communications biology

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42003-026-09856-9

Distinct adaptation and ancestral retention signals in African and European indigenous cattle genomes

Junxin Gao[1] ✉, Catarina Ginja[2], Ying Liu[1], Juha Kantanen[3], Nasser Ghanem[4], Donald Kugonza[5], Mahlako Makgahlela[6,7], Rodney Okwasiimire[3,8], Henk Bovenhuis[1], Martien A. M. Groenen[1] & Richard P. M. A. Crooijmans[1] ✉

Domestic cattle (*Bos taurus* and *Bos indicus*) underpin food security and livelihoods worldwide but face intensifying pressures from climate change, infectious disease, and inconsistent feed supplies. African and European indigenous cattle provide a natural comparative framework spanning gradients of climate, pathogen burden, and husbandry, and possess genomic mosaics comprising African taurine, European taurine, and indicine ancestry. We analyzed whole-genome sequences from 519 cattle across 24 African and European indigenous populations and 117 publicly available genomes from Africa, Asia, Europe, and the Americas. This dataset reveals admixture mosaics among major lineages and identifies 36 candidate genes exhibiting adaptive retention of ancestral alleles associated with response to heat stress (e.g., *HSPA12B*, *DDIT3*), immunity (*IRAK3*), productivity (*ACSF3*), and reproductivity (*SSMEM1*, *SPEF1*). Our study suggests that historical admixture introduced variation shaped by local ecological selection, clarifying how environmental heterogeneity drives the retention of advantageous alleles and informing sustainable breeding and diversity conservation.

Livestock worldwide are increasingly challenged by climate change, emerging infectious diseases, and resource limitations, threatening food security and rural livelihoods[1]. Cattle (*Bos taurus* and *Bos indicus*) are particularly important, providing nutrition, agricultural resources, and socio-economic and cultural value across continents[2–4]. However, sustaining cattle productivity under rapidly changing environmental and pathogenic pressures requires a deeper understanding of the genetic bases of adaptation, resilience, and productivity.

Across Africa and Europe, indigenous cattle provide a powerful natural comparative framework for studying adaptation because they span broad gradients of climate, pathogen exposure, and management intensity. Their genomes are complex mosaics of African taurine, European taurine, and indicine ancestries, each with distinct evolutionary origins and adaptive advantages[2,5,6]. The humpless taurine (*Bos taurus*) and humped indicine (*Bos indicus*) lineages diverged approximately 270-330 kya from distinct *Bos primigenius* populations[7–9] and were domesticated independently in the Near East and the Indus Valley about 10 and 8 kya[10–12]. Taurine cattle expanded westward into Europe around 8 kya[13], where artificial selection for milk and beef production produced the high-yielding European taurine breeds[14,15]. Before this westward expansion, taurine populations in both Europe and the Levant experienced secondary introgression from local aurochs, resulting in region-specific contributions of European and Near Eastern *Bos primigenius* to the taurine gene pool[16,17]. In parallel, taurine lineages also dispersed into Africa, where exposure to endemic pathogens such as East Coast fever, Rift Valley fever, and trypanosomiasis favored the retention of alleles conferring disease tolerance and resilience, exemplified by breeds such as N'Dama and Mututu[18–20]. Later, limited indicine introductions into Africa around ~3 kya are suggested by archaeological evidence[21], followed by a major wave of *Bos indicus* associated with Arab settlements along the East African coast from ~1.3 kya onward[22,23], contributing alleles associated with heat tolerance, parasite resistance, and metabolic efficiency[24,25]. These successive waves of migration, admixture,

[1]Animal Breeding and Genomics, Wageningen University & Research, Wageningen, The Netherlands. [2]BIOPOLIS, Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal and CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, Vairão, Portugal. [3]Natural Resources Institute Finland, Jokioinen, Finland. [4]Animal Production Department, Faculty of Agriculture, Cairo University, Giza, Egypt. [5]Department of Agricultural Production, College of Agricultural and Environmental Sciences, Makerere University, Kampala, Uganda. [6]Agricultural Research Council-Animal Production Institute, Irene, South Africa. [7]Department of Animal, Wildlife and Grassland Sciences, University of the Free State, Bloemfontein, South Africa. [8]Department of Agricultural Sciences, University of Helsinki, Helsinki, Finland.
✉e-mail: junxin.gao@wur.nl; richard.crooijmans@wur.nl

and local selection forged the genomic mosaics that define present-day African and European cattle[5].

Globally, more than 1000 cattle breeds are recognized. Of these, several hundred indigenous breeds across Africa and Europe are adapted to diverse agro-ecological conditions and reflect distinct genetic and adaptive histories[26–28]. Beyond the primary ancestral lineages of African taurine, European taurine, and African indicine, many African cattle represent hybrid composites formed through ancient and recent admixture. The term Sanga refers to African humped cattle that originated from early admixture between African taurine and indicine lineages following the introduction of Zebu (*Bos indicus*) into East Africa, whereas Zenga refers to crosses between Zebu and Sanga[2]. In the twentieth century, European commercial taurine breeds were further introduced into African populations to enhance production traits, such as Bonsmara (Afrikaner × Hereford × Shorthorn)[27,29]. Gene flow between Africa and Europe was bidirectional: several southern European and North African breeds retain signatures of ancient admixture linked to historical migration and trade. Archaeological and mitochondrial DNA evidence indicate African genetic contributions to Iberian cattle from Moorish occupation[30–32], while indicine ancestry in Italian breeds such as Chianina, Marchigiana, and Romagnola also reflects historical introgression into European populations[33].

Admixture between divergent lineages introduces ancestral variants that may be retained by selection when advantageous under local environmental conditions[34,35]. Such adaptive retention enhances fitness and supports local adaptation to diverse ecological niches, a process that has shaped genetic diversity in multiple species, including human Tibetan populations[36] and yaks[37]. In cattle, admixture among lineages with distinct evolutionary histories provides an opportunity to identify ancestral alleles associated with heat tolerance, disease resistance, and productivity[33,38–40]. However, the extent, distribution, and functional roles of these ancestry-retained genomic regions remain poorly characterized across breeds and continents.

Here, we analyzed whole-genome sequences from 519 cattle across 24 populations representative of African and European indigenous cattle, covering a wide ecological range, and compared these to a reference set of 117 previously sequenced cattle genomes[41,42]. By integrating local-ancestry inference, admixture dating, and selection analyses, we show that historical admixture between African and European taurine and indicine cattle has shaped the genomic architecture of indigenous breeds. Our findings highlight recurrently retained ancestry segments enriched for genes involved in immune response, thermotolerance, reproductivity, and productivity. Elucidating the genetic basis of these mosaic genomic patterns provides insight into the evolutionary mechanisms underlying cross-continental admixture and offers a foundation for future implementation of genetic intervention strategies and conservation of cattle genetic resources.

Results

Sample collection and identification of genetic variants

A total of 519 animals representing 24 indigenous cattle breeds/populations from Africa ($n = 240$) and Europe ($n = 279$) were sequenced at an average genome coverage depth of 10×, and supplemented with 117 animals from a previously sequenced genome reference set (average depth of 12×) (Supplementary Date 1). The reference set included 16 representative breeds from four distinct cattle ancestries (Supplementary Fig. 1a,b): African taurine (AFT: N'Dama and Muturu); European taurine (EUT: Hereford, Holstein Friesian, Jersey, and Angus); African indicine (AFI: Goffa, Boran, Kenana, and Ogaden); and Asian-American indicine (AAI) (American/Asian indicine: Brahman and Asian indicine: Nelore, Hariana, Tharparkar, Gir, and Red Sindhi).

Sequence reads were aligned to the ARS-UCD1.2 reference genome with an average alignment rate of 99.39% and genome coverage of 98.42%. Variant calling retained ~15.1 million high-quality autosomal single-nucleotide polymorphisms (SNPs), unevenly distributed across the whole genome (Supplementary Fig. 1c). Annotation details for these SNPs, including genomic locations and predicted functional impacts on proteins, are summarized in Supplementary Table 1.

African and European ancestral and indigenous cattle population structure

A principal component analysis (PCA) based on autosomal SNP data was conducted to investigate the genetic structure of African and European indigenous cattle, incorporating reference groups representing AFT, EUT, AFI, and AAI (Fig. 1a, b; Supplementary Fig. 2a, b). The PCA clearly separated the four reference groups, with European indigenous cattle clustering closely with the EUT reference, whereas the Portuguese breeds (Barrosã, Mertolenga, and Mirandesa) formed a distinct cluster positioned slightly apart from the main EUT group. Ugandan populations were located closer to the AFI reference cluster, while Egyptian and South African breeds occupied intermediate positions within a triangular space defined by the AFT, EUT, and AFI ancestries, suggesting historical admixture among these lineages. The neighbor-joining phylogenetic tree supported the PCA results, clearly separating African and European indigenous cattle into distinct reference clades (Fig. 1c). Within European taurine cattle, Portuguese breeds appeared closest to African taurine cattle, suggesting a closer genetic relationship or past gene flow between Iberian and African populations.

Genetic clustering analysis using ADMIXTURE at $K = 2$–4 revealed key ancestry patterns (Fig. 1d; Supplementary Fig. 2c). At $K = 2$, cattle separated into taurine and indicine lineages. At $K = 3$, three major ancestral components corresponding to AFT, EUT, and AAI were distinguished. At $K = 4$, Portuguese breeds, particularly Mertolenga, displayed bidirectional gene flow with the AFT component. Genome-wide genetic differentiation among breeds was low to moderate, with pairwise F_{st} values ranging from 0.05 to 0.41 (mean = 0.21; Supplementary Fig. 2d; Supplementary Table 2).

Genome-wide ancestry and admixture dating in African and European indigenous cattle breeds

To further investigate and quantify admixture levels in African and European cattle, we analyzed patterns of allele sharing using f_3 statistics and LD-based admixture dating. AFT (Muturu and N'Dama) showed no evidence of admixture in f_3 analysis assuming EUT and AAI as unadmixed proxies (Fig. 2a). A slightly negative f_3 value (N'Dama | Muturu, AAI) provides evidence for limited historical indicine introgression into N'Dama (Supplementary Table 3). In contrast, AFI reference, like other African indigenous cattle breeds, exhibited strong taurine-indicine mixed signals and were therefore excluded as ancestry groups (Fig. 2b). Among European breeds, the Iberian Mertolenga displayed the lowest positive f_3 value (Supplementary Table 3), suggesting a higher likelihood of historical gene flow. Egyptian and Sanga breeds further exhibited clear signatures of recent introgression from European commercial cattle (Fig. 2b; Supplementary Table 3). Admixture timing inferred using DATES indicates that the most recent detectable gene-flow episodes between EUT and AAI lineages occurred approximately 13.0–155.1 generations ago (Fig. 2c; Supplementary Table 4), noting that LD-based estimates can be weighted by more recent admixture when multiple events are present[43,44]. These estimates span very recent events in Tuli (Sanga cattle) to older signals in Kenana (African indicine cattle). Among European breeds, only the Iberian Mertolenga showed evidence consistent with introgression from African-related ancestry, with the lowest standard error observed under the AFTAAI–EUT model relative to alternative gene-flow directions (Fig. 2d; Supplementary Table 4).

Genome-wide local ancestry inference using LOTER[45] identified genomic segments derived from AFT, EUT, or AAI ancestry. Breed-level ancestry proportions were visualized as pie charts based on phased haplotypes (Fig. 2e). For example, in Afrikaner, ancestry along chromosome BTA11 (45–70 Mb) revealed a mosaic of retained ancestral components (Fig. 2f). Local ancestry profiles for representative breeds (Karamojong, Afrikaner, Egyptian cattle, and Mertolenga) illustrate distinct regional admixture patterns (Fig. 2g). Comprehensive mosaic ancestry maps computed in 50-kb windows for all breeds are provided in Supplementary Data 2.

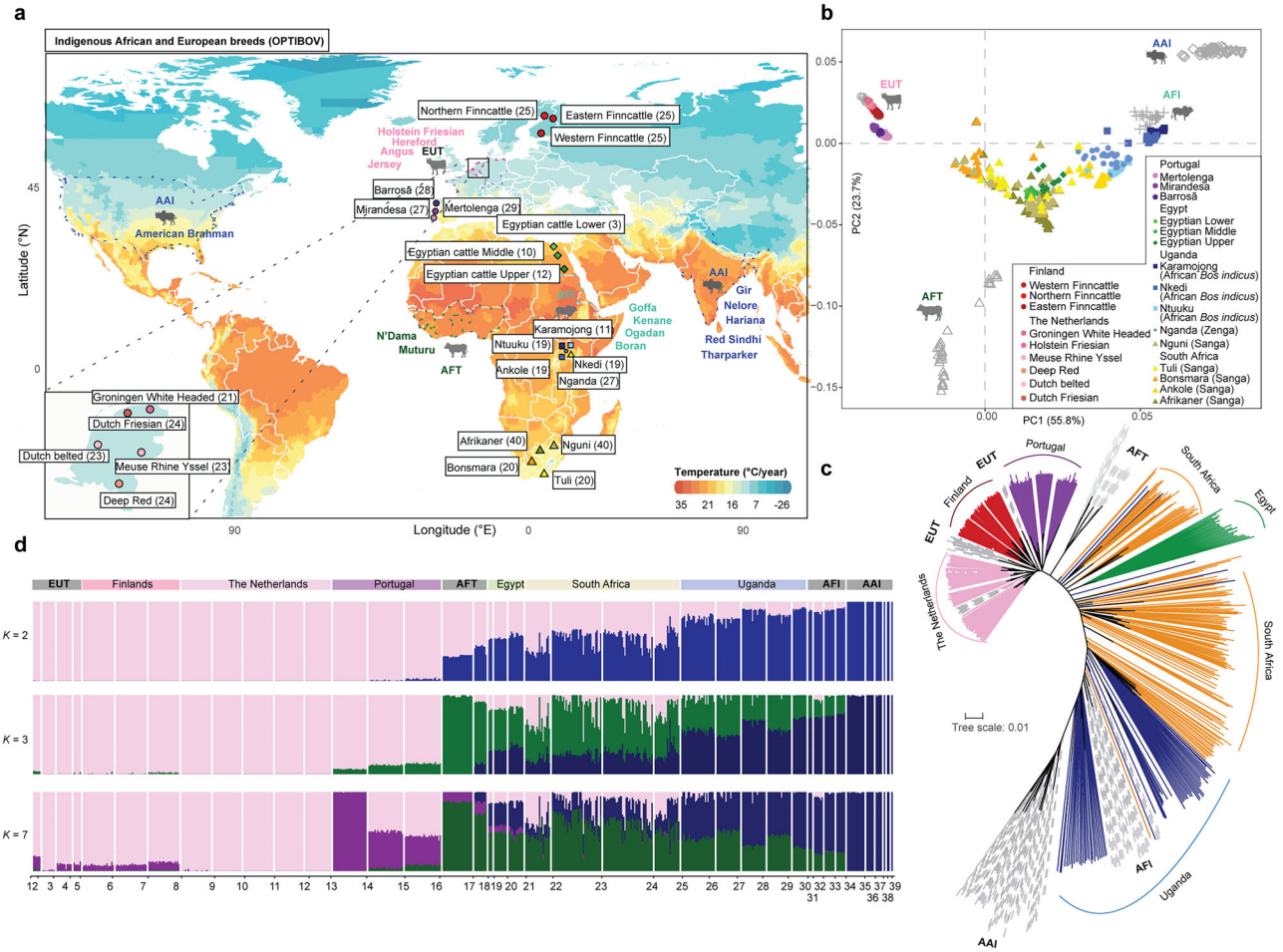

Fig. 1 | Population structure of African and European indigenous cattle. **a** Geographic distribution of sampled cattle populations across Africa and Europe. The dataset includes $n = 519$ individuals from 24 indigenous African and European breeds (OPTIBOV project) and $n = 117$ individuals from 16 reference breeds from the 1000 Bull Genomes Project. The dotted outlines indicate populations obtained from public databases, and the white boxes denote the 24 local breeds from the OPTIBOV project. Map images were generated by the authors using https://impactlab.org/map and GPS coordinates processed in the R package maps. **b** Principal component analysis (PCA) of 636 cattle representing 39 breeds. Solid symbols correspond to indigenous samples, and hollow gray symbols represent publicly available reference samples. The country of origin for each population is provided in Supplementary Data 1. A PCA of African, European and reference breeds is shown in Supplementary Fig. 1a and 2a,b. Reference breeds were categorized as follows: African taurine (AFT; N'Dama, Muturu), European taurine (EUT; Hereford, Holstein Friesian, Jersey, Angus), African indicine (AFI; Goffa, Boran, Kenana, Ogaden), and Asian-American indicine (AAI; Brahman, Nelore, Hariana, Tharparkar, Red Sindhi, Gir). **c** Neighbor-joining phylogenetic tree based on pairwise identity-by-state distances from autosomal SNPs. **d** ADMIXTURE analysis showing inferred ancestry proportions at $K = 2$, 3, and 4. Breed order: (1) Jersey, (2) Holstein Friesian, (3) Hereford, (4) Angus, (5) Western Finncattle, (6) Northern Finncattle, (7) Eastern Finncattle, (8) Deep Red, (9) Dutch Belted, (10) Dutch Friesian, (11) Meuse Rhine Yssel, (12) Groningen White Headed, (13) Mirandesa, (14) Barrosã, (15) Mertolenga, (16) Muturu, (17) N'Dama, (18) Egypt Lower, (19) Egypt Middle, (20) Egypt Upper, (21) Bonsmara, (22) Afrikaner, (23) Nguni, (24) Tuli, (25) Nganda, (26) Ankole, (27) Nkedi, (28) Ntuuku, (29) Karamojong, (30) Goffa, (31) Boran, (32) Kenana, (33)Ogaden, (34) Brahman, (35) Nelore, (36) Gir, (37) Hariana, (38) Red Sindhi, (39) Tharparkar.

Ancestry-enriched genomic regions in indigenous African cattle breeds

Genome-wide admixture analyses revealed that all African indigenous cattle breeds exhibit mosaic genomic compositions combining African taurine and indicine ancestry, with recent European taurine introgression detected in Egyptian and Sanga breeds. Ancestry-specific haplotypes conferring adaptive advantages tend to be retained over time, and their recurrence across breeds reduces the likelihood of random drift or recombination[46]. To identify ancestry-enriched regions potentially conferring adaptive advantages, local ancestry was inferred, and the top 1% of continuous 50-kb windows (651 windows/breed) were selected and compared across multiple breeds. Restricting the analysis to windows containing coding variants identified an average of 130 genes per breed enriched for AFT ancestry and 216 genes per breed enriched for AAI ancestry across the thirteen African breeds. In the seven populations with recent European taurine introgression, an average of 74 genes per breed showed enrichment for EUT ancestry. Breed-specific and ancestry-enriched genes are summarized in Supplementary Data 3.

All genes located within top AFT ancestry regions were significantly enriched ($^*p < 0.05$) for immune and metabolic processes (Fig. 3a), including interleukin-mediated signaling, antifungal innate immune response, detection of bacterium, and fatty acid β-oxidation, consistent with the long-term exposure of African taurine cattle to tropical disease pressures. For example, *AOX4* (BTA2) harbored high-impact variants and showed the highest retention rate (38.5%), overlapping across five African indigenous populations (Fig. 3b; Table 1). This gene encodes an aldehyde oxidase involved in oxidative metabolism and detoxification[47]. Genes located within top EUT ancestry were enriched ($^*p < 0.05$) for developmental and regulatory processes (Fig. 3c), such as embryonic skeletal system morphogenesis, ventricular septum development, and uterus development, suggesting contributions to tissue growth and organogenesis. Additional enrichments such as vitamin

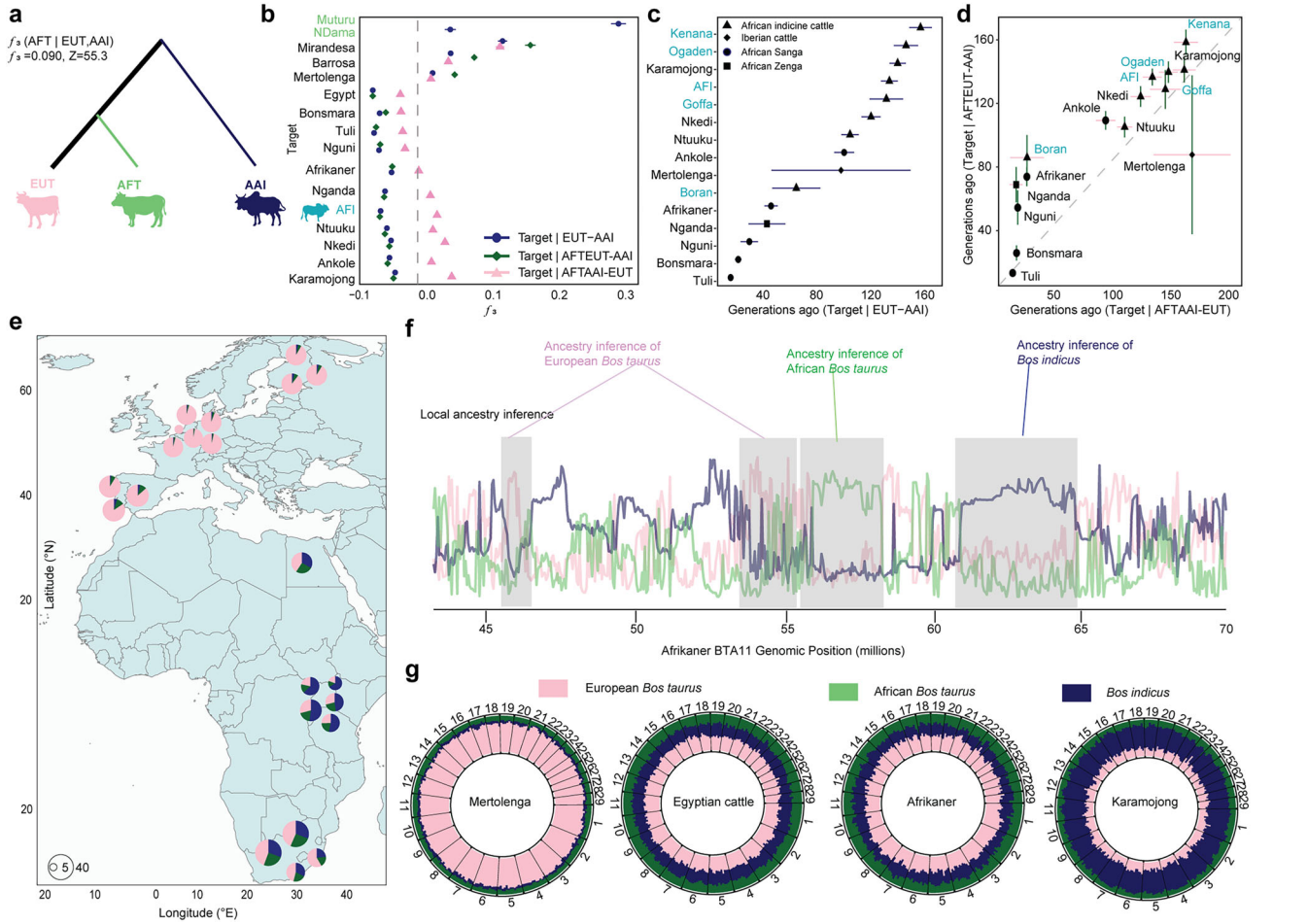

Fig. 2 | Genome-wide local ancestry inference and admixture dating in African and European indigenous cattle breeds. a f_3 statistics showing limited evidence of admixture in AFT ancestry, using EUT and AAI as ancestral taurine and indicine proxies: f_3(AFT; EUT, AAI). **b** f_3 statistics estimating admixture and gene flow among African taurine, European taurine, and indicine lineages in African and southern European breeds (X), using different ancestral proxies: EUT/AAI, AFTEUT/AAI, and AFTAAI/EUT. Each point represents the f_3 statistic; horizontal bars denote ±1 s.e. Results for (N'Dama; EUT, AAI) and (Muturu; EUT, AAI) are shown in green, and those for AFI breeds in blue. The statistics shown are population-level estimates; individual-level values are not available for qp3Pop. **c** Admixture times (in generations) estimated using DATES, with reference populations EUT ($n = 29$) and AAI ($n = 29$). Data points represent log-transformed admixture times (*flog*) ± 1 s.e. (Supplementary Table 4). **d** Admixture times (in generations) estimated using DATES for African and southern European breeds, using reference populations AFTAAI ($n = 63$)–EUT ($n = 29$) and AFTEUT ($n = 63$)–AAI ($n = 29$). Values are shown as *flog* ± 1 s.e. (Supplementary Table 4). The statistics shown are population-level estimates; individual-level values are not available for DATES. **e** LOTER-inferred ancestry proportions for each breed, displayed as pie charts. Each color represents the mean contribution of a distinct ancestral component (AFT, EUT, or AAI). **f** Example of local ancestry inference of BTA11 in Afrikaner ($n = 40$). Colored lines represent ancestry proportions across non-overlapping 50-kb windows. **g** Genome-wide mosaic ancestry composition of representative indigenous breeds from diverse geographic regions, visualized as Circos plots. Each vertical segment represents the average ancestry proportion within 10-Mb genomic intervals, colored according to ancestral origin.

A import, protein localization, and miRNA-mediated gene silencing may indicate refined control of differentiation and transcriptional regulation. Among these, *ZSCAN23* (BTA1), a zinc-finger transcription factor[48], carried moderate-impact variants and exhibited the highest retention rate (71.4%) across five African populations (Fig. 3d; Table 1). Genes with excess AAI ancestry were enriched ($^*p < 0.05$) for immune and cellular homeostasis pathways (Fig. 3e), including antigen presentation via MHC class, interleukin signaling, and cell-cycle, suggesting indicine-derived adaptation to hot, pathogen-rich environments. Genes with the highest retention frequencies (61.5%), such as *IRAK3* (Toll-like receptor–mediated immune signaling[49]; Fig. 3f), *DDIT3* (stress-induced apoptosis[50]), and *DCTN2* (Table 1; intracellular transport[51]), carried protein-altering variants and were shared across multiple African breeds ($n = 8$), likely contribute to adaptive resilience under tropical environments. The top genes showing the highest retention rates and protein-altering variants for each ancestry are summarized in Table 1.

Adaptive indicine-ancestry retention at the *DDIT3* locus

Indicine cattle are well known for their resilience to harsh tropical environments[34,52]. DNA-damage-inducible transcript 3 (*DDIT3*, also known as *CHOP*) encodes a transcription factor that functions as a central mediator of the unfolded protein response, integrating signals from endoplasmic reticulum stress, heat stress, and nutrient deprivation[53]. Among all candidate genes, *DDIT3* exhibited the most extensive indicine ancestry retention across African breeds and harbored amino acid-altering variants (Table 1).

To investigate the potential adaptive role of indicine ancestry at the *DDIT3* locus, we analyzed local ancestry patterns in cattle from four distinct African and southern European ecosystems (Fig. 4a). The indicine ancestry proportion at this locus was consistently elevated compared with the genome-wide average in all African breeds (paired *t*-test, $^*p < 0.05$), whereas European taurine breeds showed no significant deviation (ns $p = 0.58$; Fig. 4b,c; Supplementary Table 5). This pattern indicates that indicine-derived alleles at *DDIT3* have been preferentially retained in African cattle

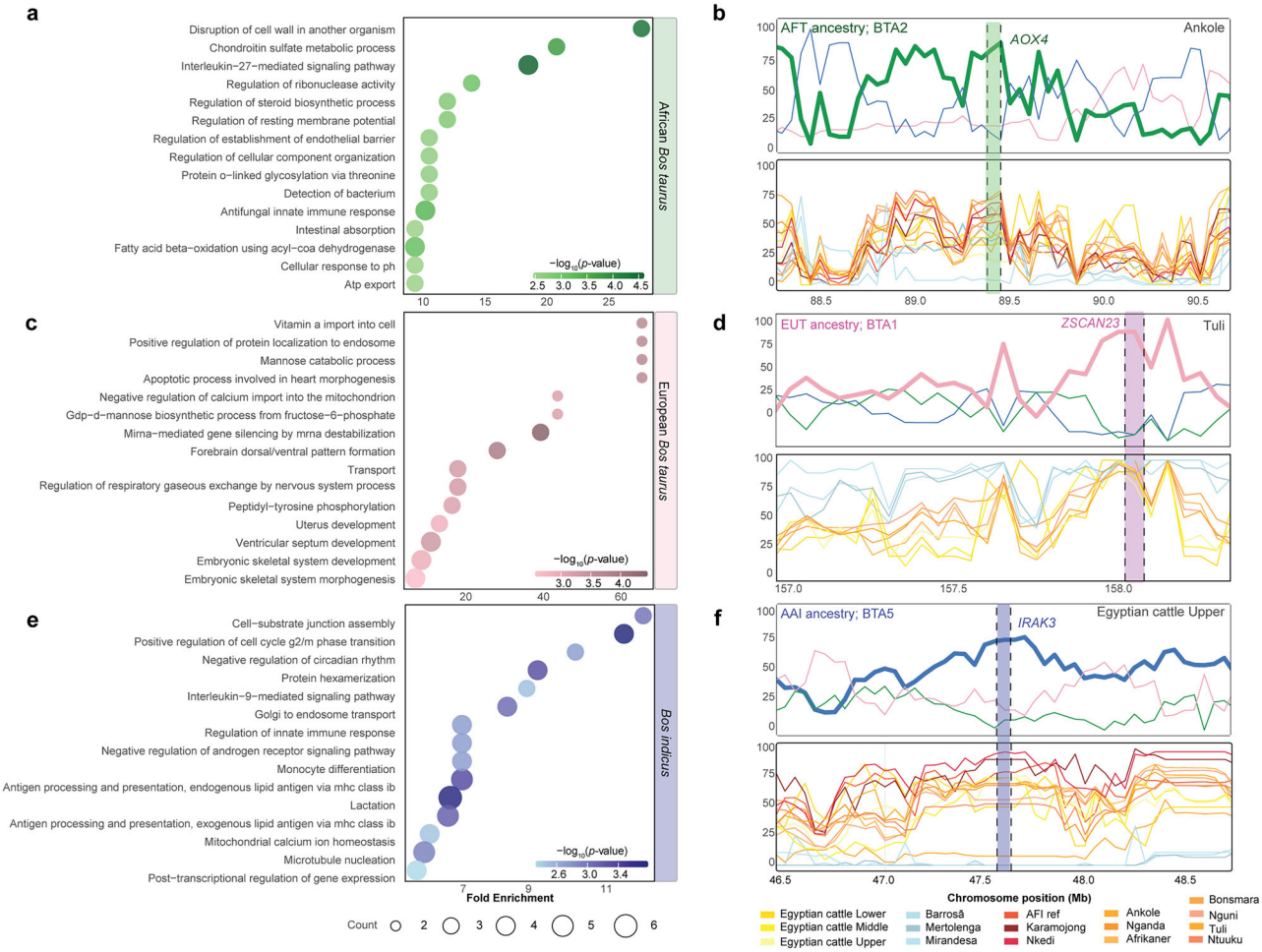

Fig. 3 | Functional enrichment and candidate genes within ancestry-enriched genomic regions. a, **c**, **e** Gene Ontology (GO) and KEGG pathway enrichment analyses of genes located within the top 1% of genomic regions showing excess ancestry, as identified from local ancestry inference using AFT, EUT, and AAI. GO terms related to biological processes are shown. Circle size represents the number of genes per term, the x-axis indicates fold enrichment, and colors denote -$\log_{10}$(*p*-value) based on modified Fisher's exact test. **b**, **d**, **f** Examples of prioritized candidate genes that harbor high-impact variants and show high retention rates, identified from the top 1% of ancestry-specific genomic windows inferred for AFT, EUT, and AAI ancestries. Candidate genes are shown for a representative breed, including all ancestry components, with ±1 Mb flanking regions around ancestry-proportion peaks. Each line of 50-kb non-overlapping windows is colored by ancestry contribution: AFT (green), EUT (pink), and AAI (dark blue). Candidate genes representing the target ancestry components across African and southern European populations are also shown, with each line representing a distinct population. The x-axis denotes genomic position (Mb, ARS-UCD1.2), and the y-axis indicates the proportional contribution of each ancestry component.

populations. Haplotype analysis revealed clear divergence between European taurine and indicine lineages at the *DDIT3* locus (Fig. 4d), encompassing both coding and regulatory regions. Two missense mutations (rs439088019 and rs210331613) displayed strong geographic and subspecies-specific allele frequency patterns (Fig. 4e; Supplementary Table 6). The lysine-to-glutamic acid substitution (rs439088019; SIFT = 0.04; deleterious) alters residue charge and may influence DNA or protein-binding affinity, whereas the leucine-to-valine substitution (rs210331613; SIFT = 0.08; tolerated) could affect protein folding and stability (Fig. 4f).

African taurine adaptive introgression in southern Portuguese cattle

Portuguese taurine cattle form a distinct genetic group within European taurine populations, likely shaped by historical introgression from AFT and subsequent geographic isolation imposed by the Pyrenees, which restricted gene flow and preserved unique haplotypes. Our admixture analyses suggest that indigenous Portuguese breeds, particularly Mertolenga, share closer genetic relationships and possible past gene flow with AFT populations (Figs. 1, 2), whereas the mountainous Barrosã and northeastern Mirandesa breeds show lower levels of African ancestry. Consistent with these genomic patterns, Mertolenga cattle have been reported to possess strong homeothermic capacity under heat stress conditions[54].

To identify putative AFT-derived adaptive introgression loci, we employed three complementary approaches. First, we inferred the top 1% of continuous 50-kb local ancestry windows with the highest AFT proportions in Mertolenga. This analysis yielded 651 candidate windows showing excess African taurine ancestry, overlapping 107 genes harboring gene-coding variants (Supplementary Data 3). These genes were significantly enriched for biological processes related to extracellular matrix organization, mitochondrial function, and metabolic regulation (Fig. 5a), suggesting potential contributions from AFT introgression to enhanced stress tolerance and energy homeostasis.

Next, we examined selective sweep signals to detect possible adaptive retention of AFT alleles, using both the Fixation index (F_{st}; top 1% = 0.24) and reduced nucleotide diversity (-$\log_2$(π_Mertolenga/π_EUT) in comparison with European taurine populations. The top 1% of 50-kb windows comprised 992 candidate regions, which were intersected with the previously identified AFT-retained gene set. Among these, nine genes carrying coding variants from the AFT retention analysis showed overlapping signatures of selection (Fig. 5b; Table 2). These genes formed the largest

https://doi.org/10.1038/s42003-026-09856-9

Table 1 | Putative genes identified within the top 1% of local ancestry–enriched regions and showing the highest retention rates, with protein-altering variants across African indigenous cattle populations

Gene	Ancestry	Variant impact	Group	Populations with signals	Retention rates (%)	Previous literature / known function
SSMEM1	African *Bos taurus* (AFT)	Moderate*	African cattle	AFI, Afrikaner, Bonsmara, Karamojong, Nganda, Nguni, Nkedi, Tuli	62%	Sperm-specific membrane protein 1; associated with male fertility[79]
XDH		Moderate	African cattle	Afrikaner, Ankole, Egyptian cattle Middle, Karamojong, Nganda, Nkedi, Ntuuku, Tuli	62%	Body's defense against infection (ROS-mediated antimicrobial activity)[75]
CALHM5		Moderate	African cattle	Afrikaner, Egyptian cattle Upper, Karamojong, Nguni, Nkedi, Tuli	46%	Electrophysiological and cellular homeostatic resilience[73]
CALHM6		Moderate	African cattle		46%	
DSE		Moderate	African cattle	Afrikaner, Egyptian cattle Upper, Karamojong, Nguni, Nkedi, Tuli	46%	Dermatan-sulfate epimerase; involved in wound healing and inflammation[76]
AOX4		High*, Moderate	African cattle	AFI, Ankole, Karamojong, Nkedi, Ntuuku	39%	Xenobiotic and redox metabolism; under selective sweeps in subtropical Holstein populations[74]
AOX2		Moderate	African cattle		39%	
SCARB2		Moderate	African cattle	AFI, Afrikaner, Karamojong, Nkedi, Tuli	39%	Immune response; linked to foot-and-mouth disease resistance in cattle[77]
TTC27		Moderate	African cattle	Ankole, Nganda, Nguni, Ntuuku, Tuli	39%	Fat metabolism, inflammatory response, and milk-protein composition[78]
ZSCAN23	European *Bos taurus* (EUT)	Moderate	Egyptian and Sanga cattle	Egyptian cattle Middle, Bonsmara, Tuli, Nguni, Afrikaner	71%	Male fertility in cattle[80]
RNF145		Moderate	Egyptian and Sanga cattle	Bonsmara, Tuli, Nguni, Afrikaner	57%	Membrane fluidity and lipid homeostasis[85]
ACSF3		Moderate	Egyptian and Sanga cattle	Bonsmara, Tuli, Afrikaner	43%	Lipid metabolism; intramuscular fat traits in beef cattle[81]
DIMT1		Moderate	Egyptian and Sanga cattle	Tuli, Nguni, Afrikaner	43%	Lifespan regulation and ribosome biogenesis[86]
MRPL13		Moderate	Egyptian and Sanga cattle	Tuli, Nguni, Afrikaner	43%	Mitochondrial function and energy metabolism; linked to feeding behavior in Nellore cattle[87]
MTBP		Moderate	Egyptian and Sanga cattle	Tuli, Nguni, Afrikaner	43%	Pregnancy maintenance in lactating dairy cows[82]
NCOA2		Moderate	Egyptian and Sanga cattle	Bonsmara, Tuli, Afrikaner	43%	Growth and reproduction via steroid/retinoid signaling[83]
TACR3		Moderate	Egyptian and Sanga cattle	Egyptian cattle Lower, Tuli, Nguni	43%	Organ development and milk quality in bovids[84]
DDIT3	*Bos indicus* (AAI)	Moderate	African cattle	Afrikaner, Ankole, Bonsmara, Egyptian cattle Upper, Nganda, Nguni, Ntuuku, Tuli	62%	Thermal and nutritional stress tolerance[146–148]
DCTN2		Moderate	African cattle	Afrikaner, Ankole, Bonsmara, Egyptian cattle Upper, Nganda, Nguni, Ntuuku, Tuli	62%	Tropical adaptation in African indicine cattle[91]
GPCPD1		Moderate	African cattle	AFI, Afrikaner, Ankole, Karamojong, Nguni, Nkedi, Ntuuku, Tuli	62%	Phospholipid metabolism[92]
IRAK3		High, moderate	African cattle	Afrikaner, Bonsmara, Egyptian cattle Middle, Egyptian cattle Upper, Nkedi, Tuli	46%	Regulation of inflammation[93]
ANTXR2		Moderate	African cattle	Egyptian cattle Lower, Egyptian cattle Middle, Egyptian cattle Upper, Karamojong, Nguni	46%	Climate adaptation[94]
SHLD1		Moderate	African cattle	AFI, Ankole, Nguni, Nkedi, Ntuuku, Tuli	46%	DNA double-strand break repair under stress[95]
ZMAT2		High	African cattle	AFI, Karamojong, Nganda, Nkedi, Ntuuku	39%	Zinc-finger matrin-type protein in spliceosome[149]

Table 1 (continued) | Putative genes identified within the top 1% of local ancestry–enriched regions and showing the highest retention rates, with protein-altering variants across African indigenous cattle populations

Gene	Ancestry	Variant impact	Group	Populations with signals	Retention rates (%)	Previous literature / known function
EIF5B		Moderate	African cattle	Egyptian cattle Lower, Egyptian cattle Middle, Egyptian cattle Upper, Nkedi, Ntuuku	39%	Regulates DNA damage-inducible protein synthesis during stress response[96]
DND1		Moderate	African cattle	AFI, Karamojong, Nganda, Nkedi, Ntuuku	39%	Germ cell maintenance in a model system[97]
PLXNB2		Moderate	African cattle	Ankole, Karamojong, Nganda, Nkedi, Ntuuku	39%	Calving performance in dairy and beef cattle[98]

[a]High-impact variants were defined as predicted disruptive changes, including stop-gain, frameshift, and splice-acceptor/donor mutations. Moderate-impact variants included non-synonymous (missense) substitutions or other amino-acid-altering changes predicted to affect protein function. Low-impact variants were those predicted to have minimal functional consequences (https://pcingola.github.io/SnpEff/).

genomic cluster on BTA13 (51.39–51.49 Mb), encompassing *SPEF1*, *CDC25B*, *SIGLEC1*, and *HSPA12B* (*HSP70*) (Fig. 5c and Supplementary Data 2; Fig. 5d,e and Supplementary Data 4), suggesting a strong candidate hotspot of AFT adaptive introgression in southern Portuguese cattle.

Discussion

We performed a comparative framework of 24 indigenous cattle populations spanning gradients in climate, pathogen pressure, and husbandry (12 from Egypt, Uganda, and South Africa; 12 from Finland, the Netherlands, and Portugal). Population-structure analyses resolved their genomic relationships and ancestry components, after which we quantified admixture evidence, dated admixture events, and mapped genome-wide mosaic ancestry from African taurine, European taurine, and indicine sources for each local population. Across admixed breeds, specific genes harboring coding variants were recurrently enriched within ancestry-biased windows, implicating disease-resistance, developmental, stress-response, and lipid or energy-metabolism pathways with distinct ancestry-linked emphases. We also identified putatively adaptive introgression of African taurine lineages into Portuguese cattle, highlighting cross-continental gene flow and its potential role in thermophysiological resilience. The mosaic patterns of ancestry and the recurrence of functionally coherent signals across populations support the hypothesis that contributions from multiple ancestral lineages have facilitated adaptation of indigenous African and northern European cattle to their diverse ecological contexts.

Today's African taurine cattle are distinctive relative to European taurine lineages, with archaeogenetic evidence suggesting region-specific domestication dynamics in North and Northeast Africa and possible genetic inputs from local aurochs, followed later by indicine introgression within Africa from lineages already adapted to hot climates in India and Pakistan[2,55,56]. In the millennia following Near Eastern domestication, African taurine cattle dispersed into ecological zones shaped by tsetse infestation, episodic drought, and endemic infectious disease pressure, leading to repeated selection on alleles associated with trypanotolerance, immune modulation, and cellular stress responses[57]. Muturu and N'Dama were treated as representative African taurine populations[58]; however, N'Dama exhibits a more complex demographic history with possible limited indicine admixture. Across our analyses, both breeds are clearly distinct from European taurine and indicine cattle, consistent with previous large-scale studies reporting that only Muturu and N'Dama lack detectable recent indicine admixture among African cattle breeds[59]. However, finer-scale analyses indicated that Muturu represents a more extreme African taurine lineage, whereas N'Dama shows limited indicine admixture relative to Muturu[22,59,60]. Accordingly, rather than assuming N'Dama to be uniformly unadmixed, our mosaic analyses using Muturu alone or both breeds as African taurine references produced nearly identical results. In contrast, European taurine lineages underwent continuous breeding for economic traits such as milk and meat production under temperate husbandry, and developing multiple world-renowned breeds[61], like Holstein Friesian and Jersey were selected for milk yield[62], Hereford and Angus were bred for beef productivity[3]. Our results are consistent with previous reports of recent cross-continental introgression from European commercial breeds into African cattle, as indicated by the evidence of European taurine ancestry in the autosomal genomes of Egyptian and Sanga populations[27,29].

Indicine lineages evolved in hot, seasonal, pathogen-dense environments of South Asia, where selection favored thermotolerance, ectoparasite resistance, and efficient water and energy balance[63], which provided strong preadaptation for African agro-ecologies. Historical livestock movements through inland routes and coastal trade, coupled with managed crossbreeding, subsequently combined the advantages of African taurine trypanotolerance, European taurine productivity, and indicine heat resistance. This admixture process facilitated the emergence and persistence of composite crossbreds across African and southern European ecosystems[57,64,65], consistent with the ancestry profiles we observed in our present-day genomic analyses of African and European indigenous cattle.

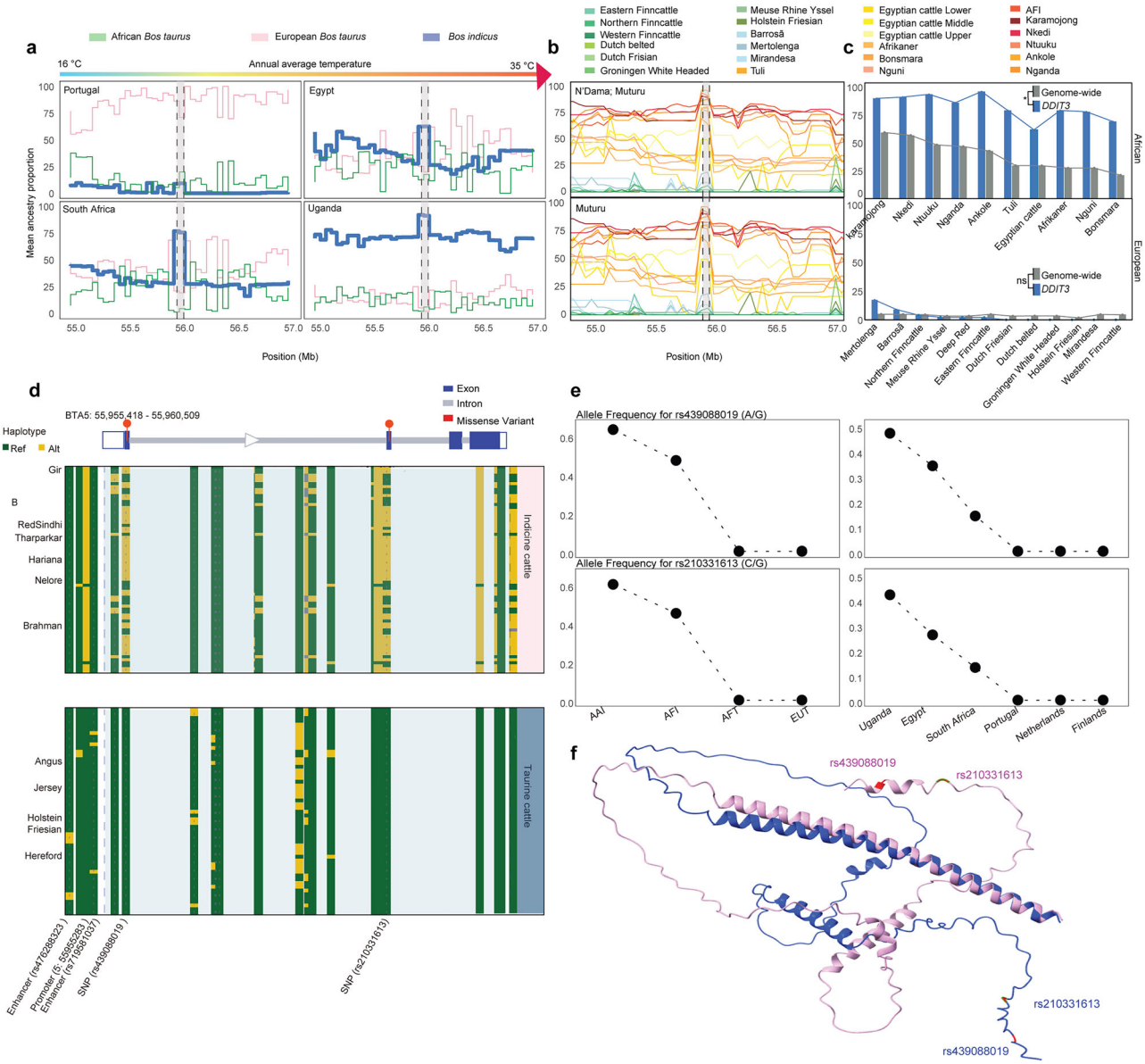

Fig. 4 | Adaptive indicine-ancestry retention at the *DDIT3* locus. a Local ancestry composition across 50-kb non-overlapping windows spanning ±1 Mb (ARS-UCD1.2) around the *DDIT3* locus in cattle breeds from four distinct African and southern European ecosystems: Portugal ($n = 84$), Egypt ($n = 25$), South Africa ($n = 120$), and Uganda ($n = 95$). The arrow indicates the annual average temperature of each country. Ancestry components are AFT (green), EUT (pink), and AAI (dark blue). The x-axis shows genomic position (Mb), and the y-axis indicates the proportion of each ancestry component. **b** Indicine-ancestry representation at the *DDIT3* locus across all African and European populations, with each line color corresponding to a different population (sample sizes are provided in Supplementary Data 1). To minimize potential effects of historical indicine introgression in N'Dama, two AFT reference configurations are shown: combined N'Dama and Muturu, and Muturu only. **c** Statistical comparison of AAI-ancestry proportions at *DDIT3* locus (BTA5: 55.95–55.97 Mb) relative to genome-wide averages. Paired *t*-tests ($^*p < 0.05$) show significantly elevated AAI ancestry in African breeds, whereas European taurine breeds show no significant deviation (ns $p > 0.05$). Breed-specific sample sizes are provided in Supplementary Data 1. **d** Haplotype-sharing structure at the *DDIT3* locus on BTA5. Haplotypes are hierarchically clustered within EUT and AAI groups. SNPs include gene-coding and putatively functional variants located in promoter and enhancer regions. SNP states are color-coded: green as the reference allele; yellow as the alternative allele. **e** Allele frequencies of two missense mutations (rs439088019 and rs210331613) are shown by reference group (AAI, $n = 29$; AFI, $n = 25$; AFT, $n = 34$; EUT, $n = 29$) and by geographic region: Uganda ($n = 95$), Egypt ($n = 25$), South Africa ($n = 120$), Portugal ($n = 84$), the Netherlands ($n = 120$), and Finland ($n = 75$). The y-axis represents the frequency of the alternative allele (Supplementary Table 6). **f** Predicted structural changes in the DDIT3 protein caused by two missense mutations differentiating EUT (pink) and AAI (dark blue) cattle.

This long history of intercontinental movement and managed breeding has left a lasting genomic imprint on both African and southern European cattle. Admixture tests and LD-decay dating revealed multiple waves of gene flow among African taurine, European taurine, and indicine lineages, reflecting complex demographic and selective histories that shaped the genomes of present-day indigenous populations. When multiple-wave admixture events have occurred, estimation of admixture times tends to get weighted towards the most recent event[44]. Across African populations, both f_3 and DATES analyses revealed extensive admixture between taurine and indicine cattle, with inferred admixture times ranging from ~13 to 155 generations ago (assuming a cattle generation time of 5–7 yr[2]), corresponding to approximately ~65–1085 yr ago. These timescales are consistent with historical records of major indicine introduction into Africa and with previous genomic estimates of extensive indicine–taurine crossing around ~750–1050 yr ago in African local cattle[2]. Together, these results reflect a continuum of admixture histories, spanning from recent Sanga and

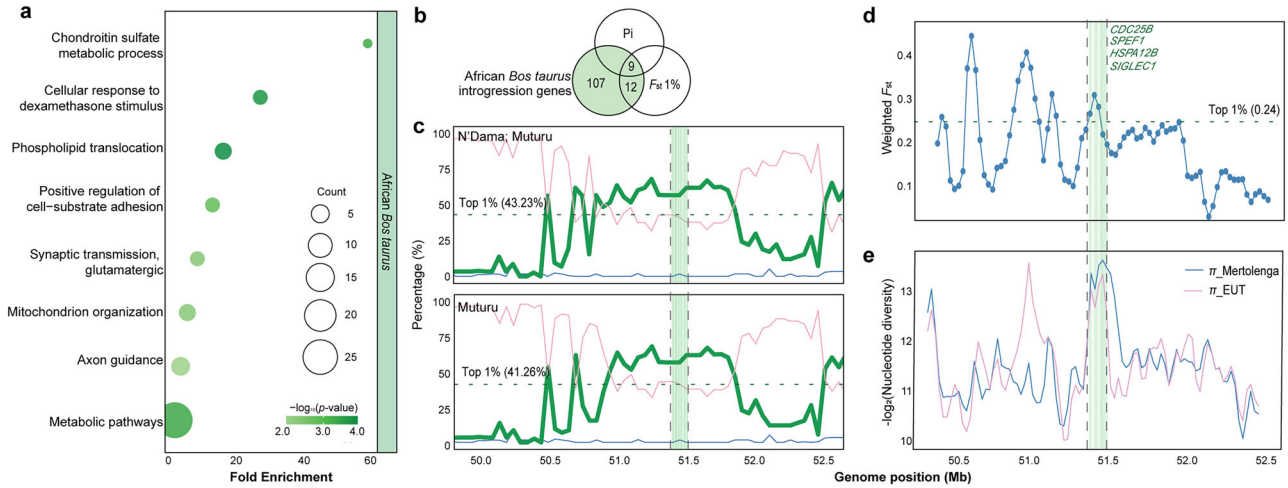

Fig. 5 | Ancestry introgression in Portuguese taurine cattle. a GO and KEGG pathway enrichment analyses of genes located within the top 1% of genomic regions showing excess AFT ancestry. GO terms related to biological processes are shown. Circle size represents the number of genes per term, the x-axis indicates fold enrichment, and colors denote -$\log_{10}$(p-value) based on modified Fisher's exact test. **b** Venn diagram showing overlap between AFT-retained genes under selective sweeps detected by F_{st} and nucleotide diversity. **c** Candidate regions introgressed from African taurine ancestry in Mertolenga ($n = 29$) using two AFT reference panels: combined N'Dama and Muturu, and Muturu only. The highlighted region (green) on BTA13 (51.39–51.49 Mb) contains a genomic cluster including *SPEF1*, *CDC25B*, *SIGLEC1*, and *HSPA12B*(*HSP70*) **d** Pairwise weighted F_{st} values calculated between Mertolenga ($n = 29$) and EUT ancestry ($n = 29$) for the genomic cluster on BTA13 (51.39-51.49 Mb). The top 1% F_{st} threshold was 0.24 (Supplementary Data 4). **e** Reduced nucleotide diversity analysis -$\log_2$(π_Mertolenga/π_EUT) for the same genomic cluster on BTA13 (51.39-51.49 Mb). Values were calculated using 50-kb windows with a 25-kb step size spanning ±1 Mb (ARS-UCD1.2) around the candidate region (Supplementary Data 4).

Table 2 | Putative genes identified within adaptive African taurine introgressed regions in Portugal

Gene	Ancestry	Variant rank	Populations	−log$_2$(ratio)*	F_{st} (top 1%: 0.24)	Previous literature / known function
SPEF1	African *Bos taurus* (AFT)	Low	Mertolenga	0.37	0.28	Reproductive resilience[105]
CDC25B		Low	Mertolenga	0.04	0.26	Reproduction[108]
SIGLEC1		Low	Mertolenga	0.3	0.28	Immune modulation and reproduction[107]
HSPA12B		Low	Mertolenga	0.37	0.28	Heat-shock protein; thermotolerance and cellular protection[99–101]
NUDT3		Moderate	Mertolenga	1.07	0.36	Adipocyte biology and lipid metabolism in pigs[102]
TMC2		Moderate	Mertolenga	0.31	0.28	Sensory transduction in auditory and vestibular hair cells[134]
PCMTD2		Low	Mertolenga	0.68	0.3	Protein repair/methyltransferase, component of ubiquitin complex[150]
MYT1		Moderate	Mertolenga	0.97	0.26	Lactation persistency[151]
ATP2C2		High	Mertolenga	0.87	0.26	Regulation of calcium (Ca2+) homeostasis[104]

*−$\log_2$(ratio): Positive values of −$\log_2$(π_Mertolenga/π_EUT) indicate reduced nucleotide diversity in Mertolenga relative to EUT across 50-kb windows, consistent with localized selection.

Zenga composite breeds (~13–98 generations ago) to older admixture events involving Egyptian (~120 generations ago) and African indicine populations (~155 generations ago). More recent gene flow in Sanga cattle corresponds to the twentieth-century introduction of intensive European commercial breeds, resulting in taurine × indicine composites with enhanced productivity traits, as supported by our detection of European taurine ancestry[39]. In southern Europe, African taurine genetic contributions were most evident in the Portuguese Mertolenga breed. DATES analyses indicated that admixture between AFT–related and EUT yielded the lowest standard error, with an inferred admixture time of ~167 generations ago, corresponding to approximately ~835–1169 yr ago. This timing overlaps with mitochondrial DNA evidence from Iberian cattle and with North African–Iberian interactions during the Moorish occupation of the Iberian Peninsula (8th–13th centuries AD)[30–32,66,67]. The geographic barrier of the Pyrenees likely restricted northward gene flow thereafter, preserving these ancestral components in southern Portuguese populations.

To uncover adaptive ancestry patterns, we integrated breed-specific local ancestry inference on extended haplotypes with variant annotation and selective statistics (e.g., retention rate, selective sweeps, and nucleotide diversity), while controlling for confounding effects of random genetic drift and recombination. Because recombination fragments ancestral haplotypes, whereas selection preferentially retains advantageous alleles[68], we observed uneven ancestry frequencies across the genome, with recurrent enrichment of functional genes. By focusing on ancestry signals consistent across breeds inhabiting similar environments, we reduced stochastic effects and false-positive signals, supporting adaptive selection as a major driver of ancestry retention. Similar patterns of uneven ancestry retention have been reported in human populations[69].

Distinct functional emphases were associated with each ancestry. Among AFT-retained genomic segments, genes harboring protein-coding variants were significantly enriched for immune, metabolic, and cellular homeostasis functions, likely shaped by prolonged exposure to pathogen-rich tropical environments[57,70]. Enriched GO/KEGG terms included interleukin-mediated signaling, antifungal innate immunity, and detection of bacterium (host defense[71]); fatty-acid β-oxidation, ATP release/export, and regulation of steroid biosynthesis (energy metabolism under nutritional

or heat stress[72]); and regulation of membrane potential and cellular pH, consistent with top-ranked genes *CALHM5/6* involved in ion and ATP conductance[73]. Top retention-rate candidate genes across African breeds included *AOX2* and *AOX4* (xenobiotic and redox metabolism; selective sweeps in subtropical Holstein populations)[74]; *XDH* (ROS-mediated antimicrobial activity in cattle)[75]; *DSE* (extracellular matrix remodeling and wound repair)[76]; *SCARB2* (immune defense and disease resistance in cattle)[77]; *TTC27* (metabolic and inflammatory regulation in cattle)[78]; and *SSMEM1* (male fertility)[79].

European taurine cattle have undergone intensive artificial selection for milk and meat production[3]. All genes harboring protein-coding variants within EUT-enriched windows were predominantly associated with developmental and regulatory processes, including skeletal and organ morphogenesis, neurodevelopment, post-transcriptional regulation, and vitamin A metabolism. Highest retention-rate candidate genes across Africa breeds with recent EUT introgression included: *ZSCAN23* (transcriptional regulation and male fertility in cattle)[80]; *ACSF3* (mitochondrial lipid metabolism and intramuscular fat content in beef cattle)[81]; *MTBP* (pregnancy maintenance in lactating dairy cows)[82]; *NCOA2* (growth and reproduction)[83]; *TACR3* (reproductive-axis signaling and milk quality in bovids)[84]; *RNF145* (cholesterol and membrane-lipid homeostasis)[85]; *DIMT1* (ribosome biogenesis and longevity in model systems)[86]; and *MRPL13* (mitochondrial energy metabolism and feeding behavior in Nellore cattle)[87].

Indicine cattle possess key adaptive traits such as heat tolerance, disease resistance, and resilience to nutritional stress, which are essential for survival in tropical environments[24,25]. Consistent with these phenotypes, AAI-retained genes harboring protein-coding variants were predominantly associated with stress tolerance, cell-cycle regulation, and immune control, including pathways related to intracellular transport, mitochondrial homeostasis, antigen presentation, and interleukin-mediated signaling. Among genes with the top retention rate across African breeds, *DDIT3 (CHOP)* exhibited the strongest indicine-retention signal. This stress-responsive transcription factor is activated by protein misfolding, heat, and nutrient deprivation and regulates apoptosis, differentiation, and metabolic adaptation[88]. Its signal has also been reported in Tharparkar and Karan-Fries cattle and African indicine cattle on drylands adaptation[89,90], suggesting a pivotal role in thermal and nutritional stress tolerance. *DCTN2* flanking with *DDIT3*, involved in intracellular transport, has likewise been reported under tropical selection in Ethiopian cattle[91]. Additional prioritized genes included *GPCPD1* (membrane remodeling and fatigue resistance[92]); *IRAK3* (negative regulator of inflammatory signaling[93]); *ANTXR2* (endothelial stability and climate adaptation in Mediterranean cattle[94]); *SHLD1* and *EIF5B* (DNA damage repair and translational control during stress responses[95,96]); *DND1* (germ cell maintenance[97]); and *PLXNB2* (developmental stability and perinatal survival in cattle[98]).

Portuguese Iberian breeds, particularly Mertolenga, represent a distinct southern European taurine lineage characterized by enhanced homeothermic capacity under heat stress[54], consistent with the AFT ancestry introgression detected in our analyses. Genes harboring protein-coding variants within AFT-derived regions were significantly enriched for pathways related to chondroitin sulfate metabolism, cellular stress responses, phospholipid translocation, mitochondrial organization, cell adhesion, axon guidance, and synaptic function. Among nine prioritized putative adaptive genes, *HSPA12B* (*HSP70* family), has been implicated as a key environmental stress–responsive chaperone involved in thermotolerance and cellular protection in cattle[99–101]; *NUDT3* is involved in adipocyte biology and lipid metabolism[102]; *TMC2* and *ATP2C2* are associated with sensory transduction[103] and calcium homeostasis[104]; and *SIGLEC1*, *CDC25B*, and *SPEF1* have been associated with immune regulation and reproductive resilience in cattle studies[102,105–108].

A major strength of this study is the broad sampling of African and European indigenous cattle, which enabled high-resolution inference of admixture and local ancestry. We identified several candidate genes that are well characterized in model organisms but remain underexplored in cattle. Moreover, our analyses provide autosomal DNA evidence consistent with historical North African-Iberian gene flow during the Moorish occupation of the Iberian Peninsula, bridging genomic and historical perspectives. However, restricting the analysis to the top present of ancestry-excess windows and retention rates, while effective in reducing false positives, may overlook biologically relevant regions or genes. Although we identified key adaptive genes exhibiting high ancestral retention rates and associated with thermotolerance, reproduction, productivity, metabolism, and immunity, many loci remain uncharacterized and will require functional validation across diverse environmental contexts.

In conclusion, despite the complexity of crossbreeding between African and European cattle and the possible historical domestication of cattle outside their original geographic regions, domestic cattle are now found across diverse agro-ecological zones in both Africa and Europe. Our findings demonstrate that admixture evidence between taurine (African/European) and indicine cattle has contributed to advantageous ancestral retention signals in present-day indigenous populations, providing comprehensive insights into local adaptation in African and European cattle. Importantly, our research highlights the critical need for conserving local indigenous breeds, as their unique ancestral compositions and adaptive genomic signatures represent invaluable genetic resources for sustainable cattle breeding programs. Preserving these local breeds will help maintain the genetic diversity essential for adaptation to future diverse and changing environments.

Methods

Ethics approval

Blood samples were collected during the animals' annual health inspections, conducted by licensed veterinarians. Prior to sample collection, written informed consent was obtained from each animal's owner. In Finland, animal handling procedures and sample collections were performed in accordance with the legislation approved by the Regional State Administrative Agency for Southern Finland (ESAVI/31854/2019). Blood sampling from Egyptian cattle was done based on animal welfare guidelines of the Institutional Animal Care and Use Committee, Cairo University (CU-IACUC), which approved this protocol under number CUIIF720. In South Africa, sampling of blood and hair was performed with the approval of the Animal Ethics Committee of the Agricultural Research Council (APAEC [2020/17]), according to guidelines for the proper handling of animals during sample collection. We have complied with all relevant ethical regulations for animal use.

Sample collection and whole-genome sequencing

Whole-genome sequences from 519 healthy cattle, unrelated within two generations and representing 24 African and European indigenous populations across six countries (Finland, the Netherlands, Portugal, Egypt, Uganda, and South Africa), were analyzed as part of the LEAP-Agri OPTIBOV project (Supplementary Data 1; Supplementary Fig. 3) (https://www.optibov.org/). OPTIBOV blood samples were collected during routine veterinary health inspections, with written informed consent obtained from animal owners. Genomic DNA was extracted from EDTA whole-blood samples using the GENTRA Blood kit (Qiagen N.V.). The quantity of the isolated DNA samples was assessed using the Qubit fluorometer (Qiagen N.V.). Subsequently, the DNA samples were used to prepare PCR-free, double-indexed genomic libraries. These libraries were then subjected to paired-end sequencing (150 bp read length) using the Illumina NovaSeq 6000 platform (Illumina Inc., USA).

Additionally, 117 publicly available whole-genome sequences representing diverse cattle populations were obtained from the 1000 Bull Genomes Project Run 9[42] and the newly released Genomic Reference Resource for African Cattle[109] (Supplementary Data 1). These included representative African taurine, European taurine, African indicine, and Asian-American indicine populations sampled from Europe, Africa, Asia, and the Americas. Variant Call Format (VCF) files were generated following the 1000 Bull Genomes Project guidelines. The combined reference dataset encompassed four major ancestry groups: African taurine (AFT; $n = 34$), consisting of the breeds N'Dama ($n = 10$) and Muturu ($n = 24$); European taurine (EUT;

$n = 29$), consisting of the breeds Hereford ($n = 12$), Holstein Friesian ($n = 5$), Jersey ($n = 6$), and Angus ($n = 6$); African indicine (AFI; $n = 25$), consisting of the breeds Goffa ($n = 3$), Boran ($n = 7$), Kenana ($n = 8$), and Ogaden ($n = 7$); and Asian-American indicine (AAI; $n = 29$), consisting of the breeds Brahman ($n = 14$), Nelore ($n = 6$), Hariana ($n = 1$), Tharparkar ($n = 1$), Red Sindhi ($n = 2$), and Gir ($n = 5$). The map location of all OPTIBOV and public cattle samples is depicted in Fig. 1.

Short read pre-processing, variant calling, and filtering

All sequenced animals from the OPTIBOV dataset ($n = 519$) were pre-processed using fastp v0.23.4[110]. This included adapter trimming, correction of mismatched bases in paired-end read overlaps, and removal of low-quality reads (average quality score < 30 or read length < 36 bases). Clean reads were then mapped to the bovine reference genome (ARS-UCD1.2) using BWA-MEM2 v2.2.1[111] with default parameters, generating mapping quality reports. Following alignment, BAM files were sorted and indexed using SAMtools v1.14[112]. PCR duplicates were identified and marked using 'MarkDuplicates' from Picard v2.20.2[112]. Quality assessment/coverage estimation on post-QC BAMs (QualiMap v2.0)[113]. Variant calling was performed using FreeBayes v1.3.1[114] using default and custom settings: calls were restricted to the two most likely alleles per site (–use-best-n-alleles 2), with at least 20% of reads supporting the alternative allele (--min-alternate-fraction 0.2) and a minimum of two alternate reads (–min-alternate-count 2). We assumed diploidy (default), required base quality > 20 (--min-base-quality 20), and disabled haplotype calling to focus on single-site polymorphisms (–haplotype-length 0). Resulting variants with QUAL < 20 or depth < 4 were removed using Bcftools v1.20 software[112].

The VCF files were then merged with the reference variant data (AFT, EUT, AFI, and AAI). Further quality filtering was performed to remove non-biallelic SNPs, SNPs with missing genotype rates greater than 0.01, and SNPs with a minor allele frequency below 0.01. Missing genotypes were imputed and phased using BEAGLE v5.5[115,116] to maximize data completeness. Genotype imputation ensured that missing genotype information was inferred accurately based on haplotype patterns, thus improving downstream analytical resolution. Finally, about 15.1 million SNPs were retained. Variant statistics were summarized with VCFstats v0.4. 3[117] and visualized in a circos plot by TBtools[118].

Population differentiation and structure

Genetic structure and relationships among cattle populations were investigated using PLINK v1.9[119]. Pairwise identity-by-descent estimates were obtained from LD-pruned SNPs (–indep-pairwise 50 10 0.2)[119,120] and indicated negligible genome-wide relatedness (mean PI_HAT = 0.006 ± 0.002 s.e.). Within-breed relatedness was also low (mean PI_HAT = 0.076 ± 0.005 s.e.), ranging from 0.000 in AFI, Egypt, and Ntuuku to 0.342 in Mirandesa, with most breeds showing PI_HAT < 0.10.

Population structure was assessed by PCA and phylogenetic tree using genotypes from 519 OPTIBOV individuals and 117 publicly available cattle genomes. LD-pruned SNPs were performed with PLINK (--indep-pairwise 50 10 0.2), and the top four principal components were extracted (–pca 4). Eigenvectors were visualized in ggplot2 v3.5.1[121]. Pairwise identity-by-state (IBS) genetic distances (–distance-matrix) were computed in PLINK v1.9[119] on LD-pruned data and used to construct a neighbor-joining (NJ) tree in APE package[122], visualized in iTOL v7.0[123].

Genome-wide ancestry composition was inferred using ADMIXTURE v1.3.0[124]. LD-pruned genotypes were used to reduce redundancy. Ancestry proportions were estimated for K values ranging from 1 to 8 with cross-validation error. Admixture plots were visualized in R with ggplot2 v3.5.1[121]. Genome-wide genetic differentiation was estimated using VCFtools v0.1.16 with a sliding-window approach (--fst-window-size 50000 --fst-window-step 25000), generating pairwise weighted F_{st} values across autosomes.

Estimation of admixture and admixture time

We used the f_3 statistic to test for admixture in African and southern European cattle breeds. Analyses were based on our variant dataset (~15.1 million SNPs) and the latest version of the ARS-UCD1.2 recombination map derived from German Holstein cattle[125]. Admixture statistics were computed with the ADMIXTOOLS[126], and standard errors were estimated by a block-jackknife approach using 5-cM block sizes. Z-scores were calculated from the block-jackknife standard errors. LD-pruned SNPs were converted from PLINK v1.9[119] binary format to EIGENSTRAT format using the convertf utility. f_3 statistics were then calculated with the ADMIXTOOLS[126] program qp3Pop to test for admixture and asymmetric allele sharing among populations. The f_3 (X; Source$_1$, Source$_2$) statistic was used to test whether a target population (X) is derived from a mixture of two source populations. We used different ancestral sources: EUT/AAI, AFTEUT/AAI, and AFTAAI/EUT to capture different admixture scenarios. A significantly negative f_3 statistic ($Z < -3$) was considered evidence of historical admixture in the target population[127].

The timing of admixture events was estimated using DATES (Distribution of Ancestry Tracts of Evolutionary Signals)[128], which infers ancestry-specific LD decay to estimate the number of generations since admixture. Analyses were performed using default settings, with a minimum genetic distance (mindis) of 0.5 cM and a maximum distance (maxdis) of 1.0 cM. Additional parameters were set as follows: binsize = 0.005, runmode = 1, mincount = 5, zdipcorrmode = YES, qbin = 10, runfit = YES, afffit = YES, and lovalfit = 0.5. For each population, pairwise LD weighted by allele frequencies in the two reference sources was modeled as an exponential decay function of genetic distance. DATES fits this decay curve and infers admixture time from the estimated decay rate. The resulting admixture times were log-transformed for visualization, and the associated uncertainty was calculated as ±1 s.e. from the Z-score variance estimates.

Inference of local ancestry admixture

Following the genome-wide admixture analyses, three unadmixed ancestral reference populations (AFT, EUT, and AAI) were identified and used for local ancestry inference in each breed. AFI populations were not treated as a separate ancestry source to avoid redundancy arising from their admixed genomic background. Local ancestry inference was performed using LOTER[45], which implements a haplotype copying model. In this model, haplotypes from admixed individuals are represented as mosaics composed of parental haplotypes drawn from the reference populations. Simulated admixed haplotypes were modeled from AFT, EUT, and AAI ancestries across non-overlapping 50-kb windows[2]. Regions within the top 1% of the continuous ancestry distribution, corresponding to extended haplotypes (>100 kb), were retained as candidate windows after excluding isolated single-window outliers. These regions were considered putative genomic segments enriched for specific ancestral lineages. Mosaic ancestry patterns were visualized using ggplot2 v3.5.1[121].

Given the historical introgression of indicine ancestry within African taurine populations, and the slightly positive f_3 value (N'Dama; Muturu, AAI)[59], small-scale hybridization events could potentially lead to false negatives in LOTER[45], which may underestimate indicine ancestry in certain genomic regions. To ensure robustness, we performed local ancestry analyses using two African taurine reference configurations: (i) N'Dama and Muturu combined, and (ii) Muturu alone. The resulting ancestry proportions and interpretations were highly consistent between the two configurations (AFT: $r = 0.97–0.98$, mean = 0.98; EUT: $r = 0.99–1.00$, mean = 0.99; AAI: $r = 0.99–1.00$, mean = 1.00; Supplementary Table 7). All downstream analyses of key genomic regions were therefore conducted using both AFT reference sets to minimize potential bias. The sex chromosome was excluded from local ancestry inference because the 1000 Bull Genomes reference panel exhibited uneven callable coverage and low SNP density on sex chromosomes, which precluded reliable phasing, imputation, and ancestry assignment.

Detection and prioritization of ancestry-retention genes

Genome-wide admixture analyses revealed that all sampled African indigenous cattle breeds possess mosaic genomic compositions comprising AFT and AAI ancestry, with signatures of recent EUT introgression observed in

Egyptian and Sanga breeds. To identify such loci and haplotypes, local ancestry inference was performed using LOTER[45]. For each breed, the top 1% of continuous 50-kb windows showing the highest ancestry proportions were selected. Variants and genes overlapping these regions were annotated using SnpEff v5.2[129], and variants classified as high, moderate, or low impact within coding regions were retained for downstream analyses.

To minimize false-positive signals arising from ancestry-enriched windows, breed-specific genetic drift, and local recombination, we focused on genes showing elevated retention rates across multiple populations: AFT and AAI across all thirteen African populations, and EUT across seven recently introgressed populations (Egyptian cattle Upper, Middle, and Lower; Bonsmara; Tuli, Nguni, and Afrikaner; Fig. 2b). Genes located within recurrent ancestry-outlier segments shared across breeds and harboring amino acid–altering variants were prioritized as candidate loci with potential functional relevance. The top genes exhibiting the highest retention rates and protein-altering variants for each ancestry across African and southern European indigenous cattle populations are summarized in Table 1. For example, *DDIT3* was prioritized because it exhibited the highest AAI-retention rates (61.5%), with significantly elevated values across eight of the thirteen African populations ($^*p < 0.05$). This gene harbored high- and moderate-impact variants predicted to alter amino acid sequence and protein function. For candidate gene and regional analyses, allele frequencies across cattle groups were computed using VCFtools v0.1.17[130], and haplotype-sharing patterns and variant distributions were visualized in ggplot2 v3.5.1[121]. A complete list of candidate genes with variant impact ranks and retention rates across each breed is provided in Supplementary Data 3.

Detection of genome-wide adaptive introgression in Portuguese taurine cattle

Our admixture analyses suggest that indigenous Portuguese breeds, particularly Mertolenga, share closer genetic relationships and possible past gene flow with African taurine populations. We used three approaches to investigate AFT adaptive-introgression loci and haplotypes. First, we inferred the top 1% of continuous 50-kb local-ancestry windows showing the highest AFT proportions in Mertolenga using LOTER[45]. Variants and genes overlapping these regions were annotated with SnpEff v5.2[129], and variants classified as high, moderate, or low impact within coding regions were retained for downstream analyses.

Genome-wide selection signatures in Mertolenga were then evaluated with two complementary statistics: the fixation index (F_{st})[131,132] and nucleotide diversity (π)[133] using VCFtools v0.1.17[130,134]. We applied a sliding-window scheme (50-kb windows, 25-kb steps) to smooth local variability and enhance detection of consistent signals across the genome. For F_{st}, we compared Portuguese taurine breeds to European taurine reference populations; weighted F_{st} was used to account for per-marker heterozygosity and unequal SNP counts across windows, reducing bias from variable SNP density and sample size. Elevated F_{st} values indicate increased allele-frequency divergence, consistent with local adaptation or gene flow from African cattle[135]. For reduced nucleotide diversity, we compared Portuguese taurine breeds to European taurine reference populations; positive $-\log_2(\pi\_\text{Mertolenga}/\pi\_\text{EUT})$ values denote locally reduced diversity, as expected under selective sweeps or recent bottlenecks[136]. Therefore, we intersected AFT-retention genes harboring protein-altering variants with the top 1% windows for F_{st} and positive $-\log_2(\pi\_\text{Mertolenga}/\pi\_\text{EUT})$. All figures were generated in ggplot2 v3.5.1[121].

Annotation analysis and functional enrichment analyses

Variant annotation for all pipeline-called SNPs was performed using SnpEff v5.2[129], which categorized effects by genomic location and predicted impact on coding sequence. Variants were classified as high, moderate, low, or modifier impact. For downstream analyses (e.g., GO/KEGG enrichment) of local ancestry retention, we retained only coding variants (high-, moderate-, and low-impact), while modifier variants were excluded. Gene- and coordinate-level annotations for the retained candidate regions were verified using Ensembl VEP v113.3 with the ARS-UCD1.2 reference genome[137].

Functional enrichment analysis was conducted using the Database for Annotation, Visualization, and Integrated Discovery (DAVID v2021) for GO Biological Process and KEGG terms under default settings. Enrichment *P*-values were computed using the modified Fisher's exact test (EASE), which applies a conservative small-sample correction[138]. Results were interpreted at nominal significance ($^*p < 0.05$). Fold enrichment was reported as the ratio of observed to expected gene counts per term. All figures were generated in ggplot2 v3.5.1[121].

Protein structure prediction for candidate genes

Coding variants in *DDIT3* were identified with Ensembl VEP v113.3[137] on the ARS-UCD1.2 bovine reference genome. Corresponding transcript and protein sequences (wild type) were retrieved from the Ensembl database (https://www.ensembl.org/Bos_taurus/Transcript/). The reference 3D structure of bovine *DDIT3* was obtained from the AlphaFold Protein Structure Database. Variants were mapped onto the AlphaFold model (https://alphafold.ebi.ac.uk/); where relevant, amino-acid substitutions were introduced in silico and side-chain rotamers inspected in UCSF ChimeraX[139]. Structural overlays, residue-level annotations (including AlphaFold pLDDT confidence), and figure rendering were performed in UCSF ChimeraX[139].

Statistics and reproducibility

Analyses were performed on WGS data from unrelated individuals. The total number of genomes analyzed and per-breed sample sizes are provided in Supplementary Data 1. Pathway enrichment *P* values were computed using the modified Fisher's exact test (EASE); $p^* < 0.05$ was considered statistically significant. Where applicable, paired *t*-tests were used for population-level comparisons. Variant-level quality control and filtering thresholds were defined above and applied consistently across datasets (short-read pre-processing, variant calling, and variant filtering). No samples were excluded from the African or European cattle cohorts.

Ancestry-retention rates were calculated as the proportion of breeds in which a segment was retained (number of retained breeds divided by the total number of breeds considered). Specifically, retention was summarized for AFT and AAI across all 13 African populations, and for EUT across seven recently introgressed populations. Genes overlapping recurrent ancestry-retention segments were ranked by retention rate; those shared across multiple populations and carrying coding variants were prioritized as candidate loci for downstream interpretation.

Reproducibility was ensured through fully scripted and version-controlled computational workflows (Linux shell, R, and Python). Code for read processing, variant calling, local-ancestry inference, selection scans, and figure generation is available via GitHub and Zenodo[140].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Population-level variant calls (VCF files) for the 636 indigenous and reference cattle analyzed in this study have been deposited in the European Variation Archive under accession PRJEB102975 (ERP184337)[141]. Whole-genome sequence data for the 519 African and European indigenous cattle analyzed in this study are available from the European Nucleotide Archive under accessions PRJEB90914[142] and PRJEB90816[143]. Publicly available reference variant datasets were obtained from the 1000 Bull Genomes Project (PRJNA391427; ERZ14211345)[144] and the African Genomic Reference Resource (PRJEB74565; Muturu)[145]. Sample IDs for the ancestry reference panels (AFT, $n = 34$; EUT, $n = 29$; AFI, $n = 25$; AAI, $n = 29$) are provided in Supplementary Data 1. Supplementary Data 2 contains genome-wide haplotype-mosaic files for locus-level inspection, provided for two AFT reference configurations (N'Dama + Muturu and Muturu only). Supplementary Data 3 lists putative ancestry-retention genes together with retention rates. Supplementary Data 4 provides window-based nucleotide

diversity (π) and genetic differentiation (weighted F_{st}) for Mertolenga and EUT populations. All supplementary datasets are available via the project repository on GitHub: https://github.com/junxingao888/ancestral_retention_signals_cattle/. Source data underlying the figures are provided as follows: Supplementary Table 3 (Fig. 2a,b), Supplementary Table 4 (Fig. 2c,d), Supplementary Table 5 (Fig. 4c), Supplementary Table 6 (Fig. 4e), and Supplementary Data 4 (Fig. 5d,e).

Code availability

All custom code and workflows (Linux shell, R, and Python) for read processing, variant calling, local-ancestry inference, selection scans, and figure generation are available at GitHub (https://github.com/junxingao888/ancestral_retention_signals_cattle/tree/main/Code_availability) and Zenodo[140].

Received: 16 July 2025; Accepted: 3 March 2026;
Published online: 19 March 2026

References

1. Godde, C. M., Mason-D'Croz, D., Mayberry, D. E., Thornton, P. K. & Herrero, M. Impacts of climate change on the livestock food supply chain; a review of the evidence. *Glob. Food Secur.* **28**, 100488 (2021).
2. Kim, K. et al. The mosaic genome of indigenous African cattle as a unique genetic resource for African pastoralism. *Nat. Genet.* **52**, 1099–1110 (2020).
3. Felius, M. *Cattle Breeds of the World*. (BRILL, 2024).
4. Malek, Ž et al. Improving the representation of cattle grazing patterns in the European Union. *Environ. Res. Lett.* **19**, 114077 (2024).
5. Friedrich, J. et al. Mapping restricted introgression across the genomes of admixed indigenous African cattle breeds. *Genet. Sel. Evol.* **55**, 91 (2023).
6. Ward, J. A. et al. Genome-wide local ancestry and evidence for mitonuclear coadaptation in African hybrid cattle populations. *iScience* **25**, 104672 (2022).
7. Chen, N. et al. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat. Commun.* **9**, 2337 (2018).
8. Achilli, A. et al. Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Curr. Biol.* **18**, R157–R158 (2008).
9. Chen, N. et al. Global genetic diversity, introgression, and evolutionary adaptation of indicine cattle revealed by whole genome sequencing. *Nat. Commun.* **14**, 7803 (2023).
10. Bollongino, R. et al. Modern taurine cattle descended from a small number of Near Eastern founders. *Mol. Biol. Evol.* **29**, 2101–2104 (2012).
11. Chen, S. et al. Zebu cattle are an exclusive legacy of the South Asia neolithic. *Mol. Biol. Evol.* **27**, 1–6 (2010).
12. Loftus, R. T., MacHugh, D. E., Bradley, D. G., Sharp, P. M. & Cunningham, P. Evidence for two independent domestications of cattle. *Proc. Natl. Acad. Sci.* **91**, 2757–2761 (1994).
13. Scheu, A. et al. The genetic prehistory of domesticated cattle from their origin to the spread across Europe. *BMC Genet.* **16**, 54 (2015).
14. Brito, L. F. et al. Review: Genetic selection of high-yielding dairy cattle toward sustainable farming systems in a rapidly changing world. *Animal* **15**, 100292 (2021).
15. Flori, L. et al. The genome response to artificial selection: a case study in dairy cattle. *PLoS One* **4**, e6595 (2009).
16. Park, S. D. et al. Genome sequencing of the extinct Eurasian wild aurochs, Bos primigenius, illuminates the phylogeography and evolution of cattle. *Genome Biol.* **16**, 234 (2015).
17. Verdugo, M. P. et al. Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. *Science* **365**, 173–176 (2019).
18. Wragg, D. et al. A locus conferring tolerance to Theileria infection in African cattle. *PLoS Genet.* **18**, e1010099 (2022).
19. Kambal, S. et al. Candidate signatures of positive selection for environmental adaptation in indigenous African cattle: A review. *Anim. Genet.* **54**, 689–708 (2023).
20. Naessens, J. Bovine trypanotolerance: A natural ability to prevent severe anaemia and haemophagocytic syndrome?. *Int J. Parasitol.* **36**, 521–528 (2006).
21. Nicolotti, M. & Guérin, C. Le zébu (Bos indicus) dans l'Egypte ancienne. *Archaeozoologia* **5**, 87–108 (1992).
22. Hanotte, O. et al. African pastoralism: genetic imprints of origins and migrations. *Science* **296**, 336–339 (2002).
23. Epstein, H. *The origin of the domestic animals of Africa*. *Vol. II*. 2 (1971).
24. Gifford-Gonzalez, D. & Hanotte, O. Domesticating animals in Africa: implications of genetic and archaeological findings. *J. World Prehist.* **24**, 1–23 (2011).
25. Fuller, D. Q. & Boivin, N. Crops, cattle and commensals across the Indian Ocean. Current and potential archaeobiological evidence. *Études Océ Indien* **13**, 46 (2009).
26. Felius, M. *Cattle breeds: An encyclopedia*. (1995).
27. Mwai, O., Hanotte, O., Kwon, Y.-J. & Cho, S. African indigenous cattle: unique genetic resources in a rapidly changing world. *Asian-Australas. J. Anim. Sci.* **28**, 911 (2015).
28. Beja-Pereira, A. et al. The origin of European cattle: evidence from modern and ancient DNA. *Proc. Natl. Acad. Sci.* **103**, 8113–8118 (2006).
29. Scholtz, M. & Theunissen, A. The use of indigenous cattle in terminal cross-breeding to improve beef cattle production in Sub-Saharan Africa. *Anim. Genet. Resour.* **46**, 33–39 (2010).
30. da Fonseca, R. R. et al. Consequences of breed formation on patterns of genomic diversity and differentiation: the case of highly diverse peripheral Iberian cattle. *BMC Genomics* **20**, 1–13 (2019).
31. Cymbron, T., Loftus, R. T., Malheiro, M. I. & Bradley, D. G. Mitochondrial sequence variation suggests an African influence in Portuguese cattle. *Proc. R. Soc. Lond. Ser. B: Biol. Sci.* **266**, 597–603 (1999).
32. Decker, J. E. et al. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet.* **10**, e1004254 (2014).
33. Upadhyay, M. et al. Deciphering the patterns of genetic admixture and diversity in southern European cattle using genome-wide SNPs. *Evolut. Appl.* **12**, 951–963 (2019).
34. Kim, J. et al. The genome landscape of indigenous African cattle. *Genome Biol.* **18**, 34 (2017).
35. Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C. & Clark, A. G. Recent and ongoing selection in the human genome. *Nat. Rev. Genet.* **8**, 857–868 (2007).
36. Huerta-Sánchez, E. et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194–197 (2014).
37. Qiu, Q. et al. The yak genome and adaptation to life at high altitude. *Nat. Genet.* **44**, 946–949 (2012).
38. Scholtz, M., Bester, J., Mamabolo, J. & Ramsay, K. Results of the national cattle survey undertaken in South Africa, with emphasis on beef. *Appl. Anim. Husb. Rural. Develop.* **1**, 1–9 (2008).
39. Kim, E.-S. & Rothschild, M. F. Genomic adaptation of admixed dairy cattle in East Africa. *Front. Genet.* **5**, 443 (2014).
40. Upadhyay, M. *Genomic variation across European cattle: contribution of gene flow*. (Wageningen University and Research, 2019).
41. Daetwyler, H. D. et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* **46**, 858–865 (2014).
42. Hayes, B. J. & Daetwyler, H. D. 1000 bull genomes project to map simple and complex genetic traits in cattle: applications and outcomes. *Annu. Rev. Anim. Biosci.* **7**, 89–102 (2019).
43. Moorjani, P. et al. The history of African gene flow into Southern Europeans, Levantines, and Jews. *PLoS Genet.* **7**, e1001373 (2011).

44. Liang, M., Shishkin, M., Mikhailova, A., Shchur, V. & Nielsen, R. Estimating the timing of multiple admixture events using 3-locus linkage disequilibrium. *PLoS Genet.* **18**, e1010281 (2022).
45. Dias-Alves, T., Mairal, J. & Blum, M. G. Loter: a software package to infer local ancestry for a wide range of species. *Mol. Biol. Evol.* **35**, 2318–2326 (2018).
46. Racimo, F., Sankararaman, S., Nielsen, R. & Huerta-Sánchez, E. Evidence for archaic adaptive introgression in humans. *Nat. Rev. Genet.* **16**, 359–371 (2015).
47. Garattini, E., Mendel, R., Romão, M. J., Wright, R. & Terao, M. Mammalian molybdo-flavoenzymes, an expanding family of proteins: structure, genetics, regulation, function and pathophysiology. *Biochem. J.* **372**, 15–32 (2003).
48. Huntley, S. et al. A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res.* **16**, 669–677 (2006).
49. Kobayashi, K. et al. IRAK-M is a negative regulator of Toll-like receptor signaling. *Cell* **110**, 191–202 (2002).
50. Harding, H. P. et al. Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol. Cell* **6**, 1099–1108 (2000).
51. Schroer, T. A. Dynactin. *Annu. Rev. Cell Dev. Biol.* **20**, 759–779 (2004).
52. Utsunomiya, Y. et al. Genomic clues of the evolutionary history of Bos indicus cattle. *Anim. Genet.* **50**, 557–568 (2019).
53. Li, M. et al. DDIT3 directs a dual mechanism to balance glycolysis and oxidative phosphorylation during glutamine deprivation. *Adv. Sci.* **8**, 2003732 (2021).
54. Pereira, A. M., Baccari, F., Titto, E. A. & Almeida, J. A. Effect of thermal stress on physiological parameters, feed intake and plasma thyroid hormones concentration in Alentejana, Mertolenga, Frisian and Limousine cattle breeds. *Int. J. Biometeorol.* **52**, 199–208 (2008).
55. Ginja, C. et al. Iron age genomic data from Althiburos-Tunisia renew the debate on the origins of African taurine cattle. *Iscience* **26** (2023).
56. Mengistie, D. Origin of cattle breeds in East Africa and introduction to general breeding science: A–review. *World N. Nat. Sci.* **49**, 88–110 (2023).
57. Murray, M., Trail, J., Davis, C. & Black, S. J. Genetic resistance to African trypanosomiasis. *J. Infect. Dis.* **149**, 311–319 (1984).
58. Zegeye, T., Belay, G., Vallejo-Trujillo, A., Han, J. & Hanotte, O. Genome-wide diversity and admixture of five indigenous cattle populations from the Tigray region of northern Ethiopia. *Front. Genet.* **14**, 1050365 (2023).
59. Kim, K. et al. Inference of Admixture Origins in Indigenous African Cattle. *Mol. Biol. Evol.* **40**, https://doi.org/10.1093/molbev/msad257 (2023).
60. Hanotte, O. et al. Geographic distribution and frequency of a taurine Bos taurus and an indicine Bos indicus Y specific allele amongst sub-saharan African cattle breeds. *Mol. Ecol.* **9**, 387–396 (2000).
61. Dai, S. et al. Global pangenome analysis highlights the critical role of structural variants in cattle improvement and identifies a unique event as a novel enhancer in IGFBP7+ cells. *Mol. Biol. Evol.* **42**, msaf205 (2025).
62. Coffey, E., Horan, B., Evans, R. & Berry, D. Milk production and fertility performance of Holstein, Friesian, and Jersey purebred cows and their respective crosses in seasonal-calving commercial farms. *J. Dairy Sci.* **99**, 5681–5689 (2016).
63. Xia, X. et al. Global dispersal and adaptive evolution of domestic cattle: a genomic perspective. *Stress Biol.* **3**, 8 (2023).
64. Taye, M. et al. Exploring the genomes of East African Indicine cattle breeds reveals signature of selection for tropical environmental adaptation traits. *Cogent Food Agric.* **4**, 1552552 (2018).
65. Jonsson, N., Piper, E. & Constantinoiu, C. Host resistance in cattle to infestation with the cattle tick R hipicephalus microplus. *Parasite Immunol.* **36**, 553–559 (2014).
66. Magee, D. A., MacHugh, D. E. & Edwards, C. J. Interrogation of modern and ancient genomes reveals the complex domestic history of cattle. *Anim. Front.* **4**, 7–22 (2014).
67. Martín-Burriel, I. et al. Genetic diversity, structure, and breed relationships in Iberian cattle. *J. Anim. Sci.* **89**, 893–906 (2011).
68. Groh, J. S. & Coop, G. The temporal and genomic scale of selection following hybridization. *Proc. Natl. Acad. Sci.* **121**, e2309168121 (2024).
69. Sankararaman, S. et al. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354–357 (2014).
70. Smetko, A. et al. Trypanosomosis: potential driver of selection in African cattle. *Front. Genet.* **6**, 137 (2015).
71. Hunter, C. A. & Kastelein, R. Interleukin-27: balancing protective and pathological immunity. *Immunity* **37**, 960–969 (2012).
72. Lamp, O. et al. Metabolic Heat Stress Adaption in Transition Cows: Differences in Macronutrient Oxidation between Late-Gestating and Early-Lactating German Holstein Dairy Cows. *PLoS One* **10**, e0125264 (2015).
73. Ma, Z., Tanis, J. E., Taruno, A. & Foskett, J. K. Calcium homeostasis modulator (CALHM) ion channels. *Pflug. Arch.* **468**, 395–403 (2016).
74. Liu, D. et al. Genome-wide selection signatures detection in Shanghai Holstein cattle population identified genes related to adaption, health and reproduction traits. *BMC Genomics* **22**, 747 (2021).
75. Kusano, T., Nishino, T., Okamoto, K., Hille, R. & Nishino, T. The mechanism and significance of the conversion of xanthine dehydrogenase to xanthine oxidase in mammalian secretory gland cells. *Redox Biol.* **59**, 102573 (2023).
76. Chen, C. et al. Dermatan sulfate: structure, biosynthesis, and biological roles. *Biomolecules* **15**, https://doi.org/10.3390/biom15081158 (2025).
77. Larkin, D. M. et al. Whole-genome resequencing of two elite sires for the detection of haplotypes under selection in dairy cattle. *Proc. Natl. Acad. Sci.* **109**, 7693–7698 (2012).
78. Nayeri, S. et al. Genome-wide association analysis for β-hydroxybutyrate concentration in Milk in Holstein dairy cattle. *BMC Genet.* **20**, 58 (2019).
79. Nozawa, K. et al. Knockout of serine-rich single-pass membrane protein 1 (Ssmem1) causes globozoospermia and sterility in male mice†. *Biol. Reprod.* **103**, 244–253 (2020).
80. Mulim, H. A. et al. Detection and evaluation of parameters influencing the identification of heterozygous-enriched regions in Holstein cattle based on SNP chip or whole-genome sequence data. *BMC Genomics* **25**, 726 (2024).
81. He, W. et al. Function identification of bovine ACSF3 gene and its Association with lipid metabolism traits in beef cattle. *Front. Vet. Sci.* **8**, 766765 (2022).
82. Dirandeh, E., Ansari-Pirsaraei, Z. & Thatcher, W. Melatonin as a smart protector of pregnancy in dairy cows. *Antioxidants.* **11**, https://doi.org/10.3390/antiox11020292 (2022).
83. de Camargo, G. M. et al. Polymorphisms in TOX and NCOA2 genes and their associations with reproductive traits in cattle. *Reprod. Fertil. Dev.* **27**, 523–528 (2015).
84. Wang, T. et al. The biological properties of the FAS and TACR3 genes and the association of single-nucleotide polymorphisms with milk quality traits in Gannan Yak. *Foods* **14**, https://doi.org/10.3390/foods14091575 (2025).
85. Jiang, L.-Y. et al. Ring finger protein 145 (RNF145) is a ubiquitin ligase for sterol-induced degradation of HMG-CoA reductase. *J. Biol. Chem.* **293**, 4047–4055 (2018).
86. Rothi, M. H. et al. The 18S rRNA methyltransferase DIMT-1 regulates lifespan in the germline later in life. *Nat. Commun.* **16**, 6944 (2025).
87. Benfica, L. F. et al. Genome-wide association study between copy number variation and feeding behavior, feed efficiency, and growth traits in Nellore cattle. *BMC Genomics* **25**, 54 (2024).

88. Singh, A. et al. Genomewide expression analysis of the heat stress response in dermal fibroblasts of Tharparkar (zebu) and Karan-Fries (zebu× taurine) cattle. *Cell Stress Chaperones* **25**, 327–344 (2020).
89. Tijjani, A. et al. Genomic signatures for drylands adaptation at gene-rich regions in African zebu cattle. *Genomics* **114**, 110423 (2022).
90. Gujar, G. et al. Characterization of thermo-physiological, hematological, and molecular changes in response to seasonal variations in two tropically adapted native cattle breeds of Bos indicus lineage in hot arid ambience of Thar Desert. *Int. J. Biometeorol.* **66**, 1515–1529 (2022).
91. Mengistie Yirsaw, D. *Genome-Wide Signature of Positive Selection, Breed-Specific SNPs and Linkage Disequilibrium in Ethiopian Indigenous and European Beef Cattle Breeds*, Addis Ababa University, (2021).
92. Liu, Y. et al. Hypoxia-induced GPCPD1 depalmitoylation triggers mitophagy via regulating PRKN-mediated ubiquitination of VDAC1. *Autophagy* **19**, 2443–2463 (2023).
93. Freihat, L. A. et al. IRAK3 modulates downstream innate immune signalling through its guanylate cyclase activity. *Sci. Rep.* **9**, 15468 (2019).
94. Flori, L. et al. A genomic map of climate adaptation in Mediterranean cattle breeds. *Mol. Ecol.* **28**, 1009–1029 (2019).
95. Shinoda, K. et al. The dystonia gene THAP1 controls DNA double-strand break repair choice. *Mol. Cell* **81**, 2611–2624.e2610 (2021).
96. Bressler, K. R. et al. Depletion of eukaryotic initiation factor 5B (eIF5B) reprograms the cellular transcriptome and leads to activation of endoplasmic reticulum (ER) stress and c-Jun N-terminal kinase (JNK). *Cell Stress Chaperones* **26**, 253–264 (2021).
97. Ruthig, V. A. et al. The RNA-binding protein DND1 acts sequentially as a negative regulator of pluripotency and a positive regulator of epigenetic modifiers required for germ cell reprogramming. *Development* **146**, dev175950 (2019).
98. Purfield, D. C., Bradley, D. G., Evans, R. D., Kearney, F. J. & Berry, D. P. Genome-wide association study for calving performance using high-density genotypes in dairy and beef cattle. *Genet. Sel. Evol.* **47**, 47 (2015).
99. Guzmán, L. F. et al. Expression of heat shock protein genes in Simmental cattle exposed to heat stress. *Anim. Biosci.* **36**, 704 (2023).
100. Kim, J. et al. Expansion of the HSP70 gene family in Tegillarca granosa and expression profiles in response to zinc toxicity. *Cell Stress Chaperones* **29**, 97–112 (2024).
101. Chen, Q. et al. A brown fat-enriched adipokine Adissp controls adipose thermogenesis and glucose homeostasis. *Nat. Commun.* **13**, 7633 (2022).
102. Bo, D. et al. Whole-genome resequencing reveals genetic diversity and growth trait-related genes in Pinan cattle. *Animals* **14**, 2163 (2024).
103. Asai, Y. et al. Transgenic Tmc2 expression preserves inner ear hair cells and vestibular function in mice lacking Tmc1. *Sci. Rep.* **8**, 12124 (2018).
104. Steiner, P. & Zierler, S. Inter-Organellar Ca(2+) Homeostasis in plant and animal systems. *Cells* **14**, https://doi.org/10.3390/cells14151204 (2025).
105. Almhanna, H. et al. Comparison of Siglec-1 protein networks and expression patterns in sperm and male reproductive tracts of mice, rats, and humans. *Vet. World* **17**, 645–657 (2024).
106. Spetter, M. J. et al. Genetic diversity, Admixture, and selection signatures in a Rarámuri Criollo cattle population introduced to the Southwestern United States. *Int. J. Mol. Sci.* **26**, 4649 (2025).
107. Almhanna, H. et al. Comparison of Siglec-1 protein networks and expression patterns in sperm and male reproductive tracts of mice, rats, and humans. *Vet. World* **17**, 645 (2024).
108. Jones, J. M. & First, N. L. Expression of the cell cycle control protein cdc25 in cleavage stage bovine embryos. *Zygote* **3**, 133–139 (1995).
109. Tijjani, A. et al. Genomic reference resource for African cattle: genome sequences and high-density array variants. *Sci. Data* **11**, 801 (2024).
110. Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta* **2**, e107 (2023).
111. Vasimuddin, M., Misra, S., Li, H. & Aluru, S. in *2019 IEEE International Parallel And Distributed Processing Symposium (IPDPS)*. 314–324 (IEEE).
112. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
113. Okonechnikov, K., Conesa, A. & García-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292–294 (2016).
114. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907* (2012).
115. Browning, B. L., Zhou, Y. & Browning, S. R. A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* **103**, 338–348 (2018).
116. Browning, B. L., Tian, X., Zhou, Y. & Browning, S. R. Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* **108**, 1880–1890 (2021).
117. Lindenbaum, P. JVarkit: Java-based utilities for Bioinformatics. *figshare* **10**, m9 (2015).
118. Chen, C. et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **13**, 1194–1202 (2020).
119. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
120. Ayalew, W. et al. Whole genome scan uncovers candidate genes related to milk production traits in Barka Cattle. *Int. J. Mol. Sci.* **25**, 6142 (2024).
121. Gómez-Rubio, V. ggplot2-elegant graphics for data analysis. *J. stat. softw.* **77**, 1–3 (2017).
122. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
123. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucl. Acids Res.*, gkae268 (2024).
124. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
125. Qanbari, S. & Wittenburg, D. Male recombination map of the autosomal genome in German Holstein. *Genet. Sel. Evol.* **52**, 73 (2020).
126. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
127. Peter, B. M. Admixture, population structure, and F-statistics. *Genetics* **202**, 1485–1501 (2016).
128. Chintalapati, M., Patterson, N. & Moorjani, P. The spatiotemporal patterns of major human admixture events during the European Holocene. *Elife* **11**, https://doi.org/10.7554/eLife.77625 (2022).
129. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *fly* **6**, 80–92 (2012).
130. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
131. Holsinger, K. E. & Weir, B. S. Genetics in geographically structured populations: defining, estimating and interpreting F ST. *Nat. Rev. Genet.* **10**, 639–650 (2009).
132. Durrett, R. & Durrett, R. *Probability models for DNA sequence evolution*. 2 (Springer, 2008).

133. Nei, M. & Li, W.-H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci.* **76**, 5269–5273 (1979).
134. Pan, B. et al. TMC1 and TMC2 are components of the mechanotransduction channel in hair cells of the mammalian inner ear. *Neuron* **79**, 504–515 (2013).
135. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *evolution*, 1358–1370 (1984).
136. Borowsky, R. L. Estimating nucleotide diversity from random amplified polymorphic DNA and amplified fragment length polymorphism data. *Mol. Phylogenet. Evol.* **18**, 143–148 (2001).
137. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 1–14 (2016).
138. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
139. Meng, E. C. et al. UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).
140. Gao, J. Ancestral retention signals cattle [Workflow]. *Zenodo* https://doi.org/10.5281/zenodo.18378574 (2026).
141. *EVA European Variation Archive*. https://identifiers.org/ena.embl:PRJEB102975 (2025).
142. *ENA European Nucleotide Archive*. https://identifiers.org/ena.embl:PRJEB90914 (2025).
143. *ENA European Nucleotide Archive*. https://identifiers.org/ena.embl:PRJEB90816 (2026).
144. *ENA European Nucleotide Archive*. https://identifiers.org/ena.embl:PRJNA391427 (2017).
145. *ENA European Nucleotide Archive*. https://identifiers.org/ena.embl:PRJEB74565 (2024).
146. Hamada, M. et al. Prognostic association of starvation-induced genes in head and neck cancer. (2021).
147. Kumar, A. et al. 2-Deoxyglucose drives plasticity via an adaptive ER stress-ATF4 pathway and elicits stroke recovery and Alzheimer's resilience. *Neuron* **111**, 2831–2846. e2810 (2023).
148. Oyadomari, S. & Mori, M. Roles of CHOP/GADD153 in endoplasmic reticulum stress. *Cell Death Differ.* **11**, 381–389 (2004).
149. Baral, K. & Rotwein, P. ZMAT2 in Humans and Other Primates: A Highly Conserved and Understudied Gene. *Evol. Bioinform. Online* **16**, 1176934320941500 (2020).
150. Warmack, R. A. et al. Human Protein-l-isoaspartate O-Methyltransferase Domain-Containing Protein 1 (PCMTD1) Associates with Cullin-RING Ligase Proteins. *Biochemistry* **61**, 879–894 (2022).
151. Sdiri, C., Ben Souf, I., Ben Salem, I., M'Hamdi, N. & Ben Hamouda, M. Assessment of Genetic and Health Management of Tunisian Holstein Dairy Herds with a Focus on Longevity. *Genes.* **14**, https://doi.org/10.3390/genes14030670 (2023).

Acknowledgements

This study was supported by the Long-term EU-Africa Research and Innovation Partnership on Food and Nutrition Security and Sustainable Agriculture (LEAP-Agri) as part of the OPTIBOV project (LEAP-Agri-326), and by the European Union's Horizon 2020 Research and Innovation Program (grant agreement No. 727715). Additional funding was provided by Fundação Nacional para a Ciência e a Tecnologia (FCT), Portugal (Leap Agri-326/LEAPAgri/0003/2017 and 2020.02754.CEECIND, C.G.); the Research Council of Finland (319987); the Netherlands Organization for Scientific Research (NWO-WOTRO, 2018/WOTRO/00488849); the Science, Technology & Innovation Funding Authority of Egypt (STDF, LEAP-Agri 326); the Ministry of Science, Technology and Innovations of Uganda (MoSTI/LEAP-11); the National Research Foundation of South Africa (NRF, 115577); and the China Scholarship Council (CSC, 202208610017). The funding bodies had no role in study design, data collection, analysis, interpretation, or manuscript writing.

We thank all collaborators for their assistance with sample collection, laboratory work, and technical support, including Bert Dibbits, Kimberley Laport, Rania Agamy, Mohamed Hamada Elsawy, Filipe Ribeiro, Ricardo Loureiro, Daniel Gaspar, Ludmilla Blaschikoff, Ana Elisabete Pires, Carolina Bruno-de-Sousa, Heli Lindberg, Tiina Reilas, Dr. Avhashoni Zwane, Khanyisani Nxumalo, Maano Malima, and all breeders and breed associations involved. We also acknowledge the 1000 Bull Genomes consortium and the African Genomic Reference Resource for providing sequence data.

Author contributions

J.G. and R.C. conceived the study. J.G. drafted the manuscript and interpreted the results. C.G. defined reference data sets. J.G. and Y.L. participated in data analysis. C.G., J.K., N.G., D.K., M.M., R.O., and R.C. collected the samples. H.B., M.G., and R.C. supervised the study. All the authors read and approved the 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/s42003-026-09856-9.

Correspondence and requests for materials should be addressed to Junxin Gao or Richard P. M. A. Crooijmans.

Peer review information *Communications Biology* thanks Alana Alexander and the other anonymous reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: George Inglis. A peer review file is available.