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

*lvfenghua@cau.edu.cn, menghua.li@cau.edu.cn

Abstract

The wild ancestor of domestic sheep, Asiatic mouflon (*O. gmelini*), consists of several subspecies mainly distributed in Iran nowadays. However, the impact of the Asiatic mouflon subspecies in Iran on sheep domestication remains unclear. 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. 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.

Key words: Iranian Asiatic mouflon, sheep, whole-genome sequences, uniparental lineage, domestication, introgression

Introduction

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[99,10], 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,

69 Demirci et al. (2013) suggested that domestic sheep could have originated from two
70 distinct ancestral populations of mtDNA lineages A+B+D and C+E of domestic
71 sheep, respectively[3]. Based on genetic, archeozoological, and geographic evidence of
72 212 Asiatic mouflon samples covering most of their distribution area, a recent study
73 suggested that the maternal origin of sheep could be traced to a subspecies of *O.*
74 *gmelini gmelini* in eastern Anatolia and central Zagros[4].
75
76 Despite the intensive focus on material inheritance, Y-chromosome and whole-
77 genome analyses have received less attention. Recently, Y-chromosome and whole-
78 genome analyses of Asiatic mouflon in Iran revealed that they could have diverged
79 from the direct ancestor of domestic sheep, and showed genetic introgression into
80 domestic sheep after domestication[8,14,15]. Also, these studies have demonstrated that
81 genomes of Asiatic mouflon in Iran are crucial for the development of morphological
82 and adaptive characteristics in domestic sheep[8,15,16]. Nevertheless, the impact and
83 extent of genetic introgression involving the Asiatic mouflon subspecies in Iran (e.g.,
84 *O. gmelini gmelini*, *O. gmelini isphahanica*, and *O. gmelini laristanica*) and their
85 hybrid populations (*O. gmelini gmelini* × *O. vignei*) on the genomes of domestic
86 sheep remains unclear.
87
88 Here, we comprehensively analyzed whole genome sequence, complete and partial
89 mtDNA sequences, and Y-chromosome genetic variants in domestic sheep and

90 Asiatic mouflon, covering a large distribution area in Asia, Europe, Africa, and
91 America. We aimed to elucidate the contributions of various Asiatic mouflon
92 subspecies in Iran to the domestication of sheep. Additionally, we investigated the
93 impact of wild introgression of the Asiatic mouflon subspecies in Iran on the genomes
94 of domestic sheep populations in South and Southeast Asia, which have adapted to the
95 hot and humid environments near the Indian Ocean.
96

97 ## Materials and Methods

98 ***Datasets***

99 Whole-genome sequences (WGS) of 780 individuals representing 158 domestic sheep
100 populations and three wild sheep species (32 Asiatic mouflon (*Ovis gmelini*) in Iran, 9
101 urial (*Ovis vignei*), and 1 bighorn sheep (*Ovis canadensis*)) were obtained from
102 publicly available data (Supplementary Table S1). Besides these, we retrieved 161
103 whole mitogenomes from the NCBI (Supplementary Table S2). Also, we included
104 333 Cytochrome b (*Cytb*) sequences of Asiatic mouflon in the integrated analysis
105 (Supplementary Table S3).
106

107 ***Variant calling***

108 We implemented quality-based trimming for WGS data with Trimmomatic v0.39 [17]

109 and mapped high-quality reads to the reference genome

110 *Oar_rambouillet_v1.0_CM022046.1*. (https://www.

111 ncbi.nlm.nih.gov/assembly/GCF_002742125.1/;

112 https://www.ncbi.nlm.nih.gov/nuccore/CM022046.1/, last accessed May 12, 2020)

113 using the Burrows–Wheeler aligner (BWA mem) v.0.7.8[18] with the default

114 parameters. Then, we removed duplicates with the *MarkDuplicates* module in GATK

115 v 4.1.2.0[19] and called SNPs following the previous procedures[15]. We used the

116 *HaplotypeCaller* module in GATK v 4.1.2.0 to generate the GVCF file of each

117 sample and merged the raw GVCF file of each individual with the *CombineGVCFs*

118 module. Afterward, we genotyped SNPs using the *GenotypeGVCFs* module and

119 filtered SNPs using the *VariantFiltering* module of GATK with the parameters

120 "QUAL < 30.0 || QD < 2.0 || MQ < 40.0 || FS > 60.0 || SOR > 3.0 || MQRankSum < -

121 12.5 || ReadPosRankSum < -8.0". We obtained a total of 121,252,917 high-quality

122 autosomal SNPs and 203,938 Y-chromosomal SNPs for the following analyses.

123

124 ***Mitochondrial DNA assembly***

125 We assessed the mitogenome depth of each sample using the SAMtools v.1.9[20] and

126 kept only the samples with depth > 100 in the below analyses. We assembled the

127 complete mitogenomes using the following protocols. We first converted BAM files

128 to the fastq format. Then, we generated the mitogenome consensus sequence from the

129 fastq files using MIA v1.0 (https://github.com/mpieva/mapping-iterative-assembler).

130

131 In addition to the 605 mitogenomes assembled here, we also included 161 whole

132 mitogenome sequences retrieved from the NCBI GenBank in the analysis, with a final

133 mitogenome dataset of 767 sequences from 727 domestic sheep, 31 Iranian Asiatic

134 mouflon, and 9 urial. All sequences were aligned using MAFFT v7.515[21] and

135 trimmed with trimAl v1.4[22].

136

137 ***Y-chromosomal SNPs***

138 To obtain single-copy Y-chromosomal SNPs, raw SNPs in the 780 samples were

139 processed as follows. Firstly, we identified ram samples from the whole samples

140 using PLINK v1.90b6.26[23] with the "--check-sex" option and kept only these ram

141 samples in the following analyses. We then excluded SNPs in the regions covering the

142 pseudoautosomal regions (PARs) and the highly repetitive sequences (around 1.06

143 Mb)[14]. Afterwards, we filtered SNPs with the mean sequencing depth ranging from

144 0.5 × to 2 × of the expected depth (half of the genome-wide depth)[14]. Further, we

145 removed SNPs that were heterozygous in at least one of the male samples. Finally, we

146 retained 4,096 Y-chromosome SNPs for the haplotype and phylogenetic analyses.

147

148 ***Population structure analysis***

149 To reveal the genetic differentiation within the Asiatic mouflon in Iran, we examined
150 the population genetic substructure for the total 779 wild and domestic samples using
151 a set of autosomal SNPs filtered as follows. First, we filtered the total 25,270,188
152 high-quality SNPs with missing rates < 0.1 and minor allele frequencies > 0.05
153 (MAF > 0.05) using PLINK v1.90b6.26[23]. Then, we pruned the remaining SNPs with
154 the options "--indep-pairwise 50 10 0.1" of PLINK v1.90b6.26[23]. Afterwards,
155 1,566,300 SNPs were retained for the population genetic structure analysis.

156

157 We computed the identity by state among samples using PLINK v1.90b6.26[23] with
158 option "distance square 1-ibs". Next, we generated a neighbor-joining (NJ) tree using
159 the function "bionj" in the R package *ape* v5.6.2[24]. The tree was rooted using urial (*O.*
160 *vignei*) as the outgroup, and its graphical representation was obtained using the R
161 package *ggtree* v3.6.2[25]. Principal component analyses (PCA) were performed using
162 the "--pca" option in PLINK v1.90b6.26[23] with the default settings. Furthermore,
163 model-based clustering was performed using the sNMF v1.2 with clusters *K* from 2 to
164 10[26].

165

166 The F_{ST} and nucleotide diversity (π) of autosomes were calculated with vcftools
167 v0.1.17[27] using 2Mbp windows and 1Mbp steps, and *P* value was calculated with

168 1000 permutation test. The F_{ST} of mitochondria and Y chromosome were calculated

169 with R package hierfstat v0.5.11[28] with 1000 permutation test. The haplotype

170 diversity of mitochondria and Y chromosome were calculated with pegas v1.1[29] with

171 1000 permutation test.

172

173 ***Phylogenetic analysis***

174 Phylogenetic analyses based on mitogenome and Y-chromosomal SNPs were

175 performed using the maximum likelihood (ML) algorithm of iqtree2 v2.0.7[30].

176 Mitogenome sequences were aligned using mafft v7.515[21] with the default setting,

177 and multiple sequence alignment (MSA) was trimmed to remove ambiguously aligned

178 regions using trimAl v1.4[22] with parameters "-automated1". To keep high-quality

179 SNPs, we further removed all the sites with gaps and ambiguous nucleotide bases and

180 only kept parsimony informative sites using *ClipKIT*[31]. We tested the potential models

181 of DNA sequence evolution using the option "-m" test in iqtree2 v2.0.7[30] and found

182 the generalized time reversible model (GTR + G)[32] was the best fitting substitution

183 model for mitogenome. We constructed a maximum likelihood (ML) tree with the site

184 model GTR + G with 1000 bootstrap replicates using iqtree2 v2.0.7[30]. The resulting

185 tree was rooted with bighorn as the outgroup and visualized using the FigTree v1.4.4

186 (http://tree.bio.ed.ac.uk/software/figtree/). Mitogenome haplotypes were inferred with

187 the function "haplotype" in the R package *pegas* v1.1[29]. We also constructed the

188 parsimony mitogenome haplotype network using POPART[33]. We transformed the
189 genotypes of Y-chromosomal 4,096 SNPs into the fasta format using vcf2phylip.py[34].
190 Phylogenetic trees were built and visualized following the same pipelines and using
191 the same programs and substitution model as above.
192

193 ***Demographics and divergence time inference***

194 The population genetic structure analyses above identified two subpopulations (*O.*
195 *gmelini*_1 and *O. gmelini*_2) in the Asiatic mouflons in Iran. We inferred the dynamic
196 changes in historical and recent effective population size (N_e) of the two
197 subpopulations using SMC++ v1.15.2 algorithms[35]. We selected 7 - 10 high-depth
198 individuals in each of the six geographically differentiated genetic groups of domestic
199 sheep (Central-and-East Asian, South-and-Southeast Asian, the Middle Eastern,
200 African, European, and American populations)[15] and the two Asiatic mouflon
201 subpopulations. We estimated the N_e during the past 1,000 to 100,000 years for each
202 group using SMC++ v1.15.2[35] under the default setting.
203

204 For the SMC++ analysis, the reference genome's low-complexity regions were
205 masked using snpable (http://lh3lh3.users.sourceforge.net/snpable.shtml, last accessed
206 October 20, 2022) using in-house scripts. We set the mutation rate and generation
207 time to be 1.00×10^{-8} per nucleotide site per generation[36] and 3 years[37,38], respectively.

208 To infer the divergence times between *O. gmelini*_1 and *O. gmelini*_2 as well as
209 between the Asian mouflon and domestic sheep, we employed the split option of
210 SMC++ with the same parameter settings as above.
211
212 Additionally, we inferred the demographic history of domestic sheep and Asiatic
213 mouflon based on mitogenomes and Y-chromosome data using Beast v2.7.3[39] and the
214 Coalescent Bayesian Skyline plot[40]. Firstly, we used a coalescent model to estimate
215 the mutation rates for mitogenome and Y chromosome chromosomes. The analysis
216 was performed using the GTR + GAMMA (4) + I site model and the strict clock
217 model. The most recent common ancestor (MRCA) of Asiatic mouflon and domestic
218 sheep was set to be 0.653 million years[41]. The chain length was set to be 40,000,000
219 with a burn-in of 2,000,000. The results were analyzed using Tracer
220 (http://beast.community/tracer), assuming mutation rates estimates of 2.10×10^{-8} and
221 7.70×10^{-9} substitutions per site per year for the mitochondrial and Y chromosome
222 SNPs, respectively.
223
224 Subsequently, we utilized the mutation rates estimated above to infer the demographic
225 history of maternal and paternal lineages in domestic sheep and Asiatic mouflon. We
226 used the same GTR + GAMMA (4) + I site model as above, with the clock model set
227 as strict and the mutation rates of 2.10×10^{-8} and 7.70×10^{-9} per site per year for
228 mitochondria and Y-chromosome, respectively. The prior model remained the

229 Coalescent Bayesian Skyline, with the chain length of 40,000,000, a burn-in of
230 2,000,000, and samples drawn every 4,000 steps. The obtained results were further
231 analyzed using Tracer, and each analysis yielded over 200 effective samples for all
232 relevant parameters.
233

234 ***Introgression analysis***

235 We conducted introgression analysis to clarify the genetic contributions of the two
236 Asiatic mouflon subpopulations to South-and-Southeast Asian (SSA) sheep. We
237 tested introgression using a combination method of *D* and f_{dM} statistics. Specifically,
238 we calculated *D* statistics[42] with a four taxon model (((P1, P2), P3), P4) using Dsuite
239 v0.5r50[43]. In this analysis, Menz sheep (MEN) was selected as the reference
240 population (P1), while SSA populations were the target population (P2), *O. gemlini*_1
241 or *O. gemlini*_2 was the donor (P3), and *O. canadensis* was the outgroup (P4). A *P*
242 value < 0.05 from a standard block-jackknife procedure was considered as evidence
243 of gene flow from P3 to P2.
244

245 We further implemented sliding window statistics analysis using the Dsuite v0.5r50[41]
246 to quantify the introgressions. We calculated f_{dM} values using a sliding window of 50
247 SNPs with a step of 25 SNPs across the genomes with the defined trios (MEN, SSA,
248 *O. gmelini*_2), where SSA represents the four South-and-Southeast Asian sheep

249 populations suggested to be the introgressed populations by the *D*-statistics and f_4-

250 ratio tests. We calculated the Z-transformed f_{dm} values, and the windows with $P <$

251 0.05 and $D > 0$ was defined as the potential introgressed genomic regions. We

252 evaluated the statistical significance using one-tailed *Z*-test in R package BSDA

253 v1.2.1[44].

254

255 ***Selective sweep tests***

256 To assess the potential roles of genomic introgression from Iranian Asiatic mouflons

257 in Southern and Southeastern Asian (SSA) sheep populations, we conducted a PBS

258 analysis[45] to identify selective signatures in four introgressed populations (i.e., Garut

259 (GUR), Garole (GAR), Bangladeshi (BGE), and Sumatra (SUM)). We pooled these

260 four populations (GUR, BGE, SUM, and GAR) as an introgressed group, and used the

261 MEN and *O. gmelini*_2 as the reference populations. We calculated the PBS values

262 using PBScan v2020.03.16 with parameters "-win 50 -step 25" described in Hämälä

263 and Savolainen (2019) [46] and estimated *P* values by 2000 permutation cycles for

264 Monte Carlo testing. and detected selective regions with top 1% PBS values.

265 Subsequently, we identified the overlaps between the selective and the introgressed

266 regions using BEDtools v2.30.0[47] and annotated functional genes in the overlapping

267 genomic regions.

268

269 ***Gene annotation and function analyses***

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

279 **Results**

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

281 A total of 121,252,917 SNPs in the 780 wild and domestic sheep samples were
282 obtained for downstream analyses. We obtained an average depth of 923.33 × (101.92
283 × - 8,324.72 ×) for the mitogenome assemblies after aligning all mitochondrial reads
284 from the 606 samples to the reference mitochondrial genome NC_001941.1
285 (Supplementary Table S2). After quality control for the SNP identification on the Y-
286 chromosome, we identified 4,096 Y-specific SNPs in a cohort of 239 rams with an
287 average read depth of 8.26 × (~4.25 × - 16.99 ×) for SNP (Supplementary Table S4).

288 ***Population genetic structure of the Asiatic mouflons in Iran***

289 We observed two subpopulations in the 32 Asiatic mouflons in Iran (Figure 1 B/C/D)
290 with autosome SNPs. The larger subpopulation (*O. gmelini*_1) comprised 25 samples
291 across eastern and southern Iran, while the smaller subpopulation (*O. gmelini*_2)
292 consisted of only 7 samples limited in northwestern Iran (Fig. 1A; Supplementary
293 Table S2). Low kinship values within two subpopulations eliminate inbreeding caused
294 by sampling (Supplementary Table S5). We also observed significant genetic
295 divergence and different genetic diversity patterns (Supplementary Table S6-S7). The
296 result makes sure that the splition is not the artefact of inbreeding.

297

298 Phylogenetic analyses of the total 780 wild and domestic sheep samples identified the
299 same two subpopulations of Asiatic mouflon (*O. gmelini*_1 and *O. gmelini*_2) and six
300 genetic clusters of domestic sheep, which were congruent with their geographic
301 distribution (Supplementary Figure 1). Admixture analysis showed inter- and intra-
302 species genetic divergence similar to the phylogenetic tree (Fig. 1D).

303

304 ***Phylogenetic and haplotype analysis based on uniparental variants***

305 We observed 593 mitochondrial haplotypes from the 767 mitogenomes of wild and
306 domestic sheep (Supplementary Table S2). The phylogenetic tree inferred from the
307 mitochondrial haplotypes revealed three monophyletic clades (clades I, II, and III)

308 (Figure 2B). Clade I comprised of haplogroups A, B, and D of domestic sheep. Clade
309 II consisted of the domestic sheep's haplogroups C and E and the Asiatic mouflon
310 samples, while the remaining Asiatic mouflon samples clustered with the urial in
311 clade III (Figure 2B).
312
313 For Y-chromosomal SNPs, the lineage HY2 of domestic sheep was closely related to
314 Asiatic mouflon in the haplogroups C and E but diverged from the main lineage HY1
315 (Figure 2C, Supplementary Table S8). However, phylogenetic and network analysis
316 revealed no shared mitochondrial haplotypes and only one shared Y-haplotypes of
317 HY2 between domestic sheep and Asiatic mouflon in Iran (Figure 2D, E). We
318 observed significant genetic divergence (*O. gmelini*_1: $F_{ST} = 0.11$, $P < 0.05$; *O.*
319 *gmelini*_2: $F_{ST} = 0.21$, $P < 0.05$) between Asiatic mouflons and domestic sheep in
320 whole genome levels (Supplementary Table S6). The observation implied that the
321 Asiatic mouflon in Iran might not be the direct ancestor of domestic sheep. Instead,
322 the relatively close phylogenetic relationships between the Asiatic mouflons in Iran
323 and domestic sheep suggested by haplogroups C, E, and Y-chromosomal
324 haplogroupY2 implied their possible gene flows.

325 ***Effective population size and divergence time***

326 Inference of demographic history showed the Asiatic mouflons and domestic sheep
327 shared a similar historical pattern 10,000 years ago (Figure 3A), confirming the

328 Asiatic mouflon as the wild ancestor of sheep[53]. SMC++ analysis revealed a continual
329 decline of N_e in domestic sheep and Asiatic mouflon during ~8,000–10,000 years ago.
330 Nevertheless, the two Asiatic mouflon populations in Iran showed slower declines
331 than domestic sheep (Figure 3A), which probably reflected the bottleneck effect of
332 domestication. A sharp decline of N_e was detected ~1,000-5,000 years ago in the *O.*
333 *gmelini*_2, demonstrating a pattern different from that for the *O. gmelini*_1 (Figure
334 3A). We observed similar estimates of divergence time between *O. gmelini*_1 and the
335 six geographically differentiated genetic groups of domestic sheep during ~9,000-
336 15,000 years ago (Figure 3B, Supplementary Table 9), which was congruent with the
337 domestication time of sheep[2]. Additionally, we found the splitting time between the
338 two Asiatic mouflon populations in Iran around 5,000 years ago (Figure 3B,
339 Supplementary Table 9), indicating a recent genetic divergence of Asiatic mouflons in
340 Iran (Figure 1). For the divergence pattern inferred by WGS of the Asian mouflons,
341 the subpopulation *O. gmelini*_2 clustered with partial individuals of the subpopulation
342 *O. gmelini*_1 in the mitogenome and Y-chromosomal phylogenetic trees (Figure 1, 3).
343 The discordant phylogenetic patterns of Asiatic mouflons in Iran inferred by
344 uniparental and autosomal genetic variants could be explained by the recent
345 hybridization events[54].
346
347 We conducted Bayesian skyline analyses to estimate N_e based on mitogenomes and
348 Y-chromosomes. Based on the estimated TMRCA (time to the most recent common

349 ancestor) of Asiatic mouflons and domestic sheep[41], three lineages (lineage o.HCE:
350 haplogroup of Asiatic mouflons in Iran and partial domestic sheep of haplogroups C
351 and E, lineage A+B+D: haplogroups A, B, and D, and lineage C+E: haplogroups C
352 and E) of mitogenome showed different patterns. Lineage A+B+D showed a steep
353 increase in N_e at approximately ~25 k years ago. For lineage C+E, it also increased at
354 approximately ~25 k years ago, while lineage o.HCE demonstrated relatively fewer
355 changes since approximately ~100 k years ago (Figure 3C). Also, we observed
356 different patterns in the three Y-chromosomal lineages (lineage o.HY2: haplogroups
357 of Asiatic mouflons in Iran and partial domestic sheep of HY2, lineage HY1, and
358 lineage HY2). Lineages o.HY2 and HY2 displayed slow decline in N_e at
359 approximately ~25 k years ago, while lineage HY1 changed sightly since
360 approximately ~100 k years ago (Figure 3D). Thus, uniparental genetic markers
361 showed similar demographic patterns between o.HY2 and HY2 as well as between
362 HCE and C/E, but different patterns between the dominant linages A+B+D and HY1
363 (Figure 2, 3). The results suggested that the Asiatic mouflons in Iran might not be the
364 direct ancestor of domestic sheep but be involved in the genetic introgression.
365

366 ***Phylogeography of Asiatic mouflon***

367 We built a *Cyt-b* dataset consisting of 247 Asiatic mouflons and 8 European
368 mouflons. The samples of Asiatic mouflon covered the current distribution of all the

369 subspecies (Supplementary Table S3, Figure 4A). To clarify phylogenetic
370 relationships between different Asiatic mouflon subspecies and domestic sheep, we
371 performed phylogenetic analysis of *Cyt-b* sequences of Asiatic mouflons and
372 domestic sheep using argali as the outgroup. In the phylogenetic tree built from *Cyt-b*,
373 Asiatic mouflon was assigned to 7 major clades supported by high bootstrap values
374 (94~100) (Figure 4B). The topology confirmed the five mitogenome clades of
375 domestic sheep. Haplogroups C+E clustered into monophyletic clade, which diverged
376 from the haplogroups A+B+D (Figure 4B). It was noteworthy that most of the hybrid
377 samples inferred above were integrated with *O. gmelini* or *O. vignei* samples, and
378 only around 10% of samples clustered with subspecies such as *O.g.ophion* and
379 *O.g.anatolica* (Figure 4B, Supplementary Table S10). Clade 3 comprised haplogroups
380 C+E of domestic sheep and the Asiatic mouflons in Iran with high frequencies of
381 shared haplotypes between them (Figure 4C, Supplementary Table S10). Haplogroups
382 A+B+D of domestic sheep and the Asiatic mouflons from Iran and Turkey fell within
383 clade 6, but no common haplotypes were detected between domestic and wild sheep
384 (Figure 4C, Supplementary Table S10). The results were in accordance with the
385 mitogenome analyses and implied that partial haplotypes C+E in domestic sheep
386 could have originated from the Asiatic mouflons in Iran by introgression.
387

388 ***Genetic introgression***

389 To investigate the impact of introgression from the Asiatic mouflons in Iran on
390 domestic sheep, we implemented introgression tests between Iranian Asiatic mouflon
391 and SSA by *D* and f_4-ratio statistics[55,56]. We first conducted an admixture analysis
392 between Asiatic mouflons and the SSA populations. We identified three genetic
393 clusters with the smallest cross validation error ($K = 3$) (*O. gmelini*_1, *O. gmelini*_2,
394 and SSA) and detected genetic components of Asiatic mouflons in Iran in some
395 domestic populations (Figure 5A). Further, we employed the ABBA model (((P1, P2),
396 P3), P4) to assess introgression between *O. gmelini*_1 or *O. gmelini*_2 (P3) and SSA
397 sheep populations (P2). We used Menz sheep (MEN) from Africa as P1, which did
398 not show genetic introgression from Asiatic mouflon and SSA populations[8,15], and
399 bighorn sheep as the outgroup P4, which inhabit North America and have no gene
400 flows with SSA sheep populations. Of the two Asiatic mouflon populations, *O.*
401 *gmelini*_1 or *O. gmelini*_2 served as P3 and 18 SSA populations as P2. We calculated
402 the estimates of Patterson's *D*-statistic and f_4-ratio for all the 18 SSA sheep
403 populations, and determined introgression based on the following criteria: $D > 0$, *P*-
404 value < 0.05, and f_4-ratio > 0. Only four sheep populations (Garut, Garole,
405 Bangladeshi, and Sumatra sheep) showed genetic introgression from *O. gmelini*_2
406 (Figure 5B, Supplementary Table S11). All the four sheep populations are distributed
407 in the regions near to the Indian Ocean.

408

409 To further examine the genetic introgression in the four sheep populations, we
410 calculated f_{dm} with a 50-SNP window and 25-SNP step across the sheep genome. We
411 identified introgressed regions genomic with top 5% f_{dm} and $D > 0$. We obtained a
412 common introgression region in the four populations spanning 6,384,279 bp (Figure
413 5C, Supplementary Tables S12 and S13). We then calculated the PBS value of the
414 four populations to identify regions under selection with MEN sheep and *O.*
415 *gmelini*_1 as the reference populations. We considered the top 1% of PBS regions as
416 being under selection (Figure 5D, Supplementary Table S14). We overlapped the
417 selected regions with the common introgressed regions and detected 19,614 bp region
418 as the selected introgressed region (Supplementary Table S15). We further overlapped
419 the introgressed genomic region with sheep QTLs (Supplementary Table S15). We
420 found that it overlapped with the QTLs for body_weight, carcass_fat, and fat_weight
421 QTLs, which might be associated with the small body size of Garole, Bangladeshi,
422 and Garut sheep[57–59]. Additionally, the Nematodirus_FEC QTL could be related to the
423 phenotypes of Sumatra sheep's susceptibility to nematode parasitism[55]. The results
424 implied that the adaptive selected introgressed region could have been introgressed
425 from *O. gmelini*_2, and then have undergone the strong selection to facilitate their
426 adaptation to the hot and humid environments near the Indian Ocean. We further
427 overlapped the introgressed genomic region with sheep QTLs (Supplementary Table
428 S15).
429

430 **Discussion**

431 Traditionally, Asiatic mouflon had been classified into five subspecies using
432 phenotypic traits, such as coat color, horns, and geographic patterns[11]. However, the
433 intraspecies taxonomy has long lacked genetic evidence. Previous phylogenetic
434 analysis based on partial mtDNA divided Asiatic mouflon into several clades[3,4,58],
435 different from the taxonomy of five subspecies defined by phenotypic traits and
436 geographic patterns[11]. Here, we identified two genetic clusters of the Asiatic mouflons
437 in Iran based on WGS data: *O. gmelini*_2 limited in Kaboudan, and *O. gmelini* _1
438 over a wide geographic area in Iran (Figure 1). We detected the divergence time of
439 two clusters around ~5,000 years ago (Figure 3B, Supplementary Table S9), which
440 indicated a recent split of them, and was further supported by the inferred phylogenies
441 based on uniparental genetic markers (Figure 2B, C). All the individuals of *O. gmelini*
442 _2 collected in Kabudan inhabited the overlap geographic region with Armenian
443 Mouflon (*O. g.gmelini*), while individuals of *O. gmelini* _1 were sampled from the
444 distribution areas of several subspecies such as Armenian Mouflon (*O. g.gmelini*),
445 Isfahan mouflon (*O.g.isphahanica*), and Laristan mouflon (*O.g.laristanica*) (Figure
446 1A)[10].
447
448 These results suggested a significant genetic divergence between Armenian Mouflon
449 and the other subspecies (Supplementary Table S6). Nevertheless, the absence of
450 phenotypic data and other sample information for the Asian mouflons and a

451 comprehensive sampling throughout Iran limited our in-depth understanding of
452 within-species genetic differentiation in Asiatic mouflon. Also, Asiatic mouflon is
453 under the status of "Near Threatened", and their population size has declined in the
454 past two decades under competition with domestic sheep, poaching, and habitat
455 deterioration[10,60]. Therefore, the population genetic structure analyses and delimitation
456 of taxonomic classification based on WGS data are fundamental for conservation
457 decisions and effective conservation management[61,62].
458
459 Genetic variants of uniparental inheritance offer insights into the domestication and
460 evolutionary history of farm animals, such as pigs[63], sheep[5,14] and goats[64,65]. In
461 domestic sheep, here we identified five maternal haplogroups (A, B, C, D, and E) in
462 two monophyletic groups (A+B+D and C+E), and two paternal clades (HY1 and
463 HY2) (Figure 2B, C). We found that haplotypes of most of the Asiatic mouflons in
464 Iran clustered with haplogroups C+E based on mitochondrial marker (24/31) and
465 grouped with haplogroup HY2 based on Y-chromosomal SNPs (12/21)
466 (Supplementary Table S8). Nevertheless, the Asiatic mouflons in Iran showed higher
467 divergence from the dominant maternal lineages A and B and paternal lineages HY1
468 (Figure C, D).
469
470 Additionally, *Cyt-b* sequences collected from the whole geographic areas of Asiatic
471 mouflon also revealed a closer relationship between Asiatic mouflon in Iran and

472 domestic sheep of maternal haplogroups C+E (Figure 4). The results supported the
473 hypothesis that mtDNA haplogroups C+E and Y-chromosomal haplogroup HY2
474 could have originated from Asiatic mouflons in Iran[14]. However, WGS analysis
475 revealed intraspecies substructure in the Asiatic mouflons in Iran, and genetic
476 introgression from the *O. gmelini*_2 showed a larger impact on domestic sheep than
477 *O. gmelini*_1 (Figure 5). Similiar results were observed from Alberto et al., (2018)
478 that IROO individuals were split into subpopulations and had admixture with some
479 domestic sheep[38]. Therefore, Iran's two Asiatic mouflon populations showed similar
480 phylogenetic patterns of uniparental inheritance but different introgression patterns of
481 WGS, implying multiple introgression events from the Asiatic mouflons to domestic
482 sheep. In summary, our results implied that Asiatic mouflons in Iran might not be the
483 ancestor of domestic sheep[4,14], but could have contributed to the gene pool of
484 domestic sheep through multiple introgressions. Also, the results implied the
485 complexity of sheep domestication[4].
486
487 Recent studies suggested that sheep diffused to South and Southeast Asia via the
488 Middle East[15]. Our results revealed the genetic introgression from Asiatic mouflon
489 into some South-and-Southeast Asian sheep populations, supporting previous
490 findings[8,15]. Further, we detected genetic introgression in the four sheep populations
491 of Garut, Garole, Bangladeshi, and Sumatra, and were only from *O. gmelini*_2 in
492 Kabudan (Figure 5B). The introgression event is congruent with the demographic

493 history of *O. gmelini*_2 and South-and-Southeast Asian sheep inferred here and
494 previously: South-and-Southeast Asian sheep descended from Central-and-East Asian
495 populations around 5.85 to 6.01 ka and then admixed from the Middle Eastern sheep
496 3.93–4.18 ka[15], and *O.gmelini*_2 split from *O.gmelini*_1 around ~5,000 years ago
497 (Figure 3B).
498
499 In the four sheep populations, we found that common introgression signals from *O.*
500 *gmelini*_2 covered about 4.29% of the total introgressed regions. In contrast to
501 population-specific introgression, common introgression could be related to a specific
502 trait or adaptation to similar environments, such as *EPAS1* for the adaptation to
503 hypoxia[45,66]. We further integrated *PBS* analyses and found four positive selective
504 regions in the common introgressed regions, which were related to
505 carcass_fat_percentage, lean_meat_yield_percentage (Figure 5D). This might have
506 contributed to the small body size phenotype of Garole, Sumatra, and Bangladeshi
507 sheep[58,67,68]. We also found introgression signals in the *PCDH9* (Figure 5E), which
508 had been reported to be associated with local adaptation (e.g., hot and arid
509 environments) in sheep and goats[69,70], and fat deposition[71]. Also, early studies showed
510 adipogenesis, ECM remodeling, inflammation, and lipid droplet dynamics in mature
511 adipose tissue played essential roles in response to climate changes in sheep[72]. Further
512 analysis from Chip-seq and RNA-Seq data will be required to uncover the function
513 and regulatory mechanisms of *PCDH9* to adapt to hot and humid environments in

514 sheep.

515

516 In summary, we conducted a comprehensive and integrated analysis of Asiatic
517 mouflon and domestic sheep based on WGS, mitogenome and Y-chromosomal
518 genetic variants. Our results suggested that the mitochondrial haplogroups C/E and Y-
519 chromosomal haplogroup HY2 of domestic sheep could have originated from the
520 Asiatic mouflons in Iran by introgression and further clarified the effects of the
521 Iranian Asiatic mouflon in the domestication processing. Additionally, our results
522 revealed the introgression signals of the Asiatic mouflons in Iran in several domestic
523 sheep populations in South and Southeast Asia and showed that the introgression
524 could have facilitated their adaptation to hot and humid environments near the Indian
525 Ocean. Our findings added new insights into the domestication and subsequent
526 evolutionary history of sheep.

527

528 **Data availability**

529 All raw sequencing reads in this study were deposited in NCBI's SRA under the
530 accession number OR160429-OR161034.

531

532 **Code availability**

533 All scripts used for this work were performed using open-source software tools and

are available from the corresponding authors upon request.

Acknowledgments

This study was financially supported by grants from the National Natural Science 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). We thank the High-performance Computing Platform of China Agricultural University for providing computing resources.

Author contributions

F.-H.L. and M.-H.L. conceived and supervised the study. D.-F.W and F.-H.L. conducted the data analysis. D.-F.W, F.-H.L. and P.O wrote the manuscript, and M.-H.L. and P.O revised the paper. All the authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

554

555 **References**

556 1. Zeder, M. A. Domestication and early agriculture in the Mediterranean Basin:
557 Origins, diffusion, and impact. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11597–11604
558 (2008).
559 2. Pedrosa, S. et al. Evidence of three maternal lineages in near eastern sheep
560 supporting multiple domestication events. *Proc. R. Soc. B: Biol. Sci.* 272, 2211–
561 2217 (2005).
562 3. Demirci, S. et al. Mitochondrial DNA Diversity of Modern, Ancient and Wild
563 Sheep (*Ovis gmelinii anatolica*) from Turkey: New Insights on the Evolutionary
564 History of Sheep. *PLoS ONE* 8, e81952 (2013).
565 4. Her, C. et al. Broad maternal geographic origin of domestic sheep in Anatolia
566 and the Zagros. *Anim. Genet.* 53, 452–459 (2022).
567 5. Lv, F.-H. et al. Mitogenomic Meta-Analysis Identifies Two Phases of Migration
568 in the History of Eastern Eurasian Sheep. *Mol. Biol. Evol.* 32, 2515–2533 (2015).
569 6. Deng, J. et al. Paternal Origins and Migratory Episodes of Domestic Sheep.
570 *Curr. Biol.* 30, 4085-4095.e6 (2020).
571 7. Kalds, P. et al. Genetics of the phenotypic evolution in sheep: a molecular look
572 at diversity-driving genes. *Genet. Sel. Evol.* 54, 61 (2022).
573 8. Cheng, H. et al. Long divergent haplotypes introgressed from wild sheep are
574 associated with distinct morphological and adaptive characteristics in domestic

575 sheep. *PLoS Genet.* 19, e1010615 (2023).

576 9. Yeomans, L., Martin, L. & Richter, T. Expansion of the known distribution of
577 Asiatic mouflon (*Ovis orientalis*) in the Late Pleistocene of the Southern Levant.
578 *R. Soc. Open Sci.* 4, 170409 (2017).

579 10. Michel, S. & Ghoddousi, A. *Ovis Gmelini*: The IUCN Red List of Threatened
580 Species 2020. e. T54940218A22147055, https://dx. doi. org/10.2305/IUCN. UK
581 2, (2020).

582 11. Blyth, E. XXVI.— An Amended List of the Species of the Genus *Ovis*. *Ann.*
583 *Mag. Nat.* 7, 195–201 (1841).

584 12. Fernandez H, Taberlet P, Mashkour M, Vigne J-D, Luikart G. The First Steps of
585 Animal Domestication: New Archaeozoological Approaches, eds Vigne JD,
586 Peters J, Helmer D (Oxbow Books, Oxford), pp 50–54 (2005) .

587 13. Hiendleder, S., Kaupe, B., Wassmuth, R. & Janke, A. Molecular analysis of wild
588 and domestic sheep questions current nomenclature and provides evidence for
589 domestication from two different subspecies. *Proc. R. Soc. B: Biol. Sci.* 269,
590 893–904 (2002).

591 14. Li, R. et al. A Hu sheep genome with the first ovine Y chromosome reveal
592 introgression history after sheep domestication. *Sci. China Life Sci.* (2020)
593 doi:10.1007/s11427-020-1807-0.

594 15. Lv, F.-H. et al. Whole-Genome Resequencing of Worldwide Wild and Domestic
595 Sheep Elucidates Genetic Diversity, Introgression, and Agronomically Important

596 Loci. *Mol. Biol. Evol.* 39, msab353 (2022).

597 16. Cao, Y.-H. et al. Historical Introgression from Wild Relatives Enhanced Climatic
598 Adaptation and Resistance to Pneumonia in Sheep. *Mol. Biol. Evol.* 38, 838–855
599 (2021).

600 17. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for
601 Illumina sequence data. *Bioinformatics* 30, 2114–2120 (2014).

602 18. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–
603 Wheeler transform. *Bioinformatics* 25, 1754–1760 (2009).

604 19. McKenna, A. et al. The Genome Analysis Toolkit: A MapReduce framework for
605 analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303
606 (2010).

607 20. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*
608 25, 2078–2079 (2009).

609 21. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software
610 Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol*. 30,
611 772–780 (2013).

612 22. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for
613 automated alignment trimming in large-scale phylogenetic analyses.
614 *Bioinformatics* 25, 1972–1973 (2009).

615 23. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger
616 and richer datasets. *GigaScience* 4, 7 (2015).

617 24. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and
618 evolutionary analyses in R. *Bioinformatics* 35, 526–528 (2019).
619 25. Yu, G. Using ggtree to Visualize Data on Tree-Like Structures. *Curr. Protoc.*
620 *Bioinformatics* 69, (2020).
621 26. Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & François, O. Fast and
622 Efficient Estimation of Individual Ancestry Coefficients. *Genetics* 196, 973–983
623 (2014).
624 27. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* 27,
625 2156–2158 (2011).
626 28. Goudet, J. hierfstat, a package for r to compute and test hierarchical F-statistics.
627 *Mol. Ecol. Notes* 5, 184–186 (2005).
628 29. Paradis, E. pegas: an R package for population genetics with an integrated–
629 modular approach. *Bioinformatics* 26, 419–420 (2010).
630 30. Minh, B. Q. et al. IQ-TREE 2: New Models and Efficient Methods for
631 Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* 37, 1530–1534
632 (2020).
633 31. Steenwyk, J. L., Buida, T. J., Li, Y., Shen, X.-X. & Rokas, A. ClipKIT: A
634 multiple sequence alignment trimming software for accurate phylogenomic
635 inference. *PLoS Biol.* 18, e3001007 (2020).
636 32. Gatto, L., Catanzaro, D. & Milinkovitch, M. C. Assessing the Applicability of
637 the GTR Nucleotide Substitution Model Through Simulations. *Evol. Bioinform.*

638 *Online* 2, 145–155 (2007).

639 33. Leigh, J. W. & Bryant, D. popart : full-feature software for haplotype network
640 construction. *Methods Ecol. Evol.* 6, 1110–1116 (2015).

641 34. Ortiz, E. M. vcf2phylip v2.0: convert a VCF matrix into several matrix formats
642 for phylogenetic analysis. (2019) doi:10.5281/zenodo.2540861.

643 35. Terhorst, J., Kamm, J. A. & Song, Y. S. Robust and scalable inference of
644 population history from hundreds of unphased whole-genomes. *Nat. Genet.* 49,
645 303–309 (2017).

646 36. Beynon, S. E. et al. Population structure and history of the Welsh sheep breeds
647 determined by whole genome genotyping. *BMC Genet.* 16, 65 (2015).

648 37. Zhao, Y.-X. et al. Genomic Reconstruction of the History of Native Sheep
649 Reveals the Peopling Patterns of Nomads and the Expansion of Early
650 Pastoralism in East Asia. *Mol. Biol. Evol.* 34, 2380–2395 (2017).

651 38. Alberto, F. J. et al. Convergent genomic signatures of domestication in sheep and
652 goats. *Nat. Commun.* 9, 813 (2018).

653 39. Bouckaert, R. et al. BEAST 2.5: An advanced software platform for Bayesian
654 evolutionary analysis. *PLoS Comput. Biol.* 15, e1006650 (2019).

655 40. Drummond, A. J. Bayesian Coalescent Inference of Past Population Dynamics
656 from Molecular Sequences. *Mol. Biol. Evol.* 22, 1185–1192 (2005).

657 41. Upadhyay, M. et al. Whole genome sequencing reveals a complex introgression
658 history and the basis of adaptation to subarctic climate in wild sheep. *Mol. Ecol.*

659 30, 6701–6717 (2021).

660 42. Patterson, N. et al. Ancient Admixture in Human History. *Genetics* 192, 1065–
661 1093 (2012).

662 43. Malinsky, M., Matschiner, M. & Svardal, H. Dsuite - Fast D-statistics and
663 related admixture evidence from VCF files. *Mol. Ecol. Resour.* 21, 584–595
664 (2021).

665 44. Arnholt, A. T. & Evans, B. BSDA: Basic statistics and data analysis. *R package*
666 *version 1* (2017).

667 45. Yi, X. et al. Sequencing of 50 Human Exomes Reveals Adaptation to High
668 Altitude. *Science* 329, 75–78 (2010).

669 46. Hämälä, T. & Savolainen, O. Genomic Patterns of Local Adaptation under Gene
670 Flow in *Arabidopsis lyrata*. *Mol. Biol. Evol.* 36, 2557–2571 (2019).

671 47. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing
672 genomic features. *Bioinformatics* 26, 841–842 (2010).

673 48. Hu, Z.-L., Park, C. A. & Reecy, J. M. Bringing the Animal QTLdb and CorrDB
674 into the future: meeting new challenges and providing updated services. *Nucleic*
675 *Acids Res.* 50, D956–D961 (2022).

676 49. Dong, S.-S. et al. LDBlockShow: a fast and convenient tool for visualizing
677 linkage disequilibrium and haplotype blocks based on variant call format files.
678 *Brief. Bioinformatics* 22, bbaa227 (2021).

679 50. Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M. & Yang, T.-L. PopLDdecay: a fast

680 and effective tool for linkage disequilibrium decay analysis based on variant call
681 format files. *Bioinformatics* 35, 1786–1788 (2019).
682 51. Korunes, K. L. & Samuk, K. pixy : Unbiased estimation of nucleotide diversity
683 and divergence in the presence of missing data. *Mol. Ecol. Resour.* 21, 1359–
684 1368 (2021).
685 52. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and
686 correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849
687 (2016).
688 53. Sanna, D. et al. The First Mitogenome of the Cyprus Mouflon (*Ovis gmelini*
689 *ophion*): New Insights into the Phylogeny of the Genus *Ovis*. *PLoS ONE* 10,
690 e0144257 (2015).
691 54. Teng, H. et al. Population Genomics Reveals Speciation and Introgression
692 between Brown Norway Rats and Their Sibling Species. *Mol. Biol. Evol.* 34,
693 2214–2228 (2017).
694 55. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing
695 Indian population history. *Nature* 461, 489–494 (2009).
696 56. Green, R. E. et al. A Draft Sequence of the Neandertal Genome. *Science* 328,
697 710–722 (2010).
698 57. Jakaria, J., Zein, M. S. A., Sulandari, S., Subandriyo, S. & Muladno, M. The use
699 of microsatellite markers to study genetic diversity in Indonesian sheep. *J.*
700 *Indones. Trop. Anim. Agric.* 37, 1–9 (2012).

58. Sharma, R. C., Arora, A. L., Narula, H. K. & Singh, R. N. Characteristics of Garole sheep in India. *Anim. Genet. Resour. Inf.* 26, (1999).

59. Ramanuj, B., Mandal, P., Pal, U. & Kunal, R. Productivity and Genetic Potential of Garole Sheep of India-A Review. *Asian J. Anim. Sci.* 4, 170–189 (2010).

60. Bleyhl, B. et al. Assessing niche overlap between domestic and threatened wild sheep to identify conservation priority areas. *Divers. Distrib*. 25, 129–141 (2019).

61. Yang, L. et al. Evolutionary Conservation Genomics Reveals Recent Speciation and Local Adaptation in Threatened Takins. *Mol. Biol. Evol*. 39, msac111 (2022).

62. Hu, Y. et al. Genomic evidence for two phylogenetic species and long-term population bottlenecks in red pandas. *Sci. Adv.* 6, eaax5751 (2020).

63. Ai, H. et al. Human-Mediated Admixture and Selection Shape the Diversity on the Modern Swine (*Sus scrofa*) Y Chromosomes. *Mol. Biol. Evol.* 38, 5051–5065 (2021).

64. Luikart, G. et al. Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proc. Natl. Acad. Sci. U.S.A.* 98, 5927–5932 (2001).

65. Naderi, S. et al. The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17659–17664 (2008).

66. Wu, D.-D. et al. Convergent genomic signatures of high-altitude adaptation

722 among domestic mammals. *Natl. Sci. Rev.* 7, 952–963 (2020).

723 67. Porter, V. Mason's World Dictionary of Livestock Breeds, Types and Varieties,
724 6th Edition. (CABI, 2020).

725 68. Putra, W. P. B., Alwi, Y. & Suryani, H. Morphometrics characterization of Thin-
726 Tail sheep in lowland and highland areas. *Trop. Anim. Sci. J.* 44, 386–398
727 (2021).

728 69. Serranito, B. et al. Local adaptations of Mediterranean sheep and goats through
729 an integrative approach. *Sci. Rep.* 11, 21363 (2021).

730 70. Kim, E.-S. et al. Multiple genomic signatures of selection in goats and sheep
731 indigenous to a hot arid environment. *Heredity* 116, 255–264 (2016).

732 71. Mastrangelo, S. et al. Genome-wide scan of fat-tail sheep identifies signals of
733 selection for fat deposition and adaptation. *Anim. Prod. Sci.* 59, 835–848 (2018).

734 72. Xu, Y.-X. et al. Whole-body adipose tissue multi-omic analyses in sheep reveal
735 molecular mechanisms underlying local adaptation to extreme environments.
736 *Commun. Biol.* 6, 159 (2023).

737

738

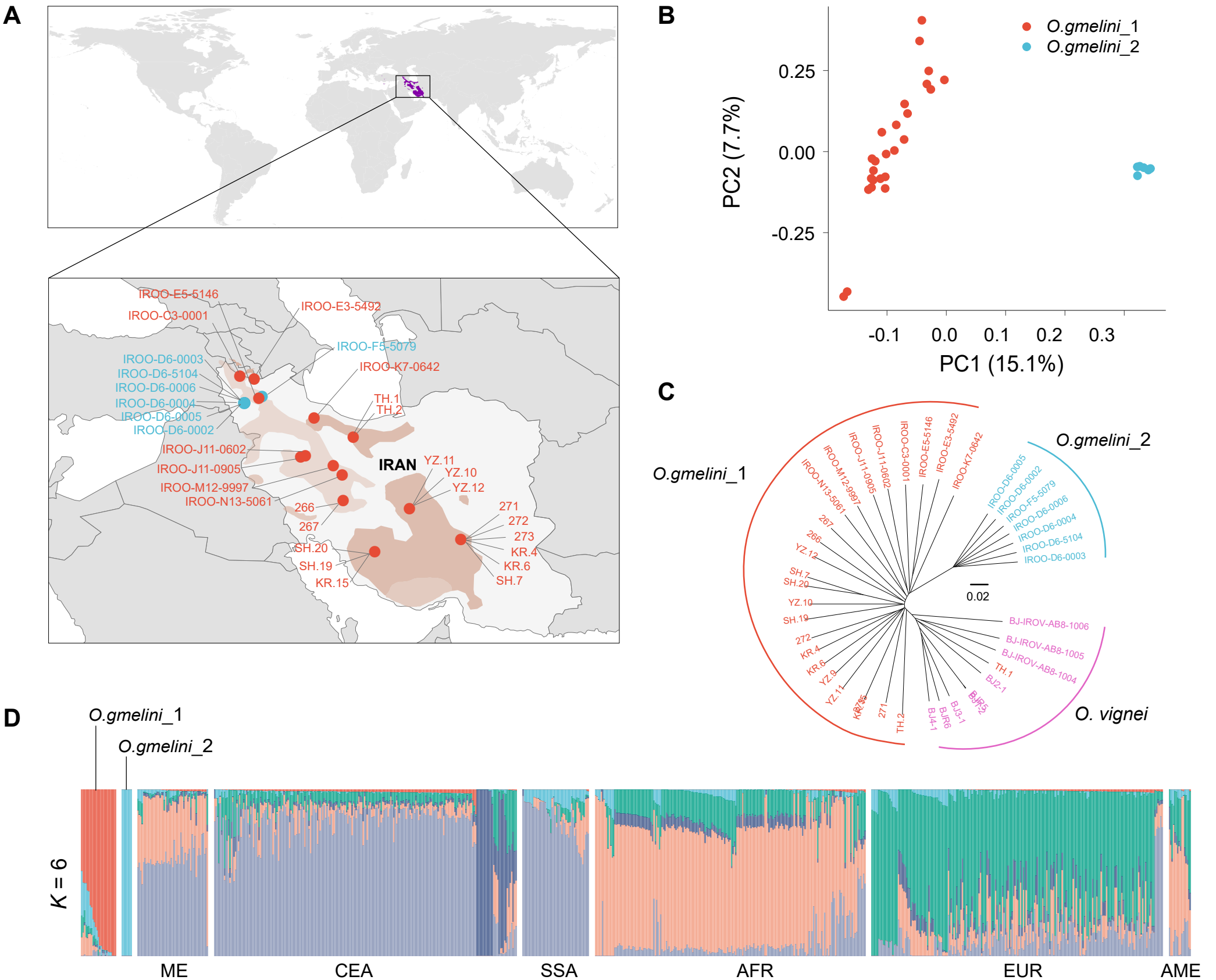

Fig. 1 Population genetics structure of Asiatic mouflons in Iran (*O. gmelini*). **a** Geographic distribution of 32 Asiatic mouflons. Purple and blue dots indicate the location of *O. gmelini*_1 and *O. gmelini_2* subpopulations; **b** Principal components 1 and 2 for the 32 Asiatic mouflon in Iran;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

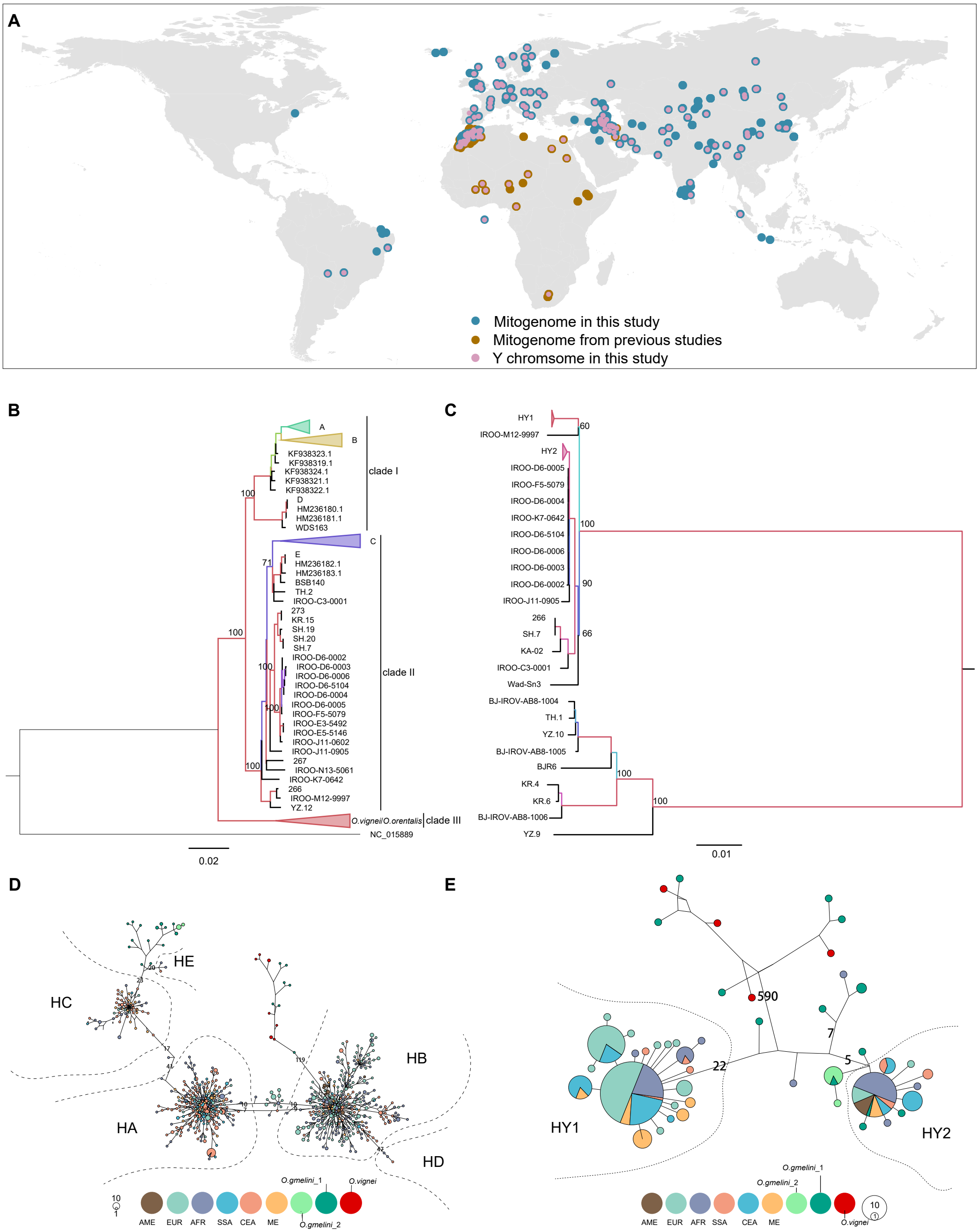

Fig. 2 Phylogeny of Asiatic mouflon, urial, and domestic sheep inferred from mitogenomes and Y-chromosomal genetic variants. a Geographical distribution of the samples in this and previous studies; **b** Phylogeny of Asiatic mouflon, urial, and domestic sheep inferred from 767 mitogenomes; **c** Phylogeny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

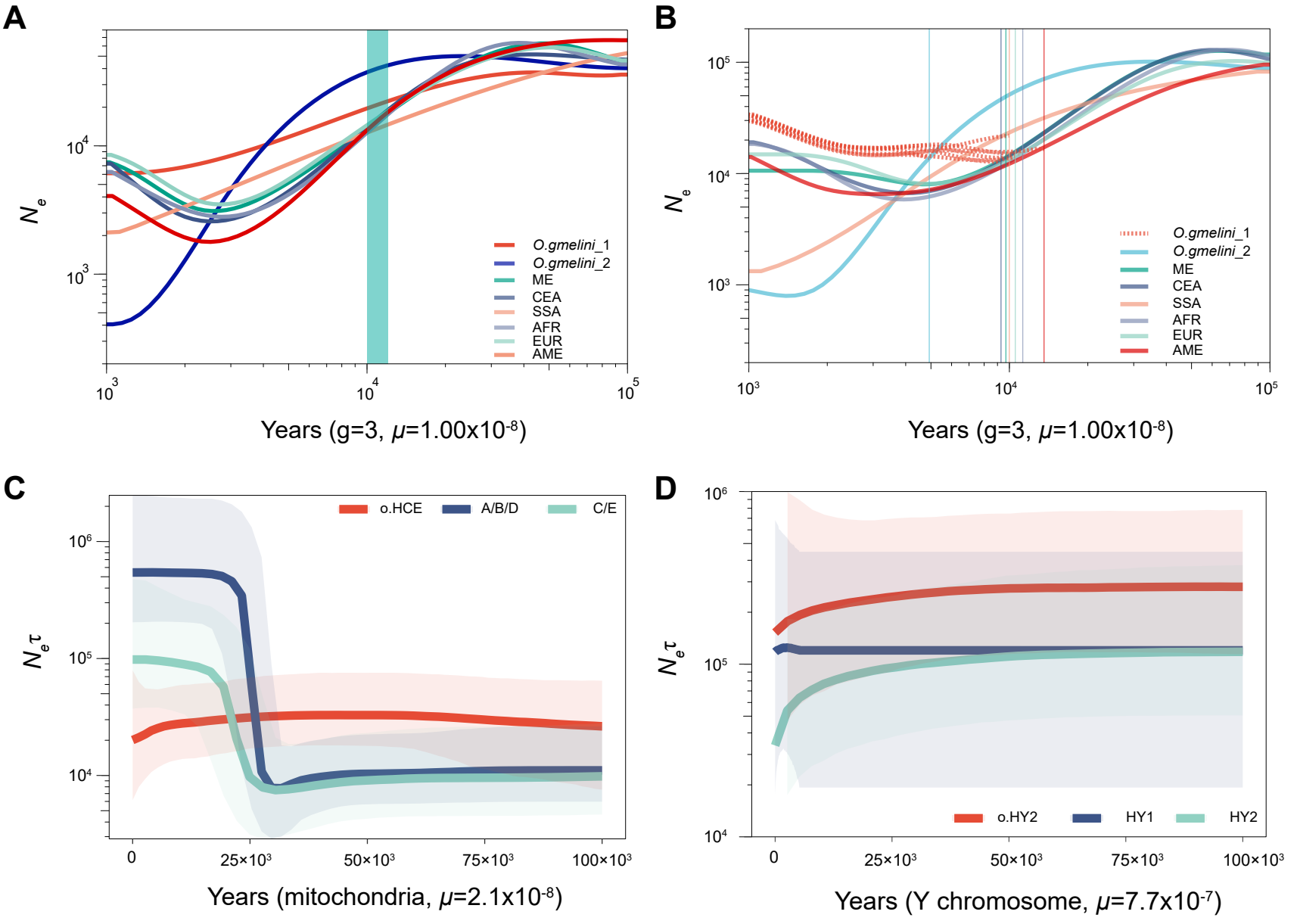

Fig. 3 Demographic history of Asiatic mouflon in Iran and domestic sheep. a Effective population sizes (N_e) across six domestic sheep populations and the two subpopulations of Asiatic mouflon in Iran inferred from SMC++; **b** Inference of split times between *O. gmelini*_1 and the other wild and domestic sheep populations. The vertical lines represent the split time between each pair of populations; **c** Bayesian skyline plot (BSP) reconstruction of historical change in effective population size (N_e) of domestic sheep of maternal haplogroups HA+B+D, and HC+E, and the Asiatic mouflons in Iran; **d** Bayesian skyline plot (BSP) reconstruction of historical changes in $N_e$of paternal lineages HY1 and HY2 of domestic sheep and the Asiatic mouflon in Iran. The middle line represents median estimates of $N_e\tau$, and the shaded area represents 95% highest posterior density (HPD) intervals.

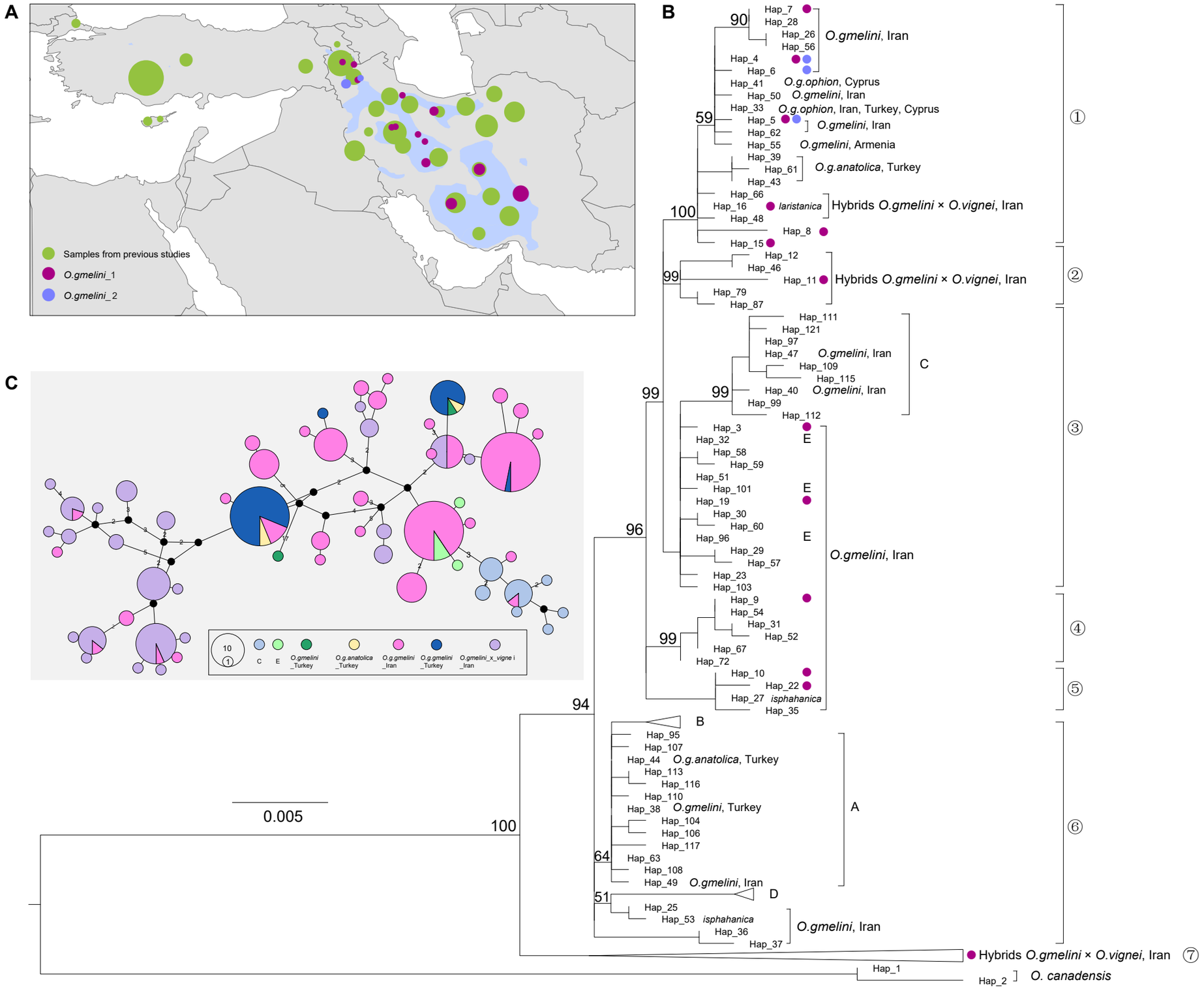

Fig. 4 Geographic distribution of Asiatic mouflon and the phylogenetic relationship among Asiatic mouflon subspecies and domestic sheep. **a** Geographical distribution of Asiatic mouflon and samples information of 32 Asiatic mouflon in Iran and 301domestic sheep samples; **b** Phylogenetic tree inferred based on complete *Cyt-b* sequences using the maximum likelihood algorithms. Bootstrap values above 50% were labeled on the phylogenetic tree. **c** Network of Asiatic mouflon in Iran and Turkey and domestic sheep lineage of C and E built from complete *Cyt-b* sequences.

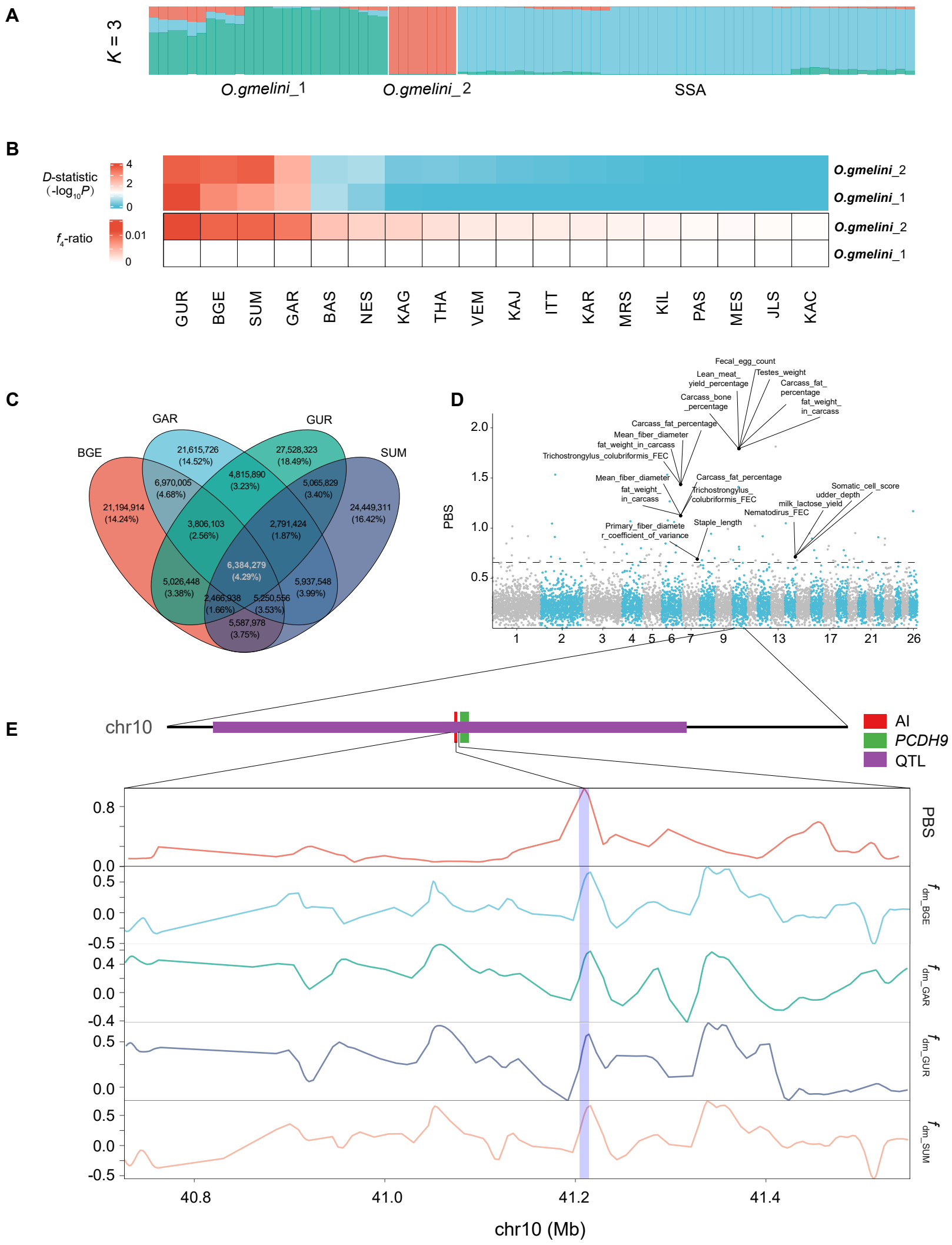

Fig. 5 Genetic introgression of Asiatic mouflon in SSA domestic populations. a Admixture proportions of Asiatic mouflon in Iran and SSA domestic populations inferred from the program the sNMF v1.2 ($K = 3$) using whole-genome SNPs. **b** Heatmap of *D*-statistics and f_4-ratio calculated by Dsuite v0.5 applied to data including *O. gmelini*_1, *O. gmelini*_2, and SSA domestic populations; **c**, Venn diagram for the shared introgression regions among four domestic South and Southeast populations; **d** Manhattan plot showing signals of selection in the four introgressed SSA populations (BGE/GAR/GUR/SUM) detected by the PBS with the dashed line marking 99.9% threshold. The introgression region, QTL, and adjacent gene *PCDH9* were labeled with blue, purple, and green bars, respectively. The sheep QTLs overlapped with the top 1% PBS regions are shown, **e** PBS and f_{dm} values around the significant introgressed genomic region on Chr.10.

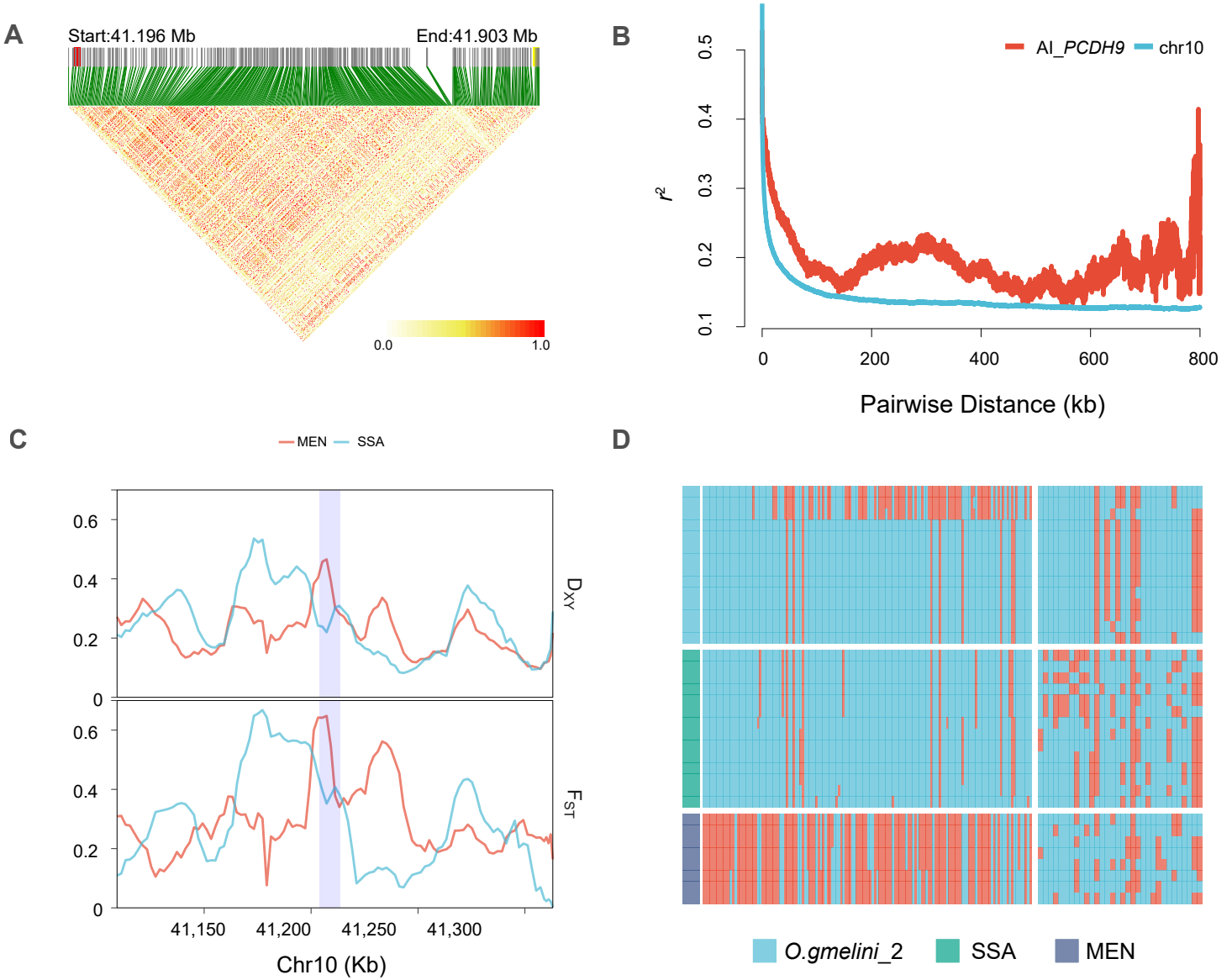

Fig. 6 Candidate gene *PCDH9* associated with the significant introgression signal on Chr. 10. **a** Linkage disequilibrium (LD) blocks in the significant introgression region and gene *PCDH9* (chr10:41,104,271-43,073,231). The red and yellow bars represent the location of the significant introgression region and gene *PCDH9*; **b** Pattern of LD decay in the introgressed populations. The black line is calculated from the whole 10 chromosomes, the red line is calculated from the beginning of the significant region to the next 800 kb (chr10:41,204,271-42,004,271); **c** D_{xy} and F_{ST} around the significant introgression region on Chr. 10 (chr10:41,104,271-41,313,804), calculated by pixy v1.2.7.beta1; **d** Haplotype pattern of the significant introgression region on Chr. 10 (chr10:41,204,271-41,213,804) (left) and exons of *PCDH9* (right).