Genomics 114 (2022) 110448

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

# Natural adaptation and human selection of northeast African sheep genomes

Abulgasim M. Ahbara [a,b,c,d,e,*,1], Hassan H. Musa [f,1], Christelle Robert [g], Ayele Abebe [h], Ahmed S. Al-Jumaili [i], Adebabay Kebede [j,k], Suliman Latairish [l], Mukhtar Omar Agoub [m], Emily Clark [g], Olivier Hanotte [b,j,**], Joram M. Mwacharo [c,e,g,***]

[a] Department of Zoology, Faculty of Sciences, Misurata University, Misurata, Libya
[b] School of Life Sciences, University of Nottingham, University Park, Nottingham, UK
[c] Small Ruminant Genomics, International Centre for Agricultural Research in the Dry Areas (ICARDA), Addis Ababa, Ethiopia
[d] LiveGene, International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia
[e] Animal and Veterinary Sciences, SRUC, The Roslin Institute Building, Midlothian, Edinburgh, UK
[f] Faculty of Medical Laboratory Sciences, University of Khartoum, Sudan
[g] Centre for Tropical Livestock Genetics and Health (CTLGH), The Roslin Institute, University of Edinburgh, UK
[h] Debre Berhan Research Centre, Debre Berhan, Ethiopia
[i] Department of Medical Laboratory Techniques, Al-Maarif University College, Ramadi, Anbar, Iraq
[j] LiveGene-CTLGH, International Livestock Research Institute (ILRI) Ethiopia, Addis Ababa, Ethiopia
[k] Amhara Regional Agricultural Research Institute, Bahir Dar, Ethiopia
[l] Department of Animal Production, Faculty of Agriculture, Misurata University, Misurata, Libya
[m] Agricultural Research Centre, Misurata, Libya

ARTICLE INFO

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

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 northeast African sheep genomes sequenced at an average depth of ~54× for 130 samples (Ethiopia, Libya) and ~20× for 20 samples (Sudan). These represented sheep from diverse environments, tail morphology and post-Neolithic introductions to Africa. Phylogenetic and model-based admixture analysis provided evidence of four genetic groups corresponding to altitudinal geographic origins, tail morphotypes and possible historical introduction and dispersal of the species into and across the continent. Running admixture at higher levels of $K$ ($6 \leq K \leq 25$), revealed cryptic levels of genome intermixing as well as distinct genetic backgrounds in some populations. Comparative genomic analysis identified targets of selection that spanned conserved haplotype structures overlapping clusters of genes and gene families. These were related to hypoxia responses, ear morphology, caudal vertebrae and tail skeleton length, and tail fat-depot structures. Our findings provide novel insights underpinning morphological variation and response to human-driven selection and environmental adaptation in African indigenous sheep.

## 1. Introduction

Following their arrival in Africa, natural and artificial selection shaped the genomes of domestic sheep, which enhanced their adaptive fitness to novel environments. Tail fat deposition likely facilitated such adaptations. Based on tail fat-depot structures, sheep exhibit five distinct phenotypes, fat-long tail, fat-short tail, fat-rumped tail, thin-long tail, and thin-short tail [1]. Three of these tail phenotypes are observed in

* Correspondence to: Abulgasim M. Ahbara, Department of Zoology, Faculty of Sciences, Misurata University, Misurata, Libya.
** Correspondence to: Olivier Hanotte, School of Life Sciences, University of Nottingham, University Park, Nottingham, UK.
*** Correspondence to: Joram M. Mwacharo, Small Ruminant Genomics, International Centre for Agricultural Research in the Dry Areas (ICARDA), Addis Ababa, Ethiopia.
*E-mail addresses:* aahbara@sruc.ac.uk, a.ahbara@sci.misuratau.edu.ly (A.M. Ahbara), olivier.hanotte@cgiar.org, j.mwacharo@cgiar.org (O. Hanotte), joram.mwacharo@sruc.ac.uk (J.M. Mwacharo).
[1] Equal contribution to the manuscript

https://doi.org/10.1016/j.ygeno.2022.110448
Received 15 January 2022; Received in revised form 25 July 2022; Accepted 7 August 2022
Available online 12 August 2022
0888-7543/Crown Copyright © 2022 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

African indigenous sheep: fat-long/short tailed, thin-long tailed and fat-rumped tail. The fat-tailed occur in northeast (Tunisia, Libya, Egypt), eastern, and southern Africa, the thin-tailed are predominantly found in northwest (Morocco, Algeria, Tunisia) and western Africa, and Sudan, while the fat-rumped are restricted to the Horn of Africa [2]. Archaeological findings from northern Africa suggest the thin-tailed sheep are the most ancient on the continent and that domestic sheep entered Africa in multiple waves from the Levant through the Isthmus of Suez and the Horn of Africa [3–5]. However, fat-tailed sheep predominate in northeast Africa today. Rock paintings from the Lake Turkana basin dated to 4500–3500 BP depict thin-tailed sheep [6,7], which today mainly occur in Ethiopia (Benishangul-Gumuz and North Gondar regions), Sudan and West Africa. Rock paintings from eastern Ethiopian highlands show fat-tailed sheep alongside humpless cattle [8]. The latter arrived on the continent earlier than humped cattle which today predominate eastern Africa [9]. Depictions of fat-rumped sheep lack in archaeological findings, thus their origin and spatiotemporal diffusion into the continent remains speculative.

Africa is home to ~418 million sheep (FAOSTAT 2020; accessed June 2022) comprising 140 phenotypically diverse populations (DAD-IS 2020; accessed June 2022). The continent has marked differences in agro-eco-climates and biophysical challenges (https://iiasa.ac.at/web/home/research/researchPrograms/water/GAEZ_v4.html). It is also home to an exceptional human ethnic agro-pastoral diversity of ancient origin. Thus, natural selection for adaptation to diverse environments, and human preference for economic, socio-cultural and aesthetic traits could have shaped the genomes of African livestock, resulting in large variations within and between populations (*e.g.*, [10,11]). The genetic control underpinning these adaptations and variations remain largely uninvestigated. We generated and analysed whole-genome sequences alongside tail skeleton and morphology of indigenous northeast African sheep from Libya, Sudan and Ethiopia. The studied populations represented fat-tailed, thin-tailed and fat-rumped African sheep tail morphotypes. The animals were sampled from different agro-ecologies, which ranged from cold high-altitude to hot arid lowland, environments. Using the data, we investigated genome-wide genetic variation and structure and inferred demographic-adaptive dynamics of the species in the continent.

## 2. Materials and methods

### 2.1. Sample collection

Blood samples were collected in EDTA coated vacutainer tubes from 150 individuals of 15 indigenous African fat-tailed, fat-rumped and thin-tailed sheep from Ethiopia, Libya and Sudan (Tables S1 and S2; Fig. S1). Genomic DNA was isolated with Qiagen's DNeasy Blood and Tissue Kit and quality checked with the Nanodrop. Using the Covaris System, 3 μg of genomic DNA were randomly sheared to generate inserts of ~300 bp. The sheared DNA was end-repaired, A-tailed, adaptor ligated, and amplified with the TruSeq DNA Sample Preparation Kit (Illumina, San Diego, CA, USA). Paired-end sequencing was conducted with the Illumina HiSeq2000. We performed sequence quality checks with fastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). The paired-end sequence reads were mapped against the Oar_v3.1.75 sheep reference genome using Burrows-Wheeler Aligner [12]. We used default parameters (except the "–no-mixed" option) to suppress unpaired alignments. Potential PCR duplicates were filtered using the "REMOVE_DUPLICATES = true" option in "MarkDuplicates" command-line of Picard (http://broadinstitute.github.io/picard). SAMtools [12] was used to create the index files for reference and bam files. The genome analysis toolkit (GATK) 3.1 (https://gatk.broadinstitute.org) was used to realign the reads to correct for any misalignments arising from the presence of indels ("RealignerTargetCreator" and "IndelRealigner" arguments). "UnifiedGenotyper" and "SelectVariants" of GATK were used to call SNPs. To filter variants and to avoid possible false positives, SNPs with: (i) a phred-scaled quality score < 30, (ii) MQ0 (mapping quality zero) > 4 and quality depth (unfiltered depth of non-reference samples) < 5, and (iii) SNPs with FS (phred-scaled *P* value using Fisher's exact test) > 200, were filtered out. Following these filtration steps, the retained SNPs were used for analysis.

### 2.2. Preparation and measurement of caudal vertebrae

We first subdivided the 15 populations into five groups based on their tail length and tail fat-depot sizes, *viz*, i) Ethiopian fat-rumped sheep (Kefis, Adane, Arabo, Segentu), ii) Ethiopian long fat-tailed sheep (Bonga, Kido, Gesses, Loya, ShubiGemo, Doyogena), iii) Ethiopian short fat-tailed sheep (Molale, Gafera), iv) Sudan long thin-tailed sheep (Hammari, Kabashi), and v) Libyan long fat-tailed sheep (Barberine).

One mature animal (≥3 permanent-pairs of incisor teeth) of Kefis (fat-rumped), Loya (long fat-tailed), Menz (short fat-tailed), Kabashi (long thin-tailed) and Barberine (Libyan long fat-tailed) was slaughtered, the tail eviscerated, and the full tail skeleton processed following [13]. The tail was deskinned, and the flesh and fat gently scraped off the caudal vertebrae (CV's). The CV's were then incubated in ethanol for five days and thereafter they were soaked in 0.5% NaOH for two days. Residual flesh and fat were removed by further soaking the CV's in petrol for three days. The number of CV's was counted and the length (cm) of each CV was determined with a Vernier calliper. The length of the complete tail skeleton was determined with a tape measure. The average length and standard deviation of the CV's were calculated in Excel.

### 2.3. Genetic diversity

VCFtools v.0.1.15 [14] was used to estimate observed ($H_O$) and expected ($H_E$) heterozygosity, and nucleotide diversity ($\pi$) as indicators of intra-population diversity. The inbreeding coefficient (*F*) and runs of homozygosity (*RoH*) were estimated as indicators of autozygosity. The *RoH* was estimated with BCFtools [15]. We estimated the *RoH*-derived genomic inbreeding coefficient ($F_{RoH}$) following [16] as:

$$F_{ROH} = \frac{L_{RoH}}{L_{AUTO}}$$

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).

### 2.4. Phylogenetic inference and genetic structure analysis

We visualized population structure using principal component analysis (PCA) performed with Plink 1.9 [17]. The first two principal components (PC) were plotted to summarise individual relationships. For this analysis, we used 17 million autosomal SNPs, which were retained after filtering SNPs with minor allele frequency (MAF) < 0.01.

To provide a graphic representation of genetic distances between individuals, and populations, we used the 17 million autosomal SNPs used in PCA analysis and constructed a Neighbour-net phylogenetic network using SplitsTree6 [18] and a cladogram phylogeny employing the Neighbour-Joining (NJ) algorithm. We also constructed a maximum likelihood (ML) phylogenetic tree with Phyml version 3.0 [19]. The best fit model for constructing the ML phylogeny was determined with jModeltest version 2.1.7 [20]. The confidence level for each bifurcation was assessed with 1000 bootstrap replications of the dataset.

We investigated the proportion of the genome arising from common ancestry using ADMIXTURE v.1.3 [21]. For this analysis, we used 2.4 million autosomal SNPs that were retained after pruning those that were in linkage disequilibrium (LD) from the 17 million used in PCA analysis. We implemented the block relaxation algorithm with the Kinship (*K*) values set from 2 to 25 and performed five runs 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 [22]. We further tested for admixture while revealing deeper insights on population and individual relationships by generating the f4 and D-statistics scores as described in [23,24].

### 2.5. Demographic history and dynamics

We estimated the LD parameter ($r^2$) with Plink v1.9 considering the genetic groups/structure inferred from the phylogenenetic and ADMIXTURE analyses. The $r^2$ values were sorted and binned in 0.05 to 10 Mb inter-SNP distances and the genome-wide LD decay plotted with 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 [25]. This was done for each genetic group/structure inferred by phylogenetic and ADMIXTURE analyses. The $N_e$ values were estimated following [26] 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 $t = 1/2c$ [27].

### 2.6. 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 (based on the length of CV's), and iii) differences in tail fat-depot size.

Based on altitude and agroecology of the geographic home ranges of the studied sheep populations (Table S2), they were classified into three groups: i) sheep populations from high altitude humid environment (Bonga, Kido, Gesses, Loya, ShubiGemo, Doyogena and Gafera from Ethiopia), ii) sheep populations from low altitude hot-dry desert environment (Hammari and Kabashi from Sudan), and iii) sheep populations from coastal hot-humid arid environment (the Barberine from Libya). Contrasting these three groups allowed for the detection of genomic signatures for adaptation to contrasting environments.

Based on the average length of the CV's and the complete tail skeleton, we classified the studied populations into two groups: i) short-tailed sheep (short fat-tailed and fat-rumped sheep), and ii) long-tailed sheep (long fat-tailed and long thin-tailed sheep). These were used to investigate genomic regions associated with differences in sizes of CV's and in the length of the tail skeletons.

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

Signatures of selection were investigated by analysing 34.8 million autosomal SNPs with three methods: i) within-group pooled heterozygosity (*Hp*) [28], ii) genetic differentiation based on $F_{ST}$ [29], and iii) the haplotype-based *XP–EHH* [30]. A sliding-window was used for the *Hp* and $F_{ST}$ analysis done with VCFtools v.0.1.15. Based on the LD decay trend, a 100 kb non-overlapping window was chosen for the *Hp* and $F_{ST}$ analysis. By exploring several distances (10, 20, 50, 75, and 100 kb) for the sliding-window, the 10 kb sliding-window distance gave the best resolved selection signals and was thus used for the analysis.

We estimated the genome-wide *Hp* statistic using the formula:

$$Hp = \frac{2\sum nMaj \sum nMin}{\left(\sum nMaj \sum nMin\right)^2}$$

where $\sum$nMaj and $\sum$nMin are the sum of the number of the major allele and the number of the minor allele, respectively. We transformed the *Hp* values into *Z*-scores using the formula:

$$ZHp = \frac{(Hp - \mu Hp)}{\sigma Hp}$$

where $\mu Hp$ is the average of the overall heterozygosity and $\sigma Hp$ is the standard deviation for all the windows in the test group.

The $F_{ST}$ [29] values were estimated for each SNP in each window between the test groups with the formulae:

$$F_{ST} = 1 - \frac{p1q1 + p2q2}{2prqr}$$

where *p1, p2* and *q1, q2* 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:

$$ZF_{ST} = \frac{F_{ST} - \mu F_{ST}}{\sigma F_{ST}}$$

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 the test groups.

The *XP-EHH* contrasts extended haplotype homozygosity (*EHH*) between populations in detecting selection signatures [30]. 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 *EHHS* for the targeted sheep group and $iES_{pop2}$ is the integrated *EHHS* for the reference sheep group.

For the analysis, haplotype phasing was inferred on all bi-allelic SNPs with BEAGLE [31]. Assuming 100 Mb = 1 Morgan [32], phased haplotypes were converted to IMPUTE format using VCFtools v.0.1.15. The generated haplotypes were then used to estimate *XP-EHH* scores in pairwise comparisons between the test groups using HAPBIN [33]. To determine the empirical significance of the *XP-EHH* statistic, *XP-EHH* scores were normalized by subtracting the mean and dividing it with the standard deviation of all the scores. Negative *XP-EHH* scores indicate selection in the reference group, while positive scores indicate selection in the test group.

The top negative 99% values for *ZHp*, and the top 0.1% of the empirical distributions of the $ZF_{ST}$ and *XP-EHH* scores were used as the cut-off threshold to detect outliers and define candidate regions under selection. For a region to be considered under selection, it had to span at least three significant SNPs. Candidate regions that overlapped between at least two methods were identified and merged with Bedtools v.2.25.0 [32].

### 2.7. Haplotype structure analysis

We investigated the candidate region intervals and putative genes for evidence of conserved haplotype structures. We estimated haplotype frequencies and visualized the haplotype structures using the *hapFLK* software [34] using scripts from the *hapFLK* homepage (https://forge-dga.jouy.inra.fr/projects/hapflk). To further explore the haplotype structures, we identified the region under selection and showing the highest genetic differentiation between populations. Haplotypes were examined at this region, by first filtering out the commonest and rarest SNPs in the populations analysed. The number of pairwise differences between every pair of haplotypes were computed. Using this statistic, the haplotypes were ordered from the most common haplotype in each population separately. To visualise the ordering, we generated a Median-Joining (MJ) network showing the relationship between the

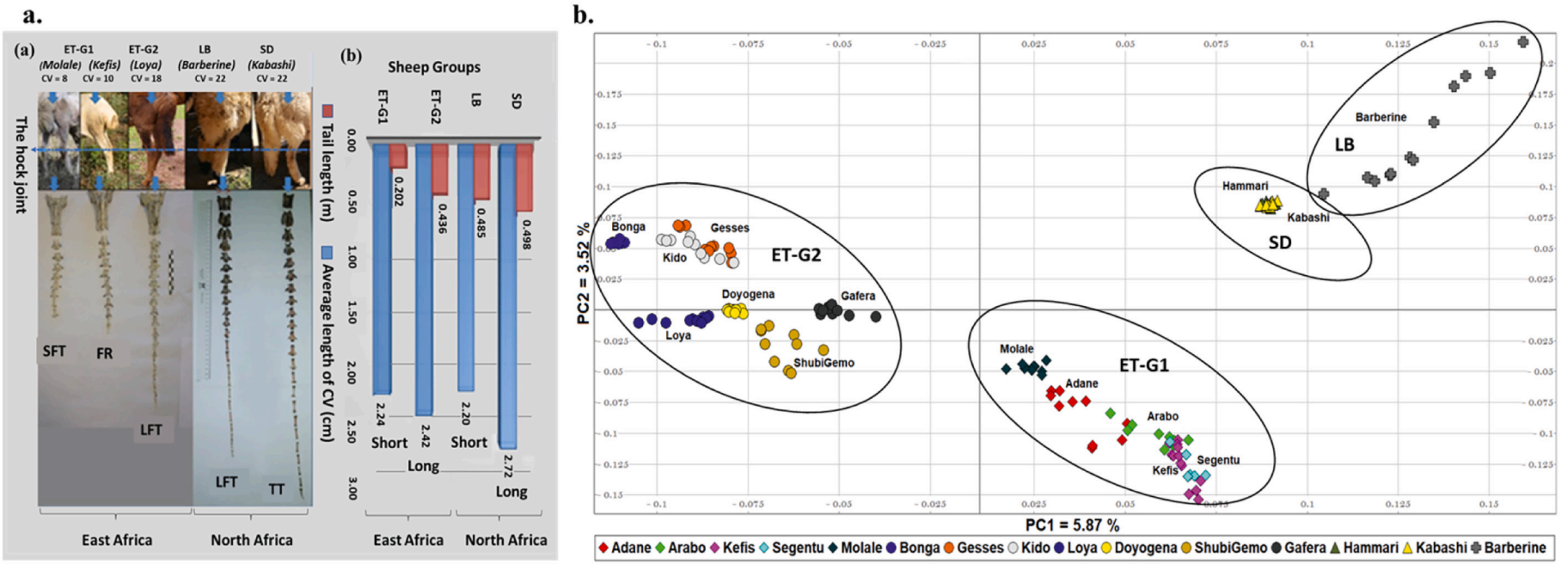


**Fig. 1. a**. Tail phenotypes of different sheep groups: **(a)** visual length and number of caudal vertebrae (CV). **(b)** Tail length (centimetres) and average length (cm) of individual CV. SFT = short fat-tailed, FR = Fat-rumped, LFT = Long fat-tailed and TT = Thin-tailed. Long = Long CVs and Short = Short CVs; **b.** Principal components analysis showing the clustering pattern of the 15 sheep populations studied.

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

| Sheep Group | Population | Observed heterozygosity (*Ho*) | Expected heterozygosity (*He*) | Nucleotide ($\pi$) | RoH Size (Mb) | Inbreeding coefficient (*F*) | $F_{RoH}$ *(Mean)* |
|---|---|---|---|---|---|---|---|
| | | (Mean ± Sd) | (Mean ± Sd) | (Mean ± Sd) | (Mean ± Sd) | (Mean ± Sd) | (Mean ± Sd) |
| ET-G1 | Kefis | 0.318 ± 0.0049 | 0.311 ± 0.0048 | 0.0030 ± 0.0014 | 0.0210 ± 0.0218 | 0.020 ± 0.0158 | 0.2056 |
| | Segentu | 0.377 ± 0.0206 | 0.368 ± 0.0201 | 0.0029 ± 0.0015 | 0.0221 ± 0.0242 | −0.023 ± 0.0578 | 0.2231 |
| | Adane | 0.357 ± 0.0354 | 0.351 ± 0.0361 | 0.0030 ± 0.0014 | 0.0239 ± 0.0311 | −0.017 ± 0.1090 | 0.2318 |
| | Arabo | 0.337 ± 0.0116 | 0.335 ± 0.0116 | 0.0030 ± 0.0015 | 0.0218 ± 0.0248 | −0.005 ± 0.0356 | 0.2096 |
| | Molale | 0.323 ± 0.0148 | 0.326 ± 0.0150 | 0.0029 ± 0.0014 | 0.0246 ± 0.0301 | 0.011 ± 0.0456 | 0.2377 |
| ET-G2 | Gafera | 0.335 ± 0.0064 | 0.329 ± 0.0063 | 0.0029 ± 0.0014 | 0.0239 ± 0.0270 | −0.019 ± 0.0197 | 0.2335 |
| | Bonga | 0.337 ± 0.0029 | 0.333 ± 0.0028 | 0.0027 ± 0.0014 | 0.0262 ± 0.0280 | −0.011 ± 0.0087 | 0.2797 |
| | Gesses | 0.348 ± 0.0052 | 0.402 ± 0.0050 | 0.0028 ± 0.0014 | 0.0238 ± 0.0265 | −0.035 ± 0.0155 | 0.2360 |
| | Kido | 0.273 ± 0.0159 | 0.263 ± 0.0155 | 0.0028 ± 0.0014 | 0.0251 ± 0.0301 | −0.025 ± 0.0474 | 0.2495 |
| | Doyogena | 0.326 ± 0.0109 | 0.396 ± 0.0110 | 0.0028 ± 0.0014 | 0.0254 ± 0.0289 | 0.009 ± 0.0331 | 0.2517 |
| | ShubiGemo | 0.332 ± 0.0126 | 0.327 ± 0.0122 | 0.0028 ± 0.0014 | 0.0239 ± 0.0270 | −0.015 ± 0.0385 | 0.2396 |
| | Loya | 0.335 ± 0.0176 | 0.334 ± 0.0175 | 0.0027 ± 0.0014 | 0.0258 ± 0.0296 | −0.004 ± 0.0528 | 0.2745 |
| SD | Hammari | 0.295 ± 0.0072 | 0.321 ± 0.0078 | 0.0029 ± 0.0014 | 0.0342 ± 0.0440 | 0.083 ± 0.0225 | 0.2094 |
| | Kabashi | 0.297 ± 0.0055 | 0.321 ± 0.0059 | 0.0029 ± 0.0014 | 0.0325 ± 0.0385 | 0.073 ± 0.0172 | 0.2042 |
| LB | Barberine | 0.303 ± 0.0151 | 0.301 ± 0.0151 | 0.0032 ± 0.0015 | 0.0182 ± 0.0224 | −0.006 ± 0.0503 | 0.1771 |

haplotypes and their frequencies in each population using NETWORK 5.0.0 [35] and PopArt [36]. This analysis was applied to the 169.9 Kb selection signature region that spanned the *PLEKHA7* gene. This region occurred on OAR15 and was specific to Libyan Barberine and thus differentiated it from the other populations. There were 271 SNPs in the region that passed all quality filters. To limit the number of haplotypes to display, we identified and plotted the MJ network for the 55 commonest haplotypes.

### 2.8. Functional annotation

The candidate regions were annotated based on the Oar-v3.1.75 genome assembly with the *Ensembl BioMart* tool (http://www.ensembl.org/biomart). Using annotated genes from all the candidate regions, DAVID v6.8 [37] was used for functional enrichment analysis with the *O. aries* annotation used as the background. To interpret the gene functions in a livestock context, we retrieved information available on their functional effects from literature.

## 3. Results

### 3.1. Variant discovery and annotation

The overall alignment rate of the reads to the Oar-v3.1.75 assembly was 98.56% with an average sequencing depth of ~54× for Ethiopian and Libyan samples and ~20× for Sudanese samples. The summary statistics for the sequence parameters are shown in Tables S3 and S4. The comparison with the *O. aries* dbSNP, revealed ~6% novel SNPs and InDels (Tables S3 and S4). Around, 36.8%, 55.9% and 0.79% of the 34,857,882 SNPs identified in this study were intronic, intergenic and exonic, respectively (Table S5; Fig. S2).

### 3.2. Tail skeleton morphometry

The characteristics of the tail and its skeleton for Menz, Kefis, Loya, Barberine and Kabashi are shown in Fig. 1a. The tail of Loya (long fat-tailed), extends to the hock joint while that of Menz (short fat-tailed) and Kefis (fat-rumped) ends above the hock joint. Kabashi (thin-tailed) and Barberine (long fat-tailed) have the longest tails, which extends beyond the hock joint; that of Kabashi almost touches the ground. The short fat-tailed and fat-rumped sheep have 8–10 CV's, the Ethiopian long-tailed has 18, and LB and SD have 22 each. SD has the longest mean CV length (2.72 ± 0.364 cm), followed by 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 skeleton length compared to SD. Taking the average lengths of individual CV's and the complete tail skeleton, we classified the populations into two groups, i) populations with long CV's (SD and Ethiopian long fat-tailed sheep) and, ii) populations with short CV's (LB long fat-tailed and Ethiopian fat-rumped sheep).

### 3.3. Genetic diversity

The average estimates of $H_O$, $H_E$, $\pi$, *F*, *RoH* and $F_{RoH}$, and their standard deviations are shown in Table 1. The least and most diverse populations followng $H_O$ and $H_E$ were Kido and Segentu, respectively. Barberine had the highest value of $\pi$ while Bonga and Loya had the lowest values. Except Hammari, Kabashi and Menz, the other populations had low values of *F* (< 1.0%). Hammari and Kabashi reported the largest mean sizes of *RoH* while the Libyan Barberine had the shortest (Table 1; Fig. S3). Barberine sheep also had the lowest $F_{RoH}$, while Bonga and Loya had the highest (Table 1).

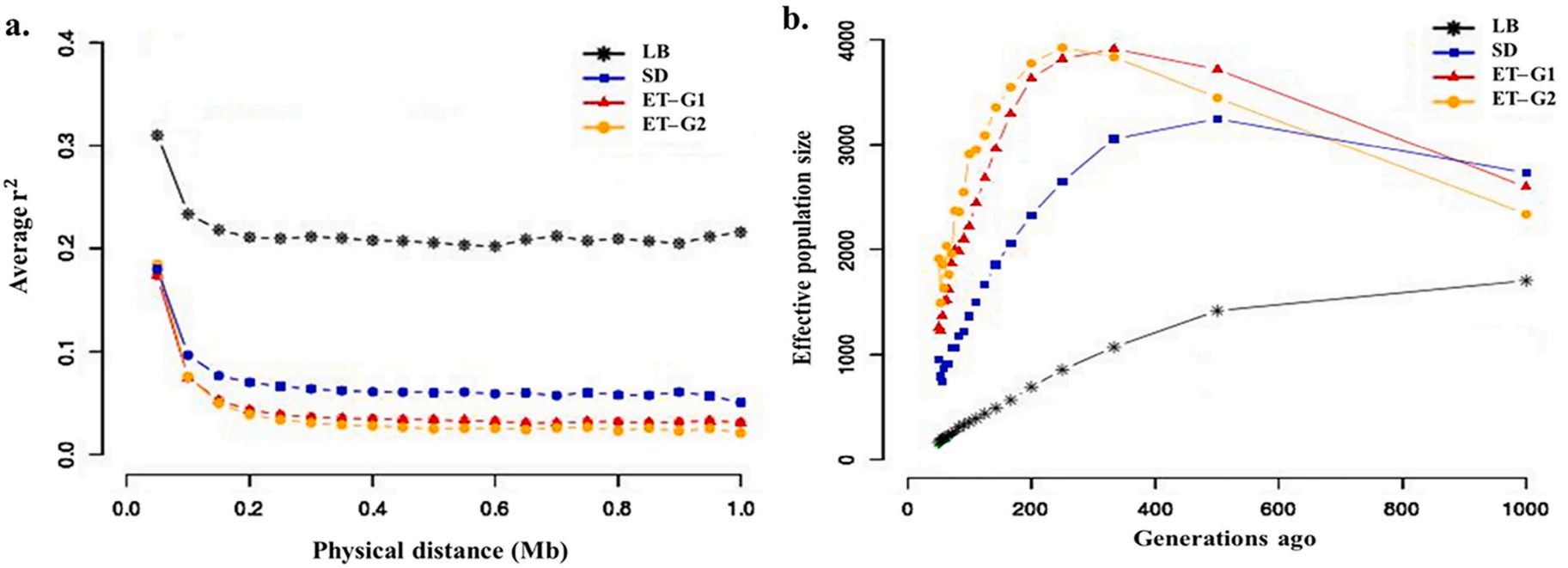


**Fig. 2.** **a.** Patterns of linkage disequilibrium ($r^2$) from 0 to 1 Mb for the four sheep groups revealed by PCA and ADMIXTURE analysis and **b.** Average estimated effective population size (*Ne*) of the four sheep groups revealed by PCA and ADMIXTURE analysis over the past 1000 generations.

**Table 2**
Candidate regions and genes associated with adaptation to diverse environments. Genes highlighted in bold are those identified in the highest-scoring widows based on the reported methodology.

| Region (base pairs) | Chr. | Comparison | Method | Genes | Function and evidence in other species |
|---|---|---|---|---|---|
| 253,550,001–253,730,000 | 1 | *ET-G2 vs. LB and SD* | $F_{ST}$ | ***TF**, RAB6B, SRPRB* | Associated with haemoglobin levels in Tibetans [55] |
| 232,750,001–232,850,000 | 2 | *LB and SD vs. ET-G2* | $F_{ST}$, *XP-EHH* | ***DIS3L2*** | Associated with height variation [72] and thermoregulation [74]. |
| 77,920,001–78,110,000 | 3 | *ET-G2 vs. LB and SD* | *XP-EHH* | ***EPAS1**, PRKCE* | Regulating cellular responses to hypoxia [58] |
| 154,130,001–154,350,000 | 3 | *ET-G2 vs. LB and SD* | *Hp*, $F_{ST}$ | ***MSRB3,** LEMD3* | High altitude adaptation in dogs [65] and Tibetan sheep [62]. |
| 55,940,001–56,140,000 | 7 | *LB and SD vs. ET-G2* | *Hp, XP-EHH* | ***GLDN**, CYP19* | Regulation of male and female reproduction in sheep [108]. |
| 28,560,001–29,540,000 | 10 | *LB vs. ET-G2* | $F_{ST}$, *XP-EHH* | ***RXFP2**, PDS5B, N4BP2L2* | implicate its dual role in the development of horns for thermoregulation and enhancing reproductive [104,106]. |
| 18,250,001–18,490,000 | 11 | *LB and SD vs. ET-G2* | $F_{ST}$, *XP-EHH* | ***NF1,** EVI2A, EVI2B, OMG* | Adaptive response to physical exhaustion [44] |
| 27,040,001–27,180,000 | 11 | *LB and SD vs. ET-G2* | *Hp*, $F_{ST}$ | ***KCNAB3**, **CNTROB**, DNAH2, NAA38, KDM6B, TMEM88, CYB5D1, CHD3, RNF227, TRAPPC1* | Correlated with cognitive performance under chronic stress [69] and is down-regulated in response to acute and chronic stress in mice [70]. |
| 42,670,001–42,800,000 | 13 | *ET-G2 vs. LB and SD* | *XP-EHH* | ***CHRNA4**, SRMS, PTK6, EEF1A2, KCNQ2, ARFGAP1* | Response to hypoxia [56,57] |
| 69,380,001–69,520,000 | 13 | *ET-G2* | *Hp* | ***PLCG1**, ZHX3* | Response to hypoxia [56,57] |
| 34,380,001–34,590,000 | 14 | *LB vs. ET-G2 and SD* | $F_{ST}$ | ***HSD11B2**, ATP6V0D1, AGRP, RIPOR1, CTCF, CARMIL2, ACD, PARD6A, ENKD1, C16orf86, GFOD2* | Associated with salt sensitivity [86] |
| 35,110,001–35,280,000 | 15 | *LB vs. ET-G2 and SD* | $F_{ST}$ | ***PLEKHA7**, C11orf58* | Associated with salt sensitivity [83,84] |
| 17,230,001–17,480,000 | 20 | *ET-G2 vs. LB and SD* | $F_{ST}$ | ***VEGFA**, MAD2L1BP, RSPH9, MRPS18A* | Cerebrovascular adaptation to chronic hypoxia in sheep [63] |
| 4,050,001–4,170,000 | 25 | *ET-G2 vs. LB* | *Hp*, $F_{ST}$, *XP-EHH* | ***EGLN1**, GNPAT, EXOC8, SPRTN* | Regulating cellular responses to hypoxia [58] |

### 3.4. Population structure

The PCA analysis (Fig. 1b) separates the Ethiopian sheep into two groups as observed by [38] from the analysis of 50 K SNP data. To ensure consistency, we adopted [38] nomenclature *viz*: "ET-G1": fat-rumped sheep (Kefis, Adane, Arabo, Segentu) and the short fat-tailed (Molale/Menz) sheep, and "ET-G2": long fat-tailed sheep (Bonga, Kido, Gesses, Loya, ShubiGemo, Doyogena) and the short fat-tailed Gafera sheep. The thin-tailed sheep (Hammari and Kabashi) from Sudan (SD) and the long fat-tailed Barberine (LB) sheep from Libya clustered close but separate from each other and from Ethiopian sheep. The Barberine sheep are more genetically diverse than the other populations. The individual and population level clustering patterns revealed by Neighbour-net phylogenetic network (Fig. S4a), the NJ cladogram (Fig. S4b) and the ML tree (Fig. S4c) were consistent with those revealed by PCA.

To determine the optimal clustering pattern generated by ADMIXTURE, we calculated and plotted the CV error for each $K$. The lowest CV error is at $K = 3$ (Fig. S5a). The distribution among the 15 populations of the three genetic backgrounds, named here A, B and C, is shown in Fig. S5b. The "A" and "B" backgrounds predominated in ET-G1 and ET-G2, respectively. The "C" background occurred in SD and LB both of which showed the "A" and "B" backgrounds. The "C" background was observed 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 the individuals of Gafera and Adane, whose genomes are defined by the "D" background at $4 \leq K \leq 5$. Furthermore, at $K = 5$, a fifth background (E) that characterizes western Ethiopian populations from the other populations in ET-G2 group is observed (Fig. S5b). To further investigate fine-scale differentiation, we extended the admixture inference to $K = 25$ (Fig. S5b). We found a strong genetic differentiation of Bonga, Doyogena and Molale with clear and distinct genetic backgrounds, from all the other populations, that persisted to $K = 25$. Gafera was predominated by one background, while the other eight Ethiopian populations show admixed genome profiles. Hammari and Kabashi are defined by one common genetically distinct background. The Barberine shows a mixed genetic background.

The f4-statistics highlighted possibilities of gene flow among various populations. The greatest negative and positive Z scores (−93.33 and 94.42) support possible admixture between Gesses and Kido and/or between Hammari and Kabashi (Table S6). Within the Ethiopian sheep the second topmost f4 statistic suggest gene flow between either Kefis and Arabo or Gesses and Kido. The D-statistic revealed significantly positive values. The 15 most extreme D-statistic scores are shown in Table S7. The highest Z score (19.43) indicates possibilities of gene flow between Bonga and Gafera or Kido and Loya sheep.

The average values of $r^2$ reflecting genome-wide LD were lowest in ET-G2 and highest in LB (Fig. 2a; Table S8). Irrespective of the genetic group, the trend in $r^2$ showed a rapid decline at the 10 kb to 100 kb distance interval. The trends in $N_e$ for each genetic group are shown in Fig. 2b and Table S9. SD and LB showed the highest and lowest $N_e$, respectively 1000 generations ago. Up to 300 generations ago, ET-G1 showed a higher $N_e$ than ET-G2, after which the opposite is observed. Except LB whose $N_e$ declined gradually up to 50 generations ago, that of ET-G1, ET-G2 and SD increased gradually up to 300 generations ago, and then declined rapidly up to 50 generations ago.

### 3.5. Genome-wide scans for signatures of selection

The PCA revealed four genetic groups (ET-G1, ET-G2, SD and LB). Because the north African LB and the northeast African SD shared a prominent genetic background (C), these four groups were reduced to three by ADMIXTURE tool which revealed the optimal CV error to be $K = 3$ (Fig. S5b). The ET-G2 comprise populations from a high-altitude humid environment, while LB and SD are sheep from a coastal hot-humid arid environment in Libya and a low altitude hot-dry desert environment in Sudan, respectively (Table S1 and S2). The ET-G1 included populations from diverse environments.

#### 3.5.1. Selection signatures for populations from high-altitude humid environments

We used the ET-G2 group to investigate selection signatures for adaptation to high-altitude African environments (Fig. S6). When ET-G2

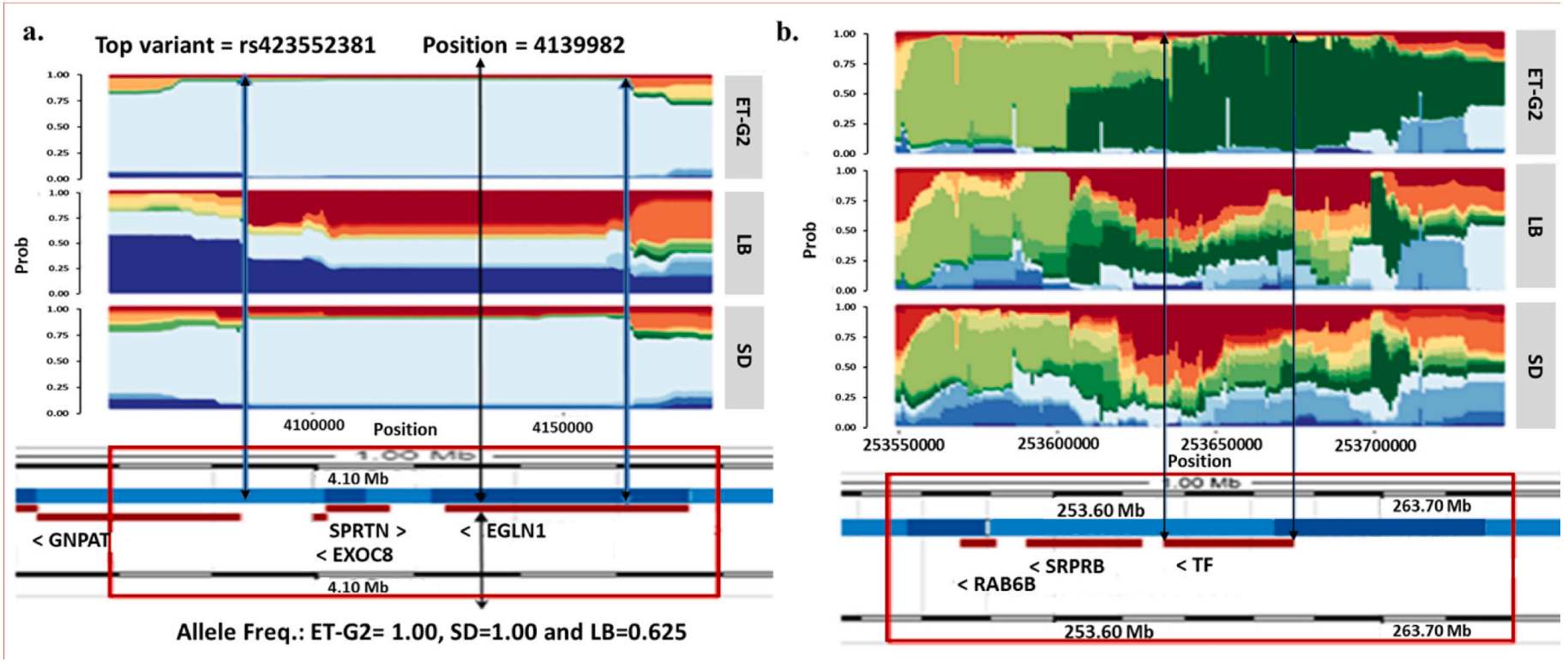


**Fig. 3.** **a.** Haplotype structure at the candidate region spanning *EGLN1* on OAR25. The position of the most significant variant is shown by the black arrow, allele frequencies in the different groups are shown below. **b.** Haplotype structure around the candidate region spanning the *TF* gene on OAR1 *(Black and Blue arrows indicate the most distinguished haplotype within the candidate gene region)*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

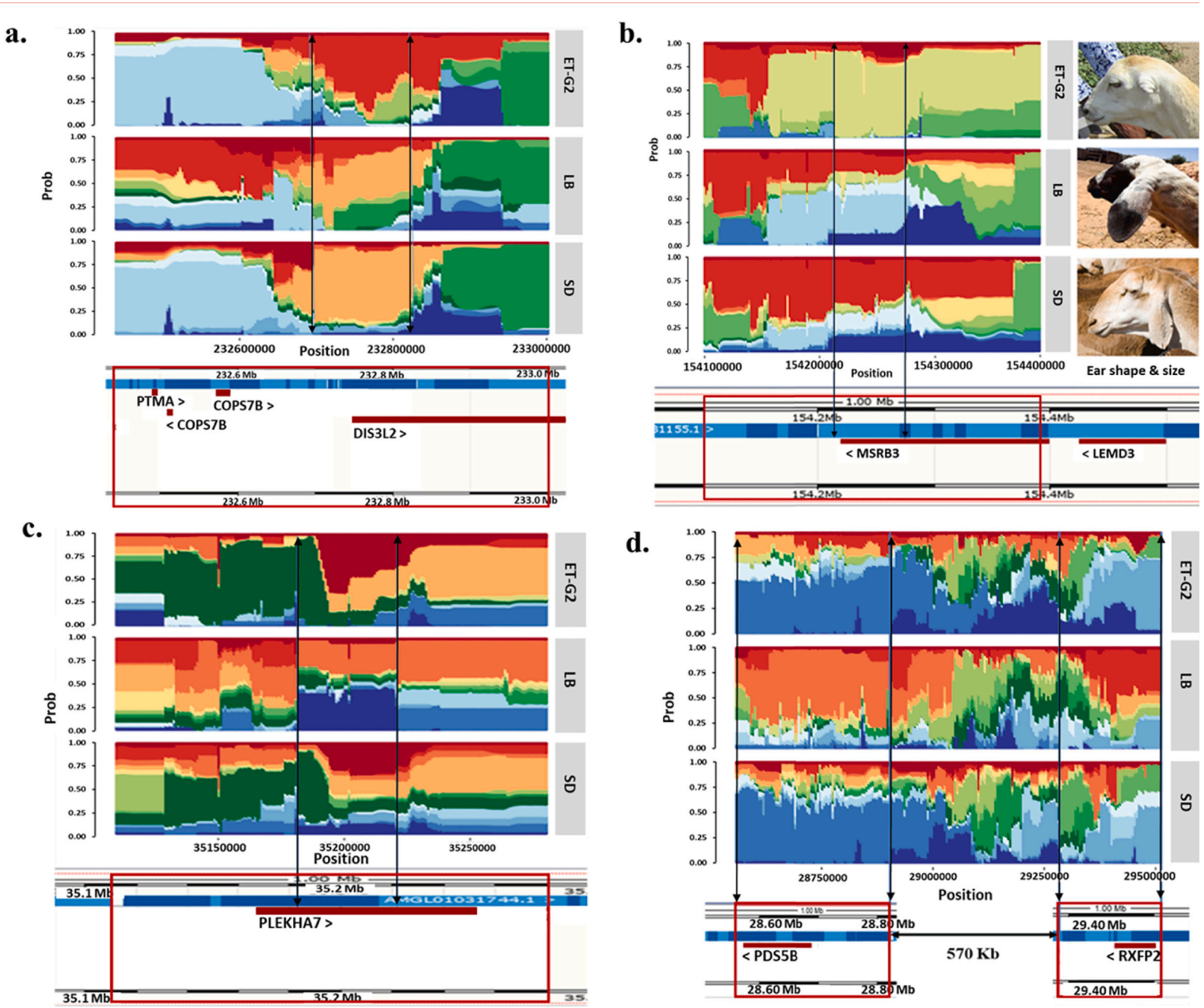


**Fig. 4.** **a.** Haplotype structure around the candidate region on OAR2 spanning *DIS3L2* gene likely associated with limb length, **b.** Haplotype structure around the candidate region on OAR3 spanning *MSRB3* gene likely associated with ear size and shape, **c.** Haplotype structure around the candidate region on OAR15 spanning *PLEKHA7* gene likely involved in salt-sensitivity and kidney function, **d.** Haplotype structure on OAR10 around *PDS5B* and *RXFP2* genes that are likely associated with heat stress and horn characters. *(Black and Blue arrows indicate the most distinguished haplotype within the candidate gene region)*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

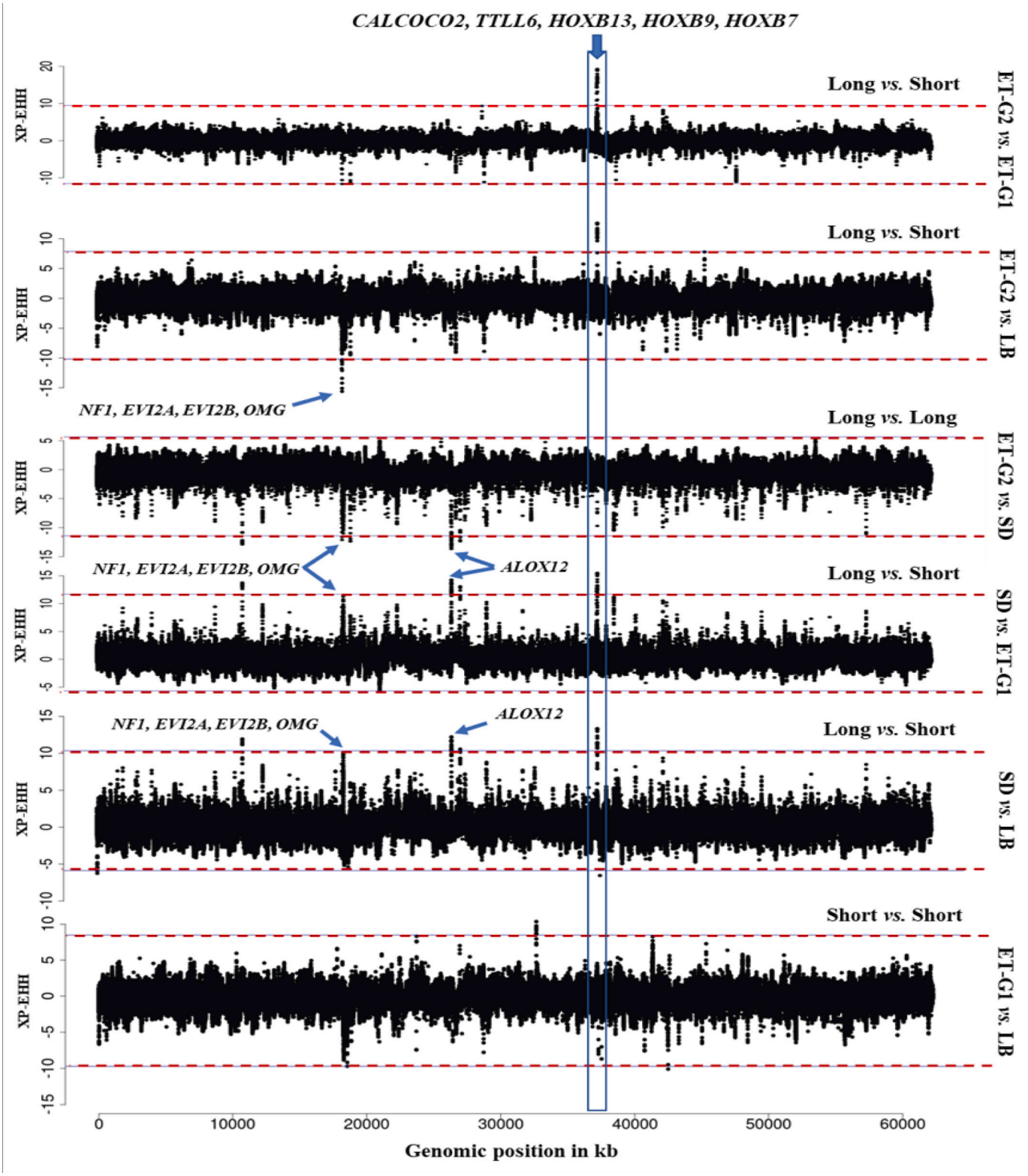


**Fig. 5.** Signatures of selection revealed by *XP-EHH* on OAR11 spanning *HOXB13*, previously associated with tail skeletal length as well as new candidate regions spanning genes related to fat metabolism (*e.g.*, *ALOX12, NF1, EVI2A, EVI2B* and *OMG*) (Long = Long CVs, Short = Short CVs).

was contrasted with LB and SD, the three analytical approaches (*Hp*, $F_{ST}$, *XP-EHH*), either independently or in combination, identified 91 candidate regions spanning 250 genes (Table S10). These genes were used for functional enrichment analysis from which we considered the genes comprising the three most significant (*P*-value $\leq$0.01) KEGG Pathways and GO terms as the primary candidates that are driving high altitude adaptation (Table S11). The genes that have been associated with high altitude environmental adaptation and which were detected in our candidate selection regions are presented in Table 2. The top-two most significant GO biological process terms were "response to hypoxia (GO:0001666)" and "positive regulation of angiogenesis (GO:0045766,)". Genes associated with "response to hypoxia" were

**Table 3**
Candidate regions and genes associated with phenotypic traits (tail length and tail fat-depots). Genes highlighted in bold are those identified in the highest-scoring widows based on the reported methodology.

| Region (base pairs) | Chr. | Comparison | Method | Genes | Function and evidence in other species |
|---|---|---|---|---|---|
| 52,290,001–52,540,000 | 2 | *SD vs. ET-G, ET-G2 and LB* | $F_{ST}$, *XP-EHH* | ***HINT2, SPAG8, NPR2****, TMEM8B, FAM221B, MSMP, RGP1, GBA2, CREB3, TLN1, TPM2* | Associated with birth and carcass weights, and fat depth, respectively, in cattle [109,110] and sheep [50,94]. |
| 10,920,001–11,320,000 | 3 | *ET-G2 and SD vs. ET-G1 and LB* | $F_{ST}$, *XP-EHH* | ***NR5A1, NR6A1****, GOLGA1, ARPC5L, WDR38, OLFML2A, oar-mir-26a, ADGRD2* | Influence tail vertebrae number in pigs [96,97] and thoracic vertebrae number in sheep [98] |
| 18,250,001–18,490,000 | 11 | *LB and SD vs. ET-G2* | $F_{ST}$, *XP-EHH* | *NF1****, EVI2A, EVI2B, OMG*** | Associated with tail morphology and fat deposition [38,49,93–95] |
| 26,350,001–26,540,000 | 11 | *SD vs. ET-G1, ET-G2 and LB* | $F_{ST}$ | ***ALOX12****, RNASEK, U6, BCL6B, SLC16A13, SLC16A11, CLEC10A, ASGR2, ASGR1* | Associated with fat deposition [38,49,93,94]. |
| 37,200,001–37,430,000 | 11 | *SD and ET-G2 vs. ET-G1 and LB* | $F_{ST}$, *XP-EHH* | ***HOXB13****, CALCOCO2, TTLL6, RF02133, HOXB9, HOXB7, HOXB6, HOXB5* | A strong candidate gene controlling tail length in mice [39]. |

*EPAS1, EGLN1, CHRNA4, VEGFA, NF1* and *ADSL*; and the ones associated with "positive regulation of angiogenesis" were *VEGFA* and *FGF2*. The top-two most significant KEGG pathways were "tuberculosis (oas05152)" and "HIF-1 signalling (oas04066)". Genes associated with the former were *IL18, PLK3, TLR1, CATHL3, CALML4, BAC5, TLR6* and *SC5*, and the ones associated with the latter were *TF, EGLN1, VEGFA* and *PLCG1* (Fig. S7).

We explored haplotype structures around *EGLN1* and *TF* (Fig. 3a and b), which have been reported in previous studies to be significant for high altitude adaptation. The analysis revealed the presence of a haplotype that was fixed in ET-G2 (frequency = 1.00) and approaching fixation in SD (frequency = 0.95) but is still segregating in LB (frequency = 0.625).

### 3.5.2. Selection signatures for low altitude hot-dry desert (Sudan), and coastal hot-humid/arid (Libya) environments

We contrasted both SD and LB with ET-G2 to identify selection signatures for adaptation to low-altitude hot-dry desert and coastal hot-humid arid African environments (Fig. S8). The three tests revealed 123 candidate regions spanning 182 genes (Table S12). Two regions were identified by *Hp* and $F_{ST}$, six by *Hp* and *XP-EHH*, three by *XP-EHH* and $F_{ST}$, two by all the three methods and one by *XP-EHH*. The regions identified by the three methods occurred on OAR6 and OAR11. The region on OAR6 spanned no genes, while the one on OAR11 spanned four genes, *NF1, EVI2A, EVI2B* and *OMG* (Fig. S8). Another candidate region supported by *Hp* and $F_{ST}$ occurred on OAR2 and spanned *DIS3L2* (Fig. S8). The haplotype structure around this gene showed a haplotype that was approaching fixation in LB and SD (Fig. 4a). Another candidate region was found on OAR3 overlapping *MSRB3* (Fig. S8 and 4b). The ET-G2 represented by Doyogena, which has vestigial ears, showed a large conserved haplotype around *MSRB3* that is absent in LB and SD sheep with long drooping ears (Fig. 4b). Several LB-specific candidate regions were also revealed. One on OAR7 spanned *GLDN* and *CYP19*. Another on OAR15 overlapped *PLEKHA7* and *C11orf58* with a distinct haplotype (Fig. 4c). There were two other regions spanning distinct haplotypes around *PDS5B* and *RXFP2* genes found 570 kb apart on OAR10 (Fig. 4d). In total, 182 genes in 123 regions (Table S12) were identified and used for functional enrichment analysis. They yielded 14 highly significant GO terms and KEGG pathways (Table S13). The top-most highly significant KEGG pathways included "Aldosterone synthesis and secretion (oas04925)" and "Vasopressin-regulated water reabsorption (oas04962)". The candidate genes associated with low altitude hot-dry desert (Sudan), and coastal hot-humid arid (Libya) environment adaptation are presented at Table 2.

### 3.5.3. Signatures of selection for differences in tail skeleton length and fat-depot sizes

To identify candidate regions associated with tail skeleton length, we took SD and ET-G2 to represent long-tailed sheep and LB and ET-G1 to represent short-tailed sheep, based on the lengths of individual CV's and the complete tail skeleton. We analysed selection signatures with *Hp* (Fig. S9), $F_{ST}$ (Fig. S10) and *XP-EHH* (Fig. S11 and Fig. 5) and identified 68 candidate regions across 21 autosomes (Table S14). Two candidate regions were identified by the three tests and 15 regions by at least two tests. The remaining 51 regions were identified by only one test; they included 29 ($F_{ST}$) and 22 (*XP-EHH*). Annotation of these candidate regions identified 225 genes (Table S14). The top-most skeletal tail-formation related candidate genes are shown in Table 3. The first region identified by all the three tests occurred on OAR3 and spanned six genes (*GAPVD1, HSPA5, RABEPK, PPP6C, SCAI, ENSOARG00000025028*). The second occurred on OAR6 spanning no genes. One strong candidate region overlapping eight genes (*CALCOCO2, TTLL6, HOXB13, RF02133, RF02132, MIR196A1, HOXB9, HOXB7*) was identified on OAR11 by $F_{ST}$ and *XP-EHH* (Table 3; Figs. S9–S11). Of the eight genes, *HOXB13* was reported to be a candidate for controlling skeletal tail length in mice [39]. An analysis of variation around this gene (Fig. 6) revealed a conserved haplotype with a frequency of 0.84 and 0.75 in ET-G1 and LB, respectively. A second strong candidate region overlapping seven genes (*GOLGA1, ARPC5L, WDR38, OLFML2A, NR5A1, NR6A1, ADGRD2*) occurred on OAR3 (Table 3; Figs. S9–S11). Two of the genes (*NR5A1* and *NR6A1*) have been associated with vertebrae number [40].

To identify selection signatures for tail fat-depot sizes, we contrasted SD (thin-tailed) with ET-G1 (fat-rumped), and LB and ET-G2 (both fat-tailed). The three tests (*Hp*, $F_{ST}$, *XP-EHH*) identified 34 candidate regions overlapping 122 genes. Twelve regions were identified by at least one test and 25 by a combination of at least two tests (Table S15; Fig. S12). Table 3 shows the candidate genes associated with fat deposition and which were present within the candidate regions identified by our analysis.

We used all the 271 SNPs present in the 169.9 Kb selection sweep region that is specific to the LB sheep and spanning the *PLEKHA7* gene on OAR15 to build a haplotype network using the 55 most common haplotypes (Fig. S13). It reveals SD and LB, which are living in arid environments (hot humid and hot dry), cluster in two groups that are separated by one haplotype (H_4) (Fig. S13) within the same lineage which is separated from the Ethiopian sheep. It suggests that this region might have been selected for the adaptation to the arid environments.

## 4. Discussion

### 4.1. Population structure and demographic history

Human and natural selection have shaped the genomes of domestic livestock over millennia. Here we investigated the genome architecture of indigenous African sheep, by analysing whole-genome sequences of 150 indigenous sheep from three northeast African countries. 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 and 55.9% were intergenic. This is consistent with observations in Eurasian wild aurochs [41] and Chinese native

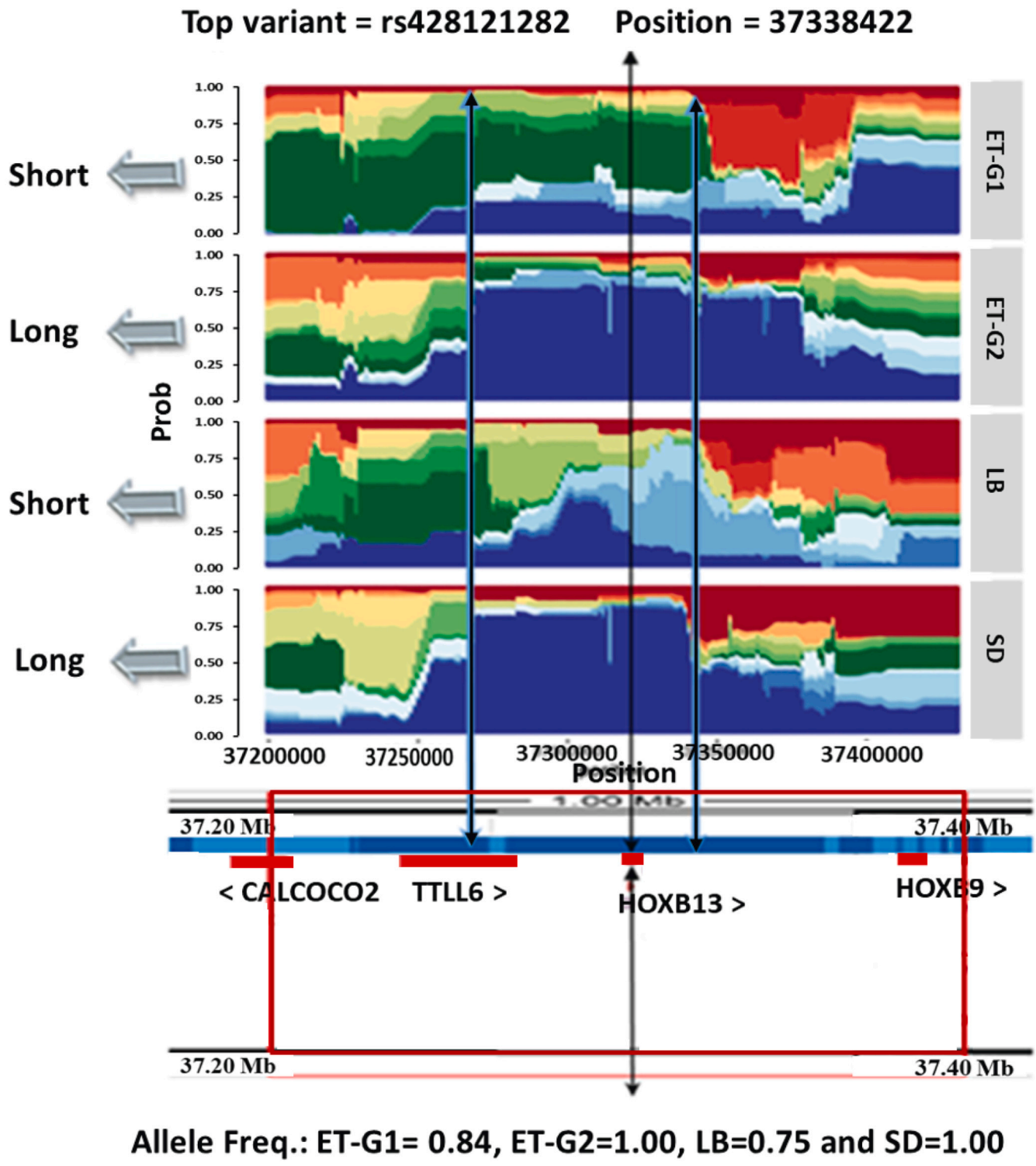


**Fig. 6.** The haplotype structure around the candidate region on OAR11 spanning *HOXB13* likely associated with vertebrae length (Long = Long CVs, Short = Short CVs) *(Black and Blue arrows indicate the most distinguished haplotype within the candidate gene region)*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sheep [42], indicating high quality and reliability of our sequences and the called SNPs.

The history of indigenous African sheep has been inferred from archaeological findings, but the picture still remains incomplete. Our phylogenetic inference identified four distinct genetic groups underlying the genetic structure of our studied populations. These genetic groups are offering interesting insights into the historical diffusion and dispersal of sheep into and across the African continent. The North African Libyan Barberine (LB) and the northeast African Sudanese Hammari and Kabashi (SD) sheep clustered in close proximity but separate from each other and from the East African sheep represented by populations from Ethiopia. The latter separated into two genetic groups (ET-G1 and ET-G2). Previous findings based on the analysis of autosomal microsatellite markers showed a closer relationship between the

Egyptian fat-tailed Ossimi sheep and the West African thin-tailed sheep than with any East and southern Africa fat-tailed sheep [43]. The analysis of 50 K SNP genotype data also revealed a clear separation of Egyptian fat-tailed sheep from East African sheep [44]. Our result therefore supports, using whole genome sequence data, two distinct entry points of sheep into the continent, the Isthmus of Suez (LB and SD) and the Horn of Africa (ET-G1 and ET-G2). The sub-grouping of Ethiopian sheep into ET-G1 (fat-rumped) and ET-G2 (long fat-tailed) is indicative of separate expansion events in the Horn of Africa. Based on the current dataset, it is however difficult to infer whether the ancestral populations of ET-G1 and ET-G2 arrived concurrently. The geographic distribution of the Ethiopian populations in each group offers indirect insights. ET-G1 populations are found in areas close to the Red Sea and the Indian Ocean coastline, while the ET-G2 populations occur in the hinterland. The ancestral populations of these two groups may thus have arrived independently; the arrival of the long fat-tailed preceding the fat-rumped sheep.

Three ancestral backgrounds were supported by ADMIXTURE. All three are present in Sudanese sheep which also show the highest $N_e$ (~2720) 1000 generations ago. The wild ancestor of domestic sheep was thin-tailed, but its tail length was shorter than that of its domestic variant [45,46]. Accordingly, as supported by archaeological thin-tailed iconography on the African continent, the thin-tailed sheep are likely to be the most ancient of the African sheep lineages [47]. The Sudanese and Libyan sheep shared a large proportion of their genome ancestry and show close genetic proximity on the PCA. They likely represent the modern-day legacy of the ancient arrival of thin-tailed sheep on the African continent followed by their dispersion along the Mediterranean Sea coast (Libya) and southwards along the Nile River Basin (Sudan) and into West Africa. Although the ET-G2 and LB sheep are both fat-tailed, they have divergent ancestral backgrounds. It supports two entry points for African fat-tailed sheep on the continent *i.e.*, the Isthmus of Suez and the Horn of Africa.

The ADMIXTURE analysis showed Doyogena, ShubiGemo and Gafera, of ET-G2 group, share ~15–20% of their genome ancestry with ET-G1. Likewise, we detect the main ancestry of ET-G2 in three populations of ET-G1 (Menz, Adane, Arabo). This pattern of intermixing can be due to flock movements/exchanges due to human socio-cultural and economic interactions [48]. This is supported by f4 and D-statistic tests which provided evidence of gene flow from fat-rumped sheep from eastern Ethiopia into long fat-tailed sheep in central and southern Ethiopia. A similar pattern of admixture was observed between six Ethiopian populations (Menz, Gafera, Gesses, Kido, Adane, Arabo) and Sudanese (SD) sheep. This admixture pattern is consistent with previous findings [38]. Half of the individuals of LB shared ~30% of their genome ancestry with those of SD and ET-G1. We suggest this shared ancestry is the result of inter-crossing fat-tailed with thin-tailed sheep to reduce the volume of tail fat in the former due to the decline in the monetary value of fat depots [49,50]. Such inter-crossings have been reported in Algeria [51] and Tunisia [52]. It is worth noting that large fat tails can account for up to 14.5% of the cold dressed carcass weight in sheep [53]. The shared ancestries between the northeast African SD and North African LB sheep in line with the equal number of tail vertebrae (Fig. 1a) and different tail sizes may support such a pattern of intermixing. On the other hand, various ancestries and varying numbers of tail vertebrae paired with diverse tail sizes might imply a different history within Ethiopian sheep (Fig. S5b). At higher values of *K*, Bonga, Doyogena and Molale showed genetically homogeneous but divergent backgrounds. This result is rather difficult to explain but we speculate that a combination of natural and human selection could be driving their divergence.

### 4.2. Genome targets for adaptation to environments

The African landscape has pronounced changes in elevation ranging from sea level to high altitudes in the East African highlands. The variation in altitude is accompanied by large variations in agro-climatic conditions ranging from dry and arid to wet and humid. Such variations in topography and environmental conditions allow studying adaptive processes in response to environmental selection pressure, *e.g.*, hypobaric hypoxia in high altitudes, and heat, water and feed stress in the drylands. It provides opportunities to investigate how natural selection shapes the genetic architecture of adaptive divergence. Here, we sampled and analysed genome variation in 15 populations of indigenous sheep from a range of ecological and altitudinal geographic regions in Northeast and North Africa. We observed a separation of the populations into four genetic groups that align well with geography. Our comparative genomic analyses involving the four genetic groups revealed several important observations that shed new insights on their biology of adaptation.

For example, hypobaric hypoxia imposes major stresses on the physiology of organisms at high altitudes and a powerful homeostasis system has evolved to offset ambient hypoxia and attain physiological homeostasis [54]. The classic response is an increase in haemoglobin concentration to compensate for the unavoidable lowered percentage of oxygen saturation. Our selective sweep mapping involving ET-G2 identified candidate regions spanning nine genes, *TF* (OAR1), *EPAS1* and *ADSL* (OAR3), *NF1* (OAR11), *PLCG1* and *CHRNA4* (OAR13), *FGF2* (OAR17), *VEGFA* (OAR20), and *EGLN1* (OAR25). These genes comprised the two most significant GO terms, “response to hypoxia” and “positive regulation of angiogenesis”, and KEGG pathways, “HIF-1 signalling pathway” and “pathways in cancer”, suggesting they may underlie adaptive changes to high-altitude environments. Indeed, *EGLN1* and *EPAS1* encode the synthesis of prolyl hydroxylase domain 2 (*PHD2*) that plays a critical role in adapting to changes in oxygen concentrations and its variation has consistently been associated with haemoglobin levels amongst high altitude dwelling Tibetans [55]. It is also a target of natural selection for hypoxia adaptation in vertebrates [56,57]. The HIF-1 pathway plays a central role in regulating cellular responses to hypoxia [58] at high altitudes (see review by [59]). The *NF1* gene, which promotes and regulates angiogenesis in response to hypoxia on its own [60] or as a component of the Ras/ERK and VEGF signalling pathways [61], occurred in a candidate region under selection in sheep living at high altitudes [42,62]. *VEGFA* has been associated with cerebrovascular adaptation to chronic hypoxia in sheep [63] and protects pregnant ewes from hypoxic effects by increasing blood flow in uterine arteries [64]. Under hypoxic conditions, elevated expression of *TF* and *VEGFA* increases oxygen delivery by stimulating iron metabolism and angiogenesis, following positive feedback from *EGLN1*.

Unexpectedly, our analytical strategy to identify genomic regions associated with differences in tail length and tail fat-depot sizes also identified selection targets related to high altitude adaptation. For instance, the gene *MSRB3* occurred on a candidate region on OAR3 that was supported by both $F_{ST}$ and *XP-EHH*. This gene has been associated with high altitude adaptation in dogs [65] and Tibetan sheep [62]. Our results appear to suggest that natural selection, driven by a hypoxic environment in the Ethiopian highlands, 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 cattle [66] and sheep [67,68] in Ethiopia did not find similar associations. These contrasting findings are rather surprising, but we speculate that they could be the result of differences in sampling strategies, analytical approaches and markers used. For instance, the studies by [66,67,68] classified their populations based on geographic sampling. [68] also used an ecological approach to analyse their data. In contrast, our analysis was informed by the genetic groups revealed by phylogenetic inference.

Haplotype analysis around the candidate regions associated with high altitude adaptation revealed some unexpected results. The haplotype structures around the *TF* locus shows a homogeneous haplotype that is approaching fixation in ET-G2, but several segregating haplotypes in SD and LB sheep (Fig. 3b), suggesting selection around the locus

in ET-G2. Haplotype structure analysis around *EGLN1* showed that SD shared identical haplotypes with ET-G2, that were approaching fixation, against an admixed haplotypes background in LB. But LB and ET-G2 shared a similar selection signature around *NF1* gene (Table 2; Fig. S6 and S8). In Egyptian sheep that are adapted to a dryland environment, [44] found candidate regions spanning genes that were highly enriched for the GO term "response to hypoxia" and "HIF-1 signaling pathway". They suggested this could be an adaptive response to physical exhaustion due to long-term long-distance trekking in search of pasture and water. Such an activity results in hypoxia-like conditions and oxygen debt in skeletal muscles. This may suggest that parallel evolution driven by distinct natural selection pressures could be acting on the same set of genes driving adaptation to different environments in the same species. This calls for further investigations.

High ambient temperatures resulting in thermal stress, represent a physiological challenge in hot arid environments. Thermal stress impacts production and reproduction and desert dwelling animals have developed physiological responses to mitigate thermal stress and its negative effects. The analysis involving SD and LB sheep groups identified 123 candidate regions spanning 182 genes (Table S12) of which 19 were reported in candidate regions in desert/arid lands dwelling Chinese sheep [42]. Among the genes identified in a candidate region on OAR11 is *KCNAB3*. This gene encodes a potassium voltage-gated channel subunit linked to cognitive performance under chronic stress [69] and it was found to be down-regulated in response to acute and chronic stress in mice [70]. This is of relevance given that LB and SD sheep are exposed to extended periods of thermal and physical stress in Libya and Sudan, respectively. Apart from thermal stress, high temperatures constrain pasture and water availability. One of the candidate regions, where SD and LB shared the same haplotype, was on OAR2 which spanned *DIS3L2* (Fig. 4a). This gene has been associated with height variation and was found to be a candidate under selection in cattle [71] and Brazilian sheep [72]. Given that large animals have higher maintenance requirements [73], 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 [74].

Thermal stress affects male reproductive performance (see review by [75]). We found *CNTROB* (centrobin, for centrosome BRCA2 interacting protein) in a candidate region on OAR11 in both SD and LB. *CNTROB* has been shown to be critical in maintaining male fertility [76]. One region on OAR10 differentiating LB from ET-G2 and SD and with a distinct haplotype approaching fixation in LB (Fig. 4d) spanned *PDS5B* (PDS5 cohesin associated factor B). This gene modulates cohensin functions in spermatocytes and spermatogonia by contributing to meiotic chromosome function and structure [77]. By occurring in a selective sweep region in dryland dwelling sheep, *CNTROB* and *PDS5B* may be critical in maintaining male fertility under extended periods of heat stress.

Soil salinisation contributes to soil degradation and compromises water and pasture quality [78]. Droughts, which exacerbate soil salinity, have increased over the years due to global warming [79,80] and worsened in arid and semi-arid environments [78]. 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 (Fig. 4c). The MJ network analysis around this gene also identified only one haplotype that anchored SD and LB within one lineage (Fig. S13). *PLEKHA7* is highly expressed in kidney and heart [81] and it was reported to attenuate salt-sensitive hypertension and renal diseases in rats [82]. In humans, *PLEKHA7* has been associated with blood pressure, kidney function and hypertension [83,84]. 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 [85]. 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 [86] and hypertension [87] in humans. *NR5A1* plays a critical role in the expression of *CYP11B2* [88] which plays a key role in salt-sensitive hypertension development [89] and in aldosterone biosynthesis in the adrenal cortex [90].

### 4.3. 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, *TTLL6, HOXB13, HOXB9* and *HOXB7* have been associated with growth and development of the spinal cord and tail vertebrae in several animal species [38,39,91,92]. It occurred upstream of two candidate regions spanning five genes (*NF1, EVI2A, EVI2B, OMG* and *ALOX12*) (Fig. 5 and S10) that are associated with fat deposition [38,49,93–95]. So at least two regions that are adjacent to each other on OAR11 span genes associated with two tail traits (fat-depot size and CV length). One region on OAR3 spanned *NR5A1* and *NR6A1*, which influence the number of vertebrae in pigs [96,97] and thoracic vertebrae in sheep [98], repectively. A candidate region spanning *PDGFRA* (platelet derived growth factor receptor alpha) was also identified by the three methods on OAR6 in LB; a distinct haplotype was found around this gene. *PDGFRA* is involved in preadipocyte differentiation and was also found in a candidate region associated with the fat-tailed phenotype in sheep [95,99]. Other than *PDGFRA*, a number of genes including *PDGFD* and *BMP2*, have been implicated in fat tail phenotype and *TBXT* has been linked to caudal vertebrae number and tail length [1]. These results suggest a possible scenario of co-evolution between tail skeleton morphology and tail fat depots in sheep.

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 sheep [94], dogs [100,101] and pigs [102]. 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 sheep, the majority of which have long drooping ears, the ET-G2 represented by Doyogena sheep with short/vestigial ears, 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 hot environments show longer extremities [103], suggesting such differences are indicative of environmental adaptation.

LB sheep differentiated from ET-G2 at two candidate regions on OAR10. One of these spans *RXFP2* (relaxin family peptide receptor 2), which determines horn size and development in ungulates [104] and primary sex characters in humans and mice [105]. Heat-exchange capacity shows an evolution trend of variation in horn core size in response to ambient temperatures [106]. Although the selective sweep spanning *RXFP2* could implicate its dual role in the development of horns for thermoregulation and enhancing male reproduction under thermal stress, in recent times sheep farmers in Libya are preferring breeding rams with larger horns and tail sizes because such rams are perceived to sire strong and healthier progenies (Fig. S14). The selection signature around *RXFP2* may be driven by human selection for horn size and natural selection for thermoregulation function, although the semi-feral nature of management of LB sheep could also be implicated as suggested for Chinese sheep [107].

## 5. Conclusion

Our results reveal four autosomal genomic backgrounds in indigenous northeast African sheep. The genotypes of most of the individuals analysed comprised of at least two genetic backgrounds which could be due to some level of current and/or historical intermixing. Our findings

corroborate archaeological pieces of evidence that have revealed domestic sheep migration events and entry points into Africa. Besides, our results support possible interaction between environmental factors and tail morphology in shaping the genomes of indigenous sheep inhabiting diverse environments. From the perspective of environmental adaptation, our studied populations illustrated the reshaping of the genome when facing diverse environmental challenges (coastal hot-humid arid, hot-dry desert and high-altitude humid). Our whole-genome sequence analysis also suggest interaction between several genomic regions for tail phenotypes in sheep. For instance, many candidate genes possibly under selection were identified, among which, the salt-sensitivity related gene (*PLEKHA7*) in Libyan sheep, thermoregulation genes (*CREB3L2, CREB3, GNAQ, DCTN4*) in Libyan and Sudanese sheep, hypoxia associated genes (*EGLN1*, *EPAS1*) in Ethiopian sheep, and skeletal tail associated genes (*HOXB13, NR5A1*, *NR6A1, KDM6B*, *NR3C1*) most likely are playing major roles in the unique adaptations of African sheep. To the best of our knowledge, we are the first to report these candidate genes in African indigenous sheep. However, future studies including thin-tailed sheep from Ethiopia and both thin- and/or fat-tailed sheep from West, northern and southern Africa could provide a comprehensive overview on the genome architecture of African indigenous sheep.

## Data accessibility

Data of the Ethiopian and Libyan sheep are publicly available at NCBI. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA523711/. The Sudanese sheep sequences are available at https://www.ncbi.nlm.nih.gov/object/PRJNA849626/.

## Ethics statement

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

## Author contributions

AMA, OH, and JMM conceived and designed the study. AMA analysed the data and wrote the initial draft manuscript with inputs from JMM. OH and JMM revised the manuscript. EC and CR contributed to sequencing and mapping data of Ethiopian and Libyan sheep and provided critical inputs on data analysis. HM contributed to sampling and sequencing Sudanese sheep. ASA contributed to haplotype network analysis. AA, SL and MOA contributed and coordinated sampling in Ethiopia and Libya, respectively. All authors read and approved the final manuscript.

## Authors statement

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript. The name of each author must appear at least once in each of the three categories below:

**Category 1 Conception and design of study:** Abulgasim M. Ahbara, Olivier Hanotte, Joram M. Mwacharo; **acquisition of data**: Hassan H. Musa, Christelle Robert, Emily Clark, Ayele Abeba, Adebabay Kebede, Suliman Latairish, Mukhtar Omar Agoub, Ahmed S. Al-Jumaili; **analysis and/or interpretation of data:** Abulgasim M. Ahbara, Joram M. Mwacharo, Olivier Hanotte.

**Category 2 Drafting the manuscript:** Abulgasim M. Ahbara, Hassan H. Musa; **revising the manuscript critically for important intellectual content:** Olivier Hanotte, Joram M. Mwacharo.

**Category 3 Approval of the version of the manuscript to be published:** Abulgasim M. Ahbara, Hassan H. Musa, Christelle Robert, Ayele Abeba, Ahmed S. Al-Jumaili, Adebabay Kebede, Suliman Latairish, Mukhtar Omar Agoub, Emily Clark, Olivier Hanotte, Joram M. Mwacharo.

## Declaration of Competing Interest

The authors declare no conflicts of interest.

## Data availability

Data will be made available on request.

## Acknowledgments

This study was conducted during AMA's PhD training supported by the Libyan Ministry of Higher Education and Scientific Research and the University of Misurata. Sampling Ethiopian sheep was supported by the CGIAR Research Program on Livestock (Livestock CRP). Accordingly, ICARDA and ILRI 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).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2022.110448.

## References

[1] P. Kalds, Q. Luo, K. Sun, S. Zhou, Y. Chen, X. Wang, Trends towards revealing the genetic architecture of sheep tail patterning: promising genes and investigatory pathways, Anim. Genet. 52 (2021) 799–812, https://doi.org/10.1111/age.13133.

[2] V. Porter, Mason's World Dictionary of Livestock Breeds, Types and Varieties, 6th ed., CABI, Wallingford, UK, 2020.

[3] A. Gautier, The evidence for the earliest livestock in North Africa: Or adventures with large Bovids, Ovicaprids, dogs and pigs, in: H. Fekri (Ed.), Droughts, Food and Culture, KLUWER ACADEMIC PUBLISHERS, London, 2006, pp. 195–207, https://doi.org/10.1007/0-306-47547-2_12.

[4] F. Marshall, The origins and spread of domestic animals in East Africa, in: R. M. Blench, K.C. MacDonald (Eds.), The Origins and Development of African Livestock: Archaeology, Genetics, Linguistics and Ethnography, University College London Press, London, UK, 2000.

[5] J. Clutton-Brock, The spread of domestic animals in Africa, in: T. Shaw, P. Sinclair, B. Andah, A. Okpoko (Eds.), The Archaeology of Africa: Foods, Metals, and Towns, Routledge, London, UK, 1993.

[6] J.W. Barthelme, Fisher-Hunters and Neolithic Pastoralists in East Turkana, Kenya, BAR, Oxford, 1985.

[7] K.C. MacDonald, R.H. MacDonald, The origins and development of domesticated animals in arid West Africa, in: K.C. MacDonald, R.M. Blench (Eds.), The Origins and Development of African Livestock: Archaeology, Genetics, Linguistics and Ethnography, Routledge, New York, 2000, pp. 127–162.

[8] J.D. Clark, M.A.J. Williams, Recent archaeological research in Southeastern Ethiopia. 1974–1975, in: Annales d'Ethiopie, 1978, 1978, pp. 19–44, https://doi.org/10.3406/ethio.1978.902.

[9] D. Gifford-Gonzalez, O. Hanotte, Domesticating animals in Africa: implications of genetic and archaeological findings, J. World Prehist. 24 (2011) 1–23, https://doi.org/10.1007/s10963-010-9042-2.

[10] K. Kim, T. Kwon, T. Dessie, D.A. Yoo, O.A. Mwai, J. Jang, S. Sung, S.B. Lee, B. Salim, J. Jung, H. Jeong, G.M. Tarekegn, A. Tijjani, D. Lim, S. Cho, S.J. Oh, H. K. Lee, J. Kim, C. Jeong, S. Kemp, O. Hanotte, H. Kim, The mosaic genome of indigenous African cattle as a unique genetic resource for African pastoralism, Nat. Genet. 52 (2020) 1099–1110, https://doi.org/10.1038/s41588-020-0694-2.

[11] A.A. Gheyas, A. Vallejo-Trujillo, A. Kebede, M. Lozano-Jaramillo, T. Dessie, J. Smith, O. Hanotte, Integrated environmental and genomic analysis reveals the drivers of local adaptation in African indigenous chickens, Mol. Biol. Evol. 38 (2021) 4268–4285, https://doi.org/10.1093/molbev/msab156.
[12] H. Li, R. Durbin, Fast and accurate short read alignment with burrows-Wheeler transform, Bioinformatics. 25 (2009) 1754–1760, https://doi.org/10.1093/bioinformatics/btp324.
[13] D. Zhi, L. Da, M. Liu, C. Cheng, Y. Zhang, X. Wang, X. Li, Z. Tian, Y. Yang, T. He, X. Long, W. Wei, G. Cao, Whole genome sequencing of Hulunbuir short-tailed sheep for identifying candidate genes related to the short-tail phenotype, G3: Genes Genomes Genet. 8 (2018) 377–383, https://doi.org/10.1534/g3.117.300307.
[14] P. Danecek, A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R. E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, G. McVean, R. Durbin, The variant call format and VCFtools, Bioinformatics. 27 (2011) 2156–2158, https://doi.org/10.1093/bioinformatics/btr330.
[15] V. Narasimhan, P. Danecek, A. Scally, Y. Xue, C. Tyler-Smith, R. Durbin, BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data, Bioinformatics. 32 (2016) 1749–1751, https://doi.org/10.1093/bioinformatics/btw044.
[16] R. McQuillan, A.L. Leutenegger, R. Abdel-Rahman, C.S. Franklin, M. Pericic, L. Barac-Lauc, N. Smolej-Narancic, B. Janicijevic, O. Polasek, A. Tenesa, A. K. MacLeod, S.M. Farrington, P. Rudan, C. Hayward, V. Vitart, I. Rudan, S. H. Wild, M.G. Dunlop, A.F. Wright, H. Campbell, J.F. Wilson, Runs of homozygosity in European populations, Am. J. Hum. Genet. 83 (2008) 359–372, https://doi.org/10.1016/j.ajhg.2008.08.007.
[17] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, P. Sklar, P.I.W. de Bakker, M.J. Daly, P.C. Sham, PLINK: a tool set for whole-genome association and population-based linkage analyses, Am. J. Hum. Genet. 81 (2007) 559–575, https://doi.org/10.1086/519795.
[18] D.H. Huson, D. Bryant, Application of phylogenetic networks in evolutionary studies, Mol. Ecol. Notes 23 (2006) 254–267, https://doi.org/10.1093/molbev/msj030.
[19] S. Guindon, O. Gascuel, A. Simple, Fast, and accurate algorithm to estimate large phylogenies by maximum likelihood, Syst. Biol. 52 (2003) 696–704, https://doi.org/10.1080/10635150390235520.
[20] D. Darriba, G.L. Taboada, R. Doallo, D. Posada, JModelTest 2: more models, new heuristics and parallel computing, Nat. Methods 9 (2012) 772, https://doi.org/10.1038/nmeth.2109.
[21] D.H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals, Genome Res. 19 (2009) 1655–1664, https://doi.org/10.1101/gr.094052.109.
[22] R. Buchmann, S. Hazelhurst, Genesis manual, Bioinf.Wits.Ac.Za. http://www.bioinf.wits.ac.za/software/genesis/Genesis.pdf, 2014 (accessed August 15, 2021).
[23] N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, D. Reich, Ancient admixture in human history, Genetics. 192 (2012) 1065–1093, https://doi.org/10.1534/genetics.112.145037.
[24] M. Malinsky, M. Matschiner, H. Svardal, Dsuite - fast D-statistics and related admixture evidence from VCF files, Mol. Ecol. Resour. 21 (2021) 584–595, https://doi.org/10.1111/1755-0998.13265.
[25] M. Barbato, P. Orozco-terWengel, M. Tapio, M.W. Bruford, SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data, Front. Genet. 6 (2015), https://doi.org/10.3389/fgene.2015.00109.
[26] J.A. Sved, Linkage disequilibrium and homozygosity of chromosome segments in finite populations, Theor. Popul. Biol. 2 (1971) 125–141, https://doi.org/10.1016/0040-5809(71)90011-6.
[27] B.J. Hayes, P.M. Visscher, H.C. McPartlan, M.E. Goddard, Novel multilocus measure of linkage disequilibrium to estimate past effective population size, Genome Res. 13 (2003) 635–643, https://doi.org/10.1101/gr.387103.
[28] C.J. Rubin, M.C. Zody, J. Eriksson, J.R.S. Meadows, E. Sherwood, M.T. Webster, L. Jiang, M. Ingman, T. Sharpe, S. Ka, F. Hallböök, F. Besnier, R. Carlborg, B. Bedhom, M. Tixier-Boichard, P. Jensen, P. Siegel, K. Lindblad-Toh, L. Andersson, Whole-genome resequencing reveals loci under selection during chicken domestication, Nature. 464 (2010) 587–591, https://doi.org/10.1038/nature08832.
[29] B.S. Weir, C.C. Cockerham, Estimating F-statistics for the analysis of population structure, Evolution (N Y). 38 (1984) 1358–1370, https://doi.org/10.1111/j.1558-5646.1984.tb05657.x.
[30] P.C. Sabeti, P. Varilly, B. Fry, J. Lohmueller, E. Hostetter, C. Cotsapas, X. Xie, E. H. Byrne, S.A. McCarroll, R. Gaudet, S.F. Schaffner, E.S. Lander, K.A. Frazer, D. G. Ballinger, D.R. Cox, D.A. Hinds, L.L. Stuve, R.A. Gibbs, J.W. Belmont, A. Boudreau, P. Hardenbol, S.M. Leal, S. Pasternak, D.A. Wheeler, T.D. Willis, F. Yu, H. Yang, C. Zeng, Y. Gao, H. Hu, W. Hu, C. Li, W. Lin, S. Liu, H. Pan, X. Tang, J. Wang, W. Wang, J. Yu, B. Zhang, Q. Zhang, H. Zhao, H. Zhao, J. Zhou, S.B. Gabriel, R. Barry, B. Blumenstiel, A. Camargo, M. Defelice, M. Faggart, M. Goyette, S. Gupta, J. Moore, H. Nguyen, R.C. Onofrio, M. Parkin, J. Roy, E. Stahl, E. Winchester, L. Ziaugra, D. Altshuler, Y. Shen, Z. Yao, W. Huang, X. Chu, Y. He, L. Jin, Y. Liu, Y. Shen, W. Sun, H. Wang, Y. Wang, Y. Wang, X. Xiong, L. Xu, M.M.Y. Waye, S.K.W. Tsui, H. Xue, J.T.F. Wong, L.M. Galver, J. B. Fan, K. Gunderson, S.S. Murray, A.R. Oliphant, M.S. Chee, A. Montpetit, F. Chagnon, V. Ferretti, M. Leboeuf, J.F. Olivier, M.S. Phillips, S. Roumy, C. Sallée, A. Verner, T.J. Hudson, P.Y. Kwok, D. Cai, D.C. Koboldt, R.D. Miller, L. Pawlikowska, P. Taillon-Miller, M. Xiao, L.C. Tsui, W. Mak, Q.S. You, P.K. H. Tam, Y. Nakamura, T. Kawaguchi, T. Kitamoto, T. Morizono, A. Nagashima, Y. Ohnishi, A. Sekine, T. Tanaka, T. Tsunoda, P. Deloukas, C.P. Bird, M. Delgado, E.T. Dermitzakis, R. Gwilliam, S. Hunt, J. Morrison, D. Powell, B.E. Stranger, P. Whittaker, D.R. Bentley, M.J. Daly, P.I.W. de Bakker, J. Barrett, Y.R. Chretien, J. Maller, S. McCarroll, N. Patterson, I. Pe'Er, A. Price, S. Purcell, D.J. Richter, R. Saxena, P.C. Sham, L.D. Stein, L. Krishnan, A.V. Smith, M.K. Tello-Ruiz, G. A. Thorisson, A. Chakravarti, P.E. Chen, D.J. Cutler, C.S. Kashuk, S. Lin, G. R. Abecasis, W. Guan, Y. Li, H.M. Munro, Z.S. Qin, D.J. Thomas, G. McVean, A. Auton, L. Bottolo, N. Cardin, S. Eyheramendy, C. Freeman, J. Marchini, S. Myers, C. Spencer, M. Stephens, P. Donnelly, L.R. Cardon, G. Clarke, D. M. Evans, A.P. Morris, B.S. Weir, T.A. Johnson, J.C. Mullikin, S.T. Sherry, M. Feolo, A. Skol, H. Zhang, I. Matsuda, Y. Fukushima, D.R. MacEr, E. Suda, C. N. Rotimi, C.A. Adebamowo, I. Ajayi, T. Aniagwu, P.A. Marshall, C. Nkwodimmah, C.D.M. Royal, M.F. Leppert, M. Dixon, A. Peiffer, R. Qiu, A. Kent, K. Kato, N. Niikawa, I.F. Adewole, B.M. Knoppers, M.W. Foster, E. W. Clayton, J. Watkin, D. Muzny, L. Nazareth, E. Sodergren, G.M. Weinstock, I. Yakub, B.W. Birren, R.K. Wilson, L.L. Fulton, J. Rogers, J. Burton, N.P. Carter, C.M. Clee, M. Griffiths, M.C. Jones, K. McLay, R.W. Plumb, M.T. Ross, S.K. Sims, D.L. Willey, Z. Chen, H. Han, L. Kang, M. Godbout, J.C. Wallenburg, P. L'Archevêque, G. Bellemare, K. Saeki, H. Wang, D. An, H. Fu, Q. Li, Z. Wang, R. Wang, A.L. Holden, L.D. Brooks, J.E. McEwen, M.S. Guyer, V.O. Wang, J. L. Peterson, M. Shi, J. Spiegel, L.M. Sung, L.F. Zacharia, F.S. Collins, K. Kennedy, R. Jamieson, J. Stewart, Genome-wide detection and characterization of positive selection in human populations, Nature. 449 (2007) 913–918, https://doi.org/10.1038/nature06250.
[31] S.R. Browning, B.L. Browning, Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering, Am. J. Hum. Genet. 81 (2007) 1084–1097, https://doi.org/10.1086/521987.
[32] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics. 26 (2010) 841–842, https://doi.org/10.1093/bioinformatics/btq033.
[33] C.A. Maclean, N.P. Chue Hong, J.G.D. Prendergast, Hapbin: an efficient program for performing haplotype-based scans for positive selection in large genomic datasets, Mol. Biol. Evol. 32 (2015) 3027–3029, https://doi.org/10.1093/molbev/msv172.
[34] M.I. Fariello, S. Boitard, H. Naya, M. SanCristobal, B. Servin, Detecting signatures of selection through haplotype differentiation among hierarchically structured populations, Genetics. 193 (2013) 929–941, https://doi.org/10.1534/genetics.112.147231.
[35] H.J. Bandelt, P. Forster, A. Röhl, Median-joining networks for inferring intraspecific phylogenies, Mol. Biol. Evol. 16 (1999) 37–48, https://doi.org/10.1093/oxfordjournals.molbev.a026036.
[36] J.W. Leigh, D. Bryant, POPART: full-feature software for haplotype network construction, Methods Ecol. Evol. 6 (2015) 1110–1116, https://doi.org/10.1111/2041-210X.12410.
[37] D.W. Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (2009) 44–57, https://doi.org/10.1038/nprot.2008.211.
[38] A. Ahbara, H. Bahbahani, F. Almathen, M. al Abri, M.O. Agoub, A. Abeba, A. Kebede, H.H. Musa, S. Mastrangelo, F. Pilla, E. Ciani, O. Hanotte, J. M. Mwacharo, Genome-wide variation, candidate regions and genes associated with fat deposition and tail morphology in Ethiopian indigenous sheep, Front. Genet. 9 (2019) 699, https://doi.org/10.3389/fgene.2018.00699.
[39] K.D. Economides, L. Zeltser, M.R. Capecchi, Hoxb13 mutations cause overgrowth of caudal spinal cordand tail vertebrae, Dev. Biol. 256 (2003) 317–330, https://doi.org/10.1016/S0012-1606(02)00137-9.
[40] G. Yang, J. Ren, Z. Zhang, L. Huang, Genetic evidence for the introgression of Western NR6A1 haplotype into Chinese Licha breed associated with increased vertebral number, Anim. Genet. 40 (2009) 247–250, https://doi.org/10.1111/J.1365-2052.2008.01820.X.
[41] S.D.E. Park, D.A. Magee, P.A. McGettigan, M.D. Teasdale, C.J. Edwards, A. J. Lohan, A. Murphy, M. Braud, M.T. Donoghue, Y. Liu, A.T. Chamberlain, K. Rue-Albrecht, S. Schroeder, C. Spillane, S. Tai, D.G. Bradley, T.S. Sonstegard, B. J. Loftus, D.E. MacHugh, Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the phylogeography and evolution of cattle, Genome Biol. 16 (2015) 1–15, https://doi.org/10.1186/s13059-015-0790-2.
[42] J. Yang, W. Li, F. Lv, S. He, S.T.-M. biology, U., Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments, Mol. Biol. Evol. 33 (2016) (2016) 2576–2592. https://academic.oup.com/mbe/article-abstract/33/10/2576/2925566 (accessed August 14, 2021).
[43] A.W.T. Muigai, Characterisation and Conservation of Indigenous Animal Genetic Resources: The Fat-Tailed and Thin-Tailed Sheep of Africa, Ph.D. thesis, Jomo Kenyatta, University of Agriculture and Technology Juja, Kenya, 2003.
[44] J.M. Mwacharo, E.S. Kim, A.R. Elbeltagy, A.M. Aboul-Naga, B.A. Rischkowsky, M. F. Rothschild, Genomic footprints of dryland stress adaptation in Egyptian fat-Tail sheep and their divergence from East African and western Asia cohorts, Sci. Rep. 7 (2017) 1–10, https://doi.org/10.1038/s41598-017-17775-3.
[45] J. Deng, X. Xie, D. Wang, C. Zhao, F. Lv, X. Li, J.Y.-C. Biology, U., Paternal origins and migratory episodes of domestic sheep, Curr. Biol. 30 (2020) (2020) 4085–4095.e6. https://www.sciencedirect.com/science/article/pii/S0960982220311362 (accessed December 13, 2021).
[46] B. Chessa, F. Pereira, F. Arnaud, A. Amorim, F. Goyache, I. Mainland, R.R. Kao, J. M. Pemberton, D. Beraldi, M.J. Stear, A. Alberti, M. Pittau, L. Iannuzzi, M. H. Banabazi, R.R. Kazwala, Y.P. Zhang, J.J. Arranz, B.A. Ali, Z. Wang, M. Uzun, M.M. Dione, I. Olsaker, L.E. Holm, U. Saarma, S. Ahmad, N. Marzanov, E. Eythorsdottir, M.J. Holland, A.M. Paolo, M.W. Bruford, J. Kantanen, T. E. Spencer, M. Palmarini, Revealing the history of sheep domestication using

retrovirus integrations, Science 324 (2009) (1979) 532–536, https://doi.org/10.1126/SCIENCE.1170587/SUPPL_FILE/CHESSA.SOM.PDF.
[47] A.W.T. Muigai, O. Hanotte, The origin of African sheep: archaeological and genetic perspectives, Afr. Archaeol. Rev. 30 (2013) 39–50, https://doi.org/10.1007/s10437-013-9129-0.
[48] S. Gizaw, J.A.M. van Arendonk, H. Komen, J.J. Windig, O. Hanotte, Population structure, genetic variation and morphological diversity in indigenous sheep of Ethiopia, Anim. Genet. 38 (2007) 621–628, https://doi.org/10.1111/j.1365-2052.2007.01659.x.
[49] B. Moioli, F. Pilla, E. Ciani, Signatures of selection identify loci associated with fat tail in sheep, J. Anim. Sci. 93 (2015) 4660–4669, https://doi.org/10.2527/jas.2015-9389.
[50] M.H. Moradi, A. Nejati-Javaremi, M. Moradi-Shahrbabak, K.G. Dodds, J. C. McEwan, Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition, BMC Genet. 13 (2012) 1–15, https://doi.org/10.1186/1471-2156-13-10.
[51] S.B.S. Gaouar, M. Lafri, A. Djaout, R. El-Bouyahiaoui, A. Bouri, A. Bouchatal, A. Maftah, E. Ciani, A.B. da Silva, Genome-wide analysis highlights genetic dilution in Algerian sheep, Heredity (Edinb). 118 (2017) 293–301, https://doi.org/10.1038/hdy.2016.86.
[52] M. Rekik, R. Aloulou, M. ben Hammouda, Small ruminant breeds of Tunisia, in: L. Iniguez (Ed.), Characterisation of Small Ruminant Breeds in West Asia and North Africa, International Center for Agricultural Research in the Dry Areas (ICARDA), Beirut, 2005, pp. 91–140.
[53] M.R. Kiyanzad, Comparison of carcass composition of Iranian fat-tailed sheep, Asian-Aust, J. Anim. Sci. 18 (2005) 1348–1352, https://doi.org/10.5713/ajas.2005.1348.
[54] K.T. Rytkönen, J.F. Storz, Evolutionary origins of oxygen sensing in animals, EMBO Rep. 12 (2011) 3–4, https://doi.org/10.1038/embor.2010.192.
[55] T.S. Simonson, Y. Yang, C.D. Huff, H. Yun, G. Qin, D.J. Witherspoon, Z. Bai, F. R. Lorenzo, J. Xing, L.B. Jorde, J.T. Prchal, R.L. Ge, Genetic evidence for high-altitude adaptation in Tibet, Science 329 (2010) (1979) 72–75, https://doi.org/10.1126/science.1189406.
[56] J.F. Storz, H. Moriyama, Mechanisms of hemoglobin adaptation to high altitude hypoxia, High Alt. Med. Biol. 9 (2008) 148–157, https://doi.org/10.1089/ham.2007.1079.
[57] M.C. Tissot Van Patot, M. Gassmann, Hypoxia: adapting to high altitude by mutating EPAS-1, the gene encoding HIF-2α, High Alt. Med. Biol. 12 (2011) 157–167, https://doi.org/10.1089/ham.2010.1099.
[58] S. Frede, J. Fandrey, Cellular and Molecular Defenses against Hypoxia, in: High Altitude, Springer New York, New York, NY, 2014, pp. 23–35, https://doi.org/10.1007/978-1-4614-8772-2_2.
[59] J. Friedrich, P. Wiener, Selection signatures for high-altitude adaptation in ruminants, Anim. Genet. 51 (2020) 157–165, https://doi.org/10.1111/age.12900.
[60] M. Wu, R. Wallace, D. Muir, Nf1 haploinsufficiency augments angiogenesis, Oncogene. 25 (2006) 2297–2303, https://doi.org/10.1038/sj.onc.1209264.
[61] E. Minet, T. Arnould, G. Michel, I. Roland, D. Mottet, M. Raes, J. Remacle, C. Michiels, ERK activation upon hypoxia: involvement in HIF-1 activation, FEBS Lett. 468 (2000) 53–58, https://doi.org/10.1016/S0014-5793(00)01181-9.
[62] C. Wei, H. Wang, G. Liu, F. Zhao, J.W. Kijas, Y. Ma, J. Lu, L. Zhang, J. Cao, M. Wu, G. Wang, R. Liu, Z. Liu, S. Zhang, C. Liu, L. Du, Genome-wide analysis reveals adaptation to high altitudes in Tibetan sheep, Sci. Rep. 6 (2016) 1–11, https://doi.org/10.1038/srep26770.
[63] M. Castillo-Melendez, T. Yawno, B.J. Allison, G. Jenkin, E.M. Wallace, S.L. Miller, Cerebrovascular adaptations to chronic hypoxia in the growth restricted lamb, Int. J. Dev. Neurosci. 45 (2015) 55–65, https://doi.org/10.1016/j.ijdevneu.2015.01.004.
[64] V. Mehta, K.N. Abi-Nader, D.M. Peebles, E. Benjamin, V. Wigley, B. Torondel, E. Filippi, S.W. Shaw, M. Boyd, J. Martin, I. Zachary, A.L. David, Long-term increase in uterine blood flow is achieved by local overexpression of VEGF-A 165 in the uterine arteries of pregnant sheep, Gene Ther. 19 (2012) 925–935, https://doi.org/10.1038/gt.2011.158.
[65] X. Gou, Z. Wang, N. Li, F. Qiu, Z. Xu, D. Yan, S. Yang, X. Jia, Z. Konga, S. Wei, L. Lu, C. Lian, X. Wu, G. Wang, M. Li, Q. Teng, X. Jiang, J. Zhao, B. Yang, D. Liu, H. Wei, J. Li, Y. Yang, G. Yan, X. Zhao, M. Dong, W. Li, J. Deng, C. Leng, C. Wei, H. Wang, H. Mao, G. Zhang, Y. Li Ding, Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia, Genome Res. 24 (2014) 1308–1315, https://doi.org/10.1101/gr.171876.113.
[66] Z. Edea, H. Dadi, S.W. Kim, J.H. Park, G.H. Shin, T. Dessie, K.S. Kim, Linkage disequilibrium and genomic scan to detect selective loci in cattle populations adapted to different ecological conditions in Ethiopia, J. Anim. Breed. Genet. 131 (2014) 358–366, https://doi.org/10.1111/jbg.12083.
[67] Z. Edea, H. Dadi, T. Dessie, K.S. Kim, Genomic signatures of high-altitude adaptation in Ethiopian sheep populations, Genes Genom. 41 (2019) 973–981, https://doi.org/10.1007/s13258-019-00820-y.
[68] P. Wiener, C. Robert, A. Ahbara, M. Salavati, A. Abebe, A. Kebede, D. Wragg, J. Friedrich, D. Vasoya, D.A. Hume, A. Djikeng, M. Watson, J.G.D. Prendergast, O. Hanotte, J.M. Mwacharo, E.L. Clark, Whole-Genome Sequence Data Suggest Environmental Adaptation of Ethiopian Sheep Populations, GBE 13 (2021), evab014, https://doi.org/10.1093/gbe/evab014.
[69] S.H. Jung, M.L. Brownlow, M. Pellegrini, R. Jankord, Divergence in Morris water maze-based cognitive performance under chronic stress is associated with the hippocampal whole transcriptomic modification in mice, Front. Mol. Neurosci. 10 (2017) 275, https://doi.org/10.3389/fnmol.2017.00275.
[70] E.E. Terenina, S. Cavigelli, P. Mormede, W. Zhao, C. Parks, L. Lu, B.C. Jones, M. K. Mulligan, Genetic factors mediate the impact of chronic stress and subsequent response to novel acute stress, Front. Neurosci. 13 (2019) 438, https://doi.org/10.3389/fnins.2019.00438.
[71] M. Gautier, L. Flori, A. Riebler, F. Jaffrézic, D. Laloé, I. Gut, K. Moazami-Goudarzi, J.L. Foulley, A whole genome Bayesian scan for adaptive genetic divergence in west African cattle, BMC Genomics 10 (2009) 1–18, https://doi.org/10.1186/1471-2164-10-550.
[72] J.J. de Simoni Gouveia, S.R. Paiva, C.M. McManus, A.R. Caetano, J.W. Kijas, O. Facó, H.C. Azevedo, A.M. de Araujo, C.J.H. de Souza, M.E.B. Yamagishi, P.L. S. Carneiro, R.N. Braga Lôbo, S.M.P. de Oliveira, M.V.G.B. da Silva, Genome-wide search for signatures of selection in three major Brazilian locally adapted sheep breeds, Livest. Sci. 197 (2017) 36–45, https://doi.org/10.1016/j.livsci.2017.01.006.
[73] R.C. Gomes, R.D. Sainz, P.R. Leme, Protein metabolism, feed energy partitioning, behavior patterns and plasma cortisol in Nellore steers with high and low residual feed intake, RBZ. 42 (2013) 44–50, https://doi.org/10.1590/S1516-35982013000100007.
[74] C. Mcmanus, H. Louvandini, T. do Prado Paim, R. Saraiva Martins, J.O. Jardim Barcellos, C. Cardoso, R. Fontes Guimaraes, O. Antunes Santana, The challenge of sheep farming in the tropics: aspects related to heat tolerance, RBZ. 40 (2011) 107–120.
[75] V. Sejian, M. Bagath, G. Krishnan, V.P. Rashamol, P. Pragna, C. Devaraj, R. Bhatta, Genes for resilience to heat stress in small ruminants: a review, Small Rumin. Res. 173 (2019) 42–53, https://doi.org/10.1016/j.smallrumres.2019.02.009.
[76] F. Liška, C. Gosele, E. Popova, B. Chylíková, D. Křenová, V. Křen, M. Bader, L. L. Tres, N. Hubner, A.L. Kierszenbaum, Overexpression of full-length Centrobin rescues limb malformation but not male fertility of the Hypodactylous (hd) rats, PLoS One 8 (2013), e60859, https://doi.org/10.1371/journal.pone.0060859.
[77] T. Fukuda, C. Hoog, The mouse cohesin-associated protein PDS5B is expressed in testicular cells and is associated with the meiotic chromosome axes, Genes (Basel) 1 (2010) 484–494, https://doi.org/10.3390/genes1030484.
[78] D.J. McFarlane, R.J. George, E.G. Barrett-Lennard, M. Gilfedder, Salinity in dryland agricultural systems: Challenges and opportunities, in: M. Farooq, K. Siddique (Eds.), Innovations in Dryland Agriculture, Springer, Cham, 2017, pp. 521–547, https://doi.org/10.1007/978-3-319-47928-6_19.
[79] J. Busby, K. White, T.G. Smith, Mapping Climate Change and Security in North Africa, The German Marshall Fund of the United States:Climate & Energy Policy Paper Series, 2010.
[80] S.P. Bindra, S. Abulifa, A. Hamid, H.S. al Reiani, K. Abdalla, Assessment of impacts on ground water resources in Libya and vulnerability to climate change, Scientific Bull. "Petru Maior" Univ. Tîrgu Mureş. 10 (2013) 63–69.
[81] P. Pulimeno, C. Bauer, J. Stutz, S. Citi, PLEKHA7 is an adherens junction protein with a tissue distribution and subcellular localization distinct from ZO-1 and E-cadherin, PLoS One 5 (2010), e12207, https://doi.org/10.1371/journal.pone.0012207.
[82] B.T. Endres, J.R.C. Priestley, O. Palygin, M.J. Flister, M.J. Hoffman, B. D. Weinberg, M. Grzybowski, J.H. Lombard, A. Staruschenko, C. Moreno, H. J. Jacob, A.M. Geurts, Mutation of Plekha7 attenuates salt-sensitive hypertension in the rat, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 12817–12822, https://doi.org/10.1073/pnas.1410745111.
[83] H.R. Taal, L.C.L. van den Hil, A. Hofman, A.J. van der Heijden, V.W.V. Jaddoe, Genetic variants associated with adult blood pressure and kidney function do not affect fetal kidney volume. The Generation R Study, Early Hum. Dev. 88 (2012) 711–716, https://doi.org/10.1016/j.earlhumdev.2012.02.014.
[84] Y. Lin, X. Lai, B. Chen, Y. Xu, B. Huang, Z. Chen, S. Zhu, J. Yao, Q. Jiang, H. Huang, J. Wen, G. Chen, Genetic variations in CYP17A1, CACNB2 and PLEKHA7 are associated with blood pressure and/or hypertension in she ethnic minority of China, Atherosclerosis. 219 (2011) 709–714, https://doi.org/10.1016/j.atherosclerosis.2011.09.006.
[85] S. Harvey, J.G. Phillips, A. Rees, T.R. Hall, Stress and adrenal function, J. Exp. Zool. 232 (1984) 633–645, https://doi.org/10.1002/jez.1402320332.
[86] R. Alikhani-Koupaei, F. Fouladkou, P. Fustier, B. Cenni, A.M. Sharma, H. Deter, B. M. Frey, F.J. Frey, Identification of polymorphisms in the human 11beta-hydroxysteroid dehydrogenase type 2 gene promoter: functional characterization and relevance for salt sensitivity, FASEB J. 21 (2007) 3618–3628, https://doi.org/10.1096/fj.07-8140com.
[87] J.M. Friedman, J. Arbiter, J.A. Epstein, D.H. Gutmann, S.J. Huot, A.E. Lin, B. McManus, B.R. Korf, Cardiovascular disease in neurofibromatosis 1: report of the NF1 cardiovascular task force, Genet. Med. 4 (2002) 105–111, https://doi.org/10.1097/00125817-200205000-00002.
[88] Y. Takeda, M. Demura, F. Wang, S. Karashima, T. Yoneda, M. Kometani, A. Hashimoto, D. Aono, S.I. Horike, M. Meguro-Horike, M. Yamagishi, Y. Takeda, Epigenetic regulation of aldosterone synthase gene by sodium and angiotensin II, J. Am. Heart Assoc. 7 (2018), e008281, https://doi.org/10.1161/JAHA.117.008281.
[89] T. Fujita, Aldosterone in salt-sensitive hypertension and metabolic syndrome, J. Mol. Med. 86 (2008) 729–734, https://doi.org/10.1007/s00109-008-0343-1.
[90] K. Curnow, M. Tusie-Luna, L.P.-M., U., The product of the CYP11B2 gene is required for aldosterone biosynthesis in the human adrenal cortex, Mol. Endocrinol. 5 (1991) (1991) 1513–1522. https://academic.oup.com/mend/article-abstract/5/10/1513/2714337 (accessed October 23, 2021).
[91] A.C. Burke, J.L. Nowicki, Hox genes and axial specification in vertebrates, Am. Zool. 41 (2001) 687–697, https://doi.org/10.1093/icb/41.3.687.

[92] M. Mallo, D.M. Wellik, J. Deschamps, Hox genes and regional patterning of the vertebrate body plan, Dev. Biol. 344 (2010) 7–15, https://doi.org/10.1016/j.ydbio.2010.04.024.
[93] Z. Yuan, E. Liu, Z. Liu, J.W. Kijas, C. Zhu, S. Hu, X. Ma, L. Zhang, L. Du, H. Wang, C. Wei, Selection signature analysis reveals genes associated with tail type in Chinese indigenous sheep, Anim. Genet. 48 (2017) 55–66, https://doi.org/10.1111/age.12477.
[94] C. Wei, H. Wang, G. Liu, M. Wu, J. Cao, Z. Liu, R. Liu, F. Zhao, L. Zhang, J. Lu, C. Liu, L. Du, Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds, BMC Genomics 16 (2015) 1–12, https://doi.org/10.1186/s12864-015-1384-9.
[95] A. Abied, A.M. Ahbara, H. Berihulay, L. Xu, R. Islam, F.M. El-Hag, M. Rekik, A. Haile, J.L. Han, Y. Ma, Q. Zhao, J.M. Mwacharo, Genome divergence and dynamics in the thin-tailed desert sheep from Sudan, Front. Genet. 12 (2021) 1296, https://doi.org/10.3389/FGENE.2021.659507/FULL.
[96] C.J. Rubin, H.J. Megens, A.M. Barrio, K. Maqbool, S. Sayyab, D. Schwochow, C. Wang, Ö. Carlborg, P. Jern, C.B. Jørgensen, A.L. Archibald, M. Fredholm, M.A. M. Groenen, L. Andersson, Strong signatures of selection in the domestic pig genome, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 19529–19536, https://doi.org/10.1073/pnas.1217149109.
[97] S. Mikawa, T. Morozumi, S.I. Shimanuki, T. Hayashi, H. Uenishi, M. Domukai, N. Okumura, T. Awata, Fine mapping of a swine quantitative trait locus for number of vertebrae and analysis of an orphan nuclear receptor, germ cell nuclear factor (NR6A1), Genome Res. 17 (2007) 586–593, https://doi.org/10.1101/gr.6085507.
[98] H. Shengwei, L. Cunyuan, L. Ming, L. Xiaoyue, N. Wei, X. Yueren, Y. Rui, W. Bin, Z. Mengdan, L. Huixiang, Z. Yue, L. Li, Y. Ullah, Y. Jiang, Whole-genome resequencing reveals loci associated with thoracic vertebrae number in sheep, Front. Genet. 10 (2019) 674, https://doi.org/10.3389/fgene.2019.00674.
[99] S. Mastrangelo, B. Moioli, A. Ahbara, S. Latairish, B. Portolano, F. Pilla, E. Ciani, Genome-wide scan of fat-tail sheep identifies signals of selection for fat deposition and adaptation, Anim. Prod. Sci. 59 (2019) 835–848, https://doi.org/10.1071/AN17753.
[100] A. Vaysse, A. Ratnakumar, T. Derrien, E. Axelsson, G.R. Pielberg, S. Sigurdsson, T. Fall, E.H. Seppälä, M.S.T. Hansen, C.T. Lawley, E.K. Karlsson, D. Bannasch, C. Vilà, H. Lohi, F. Galibert, M. Fredholm, J. Häggström, Å. Hedhammar, C. André, K. Lindblad-Toh, C. Hitte, M.T. Webster, Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping, PLoS Genet. 7 (2011), e1002316, https://doi.org/10.1371/journal.pgen.1002316.
[101] A.R. Boyko, P. Quignon, L. Li, J.J. Schoenebeck, J.D. Degenhardt, K. E. Lohmueller, K. Zhao, A. Brisbin, H.G. Parker, B.M. VonHoldt, M. Cargill, A. Auton, A. Reynolds, A.G. Elkahloun, M. Castelhano, D.S. Mosher, N.B. Sutter, G.S. Johnson, J. Novembre, M.J. Hubisz, A. Siepel, R.K. Wayne, C.D. Bustamante, E.A. Ostrander, A simple genetic architecture underlies morphological variation in dogs, PLoS Biol. 8 (2010), e1000451, https://doi.org/10.1371/journal.pbio.1000451.
[102] C. Chen, C. Liu, X. Xiong, S. Fang, H. Yang, Z. Zhang, J. Ren, Y. Guo, L. Huang, Copy number variation in the MSRB3 gene enlarges porcine ear size through a mechanism involving miR-584-5p, Genet. Sel. Evol. 50 (2018) 1–18, https://doi.org/10.1186/s12711-018-0442-6.
[103] J.A. Allen, The influence of physical conditions in the Genesis of species, Sci. Am. 63 (1907), https://doi.org/10.1038/scientificamerican05251907-26247supp.
[104] M. Kardos, G. Luikart, R. Bunch, S. Dewey, W. Edwards, S. McWilliam, J. Stephenson, F.W. Allendorf, J.T. Hogg, J. Kijas, Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep, Mol. Ecol. 24 (2015) 5616–5632, https://doi.org/10.1111/mec.13415.
[105] S.E. Johnston, J.C. McEwan, N.K. Pickering, J.W. Kijas, D. Beraldi, J. G. Pilkington, J.M. Pemberton, J. Slate, Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population, Mol. Ecol. 20 (2011) 2555–2566, https://doi.org/10.1111/j.1365-294X.2011.05076.x.
[106] M. Hoefs, The thermoregulatory potential of Ovis horn cores, Can. J. Zool. 78 (2000) 1419–1426, https://doi.org/10.1139/cjz-78-8-1419.
[107] Z. Pan, S. Li, Q. Liu, Z. Wang, Z. Zhou, R. Di, B. Miao, W. Hu, X. Wang, X. Hu, Z. Xu, D. Wei, X. He, L. Yuan, X. Guo, B. Liang, R. Wang, X. Li, X. Cao, X. Dong, Q. Xia, H. Shi, G. Hao, J. Yang, C. Luosang, Y. Zhao, M. Jin, Y. Zhang, S. Lv, F. Li, G. Ding, M. Chu, Y. Li, Whole-genome sequences of 89 Chinese sheep suggest role of RXFP2 in the development of unique horn phenotype as response to semi-feralization, Gigascience. 7 (2018) giy019, https://doi.org/10.1093/gigascience/giy019.
[108] A. Zsolnai, I. Anton, L. Fésüs, A. Estonba, M. Schwerin, J. Vanselow, Allele distributions of two novel SNPs within the sheep Cyp19 gene, J. Anim. Breed. Genet. 119 (2002) 402–405, https://doi.org/10.1046/j.1439-0388.2002.00364.x.
[109] E. Casas, S.D. Shackelford, J.W. Keele, R.T. Stone, S.M. Kappes, M. Koohmaraie, Quantitative trait loci affecting growth and carcass composition of cattle segregating alternate forms of myostatin, J. Anim. Sci. 78 (2000) 560–569, https://doi.org/10.2527/2000.783560x.
[110] M.C. McClure, N.S. Morsci, R.D. Schnabel, J.W. Kim, P. Yao, M.M. Rolf, S. D. McKay, S.J. Gregg, R.H. Chapple, S.L. Northcutt, J.F. Taylor, A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle, Anim. Genet. 41 (2010) 597–607, https://doi.org/10.1111/j.1365-2052.2010.02063.x.