nature genetics

ARTICLES

https://doi.org/10.1038/s41588-020-0694-2

The mosaic genome of indigenous African cattle as a unique genetic resource for African pastoralism

Kwondo Kim[1,2], Taehyung Kwon[1], Tadelle Dessie[3], DongAhn Yoo[4], Okeyo Ally Mwai[5], Jisung Jang[4], Samsun Sung[2], SaetByeol Lee[2], Bashir Salim[6], Jaehoon Jung[1], Heesu Jeong[4], Getinet Mekuriaw Tarekegn[7,8], Abdulfatai Tijjani[3,9], Dajeong Lim[10], Seoae Cho[2], Sung Jong Oh[11], Hak-Kyo Lee[12], Jaemin Kim[13], Choongwon Jeong[14], Stephen Kemp[5,9], Olivier Hanotte[3,9,15] ✉ and Heebal Kim[1,2,4] ✉

Cattle pastoralism plays a central role in human livelihood in Africa. However, the genetic history of its success remains unknown. Here, through whole-genome sequence analysis of 172 indigenous African cattle from 16 breeds representative of the main cattle groups, we identify a major taurine × indicine cattle admixture event dated to circa 750–1,050 yr ago, which has shaped the genome of today's cattle in the Horn of Africa. We identify 16 loci linked to African environmental adaptations across crossbred animals showing an excess of taurine or indicine ancestry. These include immune-, heat-tolerance- and reproduction-related genes. Moreover, we identify one highly divergent locus in African taurine cattle, which is putatively linked to trypanotolerance and present in crossbred cattle living in trypanosomosis-infested areas. Our findings indicate that a combination of past taurine and recent indicine admixture-derived genetic resources is at the root of the present success of African pastoralism.

Cattle play an important role across African economies and societies as a primary source of wealth[1,2]. They provide nutrition, manure and draught power, and are often used to pay as bride wealth[1,2]. Today, at least 150 indigenous cattle breeds have been recognized across the different agro-ecologies of the African continent[3], each with unique phenotypic and adaptive characteristics[4,5].

Previous studies[5,6] have indicated that the dispersion and diversity of African cattle followed the history and development of African pastoralism. It is understood that the humpless *Bos taurus* and the humped *Bos indicus* originated from domestications of distinct auroch *Bos primigenius* subspecies with an ancestral divergence time of ~200,000 to less than 1 million yr ago[7–10]. The oldest uncontroversial evidence of domestic cattle in Africa dates back to circa 5750–4550 BC in Egypt's Western Desert at Nabta–Keseiba and circa 7000 BC in Kerma, Sudan[11]. These *B. taurus* cattle were introduced through North Africa and reached the Western and Eastern continent. They remained largely confined to the Saharan–Sahelian belt[12,13], until circa 4,000–3,000 yr ago, when they reached the Tilemsi Valley tributary of the Niger River in West Africa[14], the Lake Turkana basin of East Africa[15,16] and the Horn of Africa[17]. The main arrival of *B. indicus* started around 700 AD along the Red Sea and the Indian Ocean coastal areas, at the outset of the Swahili civilization[18,19] (Fig. 1a), which subsequently led to crossbreeding between *B. indicus* and already established African taurine.

However, the timing of the taurine × indicine admixture event(s) and their impacts on the development of African pastoralism remain unknown. Archeological evidence indicates that the development of sub-Saharan cattle pastoralism was a complex process that may not have proceeded as smoothly as its modern prevalence suggests[20,21]. In particular, environmental climatic and infectious disease challenges (for example, bovine malignant catarrhal fever, East Coast fever, foot-and-mouth disease, Rift Valley fever and trypanosomosis) likely have led to patchy and delayed establishment of herding across East Africa[16,20,22].

Today, the majority of African cattle are *B. taurus* × *B. indicus* humped populations of diverse phenotypes. They are classified as African Sanga (crossbred between Taurine and Zebu cattle), African Zenga (crossbred between Sanga and Zebu) and African Zebu[3,23]. The African Sanga, an Abyssinian word meaning bull, likely originated in North-East Africa with subsequent dispersion in the Central Lake Region and Southern Africa[14]. A few taurine populations found within the tsetse-belt in West Africa are the only pure African taurine cattle left on the continent[6,24].

African humped cattle carry only taurine mitochondrial DNA haplotypes[25–27]. The Y-chromosome microsatellite indicates the presence of both indicine and taurine Y chromosomes on the continent[5,28]. Furthermore, autosomal genome-wide analyses show that African humped cattle contain taurine background with

[1]Department of Agricultural Biotechnology and Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea. [2]C&K Genomics, Seoul, Republic of Korea. [3]International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia. [4]Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Republic of Korea. [5]International Livestock Research Institute (ILRI), Nairobi, Kenya. [6]Department of Parasitology, Faculty of Veterinary Medicine, University of Khartoum, Khartoum North, Sudan. [7]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden. [8]Department of Animal Production and Technology, Bahir Dar University, Bahir Dar, Ethiopia. [9]The Centre for Tropical Livestock Genetics and Health (CTLGH), The Roslin Institute, The University of Edinburgh, Midlothian, UK. [10]Animal Genomics & Bioinformatics Division, National Institute of Animal Science, RDA, Wanju, Republic of Korea. [11]International Agricultural Development and Cooperation Center, Jeonbuk National University, Jeonju, Republic of Korea. [12]Department of Animal Biotechnology, College of Agriculture and Life Sciences, Jeonbuk National University, Jeonju, Republic of Korea. [13]Department of Animal Science, College of Agriculture and Life Sciences, Gyeongsang National University, Jinju, Republic of Korea. [14]School of Biological Sciences, Seoul National University, Seoul, Republic of Korea. [15]School of Life Sciences, University of Nottingham, Nottingham, UK. ✉e-mail: o.hanotte@cgiar.org; heebal@snu.ac.kr

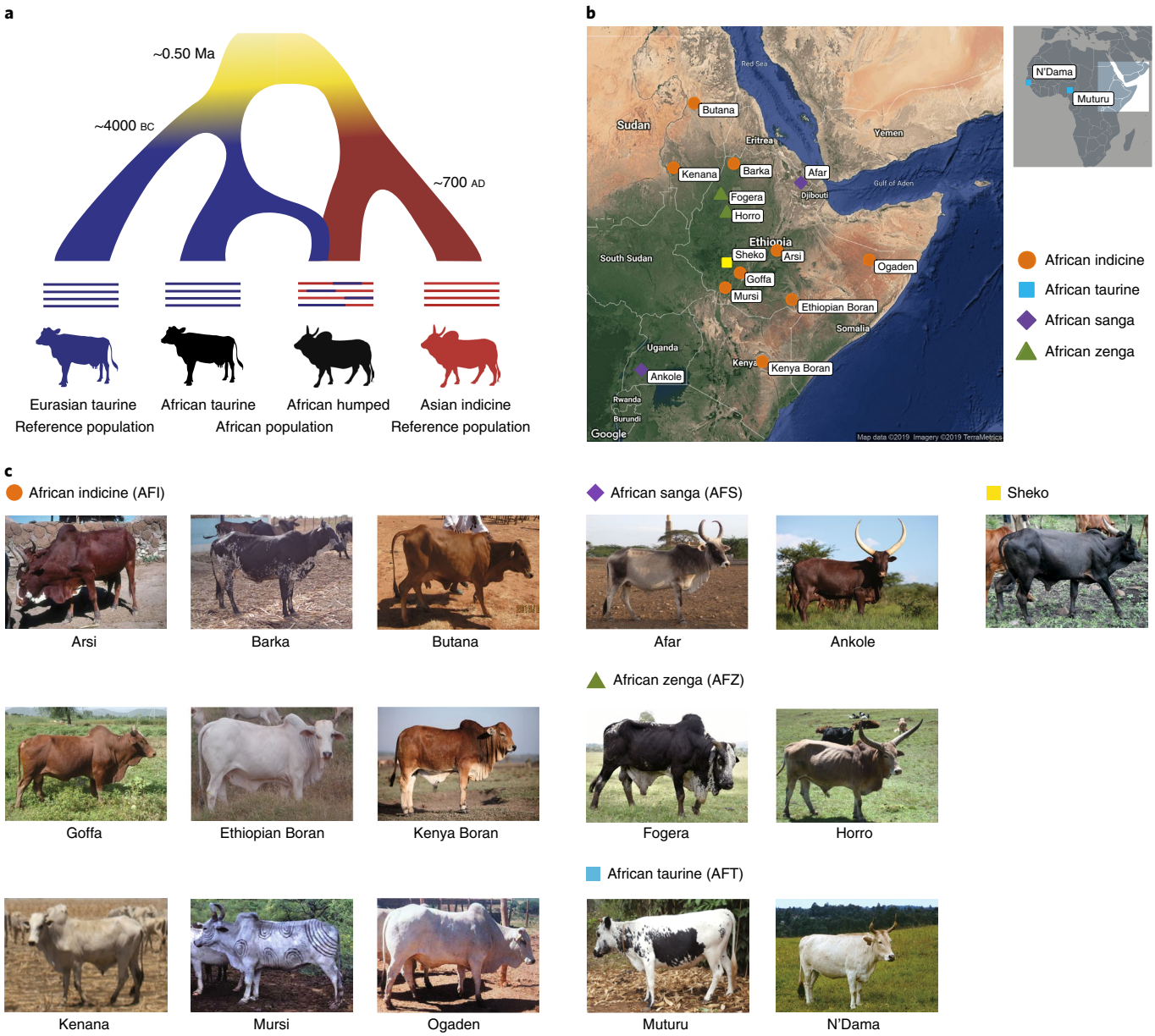


Fig. 1 | Historical and geographical origin of African cattle breeds in this study. a, Schematic diagram showing the relationships among the main cattle lineages. The divergence times are approximate estimates based on previous studies[3,10,19]. **b**, Geographical origin of the indigenous East African cattle breeds. The map in the background has been generated by R package 'ggmap'[101]. The different colors reflect the classification of the populations in different phenotypic groups, with the Sheko indicated in yellow. **c**, Photographs of each breed. Ma, million yr ago. Credits: Arsi, Mursi, Ogaden, Afar, Barka and Ethiopian Boran: ILRI Addis Ababa; Ankole, Sheko, Fogera and Horro: ILRI Nairobi Kenya; Kenya Boran and Gambian N'Dama: Steve Kemp; Muturu: Abdulfatai Tijjani; Kenana and Butana: Bashir Salim; Goffa: Chencha Chebo.

different levels of genetic contributions across populations, but with little variation within a population[29–31]. These analyses suggest that selection played a role in shaping the *B. taurus*×*B. indicus* admixture proportion in African cattle, with admixture increasing diversity and providing new genetic resources for human and natural selection[32]. This may have facilitated dispersion and colonization of new habitats[33]. Several recent studies have addressed the effects of admixture and introgression among the *Bos* species. They have identified loci derived from donor species, which have contributed to the adaptation of recipient species[34–36]. However, admixture and introgression also have a cost, as they may reduce the reproductive fitness due to genome incompatibility[37].

Here, we generated whole-genome sequences of 114 cattle that belong to 12 indigenous African cattle populations and two African buffalo. We combined these with the previously sequenced genomes of 58 cattle from four additional African populations[31,38]. These populations represent the main African cattle groups (Supplementary Note). Using this unique set, we date a main taurine×indicine admixture event and assess the present genome ancestry of African cattle, supporting that a combination of these two ancestries is at the root of the success of African pastoralism.

Results

Sequencing, mapping and identification of SNPs. Individual genomes of 114 indigenous African cattle and two African buffalo, *Syncerus caffer caffer* (AFB), were sequenced to an average of ~9.91× depth coverage and jointly genotyped with 217 publicly available genomes. A total of 45 cattle breeds or populations

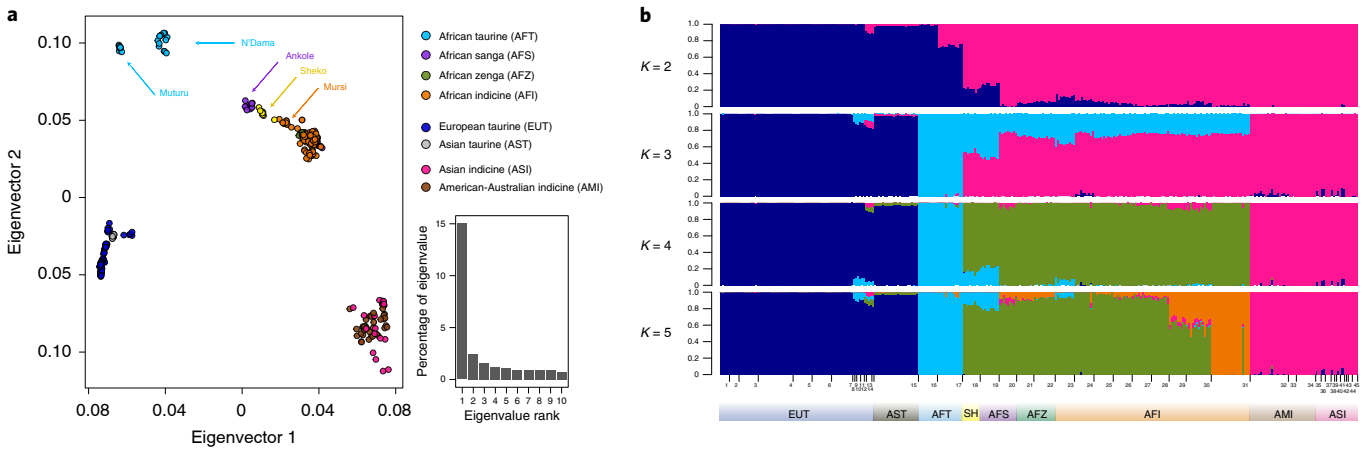


Fig. 2 | Population structure of indigenous African cattle. a, PCA results of 331 cattle samples (left), and percentage of eigenvalues (right). The Sheko is indicated in yellow. **b**, Results of admixture analysis for *K* 2–5. The 45 cattle breeds are listed from left to right as follows: (1) Eastern Finn, (2) Western Finn, (3) Angus, (4) Hereford, (5) Jersey, (6) Holstein, (7) Simmental, (8) Limia, (9) Maronesa, (10) Pajuna, (11) Sayaguesa, (12) Boskarin, (13) Maremmana, (14) Podolica, (15) Hanwoo, (16) Muturu, (17) N'Dama, (18) Sheko (SH), (19) Ankole, (20) Afar, (21) Fogera, (22) Horro, (23) Mursi, (24) Kenya Boran, (25) Goffa, (26) Arsi, (27) Ethiopian Boran, (28) Ogaden, (29) Barka, (30) Kenana, (31) Butana, (32) Brahman, (33) Gir, (34) Nelore, (35) Hariana, (36) Achai, (37) Bhagnari, (38) Cholistani, (39) Dajal, (40) Dhanni, (41) Gabrali, (42) Lohani, (43) Red Sindhi, (44) Sahiwal and (45) Tharparkar.

including 331 samples were classified according to their phenotypes as follows: African Taurine (AFT)[3], African Humped cattle (AFH) (including African Indicine (AFI)[3,31,39,40], African Sanga (AFS)[31,40], African Zenga (AFZ)[3] and Sheko), Eurasian Taurine (EAT) (including European Taurine (EUT) and Asian Taurine (AST)) and American-Australian-Asian Indicine (AAI) (including American-Australian Indicine (AMI) and Asian Indicine (ASI)) (Fig. 1, Supplementary Note and Supplementary Table 1).

We generated ~35 billion reads or ~3.50 terabytes of sequences. Sequence reads were aligned to the taurine reference genome (ARS-UCD1.2) with an average alignment rate of 99.47% (minimum: 91.70%, maximum: 99.91%) and covering 94.93% (minimum: 83.05%, maximum: 96.20%) of the reference genome. Concordant with a previous analysis of a zebu cattle, Nelore[41], the average alignment rate for AFH (99.67%) was comparable to the one obtained for AFT (99.43%) (Supplementary Table 2). Average genotype concordance of 114 samples was 95.40%, which was subsequently improved to 97.35% after genotype refinement using BEAGLE[42] (Supplementary Table 3 and Extended Data Fig. 1).

Population structure and genetic diversity of African cattle. *Population structure and relationships.* To characterize the structure of the African populations, we performed principal component analysis (PCA) of the 331 animals (Fig. 2a). All AFH position between EAT and AAI, along eigenvector 1, which explains ~15% of the total variation. AFT Muturu and N'Dama are close to EAT along the eigenvector 1. Most of the AFH cattle cluster together regardless of their breed memberships, leaving only Ankole, Mursi and Sheko outside the main cluster toward the AFT Muturu and N'Dama. The PCA results also show that Muturu and N'Dama, our representative of the AFT population, are separated from the other cattle groups (eigenvector 2, ~2.5% of total variation). Sheko positions close to the AFH, as similarly reported in other studies[5,43].

Genetic clustering analysis using ADMIXTURE[44] corroborates the pattern found in PCA (Fig. 2b and Extended Data Fig. 2). Most of AFH show a similar proportion of taurine ancestry, around 25% on average. Only a few AFH breeds have elevated taurine ancestry: Ankole (53.37 ± 1.49%), Sheko (46.28 ± 2.03%) and Mursi (35.90 ± 2.16%) (Fig. 2b; $K = 3$, light blue).

Genetic distance and diversity. Pairwise fixation index (F_{st}) values were calculated to estimate the genetic distances between populations ($n = 38$) (Extended Data Fig. 3). Taurine populations (EUT, AST and AFT) show F_{st} values of 0.1568 and 0.3287 on average against AFH and AAI, respectively. Across AFH, the pairwise F_{st} between breeds is close to zero, regardless of their phenotypic classification as African Zebu, Sanga or Zenga. Muturu and N'Dama show an F_{st} value of 0.1769, 0.1847 and 0.3734 against AFH, EAT and AAI, respectively.

The genome-wide autosomal SNPs show reduced levels of heterozygosity in the taurine (0.0021 ± 0.0005 per base pair (bp)) compared with all other populations (0.0048 ± 0.0008 per bp). Heterozygosity values of AFH are similarly higher across populations (0.0046 ± 0.0003 per bp). AAI shows a higher level of heterozygosity compared with AFH (0.0052 ± 0.0014 per bp) (Extended Data Fig. 4). The degree of inbreeding measured by runs of homozygosity (ROH) shows that taurine populations, including Muturu and N'Dama, have a higher level of inbreeding compared with the other populations. AAI shows a similar pattern of ROH distribution to AFH (Extended Data Fig. 5).

Genome-wide admixture signatures in African cattle. *Evidence of intensive admixture across African cattle.* To further analyze and quantify admixture levels in African cattle, we examined patterns of allele sharing using f_3, D and f_4 ratio statistics[45]. In the group-based analyses, we used EAT and AAI as a single group considering their genetic similarity compared with the African populations. Only Muturu and N'Dama show no evidence of admixture in f_3 analysis assuming EAT and AAI as proxies for unadmixed taurine and indicine cattle, respectively. For the D statistics, which are more robust to the effect of population-specific drift, this is only the case for the Muturu (Fig. 3a and Supplementary Tables 4 and 5). The positive f_3 statistic in N'Dama is likely due to a recent population bottleneck and subsequent allele frequency changes by genetic drift[45], as suggested by its high ROH counts and lengths (Extended Data Fig. 5). As Muturu shows no evidence of admixture (Supplementary Tables 4 and 5), we recalculated f_3 and D statistics using Muturu as a proxy for unadmixed taurine. These showed consistent results compared with those when EAT was the proxy (Supplementary Tables 4 and 5). The admixture proportions estimated by f_4 ratio statistics

(Fig. 3b and Supplementary Table 6) range from 21.03% to 26.85% in the AFH (excluding Mursi, Ankole and Sheko).

Dating taurine×indicine admixture across African cattle. Having established the level of taurine×indicine admixture among African cattle, we then estimated the timing of its generation using admixture LD decay. We first employed a single-pulse admixture model using ALDER. Across all AFH populations, excluding the Kenya Boran, admixture times range from 126.88 (Mursi) to 181.58 (Fogera) generations ago (mean 153.67) (Fig. 3c and Supplementary Table 7). Additionally, we analyzed our data using MALDER[46] to assess the possibility of multiple admixture events. After fitting a single-pulse model, MALDER analysis did not add a new admixture event with enough significance. Also, the lower significance (*Z*-score) and larger standard errors of the double-pulse model fitting compared with the single-pulse model fitting support the single-pulse admixture model for our data (Fig. 3d). When we combined AFH populations, excluding Ankole, Kenya Boran, Mursi and Sheko, we obtained a similar result (Supplementary Table 8).

Only the Kenya Boran has a different timing of admixture among the AFH populations, with a very recent admixture signal and similar significances for both the single- and double-pulse model fittings (Fig. 3d). These results support recent and ancient admixture signals in Kenya Boran (Extended Data Fig. 6). The Kenya Boran originates from the Ethiopian Boran[47,48]. After they migrated from Ethiopia to Kenya, they underwent selection and improvement with European taurine in the early twentieth century[47,48]. These recent crossbreeding events most likely correspond to the admixture signal (12.77 ± 12.96 generations ago) of the Kenya Boran (Extended Data Fig. 6). We also detect an ancient admixture signal (132.28 ± 13.60 generations ago) in the Sheko.

In N'Dama, we detected only a recent admixture signal (21.36 ± 2.50 generations ago) (Supplementary Table 7). Previous studies have shown that the N'Dama is composed of several subpopulations with varying degrees of indicine ancestry[5,24,49]. The N'Dama population here is from The Gambia, where an indicine ancestry has previously been documented[5,24,49]. Our results now provide a timescale for this recent admixture event.

We also performed GLOBETROTTER[50] analysis, based on haplotype sharing, as an alternative method to estimate admixture time. The 14 African cattle populations, excluding Muturu and N'Dama, show robust evidence of admixture (bootstrap $P < 0.01$) (Supplementary Table 9). In addition, admixture time estimates from the populations with best-guess model 'one-date' range from 94.85 to 158.08 generations ago, in agreement with the results from ALDER (Fig. 3e). The exceptions are the Kenya Boran and Kenana, with best-guess model 'multiple-dates' (Supplementary Table 9).

Selection signatures with an excess of taurine or indicine ancestry in African humped cattle. Our genome-wide analysis shows that all sampled African cattle breeds, except Muturu, have taurine and indicine ancestry, with little variation within a population. In such crossbreeds, a haplotype of either taurine or indicine ancestry may confer a relative adaptive advantage following selection pressures. Accordingly, such haplotypes will be selected in the admixed African cattle population over time.

We employed two approaches to identify such loci and haplotypes. We first explored ongoing selective sweep using the integrated haplotype score (*iHS*). Taking the top 1% windows in terms of the proportion of SNPs with $|iHS| \geq 2$ (≥60.00%), we obtained a total of 496 windows of 50 kilobase (kb) length as candidates under selection (Extended Data Fig. 7a). The 494 protein-coding genes overlapped with these windows show significant enrichment in 'defense response to bacterium' (GO:0042742) and 'keratinization' (R-BTA-6805567) (false discovery rate–adjusted $P < 0.05$) (Supplementary Table 10). These 496 windows have a lower average taurine ancestry (26.14%) than other *iHS* percentiles as well as the whole genome (32.49%) (Extended Data Figs. 8 and 9). Also, the average taurine ancestry of the windows is outside the empirical distribution generated by resampling (Extended Data Fig. 10). This indicates that the overall ancestry of these selected loci is more skewed toward indicine than the whole genome.

We then inferred local ancestry across the genome using LOTER[51] and selected the top 0.5% windows with the highest taurine or indicine ancestry (Extended Data Fig. 7b). Of these 496 windows, 63 windows identified in the previous *iHS* analysis were further considered. After filtering out windows with pairwise F_{st} value between the reference populations (EAT and AAI) less than the genome-wide level (<0.2296) and merging adjacent windows, 16 genomic regions were retained, of which three and 13 show an excess of taurine and indicine ancestry, respectively. Eleven of the regions with an excess of indicine ancestry have been identified as selection signal in previous African cattle studies (Table 1). To our knowledge, none of the regions with an excess of taurine ancestry was previously reported under selection in African cattle. The taurine and indicine excess regions overlap with nine and 51 protein-coding genes, respectively.

The longest region, 600 kb in length, is observed at BTA7 (Table 1). It includes 12 significant windows with 92.05% average indicine ancestry, which is much higher than the 67.51% genome-wide average. Downstream of this region, we found three smaller regions of 150-, 200- and 50-kb length with high average indicine ancestry of 91.28%, 91.28% and 92.62%, respectively (Table 1). This cluster of four candidate regions spans 1.40 megabases (Mb) of BTA7 (49.75–51.15 Mb). It shows a reduced level of diversity within AFH and an increased level of genetic differentiation between AFH and EAT.

Fig. 3 | Admixture signatures in African cattle genomes. a, *D* statistics estimating indicine gene flow in African breeds (*X*), using EAT/AAI as an ancestral taurine/indicine proxy and AFB as an outgroup; *D* (*EAT*, *X*; *AAI*, *AFB*). The dotted red line indicates the expected statistics at a neutral locus. Thick and thin horizontal bars represent ±1 s.e. and ±3 s.e., respectively. The Sheko is indicated in yellow. **b**, Admixture proportions are measured by the f_4 ratio; f_4 (*EATa*, *AFB*; *X*, *AAI*)/f_4 (*EATa*, *AFB*; *EATb*, *AAI*). EAT are randomly divided into two subgroups, EATa and EATb, and AFB is the outgroup. Blue and pink colors indicate taurine and indicine ancestries, respectively. **c**, Admixture times in generations are estimated by ALDER[70] with two reference populations, EAT ($n = 103$) and AAI ($n = 56$). The numbers of biologically independent animals used in this analysis for each breed are as follows: Afar (9), Ankole (10), Arsi (10), Barka (9), Butana (20), Ethiopian Boran (10), Fogera (9), Goffa (10), Horro (11), Kenya Boran (10), Kenana (13), Mursi (10), N'Dama (13), Ogaden (9) and Sheko (9). The data points are presented as estimated admixture times in generations ±s.e. Thick and thin horizontal bars represent ±1 s.e. and ±3 s.e., respectively. The Sheko is indicated in yellow. **d**, Admixture times in generations are estimated by both single- (left) and double-pulse (middle and right) models using MALDER[46] with two reference populations, EAT ($n = 103$) and AAI ($n = 56$). The numbers of biologically independent animals used in this analysis for each breed are identical to those of the ALDER analysis in **c**. The data points are presented as estimated admixture times in generations ±1 s.e. The *y* axis indicates *Z*-score for each model fitting. **e**, The comparison between estimates from the GLOBETROTTER analysis (*x* axis) and those from ALDER analysis (*y* axis). The red line indicates $y = x$. The data points are presented as estimated admixture times in generations ±1 s.e. (horizontal and vertical bars). Standard errors were estimated by leave-one-chromosome-out jackknifing (ALDER) or by bootstrapping (GLOBETROTTER). The numbers of biologically independent animals used in each of the analyses for each breed are identical to those of the ALDER analysis in **c**. The Sheko is indicated in yellow.

Shared haplotypes are more commonly observed between AFH and AAI than AFH and EAT (Fig. 4). In this cluster, we identified 18 protein-coding genes, three related to the host immune (*MATR3* (ref. [52]), *MZB1* (ref. [53]) and *STING1* (ref. [54])) and one to the environmental thermal stresses (heat shock protein gene *DNAJC18* (ref. [55])) responses. We also found one more heat shock protein gene (*HSPA9* (ref. [51])) with an excess of indicine ancestry (BTA7: 49.85–49.95 Mb; 91.30% average indicine ancestry), but here the *iHS* (36.98%) does not reach the significance threshold. Two protein-coding genes linked to reproduction (*PAIP2* (ref. [56]) and *SPATA24* (ref. [57])) are also found in this region, together with *SEPTIN2* (ref. [58]) on BTA3 (Table 1).

The region with the highest taurine ancestry (61.34%) is of 200-kb length (BTA11: 14.65–14.85 Mb) (Table 1). As for the BTA7 region, it shows reduced genetic diversity (Fig. 5). However, we observe an increased level of genetic differentiation between AFH and AAI as well as extended haplotype sharing between EAT and AFH (Fig. 5). This region overlaps with seven protein-coding genes (Table 1), one of which linked to the inflammatory response[59–61] and tick infestation[62] (*NLRC4*).

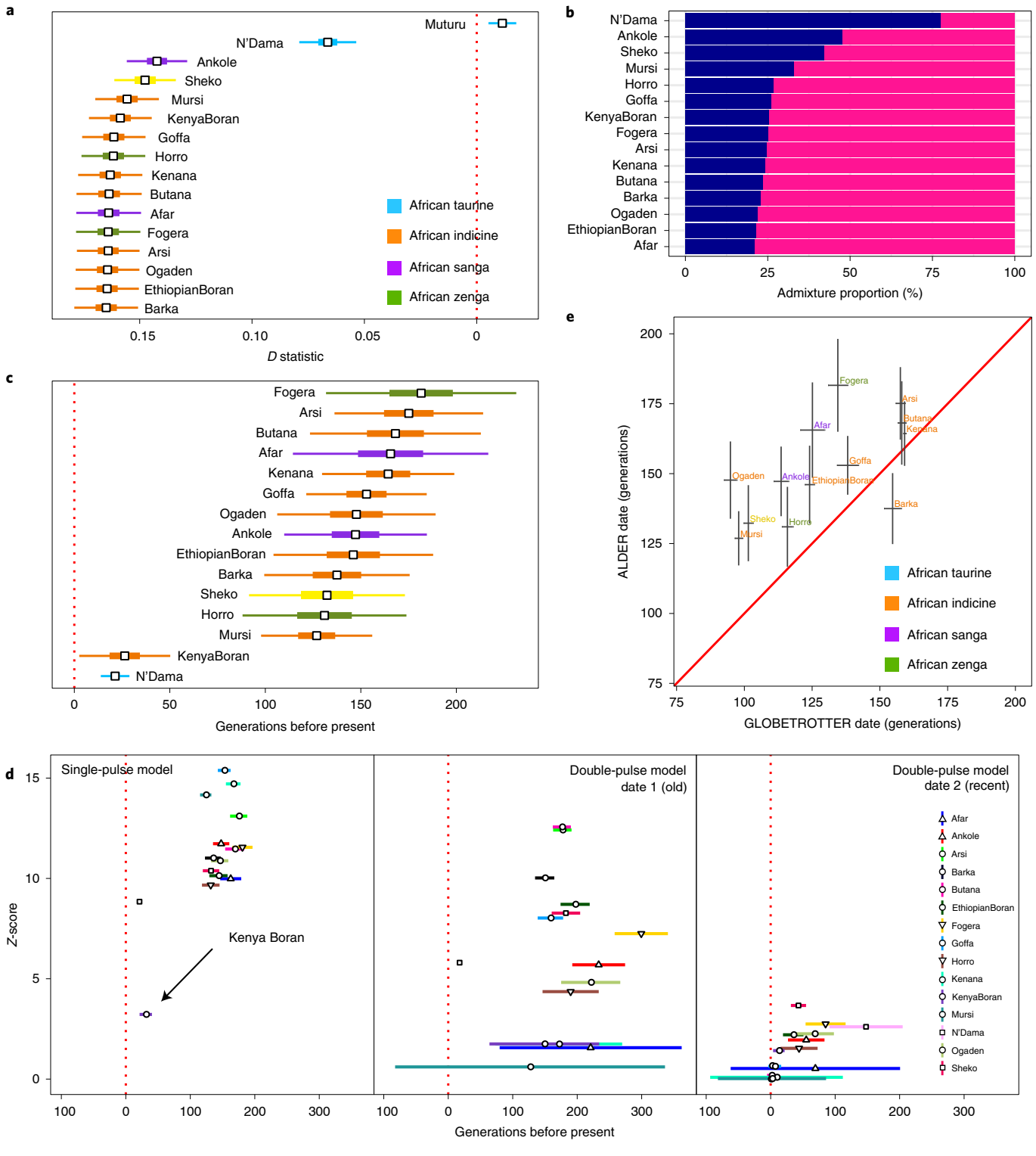


Table 1 | Common AFH candidate regions identified in the *iHS* and local ancestry (taurine or indicine) inference (LOTER, top 0.5% windows) analysis

BTA[a]	Region (Mb)	No. of windows	Proportion of SNPs with $\|iHS\| \geq 2$ (%)	Ancestry (%)	F_{st}	Genes identified	Previous studies
Regions with an excess of indicine ancestry							
3	120.30–120.40	2	67.74	93.02	0.3390	*PASK, PPP1R7, SNED1, MTERF4*	Kim et al.[31]
3	120.45–120.55	2	63.33	92.86	0.2913	*SEPTIN2, FARP2, HDLBP*	Makina et al.[99]
3	120.60–120.65	1	79.35	92.62	0.2875	*FARP2, STK25, BOK*	Makina et al.[99]
3	120.70–120.80	2	83.36	92.62	0.2553	*ING5, D2HGDH, THAP4, ATG4B, DTYMK*	Kim et al.[31] Makina et al.[99]
3	120.85–120.90	1	79.25	92.62	0.3182	*RTP5*	Makina et al.[99]
7	49.75–49.80	1	65.74	92.62	0.3817	*KDM3B*	Gautier et al.[100]
7	50.05–50.25	4	67.90	91.28	0.4179	*CTNNA1, LRRTM2, ENSBTAG00000004415*	Kim et al.[31] Gautier et al.[100]
7	50.30–50.45	3	75.17	91.28	0.6321	*SIL1*	Kim et al.[31] Gautier et al.[100]
7	50.55–51.15	12	86.06	92.05	0.4861	*PSD2, NRG2, DNAJC18, ECSCR, SMIM33, STING1, CXXC5, UBE2D2, MATR3, PAIP2, SLC23A1, MZB1, PROB1, SPATA24*	Bahbahani et al.[30] Kim et al.[31] Bahbahani et al.[76] Gautier et al.[100]
13	56.95–57.00	1	82.80	93.58	0.3090	–	–
13	57.05–57.10	1	73.94	93.76	0.2685	*EDN3*	–
13	57.15–57.65	10	81.95	92.69	0.3114	*PRELID3B, ATP5F1E, TUBB1, CTSZ, NELFCD, ZNF831, GNAS*	Kim et al.[31] Bahbahani et al.[76]
19	39.65–39.85	4	67.07	92.44	0.2982	*STAC2, FBXL20, MED1, PLXDC1, CACNB1, RPL19, ENSBTAG00000008368, ENSBTAG00000050597*	Bahbahani et al.[30] Gautier et al.[100]
Regions with an excess of taurine ancestry							
10	92.15–92.25	2	72.23	59.98	0.3211	*CEP128, ENSBTAG00000047322*	–
11	14.40–14.45	1	67.08	61.19	0.4337	–	–
11	14.65–14.85	4	78.31	61.34	0.2870	*MEMO1, DPY30, SPAST, SLC30A6, NLRC4, ENSBTAG00000048521, ENSBTAG00000049576*	–

[a]*B. taurus* autosomes. The proportion (%) of SNPs ($|iHS| \geq 2$) and ancestries are averaged values over windows. The F_{st} values are pairwise values between reference populations (EAT and AAI) averaged over windows. Dashes (–) indicate that no genes have been annotated within the region or have not overlapped with candidate selection signals in African cattle from previous studies.

African taurine-specific loci and their distribution in African humped cattle. Taurine cattle are the most ancient African cattle population. They have adapted to the local environmental challenges, as exemplified by the trypanotolerance traits of West African taurine cattle[63]. Accordingly, their unique genetic components may confer a selective advantage in crossbreed animals facing similar environmental challenges to the West African taurine.

To identify such loci, we performed population branch statistics (PBS) analysis[64], comparing AFT and EAT using AAI as an outgroup. After filtering out windows with less than 10 SNPs, we remained with 1,239,021 autosomal windows (50-kb sliding windows with 2-kb overlapping step). PBS values ranged from −0.1156 to 0.8341, with a mean of 0.0314. After removing windows with F_{st} value (AFT versus EAT) less than 0.1 (Supplementary Fig. 1) from the highest 0.1% PBS windows, we considered the remaining windows as candidate selection signal specific to AFT (Supplementary Table 11).

The strongest PBS signal (0.6740) overlaps with *SDK1* on BTA25 (40,052,001–40,102,000), approximately 300 kb upstream of *CARD11* (Fig. 6). At this region, F_{st} values between AFT and EAT ($F_{st}=0.5173$) or AAI ($F_{st}=0.5308$) are much higher than the genome-wide level ($F_{st}=0.1106$ and $F_{st}=0.1825$, respectively) (Fig. 6b). We observe a unique AFT haplotype pattern compared with EAT and AAI, which is present in some AFH breeds (Supplementary Figs. 2 and 3).

Discussion

In this study, we first highlighted the taurine × indicine admixture characteristics of 16 indigenous African cattle populations, 14 of them living in the Horn of Africa, the main entry point of Asian zebu on the African continent. Then, we identified and dated the main taurine × indicine admixture event, which has shaped today's genome of these crossbreeds, to around 150 generations ago. We also identified candidate selected regions in these admixed populations, including immune-response- and heat-tolerance-related genes in haplotypes of indicine origins and inflammatory-response-related genes in haplotypes of taurine origins. Last, we identified a locus of African taurine origin putatively linked to trypanotolerance. Together, these results support our hypothesis that the present success and dispersion of African pastoralism followed the arrival of indicine cattle and their crossbreeding with local taurine cattle.

Our estimation under a single-pulse admixture model dates back the admixture time of AFH to around 150 generations ago. Assuming a cattle generation time of 5–7 yr (refs. [65,66]), it corresponds

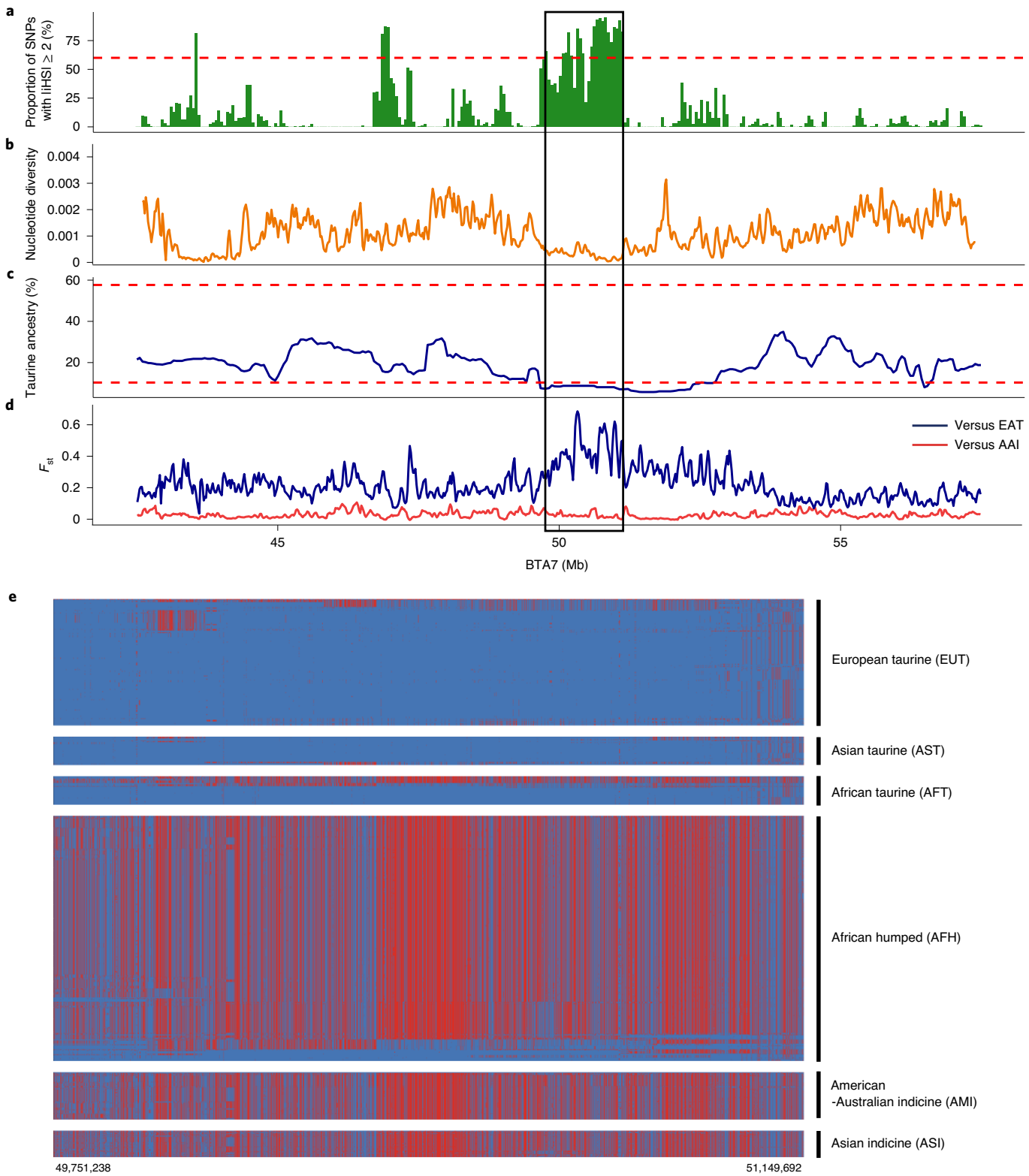


Fig. 4 | Example of candidate selective loci on BTA7 with an excess of indicine ancestry. a, Proportion of SNPs with $|iHS| \geq 2$ in each nonoverlapping 50-kb window around the candidate locus (BTA7: 49.75–51.15 Mb, the black square) including *MATR3*, *MZB1*, *STING1* and *DNAJC18*. The dashed red line indicates the top 1% proportion of SNPs with $|iHS| \geq 2$ (60.00%). **b**, Nucleotide diversity calculated using VCFtools v.0.1.17 (ref. [102]) for each 50-kb window with 20-kb step around the candidate locus. **c**, Average taurine ancestry (%) in each nonoverlapping 50-kb window around the candidate locus. The lower and upper dashed red lines indicate the lowest and highest 0.5% of average taurine ancestry, respectively (10.31% and 57.67%). **d**, Pairwise F_{st} values calculated using VCFtools v.0.1.17 (ref. [102]) for each 50-kb window with 20-kb step around the candidate locus. The blue line indicates the pairwise F_{st} values between AFH and EAT. The red line indicates the pairwise F_{st} values between AFH and AAI. **e**, Haplotype sharing at the candidate locus. The haplotypes were hierarchically clustered within each cattle group. The major allele in EAT (allele frequency ≥ 50%) is indicated in blue.

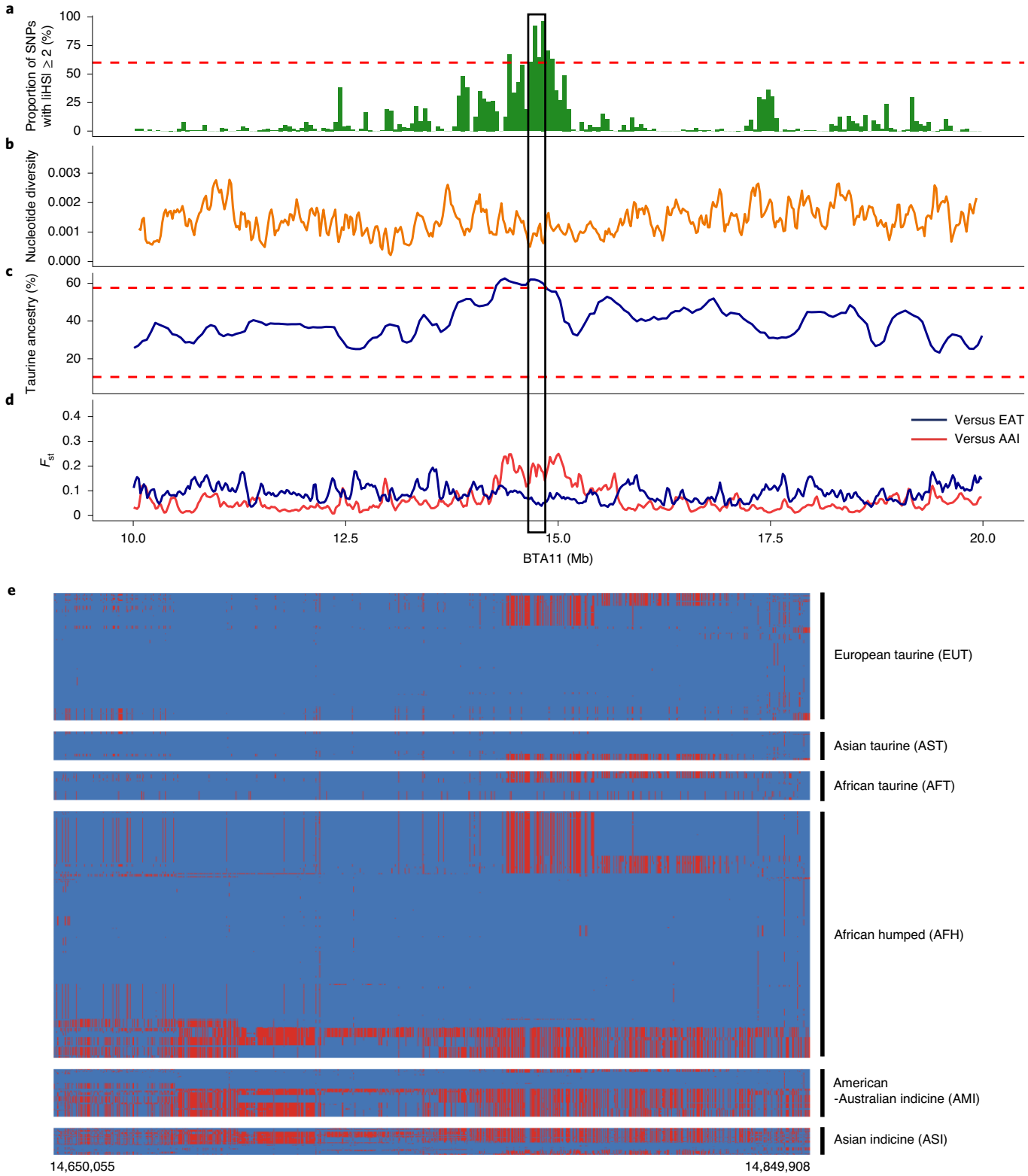


Fig. 5 | Example of candidate selective loci on BTA11 with an excess of taurine ancestry. a, The proportion of SNPs with $|iHS| \geq 2$ in each nonoverlapping 50-kb window around the candidate locus (BTA11: 14.65–14.85 Mb, the black square) including *NLRC4*. The dashed red line indicates the top 1% proportion of SNPs with $|iHS| \geq 2$ (60.00%). **b**, Nucleotide diversity calculated using VCFtools v.0.1.17 (ref. [102]) for each 50-kb window with 20-kb step around the candidate locus. **c**, Average taurine ancestry (%) in each nonoverlapping 50-kb window around the candidate locus. The lower and upper dashed red lines indicate the lowest and highest 0.5% of average taurine ancestry, respectively (10.31% and 57.67%). **d**, Pairwise F_{st} value calculated using VCFtools v.0.1.17 (ref. [102]) for each 50-kb window with 20-kb step around the candidate locus. The blue line indicates the pairwise F_{st} values between AFH and EAT. The red line indicates the pairwise F_{st} values between AFH and AAI. **e**, Haplotype sharing at the candidate locus. The haplotypes were hierarchically clustered within each cattle group. The major allele in EAT (allele frequency ≥ 50%) is indicated in blue.

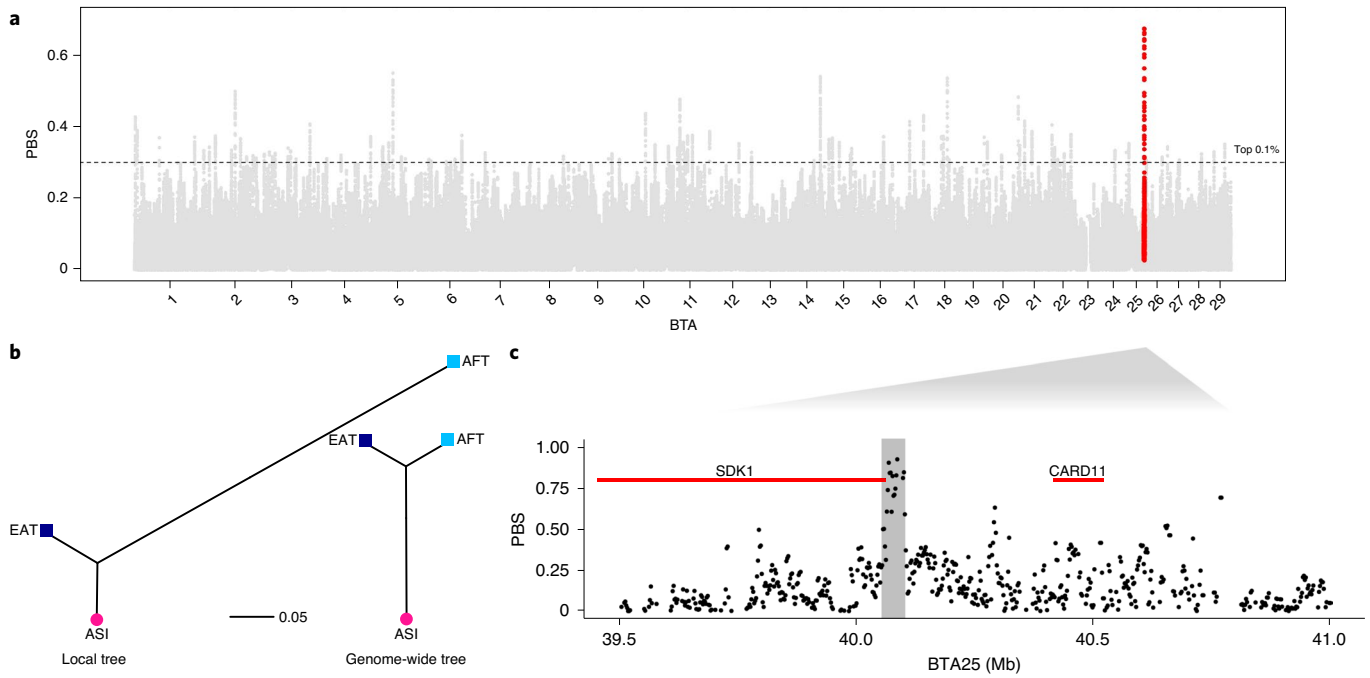


Fig. 6 | Unique selection signatures in African taurine cattle following their separation from the common ancestor with Eurasian taurine cattle. **a**, Genome-wide distribution of PBS values with 50-kb window and 2-kb step. The windows with F_{st} value (AFT versus EAT) < 0.1 or PBS < 0 are not plotted. The dashed line indicates the top 0.1% PBS values. **b**, F_{st}-based phylogeny among AFT, EAT and AAI. The branch lengths are proportional to F_{st} values. Genome-wide F_{st} values ±s.d. are as follows for each comparison; AFT versus EAT: 0.1106 ± 0.0494; AFT versus AAI: 0.1825 ± 0.0490; and EAT versus AAI: 0.2296 ± 0.0493. **c**, PBS values around the peak with the highest PBS value. The PBS values were calculated with 5-kb window and 2-kb step. BTA, *B. taurus* autosomes.

to about 750–1,050 yr ago at the beginning of the second millennium AD (950–1250 AD). According to historical records, Asian zebu arrival along the Horn of Africa started earlier, around 700 AD, following the Islamization of the East African coast and the onset of the Swahili civilization[19], in agreement with the earliest noncontroversial archeological evidence in the Horn of Africa for African humped cattle, dated around the mid-first millennium AD[18]. Therefore, our results suggest that indicine cattle remained initially confined to the East African coastal areas for at least 2–3 centuries before crossing extensively with taurine cattle. Then, during the second millennium AD, the complex human history of the Horn of Africa, characterized by multiple human population movements and dispersion[67] as well as climatic fluctuation[16,68], would have further contributed to the landscape of today's genome admixture in East African cattle. Interestingly, a previous study indicates an admixture event in two West African zebu populations at around 500 yr ago[66]. This timing is in agreement with our earlier East African dating of taurine × indicine crossbreeding, which would have been followed by the movement of East African humped cattle along the Sahelian belt and crossbreeding with local taurine cattle in West Africa. The same study identified a more recent admixture event in the West African Borgou around 20 generations ago[66]. This is at approximately the same time as the one identified in our study in the N'Dama from The Gambia. These more recent admixture events may have been linked to the rinderpest epidemics of the end of the nineteenth century[69].

We cannot exclude the possibility that more ancient taurine × indicine admixture events have contributed to the genetic composition of the AFH population from the Horn of Africa. Indeed, the haplotype sharing-based and LD-based admixture dating have limited power to detect admixture signals older than about 200 generations ago[50,70]. However, if this was the case, their admixture signals would have been likely erased by the more recent ones identified here.

The ancestry of the selection signatures in AFH was found to be more skewed toward indicine than the genome-wide average. Domestic cattle are not native to the African continent; African taurine cattle originate from the Near East[3], while indicine cattle were introduced into Africa after their domestication on the Indian subcontinent[3]. On reaching the African tropical environments, the Near East taurine cattle must have faced major environmental challenges. However, indicine cattle found across the tropical Indian subcontinent may have been better preadapted to African environments and, in particular, to its climatic characteristics[71]. These preadaptations would have facilitated indicine introgression into local inland taurine populations and the dispersion of crossbred animals. However, African livestock diseases (for example, trypanosomosis, bovine malignant catarrhal fever, East Coast fever and Rift Valley fever) would have represented major constraints to the dispersion of indicine × taurine crossbred cattle[22]. Here, the tolerance of African taurine cattle to trypanosomosis[4] as well as the resistance of indicine cattle to infestation with ticks and to heat stress have proven advantageous[72–74].

Heat tolerance, a characteristic of zebu cattle[73,74], is a candidate for indicine preadaptations to climatic challenges. We found two heat shock protein genes (*HSPA9* and *DNAJC18*) at BTA7, which were previously reported as candidate selective loci in African and Asian indicine cattle[30,75–77]. We also found a water-reabsorption-related gene, *GNAS*, at BTA13. The protein encoded by *GNAS* mediates antidiuretic hormone arginine vasopressin (AVP) to aquaporin-2 (AQP2) water channels, contributing to the water conservation pathway of the kidney[78]. Considering the adaptation of Asian zebu cattle to the arid environments[79], we infer that the indicine haplotype of *GNAS* contributes to the local adaptation of AFH to the arid areas of the continent. Also, the immune-related genes at BTA7 (*MATR3, MZB1* and *STING1*) and BTA3 (*ATG4B* (ref. [80])) (Table 1) might be related to the resistance of indicine cattle to

ticks and tick-borne diseases, such as East Coast fever. *STING1* is essential for DNA-mediated type I interferon production and host defense against DNA viral pathogens[81], and therefore might confer some tolerance to viral infections such as Rift Valley fever and foot-and-mouth disease.

The identification of an autosomal taurine background in all African cattle leads us to expect a contribution of local taurine ancestry to environmental adaptation and thus its contribution to the success of African cattle pastoralism. One example is the candidate region at BTA11, which overlaps with *NLRC4* (ref. [59]) involved in the inflammatory response. It shows extensive haplotype sharing between AFH and taurine cattle (AFT and EAT). Considering the lack of EAT ancestry in AFH cattle, this haplotype likely originates from AFT. Its presence in AFH may have resulted from selection for a better control of the inflammatory response following infections with diseases such as East Coast Fever and Rift Valley Fever[82,83].

Similarly, across large areas of sub-Saharan Africa, cattle have been exposed to the challenge of trypanosomosis, a severe obstacle to livestock productivity in Africa[84]. African taurine cattle show tolerance to *Trypanosoma sp* infection, controlling both the effect of infection (for example, anemia and weight loss) and the level of blood parasites[85]. Accordingly, we expect to detect selection signals in some of the humped cattle exposed to trypanosomosis challenges.

In our PBS analysis, a selection signature in AFT was found upstream of *CARD11*, which encodes a protein essential for the signaling of T and B cells in the innate and adaptive immune systems[86–88]. Importantly, it was reported as a differentially expressed gene between the trypanotolerant N'Dama and trypanosusceptible Kenya Boran[89]. We suggest that this candidate region plays a role in regulating *CARD11* expression and contributes to the adaptation of AFT and AFH populations to trypanosomosis challenge. Accordingly, this taurine region is expected to be observed in crossbreeds (Sheko, Horro and Mursi), whose natural habitats are infested with tsetse flies[90,91]. However, as a complex quantitative trait[92–94], the potential regulatory element upstream of *CARD11* should be regarded as one of many genetic factors contributing to trypanotolerance. Accordingly, it is worth mentioning that the windows within the highest 0.1% PBS value include several genes (*FAAP24* (ref. [95]), *WDR48* (ref. [96]), *LRRC8A* (ref. [97]) and *IFNAR1* (ref. [98])) related to anemia and immune response (Supplementary Table 11).

In conclusion, despite the environmental complexity of the African continent, and cattle domestication outside its geographic area, we currently find domestic cattle across all African agro-ecologies. The results presented here support that taurine×indicine admixture events followed by taurine and indicine ancestry selection across the genome is at the root of the success of African cattle pastoralism. These findings are far-reaching in today's context of improving livestock productivity to respond to the needs of the growing human populations, with further crossbreeding of indigenous African cattle with exotic cattle recommended as one of the pathways for the continent's food security. A complete characterization at the genome level of African cattle unique adaptations will open the door to sustainable crossbreeding programs combining local environmental adaptation and increased exotic productivity.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-020-0694-2.

Received: 18 October 2019; Accepted: 18 August 2020;

References

1. Schneider, H. K. A model of African indigenous economy and society. *Comp. Stud. Soc. Hist.* **7**, 37–55 (1964).
2. Di Lernia, S. et al. Inside the 'African cattle complex': animal burials in the Holocene central Sahara. *PLoS ONE* **8**, e56879 (2013).
3. 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**, 911–921 (2015).
4. Roberts, C. & Gray, A. Studies on trypanosome-resistant cattle. II. The effect of trypanosomiasis on N'dama, Muturu and Zebu cattle. *Trop. Anim. Health Prod.* **5**, 220–233 (1973).
5. 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).
6. Hanotte, O. et al. African pastoralism: genetic imprints of origins and migrations. *Science* **296**, 336–339 (2002).
7. 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. USA* **91**, 2757–2761 (1994).
8. MacHugh, D. E., Shriver, M. D., Loftus, R. T., Cunningham, P. & Bradley, D. G. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics* **146**, 1071–1086 (1997).
9. Achilli, A. et al. Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Curr. Biol.* **18**, R157–R158 (2008).
10. Bibi, F. A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. *BMC Evolut. Biol.* **13**, 166 (2013).
11. Gifford-Gonzalez, D. & Hanotte, O. Domesticating animals in Africa. in *The Oxford Handbook of African Archaeology* 491–506 (Oxford University Press, 2013).
12. Blench, R. & MacDonald, K. *The Origins and Development of African Livestock: Archaeology, Genetics, Linguistics and Ethnography* (Routledge, 2006).
13. Ajmone-Marsan, P., Garcia, J. F. & Lenstra, J. A. On the origin of cattle: how aurochs became cattle and colonized the world. *Evol. Anthropol.* **19**, 148–157 (2010).
14. Manning, K. The first herders of the West African Sahel: inter-site comparative analysis of zooarchaeological data from the lower Tilemsi Valley, Mali. in *People and Animals in Holocene Africa. Recent Advances in Archaeozoology* 75–85 (Africa Magna, 2011).
15. Hildebrand, E. A. & Grillo, K. M. Early herders and monumental sites in eastern Africa: dating and interpretation. *Antiquity* **86**, 338–352 (2012).
16. Chritz, K. L. et al. Climate, ecology, and the spread of herding in eastern Africa. *Quat. Sci. Rev.* **204**, 119–132 (2019).
17. Lesur, J., Hildebrand, E. A., Abawa, G. & Gutherz, X. The advent of herding in the Horn of Africa: new data from Ethiopia, Djibouti and Somaliland. *Quat. Int.* **343**, 148–158 (2014).
18. Gifford-Gonzalez, D. & Hanotte, O. Domesticating animals in Africa: implications of genetic and archaeological findings. *J. World Prehist.* **24**, 1–23 (2011).
19. Epstein, H. *The Origin of the Domestic Animals of Africa* (Africana Publishing Corporation, 1971).
20. Gifford-Gonzalez, D. Animal disease challenges to the emergence of pastoralism in sub-Saharan Africa. *Afr. Archaeol. Rev.* **17**, 95–139 (2000).
21. Sadr, K. The archaeology of herding in southernmost Africa. in *The Oxford Handbook of African Archaeology* 645–655 (Oxford University Press, 2013).
22. Gifford-Gonzalez, D. 'Animal disease challenges' fifteen years later: the hypothesis in light of new data. *Quat. Int.* **436**, 283–293 (2017).
23. Felius, M., Koolmees, P. A., Theunissen, B., European Cattle Genetic Diversity Consortium & Lenstra, J. A. On the breeds of cattle—historic and current classifications. *Diversity* **3**, 660–692 (2011).
24. Freeman, A. et al. Admixture and diversity in West African cattle populations. *Mol. Ecol.* **13**, 3477–3487 (2004).
25. Bradley, D. G., MacHugh, D. E., Cunningham, P. & Loftus, R. T. Mitochondrial diversity and the origins of African and European cattle. *Proc. Natl Acad. Sci. USA* **93**, 5131–5135 (1996).
26. Bonfiglio, S. et al. Origin and spread of *Bos taurus*: new clues from mitochondrial genomes belonging to haplogroup T1. *PLoS ONE* **7**, e38601 (2012).
27. Tarekegn, G. M. et al. Variations in mitochondrial cytochrome b region among Ethiopian indigenous cattle populations assert *Bos taurus* maternal origin and historical dynamics. *Asian-Australas. J. Anim. Sci.* **31**, 1393 (2018).
28. Pérez-Pardal, L. et al. Y-specific microsatellites reveal an African subfamily in taurine (*Bos taurus*) cattle. *Anim. Genet.* **41**, 232–241 (2010).
29. Mbole-Kariuki, M. N. et al. Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya. *Heredity* **113**, 297 (2014).

30. Bahbahani, H. et al. Signatures of selection for environmental adaptation and zebu × taurine hybrid fitness in East African Shorthorn Zebu. *Front. Genet.* **8**, 68 (2017).
31. Kim, J. et al. The genome landscape of indigenous African cattle. *Genome Biol.* **18**, 34 (2017).
32. Verhoeven, K. J., Macel, M., Wolfe, L. M. & Biere, A. Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. *Proc. R. Soc. B Biol. Sci.* **278**, 2–8 (2010).
33. Hovick, S. M. & Whitney, K. D. Hybridisation is associated with increased fecundity and size in invasive taxa: meta-analytic support for the hybridisation-invasion hypothesis. *Ecol. Lett.* **17**, 1464–1477 (2014).
34. Medugorac, I. et al. Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. *Nat. Genet.* **49**, 470 (2017).
35. 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).
36. Wu, D.-D. et al. Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nat. Ecol. Evol.* **2**, 1139–1145 (2018).
37. Wu, C.-I. & Ting, C.-T. Genes and speciation. *Nat. Rev. Genet.* **5**, 114 (2004).
38. Tijjani, A., Utsunomiya, Y. T., Ezekwe, A., Nash, O. & Hanotte, O. H. Genome sequence analysis reveals selection signatures in endangered trypano-tolerant West African Muturu cattle. *Front. Genet.* **10**, 442 (2019).
39. Bahbahani, H. et al. Signatures of positive selection in African Butana and Kenana dairy zebu cattle. *PLoS ONE* **13**, e0190446 (2018).
40. Rege, J., Ayalew, W., Getahun, E., Hanotte, O. & Dessie, T. *DAGRIS (Domestic Animal Genetic Resources Information System)* (International Livestock Research Institute, 2006).
41. Canavez, F. C. et al. Genome sequence and assembly of *Bos indicus*. *J. Heredity* **103**, 342–348 (2012).
42. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
43. Bahbahani, H., Afana, A. & Wragg, D. Genomic signatures of adaptive introgression and environmental adaptation in the Sheko cattle of southwest Ethiopia. *PLoS ONE* **13**, e0202479 (2018).
44. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
45. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
46. Pickrell, J. K. et al. Ancient west Eurasian ancestry in southern and eastern Africa. *Proc. Natl Acad. Sci. USA* **111**, 2632–2637 (2014).
47. Porter, V., Alderson, L., Hall, S. J. & Sponenberg, D. P. *Mason's World Encyclopedia of Livestock Breeds and Breeding* (Cabi, 2016).
48. Rege, J. *Zebu Cattle of Kenya: Uses, Performance, Farmer Preferences, Measures of Genetic Diversity and Options for Improved Use* (ILRI 2001).
49. 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).
50. Hellenthal, G. et al. A genetic atlas of human admixture history. *Science* **343**, 747–751 (2014).
51. 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).
52. Morchikh, M. et al. HEXIM1 and NEAT1 long non-coding RNA form a multi-subunit complex that regulates DNA-mediated innate immune response. *Mol. Cell* **67**, 387–399.e5 (2017).
53. Flach, H. et al. Mzb1 protein regulates calcium homeostasis, antibody secretion, and integrin activation in innate-like B cells. *Immunity* **33**, 723–735 (2010).
54. Patel, S. & Jin, L. *TMEM173* variants and potential importance to human biology and disease. *Genes Immun.* **20**, 82 (2019).
55. Qiu, X.-B., Shao, Y.-M., Miao, S. & Wang, L. The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cell. Mol. Life Sci.* **63**, 2560–2570 (2006).
56. Delbes, G., Yanagiya, A., Sonenberg, N. & Robaire, B. PABP interacting protein 2 (Paip2) is a major translational regulator involved in the maturation of male germ cells and male fertility. *Biol. Reprod.* **81**, 167–167 (2009).
57. McReynolds, S. et al. Toward the identification of a subset of unexplained infertility: a sperm proteomic approach. *Fertil. Steril.* **102**, 692–699 (2014).
58. Kuo, Y.-C. et al. SEPT12 orchestrates the formation of mammalian sperm annulus by organizing core octameric complexes with other SEPT proteins. *J. Cell Sci.* **128**, 923–934 (2015).
59. Zhao, Y. et al. The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus. *Nature* **477**, 596 (2011).
60. Canna, S. W. et al. An activating NLRC4 inflammasome mutation causes autoinflammation with recurrent macrophage activation syndrome. *Nat. Genet.* **46**, 1140 (2014).
61. Kitamura, A., Sasaki, Y., Abe, T., Kano, H. & Yasutomo, K. An inherited mutation in NLRC4 causes autoinflammation in human and mice. *J. Exp. Med.* **211**, 2385–2396 (2014).
62. Wang, X. et al. The tick protein Sialostatin L2 binds to Annexin A2 and inhibits NLRC4-mediated inflammasome activation. *Infect. Immun.* **84**, 1796–1805 (2016).
63. Rege, J., Aboagye, G. & Tawah, C. Shorthorn cattle of West and Central Africa. I. Origin, distribution, classification and population statistics. *World Anim. Rev.* **78**, 2–13 (1994).
64. Yi, X. et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75–78 (2010).
65. MacEachern, S., Hayes, B., McEwan, J. & Goddard, M. An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (*Bos taurus*) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in domestic cattle. *BMC Genomics* **10**, 181 (2009).
66. Flori, L. et al. Adaptive admixture in the West African bovine hybrid zone: insight from the Borgou population. *Mol. Ecol.* **23**, 3241–3257 (2014).
67. Newman, J. L. *The Peopling of Africa: A Geographic Interpretation* (Yale University Press, 1995).
68. Russell, J. M., Verschuren, D. & Eggermont, H. Spatial complexity of 'Little Ice Age' climate in East Africa: sedimentary records from two crater lake basins in western Uganda. *Holocene* **17**, 183–193 (2007).
69. Phoofolo, P. Epidemics and revolutions: the rinderpest epidemic in late nineteenth-century Southern Africa. *Past Present* **138**, 112–143 (1993).
70. Loh, P.-R. et al. Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193**, 1233–1254 (2013).
71. Boivin, N., Crowther, A., Prendergast, M. & Fuller, D. Q. Indian Ocean food globalisation and Africa. *Afr. Archaeol. Rev.* **31**, 547–581 (2014).
72. Burrow, H. M. et al. Towards a new phenotype for tick resistance in beef and dairy cattle: a review. *Anim. Prod. Sci.* **59**, 1401–1427 (2019).
73. Hansen, P. Physiological and cellular adaptations of zebu cattle to thermal stress. *Anim. Reprod. Sci.* **82**, 349–360 (2004).
74. Mirkena, T. et al. Genetics of adaptation in domestic farm animals: a review. *Livest. Sci.* **132**, 1–12 (2010).
75. Porto-Neto, L. R. et al. Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. *BMC Genomics* **14**, 876 (2013).
76. Bahbahani, H. et al. Signatures of positive selection in East African Shorthorn Zebu: a genome-wide single nucleotide polymorphism analysis. *Sci. Rep.* **5**, 11729 (2015).
77. Kasarapu, P. et al. The *Bos taurus–Bos indicus* balance in fertility and milk related genes. *PLoS ONE* **12**, e0181930 (2017).
78. Boone, M. & Deen, P. M. Physiology and pathophysiology of the vasopressin-regulated renal water reabsorption. *Pflugers Arch.* **456**, 1005 (2008).
79. Sodhi, M. et al. Microsatellite analysis of genetic population structure of Zebu cattle (*Bos indicus*) breeds from North-Western region of India. *Anim. Biotechnol.* **22**, 16–29 (2011).
80. Yang, Z. et al. ATG4B (Autophagin-1) phosphorylation modulates autophagy. *J. Biol. Chem.* **290**, 26549–26561 (2015).
81. Ishikawa, H., Ma, Z. & Barber, G. N. STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. *Nature* **461**, 788–792 (2009).
82. Yamada, S. et al. Quantitative analysis of cytokine mRNA expression and protozoan DNA load in *Theileria parva*-infected cattle. *J. Vet. Med. Sci.* **71**, 49–54 (2009).
83. McElroy, A. K. & Nichol, S. T. Rift Valley fever virus inhibits a pro-inflammatory response in experimentally infected human monocyte derived macrophages and a pro-inflammatory cytokine response may be associated with patient survival during natural infection. *Virology* **422**, 6–12 (2012).
84. Smetko, A. et al. Trypanosomosis: potential driver of selection in African cattle. *Front. Genet.* **6**, 137 (2015).
85. Murray, M., Trail, J., Davis, C. & Black, S. Genetic resistance to African trypanosomiasis. *J. Infect. Dis.* **149**, 311–319 (1984).
86. Safran, M. et al. GeneCards version 3: the human gene integrator. *Database* **2010**, 1–16 (2010).
87. Pomerantz, J. L., Denny, E. M. & Baltimore, D. CARD11 mediates factor-specific activation of NF-κB by the T cell receptor complex. *EMBO J.* **21**, 5184–5194 (2002).
88. Hara, H. et al. The MAGUK family protein CARD11 is essential for lymphocyte activation. *Immunity* **18**, 763–775 (2003).
89. Noyes, H. et al. Genetic and expression analysis of cattle identifies candidate genes in pathways responding to *Trypanosoma congolense* infection. *Proc. Natl Acad. Sci. USA* **108**, 9304–9309 (2011).
90. Cecchi, G., Paone, M., Herrero, R. A., Vreysen, M. J. & Mattioli, R. C. Developing a continental atlas of the distribution and trypanosomal infection of tsetse flies (*Glossina* species). *Parasit. Vectors* **8**, 284 (2015).

91. Lemecha et al. Response of four indigenous cattle breeds to natural tsetse and trypanosomosis challenge in the Ghibe valley of Ethiopia. *Vet. Parasitol.* **141**, 165–176 (2006).
92. Naessens, J., Teale, A. & Sileghem, M. Identification of mechanisms of natural resistance to *African trypanosomiasis* in cattle. *Vet. Immunol. Immunopathol.* **87**, 187–194 (2002).
93. Hanotte, O. et al. Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. *Proc. Natl Acad. Sci. USA* **100**, 7443–7448 (2003).
94. Courtin, D. et al. Host genetics in African trypanosomiasis. *Infect. Genet. Evol.* **8**, 229–238 (2008).
95. Ciccia, A. et al. Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. *Mol. Cell* **25**, 331–343 (2007).
96. Cohn, M. A. et al. A UAF1-containing multisubunit protein complex regulates the Fanconi anemia pathway. *Mol. Cell* **28**, 786–797 (2007).
97. Kumar, L. et al. Leucine-rich repeat containing 8A (LRRC8A) is essential for T lymphocyte development and function. *J. Exp. Med.* **211**, 929–942 (2014).
98. Ball, E. A. et al. *IFNAR1* controls progression to cerebral malaria in children and $CD8^+$ T cell brain pathology in *Plasmodium berghei*–infected mice. *J. Immunol.* **190**, 5118–5127 (2013).
99. Makina, S. O. et al. Genome-wide scan for selection signatures in six cattle breeds in South Africa. *Genet. Sel. Evol.* **47**, 92 (2015).
100. Gautier, M. et al. A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics* **10**, 550 (2009).
101. Kahle, D. & Wickham, H. ggmap: spatial visualization with ggplot2. *R. J.* **5**, 144–161 (2013).
102. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

Methods

Ethics statement. Blood samples were collected during routine veterinary treatments with the logistical support and agreement of relevant agricultural institutions in each country: the International Trypanotolerance Center, The Gambia and the International Livestock Research Institute (ILRI), Kenya (N'Dama, Kenya Boran); the Ministry of Animal Resources, Sudan (Kenana and Butana); the Ol Pejeta Conservancy, Kenya (Ankole, African Buffalo); and the Ethiopian Ministry of Agriculture, Ethiopia (Afar, Arsi, Barka, Ethiopian Boran, Fogera, Goffa, Horro, Mursi, Ogaden and Sheko). No further ethics permissions were required for this study. For European and Asian taurine, all animal work was approved by the Institutional Animal Care and Use Committee of the National Institute of Animal Science in Korea under approval numbers 2012-C-005 (Holstein and Hanwoo) and NIAS-2014-093 (Angus and Jersey). All animals were handled in strict accordance with good animal practice.

Sequencing and variant calling. All sequenced samples ($n=116$) were prepared according to the Illumina protocols (TruSeq DNA Sample Prep Kit v.2 Support (FC121-2001)). Briefly, 1 µg of genomic DNA was fragmented using a Covaris Focused-Ultrasonicator, and repaired. An 'A' was ligated to the 3′ end of the fragments, followed by Illumina adapter ligation. The product was further size-selected for 400–500 bp, PCR-amplified and validated using the Agilent Bioanalyzer. Finally, the DNA was sequenced using the HiSeq2000 platform (Illumina) by Macrogen.

Our previously published data of 53 commercial taurine[31,103,104] and 48 African[31] cattle, as well as publicly available data of 10 African taurine, 50 European taurine, 34 American-Australian zebu and 22 Asian zebu[105,106], were used in this study in addition to the generated sequence data. We generated genotype data following the 1000 Bull Genomes Project Run 8 guideline (17 October 2019) (http://www.1000bullgenomes.com/). We first examined a per-base sequence quality for the raw sequence reads using the fastQC software v.0.11.8 (ref. [107]), and removed low-quality bases and artifact sequences using Trimommatic v.0.39 (ref. [108]). The high-quality sequence reads were mapped against the bovine reference genome (ARS-UCD1.2) using bwa mem v.0.7.17 (ref. [109]) with default parameters. We then used Samtools v.1.9 (ref. [110]) to sort bam files and create index files. For the mapped reads, potential PCR duplicates were identified using the 'MarkDuplicates' of Picard v.2.20.2 (http://broadinstitute.github.io/picard). The 'BaseRecalibrator' and 'PrintReads' of the genome analysis toolkit (GATK) v.3.8 (GATK)[111] were used to perform base quality score recalibration (BQSR). The known variants file (ARS1.2PlusY_BQSR_v3.vcf.gz) provided by the 1000 Bull Genomes Project was used for masking known sites for all individuals except the two African buffalos (AFB). The before/after BQSR reports were checked by running 'AnalyzeCovariates' to ensure that base quality scores were corrected as expected. For the two AFB samples, we performed an initial round of variant calling on unrecalibrated data. We then performed BQSR by feeding the variants obtained from the initial variant calling as known sites to BaseRecalibrator and finally checked the convergence of base quality improvement.

For the calling of the candidate SNPs from the bam files, we created GVCF files using 'HaplotypeCaller' in GATK with '-ERC GVCF' option. Individual GVCF files were merged by breeds using 'CombineGVCFs' in GATK. We called and selected candidate SNPs from these combined GVCF files using 'GenotypeGVCFs' and 'SelectVariants', respectively. To avoid possible false-positive calls, we used 'VariantFiltration' of GATK as recommended by GATK best practice: (1) SNP clusters with '-clusterSize 3' and '-clusterWindowSize 10' options; (2) SNPs with mean depth (for all individuals) $<1/3x$ and $>3x$ (x, overall mean sequencing depth across all SNP sites); (3) quality by depth, QD < 2; (4) phred-scaled variant quality score, QUAL < 30; (5) strand odds ratio, SOR > 3; (6) Fisher strand, FS > 60; (7) mapping quality, MQ < 40; (8) mapping quality rank sum test, MQRankSum < −12.5; and (9) read position rank sum test, ReadPosRankSum < −8 were filtered. We then filtered out nonbiallelic SNPs or SNPs with missing genotype rates >0.01. For the remaining SNPs, genotype refinement, imputation and phasing were simultaneously performed using BEAGLE 4.0 (r1399)[42], while excluding AFB individuals. After filtering out SNPs with minor allele frequency < 0.01, the remaining high-quality SNPs were annotated according to their positions using SnpEff v.4.3 (ref. [112]) and were used in the downstream analysis (Supplementary Tables 12 and 13).

To check the confidence of variant calls from the resequencing analysis, we additionally genotyped 69 cattle samples using the BovineSNP50 Genotyping BeadChip (Illumina). After filtering out SNPs based on GeneCall score < 0.7, common loci of SNP chip and DNA resequencing data were extracted and examined to assess concordance between genotypes from the two different platforms. We also incorporated the genotype data of 45 samples from our previously published study[21] into this assessment to check the reliability of our current pipeline.

Population differentiation and structure. For PCA, we used the Genome-wide Complex Trait Analysis (GCTA)[113] tool v.1.93.0 to estimate the eigenvalues and eigenvectors, incorporating genotype data from 331 individuals, excluding two African Buffalos. For admixture analysis, we performed LD-based pruning for the genotype data using PLINK v.1.9 (ref. [114]) with '-indep-pairwise 50 10 0.1' option as recommended by the developer. Admixture v.1.3.0 (ref. [44]) was run, increasing K from 1 to 10, where K is the assumed number of ancestral populations. The delta K method was used to choose the optimal K[115]. Genetic distances between cattle breeds were estimated with the F_{st} estimator as described by Weir and Cockerham[116] using PLINK v.1.9 (ref. [114]).

Phylogenetic reconstruction and genetic diversity. For the most significant candidate region in PBS analysis (BTA25: 40,052,001 ~ 40,102,000), we split the phased variant call format and generated reference-based consensus sequences for the 50-kb window using bcftools v.1.8 (http://samtools.github.io/bcftools/bcftools.html). A maximum-likelihood tree for the generated 666 haplotypes was reconstructed using IQ-TREE v.1.6.12 (ref. [117]) with the following options: Modelfinder Plus[118] -mset phyml, -cmin 4, -cmin 6 and -mset phyml. The best-fit model was determined to TVM + F + I + G4 (TVM, transversion model; F, empirical base frequencies; I + G4, invariable site plus discrete Gamma model with default four rate categories) under the Bayesian information criterion. The reconstructed trees were visualized using FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Individual heterozygosity (theta) based on Felsenstein's model of substitutions[119] was estimated using the ATLAS v.0.9 (ref. [120]) program, which takes into account the depth coverage and sequencing error of each locus. ROH were analyzed using VCFtools v.0.1.17 (ref. [102]), filtering out ROH segments of <50 kb.

Test for admixture and estimation of admixture proportion. We used the f and D statistics to test and quantify admixture in African cattle. We used our variant calls (~17.7 million SNPs) and the linearly interpolated recombination map derived from a large US Department of Agriculture dairy cattle pedigree[121]. All statistics were computed using ADMIXTOOLS v.5.1 (ref. [45]) with standard errors obtained from a block jackknife with 5-cM block size. Z-score was calculated on the standard errors. Three types of statistics were used in these analyses with the following notations. Note that EAT was replaced with Muturu, when we used Muturu as the surrogate population close to the source population in the three statistics.

$$f_3(X; \mathrm{EAT}, \mathrm{AAI})$$

The f_3 statistic was used to test for evidence that African cattle populations are derived from the admixture of two populations (EAT and AAI). X is the target African population of interest and EAT and AAI are populations close to the source populations. A significant negative f_3 statistic is considered evidence of historical admixture in the X population. In contrast, a positive value does not always mean there is no admixture, as a high degree of drift specific to the X population can mask the negative signal[45].

$$D(\mathrm{EAT}, \mathrm{X}; \mathrm{AAI}, \mathrm{AFB})$$

The D statistic was used to evaluate gene flow between different cattle populations. X is the target African population. If we ascertain AFB as an outgroup, a significant positive value indicates gene flow between EAT and AAI, while a significant negative value indicates gene flow between X and AAI.

$$\alpha = f_4(\mathrm{EATa}, \mathrm{AFB}; \mathrm{X}, \mathrm{AAI})/f_4(\mathrm{EATa}, \mathrm{AFB}; \mathrm{EATb}, \mathrm{AAI})$$

The f_4 ratio (α) quantifies the mixing proportion of an admixture event using the ratio of two f_4 statistics. We specified X as the target African population and AFB as an outgroup. EAT is randomly divided into two subgroups, EATa and EATb, to provide a pair of populations that are completely admixed. Under this specification, the α value is interpreted as the mixing proportion of EAT ancestry in the target African population X.

Estimation of admixture time. The time of admixture was first estimated with ALDER v.1.03 (ref. [70]), which provides an LD-based admixture time, using the default parameters with a minimum genetic distance (mindis) of 0.5 cM. For this, we used our variant calls (~17.7 million SNPs) and the linearly interpolated recombination map derived from a large US Department of Agriculture dairy cattle pedigree[121]. If a population is derived from an admixture between two source populations close to the reference populations, the pairwise LD in this population, weighted by the allele frequencies in the reference populations, shows an exponential decay as a function of the genetic distance. ALDER fits this decay and then infers the admixture time from the decay rate of the fitted curve.

We additionally used the modified version of ALDER (MALDER v.1.0 (ref. [46])), which allows multiple admixture events, to compare the agreements of single- and double-pulse admixture models with our data. For estimating admixture time using ALDER and MALDER, we performed two analyses for each African cattle population using two sets of reference populations (EAT and AAI, Muturu and AAI). The fitted curve of both the single- and double-pulse admixture models for Kenya Boran was visually checked using the 'nls' function implemented in R. For all of the admixture time estimations, standard errors were estimated from a leave-one-chromosome-out jackknifing.

In addition, we used GLOBETROTTER[50] on 14 African cattle populations (AFH) to estimate haplotype sharing-based admixture time. The

GLOBETROTTER method uses a coancestry curve, in which a measure of how often pairs of haplotypes separated by a genetic distance X come from each respective source population is plotted as a function of the genetic distance X (ref. [50]). Given a single admixture event, haplotypes inherited from each source population theoretically have an exponential size distribution, which leads to an exponential decay of the coancestry curve[50]. GLOBETROTTER fits this curve, allowing us to estimate the rate of the exponential decay, which is an estimate of the admixture time[50].

We specified the 14 African humped cattle populations and the other non-African cattle populations as target and donor populations, respectively. This specification indicates that target haplotypes are allowed to be copied from the donor haplotypes, not from the other target haplotypes. This is recommended when a similar admixture history is shared across the target populations[50].

To reduce the computational load, we performed LD-based pruning for the phased data using PLINK v.1.9 (ref. [114]) with '-indep-pairwise 50 10 0.1' option. The known genetic map[121] was interpolated against this reduced data, not allowing interpolation for gaps larger than 50 kb. Using the loci of the LD-pruned data, for which the recombination rates are available on the interpolated genetic map (~0.72 million SNPs), we performed GLOBETROTTER analysis as the following: (1) first, we ran ten rounds of the expectation–maximization iterations for BTA1, 2, 7 and 12 using ChromoPainter v.2 (ref. [122]) with '-in' and '-iM' switches, which result in estimates of the switch rate and global mutation rate parameters; (2) we then averaged the estimated parameters from (1) over all individuals and chromosomes, and used these as fixed estimated values (-n 514.030 -M 0.005127882) for the second running of ChromoPainter v.2 (ref. [122]) on all individuals; (3) we summed the 'chunk length' output from (2) across chromosomes using ChromoCombine, and obtained a single 'chunk length' output; (4) we also obtained ten painting samples for each target individual by running ChromoPainter v.2 (ref. [122]) with the fixed parameters averaged over all target individuals (-n 632.949 -M 0.006501492); (5) using the summed chunk length from (3) and ten painting samples from (4), we ran GLOBETROTTER with the 'prop.ind: 1' and 'null.ind: 1' options; and (6) to check the significance of admixture evidence, bootstrapping was performed with 100 replicates using 'prop.ind: 0' and 'bootstrap.date.ind: 1' options. In the bootstrap replicates, the proportion of inferred generations(s) that were between 1 and 400 was considered as evidence of detectable admixture[50].

Detection of selection signatures in African humped cattle. To detect ongoing selection signatures in AFH genomes ($n = 149$), we employed the *iHS*[123] implemented in HAPBIN v.1.3.0 (ref. [124]) using the default settings except for the '-f 0.01' option. For each SNP, the ancestral allele was defined as the allele fixed in the AFB outgroup. After computing *iHS* values for each SNP, they were grouped into 2% frequency bins and standardized. The proportion of SNPs with $|iHS| \geq 2$ was then calculated in each nonoverlapping window of 50 kb. In this step, windows with less than 10 SNPs were removed. We considered windows within the highest 1% of the empirical distribution for the proportion of SNPs with $|iHS| \geq 2$ as candidate regions with selection signal.

Local ancestry inference in African humped cattle. Using the genotype data phased in the *iHS* analysis, we performed local ancestry inference implemented in the LOTER package[51] to infer taurine–indicine ancestry along the AFH genomes. We specified 103 individuals of EAT and 56 individuals of AAI as reference populations, assuming that a haplotype of an admixed AFH consists of a mosaic of existing haplotypes from the two reference populations. Using LOTER, we first assigned each allele to taurine or indicine ancestry and calculated the frequency of assigned taurine or indicine ancestry within AFH. The resulting frequencies were then averaged over each nonoverlapping window of 50 kb. For the windows with the highest or lowest 0.5% of the empirical distribution for averaged taurine ancestry, we additionally filtered out windows with pairwise F_{st} values between reference populations less than genome-wide level (<0.2296) to reduce false positives from the admixture in each reference population. The remaining windows were considered as candidate regions with excess or deficiency of taurine ancestry. In light of the history of indicine cattle on the Indian subcontinent and in the Americas, it is possible that they contain some taurine background, although at low frequencies[125–127]. However, this will not result in false positives. Rather, it could lead to few false negatives since there are similar haplotypes to select in the LOTER algorithm, which may mask an excess of a particular ancestry.

Detection of selection signatures in African taurine cattle. To detect selection signatures in AFT after divergence from EAT, we employed the PBS developed by Yi et al.[64]. For each window with 50-kb size and 2-kb step, we calculated the PBS as follows:

$$T = -\log(1 - F_{st})$$

$$\mathrm{PBS} = \frac{T^{AE} + T^{AO} - T^{EO}}{2}$$

where T^{ij} represents estimated branch length between i and j populations based on pairwise Weir and Cockerham[116] F_{st} estimated by VCFtools v.0.1.17 (ref. [102]). A represents the target population (AFT), while E and O represent the control population (EAT) and the outgroup (AAI), respectively. A population PBS value conceptually represents the amount of allele frequency change at a given locus since its divergence from the other two populations. From this statistic, we intended to discover selection signatures in AFT cattle following their ancestral migration into the African continent.

Annotation and functional enrichment analysis. The annotation of the candidate regions was based on the ARS-UCD1.2 Gene Transfer Format file (.gtf) from Ensembl release 99 (ref. [128]). For functional enrichment analysis of a candidate gene set, a statistical overrepresentation test in PANTHER v.15.0 (ref. [129]) was used based on the GO-Slim Biological Process terms and REACTOME pathway[130] with default settings. A false discovery rate–adjusted P value of 0.05 was used as the threshold for statistical significance.

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

Data availability

The newly generated sequences for 114 African cattle and two African buffalo samples are available from the Sequence Read Archive (SRA) with the Bioproject accession number PRJNA574857. The publicly available sequences were downloaded from the SRA and China National GeneBank (CNGB) with the following project accession numbers; CNP0000189 (Achai, Bhagnari, Cholistani, Dajal, Dhanni, Gabrali, Hariana, Lohani, Red Sindhi, Sahiwal and Tharparkar), PRJNA318087 (Angus, Ankole, Jersey, Kenya Boran, Kenana, N'Dama and Ogaden), PRJNA514237 (Boskarin, Limia, Maremmana, Maronesa, Pajuna, Podolica and Sayaguesa), PRJNA324822 (Brahman), PRJNA343262 (Brahman, Gir, Hereford, Nelore and Simmental), PRJNA432125 (Brahman), PRJEB28185 (Eastern Finn and Western Finn), PRJNA210523 (Hanwoo), PRJNA379859 (Hariana, Sahiwal and Thaparkar), PRJNA210521 (Holstein), PRJNA386202 (Muturu) and PRJNA507259 (Nelore). The known variants file (ARS1.2PlusY_BQSR_v3.vcf.gz) for base quality score recalibration was provided by the 1000 Bull Genomes Project (http://www.1000bullgenomes.com/). The annotation of the candidate regions was based on the ARS-UCD1.2 Gene Transfer Format file (.gtf) from Ensembl release 99 (http://www.ensembl.org/). The PANTHER database (http://pantherdb.org/) was used for functional enrichment analysis of a candidate gene set.

References

103. Lee, H.-J. et al. Deciphering the genetic blueprint behind Holstein milk proteins and production. *Genome Biol. Evol.* **6**, 1366–1374 (2014).
104. Shin, D.-H. et al. Deleted copy number variation of Hanwoo and Holstein using next generation sequencing at the population level. *BMC Genomics* **15**, 240 (2014).
105. Heaton, M. P. et al. Using diverse US beef cattle genomes to identify missense mutations in *EPAS1*, a gene associated with pulmonary hypertension. *F1000Res.* **5**, 2003 (2016).
106. Taylor, J. F. et al. Lessons for livestock genomics from genome and transcriptome sequencing in cattle and other mammals. *Genet. Sel. Evol.* **48**, 59 (2016).
107. Andrews, S. FastQC: a quality control tool for high throughput sequence data. *Babraham Bioinformatics* http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).
108. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
109. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
110. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
111. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
112. 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).
113. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
114. 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).
115. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
116. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
117. Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

118. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587 (2017).
119. Felsenstein, J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**, 368–376 (1981).
120. Kousathanas, A. et al. Inferring heterozygosity from ancient and low coverage genomes. *Genetics* **205**, 317–332 (2017).
121. Ma, L. et al. Cattle sex-specific recombination and genetic control from a large pedigree analysis. *PLoS Genet.* **11**, e1005387 (2015).
122. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet.* **8**, e1002453 (2012).
123. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
124. Maclean, C. A., Chue Hong, N. P. & Prendergast, J. G. hapbin: an efficient program for performing haplotype-based scans for positive selection in large genomic datasets. *Mol. Biol. Evol.* **32**, 3027–3029 (2015).
125. Utsunomiya, Y. et al. Genomic clues of the evolutionary history of *Bos indicus* cattle. *Anim. Genet.* **50**, 557–568 (2019).
126. Koufariotis, L. et al. Sequencing the mosaic genome of Brahman cattle identifies historic and recent introgression including polled. *Sci. Rep.* **8**, 17761 (2018).
127. O'brien, A. M. P. et al. Low levels of taurine introgression in the current Brazilian Nelore and Gir indicine cattle populations. *Genet. Sel. Evol.* **47**, 31 (2015).
128. Zerbino, D. R. et al. Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2017).
129. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).
130. Croft, D. et al. The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42**, D472–D477 (2014).

Acknowledgements

This work was supported by a grant from the Next-Generation BioGreen 21 Program and Post-Genome Project (Project Nos. PJ01323701 and PJ01040601), Rural Development Administration, Republic of Korea. Sampling of cattle populations was supported by the CGIAR Livestock and Fish CRP (Uganda and Ethiopia), the University of Khartoum (Sudan) and the National Biotechnology Development Agency (NABDA) (Nigeria). The following institutions and their personnel provided help for the sampling of the African cattle: the ILRI Kapiti Ranch; the Ministry of Animal Resources, Fisheries and Range (Sudan); the Ol Pejeta Conservancy (Kenya); the Institute of Biodiversity (Ethiopia); and the Directors of Veterinary Services and the cattle keepers from Ethiopia, Kenya, Uganda and Sudan. The ILRI livestock genomics program is supported by the CGIAR Research Program on Livestock (CRP Livestock), which is supported by contributors to the CGIAR Trust Fund (http://www.cgiar.org/about-us/our-funders/). This research was funded in part by the Bill & Melinda Gates Foundation and with UK aid from the UK Foreign, Commonwealth and Development Office (Grant Agreement OPP1127286) under the auspices of the Centre for Tropical Livestock Genetics and Health (CTLGH), established jointly by the University of Edinburgh, SRUC (Scotland's Rural College) and the International Livestock Research Institute. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation or the UK Government. We thank the reviewers for their critical and constructive comments on the manuscript, and D. Gifford-Gonzalez (University of California, Santa Cruz, CA, USA) for a critical reading of the manuscript in light of the current knowledge on the archeology and history of African pastoralism.

Author contributions

K.K. and O.H. devised the main conceptual ideas. O.H. and H.K. managed the project. D.L., S.C., S.J.O., H.-K.L., O.A.M., T.D., S.K., O.H. and H.K. conceived of and designed all of the described experiments. O.A.M., T.D., B.S., G.M.T. and A.T. contributed to sample collection and laboratory work. K.K., T.K., D.Y., J. Jang, S.S., S.L., J. Jung and H.J. analyzed the data. K.K., C.J., J.K. and O.H. drafted the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

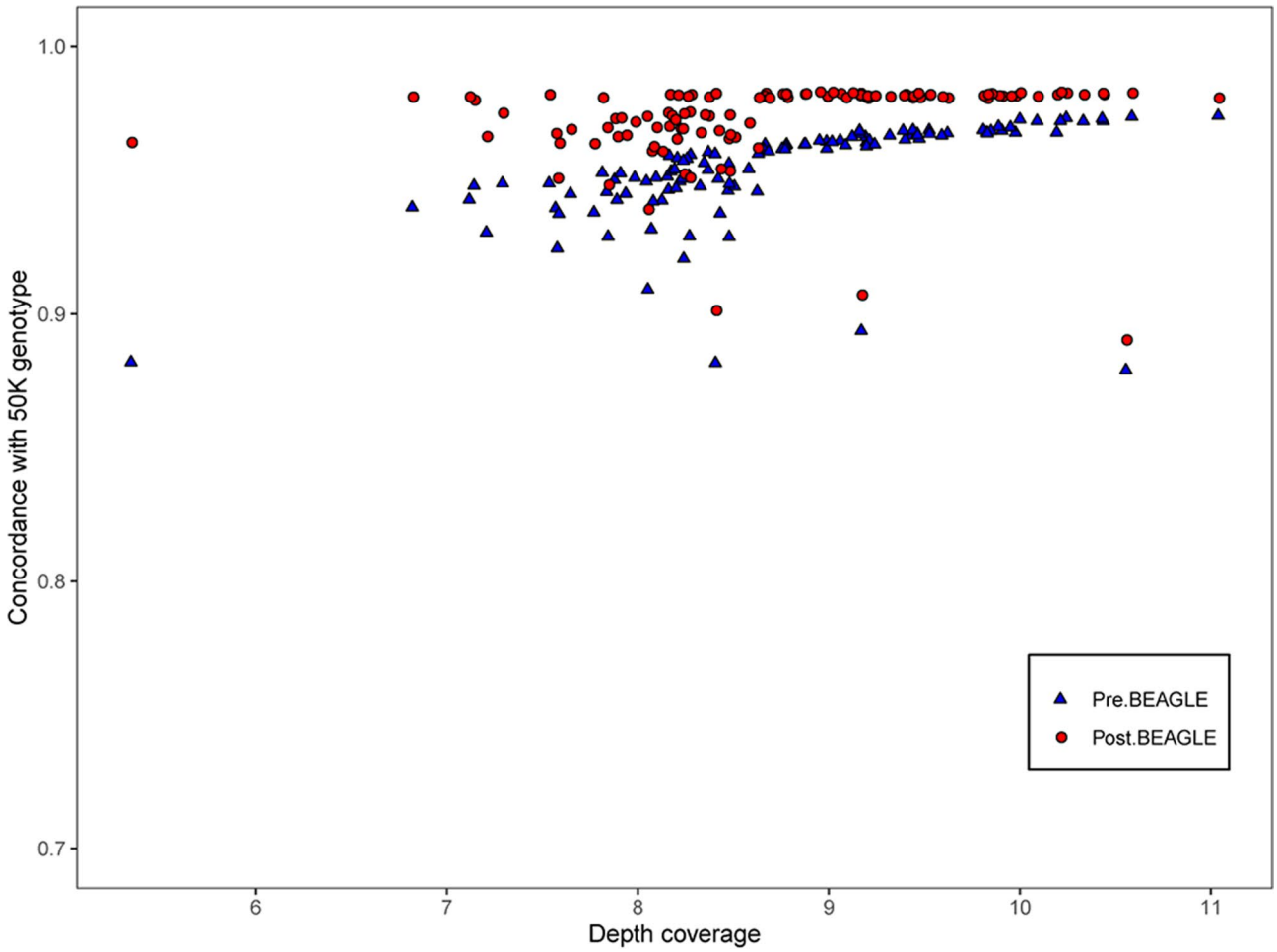
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-020-0694-2.

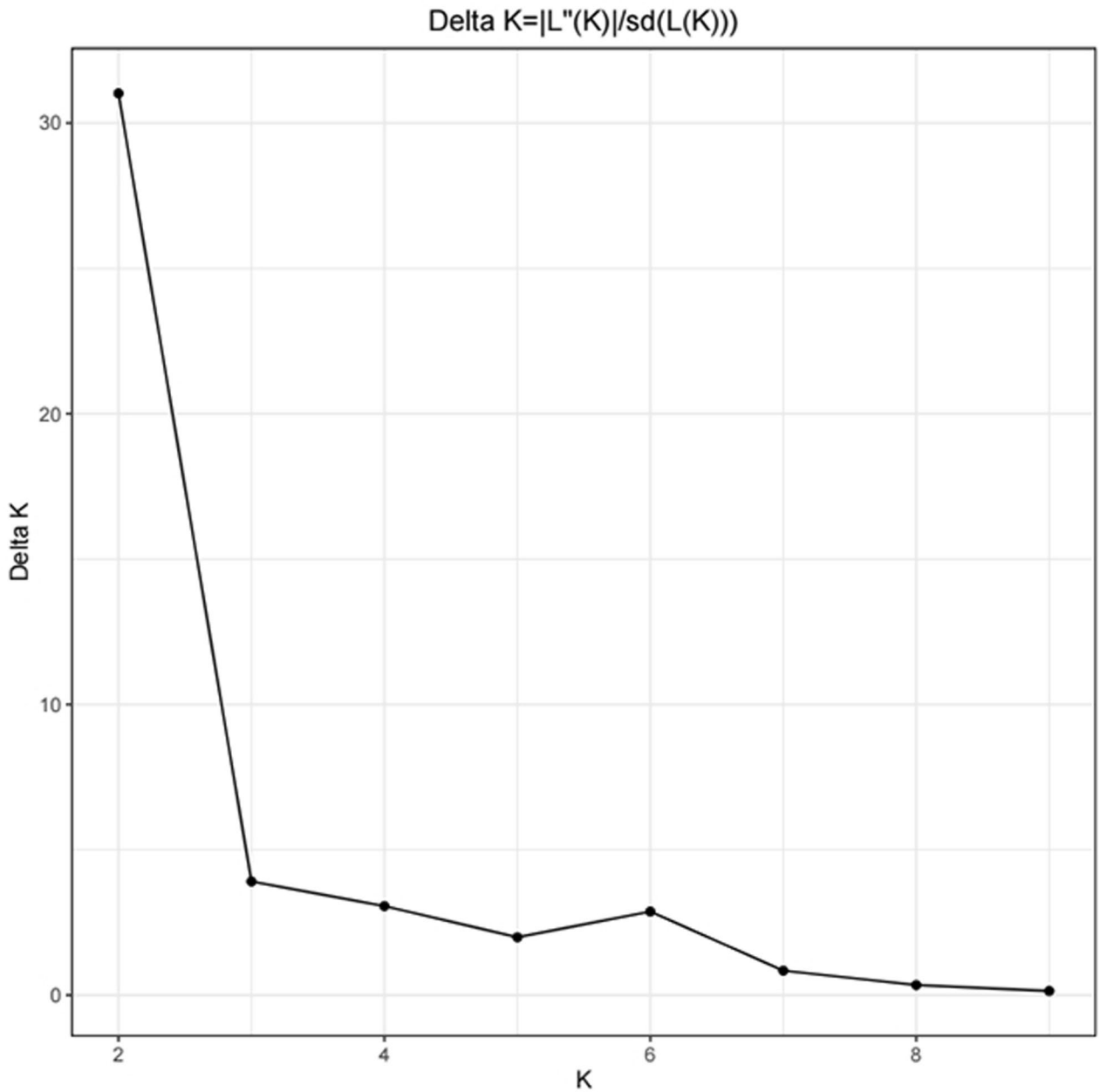
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-020-0694-2.

Correspondence and requests for materials should be addressed to O.H. or H.K.

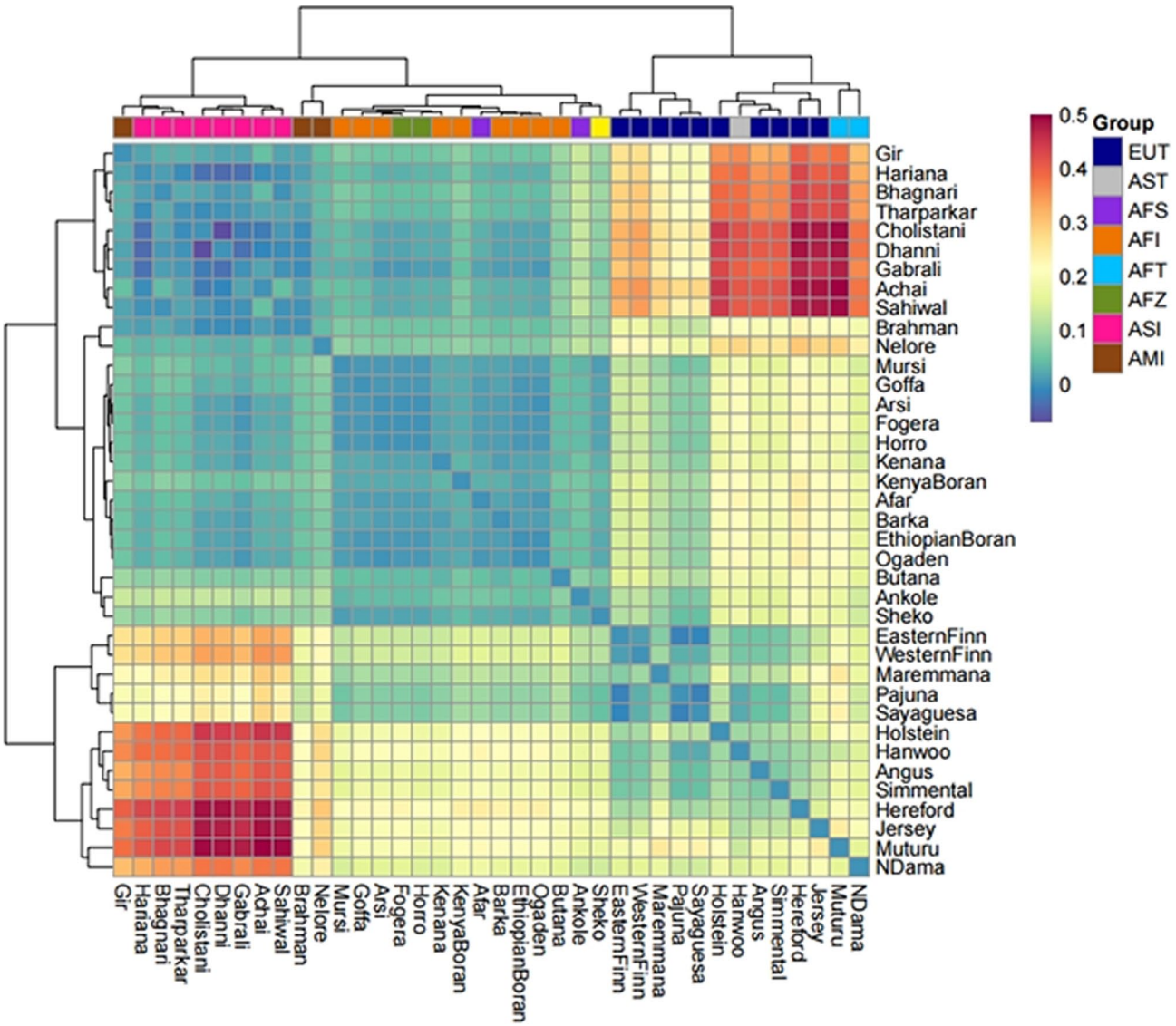
Reprints and permissions information is available at www.nature.com/reprints.



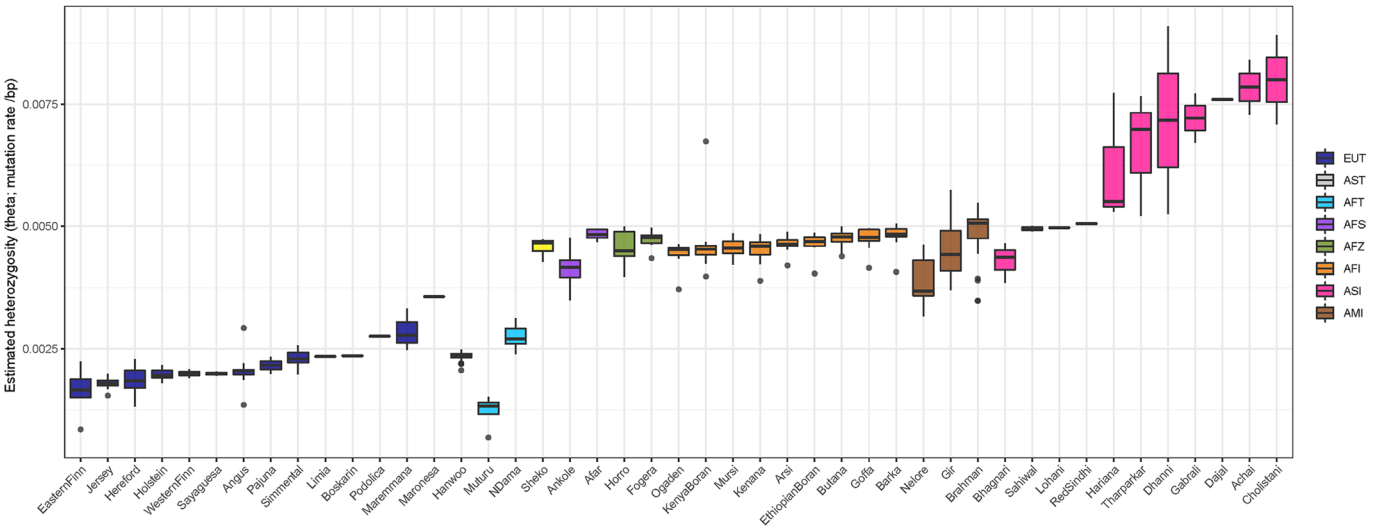
Extended Data Fig. 1 | Improvement in genotype concordance after genotype refinement using BEAGLE as a function of depth coverage. The *y*-axis shows the concordance between genotypes called from sequencing data compared to genotypes obtained using the BovineSNP50 Genotyping BeadChip.



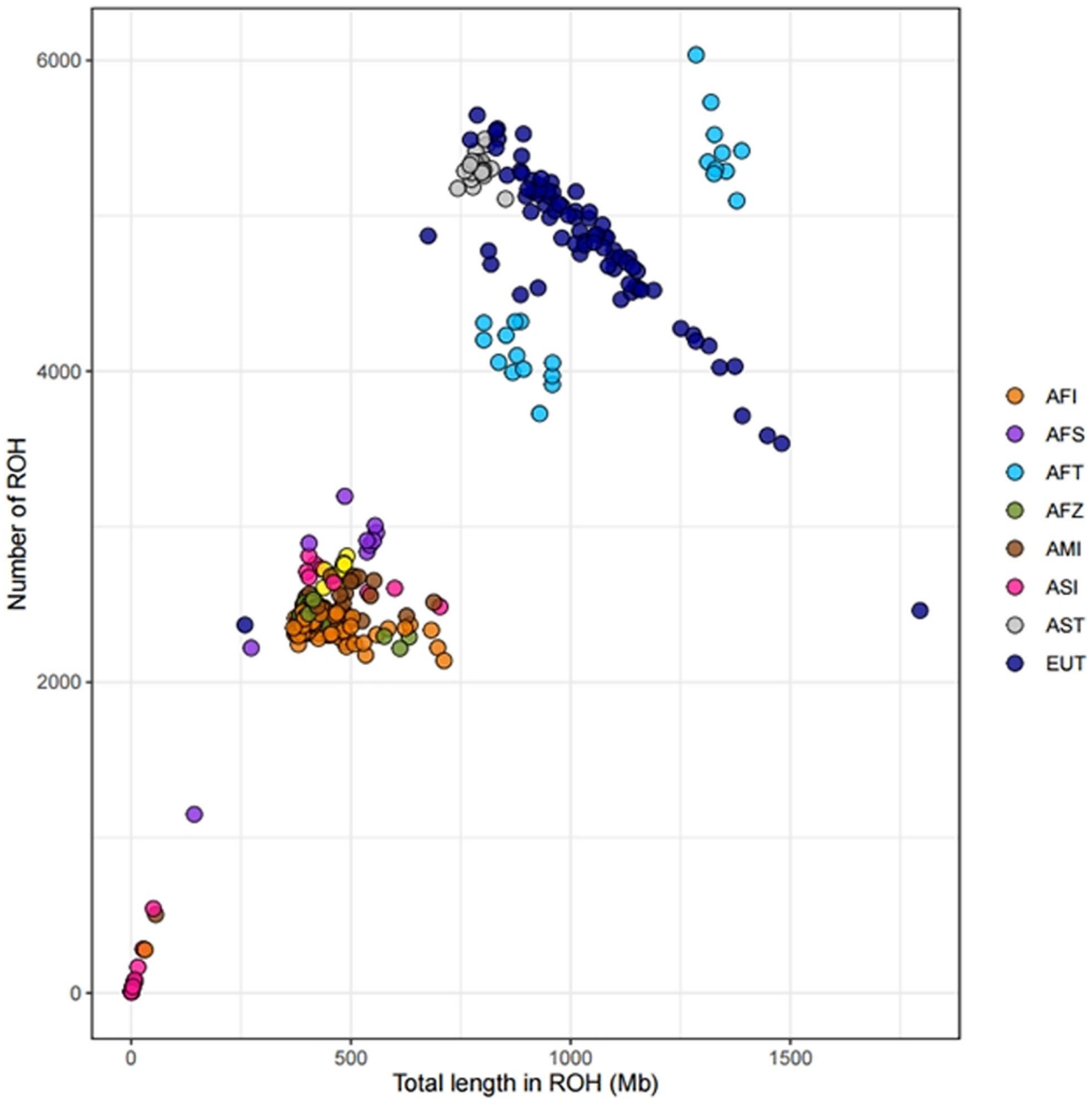
Extended Data Fig. 2 | Delta K of cluster number K in genetic clustering analysis using ADMIXTURE. A subset of ~1.6 million SNPs (linkage disequilibrium (LD)-based pruning using PLINK v1.9 with '-indep-pairwise 50 10 0.1' option) was used for K from 1 to 10. The delta K analysis suggests K = 2 as the most likely number of distinct genetic ancestries among the 10 Ks (delta K = 31.02).



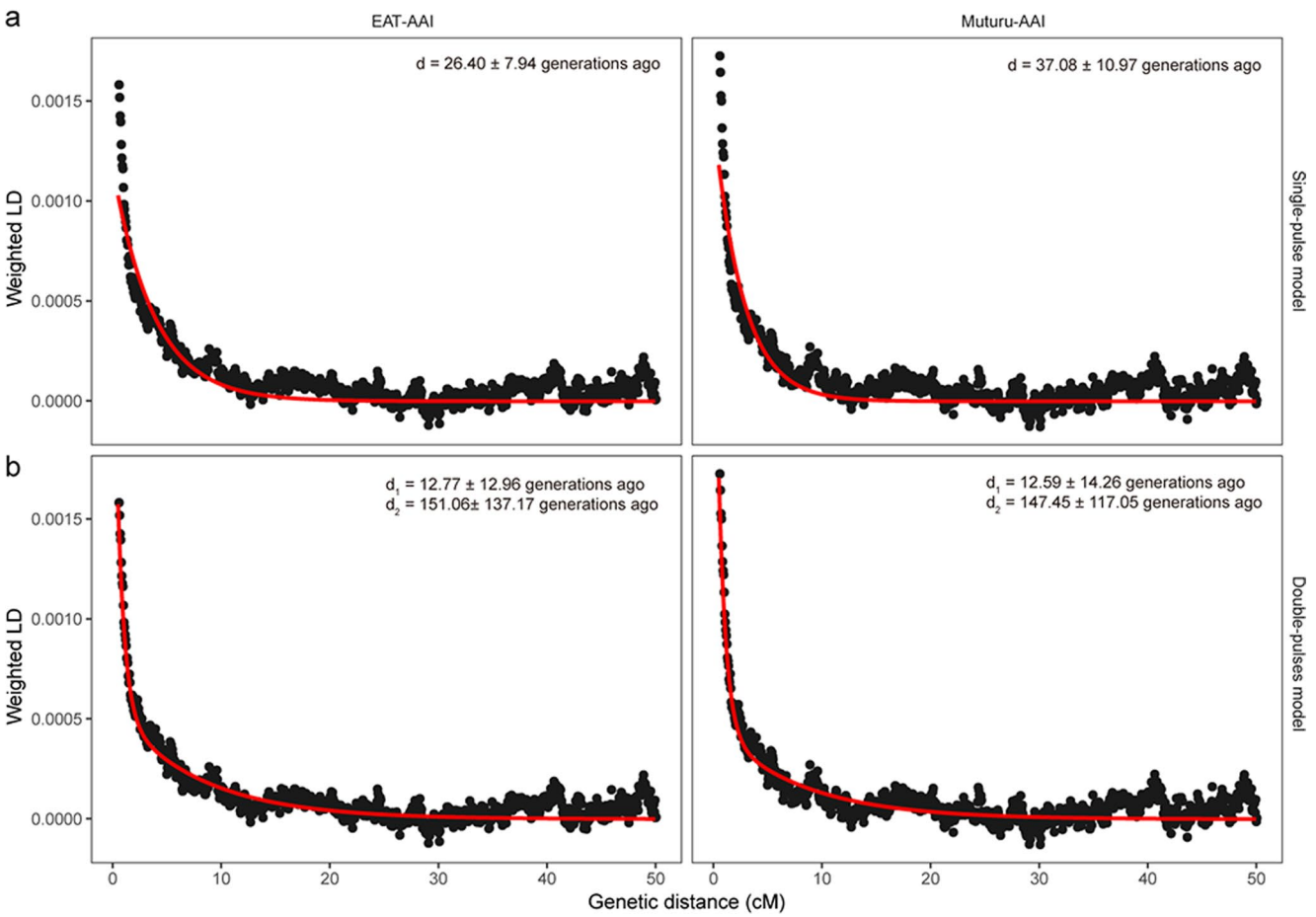
Extended Data Fig. 3 | Mean pairwise Fst values between cattle breeds represented by more than one animal. Sheko is indicated as yellow.



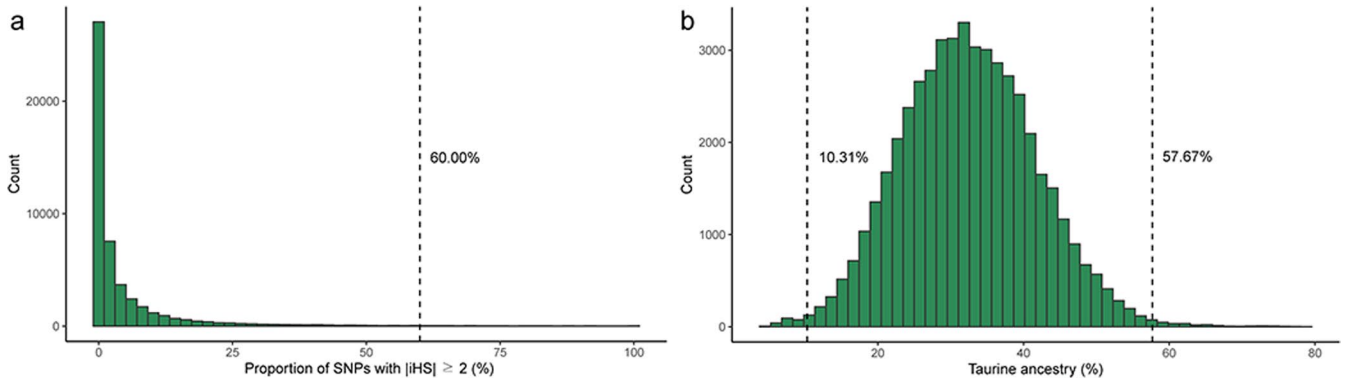
Extended Data Fig. 4 | Estimated heterozygosity of cattle breeds. The lower and upper bounds of box correspond to the first and third quartiles (the 25th and 75th percentiles), respectively. The horizontal line in the box represents the median value. The upper and lower whisker extend from the bounds to the largest and lowest value no further than 1.5 * interquartile range (IQR), respectively. The number of biologically independent animals used in this analysis for each breed is as follows: Achai (2), Afar (9), Angus (10), Ankole (10), Arsi (10), Barka (9), Bhagnari (3), Boskarin (1), Brahman (20), Butana (20), Cholistani (2), Dajal (1), Dhanni (2), Eastern Finn (5), Ethiopian Boran (10), Fogera (9), Gabrali (2), Gir (4), Goffa (10), Hanwoo (23), Hariana (3), Hereford (18), Holstein (10), Horro (11), Jersey (10), Kenya Boran (10), Kenana (13), Limia (1), Lohani (1), Maremmana (3), Maronesa (1), Mursi (10), Muturu (10), N'Dama (13), Nelore (10), Ogaden (9), Pajuna (2), Poldolica (1), Red Sindhi (1), Sahiwal (2), Sayaguesa (2), Sheko (9), Simmental (11), Tharparkar(3) and Wetern Finn (5). Sheko is indicated as yellow.



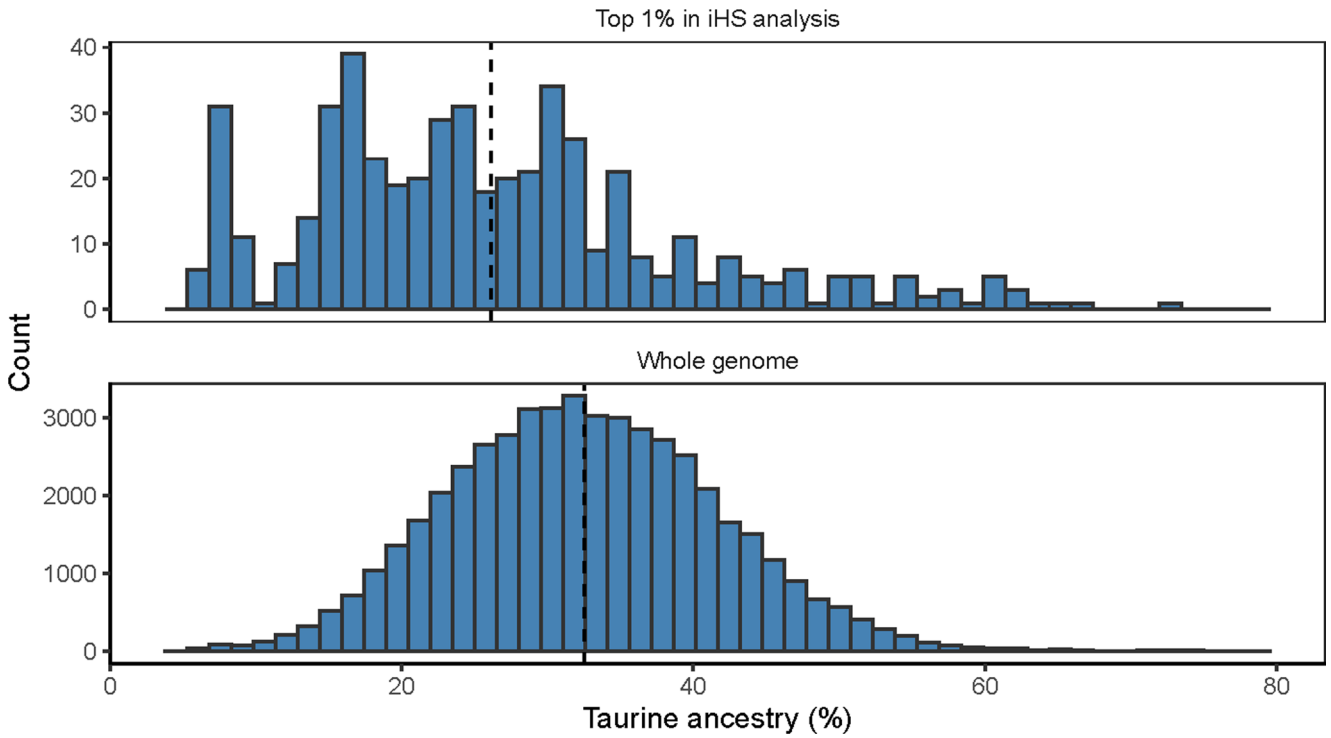
Extended Data Fig. 5 | Runs of homozygosity patterns of cattle breeds. Sheko is indicated as yellow.



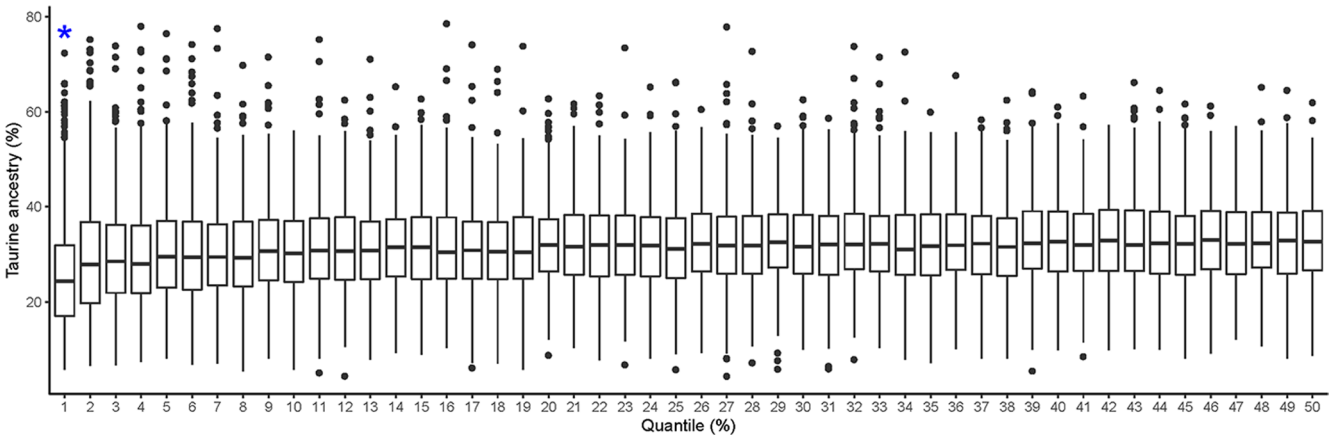
Extended Data Fig. 6 | Weighted LD decay in the Kenya Boran breed before and after fitted with a double-pulse admixture model. The red curve shows the exponential fit to the data. **a**, Weighted LD fitted by a single-pulse admixture model, when using EAT and Muturu as a reference population separately. **b**, Weighted LD fitted by a double-pulse admixture model, when using EAT and Muturu as a reference population separately.



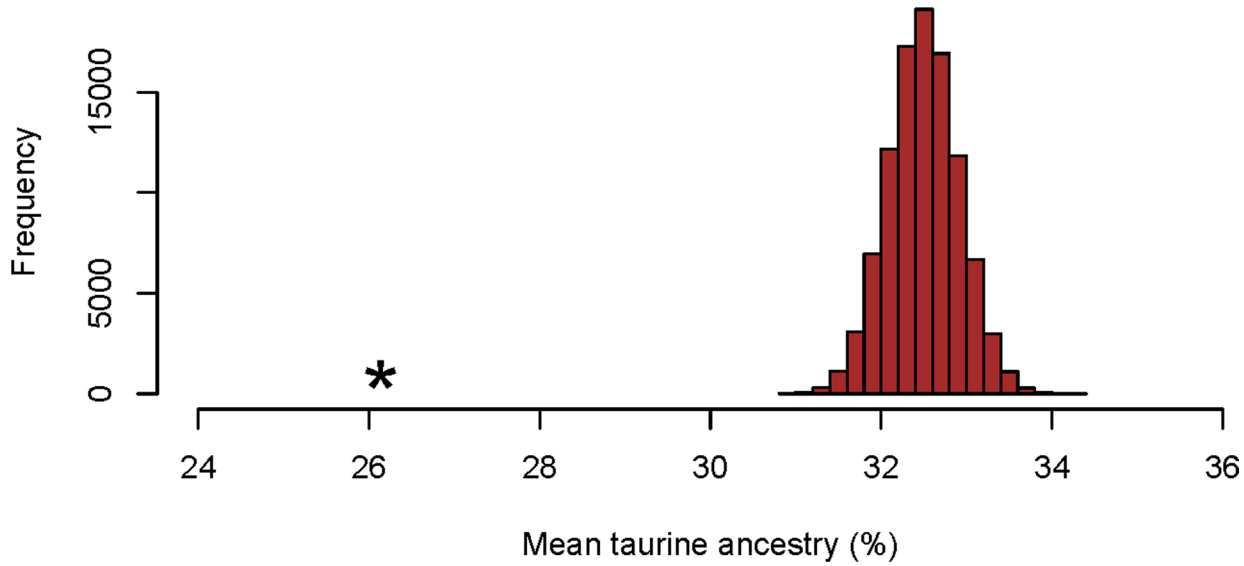
Extended Data Fig. 7 | Distribution of proportions of SNPs with |iHS| ≥ 2 and taurine ancestry in each 50-kb window. a, Distribution of proportions of SNPs with |iHS| ≥ 2. **b**, Distribution of taurine ancestry. The windows with SNPs less than 10 were removed. Dashed lines indicate the highest 1% for **a**, and highest or lowest 0.5% in **b**.



Extended Data Fig. 8 | Distribution of taurine ancestry in the candidate regions (the highest 1% for proportion of SNPs with |iHS| ≥ 2), and whole genome windows. Dashed lines indicate mean (top 1% in iHS analysis: 26.14%, and whole genome: 32.49%).



Extended Data Fig. 9 | Distribution of taurine ancestry according to quantiles of proportions of SNPs with |iHS| ≥ 2 in each 50-kb window. The lower and upper bounds of box correspond to the first and third quartiles (the 25th and 75th percentiles), respectively. The horizontal line in the box represents the median value. The upper and lower whisker extend from the bounds to the largest and lowest value no further than 1.5 * interquartile range (IQR), respectively. Asterisk indicates the highest 1% with proportions of SNPs with |iHS| ≥ 2. $n = 149$ (African humped cattle) biologically independent animals were used in this analysis.



Extended Data Fig. 10 | Distribution of average taurine ancestry generated by resampling random windows (same number of windows as the candidate) for 0.1 million times. Asterisk indicates average taurine ancestry of the candidate windows from iHS analysis.

nature research

Corresponding author(s): Olivier Hanotte and Heebal Kim

Last updated by author(s): Jul 30, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used in data collection.
Data analysis	Raw sequence quality check: FastQC v0.11.8 Raw sequence trimming: Trimommatic v0.39 Read mapping: bwa mem v0.7.17 Mark duplicates: Picard v2.20.2 Sort and indexing bam files: Samtools v1.9 Base quality score recalibration, variant calling and filtering: Genome analysis toolkit (GATK) v3.8 Haplotype phasing: BEAGLE v4.0 (r1399) SNP annotation: SnpEff v4.3 Principal component analysis: Genome-wide complex trait analysis (GCTA) v1.93.0 Admixture analysis: Admixture v1.3.0 Fst calculation and LD-based pruning: PLINK v1.9 Maximum likelihood tree reconstruction: IQ-TREE v1.6.12 Tree visualization: FigTree v1.4.4 Heterozygosity estimation: ATLAS v0.9 Runs of homozygosity and Population Branch Statistics (PBS) analysis: VCFtools v0.1.17 f3, D, f4 ratio statistics calculation: ADMIXTOOLS v5.1 Chromosome painting: ChromoPainter v2 Admixture time estimation: ALDER v1.03, MALDER v1.0, GLOBETROTTER v1.0 integrated haplotype score (iHS) analysis: HAPBIN v1.3.0 Local ancestry inference: LOTER v1

Gene set enrichment analysis: PANTHER v15.0
Reference-based consensus sequence generation: bcftools v1.8

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The newly generated sequences for 114 African cattle and 2 African buffalo samples are available from Sequence read archive (SRA) with the Bioproject accession number PRJNA574857. The publicly available sequences were downloaded from SRA and China National GeneBank (CNGB) with following project accession numbers;
CNP0000189 (Achai, Bhagnari, Cholistani, Dajal, Dhanni, Gabrali, Hariana, Lohani, Red Sindhi, Sahiwal, and Tharparkar), PRJNA318087 (Angus, Ankole, Jersey, Kenya Boran, Kenana, N'Dama, and Ogaden), PRJNA514237 (Boskarin, Limia, Maremmana, Maronesa, Pajuna, Podolica, and Sayaguesa), PRJNA324822 (Brahman), PRJNA343262 (Brahman, Gir, Hereford, Nelore, and Simmental), PRJNA432125 (Brahman), PRJEB28185 (Eastern Finn, and Western Finn), PRJNA210523 (Hanwoo), PRJNA379859 (Hariana, Sahiwal, and Thaparkar), PRJNA210521 (Holstein), PRJNA386202 (Muturu), and PRJNA507259 (Nelore)
The known variants file (ARS1.2PlusY_BQSR_v3.vcf.gz) for base quality score recalibration is provided by the 1000 bull genomes project (http://www.1000bullgenomes.com/). The annotation of the candidate regions was based on the ARS-UCD1.2 Gene Transfer Format file (.gtf) from Ensembl release 99 (http://www.ensembl.org/). PANTHER database (http://pantherdb.org/) was used for functional enrichment analysis of a candidate gene set.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	A minimum sample size of 9 animals per breed or 18 autosomal chromosomes was chosen under the expectation that for a biallelic locus of equal frequency for each allele (50%), random fixation of one allele in a breed will have a sufficiently low probability (0.5 to the power of 18) to indicate selection signal. There are no further consideration in determining sample size than the one described. Note that we collected samples, excluding 1st or 2nd degree of relatives based on the pedigree or farmer interview.
Data exclusions	We did not exclude any data.
Replication	To assess the confidence of SNPs identification, we performed SNP genotyping for a subset of whole samples (n = 114), all of which were successful.
Randomization	No randomization was required as no analyses involved in selection of subset of animals or informative SNPs.
Blinding	Blinding was not required, as no human participant was involved in our experiment or analyses.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	The study did not involve laboratory animals.
Wild animals	This specific study did not involve the capture of any wild animals.
Field-collected samples	International Livestock Research Institute (ILRI) livestock biorepository provided the cattle samples. The samples of each breed were collected at a single time point within the geographic area of the breed or in the research station as follows; Breed (sample location, country, latitude, longitude, altitude (m)) Afar (Melka Werer, Ethiopia, 09.78217, 040.44031, 804), Arsi (Bekoji, Ethiopia, 07.58051, 039.27923, 3300), Barka (Humera, Ethiopia, 14.09885, 037.21861, 895), Butana (Tamboul;Atabara, Sudan, N/A, N/A, N/A), Ethiopian Boran (Dida Tuyura ranch, Ethiopia, 04.55188, 38.10037, 1569), Fogera (Andassa Livestock Research Centre, Ethiopia, 11.30020, 37.28463, 1735), Goffa (Gamo Goffa, Ethiopia, 06.21423, 37.07276, 1100), Horro (Bako, Ethiopia, 09.07076, 37.03445, 1722), Kenana (Rabak, Sudan, N/A, N/A, N/A), Mursi (Jinka, Ethiopia, 05.47000, 36.34012, 1405), and Sheko (Masha, Ethiopia, 07.64703, 035.50526, 2240). From N'Dama cows in The Gambia, embryos were collected in 1983, with the assistance of the International Trypanotolerance Center (Banjul, The Gambia). They were implanted into recipient cows, and five N'Dama males and five N'Dama females were born in 1984. The N'Dama animals in this study are the offsprings of them. As these samples were collected for different purposes, they were collected across several years. Animals at the village level were under traditional management system of each community holding them. Animals at the reserach station were kept indoor at night or in grazing areas at daytime.
Ethics oversight	Blood samples were collected during routine veterinary treatments with the logistical support and agreement of relevant agricultural institutions in each country: International Trypanotolerance Center, The Gambia and International Livestock Research Institute (ILRI – Kenya) (N'Dama, Kenya Boran); Ministry of Animal Resources, Sudan (Kenana, and Butana); Ol Pejeta Conservancy, Kenya (Ankole, African Buffalo); Ethiopian Ministry of Agriculture, Ethiopia (Afar, Arsi, Barka, Ethiopian Boran, Fogera, Goffa, Horro, Mursi, Ogaden, and Sheko). No further ethics permissions were required for this study. For European and Asian taurine, all animal works were approved by the Institutional Animal Care and Use Committee of the National Institute of Animal Science in Korea under approval numbers 2012-C-005 (Holstein and Hanwoo) and NIAS-2014-093 (Angus and Jersey). All animals were handled in strict accordance with good animal practice.

Note that full information on the approval of the study protocol must also be provided in the manuscript.