Genomic analyses of Asiatic Mouflon in Iran provide insights into the domestication and evolution of sheep

Dong-Feng Wang[1,2,3], Pablo Orozco-terWengel[4], Meng-Hua Li[1*], Feng-Hua Lv[1*]

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

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

[3] College of Life Sciences, University of Chinese Academy of Sciences (UCAS), Beijing, China

[4] School of Biosciences, Cardiff University, Cardiff, UK

*Corresponding author

E-mail addresses:

DFW: atongsa@163.com

POW: Orozco-terWengelPA@cardiff.ac.uk

MHL: menghua.li@cau.edu.cn

FHL: lvfenghua@cau.edu.cn

Abstract

Background

Asiatic mouflon (*O. gmelini*) consists of several subspecies mainly distributed in Armenia, southern Azerbaijan, Cyprus, northern, southern, and western regions of Iran, and eastern and central regions of Turkey nowadays. Genome analyses of Asiatic mouflon in Iran revealed that they could have diverged from the direct ancestor of domestic sheep, and showed genetic introgression into domestic sheep after domestication. However, the impact of the Asiatic mouflon subspecies in Iran on sheep domestication remains unclear.

Results

Here, we conducted a comprehensive population genomics analysis of Asiatic mouflons in Iran with 780 whole-genome sequences, 767 whole mitogenomes, and 239 Y chromosomes. Whole-genome sequence analyses revealed two subpopulations of the Asiatic mouflons in Iran: *O. gmelini*_2 limited in Kabudan and *O. gmelini*_1 over a wide geographic area. Phylogenetic analyses of the Asiatic mouflons in Iran based on uniparental variants revealed a monophyletic lineage with the mitochondrial haplogroups C+E, and clustered into a monophyletic with Y-chromosomal lineage HY2 of sheep. Additionally, introgression tests detected significant signals of genetic introgression from *O. gmelini*_2 to four sheep populations (e.g., Garut, Garole, Bangladeshi, and Sumatra) in South and Southeast Asia. In the four sheep

populations, selective tests and introgression signals revealed that the wild introgression could have contributed to their small body size and local adaptation to the hot and humid environments in the Indian Peninsula.

Conclusions

Our results suggested that the maternal haplogroups C+E and paternal lineage HY2 could have originated from the Asiatic mouflon in Iran. Also, our findings provide new insights into sheep domestication and subsequent introgressions from wild ancestors to domestic populations.

Background

Archeozoological and genetic evidence suggested that Asiatic mouflon (*O. gmelini*), the wild ancestor of sheep, have underwent a period of multigenerational human management in southeastern Anatolia by ~10,500 years B.P. [1–4]. Following domestication, sheep diffused throughout Eurasia with human migrations and developed into different populations under natural and artificial selections [5–8]. Asiatic mouflon ranges over a wide geographic area in Armenia, southern Azerbaijan, Cyprus, northern, southern, and western regions of Iran, and eastern and central regions of Turkey. It has been classified into five subspecies based on morphological differences and geographic distributions [9, 10], such as Armenian mouflon (*O. gmelini gmelini*) [11], Anatolian mouflon (*O. gmelini anatolica*) [11], Isfahan

mouflon (*O. gmelini isphahanica*), Laristan mouflon (*O. gmelini laristanica*), and Cyprus mouflon (*O. gmelini ophion*) [11]. Due to the complex intraspecies taxonomy of Asiatic mouflon, the genetic contributions of various subspecies to the domestication of sheep remains unclear [3, 12–14].

Previous genetic studies revealed controversial ancestral origins of domestic sheep. Hiendleder et al. (2002) proposed that sheep were domesticated from two different subspecies in Turkey and western Iran, corresponding to mtDNA lineages A and B of domestic sheep, respectively [13]. Through genetic and phylogeographic analyses of partial mtDNA sequences of modern and ancient samples in wild and domestic sheep, Demirci et al. (2013) suggested that domestic sheep could have originated from two distinct ancestral populations of mtDNA lineages A+B+D and C+E of domestic sheep, respectively [3]. Based on genetic, archeozoological, and geographic evidence of 212 Asiatic mouflon samples covering most of their distribution area, a recent study suggested that the maternal origin of sheep could be traced to a subspecies of *O. gmelini gmelini* in eastern Anatolia and central Zagros [4].

Despite the intensive focus on material inheritance, Y-chromosome and whole-genome analyses have received less attention. Recently, Y-chromosome and whole-genome analyses of Asiatic mouflon in Iran revealed that they could have diverged from the direct ancestor of domestic sheep, and showed genetic introgression into

domestic sheep after domestication [8, 14, 15]. Also, these studies have demonstrated that genomes of Asiatic mouflon in Iran are crucial for the development of morphological and adaptive characteristics in domestic sheep [8, 15, 16]. Nevertheless, the impact and extent of genetic introgression involving the Asiatic mouflon subspecies in Iran (e.g., *O. gmelini gmelini*, *O. gmelini isphahanica*, and *O. gmelini laristanica*) and their hybrid populations (*O. gmelini gmelini* × *O. vignei*) on the genomes of domestic sheep remains unclear.

Here, we comprehensively analyzed whole genome sequence, complete and partial mtDNA sequences, and Y-chromosome genetic variants in domestic sheep and Asiatic mouflon, covering a large distribution area in Asia, Europe, Africa, and America. We aimed to elucidate the contributions of various Asiatic mouflon subspecies in Iran to the domestication of sheep. Additionally, we investigated the impact of wild introgression of the Asiatic mouflon subspecies in Iran on the genomes of domestic sheep populations in South and Southeast Asia, which have adapted to the hot and humid environments near the Indian Ocean.

5

Methods

Datasets

Whole-genome sequences (WGS) of 780 individuals representing 158 domestic sheep populations and three wild sheep species (32 Asiatic mouflon (*Ovis gmelini*) in Iran, 9 urial (*Ovis vignei*), and 1 bighorn sheep (*Ovis canadensis*)) were obtained from publicly available data [See Additional file 1, Table S1]. Besides these, we retrieved 161 whole mitogenomes from the NCBI [See Additional file 1, Table S2]. Also, we included 333 Cytochrome b (*Cytb*) sequences of Asiatic mouflon in the integrated analysis [See Additional file 1, Table S3].

Variant calling

We implemented quality-based trimming for WGS data with Trimmomatic v0.39 [17] and mapped high-quality reads to the reference genome *Oar_rambouillet_v1.0_CM022046.1*. (https://www.ncbi.nlm.nih.gov/assembly/GCF_002742125.1/; https://www.ncbi.nlm.nih.gov/nuccore/CM022046.1/, last accessed May 12, 2020) using the Burrows–Wheeler aligner (BWA mem) v.0.7.8 [18] with the default parameters. Then, we removed duplicates with the *MarkDuplicates* module in GATK v 4.1.2.0 [19] and called SNPs following the previous procedures [15]. We used the

6

HaplotypeCaller module in GATK v 4.1.2.0 to generate the GVCF file of each sample and merged the raw GVCF file of each individual with the *CombineGVCFs* module. Afterward, we genotyped SNPs using the *GenotypeGVCFs* module and filtered SNPs using the *VariantFiltering* module of GATK with the parameters "QUAL < 30.0 || QD < 2.0 || MQ < 40.0 || FS > 60.0 || SOR > 3.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0". We obtained a total of 121,252,917 high-quality autosomal SNPs and 203,938 Y-chromosomal SNPs for the following analyses.

Mitochondrial DNA assembly

We assessed the mitogenome depth of each sample using the SAMtools v.1.9 [20] and kept only the samples with depth > 100 in the below analyses. We assembled the complete mitogenomes using the following protocols. We first converted BAM files to the fastq format. Then, we generated the mitogenome consensus sequence from the fastq files using MIA v1.0 (https://github.com/mpieva/mapping-iterative-assembler).

In addition to the 605 mitogenomes assembled here, we also included 161 whole mitogenome sequences retrieved from the NCBI GenBank in the analysis, with a final mitogenome dataset of 767 sequences from 727 domestic sheep, 31 Iranian Asiatic mouflon, and 9 urial. All sequences were aligned using MAFFT v7.515 [21] and trimmed with trimAl v1.4 [22].

Y-chromosomal SNPs

To obtain single-copy Y-chromosomal SNPs, raw SNPs in the 780 samples were processed as follows. Firstly, we identified ram samples from the whole samples using PLINK v1.90b6.26 [23] with the "--check-sex" option and kept only these ram samples in the following analyses. We then excluded SNPs in the regions covering the pseudoautosomal regions (PARs) and the highly repetitive sequences (around 1.06 Mb) [14]. Afterwards, we filtered SNPs with the mean sequencing depth ranging from 0.5 × to 2 × of the expected depth (half of the genome-wide depth) [14]. Further, we removed SNPs that were heterozygous in at least one of the male samples. Finally, we retained 4,096 Y-chromosome SNPs for the haplotype and phylogenetic analyses.

Population structure analysis

To reveal the genetic differentiation within the Asiatic mouflon in Iran, we examined the population genetic substructure for the total 779 wild and domestic samples using a set of autosomal SNPs filtered as follows. First, we filtered the total 25,270,188 high-quality SNPs with missing rates < 0.1 and minor allele frequencies > 0.05 (MAF > 0.05) using PLINK v1.90b6.26 [23]. Then, we pruned the remaining SNPs with the options "--indep-pairwise 50 10 0.1" of PLINK v1.90b6.26 [23]. Afterwards, 1,566,300 SNPs were retained for the population genetic structure analysis.

8

We computed the identity by state among samples using PLINK v1.90b6.26 [23] with option “distance square 1-ibs”. Next, we generated a neighbor-joining (NJ) tree using the function “bionj” in the R package *ape* v5.6.2 [24]. The tree was rooted using urial (*O. vignei*) as the outgroup, and its graphical representation was obtained using the R package *ggtree* v3.6.2 [25]. Principal component analyses (PCA) were performed using the “--pca” option in PLINK v1.90b6.26 [23] with the default settings. Furthermore, model-based clustering was performed using the sNMF v1.2 with clusters *K* from 2 to 10 [26].

The F_{ST} and nucleotide diversity (π) of autosomes were calculated with vcftools v0.1.17 [27] using 2Mbp windows and 1Mbp steps, and *P* value was calculated with 1000 permutation test. The F_{ST} of mitochondria and Y chromosome were calculated with R package hierfstat v0.5.11 [28] with 1000 permutation test. The haplotype diversity of mitochondria and Y chromosome were calculated with pegas v1.1 [29] with 1000 permutation test.

Phylogenetic analysis

Phylogenetic analyses based on mitogenome and Y-chromosomal SNPs were performed using the maximum likelihood (ML) algorithm of iqtree2 v2.0.7 [30].

9

179 Mitogenome sequences were aligned using mafft v7.515 [21] with the default setting,
180 and multiple sequence alignment (MSA) was trimmed to remove ambiguously aligned
181 regions using trimAl v1.4 [22] with parameters "-automated1". To keep high-quality
182 SNPs, we further removed all the sites with gaps and ambiguous nucleotide bases and
183 only kept parsimony informative sites using *ClipKIT* [31]. We tested the potential
184 models of DNA sequence evolution using the option "-m" test in iqtree2 v2.0.7 [30]
185 and found the generalized time reversible model (GTR + G) [32] was the best fitting
186 substitution model for mitogenome. We constructed a maximum likelihood (ML) tree
187 with the site model GTR + G with 1000 bootstrap replicates using iqtree2 v2.0.7 [30].
188 The resulting tree was rooted with bighorn as the outgroup and visualized using the
189 FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). Mitogenome haplotypes
190 were inferred with the function "haplotype" in the R package *pegas* v1.1 [29]. We
191 also constructed the parsimony mitogenome haplotype network using POPART [33].
192 We transformed the genotypes of Y-chromosomal 4,096 SNPs into the fasta format
193 using vcf2phylip.py [34]. Phylogenetic trees were built and visualized following the
194 same pipelines and using the same programs and substitution model as above.
195

196 **Demographics and divergence time inference**

197 The population genetic structure analyses above identified two subpopulations (*O.*
198 *gmelini*_1 and *O. gmelini*_2) in the Asiatic mouflons in Iran. We inferred the dynamic
199 changes in historical and recent effective population size (N_e) of the two

200 subpopulations using SMC++ v1.15.2 algorithms [35]. We selected 7 - 10 high-depth
201 individuals in each of the six geographically differentiated genetic groups of domestic
202 sheep (Central-and-East Asian, South-and-Southeast Asian, the Middle Eastern,
203 African, European, and American populations) [15] and the two Asiatic mouflon
204 subpopulations. We estimated the N_e during the past 1,000 to 100,000 years for each
205 group using SMC++ v1.15.2 [35] under the default setting.
206
207 For the SMC++ analysis, the reference genome's low-complexity regions were
208 masked using snpable (http://lh3lh3.users.sourceforge.net/snpable.shtml, last accessed
209 October 20, 2022) using in-house scripts. We set the mutation rate and generation
210 time to be 1.00×10^{-8} per nucleotide site per generation [36] and 3 years [37, 38],
211 respectively. To infer the divergence times between *O. gmelini*_1 and *O. gmelini*_2 as
212 well as between the Asian mouflon and domestic sheep, we employed the split option
213 of SMC++ with the same parameter settings as above.
214
215 Additionally, we inferred the demographic history of domestic sheep and Asiatic
216 mouflon based on mitogenomes and Y-chromosome data using Beast v2.7.3 [39] and
217 the Coalescent Bayesian Skyline plot [40]. Firstly, we used a coalescent model to
218 estimate the mutation rates for mitogenome and Y chromosome chromosomes. The
219 analysis was performed using the GTR + GAMMA (4) + I site model and the strict
220 clock model. The most recent common ancestor (MRCA) of Asiatic mouflon and

221 domestic sheep was set to be 0.653 million years [41]. The chain length was set to be
222 40,000,000 with a burn-in of 2,000,000. The results were analyzed using Tracer
223 (http://beast.community/tracer), assuming mutation rates estimates of 2.10×10^{-8} and
224 7.70×10^{-9} substitutions per site per year for the mitochondrial and Y chromosome
225 SNPs, respectively.
226
227 Subsequently, we utilized the mutation rates estimated above to infer the demographic
228 history of maternal and paternal lineages in domestic sheep and Asiatic mouflon. We
229 used the same GTR + GAMMA (4) + I site model as above, with the clock model set
230 as strict and the mutation rates of 2.10×10^{-8} and 7.70×10^{-9} per site per year for
231 mitochondria and Y-chromosome, respectively. The prior model remained the
232 Coalescent Bayesian Skyline, with the chain length of 40,000,000, a burn-in of
233 2,000,000, and samples drawn every 4,000 steps. The obtained results were further
234 analyzed using Tracer, and each analysis yielded over 200 effective samples for all
235 relevant parameters.
236

237 **Introgression analysis**

238 We conducted introgression analysis to clarify the genetic contributions of the two
239 Asiatic mouflon subpopulations to South-and-Southeast Asian (SSA) sheep. We
240 tested introgression using a combination method of D and f_{dM} statistics. Specifically,
241 we calculated D statistics [42] with a four taxon model (((P1, P2), P3), P4) using

242 Dsuite v0.5r50 [43]. In this analysis, Menz sheep (MEN) was selected as the reference
243 population (P1), while SSA populations were the target population (P2), *O. gemlini*_1
244 or *O. gemlini*_2 was the donor (P3), and *O. canadensis* was the outgroup (P4). A *P*
245 value < 0.05 from a standard block-jackknife procedure was considered as evidence
246 of gene flow from P3 to P2.
247
248 We further implemented sliding window statistics analysis using the Dsuite v0.5r50
249 [43] to quantify the introgressions. We calculated f_{dM} values using a sliding window of
250 50 SNPs with a step of 25 SNPs across the genomes with the defined trios (MEN,
251 SSA, *O. gmelini*_2), where SSA represents the four South-and-Southeast Asian sheep
252 populations suggested to be the introgressed populations by the *D*-statistics and f_4-
253 ratio tests. We calculated the Z-transformed f_{dm} values, and the windows with $P < 0.05$
254 and $D > 0$ was defined as the potential introgressed genomic regions. We evaluated
255 the statistical significance using one-tailed *Z*-test in R package BSDA v1.2.1 [44].
256

257 **Selective sweep tests**

258 To assess the potential roles of genomic introgression from Iranian Asiatic
259 mouflons in Southern and Southeastern Asian (SSA) sheep populations, we conducted
260 a PBS analysis [45] to identify selective signatures in four introgressed populations
261 (i.e., Garut (GUR), Garole (GAR), Bangladeshi (BGE), and Sumatra (SUM)). We
262 pooled these four populations (GUR, BGE, SUM, and GAR) as an introgressed group,

263 and used the MEN and *O. gmelini*_2 as the reference populations. We calculated the
264 PBS values using PBScan v2020.03.16 with parameters "-win 50 -step 25" described
265 in Hämälä and Savolainen (2019) [46] and estimated *P* values by 2000 permutation
266 cycles for Monte Carlo testing. and detected selective regions with top 1% PBS
267 values. Subsequently, we identified the overlaps between the selective and the
268 introgressed regions using BEDtools v2.30.0 [47] and annotated functional genes in
269 the overlapping genomic regions.
270

271 **Gene annotation and function analyses**

272 We annotated genes in the introgressed regions using the sheep reference assembly
273 *Oar_rambouillet_v1.0* by BEDtools v2.30.0 [47]. We also used the sheepQTLdb
274 (https://www.animalgenome.org/cgi-bin/QTLdb/OA/index) [48] to detect QTLs
275 associated with the introgressed regions. We examined the LD blocks by
276 LDBlockShow v1.40 [49]. The LD decay was calculated with PopLDdecay v3.42
277 [50]. d_{xy} and F_{ST} for the introgressed region is calculated with Pixy 1.2.7.beta1 [51]
278 with a 20kb sliding window with 2kb steps. The haplotype heatmap of regions near
279 the introgression region was plotted with R package ComplexHeatmap v2.14.0 [52].
280

Results

Genetic variants, mitogenome assembly, and Y-chromosome specific SNPs

A total of 121,252,917 SNPs in the 780 wild and domestic sheep samples were obtained for downstream analyses. We obtained an average depth of 923.33 × (101.92 × - 8,324.72 ×) for the mitogenome assemblies after aligning all mitochondrial reads from the 606 samples to the reference mitochondrial genome NC_001941.1 [See Additional file 1, Table S2]. After quality control for the SNP identification on the Y-chromosome, we identified 4,096 Y-specific SNPs in a cohort of 239 rams with an average read depth of 8.26 × (~4.25 × - 16.99 ×) for SNP [See Additional file 1, Table S4].

Population genetic structure of the Asiatic mouflons in Iran

We observed two subpopulations in the 32 Asiatic mouflons in Iran (Fig. 1) with autosome SNPs. The larger subpopulation (*O. gmelini*_1) comprised 25 samples across eastern and southern Iran, while the smaller subpopulation (*O. gmelini*_2) consisted of only 7 samples limited in northwestern Iran (Fig. 1a) and [See Additional file 1, Table S2]. Low kinship values within two subpopulations eliminate inbreeding caused by sampling [See Additional file 1, Table S5]. We also observed significant

genetic divergence and different genetic diversity patterns [See Additional file 1, Table S6] and [See Additional file 1, Table S7]. The result makes sure that the divergence is not the artefact of inbreeding.

Phylogenetic analyses of the total 780 wild and domestic sheep samples identified the same two subpopulations of Asiatic mouflon (*O. gmelini*_1 and *O. gmelini*_2) and six genetic clusters of domestic sheep, which were congruent with their geographic distribution [See Additional file 2, Figure S1] and [See Additional file 2, Figure S2]. Admixture analysis showed inter- and intra-species genetic divergence similar to the phylogenetic tree (Fig. 1d).

[Figure 1 here]

Phylogenetic and haplotype analysis based on uniparental variants

We observed 593 mitochondrial haplotypes from the 767 mitogenomes of wild and domestic sheep (Fig. 2a) and [See Additional file 1, Table S2]. The phylogenetic tree inferred from the mitochondrial haplotypes revealed three monophyletic clades (clades I, II, and III) (Fig. 2b). Clade I comprised of haplogroups A, B, and D of domestic sheep. Clade II consisted of the domestic sheep's haplogroups C and E and the Asiatic mouflon samples, while the remaining Asiatic mouflon samples clustered with the urial in clade III (Fig. 2b).

320

321 For Y-chromosomal SNPs, the lineage HY2 of domestic sheep was closely related to
322 Asiatic mouflon in the haplogroups C and E but diverged from the main lineage HY1
323 (Fig. 2c) and [See Additional file 1, Table S8]. However, phylogenetic and network
324 analysis revealed no shared mitochondrial haplotypes and only one shared Y-
325 haplotypes of HY2 between domestic sheep and Asiatic mouflon in Iran (Fig. 2d and
326 e). We observed significant genetic divergence (*O. gmelini*_1: $F_{ST} = 0.11$, $P < 0.05$;
327 *O. gmelini*_2: $F_{ST} = 0.21$, $P < 0.05$) between Asiatic mouflons and domestic sheep in
328 whole genome levels [See Additional file 1, Table S6]. The observation implied that
329 the Asiatic mouflon in Iran might not be the direct ancestor of domestic sheep.
330 Instead, the relatively close phylogenetic relationships between the Asiatic mouflons
331 in Iran and domestic sheep suggested by haplogroups C, E, and Y-chromosomal
332 haplogroupY2 implied their possible gene flows.
333 **[Figure 2 here]**
334

335 **Effective population size and divergence time**

336 Inference of demographic history showed the Asiatic mouflons and domestic
337 sheep shared a similar historical pattern 10,000 years ago (Fig. 3a), confirming the
338 Asiatic mouflon as the wild ancestor of sheep [53]. SMC++ analysis revealed a
339 continual decline of N_e in domestic sheep and Asiatic mouflon during ~8,000–10,000

340 years ago. Nevertheless, the two Asiatic mouflon populations in Iran showed slower
341 declines than domestic sheep (Fig. 3a), which probably reflected the bottleneck effect
342 of domestication. A sharp decline of N_e was detected ~1,000-5,000 years ago in the *O.*
343 *gmelini*_2, demonstrating a pattern different from that for the *O. gmelini*_1 (Fig. 3a).
344 We observed similar estimates of divergence time between *O. gmelini*_1 and the six
345 geographically differentiated genetic groups of domestic sheep during ~9,000-15,000
346 years ago (Fig. 3b) and [See Additional file 1, Table 9], which was congruent with the
347 domestication time of sheep [2]. Additionally, we found the splitting time between the
348 two Asiatic mouflon populations in Iran around 5,000 years ago (Fig. 3b) and [See
349 Additional file 1, Table 9], indicating a recent genetic divergence of Asiatic mouflons
350 in Iran (Fig. 1). For the divergence pattern inferred by WGS of the Asian mouflons,
351 the subpopulation *O. gmelini*_2 clustered with partial individuals of the subpopulation
352 *O. gmelini*_1 in the mitogenome and Y-chromosomal phylogenetic trees (Fig. 1 and
353 3). The discordant phylogenetic patterns of Asiatic mouflons in Iran inferred by
354 uniparental and autosomal genetic variants could be explained by the recent
355 hybridization events [54].
356
357 We conducted Bayesian skyline analyses to estimate N_e based on mitogenomes
358 and Y-chromosomes. Based on the estimated TMRCA (time to the most recent
359 common ancestor) of Asiatic mouflons and domestic sheep [41], three lineages
360 (lineage o.HCE: haplogroup of Asiatic mouflons in Iran and partial domestic sheep of

361 haplogroups C and E, lineage A+B+D: haplogroups A, B, and D, and lineage C+E:
362 haplogroups C and E) of mitogenome showed different patterns. Lineage A+B+D
363 showed a steep increase in N_e at approximately ~25 k years ago. For lineage C+E, it
364 also increased at approximately ~25 k years ago, while lineage o.HCE demonstrated
365 relatively fewer changes since approximately ~100 k years ago (Fig. 3c). Also, we
366 observed different patterns in the three Y-chromosomal lineages (lineage o.HY2:
367 haplogroups of Asiatic mouflons in Iran and partial domestic sheep of HY2, lineage
368 HY1, and lineage HY2). Lineages o.HY2 and HY2 displayed slow decline in N_e at
369 approximately ~25 k years ago, while lineage HY1 changed sightly since
370 approximately ~100 k years ago (Fig. 3d). Thus, uniparental genetic markers showed
371 similar demographic patterns between o.HY2 and HY2 as well as between HCE and
372 C/E, but different patterns between the dominant linages A+B+D and HY1 (Fig. 2 and
373 3). The results suggested that the Asiatic mouflons in Iran might not be the direct
374 ancestor of domestic sheep but be involved in the genetic introgression.
375 **[Figure 3 here]**
376

377 **Phylogeography of Asiatic mouflon**

378 We built a *Cyt-b* dataset consisting of 247 Asiatic mouflons and 8 European
379 mouflons. The samples of Asiatic mouflon covered the current distribution of all the
380 subspecies (Fig. 4a) and [See Additional file 1, Table S3]. To clarify phylogenetic
381 relationships between different Asiatic mouflon subspecies and domestic sheep, we

382 performed phylogenetic analysis of *Cyt-b* sequences of Asiatic mouflons and
383 domestic sheep using argali as the outgroup. In the phylogenetic tree built from *Cyt-b*,
384 Asiatic mouflon was assigned to 7 major clades supported by high bootstrap values
385 (94~100) (Fig. 4b). The topology confirmed the five mitogenome clades of domestic
386 sheep. Haplogroups C+E clustered into monophyletic clade, which diverged from the
387 haplogroups A+B+D (Fig. 4b). It was noteworthy that most of the hybrid samples
388 inferred above were integrated with *O. gmelini* or *O. vignei* samples, and only around
389 10% of samples clustered with subspecies such as *O.g.ophion* and *O.g.anatolica* (Fig.
390 4b) and [See Additional file 1, Table S10]. Clade 3 comprised haplogroups C+E of
391 domestic sheep and the Asiatic mouflons in Iran with high frequencies of shared
392 haplotypes between them (Fig. 4c) and [See Additional file 1, Table S10].
393 Haplogroups A+B+D of domestic sheep and the Asiatic mouflons from Iran and
394 Turkey fell within clade 6, but no common haplotypes were detected between
395 domestic and wild sheep (Fig. 4c) and [See Additional file 1, Table S10]. The results
396 were in accordance with the mitogenome analyses and implied that partial haplotypes
397 C+E in domestic sheep could have originated from the Asiatic mouflons in Iran by
398 introgression.
399 **[Figure 4 here]**
400

401 ## Genetic introgression

402 To investigate the impact of introgression from the Asiatic mouflons in Iran on

403 domestic sheep, we implemented introgression tests between Iranian Asiatic mouflon
404 and SSA by *D* and f_4-ratio statistics [55, 56]. We first conducted an admixture
405 analysis between Asiatic mouflons and the SSA populations. We identified three
406 genetic clusters with the smallest cross validation error ($K = 3$) (*O. gmelini*_1, *O.*
407 *gmelini*_2, and SSA) and detected genetic components of Asiatic mouflons in Iran in
408 some domestic populations (Fig. 5a) and [See Additional file 2, Figure S3]. Further,
409 we employed the ABBA model (((P1, P2), P3), P4) to assess introgression between *O.*
410 *gmelini*_1 or *O. gmelini*_2 (P3) and SSA sheep populations (P2). We used Menz
411 sheep (MEN) from Africa as P1, which did not show genetic introgression from
412 Asiatic mouflon and SSA populations [8, 15], and bighorn sheep as the outgroup P4,
413 which inhabit North America and have no gene flows with SSA sheep populations. Of
414 the two Asiatic mouflon populations, *O. gmelini*_1 or *O. gmelini*_2 served as P3 and
415 18 SSA populations as P2. We calculated the estimates of Patterson's *D*-statistic and
416 f_4-ratio for all the 18 SSA sheep populations, and determined introgression based on
417 the following criteria: $D > 0$, P-value < 0.05, and f_4-ratio > 0. Only four sheep
418 populations (Garut, Garole, Bangladeshi, and Sumatra sheep) showed genetic
419 introgression from *O. gmelini*_2 (Fig. 5b) and [See Additional file 1, Table S11]. All
420 the four sheep populations are distributed in the regions near to the Indian Ocean.
421

422 To further examine the genetic introgression in the four sheep populations, we
423 calculated f_{dm} with a 50-SNP window and 25-SNP step across the sheep genome. We

424 identified introgressed regions genomic with top 5% f_{dm} and $D > 0$. We obtained a
425 common introgression region in the four populations spanning 6,384,279 bp (Fig. 5c),
426 [See Additional file 1, Table S12] and [See Additional file 1, Table S13]. We then
427 calculated the PBS value of the four populations to identify regions under selection
428 with MEN sheep and *O. gmelini*_1 as the reference populations. We considered the
429 top 1% of PBS regions as being under selection (Fig. 5d) and [See Additional file 1,
430 Table S14]. We overlapped the selected regions with the common introgressed
431 regions and detected 19,614 bp region as the selected introgressed region [See
432 Additional file 1, Table S15]. We further overlapped the introgressed genomic region
433 with sheep QTLs [See Additional file 1, Table S15]. We found that it overlapped with
434 the QTLs for body_weight, carcass_fat, and fat_weight QTLs, which might be
435 associated with the small body size of Garole, Bangladeshi, and Garut sheep [57–59].
436 Additionally, the Nematodirus_FEC QTL could be related to the phenotypes of
437 Sumatra sheep's susceptibility to nematode parasitism [55]. The results implied that
438 the adaptive selected introgressed region could have been introgressed from *O.*
439 *gmelini*_2, and then have undergone the strong selection to facilitate their adaptation
440 to the hot and humid environments near the Indian Ocean.
441 **[Figure 5 here]**
442

Discussion

Traditionally, Asiatic mouflon had been classified into five subspecies using phenotypic traits, such as coat color, horns, and geographic patterns [11]. However, the intraspecies taxonomy has long lacked genetic evidence. Previous phylogenetic analysis based on partial mtDNA divided Asiatic mouflon into several clades [3, 4, 58], different from the taxonomy of five subspecies defined by phenotypic traits and geographic patterns [11]. Here, we identified two genetic clusters of the Asiatic mouflons in Iran based on WGS data: *O. gmelini*_2 limited in Kaboudan, and *O. gmelini*_1 over a wide geographic area in Iran (Fig. 1). We detected the divergence time of two clusters around ~5,000 years ago (Fig. 3b) and [See Additional file 1, Table S9], which indicated a recent split of them, and was further supported by the inferred phylogenies based on uniparental genetic markers (Fig. 2b and c). All the individuals of *O. gmelini* _2 collected in Kabudan inhabited the overlap geographic region with Armenian Mouflon (*O. g.gmelini*), while individuals of *O. gmelini* _1 were sampled from the distribution areas of several subspecies such as Armenian Mouflon (*O. g.gmelini*), Isfahan mouflon (*O.g.isphahanica*), and Laristan mouflon (*O.g.laristanica*) (Fig. 1a) [10].

These results suggested a significant genetic divergence between Armenian Mouflon and the other subspecies [See Additional file 1, Table S6]. Nevertheless, the absence

23

463 of phenotypic data and other sample information for the Asian mouflons and a
464 comprehensive sampling throughout Iran limited our in-depth understanding of
465 within-species genetic differentiation in Asiatic mouflon. Also, Asiatic mouflon is
466 under the status of "Near Threatened", and their population size has declined in the
467 past two decades under competition with domestic sheep, poaching, and habitat
468 deterioration [10, 60]. Therefore, the population genetic structure analyses and
469 delimitation of taxonomic classification based on WGS data are fundamental for
470 conservation decisions and effective conservation management [61, 62].
471
472 Genetic variants of uniparental inheritance offer insights into the domestication
473 and evolutionary history of farm animals, such as pigs [63], sheep [5, 14] and goats
474 [64, 65]. In domestic sheep, here we identified five maternal haplogroups (A, B, C, D,
475 and E) in two monophyletic groups (A+B+D and C+E), and two paternal clades (HY1
476 and HY2) (Fig. 2b and c). We found that haplotypes of most of the Asiatic mouflons
477 in Iran clustered with haplogroups C+E based on mitochondrial marker (24/31) and
478 grouped with haplogroup HY2 based on Y-chromosomal SNPs (12/21) [See
479 Additional file 1, Table S8]. Nevertheless, the Asiatic mouflons in Iran showed higher
480 divergence from the dominant maternal lineages A and B and paternal lineages HY1
481 (Fig. 2c and d).
482
483 Additionally, *Cyt-b* sequences collected from the whole geographic areas of

484 Asiatic mouflon also revealed a closer relationship between Asiatic mouflon in Iran
485 and domestic sheep of maternal haplogroups C+E (Fig. 4). The results supported the
486 hypothesis that mtDNA haplogroups C+E and Y-chromosomal haplogroup HY2
487 could have originated from Asiatic mouflons in Iran [14]. However, WGS analysis
488 revealed intraspecies substructure in the Asiatic mouflons in Iran, and genetic
489 introgression from the *O. gmelini*_2 showed a larger impact on domestic sheep than
490 *O. gmelini*_1 (Fig. 5). Similar results were observed from Alberto et al., (2018) that
491 IROO individuals were split into subpopulations and had admixture with some
492 domestic sheep [38]. Therefore, Iran's two Asiatic mouflon populations showed
493 similar phylogenetic patterns of uniparental inheritance but different introgression
494 patterns of WGS, implying multiple introgression events from the Asiatic mouflons to
495 domestic sheep. In summary, our results implied that Asiatic mouflons in Iran might
496 not be the ancestor of domestic sheep [4, 14], but could have contributed to the gene
497 pool of domestic sheep through multiple introgressions. Also, the results implied the
498 complexity of sheep domestication [4, 66].
499
500 Recent studies suggested that sheep diffused to South and Southeast Asia via the
501 Middle East [15]. Our results revealed the genetic introgression from Asiatic mouflon
502 into some South-and-Southeast Asian sheep populations, supporting previous findings
503 [8, 15]. Further, we detected genetic introgression in the four sheep populations of
504 Garut, Garole, Bangladeshi, and Sumatra, and were only from *O. gmelini*_2 in

505 Kabudan (Fig. 5b). The introgression event is congruent with the demographic history
506 of *O. gmelini*_2 and South-and-Southeast Asian sheep inferred here and previously:
507 South-and-Southeast Asian sheep descended from Central-and-East Asian populations
508 around 5.85 to 6.01 ka and then admixed from the Middle Eastern sheep 3.93–4.18 ka
509 [15], and *O.gmelini*_2 split from *O.gmelini*_1 around ~5,000 years ago (Fig. 3b).
510
511 In the four sheep populations, we found that common introgression signals from *O.*
512 *gmelini*_2 covered about 4.29% of the total introgressed regions. In contrast to
513 population-specific introgression, common introgression could be related to a specific
514 trait or adaptation to similar environments, such as *EPAS1* for the adaptation to
515 hypoxia [45, 67]. We further integrated *PBS* analyses and found four positive
516 selective regions in the common introgressed regions, which were related to
517 carcass_fat_percentage, lean_meat_yield_percentage (Fig. 5d). This might have
518 contributed to the small body size phenotype of Garole, Sumatra, and Bangladeshi
519 sheep [58, 68, 69]. We also found a candidate gene *PCDH9* downstream ~679 kb of
520 the AI region (Fig. 5e), which had been reported to be associated with local adaptation
521 (e.g., hot and arid environments) in sheep and goats [70, 71] and fat deposition [72].
522 The gene showed a higher linkage with the AI region (Fig. 6a and b). The AI region
523 was confirmed by d_{XY} and F_{ST} of Menz sheep (MEN) and introgressed breeds (BGE,
524 GAR, GUR, and SUM) from SSA (Fig. 6c). We observed similar haplotype patterns
525 among introgressed breeds and *O.gmelini*_2 in the AI region contrasting with MEN.

526 In contrast, no similar haplotype pattern was found in the exon of *PCDH9* (Fig. 6d).
527 These results implied that the AI region may regulate the function of *PCDH9*. Also,
528 early studies showed adipogenesis, ECM remodeling, inflammation, and lipid droplet
529 dynamics in mature adipose tissue played essential roles in response to climate
530 changes in sheep [73]. Further analysis from Chip-seq and RNA-Seq data will be
531 required to uncover the function and regulatory mechanisms of *PCDH9* to adapt to
532 hot and humid environments in sheep.
533 **[Figure 6 here]**
534

535 Conclusions

536 In summary, we conducted a comprehensive and integrated analysis of Asiatic
537 mouflon and domestic sheep based on WGS, mitogenome and Y-chromosomal
538 genetic variants. Our results suggested that the mitochondrial haplogroups C/E and Y-
539 chromosomal haplogroup HY2 of domestic sheep could have originated from the
540 Asiatic mouflons in Iran by introgression and further clarified the effects of the
541 Iranian Asiatic mouflon in the domestication processing. Additionally, our results
542 revealed the introgression signals of the Asiatic mouflons in Iran in several domestic
543 sheep populations in South and Southeast Asia and showed that the introgression
544 could have facilitated their adaptation to hot and humid environments near the Indian
545 Ocean. Our findings added new insights into the domestication and subsequent

evolutionary history of sheep.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The whole genome re-sequence data used for the study is publicly available under the sample accession numbers listed in Additional file 1 Table S1. All mitogenomes assembled in this study were deposited in NCBI's SRA under the accession number OR160429-OR161034. All scripts used for this work were performed using open-source software tools and are available from the corresponding authors upon request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was financially supported by grants from the National Natural Science

28

Foundation of China (Nos. U21A20246, 32061133010, and 31972527), the National Key Research and Development Program-Key Projects (2021YFD1200900 and 2021YFD1300904), the Second Tibetan Plateau Scientific Expedition and Research Program (STEP) (No. 2019QZKK0501), and Initiation Fund of Sanya Institute of China Agricultural University (SYND-2022-12).

Authors' contributions

FHL and MHL conceived and supervised the study. DFW and FHL conducted the data analysis. DFW, FHL and POW wrote the manuscript, and MHL and POW revised the paper. All authors read and approved the final manuscript.

Acknowledgements

We thank the High-performance Computing Platform of China Agricultural University for providing computing resources.

References

1. Zeder MA. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. Proc Natl Acad Sci USA. 2008;105:11597–604.

582 2. Pedrosa S, Uzun M, Arranz J-J, Gutiérrez-Gil B, San Primitivo F, Bayón Y.
583 Evidence of three maternal lineages in near eastern sheep supporting multiple
584 domestication events. Proc Biol Sci.. 2005;272:2211–7.
585 3. Demirci S, Koban Baştanlar E, Dağtaş ND, Pişkin E, Engin A, Özer F, et al.
586 Mitochondrial DNA diversity of modern, ancient and wild sheep (*Ovis gmelinii*
587 *anatolica*) from Turkey: new insights on the evolutionary history of sheep. PLoS
588 One 2013;8:e81952.
589 4. Her C, Rezaei H, Hughes S, Naderi S, Duffraisse M, Mashkour M, et al. Broad
590 maternal geographic origin of domestic sheep in Anatolia and the Zagros. Anim
591 Genet. 2022;53:452–9.
592 5. Lv F-H, Peng W-F, Yang J, Zhao Y-X, Li W-R, Liu M-J, et al. Mitogenomic meta-
593 analysis identifies two phases of migration in the history of eastern eurasian sheep.
594 Mol Biol Evol. 2015;32:2515–33.
595 6. Deng J, Xie X-L, Wang D-F, Zhao C, Lv F-H, Li X, et al. Paternal origins and
596 migratory episodes of domestic sheep. Curr Biol. 2020;30:4085-4095.e6.
597 7. Kalds P, Zhou S, Gao Y, Cai B, Huang S, Chen Y, et al. Genetics of the phenotypic
598 evolution in sheep: a molecular look at diversity-driving genes. Genet Sel Evol.
599 2022;54:61.
600 8. Cheng H, Zhang Z, Wen J, Lenstra JA, Heller R, Cai Y, et al. Long divergent
601 haplotypes introgressed from wild sheep are associated with distinct morphological
602 and adaptive characteristics in domestic sheep. PLoS Genet. 2023;19:e1010615.
603 9. Yeomans L, Martin L, Richter T. Expansion of the known distribution of Asiatic

mouflon (*Ovis orientalis*) in the late Pleistocene of the southern levant. R Soc open sci. 2017;4:170409.

10. Michel S, Ghoddousi A. *Ovis gmelini:* The IUCN red list of threatened species 2020. 2020. https://dx.doi.org/10.2305/IUCN.UK.2020-2.RLTS.T54940218A22147055.en/ Accessed 23 October 2023.

11. Blyth E. XXVI.— an amended list of the species of the genus *Ovis*. Ann Mag Nat Hist. 1841;7:195–201.

12. Vigne JD, Peters J, Helmer D. The first steps of animal domestication: new archaeozoological approaches. Oxford: Oxbow Books; 2005.

13. Hiendleder S, Kaupe B, Wassmuth R, Janke A. Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. Proc R Soc Lond B. 2002;269:893–904.

14. Li R, Yang P, Li M, Fang W, Yue X, Nanaei HA, et al. A hu sheep genome with the first ovine y chromosome reveal introgression history after sheep domestication. Sci China Life Sci. 2021; 64:1116-30.

15. Lv F-H, Cao Y-H, Liu G-J, Luo L-Y, Lu R, Liu M-J, et al. Whole-genome resequencing of worldwide wild and domestic sheep elucidates genetic diversity, introgression, and agronomically important loci. Mol Biol Evol. 2022;39:msab353.

16. Cao Y-H, Xu S-S, Shen M, Chen Z-H, Gao L, Lv F-H, et al. Historical introgression from wild relatives enhanced climatic adaptation and resistance to

626 pneumonia in sheep. Mol Biol Evol. 2021;38:838–55.
627 17. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina
628 sequence data. Bioinformatics. 2014;30:2114–20.
629 18. Li H, Durbin R. Fast and accurate short read alignment with burrows–wheeler
630 transform. Bioinformatics. 2009;25:1754–60.
631 19. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al.
632 The genome analysis toolkit: a mapreduce framework for analyzing next-
633 generation DNA sequencing data. Genome Res. 2010;20:1297–303.
634 20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence
635 alignment/map format and samtools. Bioinformatics. 2009;25:2078–9.
636 21. Katoh K, Standley DM. Mafft multiple sequence alignment software version 7:
637 improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
638 22. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. Trimal: a tool for automated
639 alignment trimming in large-scale phylogenetic analyses. Bioinformatics.
640 2009;25:1972–3.
641 23. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-
642 generation plink: rising to the challenge of larger and richer datasets. GigaScience.
643 2015;4:7.
644 24. Paradis E, Schliep K. Ape 5.0: an environment for modern phylogenetics and
645 evolutionary analyses in r. Bioinformatics. 2019;35:526–8.
646 25. Yu G. Using ggtree to visualize data on tree-like structures. Curr Protoc
647 Bioinformatics. 2020. 69:e96.

26. Frichot E, Mathieu F, Trouillon T, Bouchard G, François O. Fast and efficient estimation of individual ancestry coefficients. Genetics. 2014;196:973–83.

27. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and vcftools. Bioinformatics. 2011;27:2156–8.

28. Goudet J. hierfstat, A package for r to compute and test hierarchical f-statistics. Mol Ecol Notes. 2005;5:184–6.

29. Paradis E. Pegas: an r package for population genetics with an integrated–modular approach. Bioinformatics. 2010;26:419–20.

30. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. Iq-tree 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.

31. Steenwyk JL, Buida TJ, Li Y, Shen X-X, Rokas A. Clipkit: a multiple sequence alignment trimming software for accurate phylogenomic inference. PLoS Biol. 2020;18:e3001007.

32. Gatto L, Catanzaro D, Milinkovitch MC. Assessing the applicability of the gtr nucleotide substitution model through simulations. Evol Bioinform Online. 2007;2:145–55.

33. Leigh JW, Bryant D. Popart : full-feature software for haplotype network construction. Methods Ecol Evol. 2015;6:1110–6.

34. Ortiz EM. Vcf2phylip v2.0: convert a vcf matrix into several matrix formats for phylogenetic analysis. 2019. https://zenodo.org/record/2540861 Accessed 15 March 2023.

670 35. Terhorst J, Kamm JA, Song YS. Robust and scalable inference of population
671 history from hundreds of unphased whole-genomes. Nat Genet. 2017;49:303–9.
672 36. Beynon SE, Slavov GT, Farré M, Sunduimijid B, Waddams K, Davies B, et al.
673 Population structure and history of the welsh sheep breeds determined by whole
674 genome genotyping. BMC Genet. 2015;16:65.
675 37. Zhao Y-X, Yang J, Lv F-H, Hu X-J, Xie X-L, Zhang M, et al. Genomic
676 reconstruction of the history of native sheep reveals the peopling patterns of
677 nomads and the expansion of early pastoralism in East Asia. Mol Biol Evol.
678 2017;34:2380–95.
679 38. Alberto FJ, Boyer F, Orozco-terWengel P, Streeter I, Servin B, de Villemereuil P,
680 et al. Convergent genomic signatures of domestication in sheep and goats. Nat
681 Commun. 2018;9:813.
682 39. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M,
683 Gavryushkina A, et al. Beast 2.5: an advanced software platform for bayesian
684 evolutionary analysis. PLoS Comput Biol. 2019;15:e1006650.
685 40. Drummond AJ. Bayesian coalescent inference of past population dynamics from
686 molecular sequences. Mol Biol Evol. 2005;22:1185–92.
687 41. Upadhyay M, Kunz E, Sandoval-Castellanos E, Hauser A, Krebs S, Graf A, et al.
688 Whole genome sequencing reveals a complex introgression history and the basis
689 of adaptation to subarctic climate in wild sheep. Mol Ecol. 2021;30:6701–17.
690 42. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient
691 admixture in human history. Genetics. 2012;192:1065–93.

692 43. Malinsky M, Matschiner M, Svardal H. Dsuite - fast d-statistics and related
693 admixture evidence from vcf files. Mol Ecol Resour. 2021;21:584–95.
694 44. Arnholt AT, Evans B. Bsda: basic statistics and data analysis. 2017. https://cran.r-
695 project.org/web/packages/BSDA/index.html. Accessed 28 Sep 2023.
696 45. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, et al. Sequencing of
697 50 human exomes reveals adaptation to high altitude. Science. 2010;329:75–8.
698 46. Hämälä T, Savolainen O. Genomic patterns of local adaptation under gene flow in
699 *Arabidopsis lyrata*. Mol Biol Evol. 2019;36:2557–71.
700 47. Quinlan AR, Hall IM. Bedtools: a flexible suite of utilities for comparing genomic
701 features. Bioinformatics. 2010;26:841–2.
702 48. Hu Z-L, Park CA, Reecy JM. Bringing the animal QTLdb and CorrDB into the
703 future: meeting new challenges and providing updated services. Nucleic Acids
704 Res. 2022;50:D956–61.
705 49. Dong S-S, He W-M, Ji J-J, Zhang C, Guo Y, Yang T-L. Ldblockshow: a fast and
706 convenient tool for visualizing linkage disequilibrium and haplotype blocks based
707 on variant call format files. Brief Bioinform. 2021;22:bbaa227.
708 50. Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L. Poplddecay: a fast and
709 effective tool for linkage disequilibrium decay analysis based on variant call
710 format files. Bioinformatics. 2019;35:1786–8.
711 51. Korunes KL, Samuk K. Pixy : unbiased estimation of nucleotide diversity and
712 divergence in the presence of missing data. Mol Ecol Resour. 2021;21:1359–68.
713 52. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in

714 multidimensional genomic data. Bioinformatics. 2016;32:2847–9.
715 53. Sanna D, Barbato M, Hadjisterkotis E, Cossu P, Decandia L, Trova S, et al. The
716 first mitogenome of the Cyprus mouflon (*Ovis gmelini ophion*): new insights into
717 the phylogeny of the genus *Ovis*. PLoS One. 2015;10:e0144257.
718 54. Teng H, Zhang Y, Shi C, Mao F, Cai W, Lu L, et al. Population genomics reveals
719 speciation and introgression between brown Norway rats and their sibling species.
720 Mol Biol Evol. 2017;34:2214–28.
721 55. Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing indian
722 population history. Nature. 2009;461:489–94.
723 56. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft
724 sequence of the neandertal genome. Science. 2010;328:710–22.
725 57. Jakaria J, Zein MSA, Sulandari S, Subandriyo S, Muladno M. The use of
726 microsatellite markers to study genetic diversity in Indonesian sheep. J Indones
727 Trop Anim Agric. 2012;37:1–9.
728 58. Sharma RC, Arora AL, Narula HK, Singh RN. Characteristics of Garole sheep in
729 India. Anim Genet Resour Inf. 1999;26:57-64
730 59. Banerjee R, KumarManda P, Kumar Pal U, Ray K. Productivity and genetic
731 potential of Garole sheep of India-a review. Asian J Anim Sci. 2010;4:170–89.
732 60. Bleyhl B, Arakelyan M, Askerov E, Bluhm H, Gavashelishvili A, Ghasabian M, et
733 al. Assessing niche overlap between domestic and threatened wild sheep to
734 identify conservation priority areas. Divers Distrib. 2019;25:129–41.
735 61. Yang L, Wei F, Zhan X, Fan H, Zhao P, Huang G, et al. Evolutionary

736 conservation genomics reveals recent speciation and local adaptation in threatened
737 takins. Mol Biol Evol. 2022;39:msac111.
738 62. Hu Y, Thapa A, Fan H, Ma T, Wu Q, Ma S, et al. Genomic evidence for two
739 phylogenetic species and long-term population bottlenecks in red pandas. Sci Adv.
740 2020;6:eaax5751.
741 63. Ai H, Zhang M, Yang B, Goldberg A, Li W, Ma J, et al. Human-mediated
742 admixture and selection shape the diversity on the modern swine (*Sus scrofa*) y
743 chromosomes. Mol Biol Evol. 2021;38:5051–65.
744 64. Luikart G, Gielly L, Excoffier L, Vigne J-D, Bouvet J, Taberlet P. Multiple
745 maternal origins and weak phylogeographic structure in domestic goats. Proc Natl
746 Acad Sci USA. 2001;98:5927–32.
747 65. Naderi S, Rezaei H-R, Pompanon F, Blum MGB, Negrini R, Naghash H-R, et al.
748 The goat domestication process inferred from large-scale mitochondrial DNA
749 analysis of wild and domestic individuals. Proc Natl Acad Sci USA.
750 2008;105:17659–64.
751 66. Yang L, Zhang X, Hu Y, Zhu P, Li H, Peng Z, et al. Ancient mitochondrial
752 genome depicts sheep maternal dispersal and migration in eastern Asia. J Gene
753 Genomics. 2023; https://doi.org/10.1016/j.jgg.2023.06.002 in press.
754 67. Wu D-D, Yang C-P, Wang M-S, Dong K-Z, Yan D-W, Hao Z-Q, et al.
755 Convergent genomic signatures of high-altitude adaptation among domestic
756 mammals. Natl Sci Rev. 2020;7:952–63.
757 68. Porter V. Mason's world dictionary of livestock breeds, types and varieties, 6th

edition. Wallingford: CABI Publishing; 2020.

69. Putra WPB, Alwi Y, Suryani H. Morphometrics characterization of thin-tail sheep in lowland and highland areas. Trop Anim Sci J. 2021;44:386–98.

70. Serranito B, Cavalazzi M, Vidal P, Taurisson-Mouret D, Ciani E, Bal M, et al. Local adaptations of mediterranean sheep and goats through an integrative approach. Sci Rep 2021;11:21363.

71. Kim E-S, Elbeltagy AR, Aboul-Naga AM, Rischkowsky B, Sayre B, Mwacharo JM, et al. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. Heredity (Edinb). 2016;116:255–64.

72. Mastrangelo S, Moioli B, Ahbara A, Latairish S, Portolano B, Pilla F, et al. Genome-wide scan of fat-tail sheep identifies signals of selection for fat deposition and adaptation. Anim Prod Sci. 2018;59:835–48.

73. Xu Y-X, Wang B, Jing J-N, Ma R, Luo Y-H, Li X, et al. Whole-body adipose tissue multi-omic analyses in sheep reveal molecular mechanisms underlying local adaptation to extreme environments. Commun Biol. 2023;6:159.

38

773 **Figures**

774 **Figure 1 Population genetics structure of Asiatic mouflons in Iran (*O. gmelini*). a**
775 Geographic distribution of 32 Asiatic mouflons. Purple and blue dots indicate the location of *O.*
776 *gmelini*_1 and *O. gmelini_2* subpopulations; **b** Principal components 1 and 2 for the 32 Asiatic
777 mouflon in Iran; **c** Neighbor-joining (NJ) trees of the 32 Asiatic mouflons in Iran based on
778 whole-genome SNPs using the nine urial samples as the outgroup; **d** Population genetic structure
779 of the Asiatic mouflons in Iran and domestic sheep inferred from the program the sNMF v1.2 (*k*
780 = 6) using whole-genome SNPs.

781

782 **Figure 2 Phylogeny of Asiatic mouflon, urial, and domestic sheep inferred from**
783 **mitogenomes and Y-chromosomal genetic variants. a** Geographical distribution of the
784 samples in this and previous studies; **b** Phylogeny of Asiatic mouflon, urial, and domestic sheep
785 inferred from 767 mitogenomes; **c** Phylogeny of Asiatic mouflon, urial, and domestic sheep
786 inferred from 239 Y-chromosomal SNPs; **d** Network of mitogenome haplotypes; **e** Network of
787 Y-chromosome haplotypes. Bootstrapping values above 50% were labeled on the phylogenetic
788 tree. Bighorn sheep were used as the outgroup in phylogenetic analyses based on mitogenome
789 and Y-chromosome genetic variants.

790

791 **Figure 3 Demographic history of Asiatic mouflon in Iran and domestic sheep. a** Effective
792 population sizes (N_e) across six domestic sheep populations and the two subpopulations of
793 Asiatic mouflon in Iran inferred from SMC++; **b** Inference of split times between *O. gmelini*_1
794 and the other wild and domestic sheep populations. The vertical lines represent the split time

795 between each pair of populations; **c** Bayesian skyline plot (BSP) reconstruction of historical
796 change in effective population size (N_e) of domestic sheep of maternal haplogroups HA+B+D,
797 and HC+E, and the Asiatic mouflons in Iran; **d** Bayesian skyline plot (BSP) reconstruction of
798 historical changes in $N_e$of paternal lineages HY1 and HY2 of domestic sheep and the Asiatic
799 mouflon in Iran. The middle line represents median estimates of $N_e\tau$, and the shaded area
800 represents 95% highest posterior density (HPD) intervals.
801
802 **Figure 4 Geographic distribution of Asiatic mouflon and the phylogenetic relationship**
803 **among Asiatic mouflon subspecies and domestic sheep**. **a** Geographical distribution of Asiatic
804 mouflon and samples information of 32 Asiatic mouflon in Iran and 301domestic sheep samples;
805 **b** Phylogenetic tree inferred based on complete *Cyt-b* sequences using the maximum likelihood
806 algorithms. Bootstrap values above 50% were labeled on the phylogenetic tree. **c** Network of
807 Asiatic mouflon in Iran and Turkey and domestic sheep lineage of C and E built from complete
808 *Cyt-b* sequences.
809
810 **Figure 5 Genetic introgression of Asiatic mouflon in SSA domestic populations. a**
811 Admixture proportions of Asiatic mouflon in Iran and SSA domestic populations inferred from
812 the program the sNMF v1.2 (K = 3) using whole-genome SNPs. **b** Heatmap of D-statistics and f_4-
813 ratio calculated by Dsuite v0.5 applied to data including *O. gmelini*_1, *O. gmelini*_2, and SSA
814 domestic populations; **c**, Venn diagram for the shared introgression regions among four domestic
815 South and Southeast populations; **d** Manhattan plot showing signals of selection in the four
816 introgressed SSA populations (BGE/GAR/GUR/SUM) detected by the PBS with the dashed line
817 marking 99.9% threshold. The introgression region, QTL, and adjacent gene *PCDH9* were

818 labeled with blue, purple, and green bars, respectively. The sheep QTLs overlapped with the top
819 1% PBS regions are shown, **e** PBS and f_{dm} values around the significant introgressed genomic
820 region on Chr.10.
821
822 **Figure 6 Candidate gene *PCDH9* associated with the significant introgression signal on**
823 **Chr. 10**. **a** Linkage disequilibrium (LD) blocks in the significant introgression region and gene
824 *PCDH9* (chr10:41,104,271-43,073,231). The red and yellow bars represent the location of the
825 significant introgression region and gene *PCDH9*; **b** Pattern of LD decay in the introgressed
826 populations. The black line is calculated from the whole 10 chromosomes, the red line is
827 calculated from the beginning of the significant region to the next 800 kb (chr10:41,204,271-
828 42,004,271); **c** D_{xy} and F_{ST} around the significant introgression region on Chr. 10
829 (chr10:41,104,271-41,313,804), calculated by pixy v1.2.7.beta1; **d** Haplotype pattern of the
830 significant introgression region on Chr. 10 (chr10:41,204,271-41,213,804) (left) and exons of
831 *PCDH9* (right).
832

833 # Additional files

834 **Additional file 1 Table S1**

835 Format: excel

836 Title: 780 samples information for autochromosome analysis.

837 **Additional file 1 Table S2**

838 Format: excel

839 Title: 770 samples information for mitochondria chromosome analysis.

840 **Additional file 1 Table S3**

841 Format: excel

842 Title: 337 samples information of *Cyt-b* sequence in this study.

843 **Additional file 1 Table S4**

844 Format: excel

845 Title: 239 samples for Y chromosome analysis in this study.

846 **Additional file 1 Table S5**

847 Format: excel

848 Title: Kinship of 32 *O. gmelini* samples.

849 **Additional file 1 Table S6**

850 Format: excel

851 Title: The *F*ST of autosome, mitochondria and Y chromosome of *O. gmelini*_1, *O. gmelini*_2

852 and *O. aries.*

853 **Additional file 1 Table S7**

854 Format: excel

855 Title: The nucleotide diversity / haplotype diversity of autosome, mitochondria and Y

856 chromosome of *O. gmelini*_1, *O. gmelini*_2 and *O. aries.*

857 **Additional file 1 Table S8**

858 Format: excel

859 Title: Haplotype information of mitogenomes and Y chromosomes.

860 **Additional file 1 Table S9**

861 Format: excel

862 Title: Split times estimated by SMC++.

863 **Additional file 1 Table S10**

864 Format: excel

865 Title: Haplotype information of *Cyt-b* sequences.

866 **Additional file 1 Table S11**

867 Format: excel

868 Title: *D*-statistics from Iranian mouflon to SSA populations.

869 **Additional file 1 Table S12**

870 Format: excel

871 Title: Details for chromosome-level introgression across four breeds.

872 **Additional file 1 Table S13**

873 Format: excel

874 Title: Common introgressed blocks from *O.gmelini*_2 into four SSA populations.

875 **Additional file 1 Table S14**

876 Format: excel

877 Title: Population branch statistic (PBS) of putative selective regions in the SSA population.

878 **Additional file 1 Table S15**

879 Format: excel

880 Title: Overlaps regions among selected regions identified by PBS, introgression regions and
881 QTLs regions.

882

883 **Additional file 2 Figure S1**

884 Format: word

885 Title: 780 autosome NJ-tree showing the same two cluster of *O. gmelini* as Fig.1b and domestic
886 sheep samples clustered based on their geographic distribution.

887 **Additional file 2 Figure S2**

888 Format: word

889 Title: Admixture of 770 samples for *K*=2-9.

890 **Additional file 2 Figure S3**

891 Format: word

892 Title: Admixture of SSA 18 breeds with two *O. gmelini* populations for *K*= 2-9.

Figure 1

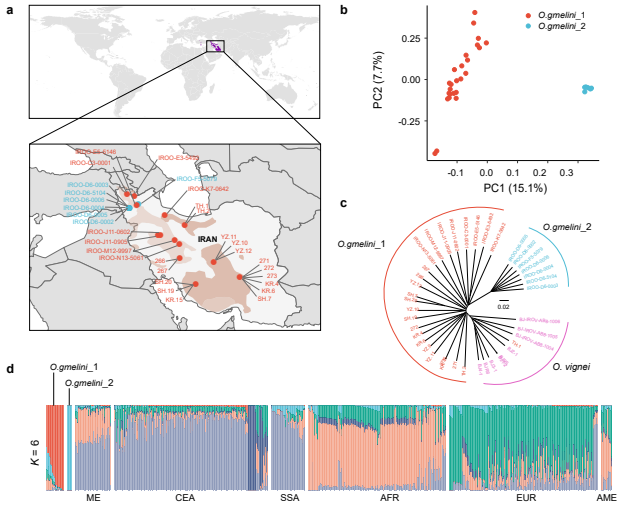
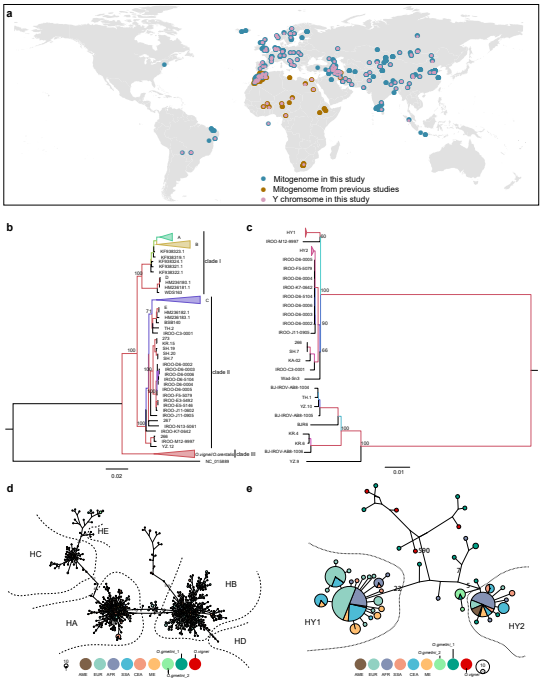


Figure 2



46

Figure 3

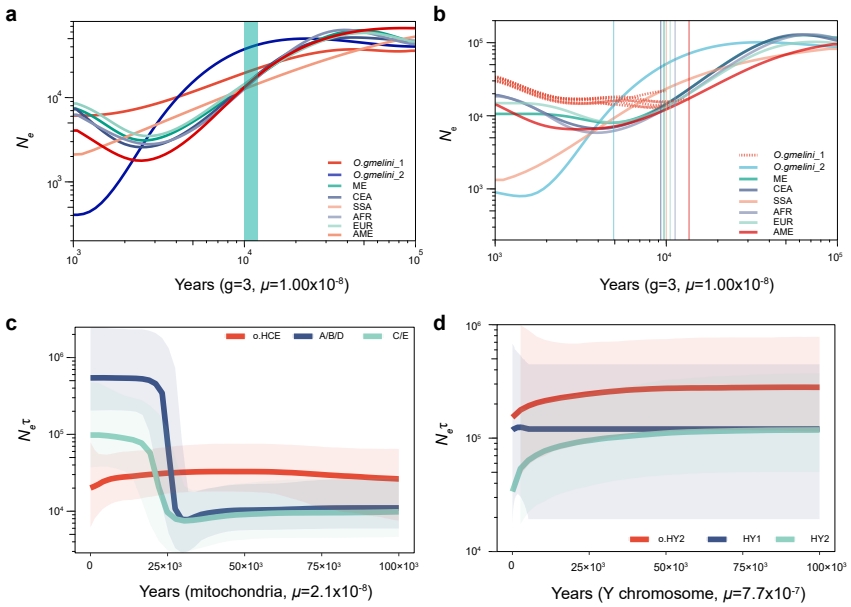
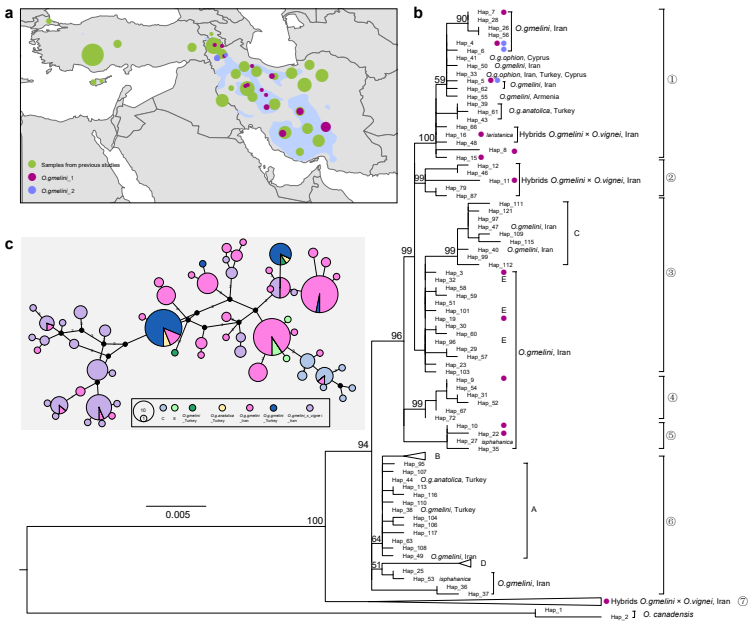


Figure 4



47

Figure 5

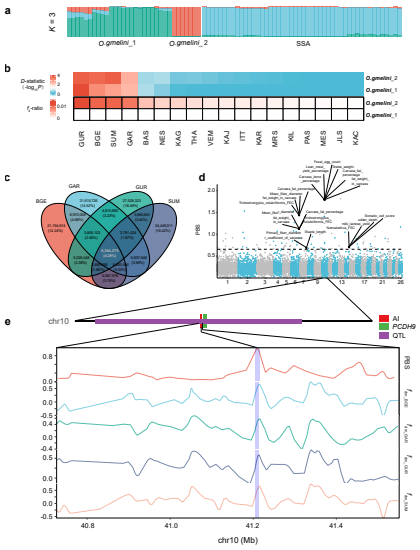
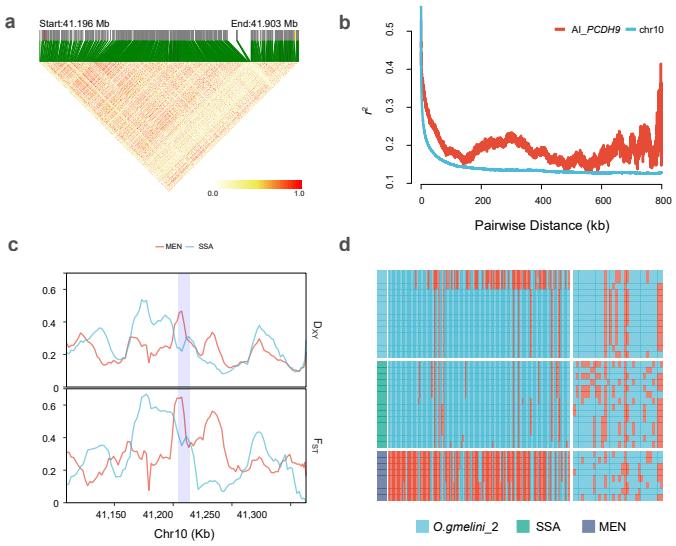


Figure 6



48