

Scotland's Rural College

## **Anthropogenic events and responses to environmental stress are shaping the genomes of Ethiopian indigenous goats**

Belay, Shumuye; Belay, Gurja; Nigussie, Helen; Ahbara, Abulgasim M; Tijjani, Abdulfatai; Dessie, Tadelle; Tarekegn, Getinet M; Jian-Lin, Han; Mor, Siobhan; Woldekiros, Helina S; Dobney, Keith; Lebrasseur, Ophelie; Hanotte, Olivier; Mwacharo, Joram M

*Published in:*  
Scientific Reports

*DOI:*  
[10.1038/s41598-024-65303-x](https://doi.org/10.1038/s41598-024-65303-x)

First published: 28/06/2024

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication](#)

### *Citation for published version (APA):*

Belay, S., Belay, G., Nigussie, H., Ahbara, A. M., Tijjani, A., Dessie, T., Tarekegn, G. M., Jian-Lin, H., Mor, S., Woldekiros, H. S., Dobney, K., Lebrasseur, O., Hanotte, O., & Mwacharo, J. M. (2024). Anthropogenic events and responses to environmental stress are shaping the genomes of Ethiopian indigenous goats. *Scientific Reports*, 14(1), 14908. Advance online publication. <https://doi.org/10.1038/s41598-024-65303-x>

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**



If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 06. Jul. 2024



OPEN

# Anthropogenic events and responses to environmental stress are shaping the genomes of Ethiopian indigenous goats

Shumuye Belay<sup>1,2,6</sup>, Gurja Belay<sup>2</sup>, Helen Nigussie<sup>2</sup>, Abulgasim M. Ahbara<sup>3,4</sup>, Abdulfatai Tijjani<sup>5,6</sup>, Tadelle Dessie<sup>6</sup>, Getinet M. Tarekegn<sup>3,7</sup>, Han Jian-Lin<sup>8,9</sup>, Siobhan Mor<sup>6,10</sup>, Helina S. Woldekiros<sup>11</sup>, Keith Dobney<sup>12,13</sup>, Ophelie Lebrasseur<sup>14</sup>, Olivier Hanotte<sup>6,15</sup> & Joram M. Mwacharo<sup>3,16</sup>

Anthropological and biophysical processes have shaped livestock genomes over Millenia and can explain their current geographic distribution and genetic divergence. We analyzed 57 Ethiopian indigenous domestic goat genomes alongside 67 equivalents of east, west, and north-west African, European, South Asian, Middle East, and wild Bezoar goats. Cluster, ADMIXTURE (K = 4) and phylogenetic analysis revealed four genetic groups comprising African, European, South Asian, and wild Bezoar goats. The Middle Eastern goats had an admixed genome of these four genetic groups. At K = 5, the West African Dwarf and Moroccan goats were separated from East African goats demonstrating a likely historical legacy of goat arrival and dispersal into Africa via the coastal Mediterranean Sea and the Horn of Africa.  $F_{ST}$ , XP-EHH, and  $H_p$  analysis revealed signatures of selection in Ethiopian goats overlaying genes for thermo-sensitivity, oxidative stress response, high-altitude hypoxic adaptation, reproductive fitness, pathogen defence, immunity, pigmentation, DNA repair, modulation of renal function and integrated fluid and electrolyte homeostasis. Notable examples include *TRPV1* (a nociception gene); *PTPMT1* (a critical hypoxia survival gene); *RETREG* (a regulator of reticulophagy during starvation), and *WNK4* (a molecular switch for osmoregulation). These results suggest that human-mediated translocations and adaptation to contrasting environments are shaping indigenous African goat genomes.

The domestication of goats (*Capra hircus*) occurred in the Fertile Crescent ~ 11,000 years ago from a mosaic of wild Bezoar (*Capra aegagrus*)<sup>1–3</sup>, ushered in one of the key milestones in the shift from hunting and gathering to the beginnings of sedentary agriculture. Following their domestication, goats demonstrated remarkable versatility in integrating and adapting to novel environments during their human-mediated dispersal occurring throughout the subsequent millenia<sup>4</sup>. Over time, their economic significance increased, and in recent times, rising demand for animal products has driven genetic-merit-based selective breeding for elite dairy (e.g., Alpine), meat (e.g., Boer), and fibre (e.g., Angora) landraces, along with other breeds with unique characteristics (<https://www.fao.org/dad-is/browse-by-country-and-species/en/>). In contrast, natural selection has given rise to significant variation

<sup>1</sup>Tigray Agricultural Research Institute, Mekelle, Ethiopia. <sup>2</sup>Department of Microbial, Cellular and Molecular Biology, Addis Ababa University, Addis Ababa, Ethiopia. <sup>3</sup>Animal and Veterinary Sciences, Scotland's Rural College (SRUC), Roslin Institute Building, Midlothian, UK. <sup>4</sup>Department of Zoology, Misurata University, Misurata, Libya. <sup>5</sup>The Jackson Laboratory, Bar Harbor, ME 04609, USA. <sup>6</sup>International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia. <sup>7</sup>Institute of Biotechnology (IoB), Addis Ababa University, Addis Ababa, Ethiopia. <sup>8</sup>CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Beijing, China. <sup>9</sup>Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing, China. <sup>10</sup>Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, UK. <sup>11</sup>Department of Anthropology, Washington University in St. Louis, St. Louis, USA. <sup>12</sup>Department of Archaeology, Classics and Egyptology, University of Liverpool, Liverpool, UK. <sup>13</sup>University of Sydney, Sydney, Australia. <sup>14</sup>Palaeogenomics and Bioarchaeology Research Network, School of Archaeology, University of Oxford, Oxford, UK. <sup>15</sup>School of Life Sciences, University of Nottingham, Nottingham, UK. <sup>16</sup>Small Ruminant Genomics, International Centre for Agricultural Research in the Dry Areas (ICARDA), Addis Ababa, Ethiopia. ✉email: shumuyeb@yahoo.com; j.mwacharo@cgiar.org

among indigenous varieties that have remained relatively nondescript in genotype and phenotype. Though less productive than elite breeds under an intensive production system, indigenous breeds carry important alleles and/or allelic combinations that confer adaptive plasticity to local environments.

Africa has a complex topography that influences agro-eco-climatic patterns and adaptive variation in domestic and wild species. These agro-eco-climates are well represented in Ethiopia which is home to a broad suite of indigenous domestic goat landraces that are uniquely adapted to diverse environments<sup>5</sup>. Domestic goats have a long history in the continent and in particular the Horn of Africa. For instance, the earliest occurrence of goats in the Ethiopian highlands and Sudanese Nile dates to the late 6th to early 5th millennia BC as attested by zooarchaeological findings<sup>6,7</sup>. Furthermore, increasing evidence indicates that north-east Africa (including the Red Sea area) and the Horn of Africa were a key arena for early socio-cultural and commercial exchange between mobile pastoralist groups from beyond the region<sup>8</sup>. A study of indigenous goats from the Horn could thus shed light on the long term dynamics of livestock-human interrelationships.

The Genome diversity of local goats in northeast Africa and the Horn has not been extensively studied. Investigations done so far, using mitochondrial DNA<sup>9–11</sup>, microsatellites<sup>12,13</sup>, and SNP genotypes<sup>14,15</sup> indicate high genetic diversity with little or no phylogeographic structure. However, mtDNA assesses only maternal divergence, microsatellites suffer low genome coverage, and SNP microarrays are prone to ascertainment bias compared to whole-genome resequencing data<sup>16</sup>. The only whole-genome study<sup>17</sup> undertaken to date analysed selection signatures in two out of 12 Ethiopian indigenous goat populations, leaving a major portion of the genomic variation poorly explored or investigated and therefore limiting interpretations at the local and global level. We therefore generated and analysed 57 whole-genomes from 12 Ethiopian indigenous domestic goat populations. The data were combined with previously sequenced genomes of 67 goats from east, west, and north-west Africa, Europe, South Asia, and the Middle East (Iran) and wild Bezoar goats. This analysis allowed us to: (1) characterise the current genomic landscape of Ethiopian indigenous goats vis a vis the non-Ethiopian ones, (2) describe a refined picture of the genetic and demographic history of Ethiopian indigenous goats in the wider context of other African and non-African goat populations, and (3) detect genome-wide signatures of positive selection for local adaptation.

## Results

### Sequence mapping quality and variant discovery

Whole-genome sequencing of the 57 Ethiopian goats yielded 2.8 Gb of paired-end raw reads (420 Gb) with a length of 150 bp. On average, 233.96 (218.84–249.22) Mb of clean reads were obtained per population. More than 99.75% of the clean reads mapped against the ARS1 goat reference genome assembly (GenBank accession number GCA\_001704415.1) with a coverage of ~99.59%. These mapped reads generated an average sequencing depth of 9.71-fold (Supplementary Table S3). A total of 24,759,579 high-quality SNPs were identified resulting in a mean SNP density of  $6.18 \pm 4.92/\text{Kb}$  (Supplementary Tables S2 and S3). Validation of the SNPs on the *C. hircus* dbSNP reference panel showed that 29.73%, 0.81%, and 40.12% were novel, exonic, and intergenic, respectively (Supplementary Table S2, Supplementary Figs. S1, S2).

### Genome-wide genetic diversity and population structure

The mean values of the genetic diversity indices are shown in Table 1. Among Ethiopian goats, the mean and standard deviation of  $H_O$ ,  $H_E$ ,  $\pi$ , and  $D_{ST}$  were 0.347, 0.371, 0.0021, and 0.303, respectively. The  $H_O$ ,  $H_E$ , and  $D_{ST}$  were highest in Ethiopian goats. The  $F_{HOM}$  and  $F_{RoH}$  averaged  $0.064 \pm 0.04$  and  $0.059 \pm 0.01$ , respectively in Ethiopian goats; the highest values were in Italian ( $F_{HOM} = 0.246 \pm 0.16$ ) and Pakistan ( $F_{HOM} = 0.243 \pm 0.11$ ) goats. In Ethiopian goats, the mean number and length of RoH were  $932.9 \pm 90.31$  and  $174.14 \pm 36.31$  Mb, respectively; the highest values were in Pakistan goats and averaged  $1663.4 \pm 811.17$  and  $410.27 \pm 233.37$  Mb, respectively.

Genetic structure and relationship were assessed at the first instance for Ethiopian indigenous goats only and then for the overall dataset that included all Ethiopian, non-Ethiopian and the wild Bezoar goats. PC1 and PC2 of the PCA explained 3.67% and 2.78% of the total genetic variation of Ethiopian goats (Fig. 1a) and each revealed two broad genetic clusters. The first cluster identified by PC1 includes Afar, Hararghe Highland, Short-eared Somali, Long-eared Somali, and Woyto-Guji. The second cluster of PC1 comprises Abergelle, Gonder, Agew, Ambo, Gumuz, Arsi-Bale and Keffa. Similarly, PC2 grouped together Afar, Hararghe highland, Short-eared Somali, Abergelle, Gonder, Agew, Ambo, Gumuz, and Arsi-Bale in its first cluster while its second cluster grouped together Long-eared Somali, Woyto-Guji, and Keffa. In combination, these two PC's reveal four clusters that likely reflect a fine-scale genetic divergence in Ethiopian indigenous goats. The first one is made up of Keffa; the second comprises Long-eared Somali and Woyto-Guji; the third comprises Afar, Hararghe Highland and Short-eared Somali, and the fourth includes Abergelle, Gonder, Agew, Ambo, Gumuz and Arsi-Bale.

To further examine the genetic structure of Ethiopian indigenous goats, we generated an NJ phylogeny using  $F_{ST}$  genetic distances (Fig. 1d). This revealed two broad genetic clusters that support the PCA results. We named these two genetic clusters A and B. The  $F_{ST}$  phylogeny also supports the fine-scale and deep genetic structure in Ethiopian goats revealed by PCA with Keffa being genetically distinct. ADMIXTURE analysis showed the lowest CV error at  $K = 1$  suggesting genetic homogeneity among Ethiopian goats. However, the genetic profile at  $K = 2$  (Fig. 1b) shows two broad genetic clusters, whose population composition corresponds to the ones identified by PC1 of the PCA and the  $F_{ST}$  phylogeny. For brevity, we also name these A and B. At the same  $K$  value (i.e.,  $K = 2$ ), ADMIXTURE analysis provides additional insights into the genetic structure of Ethiopian goats that are not apparent in the PCA and  $F_{ST}$  phylogeny; that the genomes of Arsi-Bale, Woyto-Guji and Hararghe Highland comprise different proportions of clusters A and B. Cluster A predominates in Arsi-Bale (69.0%) and, B in Woyto-Guji (71.47%) and Hararghe Highland (80.47%) (Supplementary Fig. S5). An analysis of the geographic

Population	N	$\Pi$	$H_O$	$H_E$	$D_{ST}$	RoH number	RoH Length (Mb)	$F_{HOM}$	$F_{ROH}$
		Mean	Mean	Mean	Mean	Mean $\pm$ SD	Mean $\pm$ SD	Mean $\pm$ SD	Mean $\pm$ SD
Abergelle	4	0.002	0.365	0.392	0.319	966 $\pm$ 72	184 $\pm$ 50.4	0.068 $\pm$ 0.03	0.063 $\pm$ 0.02
Afar	5	0.002	0.331	0.354	0.293	846 $\pm$ 142	162 $\pm$ 61.7	0.065 $\pm$ 0.04	0.056 $\pm$ 0.02
Agew	5	0.002	0.319	0.362	0.303	1004 $\pm$ 127	207 $\pm$ 66.8	0.117 $\pm$ 0.16	0.071 $\pm$ 0.02
Ambo	5	0.002	0.363	0.391	0.319	974 $\pm$ 123	200 $\pm$ 59.1	0.072 $\pm$ 0.03	0.068 $\pm$ 0.02
Arsi-Bale	5	0.002	0.339	0.359	0.290	958 $\pm$ 100	193 $\pm$ 41.6	0.056 $\pm$ 0.04	0.066 $\pm$ 0.01
Gonder	5	0.002	0.415	0.442	0.355	909 $\pm$ 25.8	155 $\pm$ 3.0	0.061 $\pm$ 0.02	0.053 $\pm$ 0.00
Gumuz	5	0.002	0.345	0.361	0.291	929 $\pm$ 54.8	171 $\pm$ 17.9	0.045 $\pm$ 0.02	0.059 $\pm$ 0.01
HHG	5	0.002	0.332	0.354	0.292	793 $\pm$ 32.4	131 $\pm$ 10.7	0.060 $\pm$ 0.03	0.045 $\pm$ 0.00
Keffa	5	0.002	0.344	0.363	0.293	1044 $\pm$ 65.4	199 $\pm$ 24.2	0.054 $\pm$ 0.03	0.068 $\pm$ 0.01
LESG	5	0.002	0.337	0.357	0.292	980 $\pm$ 114	172 $\pm$ 36.1	0.057 $\pm$ 0.03	0.059 $\pm$ 0.01
SESG	5	0.002	0.332	0.354	0.292	894 $\pm$ 90.4	168 $\pm$ 40.2	0.063 $\pm$ 0.04	0.058 $\pm$ 0.01
WGG	5	0.002	0.349	0.366	0.294	898 $\pm$ 137	147 $\pm$ 24.3	0.046 $\pm$ 0.03	0.050 $\pm$ 0.01
Mean	57	0.002	0.347	0.371	0.303	933 $\pm$ 90.3	174 $\pm$ 36.3	0.064 $\pm$ 0.04	0.059 $\pm$ 0.01
Boran	8	0.002	0.300	0.307	0.242	826 $\pm$ 125	130 $\pm$ 22.3	0.022 $\pm$ 0.03	0.044 $\pm$ 0.01
Moroccan	10	0.002	0.257	0.277	0.233	985 $\pm$ 474	205 $\pm$ 146	0.074 $\pm$ 0.08	0.070 $\pm$ 0.05
Mean	18	0.002	0.279	0.292	0.238	905 $\pm$ 299	167 $\pm$ 84.0	0.048 $\pm$ 0.06	0.057 $\pm$ 0.03
Bengal	7	0.002	0.292	0.320	0.261	729 $\pm$ 90.3	114 $\pm$ 17.1	0.088 $\pm$ 0.06	0.039 $\pm$ 0.01
Pakistan	10	0.002	0.226	0.299	0.259	1663 $\pm$ 811	410 $\pm$ 233	0.243 $\pm$ 0.11	0.140 $\pm$ 0.08
Iranian	10	0.002	0.233	0.260	0.228	641 $\pm$ 481	186 $\pm$ 203	0.102 $\pm$ 0.08	0.064 $\pm$ 0.07
Mean	27	0.002	0.250	0.293	0.249	1011 $\pm$ 461	237 $\pm$ 151	0.144 $\pm$ 0.08	0.081 $\pm$ 0.05
Italian	10	0.002	0.216	0.286	0.257	753 $\pm$ 339	144 $\pm$ 87.5	0.246 $\pm$ 0.16	0.05 $\pm$ 0.03
Bezoar	10	0.002	0.217	0.286	0.238	1608 $\pm$ 69.0	388 $\pm$ 133	0.242 $\pm$ 0.16	0.133 $\pm$ 0.04

**Table 1.** Indicators of genetic diversity and variation in the 25 Ethiopian and non-Ethiopian goat populations analysed in the study.  $\Pi$  = Nucleotide diversity,  $H_O$  = Observed heterozygosity,  $H_E$  = Expected heterozygosity,  $D_{ST}$  = Genetic distance,  $F_{HOM}$  and  $F_{ROH}$  = Genomic inbreeding coefficients, ROH length = length of Run of homozygosity (RoH), SD = Standard deviation, Mb = Mega base,

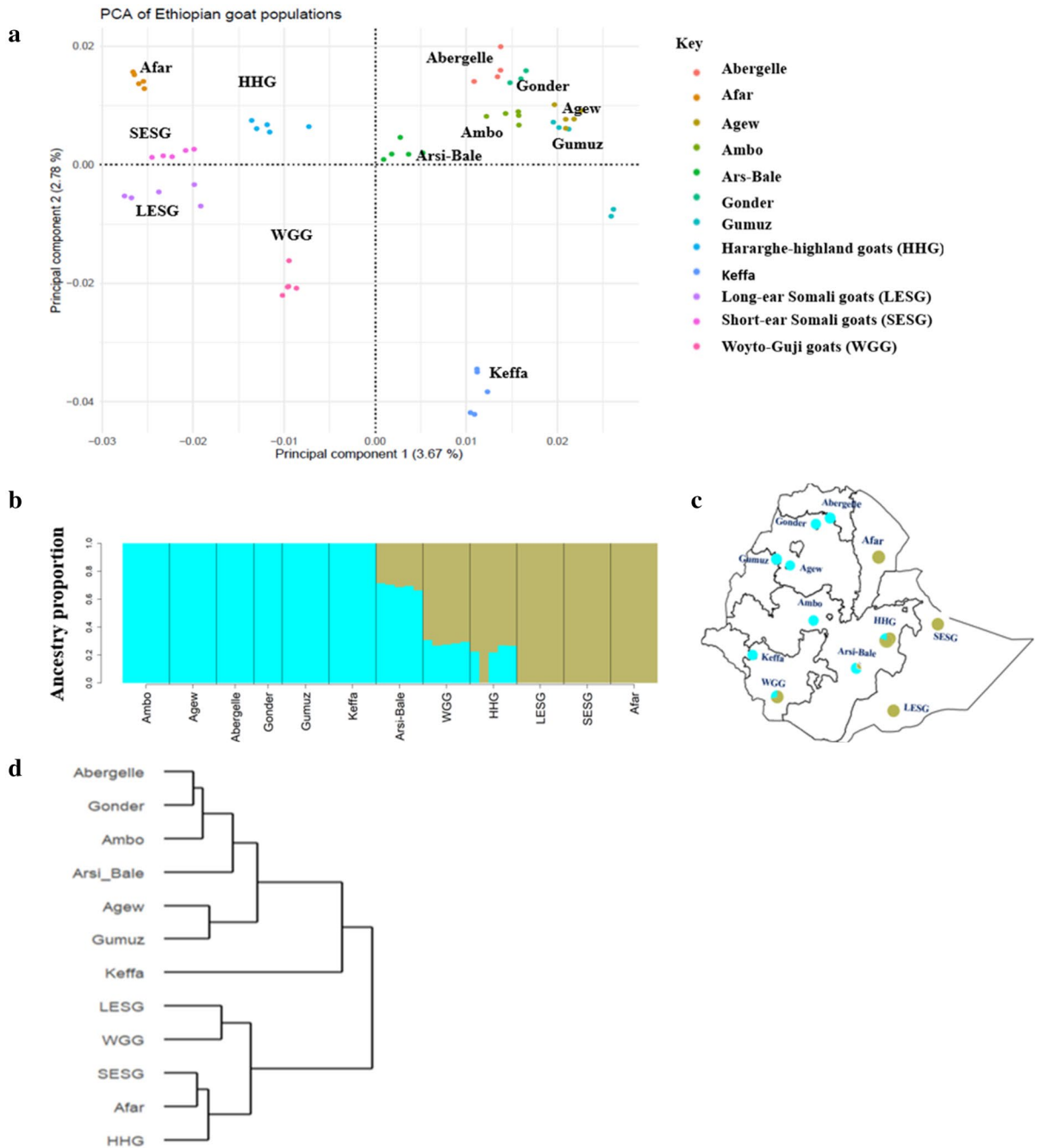
distribution of the two genetic clusters across Ethiopia shows that A predominates in the North and West, while B occurs at a higher frequency in the South and East of the country (Fig. 1c).

To explore genetic variation in Ethiopian goats in the context of their divergence from other African and non-African goat populations, we generated a PCA and ADMIXTURE profiles while including east African, west African, north-west African, European, South Asian, Middle East, and wild Bezoar goats in the analysis. PC1 and PC2 explain, respectively 11.76 and 5.31% of the total genetic variation in the dataset (Fig. 2a). It shows that Boran (Kenya) and Ethiopian goats are monophyletic. PC1 separates African and non-African goats while PC2 separates east African (Kenya, Ethiopia) goats from the west and north-west African (Dwarf and Moroccan) ones. This genetic clustering pattern is replicated by ADMIXTURE (Fig. 2b), which also reveals deeper insights into the genome architecture of the study populations. It reveals the lowest CV score is  $K=4$  (Fig. 2c), suggesting four genetic groups. East African goats share one genetic group that also occurs in the west (78.8%; Dwarf) and northwest (68.7%; Moroccan) African goats (Supplementary Fig. S6d). We refer to this as the “African genetic group”. The next dominant genetic group occurs in Pakistan and Bangladesh goats (referred to as the “South Asian genetic group”). The third predominates in Italian goats (“European genetic group”) whilst the fourth predominates in the wild Bezoar goat (“Wild genetic group”). We observed a certain level of admixture of the four genetic groups in Iranian goats. Admixture is also observed in west and north-west African, and European goats, and in a few individuals of the wild Bezoar goats (Fig. 2b; Supplementary Fig. S5d). Further scrutiny of ADMIXTURE at  $K=5$  shows that west and north-west African goats share a genome component that is different from that of their east African counterparts (Fig. 2b). This suggests further divergence of the African genetic group into two sub-groups, an east African, and a west/north-west African one. The latter is also found in Iranian and European goats and the wild Bezoar goats but at low frequencies.

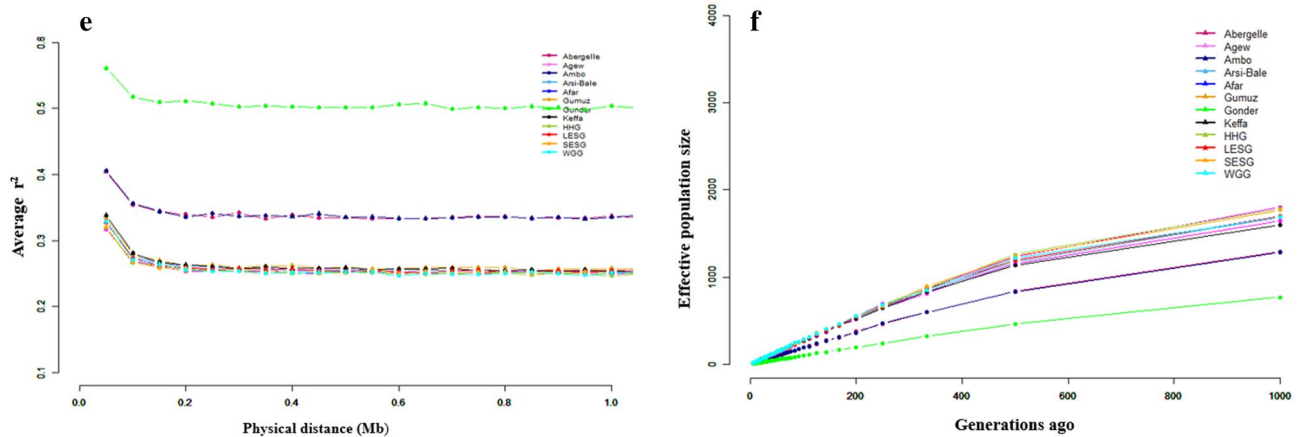
### Genome-wide dynamics

Genome-wide demographic dynamics were inferred by assessing LD patterns against genomic distances and changes in  $N_e$  over generation time for each population and genetic clusters that were revealed by PCA and ADMIXTURE. The pattern of LD decay was the same for all the population clusters (Figs. 1e, 2e). It reveals higher LD at shorter genomic distances which decays rapidly reaching a plateau around  $\sim 0.2$  Mb. Generally, Ethiopian goats show higher average LD ( $r^2 \geq 0.25$ ) (Fig. 1e) and within them, Gonder had the highest LD ( $r^2 \geq 0.5$ ), followed by Ambo and Abergelle ( $r^2 \geq 0.34$ ) and then the other populations ( $r^2 \geq 0.25$ ).

The pattern of  $N_e$  was similar across all the Ethiopian indigenous goat populations (Fig. 1f). There was a gradual decline in  $N_e$  between 1000 and 500 generations ago, following which the decline accelerated. Performing the analysis based on the genetic groups generated by cluster analysis reveals that both clusters A and B of



**Figure 1.** Population genetic structure and relationship of the Ethiopian goat populations based on (a) PCA and (b) ADMIXTURE analysis at  $K=2$ , (c) geographic distribution and genetic admixture proportion of the Ethiopian indigenous goat populations (d) phylogenetic tree constructed using  $F_{ST}$  values, (e) the pattern of linkage disequilibrium ( $r^2$ ) from 0 to 1 Mb and (f) the pattern of effective population size ( $N_e$ ) in the past 1000 generations.



**Figure 1.** (continued)

Ethiopian populations show an increase in  $N_e$  between 1000 and 400 generations ago following which there is a drastic decline (Fig. 2f).

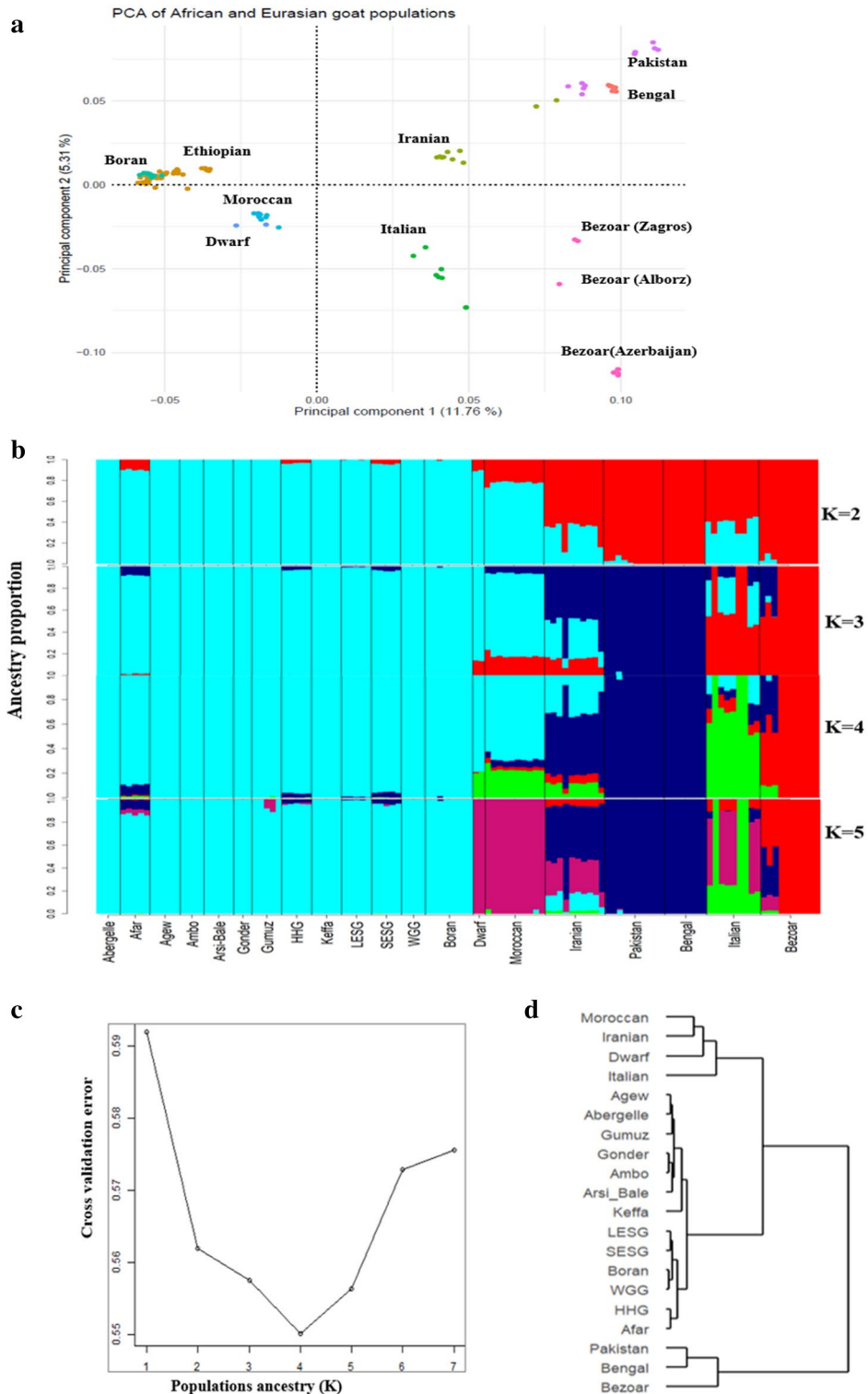
### Selection signature analysis

We investigated genome-wide selection signatures with  $H_p$ ,  $F_{ST}$ , and XP-EHH tests to explore whether the population clusters observed in Ethiopian goats are the result of adaptation to different environments. For this analysis, we selected Afar, Arsi-Bale, and Keffa goats as they occurred in different clusters on the PCA. The  $H_p$  test revealed a total of 196 candidate regions under selection, that overlapped 484 genes across Afar, Arsi-Bale, and Keffa goats (Fig. 3a; Supplementary Table S5). The  $F_{ST}$  (Fig. 3b; Supplementary Table S6) and XP-EHH (Fig. 3c; Supplementary Table S7) tests identified a total of 222 and 356 candidate regions, respectively that spanned 411 and 757 genes when contrasting Afar and Arsi-Bale, Afar and Keffa, and Arsi-Bale and Keffa. Based on the ARS1 RefSeq gene annotation, a total of 145 regions did not overlap with any gene(s) and/or spanned genes that are not yet annotated (Supplementary Tables S5–S7). Given the large number of candidate regions identified by the three tests, to retain high specificity we used the threshold score values of  $-7.0$  ( $ZH_p$ ),  $7.0$  ( $ZF_{ST}$ ), and  $6.0$  (XP-EHH) to define the top-most significant selection regions. We regarded these to be the primary selection signatures that are shaping the genomes of the study populations. Our results and discussions will be based on these top-most significant regions unless specified otherwise.

Based on the above cut-off threshold scores, eight candidate regions in Afar (overlapping 36 genes), seven in Arsi-Bale (seven genes), and eight in Keffa (55 genes) were identified by  $ZH_p$  as the top-most significant regions (Table 2). The  $ZF_{ST}$  identified nine regions in Afar vs Arsi-Bale (13 genes), nine in Afar vs Keffa (34 genes), and four in Arsi-Bale vs Keffa (4 genes) (Table 3) while those identified by XP-EHH were, 23 for Afar vs Arsi-Bale (80 genes), seven for Afar vs Keffa (12 genes), and nine for Arsi-Bale vs Keffa (12 genes) (Table 4). Of these 253 genes, eight (*SCNNIB*, *SMPD3*, *NUP43*, *ZNF609*, *COG7*, *ZFP90*, *PCMT1*, *PIFI*) overlapped between  $ZH_p$  and  $ZF_{ST}$ , one each (Fig. 4) overlapped between  $ZH_p$  and XP-EHH (*BPIFB4*), and  $ZF_{ST}$  and XP-EHH (*ALOX5AP*).

The 253 genes were compared with published literature to determine their functional significance. We categorised these genes into two groups (Tables 2, 3 and 4). Group 1 comprises genes previously reported in other livestock species: among these are five genes: *POU2F1* and *KITLG*, which relate to coat/fur/skin pigmentation, *RXFP2* which is a major gene for sheep horn status and secondary sexual characteristics, *CACNB2* which plays a role in reproduction process in dairy cows, *PPP1R14C* which is believed to contribute to trypanotolerance in Sheko cattle. Group 2 consisted of genes reported to have biological roles in other animal species. These included among others thermo-sensitivity (*TRPV1*, *PCMT1*, *IYD*, *PACRGL*), oxidative stress response and control of reactive oxygen species (*PCMT1*, *TIGAR*, *TRPM2*, *ALDH3B1*, *ALOX5AP*, *RAMP2*, *HSF4*), hypoxic survival and adaptation to high-altitude (*DDX28*, *RUNDC3B*, *PIK3CD*, *TIGAR*, *PTPMT1*, *STXBPA*), and reproduction function-related (*NUP43*, *EXOSC10*, *TARDBP*, *DPEP3*, *ESCO2*, *OAZ2*, *HSD17B1*, *PSMC3IP*, *EZH1*, *NECTIN3*, *KATNAL1*, *CACNB2*, *HASPIN*, *USP42*, *SUN5*, *KCTD19*, *ELF5*, *KITLG*, *TSNAXIP1*) genes. Generally, the majority of these genes in both classes encode proteins with multifaceted functions that range from adaptation, development, regulation, and maintenance of tissue and cellular functions.

Functional annotation and gene ontology analysis were performed with DAVID at two levels (1) for all the genes identified by  $ZH_p$  in each population and (2) for all the genes identified by the three comparative analyses (Afar vs Arsi-Bale, Afar vs Keffa, and Arsi-Bale vs Keffa) involving  $ZF_{ST}$  and XP-EHH combined as one gene list. We found several functional gene clusters that were significantly enriched (Supplementary Tables S8 and S9). In Afar, the enriched clusters were RNA degradation, endocytosis, autophagy, Rap1 signaling pathway, PI3K-Akt signaling pathway, and MAPK signaling pathways. In Arsi-Bale, enrichments were Ras signaling pathway, Rap1 signaling pathway, PI3K-Akt signaling pathway, MAPK signaling pathway, bacterial invasion of epithelial cells and melanogenesis. In Keffa, MAPK signaling pathway, glycerophospholipid metabolism, TGF-beta signaling pathway, Yersinia infection and IL-17 signaling pathway were overrepresented. In the Afar vs Arsi-Bale comparative analysis, significantly enriched clusters were for neurological system processes, apoptotic processes involved in morphogenesis, and sensory organ development. In Afar vs Keffa, the significantly enriched



**Figure 2.** Population genetic structure and relationship of the African, South Asian, Middle Eastern, European, and wild Bezoar goat populations. **(a)** PCA, **(b)** ADMIXTURE analysis at K=2, 3, 4, and 5, **(c)** Cross-validation error (CV) value at K=4, **(d)** phylogenetic tree constructed using  $F_{ST}$  values, ( $e_a$  and  $e_b$ ) the pattern of linkage disequilibrium ( $r^2$ ) from 0 to 1 Mb and ( $f_a$  and  $f_b$ ) the pattern of effective population size ( $N_e$ ) in the past 1000 generations. Note: The LD and  $N_e$  are plotted based on the admixture, PCA and phylogenetic tree results. Ethiopian-A includes Abergelle, Gonder, Agew, Ambo, Gumuz, Arsi-Bale and Keffa whereas Ethiopian-B consists of Afar, Short ear Somali, Long ear Somali, Haraghe highland, Woyto-Guji and Kenyan-Boarn goats.

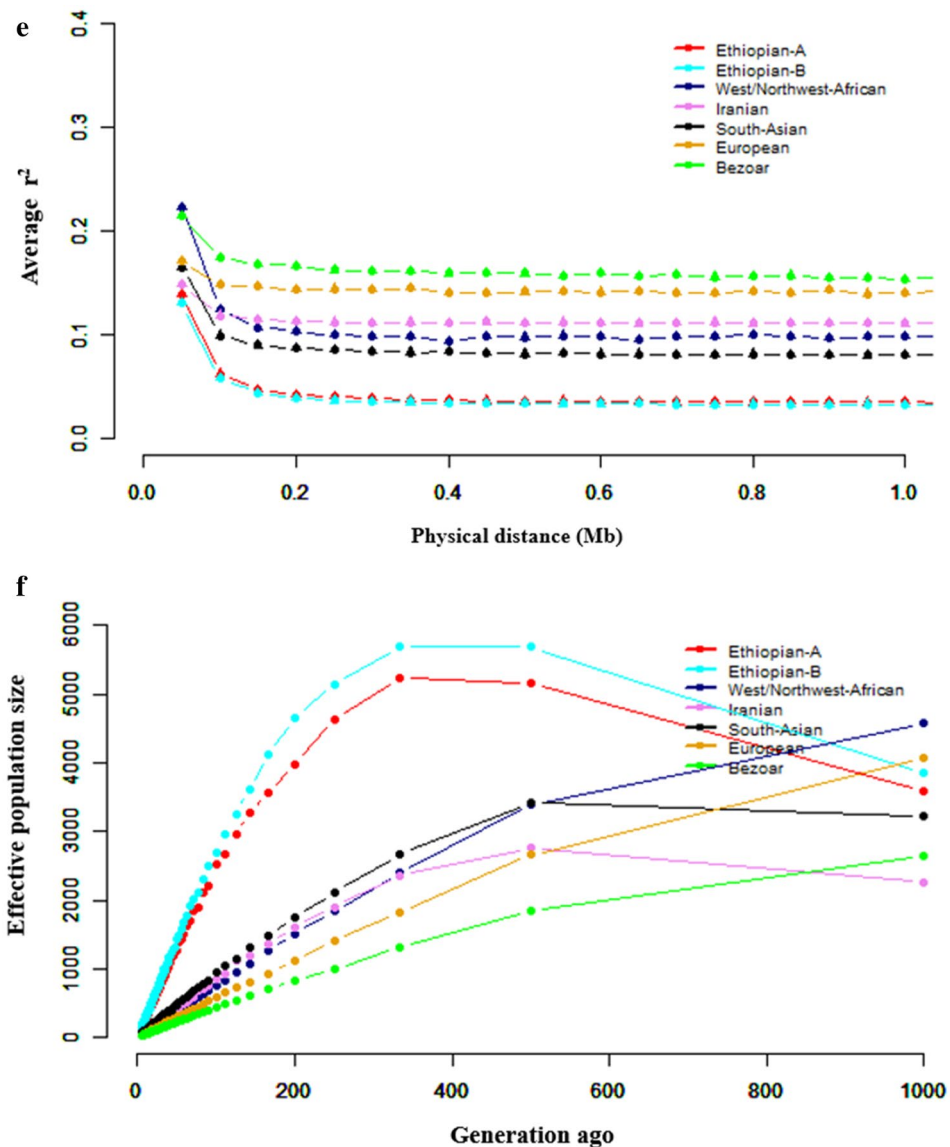


Figure 2. (continued)

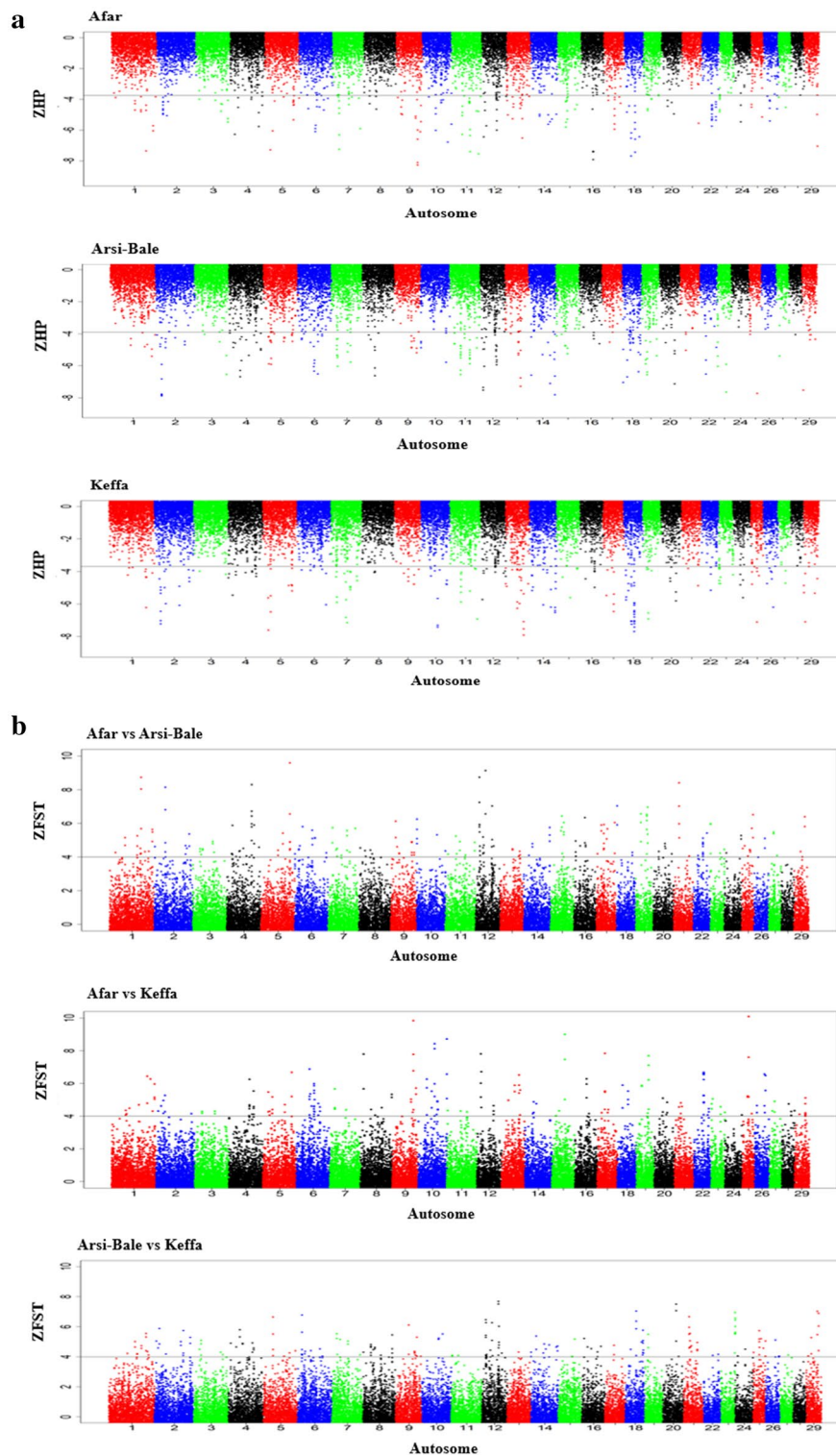
clusters were for sensory organ development, system process, and cell differentiation. Finally, Arsi-Bale vs Keffa showed distinct and extensive enrichments for 104 clusters, including growth, hematopoietic or lymphoid organ development, tissue development, and multicellular organismal homeostasis.

## Discussion

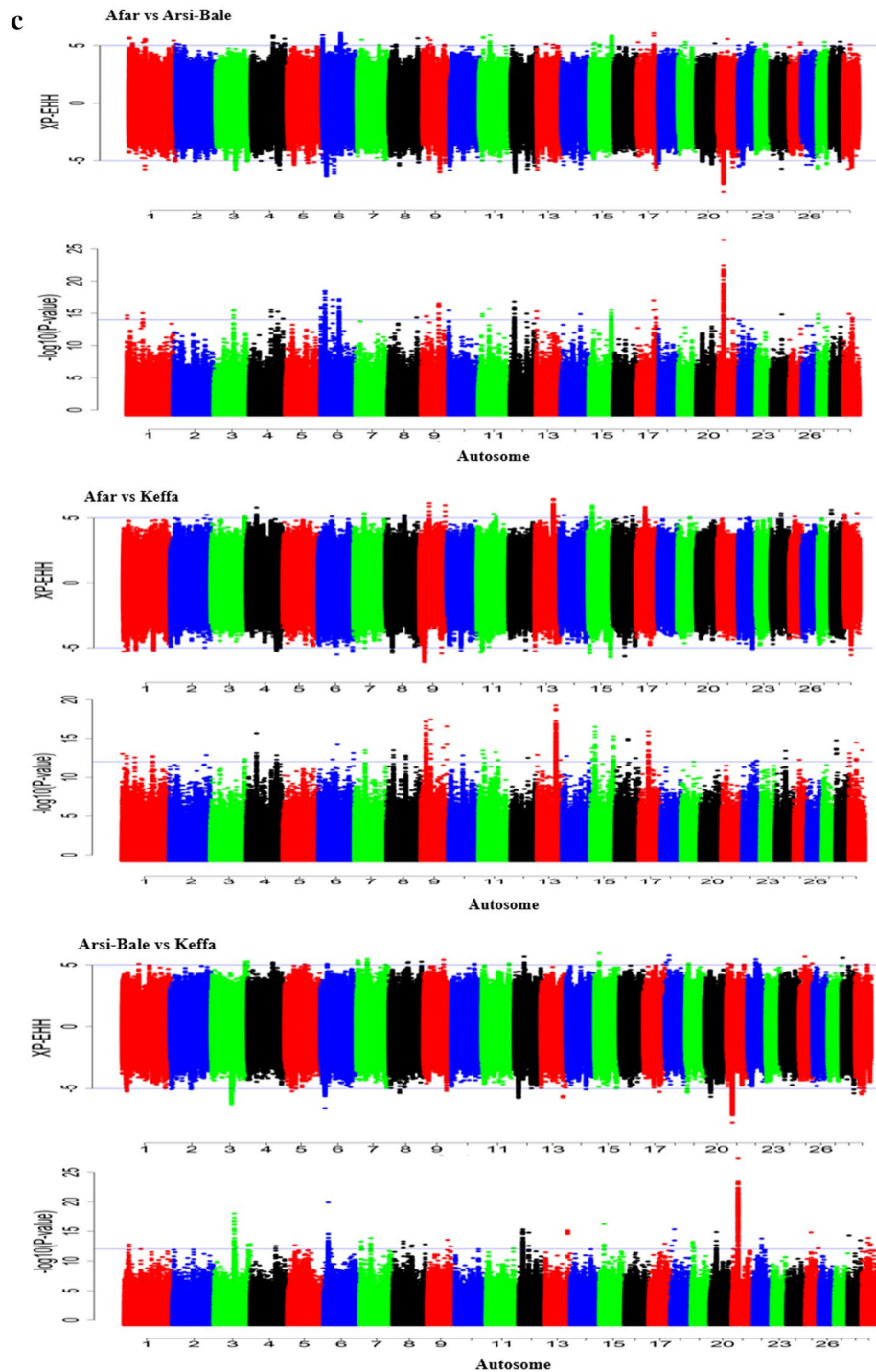
Following goat domestication from the wild Bezoar ~11,000 years ago<sup>1,2</sup>, population bottlenecks, inbreeding, intermixing, and selection (natural and artificial) have been modifying the genomes of domestic goats. This is especially the case in Africa where their initial introduction as exotics from the centre of domestication into the continent exposed them to novel environments and physiological extremes and challenges. Here, by analysing genetic divergence by exploiting SNPs from 57 individual genomes from 12 Ethiopian indigenous domestic goat populations and placing them within a comparative dataset of published African and non-African domestic goat breeds and the wild Bezoar goat allowed us to contextualize regional, continental, and intercontinental diversity and divergence. Notwithstanding differences in sequencing depth and platforms, our sequence statistics (Supplementary Table S1) were consistent with previous observations in goats<sup>18–20</sup>, sheep<sup>21,22</sup>, and cattle<sup>23,24</sup>, indicating the high quality and reliability of our dataset.

Irrespective of the species and markers used, several studies have reported high genetic diversity estimated from whole-genome sequences in indigenous livestock compared to exotic/commercial breeds<sup>22,25,26</sup>. The reduced genetic diversity in the latter is the outcome of long-term artificial selection and/or genetic drift due to a demographic history of low effective population sizes. The genome-wide autosomal SNPs showed high values in all the estimated parameters of genetic diversity in Ethiopian goats (Table 3). These values are comparable with those reported for indigenous goats in Uganda<sup>27</sup>, South Africa<sup>28</sup>, Egypt<sup>14</sup>, Spain<sup>29</sup> and Italy<sup>30</sup> estimated





**Figure 3.** Manhattan plots showing genome-wide selection signals as revealed by: (a) ZHP, (b)  $ZF_{ST}$  and (c) XP-EHH amongst Ethiopian indigenous goat populations. (a) ZHP Analysis for individual Ethiopian goat populations (Afar, Arsi-Bale and Keffa). (b) Manhattan plots for pairwise  $ZF_{ST}$  analysis results among the three Ethiopian indigenous goat populations (Afar, Arsi-Bale, Keffa) used in this analysis. (c) Manhattan plots for pairwise XP-EHH analysis results among the three Ethiopian indigenous goat populations (Afar, Arsi-Bale, Keffa) used in this analysis.



**Figure 3.** (continued)

using SNP microarray genotypes and from whole-genome sequence analysis in indigenous African goats<sup>25</sup>, cattle<sup>24</sup> and sheep<sup>22</sup>. In diploid genomes, RoH represents continuous homozygous segments of DNA sequences and can provide insights into how population history, structure, and demography have evolved over time<sup>31</sup>. In this study, estimates of  $F_{\text{RoH}}$  and  $F_{\text{HOM}}$  reveal low levels of inbreeding in all the populations analysed which is consistent with findings in Ugandan<sup>32</sup> and Egyptian Barki goats<sup>14</sup>. Our analysis also revealed a similar pattern of high frequency (> 50%) of the shortest average RoH length segments (0.1–0.25 Mb length category) in the populations analysed (Supplementary Table S4; Supplementary Figs. S3, S4). This skewed distribution agrees with other findings on goats<sup>32,33</sup>, sheep<sup>34</sup> and cattle<sup>34–36</sup> where long RoH segments were more infrequent than shorter ones. Short ROHs suggest inbreeding is not recent. Taken together, these results appear to suggest that the high heterogeneity in the study populations could be the result of a combination of factors. These include low inbreeding due to random mating and historic admixture arising from the communal use of resources and/or the

Chromosome	Region (Mb)	Highest ZHp score	No. of candidate regions	Genes
Afar				
1	119.60–119.80	–7.3551	3	
7	21.60–21.90	–7.2470	5	
9	73.85–74.10	–8.2640	4	<i>NUP43, PCMT1</i>
11	64.90–65.10	–7.3977	3	<i>TRNAC-ACA</i>
16	40.90–41.15	–7.9183	4	<i>MTOR, ANGPTL7, EXOSC10, SRM, MASP2, TARDBP</i>
18	22.95–23.15	–7.6813	3	<i>CHD9</i>
18	36.45–36.80	–7.4372	6	<i>CTCF, CARMIL2, ACD, PARD6A, ENKD1, GFOD2, RANBP10, TSNAXIP1, CENPT, THAP11, NUTF2, EDC4, NRN1L, PSKH1, PSMB10, LCAT, SLC12A4, DPEP3, DPEP2, DDX28, DUS2, NFATC3</i>
29	46.60–46.70	–7.0457	1	<i>ALDH3B1, NDUFS8, TCIRG1, CHKA</i>
Arsi-Bale				
2	21.15–21.50	–7.8828	5	
12	10.05–10.25	–7.5186	3	<i>GPR183, GPR18, UBAC2</i>
14	73.60–73.75	–7.0998	2	<i>SLA, TG</i>
	90.20–90.40	–7.8080	3	<i>LMBRD1</i>
18	2.00–2.15	–7.0364	2	
20	51.70–51.95	–7.1316	4	
23	30.45–30.60	–7.6424	2	<i>SUPT3H</i>
Keffa				
2	20.95–21.30	–7.2204	6	
5	18.05–18.60	–7.6087	2	<i>KITLG</i>
10	57.55–57.75	–7.4240	3	<i>OAZ2, ZNF609</i>
13	61.70–61.95	–7.9138	4	<i>BPIFB4</i>
18	26.10–27.00	–7.2406	6	<i>ADGRG5, CCDC102A, SLC12A3, NUP93, ADGRG1, HERPUD1, ADGRG3</i>
	36.00–37.20	–7.6909	18	<i>CBFB, C16orf70, B3GNT9, TRADD, HSF4, NOL3, KIAA0895L, EXOC3L1, E2F4, ELMO3, MIR328, TMEM208, FHOD1, SLC9A5, PLEKHG4, KCTD19, LRRC36, TPPP3, ZDHHC1, RANBP10, TSNAXIP1, CENPT, THAP11, NUTF2, EDC4, NRN1L, PSKH1, PSMB10, LCAT, SLC12A4, DPEP3, DPEP2, DDX28, DUS2, NFATC3, ESRP2, PLA2G15, SLC7A6, SLC7A6OS, PRMT7, SMPD3, ZFP90</i>
25	20.90–21.10	–7.1041	3	<i>SCNN1B, COG7</i>
29	5.00–5.15	–7.1025	2	

**Table 2.** Candidate regions revealed to be under selection and associated genes as identified by the ZHp approach in Afar, Arsi-Bale and Keffa Ethiopian indigenous goat populations.

common practice of sharing and/or gifting stock to cement social bonds and relationships, and the predominance of natural selection which favours standing genetic variation. This is supported by the demographic dynamics, which show all populations have low LD ranges (> 200 kb) and historic high  $N_e$  (Fig. 2). The highest average LD was recorded in Ethiopian (0.25–0.55) goat populations. These LD values are within the range reported for a large number of goat breeds<sup>37</sup> and East African shorthorn Zebu cattle<sup>38</sup>. However, when the data is combined the lowest and highest LD values are recorded in Ethiopian and wild Bezoar goats, respectively. The low levels of long-range LD show the lack of intensive selection or the populations have had large effective ancestral population sizes<sup>39</sup>. The decline in  $N_e$  from around 1250 years ago (500 generations) is difficult to explain as no significant events that could have affected goat populations have been reported in the Horn of Africa region. We however speculate that the start of this decline could have been driven by the Lapanarat–Mahlatale drought that occurred in the region and was characterized by a sequel of severe droughts and political upheavals<sup>40</sup>.

Though not native to Africa, goats are ubiquitous in the continent and are closely associated with the subsistence practices, socio-cultural life, and economy of many African societies. Archaeological evidence suggests that goats first arrived in Africa from Southwest Asia<sup>41</sup>. Our phylogenetic and population structure analyses at the whole genome level revealed at least two ancestral genetic groups in African goats; one was observed in west and north-west African goats and the other in east African goats. We hypothesise that these two genetic ancestries could be contiguous with the trajectories of the initial movement of the first goat-farming pastoralists into the continent from Southwest Asia. Radiocarbon dates of caprovid remains from the North African Mediterranean coastline are amongst the oldest in the continent<sup>6,42</sup>. Thus, the west/north-west African ancestry, due to its location furthest from the postulated initial entry point of goats into Africa, was likely the first in the continent. To arrive at their current locations this ancestral group may have dispersed along two routes following its entry via Egypt. The ancestry that occurs in north-west African goats dispersed along Africa side of the Mediterranean Sea while the one found in west Africa dispersed overland across the present-day Sudano-Sahelian belt<sup>43</sup>.

The East African genetic ancestry could represent goats that spread from the Near East into the central Sahara, Sudan Nile, and the Ethiopian highlands between 6500 and 5000 BP<sup>6,7</sup>. This genome ancestry could have spread

Chromosome	Region (Mb)	Highest Z <sub>ST</sub> score	No. of candidate regions	Genes
Afar x Arsi-Bale				
1	111.20–111.45	8.7369	4	<i>GMPS, SLC33A1</i>
2	39.75–40.00	8.1483	4	<i>PLEKHM3, FZD5, CREB1</i>
4	87.45–87.90	8.3029	7	<i>ABCB4, RUNDC3B</i>
5	101.50–102.10	9.5972	6	<i>FGF23, TIGAR, CCND2</i>
12	13.65–13.95	8.7414	5	
	35.15–35.30	9.1363	2	
	57.25–58.35	7.0350	3	<i>ALOX5AP, N4BP2L2, PDS5B</i>
18	2.00–2.15	7.0405	2	
21	19.00–19.20	8.4123	3	
Afar x Keffa				
8	10.70–10.85	7.7960	2	<i>SCARA5, ESCO2</i>
9	73.85–74.35	10.7976	5	<i>LATS1, NUP43, PCMT1</i>
10	57.45–57.75	8.4319	5	<i>ZNF609, OAZ2, PIF1</i>
	100.80–101.00	8.7218	3	
12	13.70–13.95	7.8119	3	
15	45.95–46.15	8.9992	3	<i>SOX6</i>
17	26.00–26.65	7.8414	4	<i>ULK1, PUS1, EP400, SNORA49, FBRSL1, LRCOL1, P2RX2</i>
19	42.25–42.65	7.6954	4	<i>NAGLU, HSD17B1, COASY, MLX, PSMC3IP, RETREG3, TUBG1, CNTNAP1, EZH1, RAMP2, VPS25, WNK4, CNTD1, PSME3, AOC2, G6PC1</i>
25	20.85–21.10	10.4511	4	<i>SCNN1B, COG7</i>
Arsi-Bale x Keffa				
12	57.20–57.40	7.6788	3	
18	37.00–37.25	7.0389	4	<i>SMPD3, ZFP90, CDH3, SA</i>
20	45.55–45.75	7.4978	3	
29	38.20–38.35	7.0217	2	

**Table 3.** Candidate regions under selection and associated genes as identified by the Z<sub>ST</sub> approach in Afar x Arsi-Bale, Afar x Keffa and Arsi-Bale x Keffa Ethiopian indigenous goat populations.

overland via the Sinai Peninsula and Nile Delta regions and/or via the Red Sea Hills region of the Egyptian Red Sea coast. The split of Ethiopian goats into two clusters, named here A and B (Fig. 1a,b,d), and their respective geographic distribution across the country could shed light on the dispersal of this genomic ancestry into East Africa. The two clusters mirror findings from the analysis of 50 K SNP genotype data of Ethiopian goats<sup>15</sup> and of mitochondrial DNA, which identified two haplogroups in the Ethiopian<sup>11</sup>, Kenyan<sup>10</sup>, Sudanese<sup>44</sup> and Egyptian<sup>45</sup> indigenous goats. Cluster A predominates in goats found in the north and west of Ethiopia, while B occurs at a higher frequency in goats found in the south and east of the country (Fig. 1d). Their spread across the country was most likely facilitated by socio-cultural and commercial interactions as can be inferred from anthropologic, linguistic and human genetic studies<sup>46</sup>. This geographic spread led us to suggest that cluster A could have dispersed to Ethiopia from Egypt following the Nile River basin or across the Red Sea Hills, while cluster B most likely arrived in Ethiopia via the Horn of Africa through the Bab el-Mandeb strait.

Worth mentioning is that the inclusion of genomes from east, west, and north-west Africa, Europe, South Asia, the Middle East, and from the wild Bezoar goats provided an interesting insight. The ADMIXTURE analysis showed that all the four genomes it revealed are present in Iranian (Middle East) goats, which supports this as the cradle of present-day goat genome diversity.

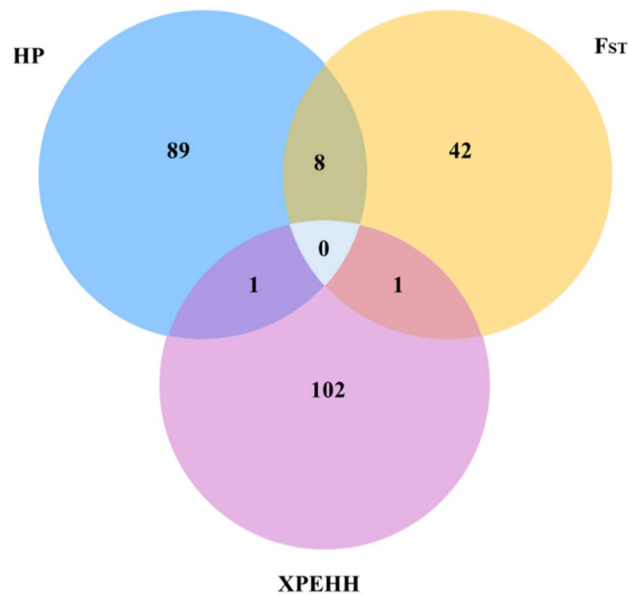
Ethiopia is characterized by a diverse combination of agro-eco-climates and ancient and modern human ethnic diversity that may have influenced the genome architecture of indigenous livestock. Our phylogenetic analysis revealed fine-scale genetic structuring in Ethiopian goats, which we hypothesize could be driven by environmental adaptation. Thus, signatures of selection were investigated within and between Afar, Arsi-Bale, and Keffa on the premise that their divergence is driven by adaptation to contrasting environments and therefore can serve as good proxies for investigating selection signatures resulting from genomic divergence. Afar goats inhabit a low altitude area (120–200 masl) with a hot semi-arid/arid agro-ecology (mean annual rainfall 150–300 mm). Arsi-Bale goats inhabit the Bale mountains (> 3000 masl) with a cool, cold sub-humid and alpine agro-ecology. These two populations clustered separately in the PCA suggesting genomic divergence. Keffa, which also occurs in a mid-altitude (≤ 1800 masl) environment, showed a clear genetic divergence from the other Ethiopian populations, suggesting it is genetically unique. This was observed previously by Tarekegn<sup>11</sup> from the analysis of 50 K SNP genotype data. The selection signature analysis revealed several candidate genomic regions under selection, some of which did not span any genes. This is not uncommon; it has been reported in cattle<sup>47</sup>, sheep<sup>14</sup>, and goats<sup>15</sup>. In line with our hypothesis, the top-most candidate regions under selection, spanned genes with roles in adaptation to different biophysical stressors rather than production. This suggests that natural rather than deliberate artificial selection is the principal driver of divergence in the studied populations and is mediated by

Chromosome	Region (Mb)	Highest XP-EHH Score	No. of candidate regions	Genes
Afar x Arsi-Bale				
1	55.00–55.15	6.3462	1	<i>NECTIN3</i>
	155.9–157.05	6.1671	2	<i>EFHB, RAB5A, PP2D1</i>
3	65.95–66.60	6.0621	2	
6	40.60–40.75	6.9709	1	<i>PACRGL</i>
9	58.50–58.70	6.5111	1	<i>EYA4</i>
11	12.85–13.00	6.5781	1	<i>DYSF</i>
	87.10–87.25	6.0839	1	<i>CYS1, KLF11</i>
12	13.20–14.10	6.1591	2	<i>DNAJC3, DZIP1, CLDN10</i>
	56.10–57.75	7.0985	3	<i>SNORA70, KATNAL1, USPL1, ALOX5AP, RXFP2</i>
13	32.25–32.40	6.5657	1	<i>CACNB2</i>
15	15.25–18.00	6.7408	4	<i>RAG2, RAG1, TRAF6, LDLRAD3, ELF5, CAT</i>
17	3.65–3.80	6.8765	1	<i>TTC28</i>
19	24.10–24.25	6.4008	1	<i>TRPV1, SHPK, CTNS, TAX1BP3, EMC6, P2RX5, ITGAE, HASPIN</i>
20	61.95–62.10	6.6412	1	<i>CTNND2</i>
21	11.60–12.45	6.2535	2	
	18.90–19.30	7.8276	1	<i>ISG20</i>
22	2.55–2.70	6.6094	1	<i>CMC1, AZI2</i>
23	32.85–33.35	6.1298	2	<i>MRPS10, GUCA1B, GUCA1A, GUCA1ANB, C6orf132, TAF8, FOXP4</i>
25	38.90–39.05	6.5969	1	<i>AIMP2, EIF2AK1, ANKRD61, USP42, CYTH3</i>
26	31.10–33.15	6.2570	2	<i>NKX23, GOT1, CNNM1, CRTAC1, RRP12, FRAT2, FRAT1, ARHGAP19</i>
29	17.35–19.40	6.1719	3	<i>ALG8, NDUFC2, THRSP, KCTD14, INTS4</i>
	21.25–24.05	6.1197	4	<i>PRMT3, HTATIP2</i>
	26.10–28.60	6.1250	3	<i>VWA5A, NRG, VSIG2, ESAM, MSANTD2, ROBO3, ROBO4, STT3A, CHEK1, ACRV1, SOLD1, PATE1, PATE2, PATE3, PUS3, HYL51</i>
Afar x Keffa				
1	1.70–2.15	6.9002	2	<i>HUNK</i>
2	26.85–27.20	6.0046	2	
	114.80–116.35	6.1629	2	
9	73.95–74.85	6.7211	2	<i>PPP1R14C, IYD</i>
11	12.85–13.00	6.5923	1	<i>DYSF</i>
13	61.60–62.00	6.5096	1	<i>SUN5, BPIFB2, BPIFB6, BPIFB3, BPIFB4</i>
15	5.60–5.75	6.3199	1	<i>NDUFS3, KBTBD4, PTPMT1</i>
Arsi-Bale x Keffa				
3	118.65–118.80	6.1374	1	<i>POU2F1</i>
8	5.30–5.65	6.1761	2	<i>GALNT7</i>
12	54.95–58.90	6.2176	5	<i>USPL1, ALOX5AP, RXFP2, FRY</i>
15	13.45–14.70	6.0397	2	
	70.85–72.6	6.2050	2	
18	3.20–3.35	6.4696	1	<i>GLG1, RFWD3</i>
19	4.95–5.10	6.4695	1	<i>STXBP4</i>
27	43.30–43.35	6.0564	1	<i>CLN8</i>
28	2.95–3.10	7.5618	1	<i>MAPK8, FRMPD2</i>

**Table 4.** Candidate regions under selection and associated genes as identified by the XP-EHH approach in Afar x Arsi-Bale, Afar x Keffa and Arsi-Bale x Keffa Ethiopian indigenous goat populations.

the concurrent action of a complex network of genes. The large number of candidate regions and genes detected is also not surprising. Similar results have been reported for livestock species from extreme environments<sup>21,48</sup>.

Our findings corroborate those of Tarekegn et al.<sup>15</sup> who reported the divergence of Keffa from other Ethiopian goats by analysing 50 K SNP genotype data suggesting it to be genetically distinct. The authors speculated that it was due to Keffa being trypanotolerant. Whereas, we found a large number of selection signals in Keffa, Tarekegn et al. found none. We attribute this difference to the higher resolution afforded by whole-genome sequences. The XP-EHH between Afar vs Keffa detected a strong selection region on CHI9 (73.95–74.85 Mb) spanning two genes, *PPP1R14C* and *IYD*. Interestingly, *PPP1R14C* was found in a selection sweep region in the Ethiopian Sheko cattle and it was postulated to be one of the candidate genes contributing to trypanotolerance in the breed<sup>49</sup>. Both Sheko cattle and Keffa goats occur in an area where trypanosomiasis is an economically important livestock disease. Whether this signature presents convergent evolution in the two species for trypanotolerance is difficult



**Figure 4.** Venn diagram showing the number of genome-wide selection signals in Ethiopian indigenous goats that are specific to, and shared between  $H_p$ ,  $F_{ST}$ , and XP-EHH approaches.

to say from the current data. Cattle and goats are both bovids with minor genomic differences<sup>50,51</sup> indicating minor genomic differences since their divergence from a common ancestor<sup>14</sup>. Therefore, it is not unusual to find genetic and biological similarities between the two species. While this signature could contribute to the uniqueness of Keffa goats, we cannot conclude it is the only factor.

Afar goats reside in a semi-arid and arid environment. It is characterized by complex interacting biophysical stressors including heat, physical exhaustion, direct solar radiation, and resource (feed and water) scarcity. It is thus unsurprising that the selection tests for Afar goats, revealed signatures that spanned genes for thermosensitivity e.g., *TRPV1*, *PCMT1*, *PACRGL*, and *IYD*. The activity of *PCMT1* was reported to peak under lethal temperatures, suggesting a role for the enzyme in short-term responses to heat extremes<sup>52</sup>. D'Alessandro et al.<sup>53</sup> demonstrated that *PCMT1* also plays an essential role that ensures normal RBC circulation during oxidative stress. *TRPV1*, a thermally activated ion channel, plays a major role in thermosensation, thermoregulation, and nociception<sup>54</sup>, and its activation provides a high-temperature noxious-heat-avoidance signal<sup>55,56</sup>. We also found strong selection signatures in regions overlaying *ANGPTL7* and *ENKD1*, which are associated with maintaining skin integrity. By regulating extracellular matrix formation coupled with its high expression in keratinocytes<sup>57</sup>, *ANGPTL7* plays a central role in skin homeostasis, repair, and regeneration<sup>58</sup>. *ENKD1* regulates spindle orientation in basal keratinocytes, which promotes epidermal stratification and regeneration<sup>59</sup>. The epidermis protects an organism from extremes of the external environment, reduces water and heat loss and pathogen entry<sup>60</sup>. Thus, in arid environments, skin integrity is important for protection against solar radiation, maintaining internal homeostasis and moisture loss.

In several of the top-most candidate regions in Afar, we also found five genes, *FGF23*, *TIGAR*, *RAMP2*, *SLC12A3*, and *WNK4*, which are critical in regulating mineral and nutritional metabolism and homeostasis, and water balance. *FGF23* regulates mineral homeostasis and plays an essential physiological role in phosphate and vitamin D metabolism<sup>61</sup>. Bensaad et al.<sup>62</sup> showed that *TIGAR* can lower intracellular reactive oxygen species in response to nutrient starvation or metabolic stress, and functions to inhibit autophagy. The adrenomedullin (AM) and its receptor-modulating protein, *RAMP2*, facilitate early adaptation to cardiovascular stress by maintaining and regulating cardiac mitochondria and cardiovascular homeostasis against cardiovascular stress<sup>63</sup>. This is critical in countering physical stress resulting from long-distance trekking in search of pasture and water which can put pressure on cardiomyocytes. The *AM-RAMP2* system also suppresses endoplasmic reticulum stress-induced tubule cell death, thereby exerting a protective effect on kidneys<sup>64</sup>. Kidneys are critical in water and electrolyte ( $Na^+$ ,  $Cl^-$ , and  $K^+$ ) homeostasis. *SLC12A3* encodes the thiazide-sensitive sodium chloride cotransporter, which is primarily expressed in the kidney, intestines, and bones. It plays a key role in sodium, potassium, and blood pressure regulation in response to various hormonal and non-hormonal stimuli<sup>65</sup>. *WNK4* functions as a molecular switch that varies the balance between  $NaCl$  reabsorption and  $K^+$  secretion to maintain integrated homeostasis<sup>66</sup>. It plays a key role in coordinating the activities of flux pathways, which are regulated by the renin–angiotensin–aldosterone system, to achieve integrated fluid and electrolyte homeostasis (osmoregulation) in the distal nephron and distal colon<sup>67,68</sup> following acute dehydration and rapid rehydration.

Arsi-Bale and Keffa goats are found in Bale and Keffa zones, respectively. Keffa goats live between 1000 and 1800 masl, while Arsi-Bale goats inhabit elevations between 3000 and 4377 masl. These zones are characterized by high levels of precipitation. High-altitude environments impose a selective constraint in the form of hypobaric hypoxia<sup>69</sup>, which results in insufficient oxygen supply in body tissues. This would affect normal physiological

functions and can result in organ failure and death<sup>70,71</sup>. Our selection signature analyses for Arsi-Bale and Keffa goats revealed several strong signatures that spanned genes such as *RUNDC3B*, *TIGAR*, *PTPMT1*, *STXBP4*, and *ALOX5AP* relating to hypoxia adaptation. It has been shown that *RUNDC3B* is amongst genes with variants associated with increased risk of high-altitude polycythemia (HAPC) in Tibetan dwellers<sup>72</sup>. HAPC is characterized by excessive proliferation of circulating erythrocytes due to high-altitude hypobaric hypoxia. The compensation for prolonged hypobaric hypoxia exposure is the main reason for the change in erythrocyte production and haemoglobin concentration that elevates oxygen retention, transportation, and exchange. Findings by Kimata et al.<sup>73</sup> showed that *TIGAR* is a significant mediator of cellular energy homeostasis (glycolysis) and cell death (apoptosis) under ischemic/hypoxic stress. Through a genome-wide CRISPR-Cas9 KO library screening, Bao et al.<sup>74</sup> identified *PTPMT1*, an important enzyme for cardiolipin synthesis, as the third most significant gene for hypoxic adaptation/survival, ranking right after *HIF-1 $\alpha$*  and *HIF-1 $\beta$* . In a genome-wide study of genetic adaptation to high altitude in feral Andean Horses, the highest region of allele frequency divergence spanned, amongst five other genes, *STXBP4* and *COX11*<sup>75</sup>. A highly significant association was also found between these two genes and are in strong LD in both humans and horses. *COX11* is up-regulated in chronic hypoxia suggesting a role in dealing with oxygen deficit, possibly by acting as a heme biosynthetic enzyme that transports copper to heme<sup>76,77</sup>. Its strong association and LD with *STXBP4* may suggest a similar function for this gene. *ALOX5AP* was identified in sheep as a potential candidate for climate-mediated adaptation<sup>78</sup>. In humans, a mutation in *ALOX5AP* was associated with lung function<sup>79</sup>. Given the high altitude and restricted oxygen concentration in the Ethiopian highlands, *ALOX5AP* may play a role in adaptation by modulating respiratory function.

Extreme environments (high-altitude, semi-arid, and arid) impose anatomical, physiological, and metabolic challenges with strong evolutionary pressure due to long-term exposure to acute and chronic stress and other factors dependent on the natural history of a population. These factors are an underlying constant in the three test populations (Arsi-Bale, Keffa, Afar) and what is likely driving their difference is their exposure to biotic and abiotic stress factors in contrasting environments. It is therefore insightful that some of the strongest selection signals spanned genes for oxidative stress mitigation (*TRPM2*, *DDX28*, *ALDH3B1*, *TIGAR*, *ALOX5AP*, *RAMP2*, *POU2F1*, *HSF4*), DNA damage repair and maintenance of genome stability (*EXOSC10*, *CTCF*, *FZD5*, *PIF1*, *FBRSL1*, *HSF4*), skin pigmentation and characteristics (*FBRSL1*, *EZH1*, *POU2F1*, *SOX5*, *KITLG*, *OAZ2*, *HSF4*) and reproduction function (*NUP43*, *EXOSC10*, *TARDBP*, *DPEP3*, *ESCO2*, *OAZ2*, *HSD17B1*, *PSMC3IP*, *EZH1*, *NECTIN3*, *KATNAL1*, *CACNB2*, *ELF5*, *HASPIN*, *USP42*, *SUN5*, *KITLG*, *KCTD19*, *TSNAXIP1*). Long-term exposure to acute and/or chronic stress results in an imbalance in the production and accumulation of free radicals (reactive oxygen species) which can induce oxidative stress. Oxidative stress causes base damage and DNA strand breaks resulting in apoptosis and necrosis. Therefore, oxidative stress response and DNA strand repair can preserve genome integrity and normal mechanisms of cellular signaling under stress.

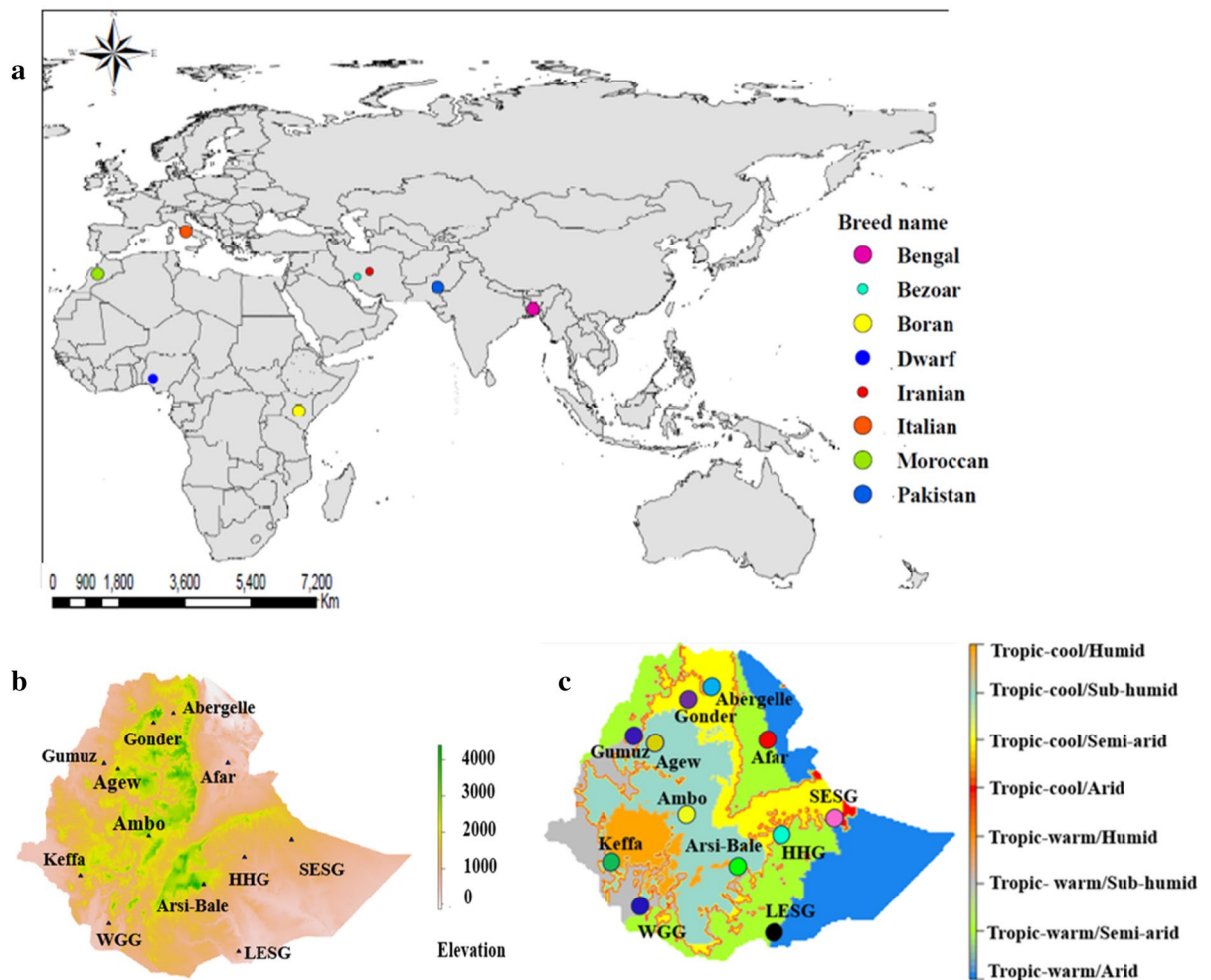
In conclusion, despite the complexity of the agro-ecological and climatic conditions of Africa, domestic goats occur across the continent. The results presented here show at least two genomic ancestries in African goats, and that genetic divergence and genomic plasticity is the driver of the successful integration of the species into African environments. These findings are significant in the context of improving livestock productivity in the continent in view of the projected consequences of climate change on biodiversity. A complete characterization of African indigenous goat's unique genomic variation and adaptation can inform the formulation and design of breeding programs that promote the long-term sustainable goat productivity.

## Materials and methods

### DNA samples and sequencing

Twelve indigenous Ethiopian goat populations (Supplementary Table 1; Fig. 5b,c) that were previously described in Tarekegn<sup>11,15</sup> were used in this study. No ethics permissions were required. DNA samples of five individuals per population were selected at random and whole genome sequenced in Novogene, China (<https://en.novogene.com/services/researchservices/genome-sequencing/whole-genome-sequencing-wgs/>). Whole-genome sequencing was performed on an Illumina NovaSeq 6000 Platform (Illumina, San Diego, CA, USA) and 150 bp of paired-end reads at a target coverage depth of 10 $\times$  were generated. The quality of the sequences was assessed with FASTQC v0.11.5<sup>80</sup> and three samples failed quality control (Supplementary Table S1). For comparative genome analysis and referencing, whole-genome sequences of 67 individuals of the east (Kenya, Boran goat), west (Nigeria, Dwarf goat) and north-west (Morocco) African (n = 20), South Asian (Pakistan and Bangladesh; n = 17), Middle East (Iran, n = 10) and European (Italy; n = 10) domestic goats, and 10 of wild Bezoar goats obtained from public databases (<https://www.goatgenome.org/vargogoats.html>) were included in the study (Supplementary Table S1; Fig. 5). The wild Bezoar goats, an extant species found in western Asia from Turkey to Pakistan is the presumed ancestor of modern-day domestic goats<sup>2</sup>. All clean reads were aligned to the *C. hircus* reference genome (ARS1; GenBank accession number GCA\_001704415.1) using the BWA tool v0.7.17<sup>81</sup>. The alignment files in SAM format were converted to BAM format with SAMtools<sup>82</sup>.

We applied the GATK v3.8 workflow<sup>83</sup> for variant calling and discovery. SAMTools was used to sort and index the alignment files and duplicate reads were purged using PICARD v2.18.2 (<https://broadinstitute.github.io/picard/>). Base Quality Score Recalibration was used to create recalibrated BAM files from the recalibrated table. Variant Quality Score Recalibration was performed using VariantRecalibrator in GATK with the "knownSites" set to the *C. hircus* dbSNP reference panel ([https://e99.ensembl.org/capra\\_hircus](https://e99.ensembl.org/capra_hircus)). Individual VCF files were called using the HaplotypeCaller in GATK followed by GenotypeGVCFs to generate consolidated GVCF files that contained the raw SNPs and Indels. Variant refinement was performed using variantRecalibrator in GATK. Variant call annotations such as Read Depth, Quality of Depth, Fisher Strand Test, Mapping Quality Score, Mapping Quality Rank Sum Test, Read Position Rank Sum Test Statistic, StrandOddsRatio Test, mode SNP and the VQSRTanchesSNP90 to 100 were used to produce recalibrated SNPs. Using ApplyRecalibration, a tranche



**Figure 5.** Map of the study areas (a) The geographic location of the African, South Asian, Middle Eastern, European, and Bezoar goat populations, (b) the geographic distribution of Ethiopian goat populations based on elevation, and (c) Agro-ecological zones (Sources: Own processed maps using global data set in ArcGIS environment version 10.8).

sensitivity threshold of 99% was used to generate filtered variants, and post-processing using SelectVariants was conducted to remove variants that failed the GATK filtering parameters. In this step, a total of 26,990,415 and 56,648,064 markers remained which, included autosomal and sex chromosomes for Ethiopian, and other (African, Eurasian, and wild Bezoar) goats, respectively. Further, quality control using the command line ‘SelectVariant’ and ‘restrictAlleles’ were used to remove sex chromosomes and multi-allelic SNPs, respectively. Finally, the subsequent analyses were performed using 24,759,579 and 42,728,409 autosomal biallelic markers found across 57 and 124 individuals from 12 and 25 populations, respectively (Table 1).

### Genome-wide genetic diversity and dynamics

Within-population variation was determined by estimating observed ( $H_O$ ) and expected ( $H_E$ ) heterozygosity, nucleotide diversity ( $\pi$ ), within-population genetic distance ( $D_{ST}$ ), and two inbreeding coefficients ( $F_{HOM}$  and  $F_{ROH}$ ). The  $H_O$ ,  $H_E$  and  $\pi$  were estimated with VCFtools v0.1.15<sup>84</sup>. To compute  $\pi$ , a 20 kb window size and a sliding step of 10 kb was used. For each population and group of populations, SNP density was also calculated with VCFtools v0.1.15 with the command line “-SNPdensity1000” and its mean and standard deviation was computed with R v4.1.0<sup>85</sup>. The  $D_{ST}$  was calculated using the command line “-genome” in PLINK v1.9<sup>86</sup>. The genetic distance between all individuals within a population was then calculated as  $D = 1 - D_{ST}$ .

Runs of homozygosity (RoH) were identified using PLINK v1.9 with the following command line parameters following Guo<sup>87</sup>, ‘-homozyg-window kb 5000 -chr-set 29 -homozyg-window-snp 50 -homozyg-window-het 1 -homozyg-snp 10 -homozyg-kb 100 -homozyg-density 10 -homozyg-gap 100’. The mean number and length of RoH were determined at four-genome length categories: 0.1–0.25, > 0.25–0.5, > 0.5–1, and > 1 Mb. The RoH-based inbreeding coefficient ( $F_{ROH}$ ) was computed as the average genome covered by RoH divided by the length of the ARS1 goat reference genome assembly.

Demographic dynamics were investigated by assessing pairwise LD between all pairs of autosomal biallelic variants (SNPs) over genomic distance through the correlation coefficient ( $r^2$ ) with PLINK v1.9. Historic



demographic dynamics were assessed by investigating the trends in effective population size ( $N_e$ ) over generation time (up to 1000 generations ago) as described in Ahbara et al.<sup>22</sup>.

### Genetic structure and relationships

Variation between populations was investigated at two levels: (1) for Ethiopian goats only and (2) for the combined dataset of Ethiopian and non-Ethiopian goat populations. The variation was explored and visualized with principal component analysis (PCA), ADMIXTURE tool, and Neighbour-Joining (NJ) phylogenetic tree. Only autosomal loci were used and those in LD were pruned with the `-indep-pairwise` function in PLINK1.9 using a window size of 50 kb, and step size of 10 kb, and  $r^2$  0.01. Out of the 24,759,579 autosomal SNPs in Ethiopian goats and the 42,728,409 autosomal SNPs in the overall dataset, 7,650,233 and 18,466,404 SNPs, respectively passed the LD filter and were used to assess population structure and relationships.

PCA was performed with PLINK v1.9 running the `-pca` command and the results were visualised by plotting the first two PCs using the tidyverse package of R v4.1.0<sup>85</sup>. The  $F_{ST}$ -distance matrices between populations were generated using VcfTools v0.1.15<sup>84</sup> and were then used to reconstruct an NJ phylogenetic tree with R software. The unsupervised block relaxation algorithm implemented in ADMIXTURE v1.3<sup>88</sup> was used to determine the proportion of shared genome ancestry between populations. A five-fold cross-validation procedure following Lawal et al.<sup>89</sup> was used to determine the optimal number of genome clusters/groups (K) and the proportion of shared genome ancestry.

### Genome annotations

To investigate whether population divergence revealed by cluster analysis, could be the outcome of adaptive radiation, we combined the results of PCA (Fig. 1a) with the topographic (Fig. 5b) and agro-eco-climatic distribution (Fig. 5c) of Ethiopian goats and selected Afar, Arsi-Bale and Keffa goats as the proxies to investigate selection signatures resulting from genomic divergence.

We used 19.28 million autosomal biallelic SNPs to run three approaches, pooled heterozygosity ( $H_p$ )<sup>90</sup>, fixation index ( $F_{ST}$ )<sup>91</sup>, and cross-population extended haplotype homozygosity (XP-EHH)<sup>92</sup>. A sliding window of 100 kb size with 50 kb sliding step was applied for  $H_p$  and  $F_{ST}$  tests. To avoid spurious signals, sliding windows with < 10 SNPs were discarded. The expected heterozygosity ( $H_p$ ) within each window was calculated using an in-house R script. For each SNP, the number of reads corresponding to the most ( $n_{MAJ}$ ) and least ( $n_{MIN}$ ) abundant alleles for each window in each population were used to calculate the  $H_p$  score as:  $H_p = 2\sum n_{MAJ}\sum n_{MIN}/(\sum n_{MAJ} + \sum n_{MIN})^2$ ; where,  $\sum n_{MAJ}$  and  $\sum n_{MIN}$  are the sums of  $n_{MAJ}$  and  $n_{MIN}$  for all the SNPs in the windows. Individual  $H_p$  values were then Z-transformed using the formulae  $ZH_p = (H_p - \mu H_p)/\sigma H_p$ . The  $F_{ST}$  value for each SNP between the three populations (Arsi-Bale, Afar, Keffa) was calculated using VCFtools (v0.1.15) to assess genetic differentiation. The  $F_{ST}$  value was then Z-transformed into  $ZF_{ST}$  with the formula:  $ZF_{ST} = (F_{ST} - \mu F_{ST})/\sigma F_{ST}$ . Putative selection targets were extracted from the extreme tail ends of the empirical distributions by applying a  $ZH_p$  score < -3.9 and the corresponding  $ZF_{ST}$  value > 4 as the cut-off thresholds. We compared the extended haplotype homozygosity (EHH) among the three populations (Arsi-Bale, Afar, Keffa) using the XP-EHH statistic estimated with the REHH package<sup>93</sup> in R. The unstandardized XP-EHH statistics were standardized using their means and variances. We estimated the p-values of the SNPs using the standard normal distribution following Sabeti et al.<sup>92</sup>. Regions falling within the top 0.001% of the empirical distribution or above XP-EHH score  $\geq 5$  were identified and taken to be the candidate selection sweep regions. All genes that either completely or partially overlapped with the candidate selection sweep regions were identified based on the ARS1 *C. hircus* reference genome gene annotations with the Ensembl BioMart (<http://www.biomart.org>) tool.

For functional classification, we retrieved genes within each candidate selective sweep region using Ensembl BioMart version 104. These gene lists were used for gene ontology (GO) (<http://geneontology.org/>) and KEGG (Kyoto encyclopedia of genes and genomes) (<http://www.genome.jp/kegg/pathway.html>) analyses implemented in DAVID version 6.8 (<http://david.ncicrf.gov/>)<sup>94</sup>. We used all RefSeq genes in the *C. hircus* genome as background. Overrepresented gene clusters were identified by Fisher's exact tests ( $p < 0.05$ ) and biological processes, cellular components and molecular functions were used as GO term categories with a significance level of  $p$ -value < 0.05.

### Data availability

The data generated herein have been deposited in NCBI under Sequence Read Archive (SRA) accession number SRP4642793<sup>95</sup>.

Received: 12 April 2023; Accepted: 19 June 2024

Published online: 28 June 2024

### References

- Zeder, M. A. The domestication of animals. *J. Anthropol. Res. Compet.* **68**, 161–190 (2012).
- Daly, K. G. et al. Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science* (80-). **361**, 85–88 (2018).
- Zheng, Z. et al. The origin of domestication genes in goats. *Sci. Adv.* **6**, 1–13 (2020).
- Pereira, F. & Amorim, A. Origin and spread of goat pastoralism. *Encycl. Life Sci. John Wiley Sons, Ltd Chichester* (2010). <https://doi.org/10.1002/9780470015902.a0022864>.
- FARM-Africa. *Goat Types of Ethiopia and Eritrea. Physical description and management systems. Published jointly by FARM-Africa, London, UK, and ILRI (International Livestock Research Institute), Nairobi, Kenya, PP 76.* (1996).
- Newman, J. L. *The Peopling of Africa: A Geographic Interpretation* (Yale University, 1995).
- Clutton-Brock, J. Cattle, sheep, and goats south of the Sahara: An archaeozoological perspective. In *The Origins and Development of African Livestock: Archaeology, Genetics, Linguistics and Ethnography* (eds Blench, R. M. & MacDonald, K. C.) 30–37 (UCL Press, 2000).

8. Boivin, N. Proto-globalisation and biotic exchange in the old world. *Human Dispersal Species Movement Prehistory Present*. <https://doi.org/10.1017/9781316686942.015> (2017).
9. Naderi, S. *et al.* Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity. *PLoS One*, **2**, (2007).
10. Kibegwa, F. M., Githui, K. E., Junga, J. O., Badamana, M. S. & Nyamu, M. N. Mitochondrial DNA variation of indigenous goats in Narok and Isiolo counties of Kenya. *J. Anim. Breed. Genet.* **133**, 238–247 (2015).
11. Tarekegn, G. M. *et al.* Mitochondrial DNA variation reveals maternal origins and demographic dynamics of Ethiopian indigenous goats. *Ecol. Evol.* **8**, 1543–1553 (2018).
12. Chenyambuga, S. W. *et al.* Genetic characterization of indigenous goats of sub-saharan Africa using microsatellite DNA markers. *Asian-Austr. J. Anim. Sci.* <https://doi.org/10.5713/ajas.2004.445> (2004).
13. Tesfaye, A. Genetic Characterization of Indigenous Goat Populations Of Ethiopia Using Microsatellite, PhD thesis submitted to the National Dairy Research Institute, Deemed University Karnal, India. (Deemed University, 2004).
14. Kim, E. *et al.* Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity (Edinb)*. **116**, 255–264 (2016).
15. Tarekegn, G. M. *et al.* Ethiopian indigenous goats offer insights into past and recent demographic dynamics and local adaptation in sub-Saharan African goats. *Evol. Appl.* <https://doi.org/10.1111/eva.13118> (2020).
16. Geibel, J. *et al.* How array design creates SNP ascertainment bias. *PLoS One* **16**, (2021).
17. Berihulay, H. *et al.* Whole genome resequencing reveals selection signatures Associated with important traits in Ethiopian indigenous goat populations. *Front. Genet.* **10**, 1–12 (2019).
18. Guo, J. *et al.* Whole-genome sequencing reveals selection signatures associated with important traits in six goat breeds. *Sci. Rep.* **8**, 1–11 (2018).
19. Wang, X. *et al.* Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. *Sci. Rep.* **6**, 1–10 (2016).
20. Guan, D. *et al.* Scanning of selection signature provides a glimpse into important economic traits in goats ( *Capra hircus* ). *Nat. Publ. Gr.* <https://doi.org/10.1038/srep36372> (2016).
21. Yang, J. *et al.* Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Mol. Biol. Evol.* **33**, 2576–2592 (2016).
22. Ahbara, A. M. *et al.* Natural adaptation and human selection of northeast African sheep genomes. *Genomics* **114**, 110448 (2022).
23. Espigolan, R. *et al.* Study of whole genome linkage disequilibrium in Nellore cattle. *BMC Genom.* **14**, 305 (2013).
24. Kim, J. *et al.* The genome landscape of indigenous African cattle. *Genome Biol.* **18**, 1–14 (2017).
25. Benjelloun, B. *et al.* Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats ( *Capra hircus* ) using WGS data. *Front. Genet.* **6**, 1–14 (2015).
26. Mwacharo, J. M. *et al.* Genomic footprints of dryland stress adaptation in Egyptian fat- tail sheep and their divergence from East African and western Asia cohorts. *Sci. Rep.* **7**, 1–10 (2017).
27. Onzima, R. B. *et al.* Genome-wide population structure and admixture analysis reveals weak differentiation among Ugandan goat breeds. 59–70 (2018). <https://doi.org/10.1111/age.12631>.
28. Mdladla, K., Dzomba, E. F., Huson, H. J. & Muchadeyi, F. C. Population genomic structure and linkage disequilibrium analysis of South African goat breeds using genome-wide SNP data. *Anim. Genet.* **47**, 471–482 (2016).
29. Manunza, A. *et al.* A genome-wide perspective about the diversity and demographic history of seven Spanish goat breeds. *Genet. Sel. Evol.* <https://doi.org/10.1186/s12711-016-0229-6> (2016).
30. Nicoloso, L. *et al.* Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip. *Genet. Sel. Evol.* <https://doi.org/10.1186/s12711-015-0140-6> (2015).
31. Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M. & Wilson, J. F. Runs of homozygosity: Windows into population history and trait architecture. *Nat. Rev. Genet.* **19**, 220–234 (2018).
32. Onzima, R. B. *et al.* Genome-wide characterization of selection signatures and runs of homozygosity in Ugandan goat breeds. *Front. Genet.* **9**, 1–13 (2018).
33. Brito, L. F. *et al.* Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *BMC Genom.* **18**, 1–20 (2017).
34. Purfield, D. C., Mcparland, S., Wall, E. & Berry, D. P. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PLoS One* **12**, 1–23 (2017).
35. Ferenc'akovic, M. *et al.* Estimates of autozygosity derived from runs of homozygosity: Empirical evidence from selected cattle populations. *Anim. Breeding Genet.* **130**, 286–293 (2013).
36. Mastrangelo, S. *et al.* Genomic inbreeding estimation in small populations: Evaluation of runs of homozygosity in three local dairy cattle breeds. *Animal* **10**, 746–754 (2016).
37. Brito, L. F. *et al.* Characterization of linkage disequilibrium, consistency of gametic phase and admixture in Australian and Canadian goats. *BMC Genet.* **16**, 1–15 (2015).
38. Mbole-Kariuki, M. N. *et al.* Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya. *Heredity (Edinb)*. **113**, 297–305 (2014).
39. Karimi, K., Koshkoiyeh, A. E. & Gondro, C. Comparison of linkage disequilibrium levels in Iranian indigenous cattle using whole genome SNPs data. *J. Anim. Sci. Technol.* **57**, 1–10 (2015).
40. Verschuren, D., Laird, K. & Cumming, B. F. Rainfall and drought in equatorial East Africa during the past 1, 100 years. *Nature* **403**, 410–414 (2000).
41. Mason, I. L. *Goat, Evolution of domesticated animals*, London; Longman group. pp 85–99 (1984).
42. Blench, R. M. & MacDonald, K. C. The origins and development of African livestock Archaeology, genetics, linguistics and ethnography. in 1–567 (London: UCL Press, 2005).
43. Pereira, F. *et al.* Tracing the history of goat pastoralism: New clues from mitochondrial and y chromosome DNA in North Africa. *Mol. Biol. Evol.* **26**, 2765–2773 (2009).
44. Sanhory, E., Giha, R. & Ibrahim, Z. H. Mitochondrial DNA diversity in three sudanese goat breeds. *Open Access Libr. J.* **1**, 1–10 (2014).
45. Naderia, S. *et al.* The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proc. Natl. Acad. Sci.* **105**, 17659–17664 (2008).
46. Pagani, L. *et al.* Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian Gene Pool. *Am. J. Hum. Genet.* **91**, 83–96 (2012).
47. Xu, L. *et al.* Genomic signatures reveal new evidences for selection of important traits in domestic cattle. *Mol. Biol. Evol.* **32**, 711–725 (2015).
48. Mwacharo, J. M. *et al.* Genomic footprints of dryland stress adaptation in Egyptian fat-Tail sheep and their divergence from East African and western Asia cohorts. *Sci. Rep.* **7**, 1–10 (2017).
49. Mekonnen, Y. A., Gültas, M., Effa, K., Hanotte, O. & Schmitt, A. O. Identification of candidate signature genes and key regulators associated with trypanotolerance in the Sheko Breed. *Front. Genet.* **10**, 1–20 (2019).
50. Dong, Y. *et al.* Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat. Biotechnol.* **31**, (2013).

51. Fontanesi, L. *et al.* A first comparative map of copy number variations in the sheep genome. *Genomics* **97**, 158–165 (2011).
52. Villa, S. T., Xu, Q., Downie, A. B. & Clarke, S. G. Arabidopsis protein repair L-isoaspartyl methyltransferases: Predominant activities at lethal temperatures. *Physiol. Plant.* **128**, 581–592 (2006).
53. D'Alessandro, A. *et al.* Protein-L-isoaspartate O-methyltransferase is required for in vivo control of oxidative damage in red blood cells. *Haematologica* **106**, 2726–2739 (2021).
54. Mishra, S. K., Tisel, S. M., Orestes, P., Bhangoo, S. K. & Hoon, M. A. TRPV1-lineage neurons are required for thermal sensation. *EMBO J.* **30**, 582–593 (2011).
55. Tan, C. H. & McNaughton, P. A. The TRPM2 ion channel is required for sensitivity to warmth. *Nature* **536**, 460–463 (2016).
56. Lawson, J. J., McIlwrath, S. L., Woodbury, C. J., Davis, B. M. & Koerber, H. R. TRPV1 unlike TRPV2 is restricted to a subset of mechanically insensitive cutaneous nociceptors responding to heat. *J. Pain* **9**, 298–308 (2008).
57. Comes, N., Buie, L. K. K. & Borrás, T. Evidence for a role of angiopoietin-like 7 (ANGPTL7) in extracellular matrix formation of the human trabecular meshwork: Implications for glaucoma. *Genes Cells* **16**, 243–259 (2011).
58. Costa, R. A., Cardoso, J. C. R. & Power, D. M. Evolution of the angiopoietin-like gene family in teleosts and their role in skin regeneration. *BMC Evol. Biol.* **17**, 1–21 (2017).
59. Zhong, T. *et al.* ENKD1 promotes epidermal stratification by regulating spindle orientation in basal keratinocytes. *Cell Death Differ.* **29**, 1719–1729 (2022).
60. Koster, M. I. & Roop, D. R. Mechanisms regulating epithelial stratification. *Annu. Rev. Cell Dev. Biol.* **23**, 93–113 (2007).
61. Shimada, T. *et al.* Targeted ablation of Fgf23 demonstrates an essential physiological role of FGF23 in phosphate and vitamin D metabolism. *J. Clin. Invest.* **113**, 561–568 (2004).
62. Bensaad, K., Cheung, E. C. & Vousden, K. H. Modulation of intracellular ROS levels by TIGAR controls autophagy. *EMBO J.* **28**, 3015–3026 (2009).
63. Cui, N. *et al.* Adrenomedullin-RAMP2 and -RAMP3 systems regulate cardiac homeostasis during cardiovascular stress. *Endocrinol. (United States)* **162**, 1–20 (2021).
64. Uetake, R. *et al.* Adrenomedullin-RAMP2 system suppresses ER stress-induced tubule cell death and is involved in kidney protection. *PLoS One* **9**, 1–12 (2014).
65. Moes, A. D., Van Der Lubbe, N., Zietse, R., Loffing, J. & Hoorn, E. J. The sodium chloride cotransporter SLC12A3: New roles in sodium, potassium, and blood pressure regulation. *Pflugers Arch. Eur. J. Physiol.* **466**, 107–118 (2014).
66. Kahle, K. T. *et al.* WNK4 regulates the balance between renal NaCl reabsorption and K<sup>+</sup> secretion. *Nat. Genet.* **35**, 372–376 (2003).
67. Ring, A. M. *et al.* An SGK1 site in WNK4 regulates Na<sup>+</sup> channel and K<sup>+</sup> channel activity and has implications for aldosterone signaling and K<sup>+</sup> homeostasis. *Proc. Natl. Acad. Sci. USA* **104**, 4025–4029 (2007).
68. Ring, A. M. *et al.* WNK4 regulates activity of the epithelial Na<sup>+</sup> channel in vitro and in vivo. *Proc. Natl. Acad. Sci. USA* **104**, 4020–4024 (2007).
69. Beall, C. M. Two routes to functional adaptation: Tibetan and Andean high-altitude natives. *Proc. Natl. Acad. Sci. USA* **104**, 8655–8660 (2007).
70. Stuart, J. A., Aibueku, O., Bagshaw, O. & Moradi, F. Hypoxia inducible factors as mediators of reactive oxygen/nitrogen species homeostasis in physiological normoxia. *Med. Hypotheses* **129**, (2019).
71. Ishibashi, M., Hayashi, A., Akiyoshi, H. & Ohashi, F. The influences of hyperbaric oxygen therapy with a lower pressure and oxygen concentration than previous methods on physiological mechanisms in dogs. *J. Vet. Med. Sci.* **77**, 297–304 (2015).
72. Zhang, Z. *et al.* Targeted sequencing identifies the genetic variants associated with high-altitude polycythemia in the Tibetan Population. *Indian J. Hematol. Blood Transfus.* **38**, 556–565 (2022).
73. Kimata, M. *et al.* p53 and TIGAR regulate cardiac myocyte energy homeostasis under hypoxic stress. *Am. J. Physiol. Hear. Circ. Physiol.* **299**, 1908–1916 (2010).
74. Bao, M. H. R. *et al.* Genome-wide CRISPR-Cas9 knockout library screening identified PTPMT1 in cardiolipin synthesis is crucial to survival in hypoxia in liver cancer. *Cell Rep.* **34**, 108676 (2021).
75. Hendrickson, S. L. A genome wide study of genetic adaptation to high altitude in feral Andean Horses of the páramo. *BMC Evol. Biol.* **13**, (2013).
76. Ghorbel, M. T. *et al.* Transcriptomic analysis of patients with tetralogy of Fallot reveals the effect of chronic hypoxia on myocardial gene expression. *J. Thorac. Cardiovasc. Surg.* **140**, 337–345.e26 (2010).
77. Hiser, L., Di Valentin, M., Hamer, A. G. & Hosler, J. P. Cox11p is required for stable formation of the Cu(B) and magnesium centers of cytochrome c oxidase. *J. Biol. Chem.* **275**, 619–623 (2000).
78. Lv, F. *et al.* Adaptations to climate-mediated selective pressures in sheep. *Mol. Biol. Evol.* **31**, 3324–3343 (2014).
79. Ro, M. *et al.* Association between arachidonate 5-lipoxygenase-activating protein (ALOX5AP) and lung function in a Korean Population. *Scand. J. Immunol.* <https://doi.org/10.1111/j.1365-3083.2012.02712.x> (2012).
80. Andrews, S. FastQC: A quality control tool for high throughput sequence data, Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. (2010).
81. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997v2* **00**, 1–3 (2013).
82. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
83. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *1297–1303* (2010) <https://doi.org/10.1101/gr.107524.110.20>.
84. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
85. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. (2021).
86. Purcell, S. *et al.* REPORT PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
87. Guo, J. *et al.* Comparative genome analyses reveal the unique genetic composition and selection signals underlying the phenotypic characteristics of three Chinese domestic goat breeds. *Genet. Sel. Evol.* **51**, 1–18 (2019).
88. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
89. Lawal, R. A. *et al.* Whole-genome resequencing of red junglefowl and indigenous village chicken reveal new insights on the genome dynamics of the species. *Front. Genet.* **9**, 1–17 (2018).
90. Rubin, C. J. *et al.* Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587–591 (2010).
91. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution (N.Y.)* **38**, 1358–1370 (1984).
92. Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
93. Gautier, M. & Vitalis, R. Reh An R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* **28**, 1176–1177 (2012).
94. Sherman, B. T. *et al.* DAVID: A web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, W216–W221 (2022).
95. Belay, S. *et al.* Whole-genome resource sequences of 57 indigenous Ethiopian goats. *Sci. data* **11**, 1–10 (2024).

## Acknowledgements

The authors would like to thank the flock owners who volunteered their animals for sampling and agricultural experts for their assistance during sampling. This project was initially supported by the BecA-ILRI Hub through the Africa Biosciences Challenge Fund (ABCF) Program and partially by the CGIAR Research Program on Livestock who supported one of the co-authors of this paper (GMT) for his PhD study and carried out all sample collection and laboratory works. We would also like to thank Tigray Agricultural Research Institute; University of Liverpool, Global Challenges Research Fund (GCRF) One Health Regional Network for the Horn of Africa (HORN) Project, from UK Research and Innovation (UKRI) and Biotechnology and Biological Sciences Research Council (BBSRC) (project number BB/P027954/1); Addis Ababa University, Department of Microbial Cellular and Molecular Biology; International Centre for Agricultural Research in the Dry Areas (ICARDA) and International Livestock Research Institute (ILRI) for their financial and logistical support.

## Author contributions

SB, GB, SM, HSW, KD, OL, OH, and JMM conceived and designed the study. GMT collected samples and extracted DNA. SB analysed the data and wrote the initial draft manuscript using the input materials from JMM and OH. GB, HN, TD, HJ, GMT, KD, OL, OH and JMM critically reviewed and interpreted the manuscript. AMA, and AT provided technical support during data analysis. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-65303-x>.

**Correspondence** and requests for materials should be addressed to S.B. or J.M.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024