# Genomic signatures of adaptive introgression and environmental adaptation in the Sheko cattle of southwest Ethiopia

**Hussain Bahbahani[1]\*, Arwa Afana[1], David Wragg[2]**

**1** Department of Biological Sciences, Faculty of Science, Kuwait University, Kuwait City, Kuwait, **2** Centre for Tropical Livestock Genetics and Health, The Roslin Institute, Edinburgh, United Kingdom

\* hussain.bahbahani@ku.edu.kw, hussain.bahbahani@gmail.com

## Abstract

Although classified as an African taurine breed, the genomes of Sheko cattle are an admixture of Asian zebu and African taurine ancestries. They populate the humid Bench Maji zone in Sheko and Bench districts in the south-western part of Ethiopia and are considered as a trypanotolerant breed with high potential for dairy production. Here, we investigate the genome of Sheko cattle for candidate signatures of adaptive introgression and positive selection using medium density genome-wide SNP data. Following locus-ancestry deviation analysis, 15 and 72 genome regions show substantial excess and deficiency in Asian zebu ancestry, respectively. Nine and 23 regions show candidate signatures of positive selection following extended haplotype homozygosity (EHH)-based analyses (*iHS* and *Rsb*), respectively. The results support natural selection before admixture for one *iHS*, one *Rsb* and three zebu ancestry-deficient regions. Genes and/or QTL associated with bovine immunity, fertility, heat tolerance, trypanotolerance and lactation are present within candidate selected regions. The identification of candidate regions under selection in Sheko cattle warrants further investigation of a larger sample size using full genome sequence data to better characterise the underlying haplotypes. The results can then support informative genomic breeding programmes to sustainably enhance livestock productivity in East African trypanosomosis infested areas.

## Introduction

The history of cattle in Africa began with the migration of humpless *Bos taurus taurus* (taurine) from their center of domestication in the Near East to the African continent through Egypt about 5000 years BC [1]. It was followed by the introduction of *Bos taurus indicus* (indicine or zebu) from their center(s) of domestication on the Indian subcontinent [2] around 2000 years BC, with further zebu arriving around 700 years AD following Arabs trading along the East coast of Africa, and the onset of the Swahili civilization [3].

Given the sole presence of taurine mitochondrial DNA haplogroups in African cattle [4, 5], together with zebu-specific *Y* chromosome alleles [6], a male-mediated pattern of zebu

introgression to the continent is the favored hypothesis [6]. Following the African rinderpest epidemic at the end of the 19th century, which led to a massive eradication of susceptible African taurine cattle, dispersal of the more resistant zebu ancestry was accelerated in the western and southern parts of the continent [3, 7].

Presently there are more than 150 recognized African cattle breeds or populations, classified as either taurine, zebu, sanga (an ancient stabilized taurine x zebu crossbreed), or a sanga x zebu crossbreed called zenga [8]. Genetically, most of the African cattle are admixed populations of zebu x taurine ancestries with a gradient of indicine ancestry peaking amongst the East African breeds and declining westward and southward to reach its minimum level in West African cattle [3, 7]. Several African trypanolerant taurine cattle with little or no zebu ancestry still populate the highly tsetse fly (the vector of trypanosomosis) infested zones of West Africa (e.g. N'Dama in Guinea and Muturu in Nigeria) [9]. A possible ancient introgression of the extinct African auroch *Bos primigenius opisthonomus* within African cattle has been suggested [7] however this requires further investigation.

Sheko cattle are indigenous to East Africa inhabiting the humid Bench Maji zone mainly in Sheko and Bench districts at the south-western part of Ethiopia [10, 11]. They were originally classified as taurine, while recent genetic analyses indicates they are more of a sanga type with African taurine and Asian zebu genetic ancestry proportions of 0.3 ± 0.014 and 0.7 ± 0.014, respectively [12]. The presence of a small cervicothoracic hump in them alludes to their zebu ancestry [11]. Sheko cattle are adapted to these highly tsetse-infested areas and are considered trypanotolerant [11, 13, 14]. They also have good potential as dairy cattle for Africa, having large teats and the ability to yield on average 2.79 ± 0.06 liters of milk daily and 850.6 ± 24.16 liters per lactation period, which is 307.69 ± 6.13 days, depending on the on-farm management practice [11, 15].

Previous studies of East African shorthorn zebu [16, 17], Butana and Kenana zebu from Sudan [18], and taurine and zebu cattle breeds from the western and eastern parts of the African continent [19] have identified signatures of positive selection in genes and quantitative trait loci (QTL) associated with adaptive traits. Many of the genes and QTL identified were found to be involved in biological pathways, such as bovine immunity, reproduction, heat tolerance and coat color. These studies applied genome-wide analyses of genotype data generated using the Illumina BovineSNP50 Beadchip [16], the Illumina BovineHD BeadChip [17, 18], as well as full genome sequence data [17, 19]. In commercial cattle breeds genes associated with milk yield and composition, muscle development and coat color have also been identified to be under positive selection [20–22].

In admixed populations, large deviations in genomic local ancestry relative to the average genome-wide admixture level may represent possible adaptive introgression [22–24], particularly if these regions are of large size and/or overlap with candidate footprints of positive selection. In humans this approach has been previously used to define positively selected genomic regions in an admixed Puerto Rican population with local ancestry deviation in the human leukocyte antigen regions on chromosomes 6, 8 and 11 [23]. A later study on a population of African–American descent identified genomic regions with significant excess of African ancestry in genes linked to the onset of diabetes, pancreatic and lung cancer [24].

Analysis of the dairy/beef dual purpose Simmental x Red Holstein-Frisian admixed Swiss Fleckvieh cattle has revealed recent responses to selection using medium density genome-wide SNP data [22]. Two regions on BTA 13 and 18 showed significant local ancestry deviation towards Simmental ancestry. These regions carry genes associated with bovine fertility (*NKD1* and *NOD2*) and the *FTO* gene, which has a pleiotropic effect being involved in milk composition and fertility.

In this study, we employ genotype data generated using Illumina's BovineSNP50 BeadChip to assess whether or not the genomic landscape of Sheko cattle has been under selection following introgression with zebu. We identify genomic regions in Sheko cattle with substantial locus-ancestry deviation and unusual extended haplotype homozygosity (EHH) and discriminate between the pre- and post-admixture selection pressures on the genome.

## Materials and methods

### SNP genotyping and quality control

Genome-wide SNP genotype data from the Illumina BovineSNP50 BeadChip version 1 [25] for 20 East African taurine Sheko, 25 West African taurine N'Dama and 21 Asian zebu Nelore cattle were obtained from the Bovine HapMap consortium [26]. Quality control analyses were carried out on 54,334 autosomal SNPs mapped to the UMD3.1 bovine reference genome using the *check.marker* function of the GenABEL package [27] for R software version 3.2.2 [28]. In total, 19,417 SNPs with minor allele frequency less than 0.05 and 6,886 SNPs with call rate less than 0.95 were removed. Among these, 5,766 SNPs failed both criteria, leaving 33,797 SNPs for downstream analyses. None of the samples had a SNP call rate < 0.95 or identity-by-state (IBS) > 0.95.

### Locus-ancestry deviation analysis

The Asian zebu and African taurine ancestry proportions were estimated in 1 Mb sliding genomic windows using the PCAdmix software version 1.0 [29]. fastPHASE software version 1.4 [30] was used to phase the genotyped SNPs into the corresponding haplotypes using K10 and T10 criteria. Population label information was provided to estimate the phased haplotype background. The Asian zebu ancestry proportion of each genomic window was estimated as the proportion of zebu haplotypes carried by the Sheko samples in that window. Windows deviating by two standard deviations (SD) from the mean zebu ancestry of all the genomic windows were considered as candidate regions with substantial excess/deficiency in Asian zebu ancestry.

### Extended haplotypes homozygosity (EHH)-derived statistics (Rsb and iHS)

Intra-population *iHS* [31] and inter-population *Rsb* [32] analyses were conducted using the rehh package [33] for R software to define candidate genomic regions with signatures of positive selection. The *iHS* analysis was carried out using genotyped SNPs with intra-population minor allele frequency $\geq$ 0.05. The *iHS* statistic is calculated by first defining the integral of the observed decay of EHH against the physical genomic position of the SNP, as one moves away from a core SNP for both the reference and alternative alleles until it reaches an arbitrary value of 0.05. These integrals, which are summed over both directions from the core SNP, are called *iHH_{Ref}* and *iHH_{Alt}* for the reference and alternative alleles, respectively. The natural log of the ratio (*iHH_{Ref}/iHH_{Alt}*) is standardized to generate an *iHS* value for each SNP with a mean 0 and variance 1, as described in Voight, Kudaravalli (31). As the standardized *iHS* values are normally distributed "Panel A in S1 Fig" and signatures of selection on both the reference and alternative alleles are equally important, a two-tailed Z test was applied to identify statistically significant SNPs. Two-sided *P*-values were derived as $-\log_{10}(1-2|\Phi(iHS)-0.5|)$, where $\Phi(iHS)$ represents the Gaussian cumulative distribution function. Inter-population *Rsb* analyses [32] were conducted separately between Sheko cattle and African taurine (N'Dama) and Asian zebu (Nelore) cattle. In the *Rsb* analysis, the EHH for the two alleles of a SNP was averaged and weighted by their squared allele frequencies, which provided the site-specific EHH

(EHHS). As with EHH, the observed decay of EHHS for each core SNP was integrated and summed over both directions in both populations (iES). An *Rsb* value for each SNP was obtained by standardizing the natural log ratio between the *iES* of Sheko population ($iES_{Sheko}$) with iES of the second reference population ($iES_{Reference}$), as described in Tang, Thornton (32). As the standardized *Rsb* values are normally distributed " Panels B and C in S1 Fig" a one-tailed Z-test was applied to identify statistically significant SNPs under selection in Sheko cattle (positive *Rsb* value). One-sided *P*-values were derived as–$\log_{10}(1- \Phi(Rsb))$, where $\Phi(Rsb)$ represents the Gaussian cumulative distribution function. In both *iHS* and *Rsb*, -$\log_{10}$ (*P*-value) = 3, equivalent to a *P*-value of 0.001, was used as a threshold to define significant *iHS* and *Rsb* values. A candidate region was defined if at least two SNPs not separated by more than 500 Kb passed the significant threshold as followed by [16], which is the extent of linkage disequilibrium determined in the genomes of different taurine and indicine breeds [34].

### Functional characterization of the candidate regions

Genes mapped on the UMD3.1 reference bovine genome within substantial excess/deficiency Asian zebu ancestry regions and candidate regions with signatures of positive selection were retrieved from the *Ensemble Genes 86* database [35]. Bovine Quantitative Trait Loci (QTL) mapped on the UMD3.1 reference genome (http://www.animalgenome.org/cgi-bin/QTLdb/BT/index) intersecting with the Sheko candidate regions were also identified.

### Genetic differentiation $F_{st}$ analysis

Genetic differentiation analysis was conducted between the Asian zebu Nelore and the African taurine N'Dama breeds using Weir and Cockerham's $F_{st}$ estimator [36] calculated by the hierfstat package [37] for R software. $F_{st}$ values were estimated for each genotyped SNP and averaged over 1 Mb sliding windows overlapping by 10 kb, in which windows with a single SNP were excluded. Genomic windows in the top 1% tail of the windows $F_{st}$ values distribution were considered as differentiated windows for further analyses. Overlapping windows were merged into candidate genomic regions.

## Results

### Asian zebu ancestry deviation on Sheko cattle genome

The locus-ancestry deviation analysis on the sliding 1 Mb genomic windows indicates a mean Asian zebu and African taurine ancestry proportions of 0.56 ± 0.18 and 0.44 ± 0.18, respectively. Out of the total 2,314 genomic windows, 15 of these distributed across 12 autosomes show substantial excess in Asian zebu ancestry. Whilst, 72 genomic windows distributed across 24 autosomes show substantial deficiency in Asian zebu ancestry "Fig 1 and S1 Table".

### Candidate iHS and Rsb regions on Sheko cattle

The intra-population *iHS* analysis reveals nine candidate regions with signatures of positive selection across six autosomes, one on BTA 2, 7 and 8, and two on BTA 3, 4 and 5 "Fig 2A". The inter-population *Rsb* analyses of Sheko with the N'Dama and Nelore cattle reveal 22 candidate regions with signatures of positive selection on 11 autosomes for the N'Dama comparison (one on BTA 3, 6, 14, 18 and 24; two on BTA 5, 11 and 12; three on BTA 2; four on BTA 7 and 13), and a single candidate region for the Nelore comparison on BTA 24 "Fig 2B and 2C and S2 Table".
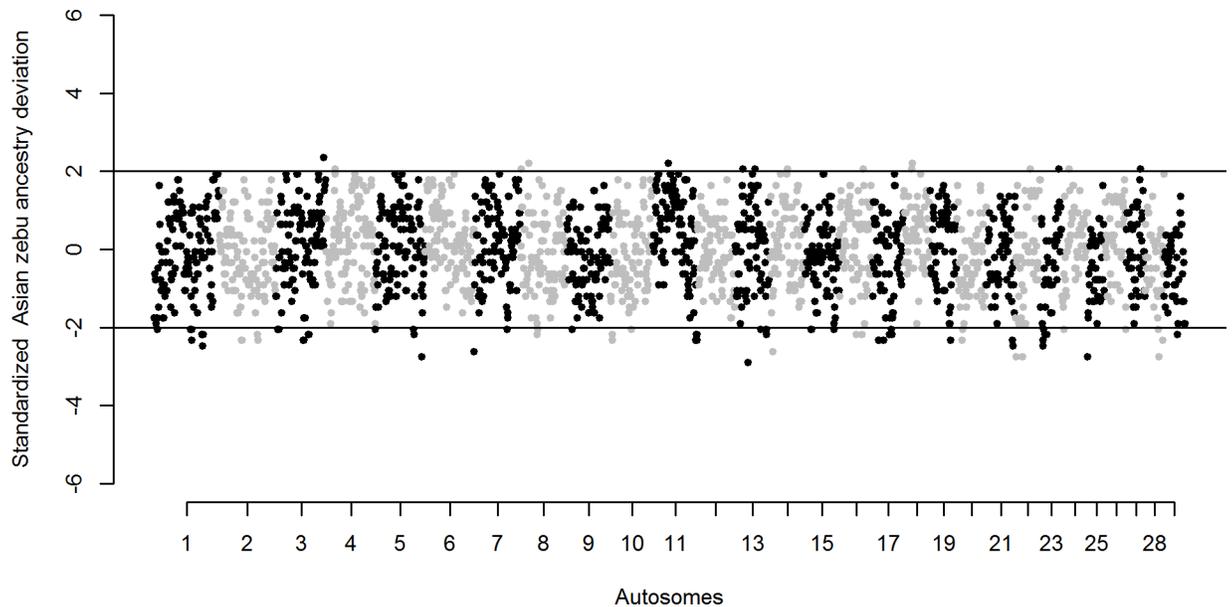
**Fig 1. Manhattan plot of standardized Asian zebu ancestry deviation on Sheko autosomes.** Sheko cattle autosomes plot showing deviation (excess/deficiency) in standardized Asian zebu ancestry in 1Mb sliding windows.

https://doi.org/10.1371/journal.pone.0202479.g001

### Overlap among the EHH-based statistics and the locus-ancestry deviation analysis

A total of four *iHS* candidate regions on BTA 2, 3, 5 and 7 overlap with the N'Dama comparison *Rsb* candidate regions. Whilst the single *Rsb* candidate region with Nelore cattle on BTA 24 overlaps with a genomic window showing substantial deficiency in Asian zebu ancestry (BTA 24: 4.46–5.43 Mb; Table 1).

### Functional annotation of the Sheko candidate regions

A total of 71 genes are found within substantial excess zebu ancestry regions and 721 genes within deficient zebu ancestry regions "S3 Table". The candidate *iHS* signatures of positive selection regions have 57 annotated genes, while 85 genes are present within the N'Dama comparison *Rsb* regions, and two genes in the single Nelore comparsion *Rsb* region "S4 Table". These genes are associated with several biological functions, such as immunity (e.g. *IL7*, *IL15*, *FCN2*, *ICOS*, *LTA4H* and *NFAM1*), fertility (e.g. *MEA1*, *CLGN* and *RXFP2*), heat tolerance (*HSPA6* and *DNAJC6*) and lactation (*PRLH*) "Table 2".

A total of 188 QTL overlap with the excess zebu ancestry regions and 706 with the deficient zebu ancestry regions "S5 Table". Moreover, 124, 284 and eight QTL intersect with *iHS*, N'Dama comparison *Rsb* and Nelore comparison *Rsb* candidate regions, respectively "S5 Table". These QTL are linked with different biological pathways, such as lactation (e.g. milk yield, milk fat percentage and milk protein yield), fertility (e.g. calving ease, gestation length and sperm motility), body composition (e.g. rump angle, foot angle and height) and immunity (e.g. tick resistance). Trypanotolerance-controlling QTL, identified by a cross between tolerant West African N'Dama and susceptible East African Boran cattle by [38], are found within two zebu ancestry excess, five zebu ancestry deficient and four N'Dama comparison *Rsb* regions "S6 Table".

A



B



C
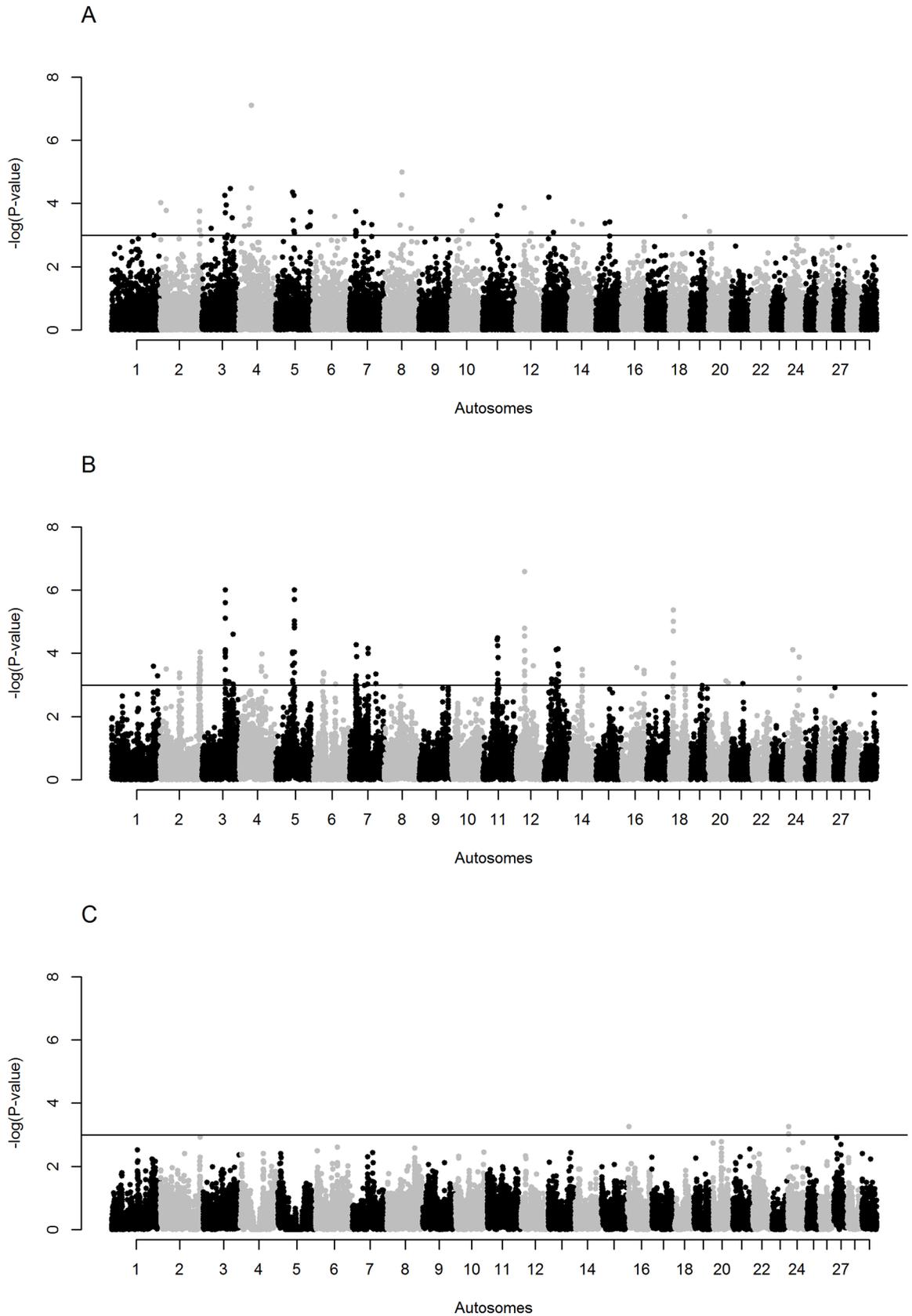
**Fig 2. Manhattan plots of genome-wide signatures of positive selection analyses.** (A) *iHS* analysis (B) *Rsb* analysis with the African taurine N'Dama cattle and (C) *Rsb* analysis with the Asian zebu Nelore cattle. The significance threshold is set at –log$_{10}$ (two-tailed *P*-value for *iHS* analysis) and (one-tailed *P*-value for *Rsb* analysis) = 3.

## Genetic differentiation regions between the Asian zebu and African taurine cattle

The mean *Fst* value of sliding 1 Mb windows between the Nelore and N'Dama cattle breeds is 0.15 ± 0.07. Upon merging genomic windows, the top 1% tail of the *Fst* values distribution contained a total of 57 regions distributed across 21 autosomes. These are considered as highly differentiated genomic regions between these two cattle breeds "Fig 3 and S7 Table".

Two of the *Fst* regions overlap with candidate regions for signatures of selection in Sheko cattle (one *iHS* region and one N'Dama comparison *Rsb* region). Three *Fst* regions overlap with regions showing substantial deficiency in zebu ancestry in the Sheko cattle genome "Table 1".

## Discussion

The genome of Sheko cattle was analyzed, using genome-wide medium density SNP data, to identify candidate genomic regions with signatures of adaptive introgression and positive selection. These regions were defined based on locus-ancestry deviation analysis and two EHH-based statistics (*iHS* and *Rsb*). We inferred the origin of these selection footprints as pre- or post-admixture based on genetic differentiation analysis between the two Sheko ancestral cattle breeds: N'Dama and Nelore.

### Genomic regions with signatures of adaptive introgression and natural selection

The first cattle on the African continent were of the taurine types. Subsequently, the spread of Asian zebu ancestry in the African continent from their center of domestication in the Indian subcontinent has led to various indigenous African cattle breeds with admixed Asian zebu x African taurine genomic structure [3]. The genome of these admixed cattle breeds would have been subjected to selective pressures to maximize the reproductive fitness of the crosses and their adaptability to the environmental challenges.

Adaptive introgression for advantageous zebu characteristics may be expected, while some taurine genomic regions previously selected for local adaptation would have resisted

**Table 1. Chromosomes and positions (in Mb) of overlapping candidate selected regions detected by locus-ancestry deviation, *iHS*, *Rsb* and *Fst* analyses.**

| Zebu deficient genome regions | His Sheko | Rsb Sheko -N'Dama | Rsb Sheko -Nelore | Fst Nelore–N'Dama |
|---|---|---|---|---|
| | 2: 129.61–129.68 | 2: 129.61–130.4 | | |
| | 3:79.18–79.22 | | | 3:78.2–79.5 |
| | 3: 75.83–76.12 | 3: 75.83–76.8 | | |
| | 5: 60.58–61.4 | 5: 60.51–61.4 | | |
| | 7: 20.66–21.6 | 7: 21.55–21.6 | | |
| 8:40.28–41.28 | | | | 8:39.6–40.8 |
| 8:42.44–43.24 | | | | 8:42.5–43.5 |
| | | 11:46.82–47.18 | | 11:46.3–48.7 |
| 23:10.5–11.48 | | | | 23:10.4–11.5 |
| 24: 4.46–5.53 | | | 24: 4.47–4.61 | |

**Table 2. Examples of candidate genes within the candidate regions of the different analyses conducted in the study.** Candidate regions are represented as (BTA: start Mb—stop Mb).

| Biological role | Candidate regions | Analysis | Gene ID |
|---|---|---|---|
| Immunity | 2:91.99–92.950 | Locus-ancestry analysis* | ICOS |
| | 5:60.51–61.4 | Rsb (Sheko-N'Dama) | LTA4H |
| | 5:112.95–113.91 | Locus-ancestry analysis* | NFAM1 |
| | 11:106.24–107.04 | Locus-ancestry analysis* | FCN2 |
| | 14:43.35–44.35 | Locus-ancestry analysis** | IL7 |
| | 17:15.94–16.87 | Locus-ancestry analysis* | IL15 |
| Fertility and reproduction | 12:29.03–29.72 | Rsb (Sheko-N'Dama) | RXFP2 |
| | 17:15.94–16.87 | Locus-ancestry analysis* | CLGN |
| | 23:16.02–17.01 | Locus-ancestry analysis* | MEA1 |
| Heat tolerance | 3:7.74–8.72 | Locus-ancestry analysis* | HSPA6 |
| | 3:80.1–80.93 | Locus-ancestry analysis* | DNAJC6 |
| Lactation | 3:116.99–117.96 | Locus-ancestry analysis** | PRLH |

*Zebu ancestry deficient region

**Zebu ancestry excess region

introgression. In the Sheko cattle, 87 candidate genomic regions showed substantial deviation in Asian zebu ancestry, of which 15 regions showed an excess and 72 showed a deficit of zebu ancestry, indicating candidate signatures of positive selection. Although the genome of Sheko cattle is mainly composed of zebu ancestry [12], about 83% of the candidate ancestry deviation regions showed deviation towards the taurine haplotypes. This supports the likelihood that these regions are of importance for the adaptability of Sheko cattle to the local environment. Interestingly, five of these zebu ancestry deficient regions overlap with five trypanotolerant QTL, while two of these regions with excess of zebu ancestry overlap with two trypanotolerant
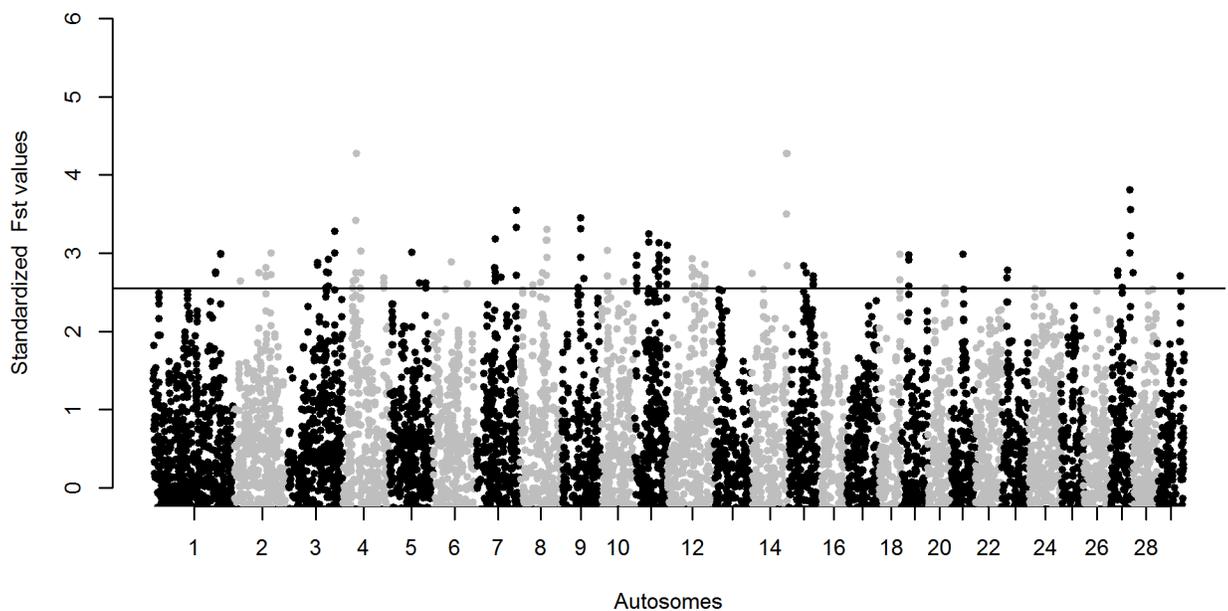


**Fig 3. Manhattan plot of the genome-wide *Fst* analysis between the Asian zebu Nelore cattle and the African taurine N'Dama cattle breeds (1 Mb sliding window).** The significance threshold is set at the top 1% of the *Fst* values distribution tail.

QTL. This is not surprising as it has been shown that both regions of zebu and taurine origin may contribute to the trypanotolerance characteristic of West African N'Dama and East African Boran crossbreeds [38].

Moreover, the two different EHH-based analyses, *iHS* and *Rsb*, identified 32 candidate regions with signatures of positive selection in Sheko cattle (nine regions for *iHS*, 22 regions for *Rsb* Sheko—N'Dama comparison, one region for *Rsb* Sheko—Nelore comparison). The *Rsb* Sheko—N'Dama analysis results support selection pressures on zebu haplotypes, whilst, the *Rsb* Sheko–Nelore analysis indicates that the taurine haplotypes within this region are the target of selection. These results require further investigation and validation using full genome sequence data of Sheko cattle and the ancestral cattle breeds.

## The confounding effect of the natural demographic history and selection

Demographic population processes, such as migration and the associated gene flow and genetic drift, also shape the genome diversity of livestock populations and may lead to similar signals as natural selection at the genome-wide level [39]. This will be the case in pure breeds as well as in admixed populations. Concerning the latter, taurine or zebu haplotypes may have become fixed following random segregation of alleles subsequent to admixture. However, it could be argued that in the absence of selection such taurine or zebu fixed regions in the crossbreed will show sequence diversity. The overlap between four *iHS* and *Rsb* candidate regions, and between a single zebu-deficient region with an *Rsb* candidate region "Table 1" supports the role of selection pressures, and not natural demographic processes, in shaping the genomic pattern of these regions. This low level of overlap between *iHS* and *Rsb* selection and introgression may be a consequence of a lack of power in the analyses performed here. A caveat of the *iHS* analysis is that it will not identify a selected haplotype which has reached or is close to reaching fixation, while the *Rsb* analysis cannot identify signatures of selection for haplotypes that are under selection in both breeds being compared [39]. The methods we have applied target different selection timeframes, with the ancestry deviation approach targeting recent post-admixture selection, while the EHH-based statistics identify much older signals of selection [40], and as such we would not expect to observe significant overlap across the results. Indeed, a study on the admixed Swiss Fleckvieh cattle breed, which is a composite of Simmental and Red Holstein-Friesian cattle breeds, also resulted in little overlap when applying the same approaches [22]. Increasing the sample size and density of the SNP data, for example through whole-genome sequencing, will greatly improve the power of these tests, enabling a more robust investigation into signatures of selection in Sheko cattle. Nonetheless, 24 candidate regions (three zebu excess, 11 zebu deficient, two *iHS* and eight Sheko—N'Dama *Rsb* regions) do overlap with candidate genomic regions under positive selection reported in previous studies on indigenous African cattle breeds such as the East African Shorthorn Zebu [16, 17], and the Butana and Kenana cattle [18], as well as commercial cattle breeds, Murray Grey, Shorthorn and Charolais [20].

## Functional annotation of the candidate regions

Several genes and QTL associated with different biological pathways, e.g. immunity, fertility and reproduction, heat stress, and the dairy production characteristics of Sheko cattle, have been identified within the candidate selected regions. These genes and QTL might be related to the adaptation of Sheko cattle to the local environment and hence can be considered as targets of natural selection in Sheko cattle. These cattle are known to be tolerant to different endemic parasitic diseases [11, 41], and so the immunity-related genes within the candidate regions identified (e.g. *LTA4H*, *IL7*, *IL15*, *FCN*, *LTA4H* and *NFAM1*) are potential targets of natural selection. An immunity-related gene identified in a Sheko—N'Dama *Rsb* candidate

region on BTA 5 is leukotriene A-4 hydrolase (*LTA4H*). This gene is associated with immune response regulation and inflammation response in mammals [42]. *LTA4H* was also been identified within a candidate region of positive selection in East African shorthorn zebu cattle (EASZ) from Kenya [16]. Two interleukin genes (*IL7* and *IL15*) were identified in zebu-excess and zebu-deficient regions on BTA 14 and BTA 17, respectively. Interleukin-7 is an important cytokine involved mainly in the early development of B- and T-cells [43]. Whilst, Interleukin-15 mediates the activation of natural killer cells [44].

Genes related to fertility and reproduction are hotspots of selection in indigenous cattle breeds living in tropical environments. The relaxin/insulin-like family peptide receptor 2 (*RXFP2*) gene is present in a Sheko—N'Dama *Rsb* candidate region on BTA 12. This gene is involved in the testicular descent development [45], and has also been found to be under selection in two different tropical-adapted admixed cattle population; EASZ [16, 17] and Creole cattle [46]. The calmegin (*CLGN*) gene, located in a zebu-deficient candidate region on BTA 17, is a testis-specific $Ca^{+2}$-binding protein involved in mediating the binding between eggs and sperms during fertilization [47]. The male-enhanced antigen-1 (MEA1) gene found within a zebu-deficient candidate region on BTA 23 is expressed mainly in spermatids indicating a possible role in late stages of spermatogenesis [48].

The agro-ecological zone of the sheko is classified as warm and humid to peri-humid, characterized by a mean annual temperature of 22.6°C and annual rainfall from 1200 to 2200 mm [49]. In such an environment tolerance to heat and humidity will be advantageous. *HSPA6* and *DNAJC6* genes are both found within zebu-deficient candidate regions on BTA 3. The heat shock protein family A member 6 (*HSPA6*) is a member of the heat shock protein (Hsp70) family which protect cells from lethal damage caused by heat stress through maintaining the folding of newly synthesized proteins and assembly of multi-protein complexes [50]. DNAJC6 acts as a co-factor for Hsp70 family to mediate their cellular function [51]. Members of these two gene families have also been found previously to be under selection in EASZ [16, 17].

Sheko cattle are considered a breed with good dairy potential. Several dairy production-related QTL (e.g. milk yield, milk fat percentage and milk protein yield) overlap the candidate regions identified, including the prolactin releasing hormone (*PRLH*) gene which overlaps with a zebu ancestry-excess region. A study in African zebu cattle also identified a candidate selection peak at the prolactin releasing hormone (*PRLH*) gene [19]. In addition, it has been shown that mutation at prolactin (*PRL*) and its receptor (*PRLR*) genes have an impact on thermoregulation and hair morphology [52]. The prolactin pathway might therefore have been selected in Sheko cattle both in relation to milk production and heat tolerance.

### Origin of selection: Pre- or post-admixture?

Highly differentiated genomic regions between the ancestral populations of an admixed population may indicate ancient signatures of selection prior to admixture [22, 24]. We found overlaps between three zebu ancestry deficient regions and a Sheko—N'Dama *Rsb* candidate region with highly differentiated regions between Nelore and N'Dama cattle. While the former suggest signals of ancient selection within African taurine prior to admixture, the later suggest an ancient zebu selected region. However, these results require further validation using a higher density genome-wide SNP chip, such as the Illumina BovineHD Genotyping BeadChip, and/or full genome sequence data.

### Conclusion

In this study we employed genome-wide medium density SNP data to investigate the genome of Sheko cattle for regions with signatures of adaptive introgression and positive selection.

Several candidate regions were identified showing excess and deficiency in zebu ancestry and unusual extended haplotype homozygosity. These regions are associated with different biological traits such as immunity, reproduction, heat tolerance and lactation. Some of these selection signals are likely to be a result of ancient selection prior to the admixture of the ancestral African taurine and Asian zebu breeds. Our findings contribute towards improving our understanding the genome of the Sheko cattle breed, and can inform breeding programmes to enhance the productivity and sustainability of the indigenous African cattle in their native environment. However, further validation and investigation using a larger sample size and high-resolution data, such as that from a high-density SNP array or full genome sequence data, is required to better characterize the favorable haplotypes or variants under selection.

## Supporting information

**S1 Fig.** Histograms showing the distribution of the (A) standardized *iHS* values, (B) standardized Sheko—N'Dama *Rsb* values and (C) standardized Sheko—Nelore *Rsb* values. (TIFF)

**S1 Table. Genomic windows on Sheko cattle showing substantial excess and deficiency in Asian zebu ancestry and overlapping with signatures of selection in other cattle breeds from previous studies.** (XLSX)

**S2 Table. Candidate signatures of selection regions identified by the *iHS* and *Rsb* analyses and overlapping with signatures of selection in other cattle breeds from previous studies.** (XLSX)

**S3 Table. Genes within the substantial excess and deficient zebu ancestry regions.** (XLSX)

**S4 Table. Genes within the candidate signatures of selection regions identified by *iHS* and *Rsb* analyses.** (XLSX)

**S5 Table. Quantitative trait Loci (QTL) within the candidate regions defined by the locus-ancestry deviation, *iHS* and *Rsb* analyses.** (XLSX)

**S6 Table. Trypanotolerance-controlling QTL overlapping zebu-ancestry excess and deficient regions and Sheko-N'Dama Rsb candidate regions.** (XLSX)

**S7 Table. Candidate genetically differentiated regions between Nelore and N'Dama cattle on the top 1% of the Fst values distribution.** (XLSX)

## Acknowledgments

## Author Contributions

**Conceptualization:** Hussain Bahbahani.

**Formal analysis:** Hussain Bahbahani.

**Investigation:** Arwa Afana.

**Methodology:** Hussain Bahbahani, David Wragg.

**Project administration:** Hussain Bahbahani.

**Software:** David Wragg.

**Supervision:** Hussain Bahbahani.

**Writing – original draft:** Hussain Bahbahani.

**Writing – review & editing:** Hussain Bahbahani, David Wragg.

## References

1. Gifford-Gonzalez D, Hanotte O. Domesticating Animals in Africa: Implications of Genetic and Archaeological Findings. Journal of World Prehistory 2011; 24:1–23.

2. Chen S, Lin BZ, Baig M, Mitra B, Lopes RJ, Santos AM, et al. Zebu cattle are an exclusive legacy of the South Asia neolithic. Molecular biology and evolution. 2010; 27(1):1–6. Epub 2009/09/23. https://doi.org/10.1093/molbev/msp213 PMID: 19770222.

3. Hanotte O, Bradley DG, Ochieng JW, Verjee Y, Hill EW, Rege JE. African pastoralism: genetic imprints of origins and migrations. Science. 2002; 296(5566):336–9. Epub 2002/04/16. https://doi.org/10.1126/science.1069878 PMID: 11951043.

4. Bonfiglio S, Ginja C, De Gaetano A, Achilli A, Olivieri A, Colli L, et al. Origin and spread of Bos taurus: new clues from mitochondrial genomes belonging to haplogroup T1. PloS one. 2012; 7(6):e38601. Epub 2012/06/12. https://doi.org/10.1371/journal.pone.0038601 PMID: 22685589; PubMed Central PMCID: PMC3369859.

5. Salim B, Taha KM, Hanotte O, Mwacharo JM. Historical demographic profiles and genetic variation of the East African Butana and Kenana indigenous dairy zebu cattle. Animal genetics. 2014; 45(6):782–90. Epub 2014/10/14. https://doi.org/10.1111/age.12225 PMID: 25308478.

6. Hanotte O, Tawah CL, Bradley DG, Okomo M, Verjee Y, Ochieng J, 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. Molecular ecology. 2000; 9(4):387–96. Epub 2000/03/29. PMID: 10736042.

7. Decker JE, McKay SD, Rolf MM, Kim J, Molina Alcala A, Sonstegard TS, et al. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. PLoS genetics. 2014; 10(3):e1004254. Epub 2014/03/29. https://doi.org/10.1371/journal.pgen.1004254 PMID: 24675901; PubMed Central PMCID: PMC3967955.

8. Rege JEO. The state of African cattle genetic resources I. Classification framework and identification of threatened and extinct breeds. Animal Genetic Resources Information. 1999; 25:1–25.

9. Felius M. Cattle breeds: an encyclopedia: Misset; 1995.

10. DAGRIS. Domestic Animal Genetic Resources Information System (DAGRIS). Addis Ababa, Ethiopia: International Livestock Research Institute; 2017.

11. Mekuriaw G, Kebede A. A review on indigenous cattle genetic resources in Ethiopia: adaptation, status and survival. Online J Anim Feed Res Vet Sci. 2015; 5(5):125–37.

12. Mbole-Kariuki MN, Sonstegard T, Orth A, Thumbi SM, Bronsvoort BM, Kiara H, et al. Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya. Heredity. 2014; 113(4):297–305. Epub 2014/04/17. https://doi.org/10.1038/hdy.2014.31 PMID: 24736786.

13. Lemecha H, Mulatu W, Hussein I, Rege E, Tekle T, Abdicho S, et al. Response of four indigenous cattle breeds to natural tsetse and trypanosomosis challenge in the Ghibe valley of Ethiopia. Veterinary parasitology. 2006; 141(1–2):165–76. https://doi.org/10.1016/j.vetpar.2006.04.035 PMID: 16797843

14. Bahbahani H, Hanotte O. Genetic resistance: tolerance to vector-borne diseases, prospect and challenges of genomics. OIE Scientific and Technical Review. 2015; 34(1):185–97.

15. Bayou E, Haile A, Gizaw S, Mekasha Y. Evaluation of non-genetic factors affecting calf growth, reproductive performance and milk yield of traditionally managed Sheko cattle in southwest Ethiopia. SpringerPlus. 2015; 4(1):568. https://doi.org/10.1186/s40064-015-1340-9 PMID: 26543703

16. Bahbahani H, Clifford H, Wragg D, Mbole-Kariuki MN, Van Tassell C, Sonstegard T, et al. Signatures of positive selection in East African Shorthorn Zebu: A genome-wide single nucleotide polymorphism

analysis. Sci Rep. 2015; 5:11729. Epub 2015/07/02. https://doi.org/10.1038/srep11729 PMID: 26130263; PubMed Central PMCID: PMCPmc4486961.

17. Bahbahani H, Tiijani A, Mukasa C, Wragg D, Almathen F, Nash O, et al. Signature of selection for environmental adaptation and zebu x taurine hybrid fitness in East African Shorthorn Zebu. Frontiers in genetics. 2017; 8:1–20. https://doi.org/10.3389/fgene.2017.00001

18. Bahbahani H, Salim B, Almathen F, Al Enezi F, Mwacharo JM, Hanotte O. Signatures of positive selection in African Butana and Kenana dairy zebu cattle. PloS one. 2018; 13(1):e0190446. https://doi.org/10.1371/journal.pone.0190446 PMID: 29300786

19. Kim J, Hanotte O, Mwai OA, Dessie T, Bashir S, Diallo B, et al. The genome landscape of indigenous African cattle. Genome biology. 2017; 18(1):34. https://doi.org/10.1186/s13059-017-1153-y PMID: 28219390

20. Kemper KE, Saxton SJ, Bolormaa S, Hayes BJ, Goddard ME. Selection for complex traits leaves little or no classic signatures of selection. BMC genomics. 2014; 15(1):246. Epub 2014/04/01. https://doi.org/10.1186/1471-2164-15-246 PMID: 24678841; PubMed Central PMCID: PMC3986643.

21. Qanbari S, Pausch H, Jansen S, Somel M, Strom TM, Fries R, et al. Classic selective sweeps revealed by massive sequencing in cattle. PLoS genetics. 2014; 10(2):e1004148. Epub 2014/03/04. https://doi.org/10.1371/journal.pgen.1004148 PMID: 24586189; PubMed Central PMCID: PMC3937232.

22. Khayatzadeh N, Mészáros G, Utsunomiya YT, Garcia JF, Schnyder U, Gredler B, et al. Locus-specific ancestry to detect recent response to selection in admixed Swiss Fleckvieh cattle. Animal genetics. 2016; 47(6):637–46. https://doi.org/10.1111/age.12470 PMID: 27435758

23. Tang H, Choudhry S, Mei R, Morgan M, Rodriguez-Cintron W, Burchard EG, et al. Recent genetic selection in the ancestral admixture of Puerto Ricans. American journal of human genetics. 2007; 81 (3):626–33. Epub 2007/08/19. https://doi.org/10.1086/520769 PMID: 17701908; PubMed Central PMCID: PMC1950843.

24. Jin W, Xu S, Wang H, Yu Y, Shen Y, Wu B, et al. Genome-wide detection of natural selection in African Americans pre- and post-admixture. Genome research. 2012; 22(3):519–27. Epub 2011/12/01. https://doi.org/10.1101/gr.124784.111 PMID: 22128132; PubMed Central PMCID: PMCPMC3290787.

25. Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, et al. Development and characterization of a high density SNP genotyping assay for cattle. PloS one. 2009; 4(4):e5350. Epub 2009/04/25. https://doi.org/10.1371/journal.pone.0005350 PMID: 19390634; PubMed Central PMCID: PMC2669730.

26. Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversole KA, Gill CA, et al. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science. 2009; 324(5926):528–32. Epub 2009/04/25. https://doi.org/10.1126/science.1167936 PMID: 19390050; PubMed Central PMCID: PMC2735092.

27. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 2007; 23(10):1294–6. Epub 2007/03/27. https://doi.org/10.1093/bioinformatics/btm108 PMID: 17384015.

28. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria 2012.

29. Brisbin A, Bryc K, Byrnes J, Zakharia F, Omberg L, Degenhardt J, et al. PCAdmix: Principal Components-Based Assignment of Ancestry along Each Chromosome in Individuals with Admixed Ancestry from Two or More Populations. Human biology. 2012; 84(4):343–64. https://doi.org/10.3378/027.084.0401 PubMed PMID: PMC3740525. PMID: 23249312

30. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. American journal of human genetics. 2006; 78(4):629–44. Epub 2006/03/15. https://doi.org/10.1086/502802 PMID: 16532393; PubMed Central PMCID: PMC1424677.

31. Voight B, Kudaravalli S, Wen X, Pritchard J. A map of recent positive selection in the human genome. PLoS biology. 2006; 4(3):e72. https://doi.org/10.1371/journal.pbio.0040072 PMID: 16494531

32. Tang K, Thornton KR, Stoneking M. A new approach for using genome scans to detect recent positive selection in the human genome. PLoS biology. 2007; 5(7):e171. Epub 2007/06/21. https://doi.org/10.1371/journal.pbio.0050171 PMID: 17579516; PubMed Central PMCID: PMC1892573.

33. Gautier M, Vitalis R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics. 2012; 28(8):1176–7. Epub 2012/03/10. https://doi.org/10.1093/bioinformatics/bts115 PMID: 22402612.

34. McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W, et al. Whole genome linkage disequilibrium maps in cattle. BMC genetics. 2007; 8:74. Epub 2007/10/27. https://doi.org/10.1186/1471-2156-8-74 PMID: 17961247; PubMed Central PMCID: PMC2174945.

**35.** Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, et al. Ensembl 2013. Nucleic acids research. 2013; 41(Database issue):D48–55. Epub 2012/12/04. https://doi.org/10.1093/nar/gks1236 PMID: 23203987; PubMed Central PMCID: PMC3531136.

**36.** Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984; 38(6):1358–70. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x PMID: 28563791

**37.** Goudet J. hierfstat, a package for r to compute and test hierarchical F-statistics. Molecular Ecology Notes. 2005; 5(1):184–6. https://doi.org/10.1111/j.1471-8286.2004.00828.x

**38.** Hanotte O, Ronin Y, Agaba M, Nilsson P, Gelhaus A, Horstmann R, et al. Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100(13):7443–8. Epub 2003/06/14. https://doi.org/10.1073/pnas.1232392100 PMID: 12805560; PubMed Central PMCID: PMC164605.

**39.** Qanbari S, Simianer H. Mapping signatures of positive selection in the genome of livestock. Livestock Science. 2014; 166:133–43. https://doi.org/10.1016/j.livsci.2014.05.003

**40.** Oleksyk TK, Smith MW, O'Brien SJ. Genome-wide scans for footprints of natural selection. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2010; 365(1537):185–205. Epub 2009/12/17. https://doi.org/10.1098/rstb.2009.0219 PMID: 20008396; PubMed Central PMCID: PMC2842710.

**41.** Taye T, Ayalew W, Hegde BP. Status of Ethiopian indigenous Sheko cattle breed and the need for participatory breed management plan. Ethiopina journal of Animal Production. 2009; 9(1).

**42.** Thunnissen MM, Nordlund P, Haeggstrom JZ. Crystal structure of human leukotriene A(4) hydrolase, a bifunctional enzyme in inflammation. Nature structural biology. 2001; 8(2):131–5. Epub 2001/02/15. https://doi.org/10.1038/84117 PubMed PMID: 11175901.

**43.** Or R, Abdul-Hai A, Ben-Yehuda A. Reviewing the potential utility of interleukin-7 as a promoter of thymopoiesis and immune recovery. Cytokines, cellular & molecular therapy. 1998; 4(4):287–94. Epub 1999/03/06. PMID: 10068062.

**44.** Sanchez-Correa B, Bergua JM, Pera A, Campos C, Arcos MJ, Banas H, et al. In Vitro Culture with Interleukin-15 Leads to Expression of Activating Receptors and Recovery of Natural Killer Cell Function in Acute Myeloid Leukemia Patients. Frontiers in immunology. 2017; 8:931. Epub 2017/08/22. https://doi.org/10.3389/fimmu.2017.00931 PMID: 28824651; PubMed Central PMCID: PMCPMC5545593.

**45.** Gorlov IP, Kamat A, Bogatcheva NV, Jones E, Lamb DJ, Truong A, et al. Mutations of the GREAT gene cause cryptorchidism. Human molecular genetics. 2002; 11(19):2309–18. Epub 2002/09/10. PMID: 12217959.

**46.** Gautier M, Naves M. Footprints of selection in the ancestral admixture of a New World Creole cattle breed. Molecular ecology. 2011; 20(15):3128–43. Epub 2011/06/22. https://doi.org/10.1111/j.1365-294X.2011.05163.x PMID: 21689193.

**47.** Ikawa M, Nakanishi T, Yamada S, Wada I, Kominami K, Tanaka H, et al. Calmegin is required for fertilin alpha/beta heterodimerization and sperm fertility. Developmental biology. 2001; 240(1):254–61. Epub 2002/01/11. https://doi.org/10.1006/dbio.2001.0462 PMID: 11784061.

**48.** Ohinata Y, Sutou S, Kondo M, Takahashi T, Mitsui Y. Male-enhanced antigen-1 gene flanked by two overlapping genes is expressed in late spermatogenesis. Biology of reproduction. 2002; 67(6):1824–31. Epub 2002/11/22. PMID: 12444059.

**49.** Bayou E, Haile A, Gizaw S, Mekasha Y. Characterizing husbandry practices and breeding objectives of Sheko cattle owners for designing conservation and improvement strategies in Ethiopia. Livestock Research for Rural Development. 2014; 26.

**50.** Radons J. The human HSP70 family of chaperones: where do we stand? Cell stress & chaperones. 2016; 21(3):379–404. Epub 2016/02/13. https://doi.org/10.1007/s12192-016-0676-6 PMID: 26865365; PubMed Central PMCID: PMCPMC4837186.

**51.** Kampinga HH, Craig EA. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nature reviews Molecular cell biology. 2010; 11(8):579–92. Epub 2010/07/24. https://doi.org/10.1038/nrm2941 PMID: 20651708; PubMed Central PMCID: PMC3003299.

**52.** Littlejohn MD, Henty KM, Tiplady K, Johnson T, Harland C, Lopdell T, et al. Functionally reciprocal mutations of the prolactin signalling pathway define hairy and slick cattle. Nat Commun. 2014; 5:5861. Epub 2014/12/19. https://doi.org/10.1038/ncomms6861 PMID: 25519203; PubMed Central PMCID: PMCPMC4284646.