The natural adaptation and human selection history of African sheep genomes

Abulgasim M. Ahbara^{1,2,3,4,12*}, Christelle Robert⁵, Adebabay Kebede^{6,7}, Ayelle Abeba⁸, Suliman Latairish⁹, Mukhtar Omar Agoub¹⁰, Hassan H. Musa¹¹, Pam Wiener⁵, Emily Clark⁵, Olivier Hanotte^{2,6*}, Joram M. Mwacharo^{3,5,12*}

¹Department of Zoology, Faculty of Sciences, Misurata University, Misurata, Libya

²School of Life Sciences, University of Nottingham, University Park, Nottingham, UK

³Small Ruminant Genomics, International Center for Agricultural Research in the Dry Areas (ICARDA), Addis Ababa, Ethiopia

⁴LiveGene, International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia

⁵Centre for Tropical Livestock Genetics and Health (CTLGH), The Roslin Institute, University of Edinburgh, UK

⁶LiveGene—CTLGH, International Livestock Research Institute (ILRI) Ethiopia, Addis Ababa, Ethiopia

⁷Amhara Regional Agricultural Research Institute, Bahir Dar, Ethiopia

⁸Debre Berhan Research Centre, Debre Berhan, Ethiopia

⁹Department of Animal Production, Faculty of Agriculture, Misurata University, Misurata, Libya

¹⁰Agricultural Research Center, Misurata, Libya

¹¹Faculty of Medical Laboratory Sciences, University of Khartoum, Sudan

¹²Animal and Veterinary Sciences, SRUC, The Roslin Institute Building, Midlothian Edinburgh, UK

*Correspondence: aahbara@sruc.ac.uk; o.hanotte@cgiar.org; j.mwacharo@cgiar.org

Abstract

African sheep manifest diverse but distinct physio-anatomical traits which are the outcomes of natural- and human-driven selection. Here, we generated 34.8 million variants from 150 indigenous African sheep genomes sequenced at an average depth of ~54x for 130 samples (Ethiopia, Libya) and ~10x for 20 samples (Sudan), representing sheep from diverse environments, tail morphology and post-Neolithic introductions to Africa. Phylogenetic and model-based admixture analysis provided evidence of four genetic groups that correspond to altitudinal geographic origins and tail morphotypes. Comparative genomic analysis identified targets of selection spanning conserved haplotype structures overlapping genes and gene families relating to hypoxia responses, caudal vertebrae and tail skeleton length, ear morphology, and tail fat-depot structure. Our findings provide novel insights underpinning variation and response to human selection and environmental adaptation, and possible pleiotropic gene interactions in indigenous African sheep genomes, which guaranteed the successful establishment of the species on the continent.

Keywords: *Diversity, Demographic history, Fat-rumped, Fat-tailed, Ovis aries, Thin-tailed*

Introduction

During their human-mediated dispersal across the globe, domestic sheep, *Ovis aries*, encountered and had to adapt to a wide range of novel agro-ecological environments. Natural and artificial selection could have shaped the genomes of domestic sheep to enhance their adaptive fitness to novel African environments. Tail fat deposition could have facilitated such adaptations, and based on tail fat-depot structures, three classes of sheep have been recognized in the continent: fat-tailed, thin-tailed and fat-rumped. Fat-tailed sheep occur in northern, eastern and southern Africa, thin-tailed sheep are predominantly found across western Africa, in Sudan and in parts of northern Africa, and the fat-rumped sheep are restricted to the Horn of Africa (Porter, 2020). Archaeological findings from northern Africa suggest that thin-tailed sheep are the most ancient in the continent and that domestic sheep entered

51 Africa from the Levant through the Isthmus of Suez in multiple waves (Clutton-Brock, 1993; Marshall,
52 2000; Gautier, 2006). However, today the fat-tailed sheep predominate in northern Africa. In eastern
53 Africa, the earliest evidence of sheep dates to 4500-3500 BP with rock paintings from the Lake Turkana
54 basin depicting thin-tailed sheep (Barthelme, 1985; MacDonald and MacDonald, 2000), which today
55 are mainly found in Ethiopia (Benishangul-Gumuz and North Gondar regions). Rock paintings from
56 eastern Ethiopian highlands show fat-tailed sheep alongside humpless cattle (Clark and Williams, 1978).
57 The latter arrived on the continent earlier than their humped counterparts which today predominate this
58 region (Gifford-Gonzalez and Hanotte, 2011). Depictions of fat-rumped sheep lack in archaeological
59 findings, thus their origin, entry point and time of diffusion into the continent remains unknown.
60
61 Today, Africa is home to ~383 million sheep (FAOSTAT 2020; accessed July 2020) distributed among
62 140 phenotypically diverse populations (DAD-IS 2020; accessed July 2020). The continent exhibits
63 marked differences in agro-eco-climates and biophysical challenges
64 (https://iiasa.ac.at/web/home/research/researchPrograms/water/GAEZ_v4.html). It is also home to an
65 exceptional human ethnic agro-pastoral diversity of ancient origin. Together with natural selection for
66 adaptation to diverse environments, human driven selection for economic, socio-cultural and aesthetic
67 traits have shaped the genomes of African livestock, resulting in unprecedented variation within and
68 between populations (e.g., Kim et al., (2020); Gheyas et al., (2021)). The genetic control underpinning
69 such adaptations and variations remain largely unknown. Here, we generated and analysed whole-
70 genome sequences alongside tail morphometric data of indigenous African sheep from Libya, Sudan
71 and Ethiopia. The studied populations represented fat-tailed, thin-tailed and fat-rumped African sheep
72 tail morphotypes and were sampled from different agro-ecologies, ranging from cold high-altitude
73 highlands to hot arid lowland environments. Using the data, we investigated genome-wide levels of
74 genetic variation and structure and used the information to infer demographic-adaptive dynamics of the
75 species in the continent.

76

77 **Materials and Methods**

78 **Sample collection**

79 Tissue samples were collected from 150 individuals of 15 indigenous African fat-tailed, fat-rumped and
 80 thin-tailed sheep from Ethiopia, Libya and Sudan (Table S1, S2; Figure S1). Genomic DNA was isolated
 81 with DNeasy Blood and Tissue Kit (Qiagen) and quality checked with the Nanodrop. Using the Covaris
 82 System, 3 µg of genomic DNA were randomly sheared to generate inserts of ~300 bp. The sheared DNA
 83 was end-repaired, A-tailed, adaptor ligated, and amplified with the TruSeq DNA Sample Preparation
 84 Kit (Illumina, San Diego, CA, USA). Paired-end sequencing was conducted with the Illumina
 85 HiSeq2000. We performed sequence quality checks with the fastQC software
 86 (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>). The paired-end sequence reads were mapped
 87 against the Oar_v3.1.75 sheep reference genome using Burrows-Wheeler Aligner software (Li and
 88 Durbin, 2009). We used default parameters (except the “--no-mixed” option) to suppress unpaired
 89 alignments. Potential PCR duplicates were filtered using the “REMOVE_DUPLICATES = true” option
 90 in “MarkDuplicates” command-line of Picard (<http://broadinstitute.github.io/picard>). SAMtools (Li and
 91 Durbin, 2009) was used to create the index files for reference and bam files. The genome analysis toolkit
 92 (GATK) 3.1 (<https://gatk.broadinstitute.org>) was used to realign the reads to correct for any
 93 misalignments arising from the presence of indels (“RealignerTargetCreator” and “IndelRealigner”
 94 arguments). “UnifiedGenotyper” and “SelectVariants” of GATK were used to call SNPs. To filter
 95 variants and to avoid possible false positives, SNPs with: (i) a phred-scaled quality score < 30, (ii) MQ0
 96 (mapping quality zero) > 4 and quality depth (unfiltered depth of non-reference samples) < 5, and (iii)
 97 SNPs with FS (phred-scaled *P* value using Fisher’s exact test) > 200, were filtered out. Following these
 98 filtration steps, the retained SNPs were used for analysis.

99

100 **Preparation and measurement of caudal vertebrae**

101 We first subdivided the 15 populations into five groups based on the physical assessment of the tail
102 length and size of fat-depots, viz, i) Ethiopian fat-rumped (Kefis, Adane, Arabo, Segentu), ii) Ethiopian
103 long fat-tailed (Bonga, Kido, Gesses, Loya, Shubin Gemo, Doyogena), iii) Ethiopian short fat-tailed
104 (Molale, Gafera) sheep, iv) Sudan long thin-tailed sheep (Hammari, Kabashi) and v) Libya long fat-
105 tailed sheep (Barbary).

106
107 One mature animal (≥ 3 permanent pairs of incisor teeth) of Kefis (fat-rumped), Loya (long fat-tailed),
108 Menz (short fat-tailed), Kabashi (long thin-tailed) and Barbary (Libyan long fat-tailed) was selected and
109 slaughtered, the tail eviscerated from the body and the full tail skeleton processed following Zhi et al.
110 (2018). The tail was deskinning, and the flesh and fat gently scrapped from the caudal vertebrae (CVs).
111 The CVs were then incubated in ethanol for five days after which they were cleaned by soaking for two
112 days in 0.5% NaOH. Residual flesh and fat were removed by further soaking the CVs in petrol for three
113 days. The number of CVs were counted and the length (cm) of each CV was determined with a Vernier
114 calliper. The length of the complete tail skeleton was determined with a tape measure. The average
115 length and standard deviation of the CV were calculated with Excel.

116

117 **Genetic diversity**

118 The VCFtools v.0.1.15 (Danecek et al., 2011) was used to estimate the observed (H_O) and expected (H_E)
119 heterozygosity, and nucleotide diversity (π) as indicators of intra-population diversity, while the
120 inbreeding coefficient (F) and runs of homozygosity (RoH) were estimated as indicators of genome-
121 wide autozygosity. The RoH was estimated with the BCFtools software package (Narasimhan et al.,
122 2016). We estimated the RoH -derived genomic inbreeding coefficient (F_{RoH}) following McQuillan et
123 al., (2008) as:

$$124 \quad F_{RoH} = \frac{L_{RoH}}{L_{AUTO}}$$

Where L_{ROH} is the total length of RoH of each individual in the genome and L_{AUTO} is the length of the sheep autosomal genome (~2600 Mb).

127

128 **Phylogenetic inference and genetic structure analysis**

We visualised population structure using principal component analysis (PCA) performed with Plink 1.9 (Purcell et al., 2007). The first two eigenvectors were plotted to summarise individual relationships. For this analysis, we used 17 million autosomal SNPs, which were retained after filtering out SNPs with minor allele frequency (MAF) < 0.01.

133

We investigated the proportion of the genome arising from common ancestry using the ADMIXTURE software v.1.3 (Alexander et al., 2009). For this analysis, we used 2.4 million autosomal SNPs that were retained after pruning the 17 million SNPs used in PCA, for linkage disequilibrium (LD). We implemented the block relaxation algorithm with the Kinship (K) values set from 2 to 15; five runs were performed for each K . A five-fold cross-validation (CV) procedure was used to determine the optimal number of ancestral genomes (K) and the proportion of admixture. The PCA and ADMIXTURE results were visualized with GENESIS (Buchmann and Hazelhurst, 2014).

141

Patterns of population splits and gene flow were assessed with the maximum likelihood tree-based procedure in TreeMix (Pickrell and Pritchard, 2012). The thin-tailed sheep from Sudan were defined as the root. The optimal number of migration events (m) were determined following Pickrell and Pritchard (2012) and the models were visualized with R using the script provided in TreeMix. The contribution of each migration vector that was added to the tree was determined using the “ f index” (Pickrell and Pritchard, 2012). We modelled the optimal number of migration events and edges (M) from 1 to 10 with the default settings of OptM package in R (<https://www.r-pkg.org/pkg/OptM>; accessed April 2020).

Five runs were performed for each “M”. The ML phylogenetic tree depicting relationships and the optimal migration event(s) was generated using the default "Evanno"-like method in R.

Demographic history and dynamics

We estimated the LD parameter (r^2) with Plink v1.9 considering the genetic groups/structure inferred from the phylogeny and ADMIXTURE analysis. The r^2 values were sorted and binned in 50 Kb to 10 Mb inter-SNP distances and the genome-wide LD decay plotted using R.

The estimated values of r^2 were used to model changes in effective population sizes (N_e) over generation time (up to 1000 generations ago) using SNeP (Barbato et al., 2015), for each genetic group/structure inferred by the phylogenetic and ADMIXTURE analysis. The N_e values were estimated following Sved (1971) as:

$$N_e = \frac{1}{4c} \left(\frac{1}{E(r^2)} - 1 \right)$$

where c is the genetic distance in centimorgans (considering 1 cM = 1 Mb), and $E(r^2)$ represents the expected LD (r^2) for a distance of c Morgan's. The time intervals denoting the number of generations in the past (t) were estimated using $t = 1/2c$ (Hayes et al., 2003).

Genome-wide signatures of selection

We assessed genome-wide signatures of selection for i) adaptation to contrasting environments (e.g., high altitude humid and low altitude dry/arid environments), ii) differences in tail length, and iii) differences in the size of tail fat-depots.

Based on the altitude and agroecology of their geographic home range (Table S2), we classified the study populations into three groups. These were i) sheep from high altitude humid environment (Bonga,

173 Kido, Gesses, Loya, Shubi Gemo, Doyogena and Gafera from Ethiopia), ii) sheep from low altitude hot-
174 dry desert environment (Hammari and Kabashi from Sudan), and iii) sheep from coastal hot-humid arid
175 environment (the Barbary from Libya). Contrasting these three groups allowed for the detection of
176 genomic signatures for adaptation to contrasting environments.

177

178 Based on the average lengths of the CVs and the length of the complete tail skeleton, we classified the
179 studied populations into two groups, i) short-tailed sheep comprising short fat-tailed and fat-rumped
180 sheep, and ii) long-tailed sheep comprising long fat-tailed and long thin-tailed sheep. These were used
181 to investigate genomic regions associated with differences in CV sizes and in the length of the tail
182 skeleton.

183

184 To identify candidate regions associated with differences in the size of fat tail-depots, we contrasted the
185 long thin-tailed sheep with the long fat-tailed, fat-rumped and short fat-tailed sheep, respectively.

186

187 Signatures of selection were investigated by analysing 34.8 million autosomal SNPs with three methods:
188 i) within-group pooled heterozygosity (H_p) (Rubin et al., 2010), ii) genetic differentiation based on F_{ST}
189 (Weir and Cockerham, 1984), and iii) the haplotype-based $XP-EHH$ (Sabeti et al., 2007). A sliding
190 window was used to perform the H_p and F_{ST} analysis with VCFtools v.0.1.15. Based on LD decay trend,
191 a 100 kb non-overlapping window was chosen for the analysis. By exploring several sliding-window
192 distances (10 kb, 20 kb, 50 kb, 75 kb, 100 kb), the 10 kb sliding-window distance gave the best resolution
193 of the selection signals and was thus chosen for our analysis.

194

195 We estimated the genome-wide H_p statistic using the formula:

196

$$H_p = \frac{2 \sum nMaj \sum nMin}{(\sum nMaj \sum nMin)^2}$$

where $\sum n_{Maj}$ and $\sum n_{Min}$ are the sum of the number of the major allele and the number of the minor allele, respectively. We transformed the H_p values into Z-scores using the formula:

$$Z_{Hp} = \frac{(Hp - \mu_{Hp})}{\sigma_{Hp}}$$

Where μ_{Hp} is the average of the overall heterozygosity and σ_{Hp} is the standard deviation for all the windows within the test group.

The F_{ST} (Weir and Cockerham, 1984) values were estimated for each SNP in each window between test groups with the formulae:

$$F_{ST} = 1 - \frac{p_1q_1 + p_2q_2}{2prqr}$$

where p_1 , p_2 and q_1 , q_2 are the frequencies of alleles “A” and “a” in the first and second test groups, respectively, and pr and qr are the frequencies of alleles “A” and “a”, respectively, across the test groups. The F_{ST} values were standardized into Z-scores as follows:

$$Z_{F_{ST}} = \frac{F_{ST} - \mu_{F_{ST}}}{\sigma_{F_{ST}}}$$

where $\mu_{F_{ST}}$ is the overall average value of F_{ST} and $\sigma_{F_{ST}}$ is the standard deviation derived from all the windows tested between test groups.

The $XP-EHH$ method contrasts extended haplotype homozygosity (EHH) between populations in detecting selection signatures (Sabeti et al., 2007). The approach estimates and contrasts the iES statistic (the pattern of integrated EHH of the same allele) between populations as follows:

$$\ln(XP - EHH) = \ln\left(\frac{iES_{pop1}}{iES_{pop2}}\right)$$

Where $XP - EHH$ is the cross-population differentiation, iES_{pop1} is the integrated EHH s for the targeted sheep group and iES_{pop2} is the integrated EHH s for the reference sheep group.

219

220 For the analysis, haplotype phasing was inferred on all bi-allelic SNPs with BEAGLE (Browning and
221 Browning, 2007). Assuming 100 Mb = 1 Morgan (Quinlan and Hall, 2010), phased haplotypes were
222 converted to IMPUTE format using VCFtools v.0.1.15. The generated haplotypes were then used to
223 estimate *XP-EHH* scores in pair-wise comparisons between test groups using HAPBIN (Maclean et al.,
224 2015). To determine the empirical significance of the *XP-EHH* statistic, *XP-EHH* scores were
225 normalized by subtracting the mean and dividing it with the standard deviation of all the scores.
226 Negative *XP-EHH* scores indicate selection in the reference group, while positive scores indicate
227 selection in the test group.

228

229 The top negative 99% values for *ZHp*, and the top 0.1% of the empirical distributions of the *ZF_{ST}* and
230 *XP-EHH* scores were used as the cut-off threshold to detect outliers and define candidate regions under
231 selection. For a region to be considered a candidate under selection, it had to span at least three
232 significant SNPs. Candidate regions that overlapped between at least two methods were identified and
233 merged using Bedtools v.2.25.0 (Quinlan and Hall, 2010).

234

235 **Haplotype structure analysis**

236 We further investigated the candidate region intervals and putative genes for evidence of conserved
237 haplotype structures in a test group. We estimated haplotype frequencies and visualized the haplotype
238 structures using the *hapFLK* software (Fariello et al., 2013) employing the scripts from the *hapFLK*
239 homepage (<https://forge-dga.jouy.inra.fr/projects/hapflk>).

240

241 **Functional annotation**

242 The candidate regions were annotated based on the Oar-v3.1.75 sheep reference genome assembly with
243 the *Ensembl BioMart* tool (<http://www.ensembl.org/biomart>). Using the annotated genes found in all

the candidate regions, DAVID v6.8 (Huang et al., 2009) and KOBAS v3.0 (Xie et al., 2011) were used to perform functional enrichment analysis. The *O. aries* annotation was used as the background species. To interpret the gene functions in a livestock context, we retrieved information available on their functional effects from literature.

Results

Variant discovery and annotation

The overall alignment rate of the reads to the ovine reference sequence was 98.56% with an average sequencing depth of ~54x for Ethiopian and Libyan samples and ~10x for Sudanese samples. The summary statistics for the sequence parameters are shown in Supplementary Table S3. In general, a comparison with the *O. aries* dbSNP database, revealed that ~6% of the SNPs and InDels were not present in the database (Table S3, S4). Of the 34,857,882 SNPs generated in this study, ~36.8%, ~55.9% and ~0.79% were intronic, intergenic and exonic, respectively (Table S5; Figure S2).

Tail skeleton morphometry

The tail and skeletal tail characteristics of Menz, Kefis, Loya, Barbary and Kabashi are shown in Figure 1a. The tail of Loya, a long fat-tailed sheep, extends to the hock joint while that of Menz and Kefis representing short fat-tailed and fat-rumped sheep, respectively does not extend beyond the hock joint. The thin-tailed Kabashi and long fat-tailed Barbary sheep have the longest tails, which extend beyond the hock joint, with the tail of SD sheep almost touching the ground. The short fat-tailed and fat-rumped sheep have 8-10 CVs, the Ethiopian long-tailed has 18, and LB and SD have 22 each. SD has the highest average length of CV (2.72 ± 0.364 cm), followed by the Ethiopian long-tailed (2.42 ± 0.243 cm), fat-rumped (2.24 ± 0.113 cm) and LB (2.20 ± 0.183 cm). Although LB has 22 CV's, their average lengths are the shortest resulting in a shorter overall tail skeletal length compared to SD. Based on the average lengths of the individual CV's and the length of the complete tail skeleton, we classified the study

populations into two groups, i) long-tail sheep comprising SD and the Ethiopian long-tail and, ii) short-tail sheep comprising LB and the Ethiopian fat-rumped.

Genetic diversity

The average estimates of H_O , H_E , π , F , RoH and F_{RoH} , and their standard deviations are shown in Table 1. The least and most diverse populations based on H_O and H_E estimates were Kido and Segentu, respectively. The highest value of π was in Barbary and the lowest in Bonga and Loya. Excepting Hammari, Kabashi and Menz, the other populations showed low values of inbreeding ($F < 1.0\%$). Hammari and Kabashi reported the largest mean sizes of RoH and Barbary had the smallest (Table 1; Figure S3). Barbary also had the lowest F_{RoH} , while Bonga and Loya had the highest (Table 1).

Population structure

The PCA (Figure 1b) separated Ethiopian sheep into two genetic groups corresponding to the ones reported by Ahbara et al. (2019) based on the analysis of 50K SNP genotypes. For consistency, we adopted Ahbara et al. (2019) nomenclature viz: “ET-G1” comprising fat-rumped sheep (Kefis, Adane, Arabo, Segentu) and the short fat-tailed Molale/Menz sheep)), and “ET-G2” comprising long fat-tailed sheep (Bonga, Kido, Gesses, Loya, Shubi Gemo, Doyogena) and the short fat-tailed Gafera sheep)). The thin-tailed (Hammari and Kabashi) sheep from Sudan (SD) and the fat-tailed Barbary (LB) sheep from Libya clustered close but separate from each other and from Ethiopian sheep. The Barbary sheep comprise a more genetically diverse group than the other populations.

To determine the optimal clustering pattern generated by ADMIXTURE, we calculated and plotted the CV error for each K . The lowest CV error occurred at $K = 3$ (Figure S4a). The distribution of the three genetic backgrounds, named here A, B and C, among the 15 study populations is shown in Figure S4b. The “A” and “B” backgrounds predominated in ET-G1 and ET-G2 populations, respectively. The “C”

background occurred in SD and LB. Interestingly, the “A” and “B” backgrounds were also present in SD and LB but, the “C” background occurred at very low frequencies in a few individuals of ET-G1 and ET-G2. There was a clear divergence of Molale sheep and roughly 50 % of Gafera and Adane populations whose genomes are defined by the D background at $4 \leq K \leq 5$. Furthermore, at $K=5$, a fifth ancestry (E) is observed that separates the western populations from the remainder of the Ethiopian populations (Figure S4b).

The pattern of population splits revealed by TreeMix mirrors that of PCA (Figure 2a). The long fat-tailed (Gesses, Kido, Bonga, Doyogena, Loya, Shubi Gemo) and the short fat-tailed (Gafera) sheep formed one group. The fat-rumped (Adane, Arabo, Kefis, Segentu) and the short fat-tailed (Menz) sheep formed another group. The Libya Barbary (LB) and the Sudan thin-tailed (SD) sheep represented two separate lineages. TreeMix also provided evidence of gene-flow. The first seven migration vertices accounted for > 90% of the model significance, with the first migration edge accounting for the highest contribution. Migration vectors ≥ 8 resulted in small incremental changes in the value of the model significance. We therefore chose $m = 7$ as the best predictive value of the optimal migration model.

Four strong (migration weight > 0.35) migration events were revealed (Figure 2a). They showed gene flow from i) central and southern Ethiopian (Bonga, Doyogena, Loya, ShubiGemo (long fat-tailed)) sheep to eastern Ethiopian (Adane, Molale, Arabo, Kefis (fat-rumped)) sheep, ii) Bonga to the cluster of Gesses and Kido sheep in the West of the country, iii) Gafera to the cluster of Menz and Adane sheep in Central Ethiopia, and iv) Segentu to Adane. Three weak (migration weight < 0.2) migration events were also revealed; two involved gene flow from Menz and Segentu, respectively to Shubi Gemo, and from the cluster of Adane and Menz to Arabo.

318 The average values of r^2 reflecting genome-wide LD were lowest in ET-G2 and highest in LB (Figure
319 2b; Table S6). Irrespective of the genetic group, the trend in r^2 showed a rapid decline at the 10 kb to
320 100 kb distance interval. The trends in N_e for each genetic group are shown in Figure 2c and Table S7.
321 SD and LB showed the highest and lowest N_e , respectively at 1000 generations ago. ET-G1 showed
322 higher N_e than ET-G2 up to 300 generations ago following which the opposite is observed. Except LB
323 whose N_e declined gradually up to 50 generations ago, that of ET-G1, ET-G2 and SD increased
324 gradually up to 300 generations ago, and then declined rapidly up to 50 generations ago.

325

326 **Genome-wide scans for signatures of selection**

327 The PCA revealed four genetic groups (ET-G1, ET-G2, SD and LB), which were supported by TreeMix.
328 Because the north African SD and LB shared a prominent background (C), these four groups were
329 reduced to only three ancestries at the optimal cross validation error value of $K=3$ (Figure S4b). The
330 ET-G2 comprised populations from a high-altitude humid environment while LB and SD comprised
331 sheep from a coastal hot-humid arid environment in Libya and low altitude hot-dry desert environment
332 in Sudan, respectively (Table S1; S2). The ET-G1 included populations from diverse environments.

333

334 *Selection signatures for populations from high-altitude humid regions*

335 We used the ET-G2 group to investigate selection signatures for adaptation to high-altitude African
336 environments (Figure 3). When this group (ET-G2) was contrasted with LB and SD, the three analytical
337 approaches (H_p , F_{ST} , $XP-EHH$), either independently or in combination, identified 91 candidate regions
338 spanning 250 genes (Table S8). These genes were used for functional enrichment analysis, and we
339 considered the genes that comprised the three most significant (P -value ≤ 0.01) KEGG Pathways and
340 GO terms as the primary candidates driving high altitude adaptation (Table S9). The genes that have
341 been associated with high altitude environmental adaptation and were detected in our candidate selection
342 regions are summarised in Table 2. The top-two most significant GO biological process terms were

343 “response to hypoxia (GO:0001666)” and “positive regulation of angiogenesis (GO:0045766,)”. Genes
 344 associated with “response to hypoxia” were *EPASI*, *EGLN1*, *CHRNA4*, *VEGFA*, *NF1* and *ADSL*; the
 345 ones associated with “positive regulation of angiogenesis” were *VEGFA* and *FGF2*. The top-two most
 346 significant KEGG pathways were “tuberculosis (oas05152)” and “HIF-1 signaling (oas04066)”. Genes
 347 associated with the former were *IL18*, *PLK3*, *TLR1*, *CATHL3*, *CALML4*, *BAC5*, *TLR6* and *SC5*, and the
 348 ones associated with the latter were *TF*, *EGLN1*, *VEGFA* and *PLCG1* (Figure S5).

349
 350 We explored haplotype structures around *EGLN1* and *TF* genes (Figure 4a, and 4b) that have been
 351 reported in previous studies to be playing a significant role in high altitude adaptation. The analysis
 352 revealed the presence of a haplotype that was fixed in ET-G2 (frequency = 1.00), almost fixed in SD
 353 (frequency = 0.95) and segregating in LB (frequency = 0.625).

354
 355 *Selection signatures for low altitude hot-dry desert (Sudan), and coastal hot-humid arid (Libya)*
 356 *environments.*

357 We contrasted both LB and SD with ET-G2 to identify selection signatures for adaptation to low-altitude
 358 hot-dry desert and coastal hot-humid arid African environments (Figure 5). The three tests revealed 123
 359 candidate regions spanning 182 genes (Table S10). Two of these regions were identified by *Hp* and *F_{ST}*,
 360 six by *Hp* and *XP-EHH*, three by *XP-EHH* and *F_{ST}*, two by all the three methods and one by *XP-EHH*.
 361 The two regions identified by all the three methods occurred on OAR6 and OAR11, respectively. The
 362 region on OAR6 spanned no genes while the one on OAR11 was amongst the strongest and spanned
 363 four genes, *NF1*, *EVI2A*, *EVI2B* and *OMG* (Figure 5). Another strong candidate region that was
 364 supported by *Hp* and *F_{ST}* occurred on OAR2 and spanned *DIS3L2* (Figure 5). The haplotype structure
 365 analysis around *DIS3L2* showed a haplotype that was approaching fixation in LB and SD (Figure 6a).
 366 Another strong candidate region was found on OAR3 and overlapped *MSRB3* (Figure 5, 6b). The ET-
 367 G2 represented here by Doyogena, which exhibits vestigial ears, showed a large, conserved haplotype

368 around *MSRB3* compared to LB and SD groups, which are characterised by long drooping ears (Figure
369 6b). Several LB-specific candidate regions were also revealed; one occurred on OAR7 and spanned
370 *GLDN* and *CYP19* genes. Another was on OAR15 and spanned *PLEKHA7* and *C11orf58*. A distinct
371 haplotype occurred around *PLEKHA7* (Figure 6c). Two other regions spanning distinct haplotypes
372 around *PDS5B* and *RXFP2* genes were found 570 kb apart on OAR10 (Figure 6d). In total, 182 genes
373 found in 123 regions (Table S10) were identified and used for functional enrichment analysis yielding
374 14 highly significant GO terms and KEGG pathways (Table S11). The top-most highly significant
375 KEGG pathways included “Aldosterone synthesis and secretion (oas04925)” and “Vasopressin-
376 regulated water reabsorption (oas04962)”. The candidate genes associated with low altitude hot-dry
377 desert (Sudan), and coastal hot-humid arid (Libya) environments adaptation are summarised in Table 2.
378

379 *Signatures of selection for differences in tail length and fat-depot size*

380 To identify candidate regions associated with tail length, we took SD and ET-G2 to represent long-tailed
381 sheep and LB and ET-G1 to represent short-tailed sheep, based on the lengths of the individual CV's
382 and the complete tail skeleton. We analysed selection signatures with *Hp* (Figure 7), *F_{ST}* (Figure 8) and
383 *XP-EHH* (Figure 9, 10) and identified 68 candidate regions across 21 autosomes (Table S12). Two
384 candidate regions were identified by the three tests and 15 regions by at least two tests. The remaining
385 51 regions were identified by only one test; they included 29 identified by *F_{ST}* and 22 by *XP-EHH*. The
386 annotation of these candidate regions identified 218 genes (Table S12). The top-most tail-formation
387 related candidate genes are shown in Table 3. The first region identified by all the three tests occurred
388 on OAR3 and spanned six genes (*GAPVD1*, *HSPA5*, *RABEPK*, *PPP6C*, *SCAI*,
389 *ENSOARG00000025028*). The second occurred on OAR6 and spanned no genes. One strong candidate
390 region that overlapped eight genes (*CALCOCO2*, *TTLL6*, *HOXB13*, *RF02133*, *RF02132*, *MIR196A1*,
391 *HOXB9*, *HOXB7*) was identified by *F_{ST}* and *XP-EHH* on OAR11 (Table 3; Figure 8, 9 and 10). Of these
392 eight genes, *HOXB13* was identified as a strong candidate gene controlling tail length in mice

(Economides et al., 2003). An analysis of haplotype structures around this gene (Figure 11) revealed a conserved haplotype with a frequency of 0.843 and 0.75 in ET-G2 and LB, respectively. A second strong candidate region overlapping seven genes (*GOLGA1*, *ARPC5L*, *WDR38*, *OLFML2A*, *NR5A1*, *NR6A1*, *ADGRD2*) occurred on OAR3 (Table 3; Figure 8, 9 and 10). Two of the genes (*NR5A1* and *NR6A1*) have been associated with vertebrae number (Yang et al., 2009).

To identify selection signatures for tail fat-depot sizes, we contrasted the thin-tailed (SD) sheep with the fat-rumped (ET-G1) and the two fat-tailed groups (LB and ET-G2). The three approaches (*Hp*, *F_{ST}*, *XP-EHH*) identified 34 candidate regions overlapping 122 genes. Twelve regions were identified by one of the three approaches and 25 by a combination of at least two approaches (Table S13; Figure 12). Table 3 shows the candidate genes associated with fat deposition and which were present in the candidate regions identified by our analysis.

Discussion

Population structure and demographic history.

Human and natural selection have shaped the genomes of domestic livestock since domestication. Here we investigated the genome architecture of indigenous African sheep, by analysing whole-genome sequences of 150 indigenous sheep from three African countries, Ethiopia, Sudan and Libya. We compared the sequence reads to the OAR_v3.1.75 reference genome assembly and obtained 34.85 million high quality variants. An average of 93.62% of the SNPs were validated in the sheep dbSNP database of which 0.79% were exonic while the majority (55.9%) were intergenic. This is consistent with observations in Eurasian wild aurochs (Park et al., 2015) and Chinese native sheep (Yang et al., 2016), indicating high quality of our sequences, high reliability of the called SNPs and completeness of the gene regions.

418 Although the history of indigenous African sheep has been inferred from archeological studies, the
 419 picture still remains incomplete. Our phylogenetic inference identified four distinct genetic groups
 420 underlying the genetic structure of our study populations. These genetic groups are offering interesting
 421 insights on the historical diffusion and dispersal of sheep into and across the African continent. The
 422 North African Libyan Barbary (LB) and the North-East African Sudanese Hammari and Kabashi (SD)
 423 sheep clustered in close proximity but separate from each other and the East African sheep represented
 424 here by Ethiopian sheep; the latter separated into two genetic groups, ET-G1 and ET-G2. This result
 425 supports for the first time, based on whole genome sequence data, two distinct entry points of sheep in
 426 the continent, the Isthmus of Suez (LB and SD) and the Horn of Africa (ET-G1 and ET-G2). The sub-
 427 grouping of Ethiopian sheep into ET-G1 (fat-rumped) and ET-G2 (long fat-tailed) is indicative of
 428 different expansion events through the Horn of Africa. Based on the current dataset, it is difficult to
 429 infer whether ET-G1 and ET-G2 arrived together or independently. However, the geographic
 430 distribution of the Ethiopian populations in each group offers indirect insights. ET-G1 populations are
 431 found in areas close to the Red Sea and Indian Ocean coastline, while ET-G2 ones occur in the country's
 432 hinterland. The ancestral populations of these two groups may therefore have arrived independently,
 433 with the arrival of the long fat-tailed sheep preceding those of the fat-rumped.

434
 435 Three ancestral backgrounds were supported following ADMIXTURE analysis. All three are present in
 436 Sudanese sheep which also show the highest N_e (~2720) 1000 generations ago. The ancestor of domestic
 437 sheep was thin-tailed, but its tail was shorter than that of the domestic variant. Accordingly, as supported
 438 by archaeological thin-tailed iconography on the African continent, the thin-tailed sheep are likely to be
 439 the most ancient of the African sheep lineages (Muigai and Hanotte, 2013). The Sudanese and Libyan
 440 sheep shared a large proportion of their genome ancestry and show close genetic proximity on the PCA.
 441 They are likely the modern-day legacy of the ancient arrival of thin-tailed sheep on the African continent
 442 followed by their dispersion along the Mediterranean Sea coast (Libya) and southwards along the Nile

443 River Basin (Sudan). Although the ET-G2 and LB sheep are both fat-tailed, they do have different
444 ancestral backgrounds, supporting two entry points for African fat-tailed sheep on the continent i.e., the
445 Isthmus of Suez and the Horn of Africa.

446

447 The ADMIXTURE analysis also revealed that three populations, Doyogena, Shubi Gemo and Gafera,
448 of ET-G2 share ~15-20% of their genome ancestry with ET-G1. Likewise, we detect the main ancestry
449 of ET-G2 in three populations of ET-G1 (Menz, Adane, Arabo). This pattern of intermixing can be
450 attributed to flock movements/exchanges arising from human socio-cultural and economic activities
451 (Gizaw et al., 2007). This is supported by TreeMix which provided evidence of gene flow from fat-
452 rumped sheep from eastern Ethiopia into long fat-tailed sheep in central and southern Ethiopia. A similar
453 pattern of admixture was observed between six Ethiopian populations (Menz, Gafera, Gesses, Kido,
454 Adane and Arabo) and the Sudanese (SD) sheep. This admixture pattern is consistent with previous
455 findings (Ahbara et al., 2019). Half of the individuals of LB shared ~30% of their genome ancestry with
456 those of SD and ET-G1. We suggest this shared ancestry is the result of inter-crossing fat-tailed with
457 thin-tailed sheep to reduce the volume of tail fat in the former due to the decline in the monetary value
458 of animal fat (Moradi et al., 2012; Moiola et al., 2015). Such inter-crossings have been reported in
459 Algeria (Gaouar et al., 2017) and Tunisia (Rekik et al., 2005). It is worth noting that large fat tails, which
460 are discarded following slaughter, can account for up to 14.5% of the cold dressed carcass weight in
461 sheep (Kiyanzad, 2005). The shared ancestries between the north African SD and LB sheep in line with
462 the equal number of tail vertebrae (Figure 1a) and different tail sizes may support such pattern of
463 intermixing. On the other hand, various ancestries and varying numbers of tail vertebrae paired with
464 diverse tail sizes might imply a different history of Ethiopian sheep (Figure S4b).

465

466 *Genome targets for adaptation to environments*

467 The African landscape has pronounced changes in elevation ranging from sea level and rising to high
 468 altitudes in the East African highlands. The variation in altitude is accompanied by large variations in
 469 agro-climatic conditions ranging from dry and arid to wet and humid. Such variations in topography
 470 and environmental conditions allow studying adaptive processes in response to environmental selection
 471 pressure, e.g., hypobaric hypoxia in high altitudes, and heat, water and feed stress in the drylands. It
 472 provides excellent opportunities to investigate how natural selection shapes the genetic architecture of
 473 adaptive divergence. Here, we sampled 15 populations of indigenous African sheep from across the
 474 continents ecological and altitudinal range. We observed a separation of the populations into four
 475 genetic groups that align well with geography. Furthermore, our comparative genomic analyses
 476 involving the four genetic groups revealed several important observations shedding new insights on
 477 their biology of adaptation.

478
 479 Hypobaric hypoxia imposes major stresses on the physiology of organisms at high altitudes and a
 480 powerful homeostasis system has evolved to offset ambient hypoxia and attain physiological
 481 homeostasis (Rytönen and Storz, 2011). The classic response is an increase in haemoglobin
 482 concentration to compensate for the unavoidable lowered percent of oxygen saturation. Our selective
 483 sweep mapping involving ET-G2 identified candidate regions spanning nine genes, *TF* (OAR1), *EPAS1*
 484 and *ADSL* (OAR3), *NF1* (OAR11), *PLCG1* and *CHRNA4* (OAR13), *FGF2* (OAR17), *VEGFA* (OAR20)
 485 and *EGLN1* (OAR25). These genes comprised the two most significant GO terms, “response to
 486 hypoxia” and “positive regulation of angiogenesis”, and KEGG pathways, “HIF-1 signalling pathway”
 487 and “pathways in cancer”, suggesting they may underlie adaptive changes in high-altitude
 488 environments. Indeed, *EGLN1* and *EPAS1* encode the synthesis of prolyl hydroxylase domain 2 (*PHD2*)
 489 that plays a critical role in adapting to changes in oxygen levels and its variation has consistently been
 490 associated with haemoglobin levels in Tibetans (Simonson et al., 2010). It is also a target of natural
 491 selection for hypoxia adaptation in humans and animals (Storz and Moriyama, 2008; Tissot Van Patot

and Gassmann, 2011). The HIF-1 pathway plays a central role in regulating cellular responses to hypoxia (Frede and Fandrey, 2014) in several species found at high altitudes (see review by Friedrich and Wiener (2020)). The *NFI* gene occurred in a candidate region under selection in sheep living at high altitudes (Wei et al., 2016; Yang et al., 2016). *NFI* promotes and regulates angiogenesis in response to hypoxia on its own (Wu et al., 2006) or as a component of the Ras/ERK and VEGF signalling pathways (Minet et al., 2000). *VEGFA* has been associated with cerebrovascular adaptation to chronic hypoxia in sheep (Castillo-Melendez et al., 2015) and protects pregnant ewes from hypoxic effects by increasing blood flow in uterine arteries (Mehta et al., 2012). Under hypoxic conditions, elevated expression of *TF* and *VEGFA* increases oxygen delivery by stimulating iron metabolism and angiogenesis, following a positive feedback from *EGLN1*. Unexpectedly, our analytical strategy to identify genomic regions associated with differences in tail length and tail fat-depot size also identified selection targets relating to high altitude adaptation. For instance, the gene encoding methionine sulfoxide reductase B3 (*MSRB3*) occurred on a candidate region on OAR3 that was supported by *F_{ST}* and *XP-EHH*. This gene has been associated with high altitude adaptation in dogs (Gou et al., 2014) and Tibetan sheep (Wei et al., 2016). Our results appear to suggest that, driven by a hypoxic environment in the Ethiopian highlands, natural selection has been acting on a group of genes associated with the HIF-pathway and angiogenesis in driving adaptive processes in the ET-G2 group. Although we do find compelling evidence implicating a genetic predisposition to high altitude adaptation in ET-G2, previous studies in humans (Alkorta-Aranburu et al., 2012), cattle (Edea et al. 2014) and sheep (Edea et al., 2019) in Ethiopia did not find similar associations with the HIF-pathway and angiogenesis. These contrasting findings are rather surprising, but we speculate that they could be the result of differences in sampling strategies, analytical approaches and the markers used. For instance, the studies by Edea et al. (2014; 2019) and Wiener et al. (2021) classified their study populations based on geographic sampling; Wiener et al. also used an ecological approach to analyse their data. In contrast, we used the genetic groups revealed by the phylogenetic inference to inform our analysis.

517

518 Haplotype analysis around the candidate regions associated with high altitude adaptation returned some
 519 unexpected results. The haplotype structures around the *TF* locus showed a homogeneous haplotype
 520 approaching fixation in ET-G2, but several segregating haplotypes in SD and LB (Figure 4b), suggesting
 521 selection around the locus in ET-G2. Haplotype structure analysis around *EGLN1* showed SD shared
 522 identical haplotypes with ET-G2 that appeared to be approaching fixation, against an admixed haplotype
 523 in LB. However, LB and ET-G2 shared a similar selection sweep occurring around the *NF1* gene (Table
 524 2; Figure 3,5). In Egyptian sheep that are adapted to a dryland environment, Mwacharo et al. (2017)
 525 found candidate regions spanning genes that were highly enriched for the GO term “response to
 526 hypoxia” and “HIF-1 signaling pathway”. They suggested this could be an adaptive response to physical
 527 exhaustion due to long-term long-distance trekking in search of pasture and water resources, an activity
 528 that results in hypoxia-like conditions and oxygen debt in skeletal muscles. These results suggest that
 529 parallel evolution driven by distinct natural selection forces could be acting on the same set of genes
 530 driving adaptation to different environments in the same species. This is rather surprising and calls for
 531 further investigations.

532

533 High ambient temperatures resulting in thermal stress, present the physiological challenge to organisms
 534 residing in hot arid environments. Thermal stress impacts production and reproduction and desert
 535 dwelling animals have to invoke physiological responses to mitigate thermal stress and its negative
 536 effects. The analysis involving SD and LB sheep groups identified 123 candidate regions spanning 182
 537 genes (Table S10); 19 were reported in candidate regions revealed in Chinese sheep living in desert and
 538 arid environments (Yang et al., 2016). Among the genes identified in a candidate region on OAR11 was
 539 *KCNAB3* which encodes a potassium voltage-gated channel subunit that is correlated with cognitive
 540 performance under chronic stress (Jung et al., 2017) and is down-regulated in response to acute and
 541 chronic stress in mice (Terenina et al., 2019). This result is enlightening given that LB and SD sheep

are exposed to extended periods of thermal stress in Libya and Sudan, respectively. Apart from thermal stress, high temperatures constrain pasture availability both in quality and quantity. One of the candidate regions in which SD and LB shared a similar haplotype, occurred in OAR2 and spanned *DIS3L2* (Figure 6a). This gene was found to be a candidate under selection in cattle (Gautier et al., 2009) and Brazilian sheep and has been associated with height variation (de Simoni Gouveia et al., 2017). Given that large animals generally have higher maintenance requirements (Gomes et al., 2013), the occurrence of *DIS3L2* in a candidate region in LB and SD may act to regulate their physical size while providing the advantage of better thermoregulation (McManus et al., 2011).

Thermal stress affects male reproductive performance through decreased semen volume, sperm motility, higher sperm defects and reduced libido (see review by Sejian et al. (2019)). We found *CNTROB* (centrobin, for centrosome BRCA2 interacting protein) in a candidate region on OAR11. Through its role in spermiogenesis (Pandey et al., 2019), *CNTROB* is critical in maintaining male fertility (Liška et al., 2013). One region that differentiated LB from ET-G2 and SD and showed a distinct haplotype approaching fixation in LB (Figure 6d) presented on OAR10 and spanned *PDS5B* (PDS5 cohesin associated factor B). This gene modulates cohesin functions in spermatocytes and spermatogonia, contributing to meiotic chromosome function and structure (Fukuda and Hoog, 2010). By virtue of occurring in a selective sweep region in dryland dwelling sheep, *CNTROB* and *PDS5B* may be critical in maintaining male fertility and thus reproductive success under prolonged heat stress.

Soil salinisation contributes to soil degradation and compromises water and pasture quality (McFarlane et al., 2017). Droughts, which exacerbate soil salinity, have increased over the years due to global warming (Busby et al., 2010; Bindra et al., 2013) and worsened in arid and semi-arid environments (McFarlane et al., 2017). Adaptive responses to increased salinity and droughts in drylands may explain the selection signatures spanning *PLEKHA7* (Pleckstrin homology domain containing A7) in a region

on OAR15 that differentiates LB (Figure 6c). *PLEKHA7* is highly expressed in kidney and heart (Pulimeno et al., 2010); it was reported to attenuate salt-sensitive hypertension and renal diseases in rats (Endres et al., 2014). In humans, *PLEKHA7* has been associated with blood pressure, kidney function and hypertension (Lin et al., 2011; Taal et al., 2012). Two of the most significant KEGG pathways were “Aldosterone synthesis and secretion (GO:0030325)” and “Vasopressin-regulated water reabsorption (oas04962)”. Aldosterone and vasopressin regulate blood pressure and maintain electrolyte and water balance thus facilitating homeostatic adaptations (Harvey et al., 1984). The overexpression of *NF1*, which we found in one of the candidate regions under selection, results in reduced transcription of *HSD11B2* possibly associated with salt sensitivity (Alikhani-Koupaei et al., 2007) and hypertension (Friedman et al., 2002) in humans. *NR5A1* plays a critical role in the expression of *CYP11B2* (Takeda et al., 2018) which plays a key role in salt-sensitive hypertension development (Fujita, 2008) and in aldosterone biosynthesis in the adrenal cortex (Curnow et al., 1991) in humans.

579

580 *Selection targets for phenotypic traits*

We observed a candidate region on OAR11 that was specific to SD and ET-G2 groups. This region included several genes among which four, *TLL6*, *HOXB13*, *HOXB9* and *HOXB7* have been associated with growth and development of the spinal cord and tail vertebrae in several animal species (Burke and Nowicki, 2001; Economides et al., 2003; Mallo et al., 2010; Ahbara et al., 2019). This region occurred upstream of two candidate regions spanning five genes (*NF1*, *EVI2A*, *EVI2B*, *OMG* and *ALOX12*) (Figure 10, 12) that have been associated with fat deposition (Moioli et al., 2015; Wei et al., 2015; Yuan et al., 2017; Ahbara et al., 2019). The fact that we observed two regions or more occurring adjacent to one another and associated with two tail traits (fat-depot size and CV length) suggests a natural selection phenomenon influencing tail traits in sheep. One region on OAR3 spanned *NR5A1* and *NR6A1* which influence vertebrae number in pigs (Mikawa et al., 2007; Rubin et al., 2012) and thoracic vertebrae number in sheep (Shengwei et al., 2019). A candidate region spanning *PDGFRA* (platelet derived

growth factor receptor alpha) was also identified by the three methods on OAR6 in LB and a distinct haplotype was found around this gene. *PDGFRA* is involved in the differentiation of preadipocytes and was also found in a candidate selection region associated with the fat-tailed phenotype in sheep (Mastrangelo et al., 2019). These results suggest possible co-evolution between tail skeleton morphology and tail fat deposition in sheep.

Unexpectedly, our analytical strategy that aimed to identify candidate regions associated with differences in tail length and tail fat-depot sizes, identified selection targets relating to other traits. For instance, the gene encoding methionine sulfoxide reductase B3 (*MSRB3*) occurred on a candidate region on OAR3 supported by F_{ST} and $XP-EHH$. An association between *MSRB3* with ear types has been reported in dogs (Boyko et al., 2010; Vaysse et al., 2011), sheep (Wei et al., 2015) and pigs (Chen et al., 2018). An examination of haplotype structures around *MSRB3* showed a pattern that was consistent with the ear phenotypes of the three sheep groups analysed here. Compared to individuals of LB and SD, majority of which have long drooping ears, the ET-G2 represented by Doyogena, revealed a large, conserved haplotype around *MSRB3*. This result is consistent with the observation that polytypic species usually exhibit shorter extremities such as limbs and ears in cold environments while the ones inhabiting warmer climates show longer extremities (Allen, 1907), suggesting such differences are indicative of environmental adaptation.

LB sheep differentiated from ET-G2 at two candidate regions on OAR10. One of these spanned *RXFP2* (relaxin family peptide receptor 2) which determines horn size and development in ungulates (Kardos et al., 2015) and primary sex characters in humans and mice (Johnston et al., 2011). Heat-exchange capacity shows an evolutionary trend of variation in horn core size in response to ambient temperatures (Hoefs, 2000). Although the selective sweep spanning *RXFP2* could implicate its dual role in the

development of horns for thermoregulation and enhancing reproductive success under thermal stress, in recent times sheep farmers in Libya are preferring breeding rams with larger horns and tail sizes because such rams are believed to result in strong and healthier progenies (Figure S6). The selection signature around *RXFP2* may therefore be driven by human selection for horn size and natural selection for the thermoregulation function of horns, although the semi-feral nature of management of LB sheep could also be implicated as suggested for Chinese sheep (Pan et al., 2018).

Selection targets for resistance to endo- and ecto-parasites

High altitude hypoxia increases vulnerability to infections by decreasing the secretion of cytokines and immunoglobulins (Mishra and Ganju, 2010). In arid environments, feed, water and heat stress can also weaken immunity (Silanikove 2000). The candidate region on OAR11 that spanned four genes (*NFI*, *EVI2A*, *EVI2B*, *OMG*) associated with tail morphology, has also been linked to gastrointestinal nematode infection resistance in traditionally managed Tunisian sheep (Ahbara et al., 2021). The *PDGFRA* gene that occurred in a candidate region on OAR6, and which has been associated with the fat-tail phenotype (Mastrangelo et al., 2019) likely through the differentiation of preadipocytes, has also been implicated in host immune responses against endoparasites and is a key gene in cytokine signaling (Beh et al., 2002; Benavides et al., 2015). The *Hp* analysis detected a candidate region on OAR1 and OAR6 in ET-G1, ET-G2 and SD sheep. The region on OAR1 includes *CD84*, reported as a candidate gene for trypanosomiasis resistance in mice (Yan et al., 2004; Martin and Tarleton, 2005). The region on OAR6 includes *LETM1* which plays an essential role in the infection stages of *Trypanosoma brucei* by maintaining its mitochondrial integrity (Nowikovsky et al., 2012; Hashimi et al., 2013; Nowikovsky and Bernardi, 2014). These regions appear to be maintained in ET-G1, ET-G2 and SD tropical sheep which may have been exposed to endo-parasite and trypanosomosis challenge at the opposite of the North African LB sheep (Table 4). Further investigations are required to confirm the significance of these results in relation to resistance/tolerance to nematode infections and trypanosomiasis.

641

642 *Possible pleiotropic effects*

643 Taken together, it may be argued based on our results that the interaction between genes, whether on
644 the same or different chromosomes, driven by environmental conditions are responsible for adaptation
645 to diverse environments and the diversity in sheep tail phenotypes. For example, on OAR11 the
646 occurrence of *KDM6B* within a candidate region that is close to other regions associated with fat
647 accumulation may regulate the expression of *HOXB13* (Figure 10).

648

649 **Conclusion**

650 Our results revealed four autosomal genomic backgrounds in indigenous East, North and North-East
651 African sheep. The genotypes of most of the individuals analyzed here comprised at least two genetic
652 backgrounds which could be accounted for by some level of current and/or historical intermixing. Our
653 findings corroborate archaeological evidence that have revealed domestic sheep migration events and
654 entry points into Africa. Besides, our results support possible interaction between environmental factors
655 and tail morphology in shaping the genomes of indigenous sheep inhabiting diverse environments. From
656 the perspective of environmental adaptation, our study populations illustrated the genome structure
657 towards environmental challenges (coastal hot-humid arid, hot-dry desert and high-altitude humid). Our
658 whole-genome sequence analysis also suggest interaction between several genomic regions for tail
659 phenotype in sheep. For instance, many candidate genes possibly under selection were identified, among
660 which, the salt-sensitivity related gene (*PLEKHA7*) in Libyan sheep, thermoregulation genes
661 (*CREB3L2*, *CREB3*, *GNAQ*, *DCTN4*) in Libyan and Sudanese sheep, hypoxia associated genes
662 (*EGLN1*, *EPAS1*) in Ethiopian sheep, and skeletal tail associated genes (*HOXB13*, *NR5A1*, *NR6A1*,
663 *KDM6B* and *NR3C1*) most likely are playing major roles in the unique adaptations of African sheep. To
664 the best of our knowledge, we are the first to report these candidate genes in African indigenous sheep.
665 However, future studies including thin-tailed sheep from Ethiopia and both thin- and/or fat-tailed sheep

666 from West, northern and southern Africa could provide a comprehensive view on the genome
667 architecture of African indigenous sheep.

668

669 **Data Accessibility**

670 The datasets analyzed in this study are available upon request from the corresponding authors.

671

672 **Ethics statement**

673 The animals used in this study are owned by farmers. Prior to sampling, the objectives of the study were
674 explained to them in their local languages for them to make informed decisions regarding giving consent
675 to sample their animals. Government veterinary, animal welfare and health regulations in Ethiopia,
676 Libya and Sudan were observed during sampling. Skin tissues importation and/or exportation was
677 permitted by the Ethiopian Ministry of Livestock and Fisheries under Certificate No: 14-160-401-16.

678

679 **Conflict of Interest**

680 The authors declare no conflicts of interest.

681

682 **Author Contributions**

683 AMA, OH, and JMM conceived and designed the study. AMA analyzed the data with inputs from JMM
684 and wrote the manuscript. OH and JMM revised the manuscript. EC, CR, and PW contributed to
685 sequencing and mapping data of Ethiopian and Libyan sheep breeds and provided critical inputs on data
686 analysis. HM contributed to sequencing samples of Sudanese sheep. AA, SL and MOA contributed and
687 coordinated the sampling in Ethiopia and Libya, respectively. All authors read and approved the final
688 manuscript.

689

690 **Acknowledgments**

This study was conducted during AMA's PhD training which was supported by the Libyan Ministry of Higher Education and Scientific Research and the University of Misurata. Sampling of Ethiopian sheep was supported by the CGIAR Research Program on Livestock. Accordingly, ICARDA and ILRI wish to thank the donors and organizations that globally support the work of the CGIAR Research Program on Livestock through their contributions to the CGIAR Trust Fund. Sampling of Libyan sheep was supported by the Libyan Agriculture Research Centre (Misurata station) and the University of Misurata (Department of Animal Production).

References

- Ahbara, A., Bahbahani, H., Almathen, F., Abri, M. al, Agoub, M. O., Abeba, A., et al. (2019). Genome-wide variation, candidate regions and genes associated with fat deposition and tail morphology in Ethiopian indigenous sheep. *Front. Genet.* 9, 699. doi:10.3389/fgene.2018.00699.
- Ahbara, A. M., Rouatbi, M., Gharbi, M., Rekik, M., Haile, A., Rischkowsky, B., et al. (2021). Genome-wide insights on gastrointestinal nematode resistance in autochthonous Tunisian sheep. *Sci. Rep.* 11, 9250. doi:10.1038/s41598-021-88501-3.
- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi:10.1101/gr.094052.109.
- Alikhani-Koupaei, R., Fouladkou, F., Fustier, P., Cenni, B., Sharma, A. M., Deter, H., et al. (2007). Identification of polymorphisms in the human 11beta-hydroxysteroid dehydrogenase type 2 gene promoter: functional characterization and relevance for salt sensitivity. *The FASEB Journal* 21, 3618–3628. doi:10.1096/fj.07-8140com.
- Alkorta-Aranburu, G., Beall, C. M., Witonsky, D. B., Gebremedhin, A., Pritchard, J. K., and di Rienzo, A. (2012). The Genetic Architecture of Adaptations to High Altitude in Ethiopia. *PLoS Genet.* 8, e1003110. doi:10.1371/journal.pgen.1003110.
- Allen, J. A. (1907). The Influence of Physical Conditions in the Genesis of Species. *Sci. Am.* 63. doi:10.1038/scientificamerican05251907-26247supp.
- Barbato, M., Orozco-terWengel, P., Tapio, M., and Bruford, M. W. (2015). SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front. Genet.* 6, 109. doi:10.3389/fgene.2015.00109.
- Barthelme, J. W. (1985). *Fisher-hunters and neolithic pastoralists in East Turkana, Kenya*. Oxford: BAR.

- 722 Beh, K. J., Hulme, D. J., Callaghan, M. J., Leish, Z., Lenane, I., Windon, R. G., et al. (2002). A genome
723 scan for quantitative trait loci affecting resistance to *Trichostrongylus colubriformis* in sheep.
724 *Anim. Genet.* 33, 97–106. doi:10.1046/j.1365-2052.2002.00829.x.
- 725 Benavides, M. V., Sonstegard, T. S., Kemp, S., Mugambi, J. M., Gibson, J. P., Baker, R. L., et al. (2015).
726 Identification of novel loci associated with gastrointestinal parasite resistance in a red Maasai x
727 Dorper backcross population. *PLoS ONE* 10, 0122797. doi:10.1371/journal.pone.0122797.
- 728 Bindra, S. P., Abulifa, S., Hamid, A., Reiani, H. S. al, and Abdalla, K. (2013). Assessment of Impacts
729 on Ground Water Resources in Libya and Vulnerability To Climate Change. *Scientific Bulletin of*
730 *the „Petru Maior” University of Tîrgu Mureş* 10, 63–69.
- 731 Boyko, A. R., Quignon, P., Li, L., Schoenebeck, J. J., Degenhardt, J. D., Lohmueller, K. E., et al. (2010).
732 A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol.* 8, e1000451.
733 doi:10.1371/journal.pbio.1000451.
- 734 Browning, S. R., and Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data
735 inference for whole-genome association studies by use of localized haplotype clustering. *Am. J.*
736 *Hum. Genet.* 81, 1084–1097. doi:10.1086/521987.
- 737 Buchmann, R., and Hazelhurst, S. (2014). Genesis manual. *bioinf.wits.ac.za*. Available at:
738 <http://www.bioinf.wits.ac.za/software/genesis/Genesis.pdf> [Accessed August 15, 2021].
- 739 Burke, A. C., and Nowicki, J. L. (2001). Hox genes and axial specification in vertebrates. *Am. Zool.* 41,
740 687–697. doi:10.1093/icb/41.3.687.
- 741 Busby, J., White, K., and Smith, T. G. (2010). *Mapping climate change and security in North Africa*.
742 The German Marshall Fund of the United States: Climate & Energy Policy Paper Series.
- 743 Casas, E., Shackelford, S. D., Keele, J. W., Stone, R. T., Kappes, S. M., and Koohmaraie, M. (2000).
744 Quantitative trait loci affecting growth and carcass composition of cattle segregating alternate
745 forms of myostatin. *J Anim. Sci.* 78, 560–569. doi:10.2527/2000.783560x.
- 746 Castillo-Melendez, M., Yawno, T., Allison, B. J., Jenkin, G., Wallace, E. M., and Miller, S. L. (2015).
747 Cerebrovascular adaptations to chronic hypoxia in the growth restricted lamb. *Int. J. Dev. Neurosci.*
748 45, 55–65. doi:10.1016/j.ijdevneu.2015.01.004.
- 749 Chen, C., Liu, C., Xiong, X., Fang, S., Yang, H., Zhang, Z., et al. (2018). Copy number variation in the
750 MSRB3 gene enlarges porcine ear size through a mechanism involving miR-584-5p. *Genet. Sel.*
751 *Evol.* 50, 1–18. doi:10.1186/s12711-018-0442-6.
- 752 Clark, J. D., and Williams, M. A. J. (1978). Recent Archaeological Research in Southeastern Ethiopia.
753 1974 - 1975. *Annales d’Ethiopie* 11, 19–44. doi:10.3406/ethio.1978.902.
- 754 Clutton-Brock, J. (1993). “The spread of domestic animals in Africa,” in *The archaeology of Africa:*
755 *Foods, metals, and towns*, eds. T. Shaw, P. Sinclair, B. Andah, and A. Okpoko (London, UK:
756 Routledge).
- 757 Curnow, K., Tusie-Luna, M., ... L. P.-M., and 1991, U. (1991). The product of the CYP11B2 gene is
758 required for aldosterone biosynthesis in the human adrenal cortex. *Mol. Endocrinol.* 5, 1513–1522.

759 Available at: <https://academic.oup.com/mend/article-abstract/5/10/1513/2714337> [Accessed
760 October 23, 2021].

761 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The
762 variant call format and VCFtools. *Bioinformatics* 27, 2156–2158.
763 doi:10.1093/bioinformatics/btr330.

764 de Simoni Gouveia, J. J., Paiva, S. R., McManus, C. M., Caetano, A. R., Kijas, J. W., Facó, O., et al.
765 (2017). Genome-wide search for signatures of selection in three major Brazilian locally adapted
766 sheep breeds. *Livest. Sci.* 197, 36–45. doi:10.1016/j.livsci.2017.01.006.

767 Economides, K. D., Zeltser, L., and Capecchi, M. R. (2003). Hoxb13 mutations cause overgrowth of
768 caudal spinal cord and tail vertebrae. *Dev. Biol.* 256, 317–330. doi:10.1016/S0012-1606(02)00137-
769 9.

770 Edea, Z., Dadi, H., Dessie, T., and Kim, K. S. (2019). Genomic signatures of high-altitude adaptation
771 in Ethiopian sheep populations. *Genes Genom.* 41, 973–981. doi:10.1007/s13258-019-00820-y.

772 Edea, Z., Dadi, H., Kim, S. W., Park, J. H., Shin, G. H., Dessie, T., et al. (2014). Linkage disequilibrium
773 and genomic scan to detect selective loci in cattle populations adapted to different ecological
774 conditions in Ethiopia. *J. Anim. Breed. Genet.* 131, 358–366. doi:10.1111/jbg.12083.

775 Endres, B. T., Priestley, J. R. C., Palygin, O., Flister, M. J., Hoffman, M. J., Weinberg, B. D., et al.
776 (2014). Mutation of *Plekha7* attenuates salt-sensitive hypertension in the rat. *Proc. Natl. Acad. Sci.*
777 *USA* 111, 12817–12822. doi:10.1073/pnas.1410745111.

778 Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M., and Servin, B. (2013). Detecting signatures of
779 selection through haplotype differentiation among hierarchically structured populations. *Genetics*
780 193, 929–941. doi:10.1534/genetics.112.147231.

781 Frede, S., and Fandrey, J. (2014). “Cellular and Molecular Defenses Against Hypoxia,” in *High Altitude*
782 (New York, NY: Springer New York). doi:10.1007/978-1-4614-8772-2_2.

783 Friedman, J. M., Arbiter, J., Epstein, J. A., Gutmann, D. H., Huot, S. J., Lin, A. E., et al. (2002).
784 Cardiovascular disease in neurofibromatosis 1: Report of the NF1 Cardiovascular Task Force.
785 *Genet. Med.* 4, 105–111. doi:10.1097/00125817-200205000-00002.

786 Friedrich, J., and Wiener, P. (2020). Selection signatures for high-altitude adaptation in ruminants.
787 *Anim. Genet.* 51, 157–165. doi:10.1111/age.12900.

788 Fujita, T. (2008). Aldosterone in salt-sensitive hypertension and metabolic syndrome. *J. Mol. Med.* 86,
789 729–734. doi:10.1007/s00109-008-0343-1.

790 Fukuda, T., and Hoog, C. (2010). The mouse cohesin-associated protein PDS5B is expressed in
791 testicular cells and is associated with the meiotic chromosome axes. *Genes* 1, 484–494.
792 doi:10.3390/genes1030484.

793 Gaouar, S. B. S., Lafri, M., Djaout, A., El-Bouyahiaoui, R., Bouri, A., Bouchatal, A., et al. (2017).
794 Genome-wide analysis highlights genetic dilution in Algerian sheep. *Heredity* 118, 293–301.
795 doi:10.1038/hdy.2016.86.

- 796 Gautier, A. (2006). “The Evidence for the Earliest Livestock in North Africa: or Adventures with Large
797 Bovids, Ovicaprids, Dogs and Pigs,” in *Droughts, Food and Culture*, ed. H. Fekri (London:
798 KLUWER ACADEMIC PUBLISHERS), 195–207. doi:10.1007/0-306-47547-2_12.
- 799 Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloé, D., Gut, I., et al. (2009). A whole genome
800 Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics* 10, 1–18.
801 doi:10.1186/1471-2164-10-550.
- 802 Gheayas, A. A., Vallejo-Trujillo, A., Kebede, A., Lozano-Jaramillo, M., Dessie, T., Smith, J., et al.
803 (2021). Integrated Environmental and Genomic Analysis Reveals the Drivers of Local Adaptation
804 in African Indigenous Chickens. *Mol. Biol. Evol.* 38, 4268–4285. doi:10.1093/molbev/msab156.
- 805 Gifford-Gonzalez, D., and Hanotte, O. (2011). Domesticating Animals in Africa: Implications of
806 Genetic and Archaeological Findings. *J. World Prehist.* 24, 1–23. doi:10.1007/s10963-010-9042-
807 2.
- 808 Gizaw, S., van Arendonk, J. A. M., Komen, H., Windig, J. J., and Hanotte, O. (2007). Population
809 structure, genetic variation and morphological diversity in indigenous sheep of Ethiopia. *Anim.*
810 *Genet.* 38, 621–628. doi:10.1111/j.1365-2052.2007.01659.x.
- 811 Gomes, R. da C., Sainz, R. D., and Leme, P. R. (2013). Protein metabolism, feed energy partitioning,
812 behavior patterns and plasma cortisol in Nellore steers with high and low residual feed intake. *RBZ*
813 42, 44–50. doi:10.1590/S1516-35982013000100007.
- 814 Gou, X., Wang, Z., Li, N., Qiu, F., Xu, Z., Yan, D., et al. (2014). Whole-genome sequencing of six dog
815 breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. *Genome Res.* 24,
816 1308–1315. doi:10.1101/gr.171876.113.
- 817 Harvey, S., Phillips, J. G., Rees, A., and Hall, T. R. (1984). Stress and adrenal function. *J. Exp. Zool.*
818 232, 633–645. doi:10.1002/jez.1402320332.
- 819 Hashimi, H., McDonald, L., Stříbrná, E., and Lukeš, J. (2013). Trypanosome letm1 protein is essential
820 for mitochondrial potassium homeostasis. *J. Biol. Chem.* 288, 26914 –26925.
821 doi:10.1074/jbc.M113.495119.
- 822 Hayes, B. J., Visscher, P. M., McPartlan, H. C., and Goddard, M. E. (2003). Novel multilocus measure
823 of linkage disequilibrium to estimate past effective population size. *Genome Res.* 13, 635–643.
824 doi:10.1101/gr.387103.
- 825 Hoefs, M. (2000). The thermoregulatory potential of Ovis horn cores. *Can. J. Zool.* 78, 1419–1426.
826 doi:10.1139/cjz-78-8-1419.
- 827 Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large
828 gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57.
829 doi:10.1038/nprot.2008.211.
- 830 Johnston, S. E., McEwan, J. C., Pickering, N. K., Kijas, J. W., Beraldi, D., Pilkington, J. G., et al. (2011).
831 Genome-wide association mapping identifies the genetic basis of discrete and quantitative
832 variation in sexual weaponry in a wild sheep population. *Mol. Ecol.* 20, 2555–2566.
833 doi:10.1111/j.1365-294X.2011.05076.x.

834 Jung, S. H., Brownlow, M. L., Pellegrini, M., and Jankord, R. (2017). Divergence in morris water maze-
835 based cognitive performance under chronic stress is associated with the hippocampal whole
836 transcriptomic modification in mice. *Front. Mol. Neurosci.* 10, 275.
837 doi:10.3389/fnmol.2017.00275.

838 Kardos, M., Luikart, G., Bunch, R., Dewey, S., Edwards, W., McWilliam, S., et al. (2015). Whole-
839 genome resequencing uncovers molecular signatures of natural and sexual selection in wild
840 bighorn sheep. *Mol. Ecol.* 24, 5616–5632. doi:10.1111/mec.13415.

841 Kim, K., Kwon, T., Dessie, T., Yoo, D. A., Mwai, O. A., Jang, J., et al. (2020). The mosaic genome of
842 indigenous African cattle as a unique genetic resource for African pastoralism. *Nat. Genet.* 52,
843 1099–1110. doi:10.1038/s41588-020-0694-2.

844 Kiyanzad, M. R. (2005). Comparison of carcass composition of Iranian fat-tailed sheep. *Asian-Aust J*
845 *Anim Sci.* 18, 1348–1352. doi:10.5713/ajas.2005.1348.

846 Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform.
847 *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324.

848 Lin, Y., Lai, X., Chen, B., Xu, Y., Huang, B., Chen, Z., et al. (2011). Genetic variations in CYP17A1,
849 CACNB2 and PLEKHA7 are associated with blood pressure and/or hypertension in she ethnic
850 minority of China. *Atherosclerosis* 219, 709–714. doi:10.1016/j.atherosclerosis.2011.09.006.

851 Liška, F., Gosele, C., Popova, E., Chyliková, B., Křenová, D., Křen, V., et al. (2013). Overexpression
852 of Full-Length Centrobin Rescues Limb Malformation but Not Male Fertility of the Hypodactylous
853 (hd) Rats. *PLoS ONE* 8, p.e60859. doi:10.1371/journal.pone.0060859.

854 MacDonald, K. C., and MacDonald, R. H. (2000). “The origins and development of domesticated
855 animals in arid West Africa,” in *The origins and development of African livestock: Archaeology,*
856 *genetics, linguistics and ethnography*, eds. K. C. MacDonald and R. M. Blench (New York:
857 Routledge), 127–162.

858 Maclean, C. A., Chue Hong, N. P., and Prendergast, J. G. D. (2015). Hapbin: An efficient program for
859 performing haplotype-based scans for positive selection in large genomic datasets. *Mol. Biol. Evol.*
860 32, 3027–3029. doi:10.1093/molbev/msv172.

861 Mallo, M., Wellik, D. M., and Deschamps, J. (2010). Hox genes and regional patterning of the vertebrate
862 body plan. *Dev. Biol.* 344, 7–15. doi:10.1016/j.ydbio.2010.04.024.

863 Marshall, F. (2000). “The origins and spread of domestic animals in East Africa,” in *The Origins and*
864 *Development of African Livestock: Archaeology, Genetics, Linguistics and Ethnography*, eds. R.
865 M. Blench and K. C. MacDonald (London, UK: University College London Press).

866 Martin, D. L., and Tarleton, R. L. (2005). Antigen-Specific T Cells Maintain an Effector Memory
867 Phenotype during Persistent Trypanosoma cruzi Infection. *J. Immunol.* 174, 1594–1601.
868 doi:10.4049/jimmunol.174.3.1594.

869 Mastrangelo, S., Moiola, B., Ahbara, A., Latairish, S., Portolano, B., Pilla, F., et al. (2019). Genome-
870 wide scan of fat-tail sheep identifies signals of selection for fat deposition and adaptation. *Anim.*
871 *Prod. Sci.* 59, 835–848. doi:10.1071/AN17753.

- 872 McClure, M. C., Morsci, N. S., Schnabel, R. D., Kim, J. W., Yao, P., Rolf, M. M., et al. (2010). A
873 genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits
874 in commercial Angus cattle. *Anim. Genet.* 41, 597–607. doi:10.1111/j.1365-2052.2010.02063.x.
- 875 McFarlane, D. J., George, R. J., Barrett-Lennard, E. G., and Gilfedder, M. (2017). “Salinity in dryland
876 agricultural Systems: Challenges and opportunities,” in *Innovations in Dryland Agriculture*, eds.
877 M. Farooq and K. Siddique (Springer, Cham.). doi:10.1007/978-3-319-47928-6_19.
- 878 Mcmanus, C., Louvandini, H., do Prado Paim, T., Saraiva Martins, R., Jardim Barcellos, J. O., Cardoso,
879 C., et al. (2011). The challenge of sheep farming in the tropics: aspects related to heat tolerance.
880 *RBZ* 40, 107–120.
- 881 McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., et
882 al. (2008). Runs of Homozygosity in European Populations. *Am. J. Hum. Genet.* 83, 359–372.
883 doi:10.1016/j.ajhg.2008.08.007.
- 884 Mehta, V., Abi-Nader, K. N., Peebles, D. M., Benjamin, E., Wigley, V., Torondel, B., et al. (2012).
885 Long-term increase in uterine blood flow is achieved by local overexpression of VEGF-A 165 in
886 the uterine arteries of pregnant sheep. *Gene Ther.* 19, 925–935. doi:10.1038/gt.2011.158.
- 887 Mikawa, S., Morozumi, T., Shimanuki, S. I., Hayashi, T., Uenishi, H., Domukai, M., et al. (2007). Fine
888 mapping of a swine quantitative trait locus for number of vertebrae and analysis of an orphan
889 nuclear receptor, germ cell nuclear factor (NR6A1). *Genome Res.* 17, 586–593.
890 doi:10.1101/gr.6085507.
- 891 Minet, E., Arnould, T., Michel, G., Roland, I., Mottet, D., Raes, M., et al. (2000). ERK activation upon
892 hypoxia: Involvement in HIF-1 activation. *FEBS Lett.* 468, 53–58. doi:10.1016/S0014-
893 5793(00)01181-9.
- 894 Mishra, K. P., and Ganju, L. (2010). Influence of high altitude exposure on the immune system: A
895 review. *Immunol. Invest.* 39, 219–234. doi:10.3109/08820131003681144.
- 896 Moioli, B., Pilla, F., and Ciani, E. (2015). Signatures of selection identify loci associated with fat tail in
897 sheep. *J. Anim. Sci.* 93, 4660–4669. doi:10.2527/jas.2015-9389.
- 898 Moradi, M. H., Nejati-Javaremi, A., Moradi-Shahrababak, M., Dodds, K. G., and McEwan, J. C. (2012).
899 Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate
900 regions associated with fat deposition. *BMC Genet.* 13, 1–15. doi:10.1186/1471-2156-13-10.
- 901 Muigai, A. W. T., and Hanotte, O. (2013). The Origin of African Sheep: Archaeological and Genetic
902 Perspectives. *Afr. Archaeol. Rev.* 30, 39–50. doi:10.1007/s10437-013-9129-0.
- 903 Mwacharo, J. M., Kim, E. S., Elbeltagy, A. R., Aboul-Naga, A. M., Rischkowsky, B. A., and Rothschild,
904 M. F. (2017). Genomic footprints of dryland stress adaptation in Egyptian fat-Tail sheep and their
905 divergence from East African and western Asia cohorts. *Sci. Rep.* 7, 1–10. doi:10.1038/s41598-
906 017-17775-3.
- 907 Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., and Durbin, R. (2016).
908 BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation
909 sequencing data. *Bioinformatics* 32, 1749–1751. doi:10.1093/bioinformatics/btw044.

910 Nowikovsky, K., and Bernardi, P. (2014). LETM1 in mitochondrial cation transport. *Front. Physiol.* 5,
911 83. doi:10.3389/fphys.2014.00083.

912 Nowikovsky, K., Pozzan, T., Rizzuto, R., Scorrano, L., and Bernardi, P. (2012). Perspectives on: SGP
913 symposium on mitochondrial physiology and medicine: the pathophysiology of LETM1. *J. Gen.*
914 *Physiol.* 139, 445–454. doi:10.1085/jgp.201110757.

915 Pan, Z., Li, S., Liu, Q., Wang, Z., Zhou, Z., Di, R., et al. (2018). Whole-genome sequences of 89 Chinese
916 sheep suggest role of RXFP2 in the development of unique horn phenotype as response to semi-
917 feralization. *Gigascience* 7, giy019. doi:10.1093/gigascience/giy019.

918 Pandey, A., Yadav, S. K., Vishvkarma, R., Singh, B., Maikhuri, J. P., Rajender, S., et al. (2019). The
919 dynamics of gene expression during and post meiosis sets the sperm agenda. *Mol. Reprod. Dev.*
920 86, 1921–1939. doi:10.1002/mrd.23278.

921 Park, S. D. E., Magee, D. A., McGettigan, P. A., Teasdale, M. D., Edwards, C. J., Lohan, A. J., et al.
922 (2015). Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the
923 phylogeography and evolution of cattle. *Genome Biol.* 16, 1–15. doi:10.1186/s13059-015-0790-2.

924 Pickrell, J. K., and Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-
925 Wide Allele Frequency Data. *Nat. Prec.* 8, 1–17. doi:10.1371/journal.pgen.1002967.

926 Porter, V. (2020). *Mason's World Dictionary of Livestock Breeds, Types and Varieties*. . 6th Edition.
927 Wallingford, UK: CABI.

928 Pulimeno, P., Bauer, C., Stutz, J., and Citi, S. (2010). PLEKHA7 is an adherens junction protein with a
929 tissue distribution and subcellular localization distinct from ZO-1 and E-Cadherin. *PLoS ONE* 5,
930 e12207. doi:10.1371/journal.pone.0012207.

931 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007).
932 PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J.*
933 *Hum. Genet.* 81, 559–575. doi:10.1086/519795.

934 Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic
935 features. *Bioinformatics* 26, 841–842. doi:10.1093/bioinformatics/btq033.

936 Rekik, M., Aloulou, R., and ben Hammouda, M. (2005). “Small Ruminant Breeds of Tunisia,” in
937 *Characterisation of Small Ruminant Breeds in West Asia and North Africa*, ed. L. Iniguez (Beirut:
938 International Center for Agricultural Research in the Dry Areas (ICARDA)), 91–140.

939 Rubin, C. J., Megens, H. J., Barrio, A. M., Maqbool, K., Sayyab, S., Schwochow, D., et al. (2012).
940 Strong signatures of selection in the domestic pig genome. *Proc. Natl Acad. Sci. USA* 109, 19529–
941 19536. doi:10.1073/pnas.1217149109.

942 Rubin, C. J., Zody, M. C., Eriksson, J., Meadows, J. R. S., Sherwood, E., Webster, M. T., et al. (2010).
943 Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*
944 464, 587–591. doi:10.1038/nature08832.

945 Rytkönen, K. T., and Storz, J. F. (2011). Evolutionary origins of oxygen sensing in animals. *EMBO*
946 *Rep.* 12, 3–4. doi:10.1038/embor.2010.192.

- 947 Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., et al. (2007). Genome-wide
948 detection and characterization of positive selection in human populations. *Nature* 449, 913–918.
949 doi:10.1038/nature06250.
- 950 Sejian, V., Bagath, M., Krishnan, G., Rashamol, V. P., Pragna, P., Devaraj, C., et al. (2019). Genes for
951 resilience to heat stress in small ruminants: A review. *Small Rumin. Res.* 173, 42–53.
952 doi:10.1016/j.smallrumres.2019.02.009.
- 953 Shengwei, H., Cunyuan, L., Ming, L., Xiaoyue, L., Wei, N., Yueren, X., et al. (2019). Whole-genome
954 resequencing reveals loci associated with thoracic vertebrae number in sheep. *Front. Genet.* 10,
955 674. doi:10.3389/fgene.2019.00674.
- 956 Simonson, T. S., Yang, Y., Huff, C. D., Yun, H., Qin, G., Witherspoon, D. J., et al. (2010). Genetic
957 evidence for high-altitude adaptation in Tibet. *Science* 329, 72–75. doi:10.1126/science.1189406.
- 958 Storz, J. F., and Moriyama, H. (2008). Mechanisms of hemoglobin adaptation to high altitude hypoxia.
959 *High Alt. Med. Biol.* 9, 148–157. doi:10.1089/ham.2007.1079.
- 960 Sved, J. A. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite
961 populations. *Theor. Popul. Biol.* 2, 125–141. doi:10.1016/0040-5809(71)90011-6.
- 962 Taal, H. R., van den Hil, L. C. L., Hofman, A., van der Heijden, A. J., and Jaddoe, V. W. V. (2012).
963 Genetic variants associated with adult blood pressure and kidney function do not affect fetal kidney
964 volume. The Generation R Study. *Early Hum. Dev.* 88, 711–716.
965 doi:10.1016/j.earlhumdev.2012.02.014.
- 966 Takeda, Y., Demura, M., Wang, F., Karashima, S., Yoneda, T., Kometani, M., et al. (2018). Epigenetic
967 regulation of aldosterone synthase gene by sodium and angiotensin II. *J. Am. Heart Assoc.* 7,
968 p.e008281. doi:10.1161/JAHA.117.008281.
- 969 Terenina, E. E., Cavigelli, S., Mormede, P., Zhao, W., Parks, C., Lu, L., et al. (2019). Genetic factors
970 mediate the impact of chronic stress and subsequent response to novel acute stress. *Front. Neurosci.*
971 13, 438. doi:10.3389/fnins.2019.00438.
- 972 Tissot Van Patot, M. C., and Gassmann, M. (2011). Hypoxia: Adapting to high altitude by mutating
973 EPAS-1, the gene encoding HIF-2 α . *High Alt. Med. Biol.* 12, 157–167.
974 doi:10.1089/ham.2010.1099.
- 975 Vaysse, A., Ratnakumar, A., Derrien, T., Axelsson, E., Pielberg, G. R., Sigurdsson, S., et al. (2011).
976 Identification of genomic regions associated with phenotypic variation between dog breeds using
977 selection mapping. *PLoS Genet.* 7, e1002316. doi:10.1371/journal.pgen.1002316.
- 978 Wei, C., Wang, H., Liu, G., Wu, M., Cao, J., Liu, Z., et al. (2015). Genome-wide analysis reveals
979 population structure and selection in Chinese indigenous sheep breeds. *BMC Genomics* 16, 1–12.
980 doi:10.1186/s12864-015-1384-9.
- 981 Wei, C., Wang, H., Liu, G., Zhao, F., Kijas, J. W., Ma, Y., et al. (2016). Genome-wide analysis reveals
982 adaptation to high altitudes in Tibetan sheep. *Sci. Rep.* 6, 1–11. doi:10.1038/srep26770.
- 983 Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population
984 structure. *Evolution* 38, 1358–1370. doi:10.1111/j.1558-5646.1984.tb05657.x.

985 Wiener, P., Robert, C., Ahbara, A., Salavati, M., Abebe, A., Kebede, A., et al. (2021). Whole-Genome
986 Sequence Data Suggest Environmental Adaptation of Ethiopian Sheep Populations. *GBE* 13,
987 evab014. doi:10.1093/gbe/evab014.

988 Wu, M., Wallace, R., and Muir, D. (2006). Nf1 haploinsufficiency augments angiogenesis. *Oncogene*
989 25, 2297–2303. doi:10.1038/sj.onc.1209264.

990 Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: A web server for
991 annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39, W316–
992 W322. doi:10.1093/nar/gkr483.

993 Yan, Y., Wang, M., Lemon, W. J., and You, M. (2004). Single nucleotide polymorphism (SNP) analysis
994 of mouse quantitative trait loci for identification of candidate genes. *J. Med. Genet.* 41, e111.
995 doi:10.1136/jmg.2004.020016.

996 Yang, G., Ren, J., Zhang, Z., and Huang, L. (2009). Genetic evidence for the introgression of Western
997 NR6A1 haplotype into Chinese Licha breed associated with increased vertebral number. *Anim.*
998 *Genet.* 40, 247–250. doi:10.1111/J.1365-2052.2008.01820.X.

999 Yang, J., Li, W., Lv, F., He, S., ... S. T.-M. biology, and 2016, U. (2016). Whole-genome sequencing
1000 of native sheep provides insights into rapid adaptations to extreme environments. *Mol. Biol. Evol.*
1001 33, 2576–2592. Available at: <https://academic.oup.com/mbe/article-abstract/33/10/2576/2925566>
1002 [Accessed August 14, 2021].

1003 Yuan, Z., Liu, E., Liu, Z., Kijas, J. W., Zhu, C., Hu, S., et al. (2017). Selection signature analysis reveals
1004 genes associated with tail type in Chinese indigenous sheep. *Anim. Genet.* 48, 55–66.
1005 doi:10.1111/age.12477.

1006 Zhi, D., Da, L., Liu, M., Cheng, C., Zhang, Y., Wang, X., et al. (2018). Whole genome sequencing of
1007 Hulunbuir short-tailed sheep for identifying candidate genes related to the short-tail phenotype.
1008 *G3: Genes Genomes Genet.* 8, 377–383. doi:10.1534/g3.117.300307.

1009 Zsolnai, A., Anton, I., Fésüs, L., Estonba, A., Schwerin, M., and Vanselow, J. (2002). Allele
1010 distributions of two novel SNPs within the sheep Cyp19 gene. *J. Anim. Breed. Genet.* 119, 402–
1011 405. doi:10.1046/j.1439-0388.2002.00364.x.

1012

1013

1014

1015

1016

1017

Table 1. Estimates of genetic diversity parameters for each of the 15 populations analyzed in this study.

Population	Observed heterozygosity (<i>Ho</i>)	Expected heterozygosity (<i>He</i>)	Nucleotide (π)	RoH Size (Mb)	Inbreeding coefficient (F)	F_{RoH} (<i>Mean</i>)
	(Mean $\pm$ Sd)	(Mean $\pm$ Sd)	(Mean $\pm$ Sd)	(Mean $\pm$ Sd)	(Mean $\pm$ Sd)	(Mean $\pm$ Sd)
Kefis	0.318 $\pm$ 0.0049	0.311 $\pm$ 0.0048	0.0030 $\pm$ 0.0014	0.0210 $\pm$ 0.0218	0.020 $\pm$ 0.0158	0.2056
Segentu	0.377 $\pm$ 0.0206	0.368 $\pm$ 0.0201	0.0029 $\pm$ 0.0015	0.0221 $\pm$ 0.0242	-0.023 $\pm$ 0.0578	0.2231
Adane	0.357 $\pm$ 0.0354	0.351 $\pm$ 0.0361	0.0030 $\pm$ 0.0014	0.0239 $\pm$ 0.0311	-0.017 $\pm$ 0.1090	0.2318
Arabo	0.337 $\pm$ 0.0116	0.335 $\pm$ 0.0116	0.0030 $\pm$ 0.0015	0.0218 $\pm$ 0.0248	-0.005 $\pm$ 0.0356	0.2096
Gafera	0.335 $\pm$ 0.0064	0.329 $\pm$ 0.0063	0.0029 $\pm$ 0.0014	0.0239 $\pm$ 0.0270	-0.019 $\pm$ 0.0197	0.2335
Molale	0.323 $\pm$ 0.0148	0.326 $\pm$ 0.0150	0.0029 $\pm$ 0.0014	0.0246 $\pm$ 0.0301	0.011 $\pm$ 0.0456	0.2377
Bonga	0.337 $\pm$ 0.0029	0.333 $\pm$ 0.0028	0.0027 $\pm$ 0.0014	0.0262 $\pm$ 0.0280	-0.011 $\pm$ 0.0087	0.2797
Gesses	0.348 $\pm$ 0.0052	0.402 $\pm$ 0.0050	0.0028 $\pm$ 0.0014	0.0238 $\pm$ 0.0265	-0.035 $\pm$ 0.0155	0.2360
Kido	0.273 $\pm$ 0.0159	0.263 $\pm$ 0.0155	0.0028 $\pm$ 0.0014	0.0251 $\pm$ 0.0301	-0.025 $\pm$ 0.0474	0.2495
Doyogena	0.326 $\pm$ 0.0109	0.396 $\pm$ 0.0110	0.0028 $\pm$ 0.0014	0.0254 $\pm$ 0.0289	0.009 $\pm$ 0.0331	0.2517
ShubiGemo	0.332 $\pm$ 0.0126	0.327 $\pm$ 0.0122	0.0028 $\pm$ 0.0014	0.0239 $\pm$ 0.0270	-0.015 $\pm$ 0.0385	0.2396
Loya	0.335 $\pm$ 0.0176	0.334 $\pm$ 0.0175	0.0027 $\pm$ 0.0014	0.0258 $\pm$ 0.0296	-0.004 $\pm$ 0.0528	0.2745
Hamhari	0.295 $\pm$ 0.0072	0.321 $\pm$ 0.0078	0.0029 $\pm$ 0.0014	0.0342 $\pm$ 0.0440	0.083 $\pm$ 0.0225	0.2094
Kabashi	0.297 $\pm$ 0.0055	0.321 $\pm$ 0.0059	0.0029 $\pm$ 0.0014	0.0325 $\pm$ 0.0385	0.073 $\pm$ 0.0172	0.2042
Barbary	0.303 $\pm$ 0.0151	0.301 $\pm$ 0.0151	0.0032 $\pm$ 0.0015	0.0182 $\pm$ 0.0224	-0.006 $\pm$ 0.0503	0.1771

Table 2. Candidate regions and genes associated with adaptation to diverse environments as detected in our study.

Region (base pairs)	Chromosome	Genes	Method	Function
253550001-253730000	1	<i>TF, RAB6B, SRPRB</i>	<i>F_{ST}</i>	Associated with haemoglobin levels in Tibetans (Simonson et al., 2010)
232750001-232850000	2	<i>DIS3L2</i>	<i>F_{ST}, XP-EHH</i>	Associated with height variation (de Simoni Gouveia et al., 2017) and thermoregulation (Mcmanus et al.76543, 2011).
77920001-78110000	3	<i>EPAS1, PRKCE</i>	<i>XP-EHH</i>	Regulating cellular responses to hypoxia (Frede and Fandrey, 2014)
154130001-154350000	3	<i>MSRB3</i>	<i>Hp, F_{ST}</i>	high altitude adaptation in dogs (Gou et al., 2014) and Tibetan sheep (Wei et al., 2016).
55940001-56140000	7	<i>GLDN, CYP19</i>	<i>Hp, XP-EHH</i>	Regulation of male and female reproduction in sheep (Zsolnai et al., 2002).
28560001- 29540000	10	<i>RXFP2, PDS5B, N4BP2L2</i>	<i>F_{ST}, XP-EHH</i>	implicate its dual role in the development of horns for thermoregulation and enhancing reproductive (Hoefs, 2000; Kardos et al., 2015).
18250001-18490000	11	<i>NF1, EVI2A, EVI2B, OMG</i>	<i>F_{ST}, XP-EHH</i>	Adaptive response to physical exhaustion (Mwacharo et al., 2017).
27040001- 27180000	11	<i>KCNAB3, CNTROB, DNAH2, NAA38, KDM6B, TMEM88, CYB5D1, CHD3, RNF227, TRAPPC1</i>	<i>Hp, F_{ST}</i>	Correlated with cognitive performance under chronic stress (Jung et al., 2017) and is down-regulated in response to acute and chronic stress in mice (Terenina et al., 2019).
42670001- 42800000	13	<i>CHRNA4, SRMS, PTK6, EEF1A2, KCNQ2, ARFGAP1</i>	<i>XP-EHH</i>	Response to hypoxia (Storz and Moriyama, 2008; Tissot Van Patot and Gassmann, 2011)
69380001- 69520000	13	<i>PLCG1, ZHX3</i>	<i>Hp</i>	Response to hypoxia (Storz and Moriyama, 2008; Tissot Van Patot and Gassmann, 2011)
34380001 - 34590000	14	<i>HSD11B2, ATP6V0D1, AGRP, RIPOR1, CTCF, CARMIL2, ACD, PARD6A, ENKD1, C16orf86, GFOD2</i>	<i>F_{ST}</i>	Associated with salt sensitivity (Alikhani-Koupaei et al., 2007)
35110001-35280000	15	<i>PLEKHA7, C11orf58</i>	<i>F_{ST}</i>	Associated with salt sensitivity (Lin et al., 2011; Taal et al., 2012)
17230001-17480000	20	<i>VEGFA, MAD2L1BP, RSPH9, MRPS18A</i>	<i>F_{ST}</i>	Cerebrovascular adaptation to chronic hypoxia in sheep (Castillo-Melendez et al., 2015)
4050001- 4170000	25	<i>EGLN1, GNPAT, EXOC8, SPRTN</i>	<i>Hp, F_{ST}, XP-EHH</i>	Regulating cellular responses to hypoxia (Frede and Fandrey, 2014)

1020
1021
1022

1023
1024
1025
1026
1027
1028
1029
1030
1031

Table 3. Candidate regions and genes associated with phenotypic traits (tail length and tail fat-depots) as detected in our study.

Region (base pairs)	Chromosome	Genes	Method	Function
52290001-52540000	2	<i>HINT2, SPAG8, NPR2, TMEM8B, FAM221B, MSMP, RGPI, GBA2, CREB3, TLN1, TPM2</i>	<i>F_{ST}, XP-EHH</i>	Associated with birth and carcass weights, and fat depth, respectively, in cattle (Casas et al., 2000; McClure et al., 2010) and sheep (Moradi et al., 2012; Wei et al., 2015).
10920001-11320000	3	<i>NR5A1, NR6A1, GOLGA1, ARPC5L, WDR38, OLFML2A, oar-mir-26a, ADGRD2</i>	<i>F_{ST}, XP-EHH</i>	Influence tail vertebrae number in pigs (Mikawa et al., 2007; Rubin et al., 2012) and thoracic vertebrae number in sheep (Shengwei et al., 2019)
18250001-18490000	11	<i>NF1, EVI2A, EVI2B, OMG</i>	<i>F_{ST}, XP-EHH</i>	Associated with tail morphology and fat deposition (Moioli et al., 2015; Wei et al., 2015; Yuan et al., 2017; Ahbara et al., 2019)
26350001-26540000	11	<i>ALOX12, RNASEK, U6, BCL6B, SLC16A13, SLC16A11, CLEC10A, ASGR2, ASGR1</i>	<i>F_{ST}</i>	Associated with fat deposition (Moioli et al., 2015; Wei et al., 2015; Yuan et al., 2017; Ahbara et al., 2019).
37200001-37430000	11	<i>HOXB13, CALCOCO2, TTLL6, RF02133, RF02132, MIR196A1, HOXB9, HOXB7</i>	<i>F_{ST}, XP-EHH</i>	A strong candidate gene controlling tail length in mice (Economides et al., 2003).

1032 **Table 4.** Candidate regions and genes associated with resistance to endo- and ecto-parasites as detected in our study.

Region (base pairs)	Chromosome	Genes	Method	Function
109920001-110020000	1	<i>CD84</i>	<i>Hp, F_{ST}, XP-EHH</i>	Candidate gene for trypanosomiasis resistance in mice (Yan et al., 2004; Martin and Tarleton, 2005)
69010001-70090000	6	<i>PDGFRA, LNX1, GSX2, CHIC2</i>	<i>Hp, F_{ST}, XP-EHH</i>	Implicated in host immune responses against endoparasites and is a key gene in cytokine signaling (Beh et al., 2002; Benavides et al., 2015).
115960001-117040000	6	<i>LETMI, NELFA, NSD2, SLC49A3, PDE6B</i>	<i>Hp, XP-EHH</i>	Plays an essential role in the infection stages of <i>Trypanosoma brucei</i> by maintaining its mitochondrial integrity (Nowikovsky et al., 2012; Hashimi et al., 2013; Nowikovsky and Bernardi, 2014)
18250001-18490000	11	<i>NF1, EVI2A, EVI2B, OMG</i>	<i>F_{ST}, XP-EHH</i>	Linked to gastrointestinal nematode infection resistance in traditionally managed Tunisian sheep (Ahbara et al., 2021)

1033

1034

Figure 1. Tail phenotypes of different sheep groups: (a) visual length and number of caudal vertebrae (CV). (b) Tail length (meters) and average length (cm) of single vertebra. SFT = short fat-tailed, FR = Fat-rumped, LFT = Long fat-tailed and TT = Thin-tailed; (c). Clustering pattern of the 15 sheep populations studied as revealed by principal component analysis.

Figure 2(a). Phylogenetic network inference based on TreeMix analysis depicting the relationships between Ethiopian, Libyan and Sudanese sheep populations. The first seven migration edges are shown with arrows pointing the direction of gene flow toward the recipient group and coloured according to the percentage of ancestry received from the donor, (b). Patterns of linkage disequilibrium (r^2) from 0 to 1Mb for the four sheep groups revealed by PCA and ADMIXTURE analysis and (c). Average estimated effective population size (N_e) of the four sheep groups revealed by PCA and ADMIXTURE analysis over the past 1000 generations.

Figure 3. Manhattan plot of genome-wide distribution of ZHp, F_{ST}, and XP-EHH values generated from the comparative analysis involving Ethiopian long fat-tailed (ET-G2) sheep with Libyan Barbary (LB) and long thin-tailed (SD) sheep group from Sudan.

Figure 4. Haplotype structure at the candidate region spanning *EGLN1* on OAR25. The position of the most significant variant is shown by the black arrow, allele frequencies in the different groups are shown below. 4b. Haplotype structure around the candidate region spanning the *TF* gene on OAR1.

Figure 6(a). Haplotype structure around the candidate region on OAR2 spanning the *DIS3L2* likely associated with limb length, (b). Haplotype structure around the candidate region on OAR3 spanning *MSRB3* likely associated with ear size and shape, (c). Haplotype structure around the candidate region on OAR15 spanning *PLEKHA7* likely involved in salt-sensitivity, blood pressure, kidney function and hypertension, (d). Haplotype structure around *PDS5B* and *RXFP2* genes likely associated with heat stress and horn characters on OAR10.

Figure 7. Manhattan plot of genome-wide distribution of ZHp values in Ethiopian fat-rumped (ET-G1), long fat-tailed (ET-G2), Libyan Barbary (LB) and Sudanese thin-tailed (SD) sheep group.

Figure 8. Manhattan plot showing genome-wide distribution of ZF_{ST} values following the pairwise comparative analysis between long-tailed Sudan sheep (SD), long-tailed Ethiopian (ET-G2), short-tailed Libyan Barbary (LB) and short-tailed Ethiopian (ET-G1) sheep groups.

Figure 9. Manhattan plot showing genome-wide distribution of XP-EHH values following pairwise comparative analysis between long-tailed Sudan sheep (SD), long-tailed Ethiopian (ET-G2), short-tailed Libyan Barbary (LB) and short-tailed Ethiopian (ET-G1) sheep groups.

Figure 10. Signatures of selection revealed by XP-EHH on OAR11 spanning *HOXB13*, previously identified in connection with tail length as well as new candidate regions spanning genes related to fat metabolism (e.g., *ALOX12*, *NF1*, *EVI2B* and *OMG*).

Figure 11. The haplotype structure around the candidate region on OAR11 spanning *HOXB13* likely associated with vertebrae length.

Figure 12. Manhattan plot showing genome-wide distribution of F_{ST} and XP-EHH values resulting from contrasting of thin-tail SD sheep group with fat-rumped (ET-G1) and fat-tailed (ET-G2 and LB) sheep groups.

Figure S1 The sampling locations of Ethiopian, Sudanese and Libyan sheep populations.

Figure S2 Distribution of SNPs following annotation categories (150 sheep samples).

Figure S3 Distribution of RoHs within range of 1.4 Mb for each sheep population.

Figure S4 Genetic admixture analysis of the studied populations (a) Cross-validation plot (b) Admixture plot.

23

24

25

26

27

28

Figure S5 Plot of HIF-1 pathway showing the key role of the positively selected genes for high-altitude adaptation in African sheep.

29

30

31

32

Figure S6 Illustration of the phenotypic criteria (horns and tail) applied by the owners of sheep in Libya when selecting ram for reproduction.