *SIN3A*) [39-42],
299 production and reproductive traits [e.g. milk and protein yield (*SH3GL3* and
300 *PAPPA2*)], fecundity (*DNAJB14* and *FSIP2*), body measurement (*SIRT3* and
301 *SH3GL3*), tail type (*HAO1*), regulation of osteogenesis (*GTF2I*), skeletal muscle
302 development (*ZNF777*) and lumbar vertebrae number traits (*NR6A1*) [43-52], and
303 environmental adaptation [e.g., superior heat tolerance (*PPP2R5E, GTF2IRD1* and
304 *DNAJB14*)] [53-55].
305
306 Of special interest was the introgressed genomic region chr3: 10980301-11211252
307 which contains gene *NR6A1* and had the highest (OUE, AMUF; SNWS, goat) f_d value
308 (Fig. 6). We also computed the mean pairwise sequence divergence (d_{xy}) of snow
309 sheep and Asiatic mouflon or Ouessant sheep. This region also had a reduced mean
310 pairwise sequence divergences (d_{xy}) of snow sheep and Asiatic mouflon, a high d_{xy} of
311 snow and Ouessant sheep, and a low differentiation (F_{ST}) of Asiatic mouflon and
312 snow sheep, all indicating introgression of snow sheep into Asiatic mouflon (Figs. 6a
313 – c).

For the 11 genes introgressed from snow sheep into Asiatic mouflon, comparisons of haplotypes of the SNPs in the introgressive regions from argali, snow sheep, Asiatic mouflon and domestic sheep were shown in Figs. 6d, e. Notably, we found that the haplotype patterns of Asiatic mouflon strongly resembled those of snow sheep and argali, but differed strikingly from the patterns observed in the domestic sheep (Figs. 6d, e). Haplotype patterns showed most of the introgressive haplotypes of genes (e.g., *NR6A1, FSIP2, ZNF777, RNF24, PPP2R5E*) have not been selected and fixed in domestic sheep (Figs. 6d, e). Since most of the domestication-related genes are associated with production traits, this scenario could be explained by that introgressions associated with adaptation rather than production traits have been mostly selected in the genetic improvement stage after domestication [38].

Common introgressions between wild and domestic sheep

In the introgression test (*D*-statistics and TreeMix analysis) between wild and domestic sheep, we found significant signatures of gene flow from (i) European mouflon into Ouessant sheep (OUE), (ii) urial and Asiatic mouflon into Shal (SHA), and (iii) argali into Tibetan sheep (GMA) (Fig. 4a, Supplementary Table 17, Supplementary Figs. 13, 17). We detected 154 genes located in the introgressed tracts from European mouflon to Ouessant sheep. These genes were significantly enriched in the GO terms and KEGG pathways of nerve conduction and development

(*e.g.*,GO:0007269, GO:0098793, GO:0043065, GO:0090129, GO:0051965 and oas04360), cell adhesion (GO:0007156), intracellular signal transduction (GO:0035556 and oas04024) and walking behavior (GO:0007628) (Supplementary Table 24).

We identified regions containing 516 and 430 introgressed genes from the Asiatic mouflon or urial into Shal sheep, 251 of these were shared between the two species (Supplementary Table 24 and Supplementary Fig. 17). All these genes were significantly ($P < 0.05$) enriched in the GO terms with functions in tissue and organ development, reproduction and morphological change. In these tests of introgression from wild sheep to their sympatric domestic relatives, shared signals were detected in 10 functional genes (e.g., *CCDC67*, *FAT3*, *PCDH15* and *NEURL1*). These 10 common genes have functions associated with arid environment adaptation (*FAT3*) [56], immune response (*PCDH15*) [57], nervous response (*NEURL1*) [58] and disease susceptibility like noise-induced hearing loss (*PCDH15*) [59]. Moreover, the genes introgressed from argali to Tibetan sheep were significantly enriched for GO terms in olfactory bulb development (e.g., *AGTPBP1*, *CRTAC1* and *RPGRIP1L*) and synaptic transmission (e.g., *GRIK2*, *PARK2* and *SHC3*) (Supplementary Table 24). Notably, two introgressed genes from Asiatic mouflon to Shal sheep (i.e., *RFX3* and *DNAJB14*) and one introgressed gene from argali to Tibetan sheep (i.e., *CAMK4*) were identified to be under domestication (Supplementary Tables 22, 23 and 24).

356

357 **Probability of incomplete lineage sorting (ILS)**

358 We estimated the probability of incomplete lineage sorting (ILS) for the introgressed
359 tracts identified from argali and snow sheep into Asiatic mouflon. The expected
360 length of a shared ancestral tract (see Online methods) is L_{snow}= 1/(1.5 × 10^{-8} × (2.3 ×
361 10^{6} × 2) /4) = 57.97 bp, L_{argali}= 1/(1.5 × 10^{-8} × (1.72 × 10^{6} × 2) /4) = 77.52 bp and the
362 probability of a length of at least 96,410 bp and 98,037 bp (*i.e.,* the observed
363 introgressed regions containing the domestication-related genes were 96,410 –
364 319,834 bp and 98,037 – 639,187 bp) is negligible (1 – GammaCDF (96,410, shape =
365 2, rate = 1/*L*) = 0) (Supplementary Table 25). Similarly, the probability of
366 introgressed tracts appearing due to ILS detected from snow sheep to European
367 mouflon and urial, as well as European mouflon, Asiatic mouflon and urial to
368 domestic sheep were all approach zero (Supplementary Table 26). Thus, the inter-
369 species introgressions detected above were unlikely due to ILS.

370

371 ## Discussion

372 In this study, we generated a novel genomic dataset of high-depth whole-genome
373 sequences of domestic sheep and all their wild relatives, including the wild ancestor
374 of sheep (Asiatic mouflon), and the vulnerable (e.g., urial) and near threatened species
375 (argali and Asiatic mouflon) according to the International Union for Conservation of
376 Nature (ICUN) Red List. Our genomic data included different types of molecular

markers such as SNPs, SVs, CNVs and INDELs, providing an important resource for the genetic improvement of sheep, as well as for ecological and evolutionary studies of the wild species.

This is the first comprehensive and in-depth investigation on phylogeny and introgressions among the whole *Ovis* genus. Different from previous studies [1,60], multiple up-to-date analyses were applied to cross-validate the obtained results. For example, to better understand the trajectories of connections between admixture events and phylogenetic relationship across the whole genome, we used sliding window-based and fitting-based methods to construct the consensus trees. Additionally, we implemented the introgression tests based on several statistical approaches such as *D*-statistics, *f*-statistics, TreeMix and admixture analyses. Further, we verified the introgression events (including introgression sources and time) using the admixturegraph fitting method and dated the introgression time using both model-based and LD-decay based methods. All the analyses showed accordant results.

We verified our SNPs based on both statistical and experimental methods, securing the dataset for the subsequent analysis. By comparing with other species, we found on average 17.61 million SNPs/per individual (9 – 21 SNPs/kb among the *Ovis* species), which is less than that in goat (53 – 54 SNPs/kb) [61], but higher than that in swine (1 SNP per 10.3 kb) [62].

Within the *Ovis* species, a relatively low diversity and effective population size (4,000 – 10,000) in the Pachyceriforms may be ascribed to long-term geographic and genetic isolation [1,21] and is relevant for its conservation. The much less genetic diversity observed in Pachyceriforms than that in Moufloniforms could be due to (*i*) the common ancestor of Pachyceriforms should have migrated out from Eurasia, the distribution region of Moufloniforms, with genetic drift and differentiation between each other; and (*ii*) the genome of domestic sheep has been used as the reference for SNP mapping, while domestic sheep is phylogenetically further from Pachyceriforms than Moufloniforms. We observed the highest diversity in Asiatic mouflon (π = 0.0044), which was much higher than in domestic sheep (π = 0.0032). This has been observed previously [63] and can be explained by the domestication bottleneck [64]. Our results also showed lower diversity estimates than previous investigations using whole-genome BeadChip SNPs and mtDNA variation[63,65]. Higher estimates from the BeadChip could be explained by the ascertainment bias in the chip design.

Previous molecular evidence for taxonomic classification has so far mostly been based on mtDNA sequences [1,60,66]. However, pervasive and frequent autosomal introgressions [14,16] probably accounts for the lower estimates of the coalescence time compared with those from mtDNA sequences [1,60]. In particular, the evidence from SMC++, CoalHMM statistics, phylogenetic trees, admixture analysis, and mean

population differentiation (F_{ST}) index indicated a more recent divergence of European mouflon and domestic sheep (~5,000 BP), than estimated on the basis of mtDNA sequences (~21,000 BP) [67]. The recent divergence of European mouflon from domestic sheep was also supported by archaeological data [24]. Our evidence confirmed that European mouflon emerged as feral domestic sheep when the earliest wave of domestic hair sheep, was displaced by a second wave of wool sheep [15,68,69].

Also, we obtained a more recent split time between argali and domestic sheep (~0.12 - 0.15 Mya) than earlier estimates. The earlier divergence time was based on orthologous genes using PAML and the node was calibrated using four fossil records such as the divergence of the opossum and human (124.6–134.8 million years ago [Mya]), human and taurine cattle (95.3–113 Mya), taurine cattle and pig (48.3–53.5 Mya), and taurine cattle and goat (18.3–28.5 Mya) [70]. Also, different estimates of 1.72 ± 0.36 Mya and ~ 2.93 Mya were obtained from mitochondrial sequence variations [1,60] using the five fossil calibration time of 18.3–28.5 Ma between Bovinae and Caprinae, 52–58Ma between Cetacea and hippopotamus, 4 34.1 Ma between baleen and toothed whales, 42.8–63.8 Ma between Caniformia and Feliformia, and 62.3–71.2 Ma between Carnivora and Perissodactyla. This difference could be due to different mutation rates of the whole-genomes and mtDNA sequences, and different calibration time points have been used in different studies. Additionally, we used two model of coalHMM (Isolation with migration and Isolation model) with full consideration of

440 migration after speciation, and the estimates have always been lower than those
441 estimated by mtDNA sequences and protein-coding genes [71]. Besides these, the more
442 recent divergence time estimated here could be attributed to the extensive genomic
443 introgressions between the sequenced genomes of the two species.
444
445 Remarkably, the Admixture and Treemix patterns, as well as *D*-statistics and f_d
446 statistics consistently showed introgression of the Pachyceriforms, comprised by the
447 snow sheep and its American relatives (bighorn and thinhorn sheep), into European
448 mouflon. The Pachyceriforms also introgressed Asiatic mouflon, but this was a more
449 recent event and involved a different set of genomic segments (Supplementary Fig.
450 15). The introgression percentage as inferred by f_d statistics were 1.47%, 1.45% and
451 1.43% from bighorn, thinhorn and snow sheep to European mouflon were, while
452 higher introgression percentages of 3.7%, 3.63% and 3.68% were from bighorn,
453 thinhorn and snow sheep to Asiatic mouflon. This indicated that the wild ancestors of
454 the European mouflon, and consequently also the first hair sheep domesticated,
455 descend from a population that differs from the Asiatic mouflons in this study, which
456 were sampled in Iran. In contrast, the Iranian Asiatic mouflons are phylogenetically
457 diverse and close to urial, which is in line with the mtDNA (Supplementary Fig. 7)
458 and Y-chromosomal phylogeny [19]. As the range of snow sheep and their American
459 relatives did not extend to Europe, our results suggested that European mouflons may
460 have partially descended from a now extinct sheep in Europe and arose through

461 hybridization events between this species and feral domesticated sheep (Fig. 4,
462 Supplementary Fig. 7). Some wild sheep species live in extreme environments, such
463 as snow sheep in the extreme cold arctic regions, and argali on the cold Qinghai-
464 Tibetan Plateau and the Pamir highland. Thus, our data may be relevant for
465 environmental adaptation. However, it's challenging to confirm the ancient
466 introgression trajectories based on modern samples, ancient samples of *Ovis* species
467 are demanding to answer this question.
468
469 Recently, there has been a strong interest in inter-species introgression, particularly
470 from wild relatives to domestic animals such as pig, goat and sheep [9,10,13,15]. For
471 example, an earlier study has shown adaptive introgression and selection on domestic
472 genes in goat [72]. In particular, *MUC6* was found to be introgressed from a West
473 Caucasian tur-like species into modern goat during domestication, and is nearly fixed
474 in domestic goat with the function of pathogen resistance [72]. In the *Ovis* genus,
475 hybridization among species has been documented in previous field and molecular
476 studies [6,9,66]. However, adaptive introgression from distantly related wild species into
477 the wild ancestors of domestic animals or into domestic animals has rarely been
478 investigated [72,73]. Genomic signature of adaptive introgression from European
479 mouflon into domestic sheep has been previously reported [15]. An earlier whole-
480 genome SNP analysis suggested that historical introgression from wild relatives is

481 associated with climatic adaptation and that introgressed alleles in *PADI2* have
482 contributed to resistance to pneumonia in sheep [38].
483
484 A strong signature of adaptive introgression from argali into Tibetan sheep was
485 detected, and the introgressive genes involved in hypoxia and ultraviolet signaling
486 pathways (e.g., *HBB* and *MITF*) and associated with morphological traits such as horn
487 size and shape (e.g., *RXFP2*). The introgressed genes were related to adaptation to the
488 extreme environment in the Qinghai-Tibetan Plateau [16]. We also identified other
489 genes in Tibetan sheep introgressed from argali, associated with disease resistance to
490 pathogens (e.g., *ACTN4*), and with olfactory development (*e.g.*, *AGTPBP1*), and
491 locomotion (*e.g.*, *OXR1*), possibly related to adaptation to the semi-wild grazing and
492 anoxic environments in plateau. Furthermore, we found patterns compatible with
493 adaptive introgression from the Pachyceriform sheep and argali into urial, Asiatic
494 mouflon and European mouflon. However, it is challenging to validate the function of
495 these genes the in vivo or vitro in the wild animals.
496
497 We detected adaptive introgression from various wild species into Asiatic mouflon,
498 covering several domestication-related genes. Inspection of these domestication-
499 related genes (e.g., *KITLG*, *CAMK4*, *NR6A1*, *RNF24*, *MBIP*, *SH3GL3*, *GMDS*,
500 *EXOC2* and *GTF2I*) indicated their functions associated with important
501 morphological, physiological and production traits such as litter size and mammary

cycle [74,75], early body weight (e.g., *PLAG1* [76]), regulation of follicular development (e.g., *NR5A1*; [77]) in sheep. Theoretically, particular functions of these domestication-related candidate genes indicated relevant traits have been the targets under intensive selective pressure during the domestication process, which eventually led to emergence of the typical morphological, production, physiological and behavioral differences between domestic sheep and their wild ancestors [78]. In practice, the highly differentiated nonsynonymous mutations in coding regions of the genes should be functionally important and could be integrated in marker-associated selection and genomic selection for related traits in future genetic improvement of domestic sheep[17].

Conclusions

In conclusion, we estimated the phylogenetic relationships of the sheep species on the basis of high-depth whole genome sequences. Our results suggested a feral origin of domestic sheep for European mouflon around 6,000 – 5,000 years BP and a genetic overlap of urial and the Iranian Asiatic mouflon. We found extensive introgression events among the *Ovis* species, which partially overlap with regions under selection in domestic sheep. Our results provide novel insights into changes in the genome landscapes of domestic sheep and their wild ancestor occurring during and after domestication.

Online Methods

Samples and DNA extraction

Seventy-two whole-genome sequences of 6 domestic breeds (*n* = 18) and all wild species (*n* = 54) of the genus *Ovis* were included in this study (Supplementary Table 1). Here we followed the classification of Nadler et al. (1973) due to the greatest taxonomy number. The classification was based on morphological traits in chromosome diploid number. These domestic breeds were selected from sheep which have showed genomic introgressions from sympatric wild relatives [15,16,38] . Thirty-five whole-genome sequences were sequenced in this study and 37 were from our previous studies [17,19]. The 35 genomes generated here consisted of 7 domestic sheep from 3 populations, including Tibetan sheep (GMA, Maqu county, Gansu), Mazekh sheep (MAZ) in Azerbaijan and Makui sheep (MAK) in Iran, and 28 wild sheep from 5 species, including *O. musimon* (*n* = 3), *O. vignei* (*n* = 5), *O. nivicola* (*n* = 8), *O. dalli* (*n* = 6) and *O. canadensis* (*n* = 6), all of which were understudied in previous studies. The 37 public genomes comprised 11 domestic sheep from the following breeds: 3 French Ouessant (OUE) sheep, sampled in the Netherlands, 3 Baidarak sheep (BAJ) from Russia, 3 Shal sheep (SHA) from Iran, 1 Tibetan sheep and 1 Makui sheep, as well as 26 wild sheep genomes from 3 species (*O. orientalis*, *n* = 16; *O. ammon*, *n* = 8 and *O. vignei*, *n* = 2; Supplementary Table 1). Historical information, geographic distribution, and morphological traits such as body size, horn morphology, color and pattern of the coat have been used in the definition of species [79] and types and varieties of hair and wool sheep [27]. Genomic DNA was extracted from the blood or

544 tissue samples using the standard methods of proteinase *K* solution and phenol-
545 chloroform extraction [80]. DNA samples with a clear band in sepharose gel, an
546 OD_{260}/OD_{280} ratio between 1.7 and 2.0 and a concentration at least 20 ng/μL were
547 used for the library construction.
548
549 **DNA sequencing and read filtering**
550 Whole-genome sequencing was performed using the Illumina Hiseq Xten. At least 1.5
551 μg of genomic DNA from each sample was sheared to a 180-500 bp range using the
552 Covaris S220 instrument (Covaris, Woburn, MA, USA) and used for Illumina library
553 preparation. Sequencing libraries were constructed using the Truseq Nano DNA HT
554 Sample preparation Kit (Illumina Inc., San Diego, CA, USA) following the
555 manufacturer's instructions. In brief, DNA fragments were end-repaired, A-tailed,
556 ligated to paired-end adapter, and the fragments with ~350 bp insert length were
557 selected for amplification by 8-12 cycles of PCR using the Platinum Pfx Taq
558 Polymerase Kit (Invitrogen, Carlsbad, CA, USA). PCR products were purified with
559 the AMPure XP system (Beckman Coulter, Brea, CA, USA), and libraries were
560 analyzed for the size distribution by the Agilent 2100 Bioanalyzer (Agilent
561 Technologies, Palo Alto, CA, USA) and quantified in real-time PCR. The constructed
562 libraries were sequenced on the Illumina HiSeq X Ten platform (Illumina Inc.) and
563 paired-end 150 bp reads were generated.
564

All the newly generated and retrieved whole genomes (*n* =72) were included in the following analyses. On average, 95.83% of the sheep reference genome was covered by the depth of ≥4×, 90.11% was covered by ≥10×, and 46.68% was covered by ≥20×.To obtain reliable reads, we removed the raw paired-reads that meet any of the following three criteria: (*i*) unidentified nucleotides (N-content) ≥ 10%; (*ii*) reads pair with adapters; and (*iii*) > 50% of the read bases with a phred quality (Q) score less than 5.

Reads mapping, variant detection, quality control and annotation

Clean reads were mapped to the sheep reference genome OARv4.0 (GCA_000298735.2) using the BWA v0.7.17 MEM module [81] with the parameters bwa -k 32 -M -R. Duplicates were removed using Picard MarkDuplicates and sorted using Picard SortSam (https://broadinstitute.github.io/picard/). To obtain reliable alignments, the reads meeting any of the following three criteria were filtered: (*i*) unmapped reads; (*ii*) reads not mapped properly according to the aligner used above; and (*iii*) the reads with RMS (root mean square) mapping quality < 20. Base quality score recalibration (BQSR) with ApplyBQSR module (default parameters) was used to detect the systematic errors during the sequencing process.

Variant discovery was carried out using the Genome Analysis Toolkit (GATK-v4.0.4.0) best practices pipeline, followed by a joint genotyping method on all

586 samples in the cohort [82]. In summary, we firstly called the variants based on each
587 sample using Haplotypecaller module in GVCF mode with the parameter -
588 genotyping-mode DISCOVERY --min-base-quality-score 20 --output-mode
589 EMIT_ALL _SITES --emit-ref-confidence GVCF. Then, we implemented the joint
590 genotyping procedure by consolidating all the GVCFs with the GenotypeGVCFs
591 module. Furthermore, we combined all the variants using CombineGVCFs. Variant
592 sites were identified for each of the eight species, separately. Within each species, the
593 following successive filtering processes were applied for the variant site and genotype
594 quality control: First, raw SNPs were hard filtered using the VariantFiltration module
595 with the strict parameters -filter-expression QUAL < 30.0 || QD < 2.0 || MQ < 40.0 ||
596 FS > 60.0 || SOR > 3.0 || HaplotypeScore > 13.0 || MQRankSum < -12.5 ||
597 ReadPosRankSum < -8.0 in each species, separately. We then merged all the 8 variant
598 datasets from eight species using the bcftools merge function after the bcftools index.
599 In addition, PLINK v1.9 [83] was used to filter SNPs which meet any of the following
600 criteria: (*i*) proportion of missing genotypes among all the individuals over 10% (geno
601 0.1); (*ii*) SNPs with minor allele frequency (MAF) higher than 0.05 (maf 0.05); (*iii*)
602 SNPs showing an excess of heterozygosity (--hwe 0.001); and (*iv*) non-biallelic sites.
603 This yielded a high-quality set of variants including 6,558,545 SNPs were obtained
604 for the genomic introgression and diversity analyses. For analyzing population genetic
605 structure, we excluded SNPs in LD with $r^2 \geq 0.2$ (--indep-pairwise 50 5 0.2). All the

606 SNPs were annotated using the ANNOVAR v.2013-06-21 software[84] and phased
607 using Shapeit v4.1.3 [85].
608
609 **SV detection and annotation**
610 To identify reliable structural variants (SVs), we detected the SVs by implementing
611 four independent calling pipelines. First, SVs were detected based on the filtered and
612 sorted BAM file using novoBreak v.1.1.3 [86], which detects deletions (DEL),
613 inversions (INV), tandem duplications (DUP) and inter-chromosomal translocations
614 (TRA). Second, SVs were identified using configManta.py in manta v.1.6.0 [87]. Manta
615 reports SVs as deletions (DEL), inversions (INV), tandem duplications (DUP),
616 insertions (INS) and inter chromosomal translocations (TRA). Third, SVs were
617 detected using GRIDSS v2.6.2 [88]. SV files in VCF format were then annotated using a
618 custom R script
619 (https://github.com/PapenfussLab/gridss/blob/master/example/simple-event-
620 annotation.R). GRIDSS generates the same variant types of SVs as those by manta.
621 These three pipelines utilized the same input of 72 BAM files. Fourth, paired-end
622 reads were re-mapped to the sheep reference genome (Oar_v4.0) using the align
623 module of SpeedSeq v.0.1.2 [89].
624
625 In addition, sorted and duplicate-marked BAMs, which contain split reads and
626 discordant read-pairs, were generated. SVs were then identified from the split reads

and discordant pairs using LUMPY v.0.2.13 [90]. CNVs were detected from the difference in read depth using CNVnator v.0.3.3 [91]. The inferred breakpoints by LUMPY were genotyped using SVTyper v.0.1.4 [89]. The variant types of SVs detected by the SpeedSeq framework are the same as those by the GRIDSS pipeline. In these two pipelines, we generated non-uniquely mappable genomic regions for autosomes and X chromosomes, respectively, using SNPable (http://lh3lh3.users.sourceforge.net/snpable.shtml), and these regions were masked in the SV detection by the two methods described above.

To reduce the false positive rate, SVs in both autosomes and X chromosomes from the four strategies (novoBreak, manta, GRIDSS and SpeedSeq) which meet the following seven criteria were retained: (*i*) at least three split reads (SR) or three spanning paired-end reads (PE) supporting the given SV event across all the samples; (*ii*) SVs with precise breakpoints by novoBreak (flag PRECISE); (*iii*) SVs passing the quality filters suggested by NovoBreak, manta and GRIDSS (flag PASS); (*iv*) SVs with more than four supporting reads (flag SU) and without ambiguous breakpoints (flag IMPRECISE) in SpeedSeq; (*v*) SVs with lengths between 50 bp and 1 Mb; (*vi*) SVs without intersections between different variant types; and (*vii*) SVs identified by at least two pipelines. For each sample, the shared SVs detected at least by two of the four independent pipelines were merged using SURVIVOR v.1.0.6 [92] with the parameters 500 2 1 1 0 50.

SVs were annotated based on their start positions using the package ANNOVAR v.2013-06-21 [84]. Species-unbalanced SVs are defined as SVs which are unevenly distributed among different species. A two-sided Fisher's exact test was utilized to determine whether the distribution of each SV is uniform. The *P*-values for all the SVs were calculated with the Fisher.test function in R followed by the Benjamini–Hochberg false discovery rate (FDR) adjustment. SVs with FDR < 0.05 were considered as species-unbalanced.

SNPs and CNVs validation

74 randomly selected SNPs of 4-12 individuals were verified by PCR amplifications and Sanger sequencing. The primers used for the PCRs were designed with the software Primer Premier 5 [93]. The PCR reactions were performed in a total volume of 25 μl, consisting of 12.5 μl 2× Taq MasterMix (Kangwei, Beijing, China), 2 μl (10 pmol/μL) reverse and forward primers, 1 μl template DNA (30 ng/μL) and 9.5 μl double-distilled water (ddH_2O) under the reacting condition of initial denaturation at 95 °C for 3 min, 35 cycles for the following three steps, such as denaturation at 95 °C for 15 sec, annealing at 60 °C for 15 sec, and extension at 72 °C for 30 sec, with a final extension at 72 °C for 5 min. Following the PCR, the amplification products were sequenced on the Applied Biosystems 3730XL DNA Analyzer (Life Technologies, Carlsbad, CA, USA), and the sequencing peaks were checked with the

669 software SEQMAN module of DNASTAR's LASERGENE [94]. Subsequently,
670 genotypes obtained from the Sanger sequencing were compared with those inferred
671 by the GATK pipelines (described above) from resequencing data for the same
672 individuals.
673
674 Moreover, 14 randomly selected CNVs (e.g., seven deletions and seven duplications;
675 Supplementary Table 13) were validated by quantitative real-time PCR (qPCR) or
676 PCR. Primers designed surrounding the deletions and within the duplications with the
677 software Primer Premier 5 (Supplementary Table 13). Deletions were genotyped by
678 PCR amplification and agarose gel electrophoresis. We measured the relative copy
679 numbers of one deletion and all duplications using qPCR on the QuantStudio™ 6
680 Flex Real-Time PCR System (Life Technologies, Carlsbad, CA, USA) using SYBR
681 Green kit (Promega, Madison, WI, USA). Following a previous study on sheep [95],
682 *DGAT2* gene was used as the internal reference gene. qPCR reaction was in 25 μl
683 volume consisting of 12.5 μl 2× SYBR Green qPCR Mix (Life Technologies,
684 Carlsbad, CA, USA), 1 μl (10 pmol/μL) each primer (forward and reverse), 2 μl
685 template DNA (30 ng/μl), and 8.5 μl ddH_2O. The thermocycling condition includes an
686 initial denaturation at 95 °C for 10 min, 40 cycles for the next three steps, such as
687 denaturation at 95 °C for 15 s, annealing at 60 °C for 15 s and extension at 72 °C for 1
688 min, and a final extension at 72 °C for 10min.
689

For qPCR, the $\Delta\Delta C_T$ method [96] was applied to estimate the relative copy numbers. Equation for $\Delta\Delta C_T$ value is $\Delta\Delta C_T = [(C_{T\ segment} - C_{T\_DGAT2})_{target\ sample} - (C_{T\ segment} - C_{T\_DGAT2})_{control\ sample}]$, where $C_{T\ segment}$ is threshold cycle (C_T) of target CNV segment and C_{T_DGAT2} is the C_T of the internal reference gene [96]. We also measured the standard deviation of the $\Delta\Delta C_T$ value using the formula: $s = (s_1^2 + s_2^2)^{1/2}$, where s_1 is the variance of target C_T value (3 replications) and s_2 is the variance of the reference C_T value (3 replications). The value of $2\times2^{-\Delta\Delta C_T}$ between 1.5 and 3 were considered to most likely represent a normal copy number of 2, below 1.5 or above 3 are considered as deletions or duplications, respectively [20]. This was used to evaluate the concordance of calling results obtained from four SVs calling strategies and the relative copy number from the qPCR.

Inference of demographic history

We inferred past temporal change in *Ne* and population split times using the pairwise sequentially Markovian coalescent (PSMC) modelling (http://github.com/lh3/psmc) and SMC++ program (https://github.com/popgenmethods/smcpp#masking). We applied the parameters of a generation time (g) of 3 years, neutral mutation rate (μ) = 2.5×10^{-8} per base pair per generation, a per-site filter of ≥10 reads and no more than 25% of missing data [97], including only autosomes from one high-coverage genomes (> 18×) per species (PSMC, Supplementary Table 2) or three individuals per population

or species (SMC++). We performed 1,000 bootstrapping simulations to estimate the variance of *Ne*.

Genomic diversity and population differentiation

For each individual, genome-wide nucleotide diversity was calculated based on the set of high-quality SNPs (n = 6,558,545) using Vcftools v0.1.13 with a window size of 200-kb. Genome-wide pairwise F_{ST} and d_{xy} genetic distance matrices between populations was estimated using in-house python scripts with a window size of 100-kb and a 20-kb step size. The matrices of pairwise distances were then plotted using the Corrplot package of R. In order to assess the genome-wide LD patterns of each species, we calculated r^2 value using the program PopLDdecay v3.30 [98] (https://github.com/BGI-shenzhen/PopLDdecay) with the default parameters and after filtering the sites with more than 10% missing genotypes among the individuals of each species cohort.

Population genetic structure and phylogenetic reconstruction

We implemented principal components analysis (PCA) using the Smartpca program [99] in the software EIGENSOFT v7.2.1 [100] without outlier removal iteration (numoutlieriter: 0) but with the default settings of the other options. The Tracy-Widom test was used to determine significance of the eigenvectors. The first two eigenvectors were plotted. We used the Ohana tool suite [101] to infer the global

731 ancestry and the covariance structure of allele frequencies among the species. The
732 number of ancestry components (*K*) was set in a range from 2 to 11. For each *K*, we
733 terminated the iteration when the likelihood improvement is smaller than 0.001 (-e
734 0.001). We only reported the ones which reached the best likelihood for each *K*.
735 Population trees at each *K* (Supplementary Fig. 12b) were plotted using the program
736 Nemetree (http://www.jade-cheng.com/trees/).
737
738 The phylogenetic tree of the nine species was constructed using the maximum
739 likelihood method implemented in the RAxML v8.2.3 [102] with the multiple nucleotide
740 substitution models. The tree was inferred based on the 12,837 protein coding
741 sequences (CDS) on autosomes and 513 CDS on X chromosome, separately. We used
742 the protein-coding gene annotation file from NCBI
743 (ftp://ftp.ncbi.nlm.nih.gov/genomes/ Ovis_aries/GFF/). Only CDS with length
744 multiple of 3 were considered in the phylogenetic inference. The consensus trees
745 (Supplementary Fig. 5) based on the whole genome was built on the concatenated
746 CDSs of autosomes (33,868,497 bp), X chromosome (1,331,184 bp) and the whole
747 mitogenomes (16,616 bp), respectively. Moreover, seventy-two haploidized whole-
748 genome sequences for all the individuals were generated using the -doFasta3 option in
749 ANGSD [103] (Fig. 2b and Supplementary Fig. 7), which uses the bases with the highest
750 effective depth (EBD) and considers both mapping quality and scores for the bases [104].
751 To examine the impact of different assembly methods on the phylogenetic inference,

752 we also tested the options of -doFasta 1 and -doFasta 2, which utilize the genomic
753 sites by randomly selecting the base or selecting the base with the highest depth.
754
755 The preliminary tree for the optimization were constructed using the GTRCAT model
756 in RAxML. Phylogenetic inference of autosomal and X chromosomal sequences was
757 then implemented based on the first two codon positions and the third codon position
758 of the whole concatenated coding sequence using the GTRGMMA model in RAxML.
759 The final trees after 200 bootstrapping replicates were generated using GTRCAT
760 model in RAxML and returned to the preliminary tree labeled with bootstrap values.
761 To clarify discordant coalescent events among different tracts in the genome, we split
762 the whole genome into 1-Mb tracts, which result in 2,598 non-overlapping windows,
763 respectively. We inferred the ML trees using the GTRGAMMA model. Finally, trees
764 were built of each 1-Mb windows (Supplementary Fig. 6). Numeration (classification
765 and ranking) of trees was conducted using all.equal function in R package of Ape
766 (analyses of phylogenetics and evolution) and plotted by in-house R scripts. The trees
767 of each tract of autosomes and X chromosome in 1-Mb window were fitted and
768 visualized by Densitree v2.0.1 [105] (Fig. 2a). Mitochondrial sequences between sheep
769 and goat were blasted using MEGA7 [106]. Genomic coordinates of goat were
770 transferred based on locations of the sheep genome after trimming the poorly mapped
771 sites. Finally, we merged the 73 mitochondrial sequences in the phylogenetic analysis
772 with goat as the outgroup.

773

774 **Estimation of split time**

775 Divergence time was estimated locally based on each 1-Mb tracts across autosomes.
776 We used the coalescent hidden Markov model (CoalHMM) [23], a framework for
777 demographic inference using a sequential Markov coalescent method, to estimate the
778 split time with or without migrations among species. We first converted the pairwise
779 sequence alignments using python scripts prepare-alignments.py
780 (https://github.com/birc-aeh/coalhmm/tree/master/scripts/). The I-CoalHMM and IM-
781 CoalHMM models were then applied to the dataset of 1-Mb tracts. The two models
782 utilized the genome alignments of two species to calculate the time of speciation. In
783 the I-CoalHMM model, a prior of split time and ancestral effective population size
784 were needed, whereas in IM-CoalHMM model extra migration rates were also needed.
785 The recombination rate was set as 1.5 cM/Mb [107]. We combined the pairwise
786 alignments between species totaling nine pairs and used 1-Mb splitting windows of
787 the whole genomes for each pair and discarded the windows with > 10% missing
788 bases. We filtered the time estimates for the windows using the following criteria: (*i*)
789 a split time of below 1,000 years or above 10,000 years for European mouflon and
790 sheep, below 1,000 years or above 20,000 years for Asiatic mouflon and sheep or
791 below 10,000 years or above 10 million years for the divergence between the other
792 seven pairs of species (Supplementary Figs. 9 – 11), (*ii*) a recombination rate below

793 0.1 cM/Mb or above 5 cM/Mb, and (*iii*) an ancestral effective population size below
794 5,000 or above 1,000,000 [14].
795
796 **Migration events by TreeMix analysis**
797 To infer migration events among the eight species, we used TreeMix v1.13 to
798 construct a ML tree with bighorn as the root using the "-noss" option to turn off the
799 sample size correction, a window size (-*K*) of 500 SNPs (around 609-kb in this study)
800 to account for the impact of LD, which is more than the average LD length of
801 approximately ~150-kb observed in sheep [7]. Blocks with 500 SNPs were resampled
802 and 100 bootstrap replications were performed. We constructed the ML trees with 0-
803 11 migration events and corresponding residuals. The proportions of explained
804 variance (Supplementary Fig. 14) for the migration numbers were calculated using in-
805 house scripts [108].
806
807 **Gene flow among species**
808 To infer the ancestral alleles, genomic comparison between domestic sheep (*O. aries*)
809 and domestic goat (*C. hircus*) was carried out using the LAST v984 program
810 (http://last.cbrc.jp/) (Supplementary Fig. 18). We aligned the sheep reference genome
811 (Oar_v4.0) to the goat reference genome (ASR.1) while masking the repeat regions.
812 Only autosomal one-to-one orthologs were considered in the alignment between the
813 two species using the lastal module with the parameters of -m 100 –E 0.05. To

visualize the corresponding orthologs between species, a synteny plot was created using the circlize function in the R package. Samtools mpileup and Bcftools call were then used to call ancestral alleles. We merged these ancestral variants with the combined SNPs of all the 72 samples using Bcftools merge after indexing the two datasets. The combined dataset was used to detect introgression among species.

To detect the potential gene flow among species, we conducted the ABBA-BABA test (*D*-statistics) based on two data panels: single high-depth genomes and high reliable SNPs among all the individuals. These two datasets can be collated between each other to reduce variants calling errors. For the first data panel, we performed the admixture analysis using ANGSD -doAbbababa 1 module with goat as the outgroup and the block size of 1,000,000 bp. For the second data panel, we examined the admixture among species using the qpDstats module of AdmixTools [109] and goat as the outgroup, which is a formal four-population test of admixture. Furthermore, we performed the three-population test using the qp3pop module of AdmixTools. The statistical significance of *D* value was evaluated using a two-tailed Z test, with |Z-score| > 3 to be significant [110]. We built the admixture graphs, fitted the graph parameters and visualized the goodness of fit using admixturegraph [111] package in R.

Inference of introgressed genomic regions

To further localize the introgressed genomic regions across the whole-genome, a window-based Patterson's four-taxon *D*-statistic test *D* (P1, P2, P3, O) and modified *f*-statistic (f_d) test with 100-kb length windows and 20-kb steps was performed using the methods of Martin *et al*. (2015) [112]. P1 was the reference population with no gene flow with P3 and is closer to P2 than P3. Here, goat was used as the outgroup (O), which was the ancestral population and shared derived alleles with populations P1, P2, and P3. The significance level (*p*-value) of Z-transformed f_d value was corrected by multiple testing using the Benjamini–Hochberg FDR method [28]. Windows with positive *D* values and *p* values (FDR adjusted) < 0.05 were selected as the significantly introgressed regions, and the adjacent windows were merged into concatenated introgressed regions [113].

We tested for genomic introgressions between different combinations of species.
(*i*) *D* (OUE, target; X, goat): The domestic population of Ouessant (OUE) serves as the reference population, the goat reference sequence was the outgroup, European mouflon, Asiatic mouflon or urial were the targets and X (bighorn, thinhorn, argali and snow sheep) was the to be tested source of introgression.
(*ii*) *D* (GMA, target; European mouflon, goat): We selected as reference the old and native Tibetan sheep (GMA) that has no potential gene flow with European mouflon and the targets were Ouessant in France, Mazekh in Azerbaijian, Makui and Shal sheep in Iran.

(*iii*) *D* (GMA, target; Asiatic mouflon, goat): the targets were the domestic Mazekh, Makui and Shal sheep.

(*iv*) *D* (OUE, Baidarak; snow sheep, goat): OUE from France was a suitable reference because its large distance to the range of snow sheep and Baidarak was the target.

(*v*) *D* (OUE, target; argali, goat): targets were domestic Tibetan sheep and Russian Baidarak.

(*vi*) *D* (OUE, target; urial goat): targets were domestic Mazekh, Makui and Shal sheep.

In addition, we calculated mean pairwise sequence divergence (d_{xy}) and F_{ST} value between the target population (P2) and the test population (P3), as well as between the test population (P3) and the reference population (P1). Introgression but not shared ancestry reduces d_{xy} in the target regions [112]. Similarly, introgressed regions have lower divergence (F_{ST}) than other regions.

Dating introgression events

We dated the time of ancient introgression using DATES [114] and Ancestry_hmm program [115]. The software DATES computed the weighted LD statistic to infer the population admixture history, which has been developed for human datasets. However, because of the short generation time for sheep, the time estimates using DATES might be younger than expected. Thus, we also applied the Ancestry_hmm program [115], using phased data and only SNPs with at least two alleles in the reference populations and applying the following filters: (*i*) SNPs with allele frequency difference lower

than 0.1 between the two reference populations; and (*ii*) SNPs with allele number less than 6 in a reference panel. Other parameters were set as default. We set the proportion of admixture (*m*) according to admixture fraction obtained above by the *f* statistics (f_d) across the whole genomes. For dating the introgression from bighorn sheep, thinhorn sheep, snow sheep or argali as source of introgression (reference population 2) into Asiatic mouflon, we used urial as the ancestor (reference population 1). We applied a single pulse model for genotype data from each population and ran 100 bootstrap replicates using a block size of 5,000 SNPs.

Incomplete lineage sorting

We calculated the probability of incomplete lineage sortings (ILSs) following the method in Huerta-Sánchez *et al*. (2014) [116]. Briefly, the expected length of a shared ancestral sequence is $L=1/(r\times t)$. The probability of a length of at least *m* follows from 1 − GammaCDF (*k*, shape = 2, *r* = 1/L), in which GammaCDF is the Gamma distribution function, *r* is the recombination rate per generation per bp, *m* is the length of introgressed tracts, and *t* is the length of the two species branch since divergence. According to the theoretical expectation, we can exclude the possibility of common ancestral source when the detected length of tracts (*m*) > *L* or the probability of a length of at least *m* infinitely approaches zero. Here, we set recombination rate of 1.5×10^{-8} [107], generation time of 4 years for Asiatic mouflon and urial [117] and 3 years for domestic sheep [118]. We set divergence times of 2.3 mya for snow sheep and Asiatic

mouflon, 1.72 mya for argali and Asiatic mouflon [21], 2.42 mya for the Pachyceriforms and the Moufloniforms (urial, Asiatic mouflon and European mouflon) [1], 5 □ 6 kya for European mouflon and domestic sheep [24], ~11 kya for Asiatic mouflon and domestic sheep [4], and ~1.26 mya [1] for urial and domestic sheep.

Functional annotation

The genes which overlapping with the concatenated introgressed regions detected by the modified f_d value were annotated. We annotated and categorized the functions of genes using DAVID v6.8 [119] (https://david.ncifcrf.gov/). FDR, Bonferroni and Benjamini-Hochberg adjusted *p*-values were estimated with p-value < 0.05 as statistically significant. GO and KEGG pathway enrichment analyses were implemented using DAVIDv6.8 [119] (https://david.ncifcrf.gov/).

Ethics statement. All animal work was conducted according to a permit (No. IOZ13015) approved by the Committee for Animal Experiments of the Institute of Zoology, Chinese Academy of Sciences (CAS), China. For domestic sheep, animal sampling was also approved by local authorities where the samples were taken.

Life Sciences Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. Raw sequencing data that support the findings of this study will deposit in the European Nucleotide Archive (ENA) with the corresponding accession codes xxxx and xxxx after acceptance. Source data for Supplementary Figs.2, 3, 15 are presented in the Supplementary Tables. Additional data such as raw image files and in-house scripts that support this study are available from the first authors upon request.

ACKNOWLEDGEMENTS

This study was financially supported by grants from the National Key Research and Development Program-Key Projects of International Innovation Cooperation between Governments (2017YFE0117900), the External Cooperation Program of Chinese Academy of Sciences (152111KYSB20190027), the National Natural Science Foundation of China (Nos. 31661143014, 31825024 and 31972527), the Second Tibetan Plateau Scientific Expedition and Research Program (STEP) (No. 2019QZKK0501), and the Taishan Scholars Program of Shandong Province (No. ts201511085). We thank Ming-Shan Wang, Sheng Wang, Hua-Jing Teng, Da-Qi Yu, Peter Wilton, Débora YC Brandt for their technical help with the statistical analysis. We express our thanks to the owners of the sheep for donating samples (see Supplementary Table 1). Thanks are also due to a number of persons for their help during sample collection.

AUTHOR CONTRIBUTIONS

M.-H.L. conceived the study. M.-H. L. and R.N. supervised the study. Z.-H.C. and Y.-X.X. conducted the laboratory work. Z.-H.C., X.-L.X., G.-J.L. contributed the data analysis. D.-F.W., D.A.G. provided the help for coding. X.-L.X., G.-J.L. performed the analysis of SVs. Z.-H.C., X.-L.X., Y.-X.X., D.W.C., A.E., J.A.L., R.N. and M.-H.L. wrote or revised the paper. K.P., I.A., D.W.C., J. K., M.N., V.R. contributed samples or provided help during the sample collection. All the authors reviewed and approved the final manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

References

In Main Text:

1. Rezaei, H.R. *et al.* Evolution and taxonomy of the wild species of the genus Ovis (Mammalia, Artiodactyla, Bovidae). *Mol. Phylogenet Evol.* **54**, 315-26 (2010).
2. Zeder, M.A. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc. Natl. Acad. Sci.* **105**, 11597-11604 (2008).
3. Wild, J.P. ML Ryder: Sheep and man. London: Duckworth, *Antiquity* **58**, 142-142 (1984).

959 4. Chessa, B. *et al.* Revealing the history of sheep domestication using retrovirus
960 integrations. *Science* **324**, 532-6 (2009).
961 5. Woronzow, N. *et al.* Chromossomi dikich baranow i proisschojdjenije domaschnich
962 owjez. *Lriroda* **3**, 74-81 (1972).
963 6. Bunch, T. & Foote, W. Evolution of the 2n = 54 karyotype of domestic sheep (*Ovis*
964 *aries*). *Ann. Genet. Sel. Anim.* **9**, 509-515 (1977).
965 7. Alberto, F.J. *et al.* Convergent genomic signatures of domestication in sheep and
966 goats. *Nat. Commun.* **9**, 1-9 (2018).
967 8. Schröder, O. *et al.* Limited hybridization between domestic sheep and the European
968 mouflon in Western Germany. *Eur. J. Wildl. Res.* **62**, 307-314 (2016).
969 9. Bagirov, V. *et al.* Cytogenetic characteristic of *Ovis ammon ammon*, *O. Nivicola*
970 borealis and their hybrids. *Сельскохозяйственная биология* **6**, 43-48 (2012).
971 10. Jones, M.R. *et al.* Adaptive introgression underlies polymorphic seasonal
972 camouflage in snowshoe hares. *Science* **360**, 1355-1358 (2018).
973 11. Chen, N. *et al.* Whole-genome resequencing reveals world-wide ancestry and
974 adaptive introgression events of domesticated cattle in East Asia. *Nat.*
975 *Commun.* **9**, 2337 (2018).
976 12. Figueiro, H.V. *et al.* Genome-wide signatures of complex introgression and
977 adaptive evolution in the big cats. *Sci. Adv.* **3**, e1700299 (2017).
978 13. Gopalakrishnan, S. *et al.* Interspecific gene flow shaped the evolution of the genus
979 *Canis*. *Curr. Biol.* **28**, 3441-3449.e5 (2018).

14. Wu, D.D. *et al.* Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nat. Ecol. Evol.* **2**, 1139-1145 (2018).

15. Barbato, M. *et al.* Genomic signatures of adaptive introgression from European mouflon into domestic sheep. *Sci. Rep.* **7**, 7623 (2017).

16. Hu, X.J. *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.* **36**, 283-303 (2019).

17. Li, X. *et al.* Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nat. Commun.* **11**, 2815 (2020).

18. Naval-Sanchez, M. *et al.* Sheep genome functional annotation reveals proximal regulatory elements contributed to the evolution of modern breeds. *Nat. Commun.* **9**, 859 (2018).

19. Deng, J. *et al.* Paternal origins and migratory episodes of domestic sheep. *Curr. Biol.* **30**, 4085-4095.e6 (2020).

20. Zhou, Y. *et al.* Genome-wide copy number variant analysis reveals variants associated with 10 diverse production traits in Holstein cattle. *BMC Genomics* **19**, 314 (2018).

21. Bunch, T.D., Wu, C., Zhang, Y.P. & Wang, S. Phylogenetic analysis of snow sheep (*Ovis nivicola*) and closely related taxa. *J Hered.* **97**, 21-30 (2006).

1000 22. Ciani, E. *et al.* On the origin of European sheep as revealed by the diversity of the
1001 Balkan breeds and by optimizing population-genetic analysis tools. *Genet. Sel.*
1002 *Evol.* **52**, 25 (2020).
1003 23. Mailund, T. *et al.* A new isolation with migration model along complete genomes
1004 infers very different divergence processes among closely related great ape
1005 species. *PLoS Genet.* **8**, e1003125 (2012).
1006 24. Vigne, J. D. Zooarchaeology and the biogeographical history of the mammals of
1007 Corsica and Sardinia since the last ice age. *Mamm. Rev.* **22**, 87-96 (1992).
1008 25. Larson, G. *et al.* Current perspectives and the future of domestication studies.
1009 *Proc. Natl. Acad. Sci. USA* **111**, 6139 (2014).
1010 26. Nielsen, R. *et al.* Tracing the peopling of the world through genomics. *Nature* **541**,
1011 302-310 (2017).
1012 27. Mason, I. A World dictionary of livestock breeds, types and varieties. CAB
1013 International. *Wallingford, UK.* (1996).
1014 28. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical
1015 and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B*
1016 *(Methodological)* **57**, 289-300 (1995).
1017 29. Yang, H. *et al.* Identification and profiling of microRNAs from ovary of estrous
1018 Kazakh sheep induced by nutritional status in the anestrous season. *Anim.*
1019 *Reprod. Sci.* **175**, 18-26 (2016).

30. Posbergh, C.J., Thonney, M.L. & Huson, H.J. P5017 Identifying genetic regions to spring ewes to lamb out of season. *J. Anim. Sci.* **94**, 123-124 (2016).

31. Peng, W.F. *et al.* A genome-wide association study reveals candidate genes for the supernumerary nipple phenotype in sheep (*Ovis aries*). *Anim. Genet.* **48**, 570-579 (2017).

32. Jia, C. *et al.* Identification of genetic loci associated with growth traits at weaning in yak through a genome-wide association study. *Anim. Genet.* **51**, 300-305 (2020).

33. Liu, G. *et al.* Expression profiling reveals genes involved in the regulation of wool follicle bulb regression and regeneration in sheep. *Int. J. Mol. Sci.* **16**, 9152-9166 (2015).

34. Tarsani, E. *et al.* Discovery and characterization of functional modules associated with body weight in broilers. *Sci. Rep.* **9**, 9125 (2019).

35. Huang, D. *et al.* Identification of the mouse and rat orthologs of the gene mutated in Usher syndrome type IIA and the cellular source of *USH2A* mRNA in retina, a target tissue of the disease. *Genomics* **80**, 195-203 (2002).

36. Iwama, E. *et al.* Cancer-related *PRUNE2* protein is associated with nucleotides and is highly expressed in mature nerve tissues. *J. Mol. Neurosci.* **44**, 103-14 (2011).

1039 37. Wallace, R.M., Pohler, K.G., Smith, M.F. & Green, J.A. Placental *PAGs*: gene
1040 origins, expression patterns, and use as markers of pregnancy. *Reproduction*
1041 **149**, R115-26 (2015).
1042 38. Cao, Y.H. *et al.* Historical introgression from wild relatives enhanced climatic
1043 adaptation and resistance to pneumonia in sheep. *Mol. Biol. Evol.* **38**, 838-855
1044 (2020).
1045 39. Al Kalaldeh, M., Gibson, J., Lee, S.H., Gondro, C. & van der Werf, J.H.J.
1046 Detection of genomic regions underlying resistance to gastrointestinal
1047 parasites in Australian sheep. *Genet. Sel. Evol.* **51**, 37 (2019).
1048 40. Wong, D. *et al.* Genomic mapping of the MHC transactivator CIITA using an
1049 integrated ChIP-seq and genetical genomics approach. *Genome Biol.* **15**, 494
1050 (2014).
1051 41. Wang, W. *et al.* Deep genome resequencing reveals artificial and natural selection
1052 for visual deterioration, plateau adaptability and high prolificacy in Chinese
1053 domestic sheep. *Front. Genet.* **10**, 300-300 (2019).
1054 42. Bouloy, M. & Weber, F. Molecular biology of rift valley Fever virus. *Open Virol.*
1055 *J.* **4**, 8-14 (2010).
1056 43. Liu, L.L., Fang, C. & Liu, W.J. Identification on novel locus of dairy traits of
1057 Kazakh horse in Xinjiang. *Gene* **677**, 105-110 (2018).
1058 44. Taye, M. *et al.* Exploring evidence of positive selection signatures in cattle breeds
1059 selected for different traits. *Mamm. Genome* **28**, 528-541 (2017).

1060 45. Yurchenko, A.A. *et al.* High-density genotyping reveals signatures of selection
1061 related to acclimation and economically important traits in 15 local sheep
1062 breeds from Russia. *BMC Genomics* **20**, 294 (2019).
1063 46. Jin, Y. *et al.* Detection of insertions/deletions within *SIRT1*, *SIRT2* and *SIRT3*
1064 Genes and their associations with body measurement traits in cattle. *Biochem.*
1065 *Genet.* **56**, 663-676 (2018).
1066 47. Yuan, Z. *et al.* Selection signature analysis reveals genes associated with tail type
1067 in Chinese indigenous sheep. *Anim. Genet.* **48**, 55-66 (2017).
1068 48. Håkelien, A.M. *et al.* The regulatory landscape of osteogenic differentiation. *Stem*
1069 *Cells* **32**, 2780-93 (2014).
1070 49. Zhang, X. *et al.* Association analysis of polymorphism in the *NR6A1* gene with
1071 the lumbar vertebrae number traits in sheep. *Genes Genom.* **41**, 1165-1171
1072 (2019).
1073 50. Ehrmann, I. *et al.* An ancient germ cell-specific RNA-binding protein protects the
1074 germline from cryptic splice site poisoning. *eLife* **8**, e39304 (2019).
1075 51. Cardoso, T.F. *et al.* RNA-seq based detection of differentially expressed genes in
1076 the skeletal muscle of Duroc pigs with distinct lipid profiles. *Sci. Rep.* **7**,
1077 40005 (2017).
1078 52. Petersen, J.L. *et al.* Genome-wide analysis reveals selection for important traits in
1079 domestic horse breeds. *PLoS genet.* **9**, e1003211-e1003211 (2013).

53. Taye, M. *et al.* Exploring the genomes of East African Indicine cattle breeds reveals signature of selection for tropical environmental adaptation traits. *Cogent Food & Agric.* **4**, 1552552 (2018).

54. Li, Y. *et al.* Heat stress-responsive transcriptome analysis in the liver tissue of Hu sheep. *Genes* **10**, 395 (2019).

55. Lyon, M.S. & Milligan, C. Extracellular heat shock proteins in neurodegenerative diseases: New perspectives. *Neurosci. Lett.* **711**, 134462 (2019).

56. Yang, J. *et al.* Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Mol. Biol. Evol.* **33**, 2576-2592 (2016).

57. Atlija, M., Arranz, J.-J., Martinez-Valladares, M. & Gutiérrez-Gil, B. Detection and replication of QTL underlying resistance to gastrointestinal nematodes in adult sheep using the ovine 50K SNP array. *Genet. Sel. Evol.* **48**, 4 (2016).

58. Nakamura, H. *et al.* Identification of a human homolog of the *Drosophila* neuralized gene within the 10q25.1 malignant astrocytoma deletion region. *Oncogene* **16**, 1009-1019 (1998).

59. Zong, S. *et al.* Association of polymorphisms in heat shock protein 70 genes with the susceptibility to noise-induced hearing loss: A meta-analysis. *PLoS One* **12**, e0188195 (2017).

60. Lv, F.H. *et al.* Mitogenomic meta-analysis identifies two phases of migration in the history of Eastern Eurasian sheep. *Mol. Biol. Evol.* **32**, 2515-2533 (2015).

1101 61. Tarekegn, G.M. *et al.* Ethiopian indigenous goats offer insights into past and
1102 recent demographic dynamics and local adaptation in sub-Saharan African
1103 goats. *Evol. Appl.* **14**, 1716-1731 (2021).
1104 62. Fang, Y. *et al.* Genome-wide detection of runs of homozygosity in Laiwu pigs
1105 revealed by sequencing data. *Front. Genet.* **12**, 629966-629966 (2021).
1106 63. Benjelloun, B. *et al.* An evaluation of sequencing coverage and genotyping
1107 strategies to assess neutral and adaptive diversity. *Mol. Ecol. Resour.* **19**,
1108 1497-1515 (2019).
1109 64. Zhang, J. *et al.* Effect of domestication on the genetic diversity and structure of
1110 *Saccharina japonica* populations in China. *Sci. Rep.* **7**, 42158 (2017).
1111 65. Demirci, S. *et al.* Mitochondrial DNA diversity of modern, ancient and wild sheep
1112 (*Ovis gmelinii anatolica*) from Turkey: new insights on the evolutionary
1113 history of sheep. *PLoS One* **8**, e81952 (2013).
1114 66. Nadler, C.F., Hoffmann, R.S. & Woolf, A. G-band patterns as chromosomal
1115 markers, and the interpretation of chromosomal evolution in wild sheep (*Ovis*).
1116 *Experientia* **29**, 117-119 (1973).
1117 67. Sanna, D. *et al.* The first mitogenome of the Cyprus mouflon (*Ovis gmelini*
1118 *ophion*): New insights into the phylogeny of the genus *Ovis. PLOS ONE* **10**,
1119 e0144257 (2015).
1120 68. Poplin, F. Origine du Mouflon de Corse dans une nouvelle perspective
1121 paléontologique: par marronnage. *Ann. Genet. Sel. Anim.***11**, 133-143 (1979).

1122 69. Vigne, J.D., Carrère, I., Briois, F. & Guilaine, J. The early process of mammal
1123 domestication in the Near East: New evidence from the pre-Neolithic and pre-
1124 Pottery Neolithic in Cyprus. *Curr. Anthropol.* **52**, S255-S271 (2011).
1125 70. Yang, Y. *et al.* Draft genome of the Marco Polo Sheep (*Ovis ammon polii*).
1126 *GigaScience* **6**(2017).
1127 71. Carling, M.D., Lovette, I.J. & Brumfield, R.T. Historical divergence and gene
1128 flow: Coalescent analyses of mitochondrial, autosomal and sex-linked loci in
1129 *Passerina Buntings*. *Evolution* **64**, 1762-1772 (2010).
1130 72. Zheng, Z. *et al.* The origin of domestication genes in goats. *Sci. Adv.* **6**, eaaz5216
1131 (2020).
1132 73. Fan, R. *et al.* Genomic analysis of the domestication and post-Spanish conquest
1133 evolution of the llama and alpaca. *Genome Biol.* **21**, 159 (2020).
1134 74. An, X. *et al.* Two mutations in the 5′□flanking region of the *KITLG* gene are
1135 associated with litter size of dairy goats. *Anim. Genet.* **46**, 308-311 (2015).
1136 75. Zhang, J. *et al.* Expression and polymorphisms of *KITLG* gene and their
1137 association with litter size in sheep (*Ovis aries*). *J. Agric. Biotechnol.* **25**, 893-
1138 900 (2017).
1139 76. Pan, Y. *et al.* Indel mutations of sheep *PLAG1* gene and their associations with
1140 growth traits. *Anim. Biotechnol.* **7**, 1-7 (2021).
1141 77. Li, Y. *et al.* Mutation-388 C> G of *NR5A1* gene affects litter size and promoter
1142 activity in sheep. *Anim.Reprod. Sci.* **196**, 19-27 (2018).

1143 78. Hunter, P. The genetics of domestication: Research into the domestication of
1144 livestock and companion animals sheds light both on their "evolution" and
1145 human history. *EMBO reports* **19**, 201-205 (2018).
1146 79. Fedosenko, A.K. & Blank, D.A. *Ovis ammon*. *Mammalian Species* **2005**, 1-15
1147 (2005).
1148 In online methods:
1149 80. Sambrook, J. & Russell, D. Molecular cloning: A laboratory manual. Cold Spring
1150 Harbor Laboratory Press, New York, (2001).
1151 81. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler
1152 transform. *Bioinformatics* **25**, 1754-60 (2009).
1153 82. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for
1154 analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297-1303
1155 (2010).
1156 83. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and
1157 richer datasets. *Gigascience* **4**, 7 (2015).
1158 84. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic
1159 variants from high-throughput sequencing data. *Nucleic. Acids. Res.* **38**, e164
1160 (2010).
1161 85. Delaneau, O. & Marchini, J. Integrating sequence and array data to create an
1162 improved 1000 Genomes Project haplotype reference panel. *Nat. Commun.* **5**,
1163 3934 (2014).

1164 86. Chong, Z. *et al.* novoBreak: local assembly for breakpoint detection in cancer
1165 genomes. *Nat. Methods* **14**, 65-67 (2017).
1166 87. Chen, X. *et al.* Manta: rapid detection of structural variants and indels for
1167 germline and cancer sequencing applications. *Bioinformatics* **32**, 1220-2
1168 (2016).
1169 88. Cameron, D.L. *et al.* GRIDSS: sensitive and specific genomic rearrangement
1170 detection using positional de Bruijn graph assembly. *Genome. Res.* **27**, 2050-
1171 2060 (2017).
1172 89. Chiang, C. *et al.* SpeedSeq: ultra-fast personal genome analysis and interpretation.
1173 *Nat. Methods* **12**, 966-8 (2015).
1174 90. Layer, R.M., Chiang, C., Quinlan, A.R. & Hall, I.M. LUMPY: a probabilistic
1175 framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).
1176 91. Abyzov, A., Urban, A.E., Snyder, M. & Gerstein, M. CNVnator: an approach to
1177 discover, genotype, and characterize typical and atypical CNVs from family
1178 and population genome sequencing. *Genome Res.* **21**, 974-84 (2011).
1179 92. Jeffares, D.C. *et al.* Transient structural variations have strong effects on
1180 quantitative traits and reproductive isolation in fission yeast. *Nat. Commun.* **8**,
1181 14061 (2017).
1182 93. Lalitha, S. Primer Premier 5. *Biotech Software & Internet Report* **1**, 270-272
1183 (2000).

1184 94. Swindell, S.R. & Plasterer, T.N. SEQMAN. Contig assembly. *Methods Mol. Biol. (Clifton, N.J.)* **70**, 75-89 (1997).

1186 95. Yuan, C. *et al.* A global analysis of CNVs in Chinese indigenous fine-wool sheep populations using whole-genome resequencing. *BMC Genomics* **22**, 78 (2021).

1188 96. Schmittgen, T.D. & Livak, K.J. Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* **3**, 1101-1108 (2008).

1190 97. Nadachowska-Brzyska, K., Burri, R., Smeds, L. & Ellegren, H. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. *Mol. Ecol.* **25**, 1058-1072 (2016).

1193 98. Zhang, C., Dong, S.S., Xu, J.Y., He, W.M. & Yang, T.L. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786-1788 (2019).

1196 99. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).

1198 100. Price, A.L., Zaitlen, N.A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* **11**, 459-63 (2010).

1201 101. Cheng, J.Y., Racimo, F. & Nielsen, R. Ohana: detecting selection in multiple populations by modelling ancestral admixture components. *bioRxiv*, 546408 (2019).

1204 102. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-
1205 analysis of large phylogenies. *Bioinformatics* **30**, 1312-3 (2014).
1206 103. Korneliussen, T.S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next
1207 Generation Sequencing Data. *BMC Bioinform.* **15**, 356 (2014).
1208 104. Wang, Y., Lu, J., Yu, J., Gibbs, R.A. & Yu, F. An integrative variant analysis
1209 pipeline for accurate genotype/haplotype inference in population NGS data.
1210 *Genome Res*. **23**, 833-42 (2013).
1211 105. Bouckaert, R.R. DensiTree: making sense of sets of phylogenetic trees.
1212 *Bioinformatics* **26**, 1372-3 (2010).
1213 106. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics
1214 Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870-1874
1215 (2016).
1216 107. Petit, M. *et al.* Variation in recombination rate and its genetic determinism in
1217 sheep populations. *Genetics* **207**, 767-784 (2017).
1218 108. Pickrell, J.K. & Pritchard, J.K. Inference of population splits and mixtures from
1219 genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
1220 109. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-93
1221 (2012).
1222 110. Durand, E.Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient
1223 admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239-52
1224 (2011).

111. Leppala, K., Nielsen, S.V. & Mailund, T. admixturegraph: an R package for admixture graph manipulation and fitting. *Bioinformatics* **33**, 1738-1740 (2017).

112. Martin, S.H., Davey, J.W. & Jiggins, C.D. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244-57 (2015).

113. Teng, H. *et al.* Population genomics reveals speciation and introgression between brown Norway rats and their sibling species. *Mol. Biol. Evol.* **34**, 2214-2228 (2017).

114. Loh, P.R. *et al.* Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193**, 1233-54 (2013).

115. Corbett-Detig, R. & Nielsen, R. A hidden markov model approach for simultaneously estimating local ancestry and admixture time using next generation sequence data in samples of arbitrary ploidy. *PLoS Genet.* **13**, e1006529 (2017).

116. Huerta-Sanchez, E. *et al.* Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194-7 (2014).

117. Guerrini, M. *et al.* Molecular DNA identity of the mouflon of Cyprus (*Ovis orientalis ophion, Bovidae*): Near Eastern origin and divergence from Western Mediterranean conspecific populations. *System. Biodivers.* **13**, 472-483 (2015).

118. Zhao, Y.X. *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.* **34**, 2380-2395 (2017).

119. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44-57 (2009).

Tables

Table 1 Summary information of whole-genome variations identified in *Ovis* species.

Sequence quality/ Variation type	***O.aries*** **(n=18)**	***O.orientalis*** **（n=17）**	***O.musimon*** **(n=3)**	***O.vignei*** **(n=7)**	***O.ammon*** **(n=8)**	***O.nivicola*** **(n=8)**	***O.dalli*** **(n=6)**	***O.canadensis*** **(n=6)**
SNPs	31,535,487	53,618,832	13,360,034	30,125,980	25,160,871	22,845,295	23,185,374	23,069,044
INDELs	4,361,226	7,173,026	3,180,207	4,748,458	4,541,250	4,204,319	4,301,920	4,255,087
SVs	123,594	161,892	55,950	81,003	84,587	75,375	77,304	75,480
CNVs	74,672	92,366	37,250	56,216	52,236	48,705	48,794	48,527
Duplications	1,814	2,403	571	915	1,185	1,071	1,122	1,070
Deletions	72,858	89,963	36,679	55,301	51,051	47,634	47,672	47,457
Insertions	11,685	12,981	6,517	8,126	8,867	7,339	8,954	8,727
Inversions	746	1,011	311	499	600	583	580	576
Translocations	36,491	55,534	11,872	16,162	22,884	18,748	18,976	17,650
Average Depth (X)	19.25	27.11	18.89	19.77	17.78	17.8	19.43	18.88
Coverage Rate (%)	97.2	98.22	97.03	97.1	96.79	96.54	96.65	96.62

Figure Legends

Figure 1 Geographic distribution and population structure of *Ovis* species. (a) Geographic map of sample location and wild sheep species distribution based on the IUCN Red list (https://www.iucnredlist.org). Here we adopted the classification of Nadler *et al.* 1973. **(b)** Principal Component Analysis (PCA) of *Ovis* species. **(c)** Admixture plot using Ohana software for K from 4 to 6. Population tree of each K indicates affinity of each ancestral component (Below, right).

Figure 2 Phylogeny of *Ovis* genus. Prevalent discordance among segmental trees on autosomes and X chromosome, totaling 2,598 1-Mb segments. The segmental trees were visualized by Densitree. Consensus tree topologies of each category are shown in purple. **(b)** Phylogenetic tree of whole autosomal coding region (CDS) of 72

individuals using RAxML. Arrows marked the introgression pairs and the corresponding Z score based on the four populations test (*D* statistics), see Supplementary Table 17. Pink arrows indicate the admixed pairs which have been selected by three or more of the tests such as *D* statistics, TreeMix, *f* statistics and admixture analysis, while purple arrows indicate the introgression pairs detected in two or less of the tests.

Figure 3 Demographic inference. (a) Ancestral dynamic change of effective population size inferred by PSMC program for eight high-depth genomes. Colors of the lines indicate different species. Plots were scaled using a mutation rate of 2.5×10^{-8} per site per generation and generation time (g) of 3. Light green shading indicates interglacials (IG) in the Pleistocene and Holocene, and light blue marked as LGM the Last Glacial Maximum and gray shading indicates the mid-Pleistocene transition (MPT) and the Plio-Pleistocene transition (PPT). **(b)** LD decay analysis for seven wild species (marked as rectangles) and six domestic breeds (marked as squares). **(c)** Dynamic change of effective population size inferred by the SMC++ program for Asiatic and European mouflon and six domestic breeds (left panel) and seven wild species (right panel), the blue shading indicates the period of domestication and the gray vertical dashed line is the potential split time point of European and Asiatic mouflon. **(d)** Dynamic change of effective population size over time for all *Ovis* species.

Figure 4 Admixturegraph fitting for introgression from the Pachyceriforms into European mouflon. (a) *D* statistics of European mouflon (EMUF) with the Pachyceriforms [snow sheep (SNWS), bighorn (BIGS) and thinhorn (THNS)]. **(b)** Prior phylogeny of wild species in *Ovis* genus. **(c)** Goodness of fit of f_4 statistics. **(d)** Admixture graph.

Figure 5 Local inference and annotation of introgression signals from snow sheep to Asiatic mouflon. (a) Treemix analysis when m=9. **(b, d)** *f*3 statistics and *D* statistics of Asiatic mouflon (AMUF) with snow sheep (SNWS) pairs. Double dashed line marked as the range of threshold from -3 to 3. **(c)** Demographic diagram of admixture from snow sheep (*O.nivicola*) to Asiatic mouflon (*O.orientalis*). **(e)** Introgressed regions identified in the Asiatic mouflon genome. A modified *f*-statistic (f_d) for (OUE,AMUF;SNWS,goat), π ratio (π_w/π_d, i.e. π of Asiatic mouflon/and π of all the domestic sheep) and *F*ST between Asiatic mouflon and all the domestic sheep for 100-kb windows with 20-kb steps is plotted along the chromosomes. Each dot represents a 100-kb window. For *f*-statistic, green and blue dots above the red horizontal line correspond to the FDR 5% and FDR 1% significance level thresholds, respectively. The regions containing genes among three indexes (f_d, π ratio and *F*ST) are plotted in red dots. For the π ratio (π_w/π_d) and *F*ST, 340 domestic selection related windows are plotted, and 62 verified candidate domestication genes are marked in the plot. Overlapped genes (n=11) with *f*-statistic are marked in purple. **(f)** Venn diagram of overlapping genes (n=11 for snow sheep and n=13 for argali) between introgressed

genes (n=497 for snow sheep and n=540 for argali) and candidate domestication genes (n=62). **(g)** GO enrichment for 62 overlapped domestication genes with the previous studies.

Figure 6 Local inference of genomic region at genes introgressed from snow sheep (SNWS) into Asiatic mouflon (AMUF). **(a)** *f* statistics (f_d) based on (OUE, AMUF; SNWS, goat) comparison with (OUE, AMUF; ARGS, goat) and (OUE, ARGS; SNWS, goat) calculated for 100-kb windows with 20-kp steps across the genome for Asiatic mouflon. Each dot represented a 100-kb window, and the dashed line indicated the significance threshold ($P < 0.05$). **(b)** Population differentiation (F_{ST}) around the introgressive genomic region between recipient (Asiatic mouflon) and donor (snow sheep). **(c)** Mean pairwise sequence divergence (d_{xy}) of the introgression region between snow sheep and either Asiatic mouflon or Ouessant (OUE) domestic sheep population. **(d, e)** Haplotype patterns among all the domestic sheep, Asiatic mouflon, argali and snow sheep for the 11 genomic regions. Genes within the introgressed segments were marked upon haplotypes.

Figures

Figure 1

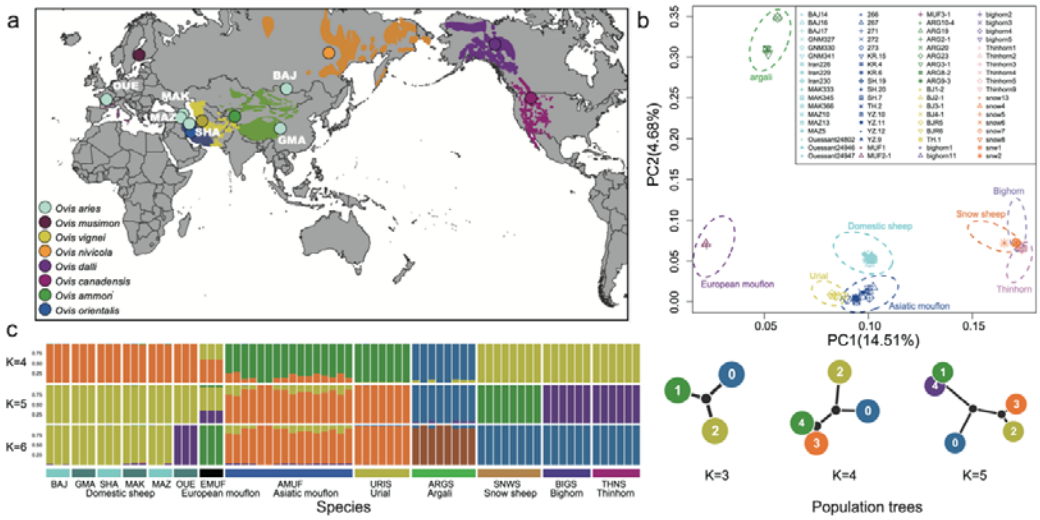

Figure 2

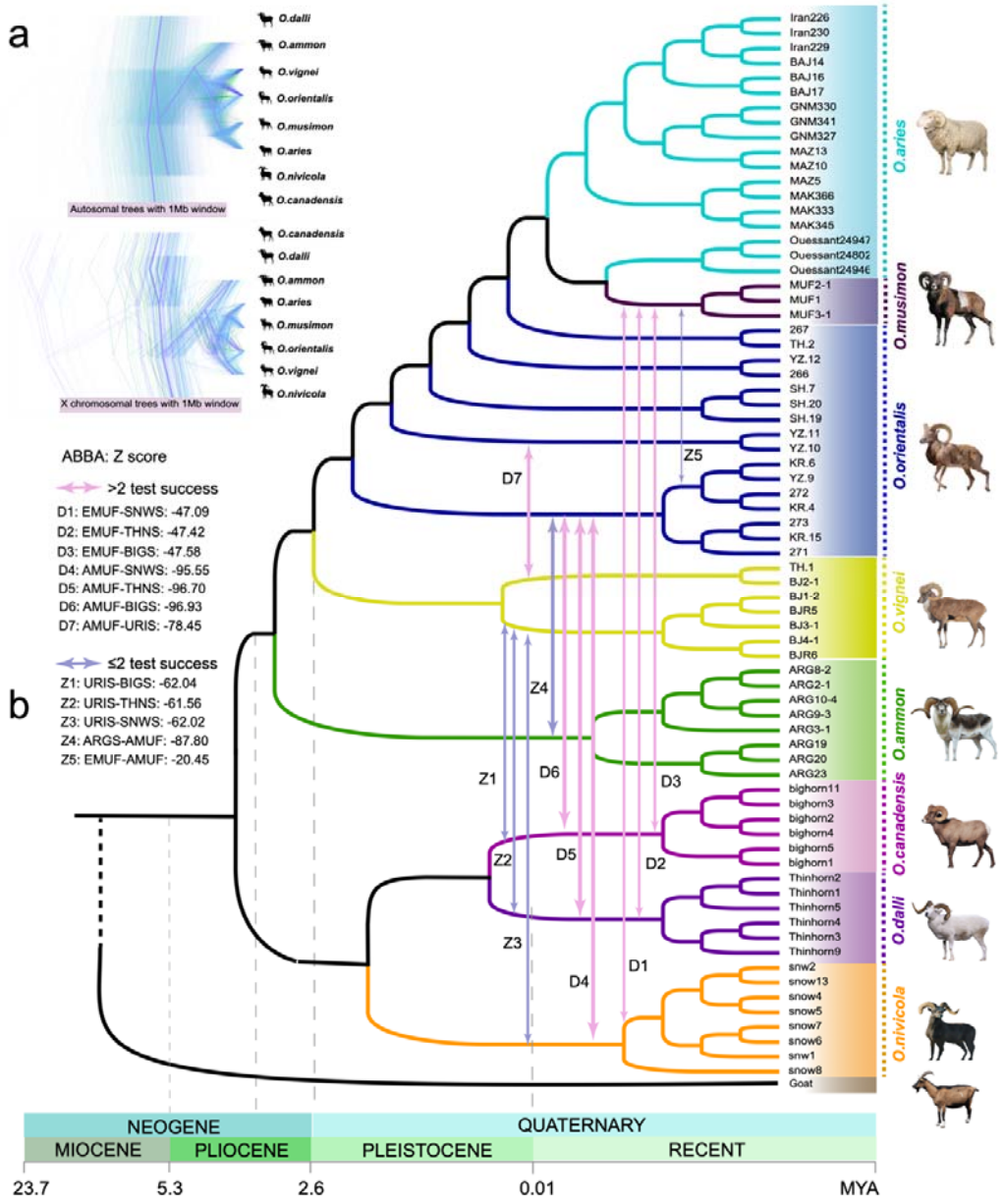

Figure 3

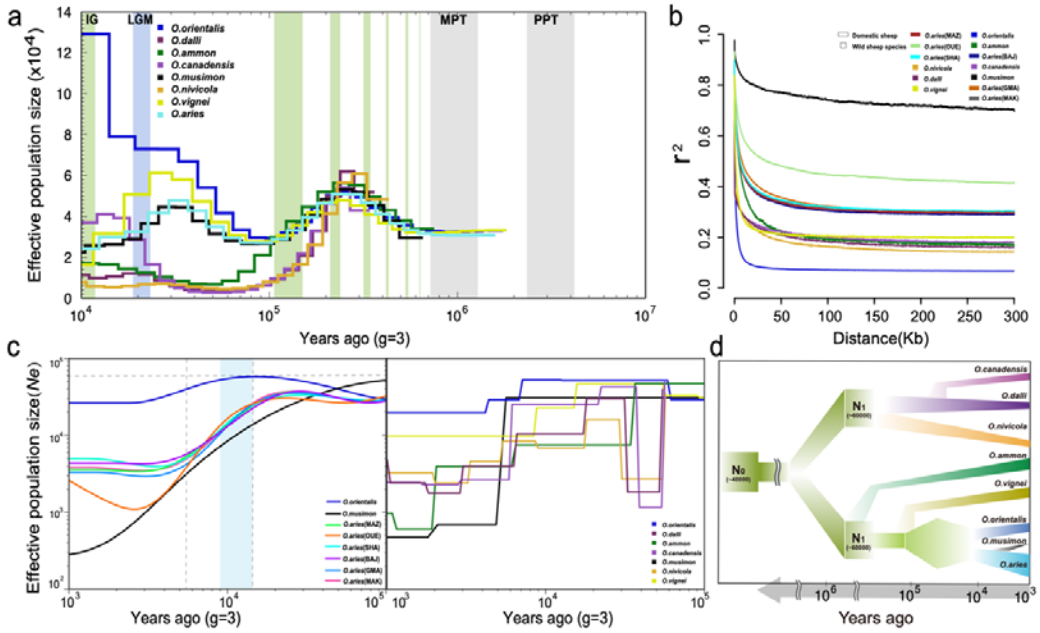

Figure 4

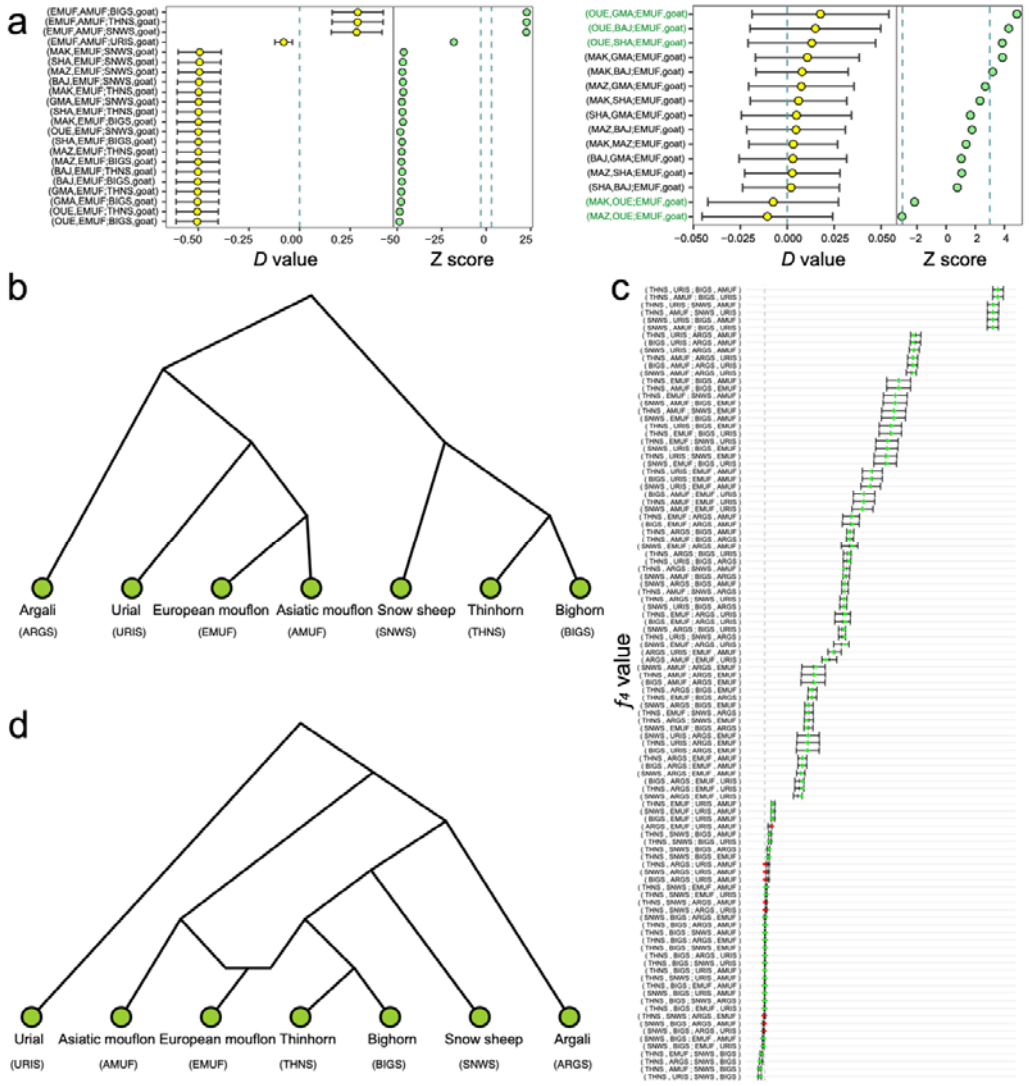

Figure 5

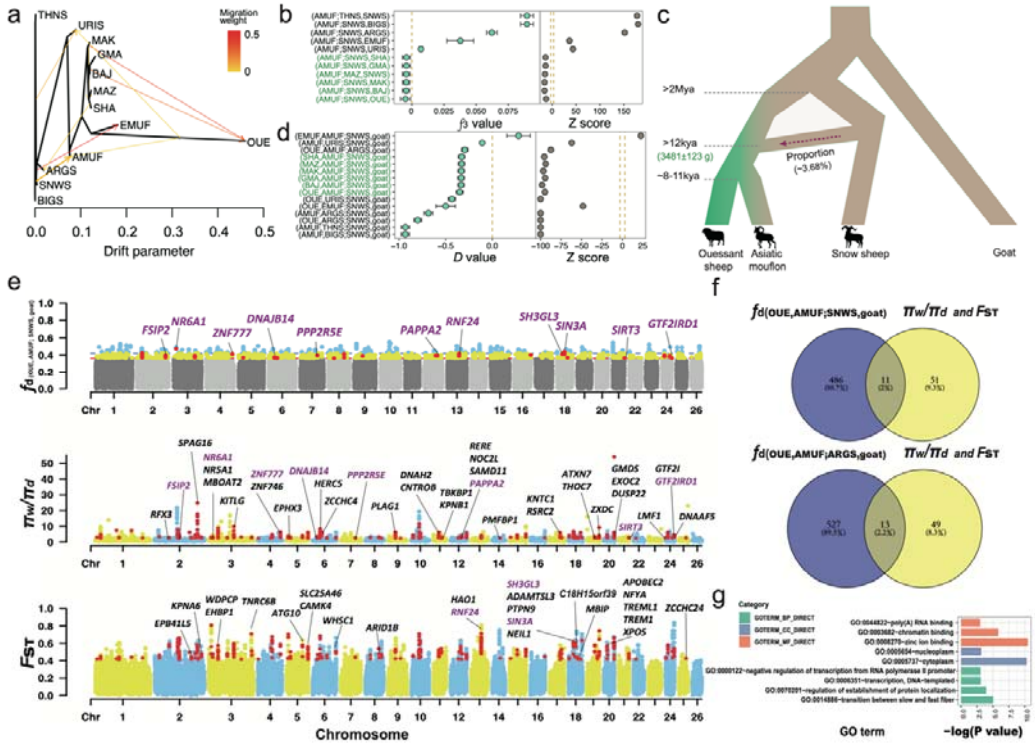

Figure 6